# SOME FUNDAMENTAL THEOREMS IN MATHEMATICS

OLIVER KNILL

Abstract. An expository hitchhikers guide to some theorems in mathematics.

Criteria for the current list of 135 theorems are whether the result can be formulated elegantly, whether it is beautiful or useful and whether it could serve as a guide [5] without leading to panic. The order is not a ranking but more like a time-line when things were written down. Since [280] stated "a mathematical theorem only becomes beautiful if presented as a crown jewel within a context" we try sometimes to give some context. Of course, any such list of theorems is a matter of personal preferences, taste and limitations. The number of theorems is arbitrary, the initial obvious goal was 42 but that number got eventually surpassed as it is hard to stop, once started. As a compensation, there are 42 "tweetable" theorems with proofs included. More comments on the choice of the theorems is included in an epilogue. For literature on general mathematics, see [110, 106, 20, 137, 312, 215, 78], for history [124, 316, 193, 40, 30, 116, 195, 188, 347, 63, 311, 43, 140, 174], for popular, beautiful or elegant things [8, 267, 113, 101, 13, 338, 339, 28, 107, 133, 229, 309, 113, 1, 70, 85, 71, 255]. For comprehensive overviews in large parts of mathematics, [41, 93, 94, 32, 297] or predictions on developments [31]. For reflections about mathematics in general [84, 233, 29, 163, 227, 56, 284]. Encyclopedic source examples are [105, 352, 336, 57, 109, 88, 126, 108, 61, 319].

## 1. Arithmetic

Let $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ be the set of **natural numbers**. A number $p \in \mathbb{N}, p > 1$ is **prime** if $p$ has no factors different from 1 and $p$. With a **prime factorization** $n = p_1 \dots p_n$, we understand the prime factors $p_j$ of $n$ to be ordered as $p_i \leq p_{i+1}$. The **fundamental theorem of arithmetic** is

**Theorem:** Every $n \in \mathbb{N}, n > 1$ has a unique prime factorization.

Euclid anticipated the result. Carl Friedrich Gauss gave in 1798 the first proof in his monograph "Disquisitiones Arithmeticae". Within abstract algebra, the result is the statement that the ring of integers $\mathbb{Z}$ is a **unique factorization domain**. For a literature source, see [182]. For more general number theory literature, see [167, 64].

## 2. Geometry

Given an **inner product space** $(V, \cdot)$ with **dot product** $v \cdot w$ leading to **length** $|v| = \sqrt{v.v}$, three non-zero vectors $v, w, v-w$ define a **right angle triangle** if $v$ and $w$ are **perpendicular** meaning that $v \cdot w = 0$. If $a = |v|, b = |w|, c = |v-w|$ are the lengths of the three vectors, then the **Pythagoras theorem** is

**Theorem:** $a^2 + b^2 = c^2$.

*Date*: 7/22/2018, last update 8/21/2018.

Anticipated by Babylonians Mathematicians in examples, it appeared independently also in Chinese mathematics [317] and was proven first by Pythagoras. It is used in many parts of mathematics like in the **Perseval equality** of Fourier theory. See [270, 231, 188].

## 3. Calculus

Let $f$ be a function of one variables which is **continuously differentiable**, meaning that the limit $g(x) = \lim_{h \to 0}[f(x+h) - f(x)]/h$ exists at every point $x$ and defines a continuous function $g$. For any such function $f$, we can form the **integral** $\int_a^b f(t)\, dt$ and the **derivative** $d/dx f(x) = f'(x)$.

**Theorem:** $\int_a^b f'(x)dx = f(b) - f(a), \quad \frac{d}{dx}\int_0^x f(t)dt = f(x)$

Newton and Leibniz discovered the result independently, Gregory wrote down the first proof in his "Geometriae Pars Universalis" of 1668. The result generalizes to higher dimensions in the form of the **Green-Stokes-Gauss-Ostogradski theorem**. For history, see [187]. [112] tells the "tongue in the cheek" proof: as the derivative is a limit of **quotient** of **differences**, the anti-derivative must be a limit of **sums** of **products**. For history, see [111]

## 4. Algebra

A **polynomial** is a complex valued function of the form $f(x) = a_0 + a_1 x + \cdots + a_n x^n$, where the entries $a_k$ are in the complex plane $\mathbb{C}$. The space of all polynomials is denoted $\mathbb{C}[x]$. The largest non-negative integer $n$ for which $a_n \neq 0$ is called the **degree** of the polynomial. Degree 1 polynomials are **linear**, degree 2 polynomials are called **quadratic** etc. The **fundamental theorem of algebra** is

**Theorem:** Every $f \in \mathbb{C}[x]$ of degree $n$ can be factored into $n$ linear factors.

This result was anticipated in the 17th century, proven first by Carl Friedrich Gauss and finalized in 1920 by Alexander Ostrowski who fixed a topological mistake in Gauss proof. The theorem assures that the field of complex numbers $\mathbb{C}$ is algebraically closed. For history and many proofs see [123].

## 5. Probability

Given a sequence $X_k$ of **independent random variables** on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ which all have the same **cumulative distribution functions** $F_X(t) = \mathrm{P}[X \leq t]$. The **normalized random variable** $\overline{X}$ = is $(X - \mathrm{E}[X])/\sigma[X]$, where $\mathrm{E}[X]$ is the **mean** $\int_\Omega X(\omega) dP(\omega)$ and $\sigma[X] = \mathrm{E}[(X - \mathrm{E}[X])^2]^{1/2}$ is the standard deviation. A sequence of random variables $Z_n \to Z$ **converges in distribution** to $Z$ if $F_{Z_n}(t) \to F_Z(t)$ for all $t$ as $n \to \infty$. If $Z$ is a **Gaussian random variable** with zero mean $\mathrm{E}[Z] = 0$ and standard deviation $\sigma[Z] = 1$, the **central limit theorem** is:

**Theorem:** $\overline{(X_1 + X_2 + \cdots + X_n)} \to Z$ in distribution.

Proven in a special case by Abraham De-Moivre for discrete random variables and then by Constantin Carathéodory and Paul Lévy, the theorem explains the importance and ubiquity of the **Gaussian density function** $e^{-x^2/2}/\sqrt{2\pi}$ defining the **normal distribution**. The Gaussian distribution was first considered by Abraham de Moivre from 1738. See [315, 199].

## 6. Dynamics

Assume $X$ is a **random variable** on a **probability space** $(\Omega, \mathcal{A}, \mathrm{P})$ for which $|X|$ has finite mean $\mathrm{E}[|X|]$. This means $X : \Omega \to \mathbb{R}$ is measurable and $\int_\Omega |X(x)| d\mathrm{P}(x)$ is finite. Let $T$ be an ergodic, measure-preserving transformation from $\Omega$ to $\Omega$. **Measure preserving** means that $T(A) = A$ for all **measurable sets** $A \in \mathcal{A}$. **Ergodic** means that that $T(A) = A$ implies $\mathrm{P}[A] = 0$ or $\mathrm{P}[A] = 1$ for all $A \in \mathcal{A}$. The **ergodic theorem** states, that for an ergodic transformation $T$ on has:

**Theorem:** $[X(x) + X(Tx) + \cdots + X(T^{n-1}(x))]/n \to \mathrm{E}[X]$ for almost all $x$.

This theorem from 1931 is due to George Birkhoff and called **Birkhoff's pointwise ergodic theorem**. It assures that "time averages" are equal to "space averages". A draft of the **von Neumann mean ergodic theorem** which appeared in 1932 by John von Neumann has motivated Birkhoff, but the mean ergodic version is weaker. See [351] for history. A special case is the **law of large numbers**, in which case the random variables $x \to X(T^k(x))$ are independent with equal distribution (IID). The theorem belongs to ergodic theory [151, 83, 298].

## 7. Set theory

A **bijection** is a map from $X$ to $Y$ which is **injective**: $f(x) = f(y) \Rightarrow x = y$ and **surjective**: for every $y \in Y$, there exists $x \in X$ with $f(x) = y$. Two sets $X, Y$ have the **same cardinality**, if there exists a bijection from $X$ to $Y$. Given a set $X$, the **power set** $2^X$ is the set of all subsets of $X$, including the **empty set** and $X$ itself. If $X$ has $n$ elements, the power set has $2^n$ elements. Cantor's theorem is

**Theorem:** For any set $X$, the sets $X$ and $2^X$ have different cardinality.

The result is due to Cantor. Taking for $X$ the natural numbers, then every $Y \in 2^X$ defines a real number $\phi(Y) = \sum_{y \in Y} 2^{-y} \in [0,1]$. As $Y$ and $[0,1]$ have the same cardinality (as **double counting pair cases** like $0.39999999\cdots = 0.400000\ldots$ form a countable set), the set $[0,1]$ is uncountable. There are different types of infinities leading to **countable infinite sets** and **uncountable infinite sets**. For comparing sets, the **Schröder-Bernstein** theorem is important. If there exist injective functions $f : X \to Y$ and $g : Y \to X$, then there exists a bijection $X \to Y$. This result was used by Cantor already. For literature, see [152].

## 8. Statistics

A **probability space** $(\Omega, \mathcal{A}, \mathrm{P})$ consists of a set $\Omega$, a **$\sigma$-algebra** $\mathcal{A}$ and a **probability measure** P. A $\sigma$-algebra is a collection of subset of $\Omega$ which contains the empty set and which is closed under the operations of taking complements, countable unions and countable intersections. The function P on $\mathcal{A}$ takes values in the interval $[0,1]$, satisfies $\mathrm{P}[\Omega] = 1$ and $\mathrm{P}[\bigcup_{A \in S} A] = \sum_{A \in S} \mathrm{P}[A]$ for any finite or countable set $S \subset \mathcal{A}$ of pairwise disjoint sets. The elements in $\mathcal{A}$ are called **events**. Given two events $A, B$ where $B$ satisfies $\mathrm{P}[B] > 0$, one can define the **conditional probability** $\mathrm{P}[A|B] = \mathrm{P}[A \cap B]/\mathrm{P}[B]$. Bayes theorem states:

**Theorem:** $\mathrm{P}[A|B] = \mathrm{P}[B|A]\mathrm{P}[A]/\mathrm{P}[B]$

The setup stated the **Kolmogorov axioms** by Andrey Kolmogorov who wrote in 1933 the "Grundbegriffe der Wahrscheinlichkeit" [209] based on measure theory built by Emile Borel and Henry Lebesgue. For history, see [291], who report that "Kolmogorov sat down to write the Grundbegriffe, in a rented cottage on the Klyaz'ma River in November 1932". Bayes theorem is more like a fantastically clever definition and not really a theorem. There is nothing to prove as multiplying with $\mathrm{P}[B]$ gives $\mathrm{P}[A \cap B]$ on both sides. It essentially restates that $A \cap B = B \cap A$, the Abelian property of the product in the ring $\mathcal{A}$. More general is the statement that if $A_1, \dots, A_n$ is a disjoint set of events whose union is $\Omega$, then $\mathrm{P}[A_i|B] = \mathrm{P}[B|A_i]\mathrm{P}[A_i]/(\sum_j \mathrm{P}[B|A_j]\mathrm{P}[A_j]$. Bayes theorem was first proven in 1763 by Thomas Bayes. It is by some considered to the theory of probability what the Pythagoras theorem is to geometry. If one measures the ratio applicability over the difficulty of proof, then this theorem even beats Pythagoras, as no proof is required. Similarly as "a+(b+c)=(a+b)+c", also Bayes theorem is essentially a definition but less intuitive as "Monty Hall" illustrates [279]. See [199].

## 9. Graph theory

A **finite simple graph** $G = (V, E)$ is a finite collection $V$ of **vertices** connected by a finite collection $E$ of **edges**, which are un-ordered pairs $(a, b)$ with $a, b \in V$. **Simple** means that no **self-loops** nor **multiple connections** are present in the graph. The **vertex degree** $d(x)$ of $x \in V$ is the number of edges containing $x$.

**Theorem:** $\sum_{x \in V} d(x)/2 = |E|$.

This formula is also called the **Euler handshake formula** because every edge in a graph contributes exactly two handshakes. It can be seen as a **Gauss-Bonnet formula** for the **valuation** $G \to v_1(G)$ counting the number of edges in $G$. A **valuation** $\phi$ is a function defined on **subgraphs** with the property that $\phi(A \cup B) = \phi(A) + \phi(B) - \phi(A \cap B)$. Examples of valuations are the number $v_k(G)$ of **complete sub-graphs** of dimension $k$ of $G$. An other example is the **Euler characteristic** $\chi(G) = v_0(G) - v_1(G) + v_2(G) - v_3(G) + \cdots + (-1)^d v_d(G)$. If we write $d_k(x) = v_k(S(x))$, where $S(x)$ is the unit sphere of $x$, then $\sum_{x \in V} d_k(x)/(k+1) = v_k(G)$ is the **generalized handshake formula**, the Gauss-Bonnet result for $v_k$. The Euler characteristic then satisfies $\sum_{x \in V} K(x) = \chi(G)$, where $K(x) = \sum_{k=0}^{\infty} (-1)^k v_k(S(x))/(k+1)$. This is the **discrete Gauss-Bonnet result**. The handshake result was found by Euler. For more about graph theory, [37, 241, 24, 144] about Euler: [122].

## 10. Polyhedra

A **finite simple graph** $G = (V, E)$ is given by a finite vertex set $V$ and edge set $E$. A subset $W$ of $V$ **generates** the sub-graph $(W, \{\{a, b\} \in E \mid a, b \in W\})$. The **unit sphere** of $v \in V$ is the sub graph generated by $S(x) = \{y \in V \mid \{x, v\} \in E\}$. The **empty graph** $0 = (\emptyset, \emptyset)$ is called the $(-1)$**-sphere**. The 1-point graph $1 = (\{1\}, \emptyset) = K_1$ is the smallest contractible graph. Inductively, a graph $G$ is called **contractible**, if it is either 1 or if there exists $x \in V$ such that both $G - x$ and $S(x)$ are contractible. Inductively, a graph $G$ is called a $d$**-sphere**, if it is either 0 or if every $S(x)$ is a $(d-1)$-sphere and if there exists a vertex $x$ such that $G - x$ is contractible. Let $v_k$ denote the number of complete sub-graphs $K_{k+1}$ of $G$. The vector $(v_0, v_1, \dots)$ is the $f$**-vector** of $G$ and $\chi(G) = v_0 - v_1 + v_2 - \dots$ is the **Euler characteristic** of $G$. The generalized **Euler gem** formula due to Schläfli is:

**Theorem:** For $d = 2$, $\chi(G) = v - e + f = 2$. For $d$-spheres, $\chi(G) = 1 + (-1)^d$.

Convex Polytopes were studied already in ancient Greece. The Euler characteristic relations were discovered in dimension 2 by Descartes [3] and interpreted topologically by Euler who proved the case $d = 2$. This is written as $v - e + f = 2$, where $v = v_0, e = v_1, f = v_2$. The two dimensional case can be stated for **planar graphs**, where one has a clear notion of what the two dimensional cells are and can use the topology of the ambient sphere in which the graph is embedded. Historically there had been confusions [73, 275] about the definitions. It was Ludwig Schläfli [289] who covered the higher dimensional case. The above set-up is a modern reformulation of his set-up, due essentially to Alexander Evako. Multiple refutations [217] can be blamed to ambiguous definitions. Polytopes are often defined through convexity [146, 350] and there is not much consensus on a general definition [145], which was the reason in this entry to formula Schläfli's theorem using here a maybe a bit restrictive (as all cells are simplices), but clear combinatorial definition of what a "sphere" is.

## 11. Topology

The **Zorn lemma** assures that that the Cartesian product of a non-empty family of non-empty sets is non-empty. The **Zorn lemma** is equivalent to the **axiom of choice** in the **ZFC axiom system** and to the **Tychonov theorem** in topology as below. Let $X = \prod_{i \in I} X_i$ denote the **product** of topological spaces. The **product topology** is the **weakest topology** on $X$ which renders all **projection functions** $\pi_i : X \to X_i$ continuous.

**Theorem:** If all $X_i$ are compact, then $\prod_{i \in I} X_i$ is compact.

**Zorn's lemma** is due to Kazimierz Kuratowski in 1922 and Max August Zorn in 1935. Andrey Nikolayevich Tykhonov proved his theorem in 1930. One application of the Zorn lemma is the **Hahn-Banach theorem** in functional analysis, the existence of **spanning trees** in infinite graphs or to the fact that commutative rings with units have **maximal ideals**. For literature, see [175].

## 12. Algebraic geometry

The **algebraic set** $V(J)$ of an **ideal** $J$ in the commutative ring $R = k[x_1, \dots, x_n]$ over an **algebraically closed field** $k$ defines the ideal $I(V(J))$ containing all polynomials that vanish on $V(J)$. The **radical** $\sqrt{J}$ of an ideal $J$ is the set of polynomials in $R$ such that $r^n \in J$ for some positive $n$. (An **ideal** $J$ in a ring $R$ is a subgroup of the additive group of $R$ such that $rx \in I$ for all $r \in R$ and all $x \in I$. It defines the quotient ring $R/I$ and is so the kernel of a ring homomorphism from $R$ to $R/I$. The algebraic set $V(J) = \{x \in k^n \mid f(x) = 0, \forall f \in J\}$ of an ideal $J$ in the polynomial ring $R$ is the set of common roots of all these functions $f$. The algebraic sets are the closed sets in the **Zariski topology** of $R$. The ring $R/I(V)$ is the **coordinate ring** of the algebraic set $V$.) The **Hilbert Nullstellensatz** is

**Theorem:** $I(V(J)) = \sqrt{J}$.

The theorem is due to Hilbert. A simple example is when $J = \langle p \rangle = \langle x^2 - 2xy + y^2 \rangle$ is the ideal $J$ generated by $p$ in $\mathbb{R}[x, y]$; then $V(J) = \{x = y\}$ and $I(V(J))$ is the ideal generated by $x - y$. For literature, see [158].

## 13. Cryptology

An integer $p > 1$ is **prime** if 1 and $p$ are the only factors of $p$. The number $k$ mod $p$ is the **reminder** when dividing $k$ by $p$. **Fermat's little theorem** is

**Theorem:** $a^p = a$ mod $p$ for every prime $p$ and every integer $a$.

The theorem was found by Pierre de Fermat in 1640. A first proof appeared in 1683 by Leibniz. Euler in 1736 published the first proof. The result is used in the **Diffie-Helleman key exchange**, where a large public prime $p$ and a public base value $a$ are taken. Ana chooses a number $x$ and publishes $X = a^x \text{mod} p$ and Bob picks $y$ publishing $Y = a^y \text{mod} p$. Their secret key is $K = X^y = Y^x$. An adversary Eve who only knows $a, p, X$ and $Y$ can from this not get $K$ due to the difficulty of the **discrete log problem**. More generally, for possibly composite numbers $n$, the theorem extends to the fact that $a^{\phi(n)} = 1$ modulo $p$, where the **Euler's totient function** $\phi(n)$ counts the number of positive integers less than $n$ which are **coprime** to $n$. The generalization is key the RSA **crypto systems**: in order for Ana and Bob to communicate. Bob publishes the product $n = pq$ of two large primes as well as some base integer $a$. Neither Ana nor any third party Eve do know the factorization. Ana communicates a message $x$ to Bob by sending $X = a^x \text{mod} n$ using **modular exponentiation**. Bob, who knows $p, q$, can find $y$ such that $xy = 1$ mod $\phi(n)$. This is because of Fermat $a^{(p-1)(q-1)} = a$ mod $n$. Now, he can compute $x = y^{-1} \text{mod } \phi(n)$. Not even Ana herself could recover $x$ from $X$.

## 14. Analysis

A bounded linear operator $A$ on a **Hilbert space** is called **normal** if $AA^* = A^*A$, where $A^* = \overline{A}^T$ is the **adjoint** and $A^T$ is the **transpose** and $\overline{A}$ is the **complex conjugate**. Examples of normal operators are **self-adjoint** operators (meaning $A = A^*$) or **unitary operators** (meaning $AA^* = 1$).

**Theorem:** $A$ is normal if and only if $A$ is unitarily diagonalizable.

In finite dimensions, any unitary $U$ diagonalizing $A$ using $B = U^*AU$ contains an **orthonormal eigenbasis** of $A$ as column vectors. The theorem is due to Hilbert. In the self-adjoint case, all the eigenvalues are real and in the unitary case, all eigenvalues are on the unit circle. The result allows a **functional calculus** for normal operators: for any continuous function $f$ and any bounded linear operator $A$, one can define $f(A) = Uf(B)U^*$, if $B = U^*AU$. See [77].

## 15. Number systems

A **monoid** is a set $X$ equipped with an **associative operation** $*$ and an **identity element** 1 satisfying $1 * x = x$ for all $x \in X$. **Associativity** means $x * (y * z) = (x * y) * z$ for all $x, y, z \in X$. The monoid structure belongs to a collection of mathematical structures **magmas** $\supset$ **semigroups** $\supset$ **monoids** $\supset$ **groups**. A monoid is **commutative**, if $x * y = y * x$ for all $x, y \in X$. A **group** is a monoid in which every element $x$ has an **inverse** $y$ satisfying $x * y = y * x = 1$.

**Theorem:** Every commutative monoid can be extended to a group.

The general result is due to Alexander Grothendieck from around 1957. The group is called the **Grothendieck group completion** of the monoid. For example, the additive monoid of natural numbers can be extended to the group of integers, the multiplicative monoid of non-zero integers can be extended to the group of rational numbers. The construction of the group is used in **K-theory** [19] For insight about the philosophy of Grothendieck's mathematics, see [239].

## 16. Combinatorics

Let $|X|$ denote the **cardinality** of a finite set $X$. This means that $|X|$ is the number of elements in $X$. A function $f$ from a set $X$ to a set $Y$ is called **injective** if $f(x) = f(y)$ implies $x = y$. The **pigeonhole principle** tells:

**Theorem:** If $|X| > |Y|$ then no function $X \to Y$ can be injective.

This implies that if we place $n$ items into $m$ boxes and $n > m$, then one box must contain more than one item. The principle is believed to be formalized first by Peter Dirichlet. Despite its simplicity, the principle has many applications, like proving that something exists. An example is the statement that there are two trees in New York City streets which have the same number of leaves. The reason is that the U.S. Forest services states 592'130 trees in the year 2006 and that a mature, healthy tree has about 200'000 leaves. One can also use it for less trivial statements like that in a cocktail party there are at least two with the same number of friends present at the party. A mathematical application is the **Chinese remainder Theorem** stating that that there exists a solution to $a_i x = b_i \bmod m_i$ all disjoint pairs $m_i, m_j$ and all pairs $a_i, m_i$ are relatively prime [95, 234]. The principle generalizes to infinite set if $|X|$ is the cardinality. It implies then for example that there is no injective function from the real numbers to the integers. For literature, see for example [50], which states also a stronger version which for example allows to show that any sequence of real $n^2 + 1$ real numbers contains either an increasing subsequence of length $n + 1$ or a decreasing subsequence of length $n + 1$.

## 17. Complex analysis

Assume $f$ is an **analytic function** in an **open domain** $G$ of the **complex plane** $\mathbb{C}$. Such a function is also called **holomorphic** in $G$. Holomorphic means that if $f(x + iy) = u(x + iy) + iv(x + iy)$, then the **Cauchy-Riemann** differential equations $u_x = v_y, u_y = -v_x$ hold in $G$. Assume $z$ is in $G$ and assume $C \subset G$ is a **circle** $z + re^{i\theta}$ centered at $z$ which is bounding a disc $D = \{w \in \mathbb{C} \mid |w - z| < r\} \subset G$.

**Theorem:** For analytic $f$ in $G$ and a circle $C \subset G$, one has $f(w) = \int_C \frac{f(z)dz}{(z-w)}$.

This **Cauchy integral formula** of Cauchy is used for other results and estimates. It implies for example the **Cauchy integral theorem** assuring that $\int_C f(z)dz = 0$ for any simple closed curve $C$ in $G$ bounding a simply connected region $D \subset G$. **Morera's theorem** assures that for any domain $G$, if $\int_C f(z)\ dz = 0$ for all simple closed smooth curves $C$ in $G$, then $f$ is holomorphic in $G$. An other generalization is **residue calculus**: For a simply connected region $G$ and a function $f$ which is analytic except in a finite set $A$ of points. If $C$ is piecewise smooth continuous closed curve not intersecting $A$, then $\int_C f(z)\ dz = 2\pi i \sum_{a \in A} I(C, a)\text{Res}(f, a)$, where $I(C, a)$ is the **winding number** of $C$ with respect to $a$ and $\text{Res}(f, a)$ is the **residue** of $f$ at $a$ which is in the case of poles given by $\lim_{z \to a}(z - a)f(z)$. See [60, 7, 76].

## 18. Linear algebra

If $A$ is a $m \times n$ **matrix** with **image** $\mathrm{ran}(A)$ and **kernel** $\ker(A)$. If $V$ is a linear subspace of $\mathbb{R}^m$, then $V^\perp$ denotes the **orthogonal complement** of $V$ in $\mathbb{R}^m$, the linear space of vectors perpendicular to all $x \in V$.

> **Theorem:** $\dim(\ker A) + \dim(\mathrm{ran} A) = n, \dim((\mathrm{ran} A)^\perp) = \dim(\ker A^T)$.

The result is used in **data fitting** for example when understanding the **least square solution** $x = (A^T A)^{-1} A^T b$ of a **system of linear equations** $Ax = b$. It assures that $A^T A$ is invertible if $A$ has a trivial kernel. The result is a bit stronger than the **rank-nullity theorem** $\dim(\mathrm{ran}(A)) + \dim(\ker(A)) = n$ alone and implies that for finite $m \times n$ matrices the **index** $\dim(\ker A) - \dim(\ker A^*)$ is always $n - m$, which is the value for the 0 matrix. For literature, see [313]. The result has an abstract generalization in the form of the group isomorphism theorem for a group homomorphism $f$ stating that $G/\ker(f)$ is isomorphic to $f(G)$. It can also be described using the **singular value decomposition** $A = UDV^T$. The number $r = \mathrm{ran} A$ has as a basis the first $r$ columns of $U$. The number $n - r = \ker A$ has as a basis the last $n - r$ columns of $V$. The number $\mathrm{ran} A^T$ has as a basis the first $r$ columns of $V$. The number $\ker A^T$ has as a basis the last $m - r$ columns of $U$.

## 19. Differential equations

A **differential equation** $\frac{d}{dt} x = f(x)$ and $x(0) = x_0$ in a **Banach space** $(X, || \cdot ||)$ (a normed, complete vector space) defines an **initial value problem**: we look for a solution $x(t)$ satisfying the equation and given initial condition $x(0) = x_0$ and $t \in (-a, a)$ for some $a > 0$. A function $f$ from $\mathbb{R}$ to $X$ is called **Lipschitz**, if there exists a constant $C$ such that for all $x, y \in X$ the inequality $||f(x) - f(y)|| \leq C|x - y|$ holds.

> **Theorem:** If $f$ is Lipschitz, a unique solution of $x' = f(x), x(0) = x_0$ exists.

This result is due to Picard and Lindelöf from 1894. Replacing the Lipschitz condition with continuity still gives an **existence theorem** which is due to Giuseppe Peano in 1886, but uniqueness can fail like for $x' = \sqrt{x}, x(0) = 0$ with solutions $x = 0$ and $x(t) = t^2/4$. The example $x'(t) = x^2(t), x(0) = 1$ with solution $1/(1 - t)$ shows that we can not have solutions for all $t$. The proof is a simple application of the Banach fixed point theorem. For literature, see [72].

## 20. Logic

An **axiom system** $A$ is a collection of formal statements assumed to be true. We assume it to contain the basic **Peano axioms** of arithmetic. An axiom system is **complete**, if every true statement can be proven within the system. The system is **consistent** if one can not prove $1 = 0$ within the system. It is **provably consistent** if one can prove a theorem **"The axiom system $A$ is consistent."** within the system.

> **Theorem:** An axiom system is neither complete nor provably consistent.

The result is due to Kurt Goedel who proved it in 1931. In this thesis, Goedel had proven a completeness theorem of first order predicate logic. The incompleteness theorems of 1931 destroyed the dream of **Hilbert's program** which aimed for a complete and consistent **axiom**

**system** for mathematics. A commonly assumed axiom system is the **Zermelo-Frenkel axiom system** together with the axiom of choice ZFC. Other examples are Quine's **new foundations** NF or Lawvere's **elementary theory of the category of sets** ETCS. For a modern view on Hilbert's program, see [322]. For Goedel's theorem [127, 252]. Hardly any other theorem had so much impact outside of mathematics.

## 21. Representation theory

For a **finite group** or **compact topological group** $G$, one can look at **representations**, group homomorphisms from $G$ to the automorphisms of a **vector space** $V$. A representation of $G$ is **irreducible** if the only $G$-invariant subspaces of $V$ are 0 or $V$. The **direct sum** of of two representations $\phi, \psi$ is defined as $\phi \oplus \psi(g)(v \oplus w) = \phi(g)(v) \oplus \phi(g)(w)$. A representation is **semi simple** if it is a unique direct sum of irreducible finite-dimensional representations:

**Theorem:** Representations of compact topological groups are semi simple.

For representation theory, see [340]. Pioneers in representation theory were Ferdinand Georg Frobenius, Herman Weyl, and Élie Cartan. Examples of compact groups are **finite group**, or **compact Lie groups** (a smooth manifold which is also a group for which the multiplications and inverse operations are smooth) like the **torus group** $T^n$, the orthogonal groups $O(n)$ of all orthogonal $n \times n$ matrices or the **unitary groups** $U(n)$ of all unitary $n \times n$ matrices or the group $\mathrm{Sp}(n)$ of all **symplectic** $n \times n$ matrices. Examples of groups that are not Lie groups are the groups $Z_p$ of $p$**-adic integers**, which are examples of **pro-finite groups**.

## 22. Lie theory

Given a **topological group** $G$, a **Borel measure** $\mu$ on $G$ is called **left invariant** if $\mu(gA) = \mu(A)$ for every $g \in G$ and every measurable set $A \subset G$. A left-invariant measure on $G$ is also called a **Haar measure**. A topological space is called **locally compact**, if every point has a compact neighborhood.

**Theorem:** A locally compact group has a unique Haar measure.

Alfréd Haar showed the existence in 1933 and John von Neumann proved that it is unique. In the compact case, the measure is finite, leading to an inner product and so to **unitary representations**. Locally compact **Abelian** groups $G$ can be understood by their **characters**, continuous group homomorphisms from $G$ to the circle group $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. The set of characters defines a new locally compact group $\hat{G}$, the **dual** of $G$. The multiplication is the pointwise multiplication, the inverse is the complex conjugate and the topology is the one of **uniform convergence** on compact sets. If $G$ is compact, then $\hat{G}$ is discrete, and if $G$ is discrete, then $\hat{G}$ is compact. In order to prove **Pontryagin duality** $\hat{\hat{G}} = G$, one needs a generalized **Fourier transform** $\hat{f}(\chi) = \int_G f(x)\overline{\chi(x)} d\mu(x)$ which uses the Haar measure. The **inverse Fourier transform** gives back $f$ using the **dual Haar measure**. The Haar measure is also used to define the **convolution** $f \star g(x) = \int_G f(x-y)g(y) d\mu(y)$ rendering $L^1(G)$ a **Banach algebra**. The Fourier transform then produces a homomorphism from $L^1(G)$ to $C_0(\hat{G})$ or a unitary transformation from $L^2(G)$ to $L^2(\hat{G})$. For literature, see [67, 332].

## 23. Computability

The class of **general recursive functions** is the smallest class of functions which allows **projection**, **iteration**, **composition** and **minimization**. The class of **Turing computable functions** are the functions which can be implemented by a **Turing machine** possessing finitely many states. Turing introduced this in 1936 [265].

**Theorem:** The generally recursive class is the Turing computable class.

Kurt Goedel and Jacques Herbrand defined the class of general recursive functions around 1933. They were motivated by work of Alonzo Church who then created $\lambda$ **calculus** later in 1936. Alan Turing developed the idea of a **Turing machine** which allows to replace Herbrand-Goedel recursion and $\lambda$ calculus. The **Church thesis** or **Church-Turing thesis** states that everything we can compute is generally recursive. As "whatever we can compute" is not formally defined, this always will remain a thesis unless some more effective computation concept would emerge.

## 24. Category theory

Given an element $A$ in a **category** $C$, let $h^A$ denote the **functor** which assigns to a set $X$ the set $\mathrm{Hom}(A, X)$ of all **morphisms** from $A$ to $X$. Given a **functor** $F$ from $C$ to the category $S = \mathrm{Set}$, let $N(G, F)$ be the set of **natural transformations** from $G = h^A$ to $F$. (A **natural transformation** between two functors $G$ and $F$ from $C$ to $S$ assigns to every object $x$ in $C$ a morphism $\eta_x : G(x) \to F(x)$ such that for every morphism $f : x \to y$ in $C$ we have $\eta_y \circ G(f) = F(f) \circ \eta_x$.) The **functor category** defined by $C$ and $S$ has as objects the functors $F$ and as morphisms the natural transformations. The **Yoneda lemma** is

**Theorem:** $N(h^A, F)$ can be identified with $F(A)$.

Category theory was introduced in 1945 by Samuel Eilenberg and Sounders Mac Lane. The lemma above is due to Nobuo Yoneda from 1954. It allows to see a category embedded in a **functor category** which is a **topos** and serves as a sort of completion. One can identify a set $S$ for example with $\mathrm{Hom}(1, S)$. An other example is **Cayley's theorem** stating that the category of groups can be completely understood by looking at the group of permutations of $G$. For category theory, see [238, 218]. For history, [216].

## 25. Perturbation theory

A function $f$ of several variables is called **smooth** if one can take **first partial derivatives** like $\partial_x, \partial_y$ and second partial derivatives like $\partial_x \partial_y f(x, y) = f_{xy}(x, y)$ and still get continuous function. Assume $f(x, y)$ is a **smooth function** of two Euclidean variables $x, y \in \mathbb{R}^n$. If $f(a, 0) = 0$, we say $a$ is a **root** of $x \to f(x, y)$. If $f_y(x_0, y)$ is invertible, the root is called **non-degenerate**. If there is a solution $f(g(y), y) = 0$ such that $g(0) = a$ and $g$ is continuous, the root $a$ has a **local continuation** and say that it **persists** under perturbation.

**Theorem:** A non-degenerate root persists under perturbation.

This is the **implicit function theorem**. There are concrete and fast algorithms to compute the continuation. An example is the **Newton method** which iterates $T(x) = x - f(x, y)/f_x(x, y)$ to find the roots of $x \to f(x, y)$ for fixed $y$. The importance of the implicit function theorem is both theoretical as well as applied. The result assures that one can makes statements about a complicated theory near some model, which is understood. There are related situations, like if we want to continue a solution of $F(x,y) = (f(x,y), g(x,y)) = (0,0)$ giving **equilibrium points** of the **vector field** $F$. Then the Newton step $T(x,y) = (x,y) - dF^{-1}(x,y) \cdot F(x,y)$ method allows a continuation if $dF(x,y)$ is invertible. This means that small deformations of $F$ do not lead to changes of the nature of the equilibrium points. When equilibrium points change, the system exhibits **bifurcations**. This in particular applies to $F(x,y) = \nabla f(x,y)$, where equilibrium points are **critical points**. The derivative $dF$ of $F$ is then the **Hessian**.

## 26. Counting

A **simplicial complex** $X$ is a finite set of non-empty sets that is closed under the operation of taking finite non-empty subsets. The **Euler characteristic** $\chi$ of a simplicial complex $G$ is defined as $\chi(X) = \sum_{x \in X} (-1)^{\dim(x)}$, where the **dimension** $\dim(x)$ of a set $x$ is its cardinality $|x|$ minus 1.

**Theorem:** $\chi(X \times Y) = \chi(X)\chi(Y)$.

For **zero-dimensional simplicial complexes** $G$, (meaning that all sets in $G$ have cardinality 1), we get the **rule of product**: if you have $m$ ways to do one thing and $n$ ways to do an other, then there are $mn$ ways to do both. This **fundamental counting principle** is used in probability theory for example. The **Cartesian product** $X \times Y$ of two complexes is defined as the set-theoretical product of the two finite sets. It is not a simplicial complex any more in general but has the same Euler characteristic than its Barycentric refinement $(X \times Y)_1$, which is a simplicial complex. The maximal dimension of $A \times B$ is $\dim(A) + \dim(B)$ and $p_X(t) = \sum_{k=0}^n v_k(X)t^k$ is the generating function of $v_k(X)$, then $p_{X \times Y}(t) = p_X(t)p_Y(t)$ implying the counting principle. as $p_X(-1) = \chi(X)$. The function $p_X(t)$ is called the **Euler polynomial** of $X$. The importance of Euler characteristic as a **counting tool** lies in the fact that only $\chi(X) = p_X(-1)$ is invariant under **Barycentric subdivision** $\chi(X) = X_1$, where $X_1$ is the complex which consists of the vertices of all complete subgraphs of the graph in which the sets of $X$ are the vertices and where two are connected if one is contained in the other. The concept of Euler characteristic goes so over to continuum spaces like **manifolds** where the product property holds too. See for example [10].

## 27. Metric spaces

A continuous map $T : X \to X$, where $(X, d)$ is a **complete** non-empty **metric space** is called a **contraction** if there exists a real number $0 < \lambda < 1$ such that $d(T(x), T(y)) \leq \lambda d(x, y)$ for all $x, y \in X$. The space is called **complete** if every **Cauchy sequence** in $X$ has a **limit**. (A sequence $x_n$ in $X$ is called **Cauchy** if for all $\epsilon > 0$, there exists $n > 0$ such that for all $i, j > n$, one has $d(x_i, x_j) < \epsilon$.)

**Theorem:** A contraction has a unique fixed point in $X$.

This result is the **Banach fixed point theorem** proven by Stefan Banach from 1922. The example case $T(x) = (1 - x^2)/2$ on $X = \mathbb{Q} \cap [0.3, 0.6]$ having contraction rate $\lambda = 0.6$ and $T(X) = \mathbb{Q} \cap [0.32, 0.455] \subset X$ shows that completeness is necessary. The unique fixed point of $T$ in $X$ is $\sqrt{2} - 1 = 0.414...$ which is not in $\mathbb{Q}$ because $\sqrt{2} = p/q$ would imply $2q^2 = p^2$, which is not possible for integers as the left hand side has an odd number of prime factors 2 while the right hand side has an even number of prime factors. See [261]

## 28. Dirichlet series

The **abscissa of simple convergence** of a **Dirichlet series** $\zeta(s) = \sum_{n=1}^{\infty} a_n e^{-\lambda_n s}$ is $\sigma_0 = \inf\{a \in \mathcal{R} \mid \zeta(z)$ converges for all $\text{Re}(z) > a\ \}$. For $\lambda_n = n$ we have the **Taylor series** $f(z) = \sum_{n=1}^{\infty} a_n z^n$ with $z = e^{-s}$. For $\lambda_n = \log(n)$ we have the **standard Dirichlet series** $\sum_{n=1}^{\infty} a_n/n^s$. For example, for $a_n = z^n$, one gets the **poly-logarithm** $\text{Li}_s(z) = \sum_{n=1}^{\infty} z^n/n^s$ and especially $\text{Li}_s(1) = \zeta(s)$, the **Riemann zeta function** or the **Lerch transcendent** $\Phi(z, s, a) = \sum_{n=1}^{\infty} z^n/(n + a)^s$. Define $S(n) = \sum_{k=1}^{n} a_k$. The **Cahen's formula** applies if the series $S(n)$ does not converge.

**Theorem:** $\sigma_0 = \limsup_{n\to\infty} \frac{\log|S(n)|}{\lambda_n}$ .

There is a similar formula for the **abscissa of absolute convergence** of $\zeta$ which is defined as $\sigma_a = \inf\{a \in \mathcal{R} \mid \zeta(z)$ converges absolutely for all $\text{Re}(z) > a\ \}$. The result is $\sigma_a = \limsup_{n\to\infty} \frac{\log(\overline{S}(n))}{\lambda_n}$, For example, for the **Dirichlet eta function** $\zeta(s) = \sum_{n=1}^{\infty}(-1)^{n-1}/n^s$ has the abscissa of convergence $\sigma_0 = 0$ and the absolute abscissa of convergence $\sigma_a = 1$. The series $\zeta(s) = \sum_{n=1}^{\infty} e^{in^{\alpha}}/n^s$ has $\sigma_a = 1$ and $\sigma_0 = 1 - \alpha$. If $a_n$ is multiplicative $a_{n+m} = a_n a_m$ for relatively prime $n, m$, then $\sum_{n=1}^{\infty} a_n/n^s = \prod_p (1 + a_p/p^s + a_{p^2}/p^{2s} + \cdots)$ generalizes the **Euler golden key formula** $\sum_n 1/n^s = \prod_p (1 - 1/p^s)^{-1}$. See [153, 155].

## 29. Trigonometry

Mathematicians had a long and painful struggle with the concept of **limit**. One of the first to ponder the question was Zeno of Elea around 450 BC. Archimedes of Syracuse made some progress around 250 BC. Since Augustin-Louis Cauchy, one defines the **limit** $\lim_{x\to a} f(x) = b$ **to exist** if and only if for all $\epsilon > 0$, there exists a $\delta > 0$ such that if $|x-a| < \delta$, then $|f(x)-b| < \epsilon$. A place where limits appear are when computing **derivatives** $g'(0) = \lim_{x\to 0}[g(x)-g(0)]/x$. In the case $g(x) = \sin(x)$, one has to understand the limit of the function $f(x) = \sin(x)/x$ which is the **sinc** function. A prototype result is the **fundamental theorem of trigonometry** (called as such in some calculus texts like [47]).

**Theorem:** $\lim_{x\to 0} \sin(x)/x = 1$.

It appears strange to give weight to such a special result but it explains the difficulty of limit and the **l'Hôpital rule** of 1694, which was formulated in a book of Bernoulli commissioned to Hôpital: the limit can be obtained by differentiating both the denominator and nominator and taking the limit of the quotients. The result allows to derive (using trigonometric identities) that in general $\sin'(x) = \cos(x)$ and $\cos'(x) = -\sin(x)$. One single limit is the gateway. It is important also culturally because it embraces thousands of years of struggle. It was Archimedes, who when computing the **circumference of the circle formula** $2\pi r$ using **exhaustion** using regular polygons from the inside and outside. Comparing the lengths of the approximations essentially battled that fundamental theorem of trigonometry. The identity is therefore the epicenter around the development of **trigonometry**, **differentiation** and **integration**.

## 30. Logarithms

The **natural logarithm** is the inverse of the **exponential function** $\exp(x)$ establishing so a **group homomorphism** from the additive group $(\mathbb{R}, +)$ to the multiplicative group $(\mathbb{R}^+, *)$. We have:

**Theorem:** $\log(uv) = \log(u) + \log(v)$.

This follows from $\exp(x + y) = \exp(x)\exp(y)$ and $\log(\exp(x)) = \exp(\log(x)) = x$ by plugging in $x = \log(u), y = \log(v)$. The logarithms were independently discovered by Jost Bürgi around 1600 and John Napier in 1614 [306]. The **logarithm** with base $b > 0$ is denoted by $\log_b$. It is the inverse of $x \to b^x = e^{x\log(b)}$. The concept of logarithm has been extended in various ways: in any **group** $G$, one can define the **discrete logarithm** $\log_b(a)$ to base $b$ as an **integer** $k$ such that $b^k = a$ (if it exists). For complex numbers the **complex logarithm** $\log(z)$ as any solution $w$ of $e^w = z$. It is **multi-valued** as $\log(|z|) + i\arg(z) + 2\pi ik$ all solve this with some integer $k$, where $\arg(z) \in (-\pi, \pi)$. The identity $\log(uv) = \log(u)+\log(v)$ is now only true up to $2\pi ki$. Logarithms can also be defined for matrices. Any matrix $B$ solving $\exp(B) = A$ is called a **logarithm** of $A$. For $A$ close to the identity $I$, can define $\log(A) = (A - I) - (A - I)^2/2 + (A - I)^3/3 - ...$, which is a **Mercator series**. For **normal invertible matrices**, one can define logarithms using the **functional calculus** by diagonalization. On a **Riemannian manifold** $M$, one also has an exponential map: it is a diffeomorphim from a small ball $B_r(0)$ in the **tangent space** $x \in M$ to $M$. The map $v \to \exp_x(v)$ is obtained by defining $\exp_x(0) = x$ and by taking for $v \neq 0$ a **geodesic** with initial direction $v/|v|$ and running it for time $|v|$. The logarithm $\log_x$ is now defined on a **geodesic ball** of radius $r$ and defines an element in the tangent space. In the case of a Lie group $M = G$, where the points are matrices, each tangent space is its **Lie algebra**.

## 31. Geometric probability

A subset $K$ of $\mathbb{R}^n$ is called **compact** if it is **closed** and **bounded**. By **Bolzano-Weierstrass** this is equivalent to the fact that every infinite sequence $x_n$ in $K$ has a **subsequence** which converges. A subset $K$ of $\mathbb{R}^n$ is called **convex**, if for any two given points $x, y \in K$, the interval $\{x + t(y - x), t \in [0, 1]\}$ is a subset of $K$. Let $G$ be the set of all **compact convex subsets** of $\mathbb{R}^n$. An **invariant valuation** $X$ is a function $X : G \to \mathbb{R}$ satisfying $X(A \cup B) + X(A \cap B) = X(A) + X(B)$, which is continuous in the **Hausdorff metric** $d(K, L) = \max(\sup_{x\in K} \inf_{y\in L} d(x, y) + \sup_{y\in K} \inf_{x\in L} d(x, y))$ and invariant under **rigid motion** generated by rotations, reflections and translations in the linear space $\mathbb{R}^n$.

**Theorem:** The space of valuations is $(n + 1)$-dimensional.

The theorem is due to Hugo Hadwiger from 1937. The coefficients $a_j(G)$ of the polynomial $\mathbf{Vol}(G + tB) = \sum_{j=0}^{n} a_j t^j$ are a basis, where $B$ is the **unit ball** $B = \{|x| \leq 1\}$. See [192].

## 32. Partial differential equations

A **quasilinear partial differential equation** is a differential equation of the form $u_t(x, t) = F(x, t, u) \cdot \nabla_x u(x, t) + f(x, t, u)$ with initial condition $u(x, 0) = u_0$ and an **analytic vector field** $F$. It defines a **quasi-linear Cauchy problem**.

**Theorem:** A quasi-linear Cauchy problem has a unique analytic solution.

This is the **Cauchy-Kovalevskaya theorem**. It was initiated by Augustin-Louis Cauchy in 1842 and proven in 1875 by Sophie Kowalevskaya. Smoothness alone is not enough. For a shorter introduction to partial differential equations, see [18].

## 33. Game theory

If $S = (S_1, \dots, S_n)$ are $n$ **players** and $f = (f_1, \dots, f_n)$ is a **payoff function** defined on a **strategy profile** $x = (x_1, \dots, x_n)$. A point $x^*$ is called an **equilibrium** if $f_i(x^*)$ is **maximal** with respect to changes of $x_i$ alone in the profile $x$ for every player $i$.

**Theorem:** There is an equilibrium for any game with mixed strategy

The equilibrium is called a **Nash equilibrium**. It tells us what we would see in a world if everybody is doing their best, given what everybody else is doing. John Forbes Nash used in 1950 the **Brouwer fixed point theorem** and later in 1951 the **Kakutani fixed point theorem** to prove it. The Brouwer fixed point theorem itself is generalized by the **Lefschetz fixed point theorem** which equates the super trace of the induced map on cohomology with the sum of the indices of the fixed points. About John Nash and some history of game theory, see [293]: game theory started maybe with Adam Smith's the Wealth of Nations published in 1776, Ernst Zermelo in 1913 (Zermelo's theorem), Émile Borel in the 1920s and John von Neumann in 1928 pioneered mathematical game theory. Together with Oskar Morgenstern von Neumann merged game theory with economics in 1944. Nash published his thesis in a paper of 1951. For the mathematics of games, see [335].

## 34. Measure theory

A topological space with open sets $\mathcal{O}$ defines the **Borel $\sigma$-algebra**, the smallest $\sigma$ algebra which contains $\mathcal{O}$. For the metric space $(\mathbb{R}, d)$ with $d(x, y) = |x - y|$, already the intervals generate the Borel $\sigma$ algebra $\mathcal{A}$. A **Borel measure** is a measure defined on a Borel $\sigma$-algebra. Every **Borel measure** $\mu$ on the real line $\mathbb{R}$ can be decomposed uniquely into an **absolutely continuous** part $\mu_{ac}$, a **singular continuous** part $\mu_{sc}$ and a **pure point** part $\mu_{pp}$:

**Theorem:** $\mu = \mu_{ac} + \mu_{sc} + \mu_{pp}$.

This is called the **Lebesgue decomposition theorem**. It uses the **Radon-Nikodym theorem**. The decomposition theorem implies the decomposition theorem of the **spectrum** of a linear operator. See [297] (like page 259). Lebesgue's theorem was published in 1904. A generalization due to Johann Radon and Otto Nikodym was done in 1913.

## 35. Geometric number theory

If $\Gamma$ is a **lattice** in $\mathbb{R}^n$, denote with $\mathbb{R}^n/\Gamma$ the **fundamental region** and by $|\Gamma|$ its **volume**. A set $K$ is **convex** if $x, y \in K$ implies $x + t(x - y) \in K$ for all $0 \leq t \leq 1$. A set $K$ is **centrally symmetric** if $x \in K$ implies $-x \in K$. A region is **Minkowski** if it is convex and centrally symmetric. Let $|K|$ denote the volume of $K$.

**Theorem:** If $K$ is Minkowski and $|K| > 2^n|\Gamma|$ then $K \cap \Gamma \neq \emptyset$.

The theorem is due to Hermann Minkowski in 1896. It lead to a field called **geometry of numbers**. [62]. It has many applications in number theory and **Diophantine analysis** [54, 167]

## 36. Fredholm

An **integral kernel** $K(x,y) \in L^2([a,b]^2)$ defines an **integral operator** $A$ defined by $Af(x) = \int_a^b K(x,y)f(y)\ dy$ with adjoint $T^*f(x) = \int_a^b \overline{K(y,x)}f(y)\ dy$. The $L^2$ assumption makes the function $K(x,y)$ what one calls a **Hilbert-Schmidt kernel**. Fredholm showed that the **Fredholm equation** $A^*f = (T^* - \overline{\lambda})f = g$ has a solution $f$ if and only if $f$ is perpendicular to the kernel of $A = T - \lambda$. This identity $\ker(A)^\perp = \mathrm{im}(A^*)$ is in finite dimensions part of the **fundamental theorem of linear algebra**. The **Fredholm alternative** reformulates this in a more catchy way as an **alternative**:

**Theorem:** Either $\exists f \neq 0$ with $Af = 0$ or for all $g$, $\exists f$ with $Af = g$.

In the second case, the solution depends continuously on $g$. The alternative can be put more generally by stating that if $A$ is a **compact operator** on a Hilbert space and $\lambda$ is not an eigenvalue of $A$, then the **resolvent** $(A-\lambda)^{-1}$ is bounded. A bounded operator $A$ on a Hilbert space $H$ is called **compact** if the image of the unit ball is relatively compact (has a compact closure). The Fredholm alternative is part of **Fredholm theory**. It was developed by Ivar Fredholm in 1903.

## 37. Prime distribution

The **Dirichlet theorem** about the primes along an arithmetic progression tells that if $a$ and $b$ are **relatively prime** meaning that there largest common divisor is 1, then there are infinitely many primes of the form $p = a \bmod b$. The Green-Tao theorem strengthens this. We say that a set $A$ contains **arbitrary long arithmetic progressions** if for every $k$ there exists an **arithmetic progression** $\{a + bj, j = 1, \cdots, k\}$ within $A$.

**Theorem:** The set of primes contains arbitrary long arithmetic progressions.

The **Dirichlet prime number theorem** in 1837. The **Green-Tao theorem** was done in 2004 and appeared in 2008 [141]. It uses **Szemerédi's theorem** [129] which shows that any set $A$ of positive upper density $\limsup_{n\to\infty} |A \cap \{1 \cdots n\}|/n$ has arbitrary long arithmetic progressions. So, any subset $A$ of the primes $P$ for which the **relative density** $\limsup_{n\to\infty} |A \cap \{1 \cdots n\}|/|P \cap \{1 \cdots n\}|$ is positive has arbitrary long arithmetic progressions. For non-linear sequences of numbers the problems are wide open. The **Landau problem** of the infinitude of primes of the form $x^2 + 1$ illustrates this. The Green-Tao theorem gives hope to tackle the **Erdös conjecture on arithmetic progressions** telling that a sequence $\{x_n\}$ of integers satisfying $\sum_n x_n = \infty$ contains arbitrary long arithmetic progressions.

## 38. Riemannian geometry

A **Riemannian manifold** is a smooth finite dimensional manifold $M$ equipped with a **symmetric**, **positive definite tensor** $(u,v) \to g_x(u,v)$ defining on each **tangent space** $T_xM$ an **inner product** $(u,v)_x = (g_x(u,v)u,v)$, where $(u,v)$ is the **standard inner product**. Let $\Omega$ be the space of **smooth vector fields**. A **connection** is a bilinear map $(X,Y) \to \nabla_X Y$ from $\Omega \times \Omega$ to $\Omega$ satisfying the differentiation rules $\nabla_{fX} Y = f \nabla_X Y$ and **Leibniz rule**

$\nabla_X(fY) = df(X)Y + f\nabla_X Y$. It is **compatible with the metric** if the **Lie derivative** satisfies $\delta_X(Y,Z) = (\Gamma_X Y, Z) + (Y, \Gamma_X Z)$. It is **torsion-free** if $\nabla_X Y - \nabla_Y X = [X,Y]$ is the **Lie bracket** on $\Omega$.

**Theorem:** There is exactly one torsion-free connection compatible with $g$.

This is the **fundamental theorem of Riemannian geometry.** The connection is called the **Levi-Civita connection**, named after Tullio Levi-Civita. See for example [100, 2, 304, 87].

## 39. Symplectic geometry

A **symplectic manifold** $(M,\omega)$ is a smooth $2n$-manifold $M$ equipped with a non-degenerate closed 2-form $\omega$. The later is called a **symplectic form**. As a 2-form, it satisfies $\omega(x,y) = -\omega(y,x)$. **Non-degenerate** means $\omega(u,v) = 0$ for all $v$ implies $u = 0$. The **standard symplectic form** is $\omega_0 = \sum_{i<j} dx_i \wedge dx_j$.

**Theorem:** Every symplectic form is locally diffeomorphic to $\omega_0$.

This theorem is due to Jean Gaston Darboux from 1882. Modern proofs use **Moser's trick** from 1965. The Darboux theorem assures that locally, two symplectic manifolds of the same dimension are symplectic equivalent. It also implies that **symplectic matrices** have **determinant** 1. In contrast, for **Riemannian manifolds**, one can not trivialize the Riemannian metric in a neighborhood one can only render it the standard metric at the point itself. See [165].

## 40. Differential topology

Given a **smooth function** $f$ on a **differentiable manifold** $M$. Let $df$ denote the **gradient** of $f$. A point $x$ is called a **critical point**, if $df(x) = 0$. We assume $f$ has only finitely many **critical points** and that all of them are **non-degenerate**. The later means that the **Hessian** $d^2f(x)$ is invertible at $x$. One calls such functions **Morse functions**. The **Morse index** of a critical point $x$ is the number of negative eigenvalues of $d^2f$. The **Morse inequalities** relate the number $c_k(f,K)$ of critical points of index $k$ of $f$ with the **Betti numbers** $b_k(M)$, defined as the nullity of the **Hodge star operator** $dd^* + d^*d$ restricted to $k$-forms $\Omega_k$, where $d_k : \Omega_k \to \Omega_{k+1}$ is the **exterior derivative**.

**Theorem:** $c_k - c_{k-1} + \cdots + (-1)^k c_0 \geq b_k - b_{k-1} + \cdots + (-1)^k b_0$.

These are the **Morse inequalities** due to Marston Morse from 1934. It implies in particular the **weak Morse inequalities** $b_k \leq c_k$. Modern proofs use **Witten deformation** [87] of the exterior derivative $d$.

## 41. Non-commutative geometry

A **spectral triple** $(A,H,D)$ is given by a **Hilbert space** $H$, a $C^*$**-algebra** $A$ of operators on $H$ and a densely defined self-adjoint operator $D$ satisfying $||[D,a]|| < \infty$ for all $a \in A$ such that $e^{-tD^2}$ is **trace class**. The operator $D$ is called a **Dirac operator**. The set-up generalizes Riemannian geometry because of the following result dealing with the **exterior derivative** $d$ on a Riemannian manifold $(M,g)$, where $A = C(M)$ is the $C^*$-algebra of continuous functions and $D = d + d^*$ is the Dirac operator, defining the spectral triple of $(M,g)$. Let $\delta$ denote the **geodesic distance** in $(M,g)$:

**Theorem:** $\delta(x,y) = \sup_{f \in A, ||[D,f]|| \leq 1} |f(x) - f(y)|$.

This formula of Alain Connes tells that the spectral triple determines the geodesic distance in $(M,g)$ and so the metric $g$. It justifies to look at spectral triples as non-commutative generalizations of Riemannian geometry. See [74].

## 42. Polytopes

A **convex polytop** $P$ in dimension $n$ is the **convex hull** of finitely many points in $R^n$. One assumes all vertices to be **extreme points**, points which do not lie in an open line segment of $P$. The **boundary** of $P$ is formed by $(n-1)$ dimensional boundary facets. The notion of **Platonic solid** is recursive. A convex polytop is **Platonic**, if all its facets are Platonic $(n-1)$-dimensional polytopes and vertex figures. Let $p = (p_2, p_3, p_4, \dots)$ encode the number of Platonic solids meaning that $p_d$ is the number of Platonic polytops in dimension $d$.

**Theorem:** There are 5 platonic solids and $p = (\infty, 5, 6, 3, 3, 3, \dots)$

In dimension 2, there are infinitely many. They are the **regular polygons**. The list of Platonic solids is "octahedron", "dodecahedron", "icosahedron", "tetrahedron" and "cube" has been known by the Greeks already. Ludwig Schläfli first classified the higher dimensional case. There are six in dimension 4: they are the "5 cell", the "8 cell" (**tesseract**), the "16 cell", the "24 cell", the "120 cell" and the "600 cell". There are only three regular polytopes in dimension 5 and higher, where only the analog of the tetrahedron, cube and octahedron exist. For literature, see [146, 350, 275].

## 43. Descriptive set theory

A **metric space** $(X,d)$ is a set with a **metric** $d$ (a function $X \times X \to [0,\infty)$ satisfying **symmetry** $d(x,y) = d(y,x)$, the **triangle inequality** $d(x,y) + d(y,z) \geq d(x,z)$, and $d(x,y) = 0 \leftrightarrow x = y$.) A metric space $(X,d)$ is **complete** if every **Cauchy sequence** converges in $X$. A metric space is of **second Baire category** if the intersection of a countable set of open dense sets is dense. The **Baire Category theorem** tells

**Theorem:** Complete metric spaces are of second Baire category.

One calls the intersection $A$ of a countable set of open dense sets $A$ in $X$ also a **generic set** or **residual set**. The complement of a generic set is also called a **meager set** or **negligible** or a set of **first category**. It is the union of countably many nowhere dense sets. Like measure theory, Baire category theory allows for existence results. There can be surprises: a generic continuous function is not differentiable for example. For descriptive set theory, see [191]. The frame work for classical descriptive set theory often are **Polish spaces**, which are separable complete metric spaces. See [44].

## 44. Calculus of variations

Let $X$ be the vector space of **smooth**, **compactly supported** functions $h$ on an interval $(a,b)$. The **fundamental lemma of calculus of variations** tells

**Theorem:** $\int_a^b f(x)g(x)dx = 0$ for all $g \in X$, then $f = 0$.

The result is due to Joseph-Louis Lagrange. One can restate this as the fact that if $f = 0$ **weakly** then $f$ is actually zero. It implies that if $\int_a^b f(x)g'(x)\,dx = 0$ for all $g \in X$, then $f$ is constant. This is nice as $f$ is not assumed to be differentiable. The result is used to prove that extrema to a **variational problem** $I(x) = \int_a^b L(t,x,x')\,dt$ are weak solutions of the **Euler Lagrange equations** $L_x = d/dt L_{x'}$. See [132, 248].

## 45. Integrable systems

Given a **Hamilton differential equation** $x' = J\nabla H(x)$ on a compact **symplectic** $2n$-**manifold** $(M,\omega)$. The **almost complex structure** $J : T^*M \to TM$ is tied to $\omega$ using a Riemannian metric $g$ by $\omega(v,w) = \langle v, Jg\rangle$. A function $F : M \to \mathbb{R}$ is called an **first integral** if $d/dt F(x(t)) = 0$ for all $t$. An example is the **Hamiltonian function** $H$ itself. A set of integrals $F_1,\dots,F_k$ **Poisson commutes** if $\{F_j, F_k\} = J\nabla F_j \cdot \nabla F_k = 0$ for all $k,j$. They are **linearly independent**, if at every point the vectors $\nabla F_j$ are linearly independent in the sense of linear algebra. A system is **Liouville integrable** if there are $d$ linearly independent, Poisson commuting integrals. The following theorem due to Liouville and Arnold characterizes the **level surfaces** $\{F = c\} = \{F_1 = c_1, \dots F_d = c_d\}$:

**Theorem:** For a Liouville integrable system, level surfaces $F = c$ are tori.

An example how to get integrals is to write the system as an **isospectral deformation** of an operator $L$. This is called a **Lax system**. Such a differential equation has the form $L' = [B, L]$, where $B = B(L)$ is skew symmetric. An example is the **periodic Toda system** $\dot{a}_n = a_n(b_{n+1} - b_n)$, $\dot{b}_n = 2(a_n^2 - a_{n-1}^2)$, where $(Lu)_n = a_n u_{n+1} + a_{n-1}u_{n-1} + b_n u_n$ and $(Bu)_n = a_n u_{n+1} - a_{n-1}u_{n-1}$. An other example is the motion of a **rigid body** in $n$ dimensions if the center of mass is fixed. See [17].

## 46. Harmonic analysis

On the vector space $X$ of continuously differentiable $2\pi$ periodic, complex- valued functions, define the **inner product** $(f,g) = (2\pi)^{-1}\int f(x)\overline{g}(x)\,dx$. The **Fourier coefficients** of $f$ are $\hat{f}_n = (f, e_n)$, where $\{e_n(x) = e^{inx}\}_{n\in\mathbb{Z}}$ is the **Fourier basis**. The **Fourier series** of $f$ is the sum $\sum_{n\in\mathbb{Z}} \hat{f}_n e^{inx}$.

**Theorem:** The Fourier series of $f \in X$ converges point-wise to $f$.

Already Fourier claimed this always to be true in his "Théorie Analytique de la Chaleur". After many fallacious proofs, Dirichlet gave the first proof of convergence [208]. The case is subtle as there are continuous functions for which the convergence fails at some points. Lipót Féjer was able to show that for a continuous function $f$, the coefficients $\hat{f}_n$ nevertheless determine the function using **Césaro convergence**. See [190].

## 47. Jensen inequality

If $V$ is a **vector space**, a set $X$ is called **convex** if for all points $a,b \in X$, the **line segment** $\{tb + (1-t)a \mid t \in [0,1]\}$ is contained in $X$. A real-valued function $\phi : X \to \mathbb{R}$ is called **convex** if $\phi(tb + (1-t)a) \le t\phi(b) + (1-t)\phi(a)$ for all $a,b \in X$ and all $t \in [0,1]$. Let now $(\Omega, \mathcal{A}, \mathrm{P})$ be a **probability space**, and $f \in L^1(\Omega,\mathrm{P})$ an integrable function. We write $\mathrm{E}[f] = \int_\omega f(x)\,dP(x)$

for the **expectation** of $f$. For any convex $\phi : \mathbb{R} \to \mathbb{R}$ and $f \in L^1(\Omega, P)$, we have the **Jensen inequality**

**Theorem:** $\phi(\mathrm{E}[f]) \leq \mathrm{E}[\phi(f)]$.

For $\phi(x) = \exp(x)$ and a finite probability space $\Omega = \{1, 2, \dots, n\}$ with $f(k) = x_k = \exp(y_k)$ and $\mathrm{P}[\{x\}] = 1/n$, this gives the **arithmetic mean- geometric mean inequality** $(x_1 \cdot x_2 \cdots x_n)^{1/n} \leq (x_1 + x_2 + \cdots + x_n)/n$. The case $\phi(x) = e^x$ is useful in general as it leads to the inequality $e^{\mathrm{E}[f]} \leq \mathrm{E}[e^f]$ if $e^f \in L^1$. For $f \in L^2(\omega, P)$ one gets $(\mathrm{E}[f])^2 \leq \mathrm{E}[f^2]$ which reflects the fact that $\mathrm{E}[f^2] - (\mathrm{E}[f])^2 = \mathrm{E}[(f - \mathrm{E}[f])^2] = \mathrm{Var}[f] \geq 0$ where $\mathrm{Var}[f]$ is the **variance** of $f$.

## 48. Jordan curve theorem

A **closed curve** in the image of a continuous map $\mathbb{T} \to \mathbb{R}^2$. It is called **simple**, if this map is injective. One then calls the map an **embedding** and the image a **topological 1-sphere** or a **Jordan curve**. The **Jordan curve theorem** deals with simple closed curves $S$ in the two-dimensional plane.

**Theorem:** A simple closed curve divides the plane into two regions.

The Jordan curve theorem is due to Camille Jordan. His proof [178] was objected at first [194] but rehabilitated in [149]. The theorem can be strengthened, a **theorem of Schoenflies** tells that each of the two regions is homeomorphic to the disk $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$. In the smooth case, it is even possible to extend the map to a diffeomorphism in the plane. In higher dimensions, one knows that an embedding of the $(d-1)$ dimensional sphere in a $\mathbb{R}^d$ divides space into two regions. This is the **Jordan-Brouwer** separation theorem. It is no more true in general that the two parts are homeomorphic to $\{x \in \mathbb{R}^d \mid |x| < 1\}$: a counter example is the **Alexander horned sphere** which is a topological 2-sphere but where the unbounded component is not simply connected and so not homeomorphic to the complement of a unit ball. See [44].

## 49. Chinese remainder theorem

Given integers $a, b$, a **linear modular equation** or **congruence** $ax + b = 0 \mod \mathrm{m}$ asks to find an integer $x$ such that $ax + b$ is divisible by $m$. This linear equation can always be solved if $a$ and $m$ are coprime. The **Chinese remainder theorem** deals with the **system of linear modular equations** $x = b_1 \bmod m_1, x = b_2 \bmod m_2, \dots, x = b_n \bmod m_n$, where $m_k$ are the **moduli**. More generally, for an integer $n \times n$ matrix $A$ we call $Ax = b \mathrm{mod}\ m$ a **Chinese remainder theorem system** or shortly **CRT system** if the $m_j$ are pairwise relatively prime and in each row there is a matrix element $A_{ij}$ relatively prime to $m_i$.

**Theorem:** Every Chinese remainder theorem system has a solution.

The classical single variable case case is when $A_{i1} = 1$ and $A_{ij} = 0$ for $j > 1$. Let $M = m_1 \cdots m_2 \cdots m_n$ be the product. In this one-dimensional case, the result implies that $x \mathrm{mod}\ M \to (x \bmod \ m_1, \dots, (x \bmod m_n)$ is a ring isomorphism. Define $M_i = M/m_i$. An explicit algorithm is to finding numbers $y_i, z_i$ with $y_i M_i + z_i m_i = 1$ (finding $y, z$ solving $ay + bz = 1$ for coprime $a, b$ is computed using the **Euclidean algorithm**), then finding $x = b_1 m_1 y_1 + \cdots + b_n m_n y_n$. [95, 234]. The multi-variable version appeared in 2005 [198, 200].

## 50. Bézout's theorem

A polynomial is **homogeneous** if the total degree of all its **monomials** is the same. A **homogeneous polynomial** $f$ in $n+1$ variables of degree $d \geq 1$ defines a **projective hypersurface** $f = 0$. Given $n$ projective irreducible hypersurfaces $f_k = c_k$ of degree $d_k$ in a **projective space** $\mathbb{P}^n$ we can look at the solution set $\{f = c\} = \{f_1 = c_1, \cdots, f_k = c_k\}$ of a system of nonlinear equations. The **Bézout's bound** is $d = d_1 \cdots d_k$ the product of the degrees. **Bézout's theorem** allows to count the number of solutions of the system, where the number of solutions is counted with multiplicity.

**Theorem:** The set $\{f = c\}$ is either infinite or has $d$ elements.

Bézout's theorem was stated in the "Principia" of Newton in 1687 but proven fist in 1779 by Étienne Bézout. If the hypersurfaces are all **irreducible** and in "general position", then there are exactly $d$ solutions and each has multiplicity 1. This can be used also for affine surfaces. If $y^2 - x^3 - 3x - 5 = 0$ is an **elliptic curve** for example, then $y^2z - x^3 - 3xz^2 - 5z^3 =$ is a projective hypersurface, its **projective completion**. Bézout's theorem implies part the fundamental theorem of algebra as for $n = 1$, when we have only one homogeneous equation we have $d$ roots to a polynomial of degree $d$. The theorem implies for example that the intersection of two **conic sections** have in general 2 intersection points. The example $x^2 - yz = 0, x^2 + z^2 - yz = 0$ has only the solution $x = z = 0, y = 1$ but with multiplicity 2. As non-linear systems of equations appear frequently in **computer algebra** this theorem gives a lower bound on the computational complexity for solving such problems.

## 51. Group theory

A **finite group** $(G, *, 1)$ is a finite set containing a **unit** $1 \in G$ and a binary operation $* : G \times G \to G$ satisfying the **associativity property** $(x * y) * z = x * (y * z)$ and such that for every $x$, there exists a unique $y = x^{-1}$ such that $x * y = y * x = 1$. The **order** $n$ of the group is the number of elements in the group. An element $x \in G$ generates a **subgroup** formed by $1, x, x^2 = x * x, \dots$. This is the **cyclic subgroup** $C(x)$ generated by $x$. **Lagrange's theorem** tells

**Theorem:** $|C(x)|$ is a factor of $|G|$

The origins of group theory go back to Joseph Louis Lagrange, Paulo Ruffini and Évariste Galois. The concept of abstract group appeared first in the work of Arthur Cayley. Given a subgroup $H$ of $G$, the **left cosets** of $H$ are the equivalence classes of the equivalence relation $x \sim y$ if there exists $z \in H$ with $x = z * y$. The equivalence classes $G/N$ partition $G$. The number $[G : N]$ of elements in $G/H$ is called the **index** of $H$ in $G$. It follows that $|G| = |H||G : H|$ and more generally that if $K$ is a subgroup of $H$ and $H$ is a subgroup of $G$ then $[G : K] = [G : H][H : K]$. The group $N$ generated by $x$ is a called a **normal group** $N \triangleleft G$ if for all $a \in N$ and all $x$ in $G$ the element $x * a * x^{-1}$ is in $N$. This can be rewritten as $H * x = x * H$. If $N$ is a normal group, then $G/H$ is again a group, the **quotient group**. For example, if $f : G \to G'$ is a group homomorphism, then the kernel of $f$ is a normal subgroup and $|G| = |\ker(f)||\mathrm{im}(f)|$ because of the **first group isomorphism theorem**.

## 52. Primes

A **prime** is an integer larger than 1 which is only divisible by 1 or itself. **The Wilson theorem** allows to define a prime as a number $n$ for which $(n-1)!+1$ is divisible by $n$. Euclid already knew that there are infinitely many primes (if there were finitely many $p_1,\dots,p_n$, the new number $p_1 p_2 \cdots p_n + 1$ would have a prime factor different from the given set). It also follows from the **divergence** of the **harmonic series** $\zeta(1) = \sum_{n=1}^\infty 1/n = 1 + 1/2 + 1/3 + \cdots$ and the **Euler golden key** or **Euler product** $\zeta(s) = \sum_{n=1}^\infty 1/n^2 = \sum_{p \text{ prime}} (1-1/p^s)^{-1}$ for the **Riemann zeta function** $\zeta(s)$ that there are infinitely many primes as otherwise, the product to the right would be finite.

Let $\pi(x)$ be the **prime-counting function** which gives the number of primes smaller or equal to $x$. Given two functions $f(x), g(x)$ from the integers to the integers, we say $f \sim g$, if $\lim_{x\to\infty} f(x)/g(x) = 1$. The **prime number theorem** tells

**Theorem:** $\pi(x) \sim x/\log(x)$.

The result was investigated experimentally first by Anton Ferkel and Jurij Vega, Adrien-Marie Legendre first conjectured in 1797 a law of this form. Carl Friedrich Gauss wrote in 1849 that he experimented independently around 1792 with such a law. The theorem was proven in 1896 by Jacques Hadamard and Charles de la Vallée Poussin. Proofs without complex analysis were put forward by Atle Selberg and Paul Erdös in 1949. The prime number theorem also assures that there are infinitely many primes but it makes the statement **quantitative** in that it gives an idea how fast the number of primes grow asymptotically. Under the assumption of the Riemann hypothesis, Lowell Schoenfeld proved $|\pi(x) - \mathrm{li}(x)| < \sqrt{x}\log(x)/(8\pi)$, where $\mathrm{li}(x) = \int_0^x dt/\log(t)$ is the **logarithmic integral**.

## 53. Cellular automata

A finite set $A$ called **alphabet** and an integer $d \geq 1$ defines the compact topological space $\Omega = A^{\mathbb{Z}^d}$ of all infinite d-dimensional configurations. The topology is the product topology which is compact by the Tychonov theorem. The translation maps $T_i(x)_n = x_{n+e_i}$ are homeomorphisms of $\Omega$ called **shifts**. A closed $T$ invariant subset $X \subset \Omega$ defines a **subshift** $(X, T)$. An automorphism $T$ of $\Omega$ which commutes with the translations $T_i$ is called a **cellular automaton**, abbreviated $CA$. An example of a cellular automaton is a map $Tx_n = \phi(x_{n+u_1}, \dots x_{n+u_k})$ where $U = \{u_1, \dots u_k\} \subset \mathbb{Z}^d$ is a fixed finite set. It is called an **local automaton** because it is defined by a finite rule so that the status of the cell $n$ at the next step depends only on the status of the "neighboring cells" $\{n+u \mid u \in U\}$. The following result is the **Curtis-Hedlund-Lyndon theorem**:

**Theorem:** Every cellular automaton is a local automaton.

Cellular automata were introduced by John von Neumann and mathematically in 1969 by Hedlund [160]. The result appears there. Hedlund saw cellular automata also as maps on subshifts. One can so look at cellular automata on subclasses of subshifts. For example, one can restrict the cellular automata map $T$ on almost periodic configurations, which are subsets $X$ of $\Omega$ on which $(X, T_1, \dots T_j)$ has only invariant measures $\mu$ for which the Koopman operators $U_i f = f(T_i)$ on $L^2(X,\mu)$ have pure point spectrum. A particularly well studied case is $d=1$ and $A = \{0,1\}$, if $U = \{-1,0,1\}$, where the automaton is called an **elementary cellular automaton**. The **Wolfram numbering** labels the $2^8$ possible elementary automata with a number between 1 and 255. The **game of life** of Conway is a case for $d = 2$ and $A = \{-1, 0, 1\} \times \{-1, 0, 1\}$. For literature on cellular automata see [345] or as part of complex systems [346] or evolutionary dynamics [257]. For topological dynamics, see [90].

## 54. Topos theory

A **category** has **objects** as **nodes** and **morphisms** as **arrows** going from one object to an other object. There can be multiple connections and self-loops so that one can visualize a category as a **quiver**. Every object has the identity arrow $1_A$. A **topos** $X$ is a **Cartesian closed** category $C$ in which **finite limits** exists and which has a **sub-object classifier** $\Omega$ allowing to identify sub-objects with morphisms from $X$ to $\Omega$. **Cartesian closed** means that one can define for any pair of objects $A, B$ in $C$ the **product** $A \times B$ and an **equalizer** representing solutions $f = g$ to arrows $f : A \to B, G : A \to B$ as well as an **exponential** $B^A$ representing all arrows from $A$ to $B$. An example is the topos of sets. An example of a sub-object classifier is $\Omega = \{0, 1\}$ encoding "true or false".
The **slice category** $E/X$ of a category $E$ with an object $X$ in $E$ is a category, where the objects are the arrows from $E \to X$. An $E/X$ arrow between objects $f : A \to X$ and $g : B \to X$ is a map $s : A \to B$ which produces a commutative triangle in $E$. The composition is pasting triangles together. The **fundamental theorem of topos theory** is:

**Theorem:** The slice category $E/X$ of a topos $E$ is a topos.

For example, if $E$ is the topos of sets, then the slice category is the category of **pointed sets**: the objects are then sets together with a function selecting a point as a "base point". A morphism $f : A \to B$ defines a functor $E/B \to E/A$ which preserves exponentials and the **subobject classifier** $\Omega$. Topos theory was motivated by geometry (Grothendieck), physics (Lawvere), topology (Tierney) and algebra (Kan). It can be seen as a generalization and even a replacement of set theory: the Lawvere's **elementary theory of the category of sets** ETCS is seen as part of ZFC which are less likely to be inconsistent [222]. For a short introduction [170], for textbooks [238, 58], for history of topos theory in particular, see [237].

## 55. Transcendentals

A **root** of an equation $f(x) = 0$ with integer polynomial $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ with $n \geq 0$ and $a_j \in \mathbb{Z}$ is called an **algebraic number**. The set $A$ of **algebraic numbers** is sub-field of the field $\mathbb{R}$ of **real numbers**. The field $A$ s the **algebraic closure** of the rational numbers $\mathbb{Q}$. It is of number theoretic interest as it contains all **algebraic number fields**, finite degree field extensions of $\mathbb{Q}$. The complement $\mathbb{R} \setminus A$ is the set of **transcendental numbers**. Transcendental numbers are necessarily irrational because every rational number $x = p/q$ is algebraic, solving $qx - p = 0$. Because the set of algebraic numbers is countable and the real numbers are not, most numbers are transcendental. The group of all automorphisms of $A$ which fix $\mathbb{Q}$ is called the **absolute Galois group** of $\mathbb{Q}$.

**Theorem:** $\pi$ and $e$ are transcendental

This result is due to Ferdinand von Lindemann. He proved that $e^x$ is transcendental for every non-zero algebraic number $x$. This immediately implies $e$ is transcendental. Now, if $\pi$ were algebraic, then $\pi i$ would be algebraic and $e^{i\pi} = -1$ would be transcendental. But $-1$ is rational. Lindemann's result was extended in 1885 by Karl Weierstrass to the statement telling that if $x_1, \dots x_n$ are linearly independent algebraic numbers, then $e^{x_1}, \dots e^{x_n}$ are algebraically independent. The transcendental property of $\pi$ also proves that $\pi$ is irrational. This is easier to prove directly. See [167].

## 56. Recurrence

A **homeomorphism** $T : X \to X$ of a compact topological space $X$ defines a **topological dynamical system** $(X, T)$. We write $T^j(x) = T(T(\dots T(x)))$ to indicate that the map $T$ is applied $j$ times. For any $d > 0$, we get from this a set $(T_1, T_2, \dots, T_d)$ of commuting homeomorphisms on $X$, where $T_j(x) = T^j x$. A point $x \in X$ is called **multiple recurrent** for $T$ if for every $d > 0$, there exists a sequence $n_1 < n_2 < n_3 < \cdots$ of integers $n_k \in \mathbb{N}$ for which $T_j^{n_k} x \to x$ for $k \to \infty$ and all $j = 1, \dots, d$. Fürstenberg's **multiple recurrence theorem** states:

**Theorem:** Every topological dynamical system is multiple recurrent.

It is known even that the set of multiple recurrent points are Baire generic. Hillel Fürstenberg proved this result in 1975. There is a parallel theorem for **measure preserving systems**: an automorphism $T$ of a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ is called **multiple recurrent** if there exists $A \in \mathcal{A}$ and an integer $n$ such that $\mathrm{P}[A \cap T_1(A) \cap \cdots \cap T_d(A)] > 0$. This generalizes the **Poincaré recurrence theorem**, which is the case $d = 1$. Recurrence theorems are related to the **Szemerédi theorem** telling that a subset $A$ of $\mathbb{N}$ of positive **upper density** contains arithmetic progressions of arbitrary finite length. See [129].

## 57. Solvability

A basic task in mathematics is to solve **polynomial equations** $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$ with complex coefficients $a_k$ using explicit formulas involving **roots**. One calls this an **explicit algebraic solution**. The linear case $ax + b = 0$ with $x = -b/a$, the quadratic case $ax^2 + bx + c = 0$ with $x = (-b \pm \sqrt{b^2 - 4ac})/(2a)$ were known since antiquity. The cubic $x^3 + ax^2 + bx + C = 0$ was solved by Niccolo Tartaglia and Cerolamo Cardano: a first substitution $x = X - a/3$ produces the **depressed cubic** $X^3 + pX + q$ (first solved by Scipione dal Ferro). The substitution $X = u - p/(3u)$ then produces a quadratic equation for $u^3$. Lodovico Ferrari solved finally the quartic by reducing it to the cubic. It was Paolo Ruffini, Niels Abel and Évariste Galois who realized that there are no algebraic solution formulas any more for polynomials of degree $n \geq 5$.

**Theorem:** Explicit algebraic solutions to $p(x) = 0$ exist if and only if $n \leq 4$.

The quadratic case was settled over a longer period in independent development in Babylonian, Egyptian, Chinese and Indian mathematics. The cubic and quartic discoveries were dramatic culminating with Cardano's book of 1545, marking the beginning of modern algebra. After centuries of failures of solving the quintic, Paolo Ruffini published the first proof in 1799, a proof which had a gap but who paved the way for Niels Hendrik Abel and Évariste Galois. For further discoveries see [235, 226, 9].

## 58. Galois theory

If $F$ is sub-field of $E$, then $E$ is a vector space over $F$. The dimension of this vector space is called the **degree** $[E : F]$ of the **field extension** $E/F$. The field extension is called **finite** if $[E : F]$ is finite. A field extension is called **transcendental** if there exists an element in $E$ which is not a root of an integral polynomial $f$ with coefficients in $F$. Otherwise, the extension is called **algebraic**. In the later case, there exists a unique monique polynomial $f$ which is irreducible over $F$ and the field extension is finite. An algebraic field extension $E/F$ is called **normal** if every irreducible polynomial over $K$ with at least one root in $E$ **splits** over $F$ into linear factors. An algebraic field extension $E/F$ is called **separable** if the associated irreducible polynomial $f$ is separable, meaning that $f'$ is not zero. This means, that $F$ has zero characteristic or that $f$ is not of the form $\sum_k a_k x^{pk}$ if $F$ has characteristic $p$. A field extension is called **Galois** if it normal and separable. Let **Fields**$(E/F)$ be the set of subfields of $E/F$ and **Groups**$(E/F)$) the set of subgroups of the automorphism group $\text{Aut}(E/F)$. The **Fundamental theorem of Galois theory** assures:

**Theorem: Fields**$(E/F) \overset{bijective}{\leftrightarrow}$ **Groups**$(E/F)$ if $E/F$ is Galois.

The **intermediate fields** of $E/F$ are so described by groups. It implies the **Abel-Ruffini theorem** about the non-solvability of the quintic by radicals. The fundamental theorem demonstrates that solvable extensions correspond to solvable groups. The **symmetry groups** of permutations of 5 or more elements are no more solvable. See [310].

## 59. Metric spaces

A **topological space** $(X, \mathcal{O})$ is given by a set $X$ and a finite collection $\mathcal{O}$ of subsets of $X$ with the property that the **empty set** $\emptyset$ and $\Omega$ both belong to $\mathcal{O}$ and that $\mathcal{O}$ is closed under arbitrary unions and finite intersections. The sets in $\mathcal{O}$ are called **open sets**. **Metric spaces** $(X, d)$ are special topological spaces. In that case, $\mathcal{O}$ consists of all sets $U$ such that for every $x \in U$ there exists $r > 0$ such that the **open ball** $B_r(x) = \{y \in X \mid d(x, y) < r\}$ is contained in $U$. Two topological spaces $(X, \mathcal{O})$, $(Y, \mathcal{Q})$ are **homeomorphic** if there exists a bijection $f : X \to Y$, such that $f$ and $f^{-1}$ are both continuous. A function $f : X \to Y$ is **continuous** if $f^{-1}(A) \in \mathcal{O}$ for all $A \in Q$. When is a topological space homeomorphic to a metric space? The **Urysohn metrization theorem** gives an answer: we need the **regular Hausdorff property** meaning that a closed set $K$ and a point $x$ can be separated by disjoint neighborhoods $K \subset U, y \in V$. We also need the space to be **second countable** meaning that there is a countable base (a base in $\mathcal{O}$ is a subset $\mathcal{B} \subset \mathcal{O}$ such that every $U \in \mathcal{O}$ can be written as a union of elements in $\mathcal{B}$.)

**Theorem:** A second countable regular Hausdorff space is metrizable.

The result was proven by Pavel Urysohn in 1925 with "regular" replaced by "normal" and by Andrey Tychonov in 1926. It follows that a compact Hausdorff space is metrizable if and only if it is second countable. For literature, see [44].

## 60. Fixed point

Given a continuous **transformation** $T : X \to X$ of a compact topological space $X$, one can look for the **fixed point** set $\text{Fix}_T(X) = \{x \mid T(x) = x\}$. This is useful for finding **periodic points** as fixed points of $T^n = T \circ T \circ T \cdots \circ T$ are periodic points of period $n$. If $X$ has a finite **cohomology** like if $X$ is a compact $d$-manifold with boundary, one can look at the **linear map** $T_p$ induced on the cohomology groups $H^p(X)$. The **super trace** $\chi_T(X) = \sum_{p=0}^d (-1)^p \text{tr}(T_p)$ is called the **Lefschetz number** of $T$ on $X$. If $T$ is the identity, this is the **Euler characteristic**. Let $\text{ind}_T(x)$ be the **Brouwer degree** of the map $T$ induced on a small $(d-1)$-sphere $S$ around $x$. This is the **trace** of the linear map $T_{d-1}$ induced from

$T$ on the cohomology group $H^{d-1}(S)$ which is an integer. If $T$ is differentiable and $dT(x)$ is invertible, the Brouwer degree is $\text{ind}_T(x) = \text{sign}(\det(dT))$. Let $\text{Fix}_T(X)$ denote the set of fixed points of $T$. The **Lefschetz-Hopf fixed point theorem** is

**Theorem:** If $\text{Fix}_T(X)$ is finite, then $\chi_T(X) = \sum_{x \in \text{Fix}_T(X)} \text{ind}_T(x)$.

A special case is the **Brouwer fixed point theorem**: if $X$ is a compact convex subset of Euclidean space. In that case $\chi_T(X) = 1$ and the theorem assures the existence of a fixed point. In particular, if $T : D \to D$ is a continuous map from the disc $D = \{x^2 + y^2 \leq 1\}$ onto itself, then $T$ has a fixed point. The **Brouwer fixed point theorem** was proved in 1910 by Jacques Hadamard and Luitzen Egbertus Jan Brouwer. The **Schauder fixed point theorem** from 1930 generalizes the result to convex compact subsets of Banach spaces. The Lefschetz-Hopf fixed point theorem was given in 1926. For literature, see [97, 39].

## 61. Quadratic reciprocity

Given a prime $p$, a number $a$ is called a **quadratic residue** if there exists a number $x$ such that $x^2$ has remainder $a$ modulo $p$. In other words quadratic residues are the squares in the field $\mathbb{Z}_p$. The **Legendre symbol** $(a|p)$ is defined by be 0 if $a$ is 0 or a multiple of $p$ and 1 if $a$ is a non-zero residue of $p$ and $-1$ if it is not. While the integer 0 is sometimes considered to be a quadratic residue we don't include it as it is a special case. Also, in the multiplicative group $\mathbb{Z}_p^*$ without zero, there is a symmetry: there are the same number of quadratic residues and non-residues. This is made more precise in the **law of quadratic reciprocity**

**Theorem:** For any two odd primes $(p|q)(q|p) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}$.

This means that $(p|q) = -(q|p)$ if and only if both $p$ and $q$ have remainder 3 modulo 4. The odd primes with of the form $4k + 3$ are also prime in the Gaussian integers. To remember the law, one can think of them as "Fermions" and quadratic reciprocity tells they Fermions are anti-commuting. The odd primes of the form $4k + 1$ factor by the **4-square theorem** in the Gaussian plane to $p = (a + ib)(a - ib)$ and are as a product of two Gaussian primes and are therefore Bosons. One can remember the rule because Boson commute both other particles so that if either $p$ or $q$ or both are "Bosonic", then $(p|q) = (q|p)$. The law of quadratic reciprocity was first conjectured by Euler and Legendre and published by Carl Friedrich Gauss in his Disquisitiones Arithmeticae of 1801. (Gauss found the first proof in 1796). [156, 167].

## 62. Quadratic map

Every quadratic map $z \to f(z) = z^2 + bz + d$ in the complex plane is conjugated to one of the quadratic family maps $T_c(z) = z^2 + c$. The **Mandelbrot set** $M = \{c \in \mathbb{C},\ T_c^n(0)$ stays bounded $\}$ is also called the **connectedness locus** of the quadratic family because for $c \in M$, the **Julia set** $J_c = \{z \in \mathbb{C}; T^n(z)$ stays bounded $\}$ is connected and for $c \notin M$, the Julia set $J_c$ is a **Cantor set**. The fundamental theorem for quadratic dynamical systems is:

**Theorem:** The Mandelbrot set is connected.

Mandelbrot first thought after seeing experiments that it was disconnected. The theorem is due to Adrien Duady and John Hubbard in 1982. One can also look at the connectedness locus for $T(z) = z^d + c$, which leads to **Multibrot sets** or the map $z \to \overline{z} + c$, which leads to the **tricorn** or **mandelbar** which is not path connected. One does not know whether the Mandelbrot set $M$ is locally connected, nor whether it is path connected. See [242, 59, 27]

## 63. Differential equations

Let us say that a differential equation $x'(t) = F(x)$ is **integrable** if a trajectory $x(t)$ either converges to infinity, or to an **equilibrium point** or to a **limit cycle** or **limiting torus**, where it is a periodic or almost periodic trajectory. We assume $F$ has global solutions. The **Poincaré-Bendixon** theorem is:

> **Theorem:** Any differential equation in the plane is integrable.

This changes in dimensions 3 and higher. The **Lorenz attractor** or the **Rössler attractor** are examples of **strange attractors**, limit sets on which the dynamics can have positive topological entropy and is therefore no more integrable. The theorem also does not hold any more on two dimensional tori as there can be recurrent non-periodic orbits and even weak mixing. The proof of the Poincaré-Bendixon theorem relies on the Jordan curve theorem. [72, 184].

## 64. Approximation theory

A function $f$ on a closed interval $I = [a, b]$ is called **continuous** if for every $\epsilon > 0$ there exists a $\delta > 0$ such that if $|x - y| < \delta$ then $|f(x) - f(y)| < \epsilon$. In the space $X = C(I)$ of all continuous functions, one can define a distance $d(f, g) = \max_{x \in I} |f(x) - g(x)|$. A subset $Y$ of $X$ is called **dense** if for every $\epsilon > 0$ and every $x \in X$, there exists $y \in Y$ with $d(x, y) < \epsilon$. Let $P$ denote the class of **polynomials** in $X$. The **Weierstrass approximation theorem** tells that

> **Theorem:** $P$ is dense in $C(I)$.

The Weierstrass theorem has been proven in 1885 by Karl Weierstrass. A constructive proof uses **Bernstein polynomials** $f_n(x) = \sum_{k=0}^{n} f(k/n) B_{k,n}(x)$ with $B_{k,n}(x) = B(n, k) x^k (1 - x)^{n-k}$, where $B(n, k)$ denotes the Binomial coefficients. The result has been generalized to compact Hausdorff spaces $X$ and more general subalgebras of $C(X)$. The **Stone-Weierstrass approximation theorem** was proven by Marshall Stone in 1937 and simplified in 1948 by Stone. In the complex, there is **Runge's theorem** from 1885 approximating functions holmomorphic on a bounded region $G$ with rational functions uniformly on a compact subset $K$ of $G$ and **Mergelyan's theorem** from 1951 allowing approximation uniformly on a compact subset with polynomials if the region $G$ is simply connected. In **numerical analysis** one has the task to approximate a given function space by functions from a simpler class. Examples are approximations of smooth functions by polynomials, trigonometric polynomials. There is also the **interpolation problem** of approximating a given data set with polynomials or piecewise polynomials like **splines** or **Bézier curves**. See [324, 253].

## 65. Diophantine approximation

An **algebraic number** is a root of a polynomial $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ with **integer coefficients** $a_k$. A real number $x$ is called **Diophantine** if there exists $\epsilon > 0$ and a positive constant $C$ such that the **Diophantine condition** $|x - p/q| > C/q^{2+\epsilon}$ is satisfied for all $p$, and all $q > 0$. **Thue-Siegel-Roth theorem** tells:

> **Theorem:** Any irrational algebraic number is Diophantine.

The **Hurwitz's theorem** from 1891 assures that there are infinitely many $p, q$ with $|x-p/q| < C/q^2$ for $C = 1/\sqrt{5}$. This shows that the Tue-Siegel-Roth Theorem can not be extended to $\epsilon = 0$. The **Hurwitz constant** $C$ is optimal. For any $C < 1/\sqrt{5}$ one can with the **golden ratio** $x = (1+\sqrt{5})/2$ have only finitely many $p, q$ with $|x - p/q| < C/q^2$. The set of **Diophantine numbers** has full Lebesgue measure. A slightly larger set is the **Brjuno set** of all numbers for which the continued fraction **convergent** $p_n/q_n$ satisfies $\sum_n \log(q_{n+1})/q_n < \infty$. A Brjuno rotation number assures the **Siegel linearization theorem** still can be proven. For quadratic polynomials, Jean-Christophe Yoccoz showed that linearizability implies the rotation number must be Brjuno. [59, 162]

## 66. Almost periodicity

If $\mu$ is a **probability measure** of compact support on $\mathbb{R}$, then $\hat{\mu}_n = \int e^{inx} \; d\mu(x)$ are the **Fourier coefficients** of $\mu$. The **Riemann-Lebesgue lemma** tells that if $\mu$ is absolutely continuous, then $\hat{\mu}_n$ goes to zero. The pure point part can be detected with the following **Wiener theorem**:

**Theorem:** $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} |\hat{\mu}_k|^2 = \sum_{x \in \mathbb{T}} |\mu(\{x\})|^2$.

This looks a bit like the **Poisson summation formula** $\sum_n f(n) = \sum_n \hat{f}(n)$, where $\hat{f}$ is the Fourier transform of $f$. [The later follows from $\sum_n e^{2\pi ikx} = \sum_n \delta(x-n)$, where $\delta(x)$ is a Dirac delta function. The Poisson formula holds if $f$ is uniformly continuous and if both $f$ and $\hat{f}$ satisfy the growth condition $|f(x)| \leq C/|1+|x||^{1+\epsilon}$. ] More generally, one can read off the **Hausdorff dimension** from decay rates of the Fourier coefficients. See [190].

## 67. Shadowing

Let $T$ be a **diffeomorphism** on a smooth **Riemannian manifold** $M$ with geodesic metric $d$. A $T$**-invariant set** is called **hyperbolic** if for each $x \in K$, the tangent space $T_xM$ splits into a **stable and unstable bundle** $E_x^+ \oplus E_x^-$ such that for some $0 < \lambda < 1$ and constant $C$, one has $dTE_x^{\pm} = E_{Tx}^{\pm}$ and $|dT^{\pm n}v| \leq C\lambda^n$ for $v \in E^{\pm}$ and $n \geq 0$. An $\epsilon$**-orbit** is a sequence $x_n$ of points in $M$ such that $x_{n+1} \in B_\epsilon(T(x_n))$, where $B_\epsilon$ is the geodesic ball of radius $\epsilon$. Two sequences $x_n, y_n \in M$ are called $\delta$**-close** if $d(y_n, x_n) \leq \delta$ for all $n$. We say that a set $K$ has the **shadowing property**, if there exists an open neighborhood $U$ of $K$ such that for all $\delta > 0$ there exists $\epsilon > 0$ such that every $\epsilon$-pseudo orbit of $T$ in $U$ is $\delta$-close to true orbit of $T$.

**Theorem:** Every hyperbolic set has the shadowing property.

This is only interesting for infinite $K$ as if $K$ is a finite periodic hyperbolic orbit, then the orbit itself is the orbit. It is interesting however for a hyperbolic invariant set like a **Smale horse shoe** or in the **Anosov case**, when the entire manifold is hyperbolic. See [184].

## 68. Partition function

Let $p(n)$ denote the number of ways we can write $n$ as a sum of positive integers without distinguishing the order. Euler used its **generating function** which is $\sum_{n=0}^{\infty} p(n)x^n = \prod_{k=1}^{\infty}(1-x^k)^{-1}$. The reciprocal function $(1-x)(1-x^2)+(1-x^3)\cdots$ is called the **Euler function** and generates the **generalized Pentagonal number theorem** $\sum_{k\in\mathbb{Z}}(-1)^k x^{k(3k-1)/2} = 1 - x - x^2 + x^5 - x^7 - x^{12} - x^{15} \cdots$ leading to the recursion $p(n) = p(n-1) + p(n-2) - p(n-5) - p(n-7) + p(n-12) + p(n-15) \cdots$. The **Jacobi triple product** identity is

**Theorem:** $\prod_{n=1}^{\infty}(1-x^{2m})(1-x^{2m-1}y^2)(1-x^{2m-1}y^{-2}) = \sum_{n=-\infty}^{\infty} x^{n^2}y^{2n}$.

The formula was found in 1829 by Jacobi. For $x = z\sqrt{z}$ and $y^2 = -\sqrt{z}$ the identity reduces to the **pentagonal number theorem**. See [15].

## 69. Burnside lemma

If $G$ is a finite group acting on a finite set $X$, let $X/G$ denote the number of disjoint **orbits** and $X^g = \{x \in X \mid g.x = x, \forall g \in G\}$ the **fixed point set** of elements which are fixed by $g$. The number $|X/G|$ of orbits and the **group order** $|G|$ and the size of the **fixed point sets** are related by the **Burnside lemma**:

**Theorem:** $|X/G| = \frac{1}{|G|}\sum_{g\in G}|X^g|$

The result was first proven by Frobenius in 1887. Burnside popularized it in 1897 [55].

## 70. Taylor series

A complex-valued function $f$ which is **analytic** in a disc $D = D_r(a) = \{|x-a| < r\}$ can be written as a series involving the $n$'th derivatives $f^{(n)}(a)$ of $f$ at $a$. If $f$ is real valued on the real axes, the function is called **real analytic** in $(x-a, x+a)$. In several dimensions we can use multi-index notation $a = (a_1, \dots, a_d)$, $n = (n_1, \dots, n_d)$, $x = (x_1, \dots, x_d)$ and $x^n = x_1^{n_1} \cdots x_d^{n_d}$ and $f^{(n)}(x) = \partial_{x_1}^{n_1} \cdots \partial_{x_d}^{n_d}$ and use a **polydisc** $D = D_r(a) = \{|x_1 - a_1| < r_1, \dots |x_d - a_d| < r_d\}$. The **Taylor series formula** is:

**Theorem:** For analytic $f$ in $D$, $f(x) = \sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(x-a)^n$.

Here, $T_r(a) = \{|x_i - a_1| = r_1 \dots |x_d - a_d| = r_d\}$ is the boundary torus. For example, for $f(x) = \exp(x)$, where $f^{(n)}(0) = 1$, one has $f(x) = \sum_{n=0}^{\infty} x^n/n!$. Using the **differential operator** $Df(x) = f'(x)$, one can see $f(x+t) = \sum_{n=0}^{\infty}\frac{f^{(n)}(x)}{n!}t^n = e^{Dt}f(x)$ as a solution of the **transport equation** $f_t = Df$. One can also represent $f$ as a **Cauchy formula** for polydiscs $1/(2\pi i)^d \int_{|T_r(a)|} f(z)/(z-a)^d dz$ integrating along the boundary torus. Finite Taylor series hold in the case if $f$ is $m+1$ times differentiable. In that case one has a finite series $S(x) = \sum_{n=0}^{m}\frac{f^{(n)}(a)}{n!}(x-a)^n$ such that the **Lagrange rest term** is $f(x) - S(x) = R(x) = f^{m+1}(\xi)(x-a)^{m+1}/((m+1)!)$, where $\xi$ is between $x$ and $a$. This generalizes the **mean value theorem** in the case $m = 0$, where $f$ is only differentiable. The remainder term can also be written as $\int_a^x f^{(m+1)}(s)(x-a)^m/m!\, ds$. Taylor did state but not justify the formula in 1715 which was actually a difference formula. 1742 Maclaurin uses the modern form. [212].

## 71. Isoperimetric inequality

Given a smooth $S$ in $\mathbb{R}^n$ homeomorphic to a sphere and bounding a region $B$. Assume that the **surface area** $|S|$ is fixed. How large can the **volume** $|B|$ of $B$ become? If $B$ is the unit ball $B_1$ with volume $|B_1|$ the answer is given by the **isoperimetric inequality**:

**Theorem:** $n^n|B|^{n-1} \le |S|^n/|B_1|$.

If $B = B_1$, this gives $n|B| \leq |S|$, which is an equality as then the **volume of the ball** $|B| = \pi^{n/2}/\Gamma(n/2+1)$ and the **surface area of the sphere** $|S| = n\pi^{n/2}/\Gamma(n/2+1)$ which Archimedes first got in the case $n = 3$, where $|S| = 4\pi$ and $|B| = 4\pi/3$. The classical **isoperimetric problem** is $n = 2$, where we are in the plane $\mathbb{R}^2$. The inequality tells then $4|B| \leq |S|^2/\pi$ which means $4\pi\text{Area} \leq \text{Length}^2$. The ball $B_1$ with area 1 maximizes the functional. For $n = 3$, with usual Euclidean space $\mathbb{R}^3$, the inequality tells $|B|^2 \leq (4\pi)^3/(27 \cdot 4\pi/3)$ which is $|B| \leq 4\pi/3$. The first proof in the case $n = 2$ was attempted by Jakob Steiner in 1838 using the **Steiner symmetrization** process which is a refinement of the **Archimedes-Cavalieri principle**. In 1902 a proof by Hurwitz was given using Fourier series. The result has been extended to geometric measure theory [121]. One can also look at the discrete problem to maximize the area defined by a polygon: if $\{(x_i, y_i), i = 0, \dots n-1\}$ are the points of the polygon, then the area is given by Green's formula as $A = \sum_{i=0}^{n-1} x_i y_{i+1} - x_{i+1} y_i$ and the length is $L = \sum_{i=0}^{n-1} (x_i - x_{i+1})^2 + (y_i - y_{i+1})^2$ with $(x_n, y_n)$ identified with $(x_0, y_0)$. The **Lagrange equations** for $A$ under the constraint $L = 1$ together with a fix of $(x_0, y_0)$ and $(x_1 = 1/n, 0)$ produces two maxima which are both **regular polygons**. A generalization to $n$-dimensional Riemannian manifolds is given by the Lévi-Gromov isoperimetric inequality.

## 72. Riemann Roch

A Riemann surface is a one-dimensional complex manifold. It is a two-dimensional real analytic manifold but it has also a **complex structure** forcing it to be orientable for example. Let $G$ be a compact connected **Riemann surface** of Euler characteristic $\chi(G) = 1 - g$, where $g = b_1(G)$ is the **genus**, the number of handles of $G$ (and $1 = b_0(G)$ indicates that we have only one connected component). A **divisor** $D = \sum_i a_i z_i$ on $G$ is an element of the free Abelian group on the points of the surface. These are finite formal sums of points in $G$, where $a_i \in \mathbb{Z}$ is the multiplicity of the point. Its **degree** is defined as $\deg(X) = \sum_i a_i$. Let us write $\chi(X) = \deg(D) + \chi(G) = \deg(D) + 1 - g$ and call this the **Euler characteristic** of the divisor as one can see a divisor as a geometric object by itself generalizing the complex manifold itself, where $D = 0$. A **meromorphic function** $f$ on $G$ defines the **principal divisor** $(f) = \sum_i a_i z_i - \sum_j b_j w_j$, where $a_i$ are the multiplicities of the **roots** $z_i$ of $f$ and $b_j$ the multiplicities of the **poles** $w_j$ of $f$. The principal divisor of a global meromorphic 1-form $dz$ is called the **canonical divisor** $K$. Let $l(D)$ be the dimension of the linear space of meromorphic functions $f$ on $G$ for which $(f) + D \geq 0$ (meaning that all coefficients are non-negative, one calls this **effective**). The **Riemann-Roch** theorem is

**Theorem:** $l(D) - l(K - D) = \chi(D)$

The idea of a Riemann surfaces was defined by Bernhard Riemann. Riemann-Roch was proven for Riemann surfaces by Bernhard Riemann in 1857 and Gustav Roch in 1865. It is possible to see this as a **Euler-Poincaré type relation** by identifying the left hand side as a signed cohomological Euler characteristic and the right hand side as a combinatorial Euler characteristic. There are various generalizations, to arithmetic geometry or to higher dimensions. See [142, 288].

## 73. Optimal transport

Given two probability spaces $(X, P), (Y, Q)$ and a continuous **cost function** $c : X \times Y \to [0, \infty]$, the **optimal transport problem** or **Monge-Kantorovich minimization problem** is to find the minimum of $\int_X c(x, T(x))\, dP(x)$ among all **coupling transformations** $T : X \to Y$ which have the property that it transports the measure $P$ to the measure $Q$. More generally, one looks at a measure $\pi$ on $X \times Y$ such that the projection of $\pi$ onto $X$ it is $P$ and the projection of $\pi$ onto $Y$ is $Q$. The function to optimize is then $I(\pi) = \int_{X \times Y} c(x, y)\, d\pi(x, y)$. One of the fundamental results is that optimal transport exists. The technical assumption is that if the two probability spaces $X, Y$ are **Polish** (=separable complete metric spaces) and that the cost function $c$ is continuous.

**Theorem:** For continuous cost functions $c$, there exists a minimum of $I$.

In the simple set-up of probability spaces, this just follows from the compactness (Alaoglu theorem for balls in the weak star topology of a Banach space) of the set of probability measures: any sequence $\pi_n$ of probability measures on $X \times Y$ has a convergent subsequence. Since $I$ is continuous, picking a sequence $\pi_n$ with $I(\pi_n)$ decreasing produces to a minimum. The problem was formalized in 1781 by Gaspard Monge and worked on by Leonid Kantorovich. Tanaka in the 1970ies produced connections with partial differential equations like the Bolzmann equation. There are also connections to **weak KAM theory** in the form of Aubry-Mather theory. The above existence result is true under substancial less regularity. The question of uniqueness or the existence of a Monge coupling given in the form of a transformation $T$ is subtle [328].

## 74. Structure from motion

Given $m$ hyper planes in $\mathbf{R}^d$ serving as retinas or photographic plates for **affine cameras** and $n$ points in $\mathbf{R}^d$. The **affine structure from motion** problem is to understand under which conditions it is possible to recover both the points and planes when knowing the orthogonal projections onto the planes. It is a model problem for the task to reconstruct both the scene as well as the camera positions if the scene has $n$ points and $m$ camera pictures were taken. Ullman's theorem is a prototype result with $n = 3$ different cameras and $m = 3$ points which are not collinear. Other setups are **perspective cameras** or **omni-directional cameras**. The **Ullman** map $F$ is a nonlinear map from $R^{d \cdot 2} \times SO_d^2$ to $(R^{3d-3})^2$ which is a map between equal dimensional spaces if $d = 2$ and $d = 3$. The group $SO_d$ is the rotation group in $\mathbb{R}$ describing the possible ways in which the affine camera can be positioned. Affine cameras capture the same picture when translated so that the planes can all go through the origin. In the case $d = 2$, we get a map from $R^4 \times SO_2^2$ to $R^6$ and in the case $d = 3$, $F$ maps $\mathbf{R}^6 \times SO_3^2$ into $\mathbf{R}^{12}$.

**Theorem:** The structure from motion map is locally invertible.

In the case $d = 2$, there is a reflection ambiguity. In dimension $d = 3$, the number of ambiguities is typically 64. Ullman's theorem appeared in 1979 in [326]. Ullman states the theorem for d=3 with 4 points as adding a four point cuts the number of ambiguities from 64 to 2. See [205] both in dimension d=2 and d=3 the Jacobean $dF$ of the Ullman map is seen to be invertible and the inverse of $F$ is given explicitly. For structure from motion problems in computer vision in general, see [120, 157, 325]. In applications one takes $n$ and $m$ large and reconstructs both the points as well as the camera parameters using **statistical data fitting**.

## 75. Poisson equation

What functions $u$ solve the **Poisson equation** $-\Delta u = f$, a partial differential equation? The right hand side can be written down for $f \in L^1$ as $K_f(x) = \int_{\mathbb{R}^n} G(x, y) f(y)\, dy + h$, where $h$ is

**harmonic**. If $f = 0$, then the Poisson equation is the **Laplace equation**. The function $G(x, y)$ is the **Green's function**, an **integral kernel**. It satisfies $-\Delta G(x, y) = \delta(y - x)$, where $\delta$ is the **Dirac delta function**, a distribution. It is given by $G(x, y) = -\log|x - y|/(2\pi)$ for $n = 2$ or $G(x, y) = |x - y|^{-1}/(4\pi)$ for $n = 3$. In **elliptic regularity theory**, one replaces the Laplacian $-\Delta$ with an **elliptic** second order **differential operator** $L = A(x) \cdot D \cdot D + b(x) \cdot D + V(x)$ where $D = \nabla$ is the gradient and $A$ is a positive definite matrix, $b(x)$ is a vector field and $c$ is a scalar field.

**Theorem:** For $f \in L^p$ and $p > n$, then $K_f$ is differentiable.

The result is much more general and can be extended. If $f$ is in $C^k$ and has compact support for example, then $K_f$ is in $C^{k+1}$. An example of the more general set up is the **Schrödinger operator** $L = -\Delta + V(x) - E$. The solution to $Lu = 0$, solves then an eigenvalue problem. As one looks for solutions in $L^2$, the solution only exists if $E$ is an **eigenvalue** of $L$. The Euclidean space $\mathbb{R}^n$ can be replaced by a bounded domain $\Omega$ of $\mathbb{R}^n$ where one can look at boundary conditions like of Dirichlet or von Neumann type. Or one can look at the situation on a general Riemannian manifold $M$ with or without boundary. On a Hilbert space, one has then **Fredholm theory**. The equation $u = \int G(x, y)f(y)dy$ is called a **Fredholm integral equation** and $\det(1 - sG) = \exp(-\sum_n s^n \text{tr}(G^n)/n!)$ the **Fredholm determinant** leading to the **zeta function** $1/\det(1 - sG)$. See [274, 225].

## 76. Four square theorem

**Waring's problem** asked whether there exists for every $k$ an integer $g(k)$ such that every positive integer can be written as a sum of $g(k)$ powers $x_1^k + \cdots + x_{g(k)}^k$. Obviously $g(1) = 1$. David Hilbert proved in 1909, that $g(k)$ is finite. This is the **Hilbert-Waring theorem**. The following **theorem of Lagrange** tells that $g(2) = 4$:

**Theorem:** Every positive integer is a sum of four squares

.

The result needs only to be verified for prime numbers as $N(a, b, c, d) = a^2 + b^2 + c^2 + d^2$ is a norm for **quaternions** $q = (a, b, c, d)$ which has the property $N(pq) = N(p)N(q)$. This property can be seen also as a **Cauchy-Binet formula**, when writing quaternions as complex $2 \times 2$ matrices. The four-square theorem had been conjectured already by Diophantus, but was proven first by Lagrange in 1770. The case $g(3) = 9$ was done by Wieferich in 1912. It is conjectured that $g(k) = 2^k + [(3/2)^k] - 2$, where $[x]$ is the integral part of a real number. See [91, 92, 167].

## 77. Knots

A **knot** is a closed curve in $\mathbb{R}^3$, an embedding of the circle in three dimensional Euclidean space. One also draws knots in the 3-sphere $S^3$. As the **knot complement** $S^3 - K$ of a knot $K$ characterizes the knot up to mirror reflection, the theory of knots is part of **3-manifold theory**. The **HOMFLYPT** polynomial $P$ of a knot or **link** $K$ is defined recursively using **skein relations** $lP(L_+) + l^{-1}P(L^-) + mP(L_0) = 0$. Let $K\#L$ denote the **knot sum** which is a **connected sum**. Oriented knots form with this operation a commutative monoid with **unknot** as unit and which features a unique prime factorization. The **unknot** has $P(K) = 1$, the **unlink** has $P(K) = 0$. The **trefoil knot** has $P(K) = 2l^2 - l^4 + l^2m^2$.

**Theorem:** $P(K\#L) = P(K)P(L)$.

The **Alexander polynomial** was discovered in 1928 and initiated classical knot theory. John Conway showed in the 60ies how to compute the Alexander polynomial using a recursive **skein relations** (skein comes from French escaigne=hank of yarn). The Alexander polynomial allows to compute an invariant for knots by looking at the projection. The Jones polynomial found by Vaughan Jones came in 1984. This is generalized by the HOMFLYPT polynomial named after Jim Hoste, Adrian Ocneanu, Kenneth Millett, Peter J. Freyd and W.B.R. Lickorish from 1985 and J. Przytycki and P. Traczyk from 1987. See [4]. Further invariants are **Vassiliev invariants** of 1990 and **Kontsevich invariants** of 1993.

## 78. Hamiltonian dynamics

Given a probability space $(M, \mathcal{A}, m)$ and a smooth Lie manifold $N$ with potential function $V : N \to \mathbb{R}$, the **Vlasov Hamiltonian differential equations** on all maps $X = (f, g) : M \to T^*N$ is $f' = g, g' = \int_N \nabla V(f(x) - f(y))\ dm(y)$. Starting with $X_0 = Id$, we get a flow $X_t$ and by push forward an evolution $P^t = X_t^* m$ of probability measures on $N$. The Vlasov intro-differential equations on measures in $T^*N$ are $\dot{P}^t(x, y) + y \cdot \nabla_x P^t(x, y) - W(x) \cdot \nabla_y P^t(x, y) = 0$ with $W(x) = \int_M \nabla_x V(x - x')P^t(x', y'))\ dy'dx'$. Note that while $X_t$ is an infinite dimensional **ordinary differential equations** evolving maps $M \to T^*N$, the path $P^t$ is an **integro differential equation** describing the evolution of measures on $T^*N$.

**Theorem:** If $X_t$ solves the Vlasov Hamiltonian, then $P^t = X_t^* m$ solves Vlasov.

This is a result which goes back to James Clerk Maxwell. Vlasov dynamics was introduced in 1938 by Anatoly Vlasov. An existence result was proven by W. Brown and Klaus Hepp in 1977. The maps $X_t$ will stay perfectly smooth if smooth initially. However, even if $P^0$ is smooth, the measure $P^t$ in general rather quickly develops singularities so that the partial differential equation has only **weak solutions**. The analysis of $P$ directly would involve complicated function spaces. The **fundamental theorem of Vlasov dynamics** therefore plays the role of the **method of characteristics** in this field. If $M$ is a finite probability space, then the Vlasov Hamiltonian system is the **Hamiltonian $n$-body problem** on $N$. An other example is $M = T^*N$ and where $m$ is an initial phase space measure. Now $X_t$ is a one parameter family of diffeomorphisms $X_t : M \to T^*N$ pushing forward $m$ to a measure $P^t$ on the cotangent bundle. If $M$ is a circle then $X^0$ defines a closed curve on $T^*N$. In particular, if $\gamma(t)$ is a curve in $N$ and $X^0(t) = (\gamma(t), 0)$, we have a continuum of particles initially at rest which evolve by interacting with a force $\nabla V$. About interacting particle dynamics, see [305].

## 79. Hypercomplexity

A **hypercomplex algebra** is a finite dimensional algebra over $\mathbb{R}$ which is **unital** and distributive. The classification of hypercomplex algebras (up to isomorphism) of two-dimensional hypercomplex algebras over the reals are the **complex numbers** $x + iy$ with $i^2 = -1$, the **split complex numbers** $x + jy$ with $j^2 = -1$ and the **dual numbers** (the exterior algebra) $x + \epsilon y$ with $\epsilon^2 = 0$. A **division algebra** over a field $F$ is an algebra over $F$ in which division is possible. **Wedderburn's little theorem** tells that a finite division algebra must be a finite field. Only $\mathbb{C}$ is the only two dimensional **division algebra** over $\mathbb{R}$. The following theorem of Frobenius classifies the class $\mathcal{X}$ of finite dimensional associative division algebras over $\mathbb{R}$:

**Theorem:** $\mathcal{X}$ consists of the algebras $\mathbb{R}, \mathbb{C}$ and $\mathbb{H}$.

Hypercomplex numbers like **quaternions**, **tessarines** or **octonions** extend the algebra of complex numbers. Cataloging them started with Benjamin Peirce 1872 "Linear associative algebra". **Dual numbers** were introduced in 1873 by William Clifford. The **Cayley-Dickson constructions** generates iteratively algebras of twice the dimensions: like the complex numbers from the reals, the quaternions from the complex numbers or the octonions from the quaternions. The next step leads to **sedenions** but multiplicativity is lost. The Hurwitz and Frobenius theorems limit the number in the case of normed division algebras. Ferdinand George Frobenius classified in 1877 the finite-dimensional associative division algebras. Adolf Hurwitz proved in 1923 (posthumously) that unital finite dimensional real algebra endowed with a positive-definite quadratic form (a **normed division algebra** must be $\mathbb{R}, \mathbb{C}, \mathbb{H}$ or $\mathbb{O}$). These four are the only **Euclidean Hurwitz algebras**. In 1907, Joseph Wedderburn classified simple algebras (simple meaning that there are no non-trivial two-sided ideals and $ab = 0$ implies $a = 0$ or $b = 0$). In 1958 J. Frank Adams showed topologically that $\mathbb{R}, \mathbb{C}, \mathbb{H}, \mathbb{O}$ are the only finite dimensional real division algebras. In general, division algebras have dimension $1, 2, 4$ or $8$ as Michel Kervaire and Raoul Bott and John Milnor have shown in 1958 by relating the problem to the **parallelizability of spheres**. The problem of classification of division algebras over a field $F$ led Richard Brauer to the **Brauer group** $BR(F)$, which Jean Pierre Serre identified it with **Galois cohomology** $H^2(K, K^*)$, where $K^*$ is the multiplicative group of $K$ seen as an algebraic group. Each Brauer equivalence class among central simple algebras (**Brauer algebras**) contains a unique division algebra by the Artin-Wedderburn theorem. Examples: the Brauer group of an algebraically closed field or finite field is trivial, the Brauer group of $\mathbb{R}$ is $Z_2$. Brauer groups were later defined for commutative rings by Maurice Auslander and Oscar Goldman and by Alexander Grothendieck in 1968 for schemes. Ofer Gabber extended the Serre result to schemes with ample line bundles. The finiteness of the Brauer group of a proper integral scheme is open. See [23, 118].

## 80. Approximation

The **Kolmogorov-Arnold superposition theorem** shows that **continuous functions** $C(\mathbb{R}^n)$ of several variables can be written as a composition of continuous functions of two variables:

**Theorem:** Every $f \in C(\mathbb{R}^n)$ composition of continuous functions in $C(\mathbb{R}^2)$.

More precisely, it is now known since 1962 that there exist functions $f_{k,l}$ and a function $g$ in $C(\mathbb{R})$ such that $f(x_1, \dots, x_n) = \sum_{k=0}^{2n} g(f_{k,1}(x_1) + \cdots + f_{k,n}x_n)$. As one can write finite sums using functions of two variables like $h(x, y) = x + y$ or $h(x + y, z) = x + y + z$ two variables suffice. The above form was given by by George Lorentz in 1962. Andrei Kolmogorov reduced the problem in 1956 to functions of three variables. Vladimir Arnold showed then in 1957 that one can do with two variables. The problem came from a more specific problem in algebra, the problem of finding roots of a polynomial $p(x) = x^n + a_1 x^{n-1} + \cdots a_n$ using radicals and arithmetic operations in the coefficients is not possible in general for $n \geq 5$. Erland Samuel Bring shows in 1786 that a quintic can be reduced to $x^5 + ax + 1$. In 1836 William Rowan Hamilton showed that the sextic can be reduced to $x^6 + ax^2 + bx + 1$ to $x^7 + ax^3 + bx^2 + cx + 1$ and the degree 8 to a 4 parameter problem $x^8 + ax^4 + bx^3 + cx^2 + dx + 1$. Hilbert conjectured that one can not do better. They are the **Hilbert's 13th problem**, the **sextic conjecture** and **octic conjecture**. In 1957, Arnold and Kolmogorov showed that no topological obstructions exist to reduce the number of variables. Important progress was done in 1975 by Richard Brauer. Some history is given in [119]:

## 81. Determinants

The **determinant** of a $n \times n$ matrix $A$ is defined as the sum $\sum_\pi (-1)^{\text{sign}(\pi)} A_{1\pi(1)} \cdots A_{n\pi(n)}$, where the sum is over all $n!$ permutations $\pi$ of $\{1, \dots, n\}$ and $\text{sign}(\pi)$ is the **signature** of the permutation $\pi$. The determinant functional satisfies the **product formula** $\det(AB) = \det(A)\det(B)$. As the determinant is the constant coefficient of the **characteristic polynomial** $p_A(x) = \det(A - x1) = p_0(-x)^n + p_1(-x)^{n-1} + \cdots + p_k(-x)^{n-k} + \cdots + p_n$ of $A$, one can get the coefficients of the product $F^T G$ of two $n \times m$ matrices $F, G$ as follows:

**Theorem:** $p_k = \sum_{|P|=k} \det(F_P) \det(G_P)$.

The right hand side is a sum over all minors of length $k$ including the empty one $|P| = 0$, where $\det(F_P) \det(G_P) = 1$. This implies $\det(1 + F^T G) = \sum_P \det(F_P) \det(G_P)$ and so $\det(1 + F^T F) = \sum_P \det^2(F_P)$. The classical Cauchy-Binet theorem is the special case $k = m$, where $\det(F^T G) = \sum_P \det(F_P)\det(G_P)$ is a sum over all $m \times m$ patterns if $n \geq m$. It has as even more special case the Pythagorean consequence $\det(A^T A) = \sum_P \det(A_P^2)$. The determinant product formula is the even more special case when $n = m$. [171, 201, 166].

## 82. Triangles

A **triangle** $T$ on a surface $S$ consists of three points $A, B, C$ joined by three geodesic paths. If $\alpha, \beta, \gamma$ are the **inner angles** of a **triangle** $T$ located on a surface with **curvature** $K$, there is the Gauss-Bonnet formula $\int_S K(x) dA(x) = \chi(S)$, where $dA$ denotes the **area element** on the surface. This implies a relation between the integral of the curvature over the triangle and the angles:

**Theorem:** $\alpha + \beta + \gamma = \int_T K \, dA + \pi$

This can be seen as a special Gauss-Bonnet result for **Riemannian manifolds with boundary** as it is equivalent to $\int_T K \; dA + \alpha' + \beta + \gamma' = 2\pi$ with **complementary angles** $\alpha' = \pi - \alpha, \beta' = \pi - \beta, \gamma' = \pi - \gamma$. One can think of the vertex contributions as **boundary curvatures** (generalized function). In the case of **constant curvature** $K$, the formula becomes $\alpha + \beta + \gamma = KA + \pi$, where $A$ is the **area of the triangle**. Since antiquity, one knows the flat case $K = 0$, where $\pi = \alpha + \beta + \gamma$ taught in elementary school. On the **unit sphere** this is $\alpha + \beta + \gamma = A + \pi$, result of Albert Girard which was predated by Thomas Harriot. In the **Poincaré disk model** $K = -1$, this is $\alpha + \beta + \gamma = -A + \pi$ which is usually stated that the area of a triangle in the disk is $\pi - \alpha - \beta - \gamma$. This was proven by Johann Heinrich Lambert. See [51] for spherical geometry and [14] for hyperbolic geometry, which are both part of **non-Euclidean geometry** and now part of **Riemannian geometry**. [33, 179]

## 83. KAM

An **area preserving map** $T(x, y) = (2x - y + cf(x), x)$ has an orbit $(x_{n+1}, x_n)$ on $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ which satisfies the recursion $x_{n+1} - 2x_n + x_{n-1} = cf(x_n)$. The 1-periodic function $f$ is assumed to be real-analytic, non-constant satisfying $\int_0^1 f(x) \; dx = 0$. In the case $f(x) = \sin(2\pi x)$, one has the **Standard map**. When looking for invariant curves $(q(t + \alpha), q(t))$ with smooth $q$, we seek a solution of the nonlinear equation $F(q) = q(t + \alpha) - 2q(t) + q(t - \alpha) - cf(q(t)) = 0$. For

$c = 0$, there is the solution $q(t) = t$. The **linearization** $dF(q)(u) = Lu = u(t+\alpha) - 2u(t) + u(t-\alpha) - cf'(q(t))u(t)$ is a bounded linear operator on $L^2(\mathbb{T})$ but not invertible for $c = 0$ so that the **implicit function theorem** does not apply. The map $Lu = u(t+\alpha)-2u(t)+u(t-\alpha)$ becomes after a Fourier transform the diagonal matrix $\hat{L}\hat{u}_n = [2\cos(n\alpha) - 2]\hat{u}_n$ which has the inverse diagonal entries $[2\cos(n\alpha) - n]^{-1}$ leading to **small divisors**. A real number $\alpha$ is called **Diophantine** if there exists a constant $C$ such that for all integers $p, q$ with $q \neq 0$, we have $|\alpha - p/q| \geq C/q^2$. **KAM theory** assures that the solution $q(t) = t$ persists and remains smooth if $c$ is small. With **solution** the theorem means a **smooth solution**. For real analytic $F$, it can be real analytic. The following result is a special case of the **twist map theorem**.

**Theorem:** For Diophantine $\alpha$, there is a solution of $F(q) = 0$ for small $|c|$.

The KAM theorem was predated by the **Poincaré-Siegel theorem** in complex dynamics which assured that if $f$ is analytic near $z = 0$ and $f'(0) = \lambda = \exp(2\pi i\alpha)$ with Diophantine $\alpha$, then there exists $u(z) = z + q(z)$ such that $f(u(z)) = u(\lambda z)$ holds in a small disk 0: there is an analytic solution $q$ to the **Schröder equation** $\lambda z + g(z + q(z)) = q(\lambda z)$. The question about the existence of invariant curves is important as it determines the **stability**. The twist map theorem result follows also from a **strong implicit function theorem** initiated by John Nash and Jürgen Moser. For larger $c$, or non-Diophantine $\alpha$, the solution $q$ still exists but it is no more continuous. This is **Aubry-Mather theory**. For $c \neq 0$, the operator $\hat{L}$ is an almost periodic **Toeplitz matrix** on $l^2(\mathbb{Z})$ which is a special kind of **discrete Schrödinger operator**. The decay rate of the off diagonals depends on the smoothness of $f$. Getting control of the inverse can be technical [42]. Even in the **Standard map** case $f(x) = \sin(x)$, the composition $f(q(t))$ is no more a trigonometric polynomial so that $\hat{L}$ appearing here is not a **Jacobi matrix in a strip**. The first breakthrough of the theorem in a frame work of Hamiltonian differential equations was done in 1954 by Andrey Kolmogorov. Jürgen Moser proved the discrete twist map version and Vladimir Arnold in 1963 for Hamiltonian systems. The above result generalizes to higher dimensions where one looks for **invariant tori** called **KAM tori**. One needs some non-degeneracy conditions See [59, 247, 248].

## 84. Continued Fraction

Given a positive **square free** integer $d$, the **Diophantine equation** $x^2 - dy^2 = 1$ is called **Pell's equation**. Solving it means to find a nontrivial unit in the ring $\mathbb{Z}[\sqrt{d}]$ because $(x + y\sqrt{d})(x - y\sqrt{d}) = 1$. The trivial solutions are $x = \pm 1, y = 0$. Solving the equation is therefore part of the **Dirichlet unit problem** from algebraic number theory. Let $[a_0; a_1, \dots]$ denote the **continued fraction expansion** of $x = \sqrt{d}$. This means $a_0 = [x]$ is the integer part and $[1/(x - a_0)] = a_1$ etc. If $x = [a_0; a_1, \dots, a_n + b_n]$, then $a_{n+1} = [1/b_n]$. Let $p_n/q_n = [a_0; a_1, a_2, \dots, a_n]$ denote the $n$'th **convergent** to the regular continued fraction of $\sqrt{d}$. A solution $(x_1, y_1)$ which minimizes $x$ is called the **fundamental solution**. The theorem tells that it is of the form $(p_n, q_n)$:

**Theorem:** Any solution to the Pell's equation is a convergent $p_n/q_n$.

One can find more solutions recursively because the ring of units in $\mathbb{Z}[\sqrt{d}]$ is $\mathbb{Z}_2 \times C_n$ for some cyclic group $C_n$. The other solutions $(x_k, y_k)$ can be obtained from $x_k + \sqrt{d}y_k = (x_1 + \sqrt{d}y_1)^k$. One of the first instances, where the equation appeared is in the **Archimedes cattle problem** which is $x^2 - 410286423278424y^2 = 1$. The equation is named after John Pell, who has nothing to do with the equation. It was Euler who attributed the solution by mistake to Pell. It was first found by William Brouncker. The approach through continued fractions started with Euler and Lagrange. See [276, 224].

## 85. Gauss-Bonnet-Chern

Let $(M, g)$ be a **Riemannian manifold** of dimension $d$ with **volume element** $d\mu$. If $R^{ij}_{kl}$ is **Riemann curvature tensor** with respect to the metric $g$, define the constant $C = ((4\pi)^{d/2}(-2)^{d/2}(d/2)!)^{-1}$ and the **curvature** $K(x) = C\sum_{\sigma,\pi} \text{sign}(\sigma)\text{sign}(\pi)R^{\sigma(1)\sigma(2)}_{\pi(1)\pi(2)} \cdots R^{\sigma(d-1)\sigma(d)}_{\pi(d-1)\pi(d)}$, where the sum is over all permutations $\pi, \sigma$ of $\{1, \dots, d\}$. It can be interpreted as a **Pfaffian**. In odd dimensions, the curvature is zero. Denote by $\chi(M)$ the **Euler characteristic** of $M$.

**Theorem:** $\int_M K(x)\, d\mu(x) = 2\pi\chi(M)$.

The case $d = 2$ was solved by Karl Friedrich Gauss and by Pierre Ossian Bonnet in 1848. Gauss knew the theorem but never published it. In the case $d = 2$, the curvature $K$ is the **Gaussian curvature** which is the product of the **principal curvatures** $\kappa_1, \kappa_2$ at a point. For a sphere of radius $R$ for example, the Gauss curvature is $1/R^2$ and $\chi(M) = 2$. The **volume form** is then the usual **area element** normalized so that $\int_M 1\, d\mu(x) = 1$. Allendoerfer-Weil in 1943 gave the first proof, based on previous work of Allendoerfer, Fenchel and Weil. Chern finally, in 1944 proved the theorem independent of an embedding. See [87], which features a proof of Vijay Kumar Patodi.

## 86. Atiyah-Singer

Assume $M$ is a compact orientable finite dimensional **manifold** of dimension $n$ and assume $D$ is an **elliptic differential operator** $D : E \to F$ between two smooth **vector bundles** $E, F$ over $M$. Using multi-index notation $D^k = \partial^{k_1}_{x_1} \cdots \partial^{k_n}_{x_n}$, a **differential operator** $\sum_k a_k(x)D^k x$ is called **elliptic** if for all $x$, its **symbol** the polynomial $\sigma(D)(y) = \sum_{|k|=n} a_k(x)y^k$ is not zero for nonzero $y$. **Elliptic regularity** assures that both the kernel of $D$ and the kernel of the **adjoint** $D^* : F \to E$ are both finite dimensional. The **analytical index** of $D$ is defined as $\chi(D) = \dim(\ker(D)) - \dim(\ker(D^*))$. We think of it as the Euler characteristic of $D$. The **topological index** of $D$ is defined as the integral of the $n$-form $K_D = (-1)^n \text{ch}(\sigma(D)) \cdot \text{td}(TM)$, over $M$. This $n$-form is the cup product $\cdot$ of the **Chern character** $\text{ch}(\sigma(D))$ and the **Todd class** of the complexified tangent bundle $TM$ of $M$. We think about $K_D$ as a **curvature**. Integration is done over the **fundamental class** $[M]$ of $M$ which is the natural **volume form** on $M$. The Chern character and the Todd classes are both mixed rational cohomology classes. On a complex vector bundle $E$ they are both given by concrete power series of **Chern classes** $c_k(E)$ like $\text{ch}(E) = e^{a_1(E)} + \cdots + e^{a_n(E)}$ and $\text{td}(E) = a_1(1 + e^{-a_1})^{-1} \cdots a_n(1 + e^{-a_n})^{-1}$ with $a_i = c_1(L_i)$ if $E = L_1 \oplus \cdots \oplus L_n$ is a direct sum of **line bundles**.

**Theorem:** The analytic index and topological indices agree: $\chi(D) = \int_M K_D$.

In the case when $D = d + d^*$ from the vector bundle of even forms $E$ to the vector bundle of odd forms $F$, then $K_D$ is the Gauss-Bonnet curvature and $\chi(D) = \chi(M)$. Israil Gelfand conjectured around 1960 that the analytical index should have a topological description. The Atiyah-Singer index theorem has been proven in 1963 by Michael Atiyah and Isadore Singer. The result generalizes the Gauss-Bonnet-Chern and Riemann-Roch-Hirzebruch theorem. According to

[278], "the theorem is valuable, because it connects analysis and topology in a beautiful and insightful way". See [260].

## 87. Complex multiplication

A **n'th root of unity** is a solution to the equation $z^n = 1$ in the complex plane $\mathbb{C}$. It is called **primitive** if it is not a solution to $z^k = 1$ for some $1 \leq k < n$. A **cyclotomic field** is a number field $\mathbb{Q}(\zeta_n)$ which is obtained by adjoining a complex **primitive root of unity** $\zeta_n$ to $\mathbb{Q}$. Every cyclotomic field is an Abelian field extension of the field of rational numbers $\mathbb{Q}$. The **Kronecker-Weber** theorem reverses this. It is also called the main theorem of **class field theory over** $\mathbb{Q}$

**Theorem:** Every Abelian extension $L/\mathbb{Q}$ is a subfield of a cyclotomic field.

Abelian field extensions of $\mathbb{Q}$ are also called **class fields**. It follows that any **algebraic number field** $K/Q$ with Abelian **Galois group** has a **conductor**, the smallest $n$ such that $K$ lies in the field generated by $n$'th roots of unity. Extending this theorem to other base number fields is **Kronecker's Jugendtraum** or **Hilbert's twelfth problem**. The theory of **complex multiplication** does the generalization for **imaginary quadratic fields**. The theorem was stated by Leopold Kronecker in 1853 and proven by Heinrich Martin Weber in 1886. A generalization to **local fields** was done by Jonathan Lubin and John Tate in 1965 and 1966. (A **local field** is a locally compact topological field with respect to some non-discrete topology. The list of local fields is $\mathbb{R}, \mathbb{C}$, field extensions of the **p-adic numbers** $\mathbb{Q}_p$, or formal Laurent series $F_q((t))$ over a finite field $F_q$.) The study of **cyclotomic fields** came from elementary geometric problems like the construction of a regular $n$-gon with **ruler and compass**. Gauss constructed a regular 17-gon and showed that a **regular $n$-gon** can be constructed if and only if $n$ is a **Fermat prime** $F_n = 2^{2^n}+1$ (the known ones are $3, 6, 17, 257, 65537$ and a problem of Eisenstein of 1844 asks whether there are infinitely many). Further interest came in the context of **Fermat's last theorem** because $x^n + y^n = z^n$ can be written as $x^n + y^n = (x+y)(x+\zeta y) + \cdots (x+\zeta^{n-1}y)$, where $\zeta$ is a $n$'th root of unity.

## 88. Choquet theory

Let $K$ be a **compact** and **convex** set in a Banach space $X$. A point $x \in K$ is called **extreme** if $x$ is not in an open interval $(a, b)$ with $a, b \in K$. Let $E$ be the set of extreme points in $K$. The **Krein-Milman theorem** assures that $K$ is the convex hull of $E$. Given a probability measure $\mu$ on $E$, it defines the point $x = \int y d\mu(y)$. We say that $x$ is the **Barycenter** of $\mu$. The **Choquet theorem** is

**Theorem:** Every point in $K$ is a Barycenter of its extreme points.

This result of Choquet implies the Krein-Milman theorem. It generalizes to **locally compact topological spaces**. The measure $\mu$ is not unique in general. It is in finite dimensions if $K$ is a simplex. But in general, as shown by Heinz Bauer in 1961, for an extreme point $x \in K$ the measure $\mu_x$ is unique. It has been proven by **Gustave Choquet** in 1956 and was generalized by Erret Bishop and Karl de Leeuw in 1959. [266]

## 89. Helly's theorem

Given a family $\mathcal{K} = \{K_1, \dots K_n\}$ of **convex** sets $K_1, K_2, \dots, K_n$ in the **Euclidean space** $\mathbb{R}^d$ and assume that $n > d$. Let $\mathcal{K}_m$ denote the set of subsets of $\mathcal{K}$ which have exactly $m$ elements. We say that $\mathcal{K}_m$ has the **intersection property** if every of its elements has a non-empty common intersection. The **theorem of Kelly** assures that

**Theorem:** $\mathcal{K}_n$ has the intersection property if $\mathcal{K}_{d+1}$ has.

The theorem was proven in 1913 by Eduard Kelly. It generalizes to an infinite collection of compact, convex subsets. This theorem led Johann Radon to prove in 1921 the **Radon theorem** which states that any set of $d+2$ points in $\mathbb{R}^d$ can be partitioned into two disjoint subsets whose convex hull intersect. A nice application of Radon's theorem is the **Borsuk-Ulam theorem** which states that a continuous function $f$ from the $d$-dimensional sphere $S^n$ to $\mathbb{R}^d$ must some pair of **antipodal points** to the same point: $f(x) = f(-x)$ has a solution. For example, if $d = 2$, this implies that on earth, there are at every moment two antipodal points on the Earth's surface for which the temperature and the pressure are the same. The **Borsuk-Ulam** theorem appears first have been stated in work of Lazar Lyusternik and Lev Shnirelman in 1930, and proven by Karol Borsuk in 1933 who attributed it to Stanislav Ulam.

## 90. Weak Mixing

An **automorphism** $T$ of a probability space $(X, \mathcal{A}, m)$ is a measure preserving invertible measurable transformation from $X$ to $X$. It is called **ergodic** if $T(A) = A$ implies $m(A) = 0$ or $m(A) = 1$. It is called **mixing** if $m(T^n(A) \cap B) \to m(A) \cdot m(B)$ for $n \to \infty$ for all $A, B$. It is called **weakly mixing** if $n^{-1} \sum_{k=0}^{n-1} |m(T^k(A) \cap B) - m(A) \cdot m(B)| \to 0$ for all $A, B \in \mathcal{A}$ and $n \to \infty$. This is equivalent to the fact that the unitary operator $Uf = f(T)$ on $L^2(X)$ has no point spectrum when restricted to the orthogonal complement of the constant functions. A topological transformation (a continuous map on a locally compact topological space) with a weakly mixing invariant measure is **not integrable** as for integrability, one wants every invariant measure to lead to an operator $U$ with pure point spectrum and conjugating it so to a group translation. Let $\mathcal{G}$ be the complete topological group of automorphisms of $(X, \mathcal{A}, m)$ with the weak topology: $T_j$ converges to $T$ **weakly**, if $m(T_j(A) \Delta T(A)) \to 0$ for all $A \in \mathcal{A}$; this topology is metrizable and completeness is defined with respect to an equivalent metric.

**Theorem:** A generic $T$ is weakly mixing and so ergodic.

Anatol Katok and Anatolii Mikhailovich Stepin in 1967 [185] proved that purely singular continuous spectrum of $U$ is generic. A new proof was given by [68] and a short proof in using **Rokhlin's lemma**, Halmos conjugacy lemma and a Simon's **"wonderland theorem"** establishes both genericity of weak mixing and genericity of singular spectrum. On the topological side, a generic volume preserving homeomorphism of a manifold has purely singular continuous spectrum which strengthens Oxtoby-Ulam's theorem [259] about generic ergodicity. [186, 151] The Wonderland theorem of Simon [295] also allowed to prove that a generic invariant measure of a shift is singular continuous [196] or that zero-dimensional singular continuous spectrum is generic for open sets of flows on the torus allowing also to show that open sets of Hamiltonian systems contain generic subset with both quasi-periodic as well as weakly mixing invariant tori [197]

## 91. Universality

The space $X$ of **unimodular maps** is the set of twice continuously differentiable even maps $f : [-1,1] \to [-1,1]$ satisfying $f(0) = 1$ $f''(x) < 0$ and $\lambda = g(1) < 0$. The **Feigenbaum-Cvitanović functional equation** (FCE) is $g = Tg$ with $T(g)(x) = \frac{1}{\lambda} g(g(\lambda x))$. The map $T$ is a **renormalization map.**

**Theorem:** There exists an analytic hyperbolic fixed point of $T$.

The first proof was given by Oscar Lanford III in 1982 (computer assisted). See [168, 169]. That proof also established that the fixed point is hyperbolic with a one-dimensional unstable manifold and positive expanding eigenvalue. This explains some **universal features** of unimodular maps found experimentally in 1978 by Mitchell Feigenbaum and which is now called **Feigenbaum universality**. The result has been ported to area preserving maps [102].

## 92. Compactness

Let $X$ be a compact metric space $(X, d)$. The Banach space $C(X)$ of real-valued continuous functions is equipped with the supremum norm. A closed subset $F \subset C(X)$ is called **uniformly bounded** if for every $x$ the supremum of all values $f(x)$ with $f \in F$ is bounded. The set $F$ is called **equicontinuous** if for every $x$ and every $\epsilon > 0$ there exists $\delta > 0$ such that if $d(x, y) < \delta$, then $|f(x) - f(y)| < \epsilon$ for all $f \in F$. A set $F$ is called **precompact** if its closure is compact. The **Arzelà-Ascoli theorem** is:

**Theorem:** Equicontinuous uniformly bounded sets in $C(X)$ are precompact.

The result also holds on **Hausdorff spaces** and not only metric spaces. In the complex, there is a variant called **Montel's theorem** which is the fundamental normality test for holomorphic functions: an uniformly bounded family of holomorphic functions on a complex domain $G$ is **normal** meaning that its closure is compact with respect to the **compact-open topology**. The compact-open topology in $C(X, Y)$ is the topology defined by the **sub-base** of all continuous maps $f_{K,U} : f : K \to U$, where $K$ runs over all compact subsets of $X$ and $U$ runs over all open subsets of $Y$.

## 93. Geodesic

The **geodesic distance** $d(x, y)$ between two points $x, y$ on a **Riemannian manifold** $(M, g)$ is defined as the length of the shortest geodesic $\gamma$ connecting $x$ with $y$. This renders the manifold a metric space $(M, d)$. We assume it is **locally compact**, meaning that every point $x \in M$ has a compact neighborhood. A metric space is called **complete** if every **Cauchy sequence** in $M$ has a convergent subsequence. (A sequence $x_k$ is called a Cauchy sequence if for every $\epsilon > 0$, there exists $n$ such that for all $i, j > n$ one has $d(x_i, x_j) < \epsilon$.) The local existence of differential equations assures that the geodesic equations exist for small enough time. This can be restated that the **exponential map** $v \in T_xM \to M$ assigning to a point $v \neq 0$ in the tangent space $T_xM$ the solution $\gamma(t)$ with initial velocity $v/|v|$ and $t \leq |v|$, and $\gamma(0) = x$. A Riemannian manifold $M$ is called **geodesically complete** if the exponential map can be extended to the entire tangent space $T_xM$ for every $x \in M$. This means that geodesics can be continued for all times. The Hopf-Rinov theorem assures:

**Theorem:** Completeness and geodesic completeness are equivalent.

The theorem was named after Heinz Hopf and his student Willi Rinov who published it in 1931. See [96].

## 94. Crystallography

A **wall paper group** is a discrete subgroup of the **Euclidean symmetry group** $E_2$ of the plane. Wall paper groups classify two-dimensional patterns according to their symmetry. In the plane $\mathbb{R}^2$, the underlying group is the group $E_2$ of **Euclidean plane symmetries** which contain **translations rotations** or **reflections** or **glide reflections**. This group is the group of rigid motions. It is a three dimensional **Lie group** which according to Klein's **Erlangen program** characterizes **Euclidean geometry**. Every element in $E_2$ can be given as a pair $(A, b)$, where $A$ is an orthogonal matrix and $b$ is a vector. A subgroup $G$ of $E_2$ is called **discrete** if there is a positive minimal distance between two elements of the group. This implies the **crystallographic restriction theorem** assuring that only rotations of order $2, 3, 4$ or $6$ can appear. This means only rotations by $180, 120, 90$ or $60$ degrees can occur in a Wall paper group.

**Theorem:** There are 17 wallpaper groups

The first proof was given by Evgraf Fedorov in 1891 and then by George Polya in 1924. in three dimensions there are 230 **space groups** and 219 types if **chiral copies** are identified. In space there are 65 space groups which preserve the orientation. See [258, 147, 177].

## 95. Quadratic forms

A symmetric square matrix $Q$ of size $n \times n$ with integer entries defines a **integer quadratic form** $Q(x) = \sum_{i,j=1}^n Q_{ij} x_i x_j$. It is called **positive** if $Q(x) > 0$ whenever $x \neq 0$. A positive integral quadratic form is called **universal** if its range is $\mathbb{N}$. For example, by the **Lagrange four square theorem**, the form $Q(x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 + x_3^2 + x_4^2$ is universal. The **Conway-Schneeberger fifteen theorem** tells

**Theorem:** $Q$ is universal if it has $\{1, \dots 15\}$ in the range.

The interest in quadratic forms started in the 17'th century especially about numbers which can be represented as sums $x^2 + y^2$. Lagrange, in 1770 proved the four square theorem. In 1916, Ramajujan listed all diagonal quaternary forms which are universal. The 15 theorem was proven in 1993 by John Conway and William Schneeberger (a student of Conway's in a graduate course given in 1993). There is an analogue theorem for **integral positive quadratic forms**, these are defined by positive definite matrices $Q$ which take only integer values. The binary quadratic form $x^2 + xy + y^2$ for example is integral but not an integer quadratic form because the corresponding matrix $Q$ has fractions $1/2$. In 2005, Bhargava and Jonathan Hanke proved the 290 theorem, assuring that an integral positive quadratic form is universal if it contains $\{1, \dots, 290\}$ in its range. [75].

## 96. Sphere packing

A **sphere packing** in $\mathbb{R}^d$ is an arrangement of non-overlapping unit spheres in the $d$-dimensional Euclidean space $\mathbb{R}^d$ with volume measure $\mu$. It is known since [143] that packings with maximal densities exist. Denote by $B_r(x)$ the ball of radius $r$ centered at $x \in \mathbb{R}^d$. If $X$ is the set of centers of the sphere and $P = \bigcup_{x \in X} B_1(x)$ is the union of the unit balls centered at points in

$X$, then the **density** of the packing is defined as $\Delta_d = \limsup \int_{B_r(0)} P \; d\mu / \int_{B_r(0)} 1 \; d\mu$. The sphere packing problem is now solved in 5 different cases:

**Theorem:** Optimal sphere packings are known for $d = 1, 2, 3, 8, 24$.

The one-dimensional case $\Delta_1 = 1$ is trivial. The case $\Delta_2 = \pi/\sqrt{12}$ was known since Axel Thue in 1910 but proven only by Lásló Fejes Toóth in 1943. The case $d = 3$ was called the **Kepler conjecture** as Johannes Kepler conjectured $\Delta_3 = \pi/\sqrt{18}$. It was settled by Thomas Hales in 1998 using computer assistance. A complete formal proof appeared in 2015. The case $d = 8$ was settled by Maryna Viazovska who proved $\Delta_8 = \pi^4/384$ and also established uniqueness. The densest packing in the case $d = 8$ is the $E_8$ lattice. The proof is based on linear programming bounds developed by Henry Cohn and Noam Elkies in 2003. Later with other collaborators, she also covered the case $d = 24$. the densest packing in dimension 24 is the **Leech lattice**. For sphere packing see [82, 81].

## 97. Sturm theorem

Given a square free **real-valued polynomial** $p$ let $p_k$ denote the **Sturm chain**, $p_0 = p$, $p_1 = p'$, $p_2 = p_0 \bmod p_1$, $p_3 = p_1 \bmod p_2$ etc. Let $\sigma(x)$ be the number of **sign changes** ignoring zeros in the sequence $p_0(x), p_1(x), \dots, p_m(x)$.

**Theorem:** The number of distinct roots of $p$ in $(a, b]$ is $\sigma(b) - \sigma(a)$.

Sturm proved the theorem in 1829. He found his theorem on sequences while studying solutions of differential equations **Sturm-Liouville theory** and credits Fourier for inspiration. See [272].

## 98. Smith Normal form

A integer $m \times n$ matrix $A$ is said to be expressible in **Smith normal form** if there exists an invertible $m \times m$ matrix $S$ and an invertible $n \times n$ matrix $T$ so that $SMT$ is a diagonal matrix $\mathrm{Diag}(\alpha_1, \dots, \alpha_r, 0, 0, 0)$ with $\alpha_i | \alpha_{i+1}$. The integers $\alpha_i$ are called **elementary divisors**. They can be written as $\alpha_i = d_i(A)/d_{i-1}(A)$, where $d_0(A) = 1$ and $d_k(A)$ is the greatest common divisor of all $k \times k$ minors of $A$. The Smith normal form is called **unique** if the elementary divisors $\alpha_i$ are determined up to a sign.

**Theorem:** Any integer matrix has a unique Smith normal form.

The result was proven by Henry John Stephen Smith in 1861. The result holds more generally in a **principal ideal domain**, which is an **integral domain** (a ring $R$ in which $ab = 0$ implies $a = 0$ or $b = 0$) in which every **ideal** (an additive subgroup $I$ of the ring such that $ab \in I$ if $a \in I$ and $b \in R$) is generated by a single element.

## 99. Spectral perturbation

A complex valued matrix $A$ is **self-adjoint** = Hermitian if $A^* = A$, where $A^*_{ij} = \overline{A}_{ji}$. The spectral theorem assures that $A$ has real eigenvalues Given two selfadjoint complex $n \times n$ matrices $A, B$ with eigenvalues $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$ and $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$, one has the Lidskii-Last theorem:

**Theorem:** $\sum_{j=1}^n |\alpha_j - \beta_j| \leq \sum_{i,j=1}^n |A - B|_{ij}$.

The result has been deduced by Yoram Last (around 1993) from **Lidskii's inequality** found in 1950 by Victor Lidskii $\sum_j |\alpha_j - \beta_j| \leq \sum_j |\gamma_j|$ where $\gamma_j$ are the eigenvalues of $C = B - A$ (see [296] page 14). The original Lidskii inequality also holds for $p \geq 1$: $\sum_j |\alpha_j - \beta_j|^p \leq \sum_j |\gamma_j|^p$. Last's spin on it allows to estimate the $l^1$ spectral distance of two self-adjoint matrices using the $l^1$ distance of the matrices. This is handy as we often know the matrices $A, B$ explicitly rather than the eigenvalues $\gamma_j$ of $A - B$.

## 100. Radon transform

In order to solve the **tomography problem** like MRI of finding the density function $g(x, y, z)$ of a three dimensional body, one looks at a **slice** $f(x, y) = g(x, y, c)$, where $z = c$ is kept constant and measures the **Radon transform** $R(f)(p, \theta) = \int_{\{x\cos(\theta)+y\sin(\theta)=p\}} f(x, y)\, ds$. This quantity is the **absorption rate** due to **nuclear magnetic resonance** along the line $L$ of polar angle $\alpha$ in distance $p$ from the center. Reconstructing $f(x, y) = g(x, y, c)$ for different $c$ allows to recover the **tissue density** $g$ and so to "see inside the body".

**Theorem:** The Radon transform can be diagonlized and so pseudo inverted.

We only need that the Fourier series $f(r, \phi) = \sum_n f_n(r)e^{in\phi}$ converges uniformly for all $r > 0$ and that $f_n(r)$ has a Taylor series. The expansion $f(r, \phi) = \sum_{n\in\mathbb{Z}} \sum_{k=1}^\infty f_{n,k}\psi_{n,k}$ with $\psi_{n,k}(r, \phi) = r^{-k}e^{in\phi}$ is an eigenfunction expansion with eigenvalues $\lambda_{n,k} = 2\int_0^{\pi/2} \cos(nx)\cos(x)^{(k-1)}\, dx = \frac{\pi}{2^{k-1}\cdot k} \cdot \frac{\Gamma(k+1)}{\Gamma(\frac{k+n+1}{2})\Gamma(\frac{k-n+1}{2})}$. The **inverse problem** is subtle due to the existence of a **kernel** spanned by $\{\psi_{n,k} \mid (n + k) \text{ odd }, |n| > k\}$. One calls it an **ill posed problem** in the sense of Hadamard. The Radon transform was first studied by Johann Radon in 1917 [161].

## 101. Linear programming

Given two vectors $c \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$, and a $n \times m$ matrix $A$, a **linear program** is the variational problem on $\mathbb{R}^m$ to maximize $f(x) = c \cdot x$ subject to the linear constraints $Ax \leq b$ and $x \geq 0$. The dual problem is to minimize $b \cdot y$ subject to to $A^T y \geq c, y \geq 0$. The **maximum principle** for linear programming is tells that the solution is on the boundary of the **convex polytop** formed by the **feasable region** defined by the constraints.

**Theorem:** Local optima of linear programs are global and on the boundary

.

Since the solutions are located on the vertices of the polytope defined by the constraints the **simplex algorithm** for solving linear programs works: start at a vertex of the polytop, then move along the edges along the gradient until the optimum is reached. If $A = [2, 3]$ and $x = [x_1, x_2]$ and $b = 6$ and $c = [3, 5]$ we have $n = 1, m = 2$. The problem is to maximize $f(x_1, x_2) = 3x_1 + 5x_2$ on the triangular region $2x_1 + 3x_2 \leq 6, x_1 \geq 0, x_2 \geq 0$. Start at $(0, 0)$, the best improvement is to go to $(0, 2)$ which is already the maximum. Linear programming is used to solve practical problems in operations research. The simplex algorithm was formulated by George Dantzig in 1947. It solves random problems nicely but there are expensive cases in general and it is possible that cycles occur. One of the open problems of Steven Smale asks for a strongly polynomial time algorithm deciding whether a solution of a linear programming problem exists. [250]

## 102. Random Matrices

A **random matrix** $A$ is given by an $n \times n$ array of independent, identically distributed random variables $A_{ij}$ of zero mean and standard deviation 1. The eigenvalues $\lambda_j$ of $A/\sqrt{n}$ define a discrete measure $\mu_n = \sum_j \delta_{\lambda_j}$ called **spectral measure** of $A$. The **circular law** on the complex plane $\mathbb{C}$ is the probability measure $\mu_0 = 1_D/\pi$, where $D = \{|z| \leq 1\}$ is the unit disk. A sequence $\nu_n$ of probability measures converges **weakly** or **in law** to $\nu$ if for every continuous function $f : \mathbb{C} \to \mathbb{C}$ one has $\int f(z)\, d\nu_n(z) \to \int f(z)\, d\nu(z)$. The **circular law** is:

**Theorem:** Almost surely, the spectral measures converge $\mu_n \to \mu_0$.

One can think of $A_n$ as a sequence of larger and larger matrix valued random variables. The circular law tells that the eigenvalues fill out the unit disk in the complex plane uniformly when taking larger and larger matrices. It is a kind of central limit theorem. An older version due to Eugene Wigner from 1955 is the **semi-circular law** telling that in the self-adjoint case, the now real measures $\mu_n$ converge to a distribution with density $\sqrt{4-x^2}/(2\pi)$ on $[-2,2]$. The circular law was stated first by Jean Ginibre in 1965 and Vyacheslav Girko 1984. It was proven first by Z.D. Bai in 1997. Various authors have generalized it and removed more and more moment conditions. The latest condition was removed by Terence Tao and Van Vu in 2010, proving so to the above "fundamental theorem of random matrix theory". See [321].

## 103. Diffeomorphisms

Let $M$ be a compact Riemannian surface and $T : M \to M$ a $C^2$-diffeomorphism. A Borel probability measure $\mu$ on $M$ is $T$-invariant if $\mu(T(A)) = \mu(A)$ for all $A \in \mathcal{A}$. It is called **ergodic** if $T(A) = A$ implies $\mu(A) = 1$ or $\mu(A) = 0$. The **Hausdorff dimension** $\dim(\mu)$ of a measure $\mu$ is defined as the Hausdorff dimension of the smallest Borel set $A$ of full measure $\mu(A) = 1$. The **entropy** $h_\mu(T)$ is the **Kolmogorov-Sinai entropy** of the measure-preserving dynamical system $(X, T, \mu)$. For an ergodic surface diffeomorphism, the **Lyapunov exponents** $\lambda_1, \lambda_2$ of $(X, T, \mu)$ are the logarithms of the eigenvalues of $A = \lim_{n\to\infty} [(dT^n(x))^* dT^n(x)]^{1/(2n)}$, which is a limiting Oseledec matrix and constant $\mu$ almost everywhere due to ergodicity. Let $\lambda(T, \mu)$ denote the Harmonic mean of $\lambda_1, -\lambda_2$. The **entropy-dimension-Lyapunov theorem** tells that for every $T$-invariant ergodic probability measure $\mu$ of $T$, one has:

**Theorem:** $h_\mu = \dim(\mu)\lambda/2$.

This formula has become famous because it relates "entropy", "fractals" and "chaos", which are all "rock star" notions also outside of mathematics. The theorem implies in the case of Lebesgue measure preserving symplectic transformation, where $\dim(\mu) = 2$ and $\lambda_1 = -\lambda_2$ that "entropy = Lyaponov exponent" which is a **formula of Pesin** given by $h_\mu(T) = \lambda(T, \mu)$. A similar result holds for **circle diffeomorphims** or smooth interval maps, where $h_\mu(T) = \dim(\mu)\lambda(T, \mu)$. The notion of Hausdorff dimension was introduced by Felix Hausdoff in 1918. Entropy was defined in 1958 by Nicolai Kolmogorov and in general by Jakov Sinai in 1959. Lyapunov exponents were introduced with the work of Valery Oseledec in 1965. The above theorem is due to Lai-Sang Young who proved it in 1982. Francois Ledrapier and Lai-Sang Young proved in 1985 that in arbitrary dimensions, $h_\mu = \sum_j \lambda_j \gamma_j$, where $\gamma_j$ are dimensions of $\mu$ in the direction of the Oseledec spaces $E_j$. This is called the **Ledrappier-Young formula**. It implies the **Margulis-Ruelle inequality** $h_\mu(T) \leq \sum_j \lambda_j^+(T)$, where $\lambda_j^+ = \max(\lambda_j, 0)$ and $\lambda_j(T)$ are the Lyapunov exponents. In the case of a smooth $T$-invariant measure $\mu$ or more generally, for SRB measures, there is an equality $h_\mu(T) = \sum_j \lambda_j^+(T)$ which is called the **Pesin formula**. See [184, 103].

## 104. Linearization

If $F : M \to M$ is a globally Lipschitz continuous function on a finite dimensional vector space $M$, then the differential equation $x' = F(x)$ has a global solution $x(t) = f^t(x(0))$ (a local by **Picard's existence theorem** and global by the **Grönwall inequality**). An **equilibrium point** of the system is a point $x_0$ for which $F(x_0) = 0$. This means that $x_0$ is a fixed point of a differentiable mapping $f = f^1$, the **time-1-map**. We say that $f$ is **linearizable** near $x_0$ if there exists a homeomorphism $\phi$ from a neighborhood $U$ of $x_0$ to a neighborhood $V$ of $x_0$ such that $\phi \circ f \circ \phi^{-1} = df$. The **Sternberg-Grobman-Hartman linearization theorem** is

**Theorem:** If $f$ is hyperbolic, then $f$ is linearizable near $x_0$.

The theorem was proven by D.M. Grobman in 1959 Philip Hartman in 1960 and by Shlomo Sternberg in 1958. This implies the existence of **stable and unstable manifolds** passing through $x_0$. One can show more and this is due to Sternberg who wrote a series of papers starting 1957 [307]: if $A = df(x_0)$ satisfies **no resonance condition** meaning that no relation $\lambda_0 = \lambda_1 \cdots \lambda_j$ exists between eigenvalues of $A$, then a **linearization to order** $n$ is a $C^n$ map $\phi(x) = x + g(x)$, with $g(0) = g'(0) = 0$ such that $\phi \circ f \circ \phi^{-1}(x) = Ax + o(|x|^n)$ near $x_0$. We say then that $f$ can be $n$**-linearized** near $x_0$. The generalized result tells that non-resonance fixed points of $C^n$ maps are $n$-linearizable near a fixed point. See [219].

## 105. Fractals

An **iterated function system** is a finite set of contractions $\{f_i\}_{i=1}^n$ on a complete metric space $(X, d)$. The corresponding **Huntchingson operator** $H(A) = \sum_i f_i(A)$ is then a contraction on the **Hausdorff metric** of sets and has a unique fixed point called the **attractor** $S$ of the iterated function system. The definition of **Hausdorff dimension** is as follows: define $h_\delta^s(A) = \inf_{U \in \mathcal{U}} \sum_i |U_i|^s$, where $\mathcal{U}$ is a $\delta$-cover of $A$. And $h^s(A) = \lim_{\delta \to 0} H_\delta^s(A)$. The **Hausdorff dimension** $\dim_H(S)$ finally is the value $s$, where $h^s(S)$ jumps from $\infty$ to 0. If the contractions are maps with contraction factors $0 < \lambda_j < 1$ then the Hausdorff dimension of the attractor $S$ can be estimated with the the **similarity dimension** of the contraction vector $(\lambda_1, \ldots, \lambda_n)$: this number is defined as the solution $s$ of the equation $\sum_{i=1}^n \lambda_i^{-s} = 1$.

**Theorem:** $\dim_{\text{hausdorff}}(S) \leq \dim_{\text{similarity}}(S)$.

There is an equality if $f_i$ are all affine contractions like $f_i(x) = A_i \lambda x + \beta_i$ with the same contraction factor and $A_i$ are orthogonal and $\beta_i$ are vectors (a situation which generates a large class of popular fractals). For equality one also has to assume that there is an open non-empty set $G$ such that $G_i = f_i(G)$ are disjoint. In the case $\lambda_j = \lambda$ are all the same then $n\lambda^{-\dim} = 1$ which implies $\dim(S) = -\log(n)/\log(\lambda)$. For the **Smith-Cantor set** $S$, where $f_1(x) = x/3 + 2/3, f_2(x) = x/3$ and $G = (0, 1)$. One gets with $n = 2$ and $\lambda = 1/3$ the dimension $\dim(S) = \log(2)/\log(3)$. For the **Menger carpet** with $n = 8$ affine maps $f_{ij}(x, y) = (x/3 + i/3, y/3 + j/3)$ with $0 \leq i \leq 2, 0 \leq j \leq 2, (i, j) \neq (1, 1)$, the dimension is

$\log(8)/\log(3)$. The **Menger sponge** is the analogue object with $n = 20$ affine contractions in $\mathbb{R}^3$ and has dimension $\log(20)/\log(3)$. For the **Koch curve** on the interval, where $n = 4$ affine contractions of contraction factor $1/3$ exist, the dimension is $\log(4)/\log(3)$. These are all **fractals**, sets with Hausdorff dimension different from an integer. The modern formulation of iterated function systems is due to John E. Hutchingson from 1981. Michael Barnsley used the concept for a **fractal compression algorithms**, which uses the idea that storing the rules for an iterated function system is much cheaper than the actual attractor. Iterated function systems appear in complex dynamics in the case when the **Julia set** is completely disconnected, they have appeared earlier also in work of Georges de Rham 1957. See [230, 117].

## 106. Strong law of Small numbers

Like the Bayes theorem or the Pigeon hole principle which both are too simple to qualify as "theorems" but still are of utmost importance, the "Strong law of large numbers" is not really a theorem but a **fundamental mathematical principle**. It is more fundamental than a specific theorem as it applies throughout mathematics. It is for example important in Ramsey theory: The statement is put in different ways like "There aren't enough small numbers to meet the many demands made of them". [148] puts it in the following catchy way:

**Theorem:** You can't tell by looking.

The point was made by Richard Guy in [148] who states two "corollaries": **"superficial similarities spawn spurious statements"** and **"early exceptions eclipse eventual essentials"**. The statement is backed up with countless many examples (a list of 35 are given in [148]). Famous are Fermat's claim that all **Fermat primes** $2^{2^n}+1$ are prime or the claim that the number $\pi_3(n)$ of primes of the form $4k+3$ in $\{1,\dots,n\}$ is larger than $\pi_1(n)$ of primes of the form $4k+1$ so that the $4k+3$ primes win the **prime race**. Hardy and Littlewood showed however $\pi_3(n) - \pi_1(n)$ changes sign infinitely often. The prime number theorem extended to arithmetic progressions shows $\pi_1(n) \sim n/(2\log(n))$ and $\pi_3(n) \sim n/(2\log(n))$ but the density of numbers with $\pi_3(n) > \pi_1(n)$ is larger than $1/2$. This is the **Chebychev bias**. Experiments then suggested the density to be 1 but also this is false: the density of numbers for which $\pi_3(n) > \pi_1(n)$ is smaller than 1. The principle is important in a branch of combinatorics called **Ramsey theory**. But it not only applies in discrete mathematics. There are many examples, where one can not tell by looking. When looking at the boundary of the Mandelbrot set for example, one would tell that it is a fractal with Hausdorff dimension between 1 and 2. In reality the Hausdorff dimension is 2 by a result of Mitsuhiro Shishikura. Mandelbrot himself thought first "by looking" that the Mandelbrot set $M$ is disconnected. Douady and Hubbard proved $M$ to be connected.

## 107. Ramsey Theory

Let $G$ be the complete graph with $n$ vertices. An **edge labeling** with $r$ colors is an assignment of $r$ numbers to the **edges** of $G$. A complete sub-graph of $G$ is called a **clique**. If it is has $s$ vertices, it is denoted by $K_s$. A graph $G$ is called **monochromatic** if all edges in $G$ have the same color. (We use in here **coloring** as a short for **edge labeling** and not in the sense of chromatology where an edge coloring assumes that intersecting edges have different colors.) Ramsey's theorem is:

**Theorem:** For large $n$, every $r$-colored $K_n$ contains a monochromatic $K_s$.

So, there exist **Ramsey numbers** $R(r,s)$ such that for $n \geq R(r,s)$, the edge coloring of one of the $s$-cliques can occur. A famous case is the identity $R(3,3)=6$. Take $n=6$ people. It defines the complete graph $G$. If two of them are friends, color the edge blue, otherwise red. This **friendship graph** therefore is a $r=2$ coloring of $G$. There are 78 possible colorings. In each of them, there is a triangle of friends or a triangle of strangers. In a group of 6 people, there are either a clique with 3 friends or a clique of 3 complete strangers. The Theorem was proven by Frank Ramsey in 1930. Paul Erdoes asked to give explicit estimated $R(s)$ which is the least integer $n$ such that any graph on $n$ vertices contains either a **clique** of size $s$ (a set where all are connected to each other) or an independent set of size $s$ (a set where none are connected to each other). Graham for example asks whether the limit $R(n)^{1/n}$ exists. Ramsey theory also deals other sets: **van der Waerden's theorem** from 1927 for example tells that if the positive integers $\mathbb{N}$ are colored with $r$ colors, then for every $k$, there exists an $N$ called $W(r,k)$ such that the finite set $\{1\dots,N\}$ has an arithmetic progression with the same color. For example, $W(2,3)=9$. Also here, it is an open problem to find a formula for $W(r,k)$ or even give good upper bounds. [139] [138]

## 108. Poincaré Duality

For a differentiable **Riemannian $n$-manifold** $(M,g)$ there is an **exterior derivative** $d=d_p$ which maps $p$-forms $\Lambda^p$ to $(p+1)$-forms $\Lambda_{p+1}$. For $p=0$, the derivative is called the **gradient**, for $p=1$, the derivative is called the **curl** and for $p=d-1$, the derivative is the adjoint of **divergence**. The Riemannian metric defines an inner product $\langle f,h\rangle$ on $\Lambda^p$ allowing so to see $\Lambda^p$ as part of a Hilbert space and to define the adjoint $d^*$ of $d$. It is a linear map from $\Lambda^{p+1}$ to $\Lambda^p$. The exterior derivative defines so the self-adjoint **Dirac operator** $D=d+d^*$ and the **Hodge Laplacian** $L=D^2=dd^*+d^*d$ which now leaves each $\Lambda^p$ invariant. **Hodge theory** assures that $\dim(\ker(L|\Lambda^p))=b_p=\dim(H^p(M))$, where $H^p(M)$ are the $p$'th **cohomology group**, the kernel of $d_p$ modulo the image of $d_{p-1}$. **Poincaré duality** is:

**Theorem:** If $M$ is orientable n-manifold, then $b_k(M)=b_{n-k}(M)$.

The **Hodge dual** of $f\in\Lambda^p$ is defined as the unique $*g\in\Lambda^{n-p}$ satisfying $\langle f,*g\rangle=\langle f\wedge g,\omega\rangle$ where $\omega$ is the volume form. One has $d^*f=(-1)^{d+dp+1}*d*f$ and $L*f=*Lf$. This implies that $*$ is a unitary map from $\ker(L|\Lambda^p)$ to $\ker(L|\Lambda^{d-p})$ proving so the duality theorem. For $n=4k$, one has $*^2=1$, allowing to define the **Hirzebruch signature** $\sigma := \dim\{u|Lu=0,*u=u\}-\dim(u|Lu=0,*u=-u\}$. The Poinaré duality theorem was first stated by Henry Poincaré in 1895. It took until the 1930ies to clean out the notions and make it precise. The Hodge approach establishing an explicit isomorphism between harmonic $p$ and $n-p$ forms appears for example in [87].

## 109. Rokhlin-Kakutani approximation

Let $T$ be an automorphism of a probability space $(\Omega,\mathcal{A},\mu)$. This means $\mu(A)=\mu(T(A))$ for all $A\in\mathcal{A}$. The system $T$ is called **aperiodic**, if the set of **periodic points** $P=\{x\in\Omega \mid \exists n>0, T^nx=x\}$ has measure $\mu(P)=0$. A set $B\in\mathcal{A}$ which has the property that $B,T(B),\dots,T^{n-1}(B)$ are disjoint is called a **Rokhlin tower**. The measure of the tower is $\mu(B\cup\dots\cup T^{n-1}(B))=n\mu(B)$. We call it an $(1-\epsilon)$-Rokhlin tower. We say $T$ can be

**approximated arbitrary well** by Rokhlin towers, if for all $\epsilon > 0$, there is an $(1-\epsilon)$ Rokhlin tower.

**Theorem:** An aperiodic $T$ can be approximated well by Rokhlin towers.

The result was proven by Vladimir Abramovich Rokhlin in his thesis 1947 and independently by Shizuo Kakutani in 1943. The lemma can be used to build **Kakutani skyscrapers**, which are nice partitions associated to a transformation. This lemma allows to approximate an aperiodic transformation $T$ by a periodic transformations $T_n$. Just change $T$ on $T^{n-1}(B)$ so that $T_n^n(x) = x$ for all $x$. The theorem has been generalized by Donald Ornstein and Benjamin Weiss to higher dimensions like $\mathbb{Z}^d$ actions of measure preserving transformations where the periodicity assumption is replaced by the assumption that the action is **free**: for any $n \neq 0$, the set $T^n(x) = x$ has zero measure. See [83, 128, 151].

## 110. Lax approximation

On the group $\mathcal{X}$ of all measurable, invertible transformations on the $d$-dimensional **torus** $X = \mathbb{T}^d$ which preserve the Lebesgue volume measure, one has the metric

$$\delta(T, S) = |\delta(T(x), S(x))|_\infty ,$$

where $\delta$ is the geodesic distance on the flat torus and where $|\cdot|_\infty$ is the $L^\infty$ supremum norm. Lets call $(\mathbb{T}^d, T, \mu)$ a **toral dynamical system** if $T$ is a **homeomorphism**, a continuous transformation with continuous inverse. A **cube exchange transformation** on $\mathbb{T}^d$ is a periodic, piecewise affine measure-preserving transformation $T$ which permutes rigidly all the cubes $\prod_{i=1}^d [k_i/n, (k_i+1)/n]$, where $k_i \in \{0, \dots, n-1\}$. Every point in $\mathbb{T}^d$ is $T$ periodic. A cube exchange transformation is determined by a permutation of the set $\{1, \dots, n\}^d$. If it is cyclic, the exchange transformation is called **cyclic**. A theorem of Lax [221] states that every toral dynamical system can approximated in the metric $\delta$ by cube exchange transformations. The approximations can even be cyclic [12].

**Theorem:** Toral systems can be approximated by cyclic cube exchanges

The result is due to Peter Lax [221]. The proof of this result uses Hall's marriage theorem in graph theory (for a 'book proof' of the later theorem, see [8]). Periodic approximations of symplectic maps work surprisingly well for relatively small $n$ (see [273]). On the Pesin region this can be explained in part by the shadowing property [184]. The approximation by cyclic transformations make long time stability questions look different [150].

## 111. Sobolev embedding

All functions are defined on $\mathbb{R}^n$, integrated $\int$ over $\mathbb{R}^n$ and assumed to be **locally integrable** meaning that for every compact set $K$ the **Lebesgue integral** $\int_K |f|\, dx$ is finite. For functions in $C_c^\infty$ which serve as **test functions**, **partial derivatives** $\partial_i = \partial/\partial_{x_i}$ and more general **differential operators** $D^k = \partial_{x_1}^{k_1} \cdots \partial_{x_n}^{k_n}$ can be applied. A function $g$ is a **weak partial derivative** of $f$ if $\int f \partial_i \phi dx = -\int g \phi dx$ for all test functions $\phi$. For $p \in [1, \infty)$, the $L^p$ space is $\{f \mid \int |f|^p dx < \infty\}$. The **Sobolev space** $W^{k,p}$ is the set of functions for which all $k$'th weak derivatives are in $L^p$. So $W^{0,p} = L^p$. The **Hölder space** $C^{r,\alpha}$ with $r \in \mathbb{N}, \alpha \in (0,1]$ is defined as the set of functions for which all $r$'th derivatives are $\alpha$-Hölder continuous. It is a Banach space with norm $\max_{|k| \leq r} ||D^k f||_\infty + \max_{|k|=r} ||D^k f||_\alpha$, where $||f||_\infty$ is the **supremum norm** and $||f||_\alpha$ is the **Hölder coefficient** $\sup_{x \neq y} |f(x) - f(y)|/|x-y|^\alpha$. The **Sobolev embedding theorem** is

**Theorem:** If $n < p$ and $l = r + \alpha < k - n/p$, one has $W^{k,p} \subset C^{r,\alpha}$.

([297] states this as Theorem 6.3.6) gives some history: **generalized functions** appeared first in the work of Oliver Heaviside in the form of "operational calculus. Paul Dirac used the formalism in quantum mechanics. In the 1930s, Kurt Otto Friedrichs, Salomon Bocher and Sergei Sobolev define weak solutions of PDE's. Schwartz used the $C_c^\infty$ functions, smooth functions of compact support. This means that the existence of $k$ weak derivatives implies the existence of actual derivatives. For $p = 2$, the spaces $W^k$ are Hilbert spaces and the theory a bit simpler due to the availability of Fourier theory, where tempered distributions flourished. In that case, one can define for any real $s > 0$ the Hilbert space $H^s$ as the subset of all $f \in S'$ for which $(1 + |\xi|^2)^{s/2} \hat{f}(\xi)$ is in $L^2$. The Schwartz test functions $S$ consists of all $C^\infty$ functions having bounded semi norms $||\phi||_k = \max_{|\alpha|+|\beta| \leq k} ||x^\beta D^\alpha \phi||_\infty < \infty$ where $\alpha, \beta \in \mathbb{N}^n$. Since $S$ is larger than the set of smooth functions of compact support, the dual space $S'$ is smaller. They are **tempered distributions**. Sobolev emedding theorems like above allow to show that weak solutions of PDE's are smooth: for example, if the Poisson problem $\Delta f = Vf$ with smooth $V$ is solved by a distribution $f$, then $f$ is smooth. [48, 297]

## 112. Whitney embedding

A smooth $n$-**manifold** $M$ is a metric space equipped with a cover $U_j = \phi_j^{-1}(B)$ with $B = \{x \in \mathbb{R}^n \mid |x|^2 < 1\}$) or $U_j = \phi_j^{-1}(H)$ with $H = \{x \in \mathbb{R}^n \mid |x|^2 < 1, x_0 \geq 0\}$) with $\delta H = \{x \in H \mid x_0 = 0\}$ such that the homeomorphisms $\phi_j : U_j \to B$ or $\phi_j : U_j \to H$ lead to smooth transition maps $\phi_{kj} = \phi_j \phi_k^{-1}$ from $\phi_k(U_j \cap U_k)$ to $\phi_j(U_j \cap U_k)$ which have the property that all restrictions of $\phi_{kj}$ from $\delta\phi_k(U_j \cap U_k)$ to $\delta\phi_j(U_j \cap U_k)$ are smooth too. The **boundary** $\delta M$ of $M$ now naturally is a smooth $(n-1)$ manifold, the atlas being given by the sets $V_j = \phi_j(\delta H)$ for the indices $j$ which map $\phi_j : U_j \to H$. Two manifolds $M, N$ are **diffeomorphic** if there is a refinement $\{U_j, \phi_j\}$ of the atlas in $M$ and a refinement $\{V_j, \psi_j\}$ of the atlas in $N$ such that $\phi_j(U_j) = \psi_j(V_j)$. A manifold $M$ can be **smoothly embedded** in $\mathbb{R}^k$ if there is a smooth injective map $f$ from $M$ to $\mathbb{R}^k$ such that the image $f(M)$ is diffeomorphic to $M$.

**Theorem:** Any $n$-manifold $M$ can be smoothly embedded in $\mathbb{R}^{2n}$.

The theorem has been proven by Hassler Whitney in 1926 who also was the first to give a precise definition of manifold in 1936. The standard assumption is that $M$ is second countable Hausdorff but as every smooth finite dimensional manifold can be upgraded to be Riemannian, the simpler metric assumption is no restriction of generality. The modern point of view is to see $M$ as a **scheme** over Euclidean $n$-space, more precisely as a **ringed space**, that is locally the spectrum of the commutative ring $C^\infty(B)$ or $C^\infty(H)$. The set of manifolds is a **category** in which the smooth maps $M \to N$ are the **morphisms**. The cover $U_j$ defines an **atlas** and the transition maps $\phi_j$ allow to port notions like smoothness from Euclidean space to $M$. The maps $\phi_j^{-1} : B \to M$ or $\phi_j^{-1} : H \to M$ parametrize the sets $U_j$. [341].

## 113. Artificial intelligence

Like **meta mathematics** or **reverse mathematics**, the field of **artificial intelligence** (AI) can be considered as part of mathematics. It is related of **data science** (algorithms for data mining, and statistics) **computation theory** (like complexity theory) **language theory** and especially **grammar** and **evolutionary dynamics**, **optimization problems** (like solving optimal transport or extremal problems) **solving inverse problems** (like developing algorithms for computer vision or optical character or speech recognition), **cognitive science** as well as **pedagogy** in education (human or machine learning and human motivation). There is no apparent "fundamental theorem" of AI, (except maybe Marvin Minsky's *"The most efficient way to solve a problem is to already know how to solve it."* [243], which is a surprisingly deep statement as modern AI agents like **Alexa**, **Siri**, **Google Home**, **IBM Watson** or **Cortana** demonstrate; they compute little, they just know or look up - or annoy you to look it up yourself...). But there is a **theorem of Lebowski on machine super intelligence** which taps into the uncharted territory of **machine motivation**

**Theorem:** No AI will bother after hacking its own reward function.

The picture [211] is that once the AI has figured out the philosophy of the "Dude" in the Cohen brothers movie Lebowski, also repeated mischiefs does not bother it and it "goes bowling". Objections are brushed away with "Well, this is your, like, opinion, man". Two examples of human super intelligent units who have succeeded to hack their own reward function are Alexander Grothendieck or Grigori Perelman. The Lebowski theorem is due to Joscha Bach [22], who stated this **theorem of super intelligence** in a tongue-in-cheek tweet. From a mathematical point of view, the smartest way to "solve" an optimal transport problem is to change the utility function. On a more serious level, the smartest way to "solve" the continuum hypothesis is to change the axiom system. This is a cheat, but on a meta level, more creativity is possible. A precursor is Stanislav Lem's notion of a **mimicretin** [223], a computer that plays stupid in order, once and for all, to be left in peace or the machine in [5] who develops humor and enjoys fooling humans with the answer to the ultimate question: "42". This document entry is the analogue to the ultimate question: "What is the fundamental theorem of AI"?

## 114. Stokes Theorem

On a smooth orientable $n$-dimensional manifold $M$, one has $\Lambda^p$, the vector bundle of smooth **differential $p$-forms**. As any $p$-form $F$ induces an **induced volume form** on a $p$-dimensional **sub-manifold** $G$ defining so an **integral** $\int_G F$. The **exterior derivative** $d : \Lambda^p \to \Lambda^{p+1}$ satisfies $d^2 = 0$ and defines an **elliptic complex**. There is a natural **Hodge duality** isomorphism given called "Hodge star" $* : \Lambda^p \to \Lambda^{n-p}$. Given a $p$-form $F \in \Lambda^p$ and a $(p+1)$-dimensional compact oriented sub-manifold $G$ of $M$ with boundary $\delta G$ compatible with the orientation of $G$, we have **Stokes theorem**:'

**Theorem:** $\langle G, dF\rangle = \int_G dF = \int_{\delta G} F = \langle \delta G, F\rangle$.

The theorem states that the exterior derivative $d$ is dual to the boundary operator $\delta$. If $G$ is a connected 1-manifold with boundary, it is a curve with boundary $\delta G = \{A, B\}$. A 1-form can be integrated over the curve $G$ by choosing the on $G$ induced volume form $r'(t)dt$ given by a **curve parametrization** $[a, b] \to G$ and integrate $\int_a^b F(r(t)) \cdot r'(t)dt$, which is the **line integral**. Stokes theorem is then the **fundamental theorem of line integrals**. Take a 0-form $f$ which is a **scalar function** the derivative $df$ is the gradient $F = \nabla f$. Then $\int_a^b \nabla f(r(t)) \cdot r'(t)\, dt = f(B) - f(A)$. If $G$ is a two dimensional surface with boundary $\delta G$ and $F$ is a 1-form, then the 2-form $dF$ is the **curl** of $F$. If $G$ is given as a **surface parametrization**

$r(u,v)$, one can apply $dF$ on the pair of tangent vectors $r_u, r_v$ and integrate this $dF(r_u, r_v)$ over the surface $G$ to get $\int_G dF$. The **Kelvin-Stokes theorem** tells that this is the same than the line integral $\int_{\delta G} F$. In the case of $M = \mathbb{R}^3$, where $F = Pdx + Qdy + Rdz$ can be identified with a vector field $F = [P, Q, R]$ and $dF = \nabla \times F$ and integration of a 2-form $H$ over a parametrized manifold $G$ is $\int \int_R H(r(u,v))(r_u, r_v) = \int \int_R H(r(u,v) \cdot r_u \times r_v dudv$ we get the **classical Kelvin-Stokes theorem.** If $F$ is a 2-form, then $dF$ is a 3-form which can be integrated over a 3-manifold $G$. As $d : \Lambda^2 \to \Lambda^3$ can via Hodge duality naturally be paired with $d_0^* : \Lambda^1 \to \Lambda^0$, which is the **divergence**, the **divergence theorem** $\int \int \int_G \text{div}(F)\ dxdydz = \int \int_{\delta G} F \cdot dS$ relates a triple integral with a flux integral. Historical milestones start with the development of the **fundamental theorem of calculus** (1666 Isaac Newton, 1668 James Gregory, Isaac Barrow 1670 and Gottfried Leibniz 1693); the first rigorous proof was done by Cauchy in 1823 (the first textbook appearance in 1876 by Paul du Bois-Reymond). See [46]. In 1762, Joseph-Louis Lagrange and in 1813 Karl-Friedrich Gauss look at special cases of divergence theorem, Mikhail Ostogradsky in 1826 and George Green in 1828 cover the general case. Green's theorem in two dimensions was first stated by Augustin-Louis Cauchy in 1846 and Bernhard Riemann in 1851. Stokes theorem first appeared in 1854 an exam question but the theorem has appeared already in a letter of William Thomson to Lord Kelvin in 1850, hence also the name **Kelvin-Stokes theorem**. Vito Volterra in 1889 and Henri Poincaré in 1899 generalized the theorems to higher dimensions. Differential forms were introduced in 1899 by Élie Cartan. The $d$ notation for exterior derivative was introduced in 1902 by Theodore de Donder. The ultimate formulation above is from Cartan 1945. We followed Katz [189] who noticed that only in 1959, this version has started to appear in textbooks.

## 115. Moments

The **Hausdorff moment problem** asks for necessary and sufficient conditions for a sequence $\mu_n$ to be realizable as a moment sequence $\int_0^1 x^n\ d\mu(x)$ for a Borel probability measure on $[0,1]$. One can study the problem also in higher dimensions: for a multi-index $n = (n_1, \dots, n_d)$ denote by $\mu_n = \int x_1^{n_1} \dots x_d^{n_d}\ d\mu(x)$ the $n$**'th moment** of a **signed Borel measure** $\mu$ on the unit cube $I^d = [0,1]^d \subset \mathbb{R}^d$. We say $\mu_n$ is a **moment configuration** if there exists a measure $\mu$ which has $\mu_n$ as moments. If $e_i$ denotes the standard basis in $\mathbb{Z}^d$, define the **partial difference** $(\Delta_i a)_n = a_{n-e_i} - a_n$ and $\Delta^k = \prod_i \Delta_i^{k_i}$. We write $\frac{k}{n} = \prod_{i=1}^n \frac{k_i}{n_i}$ and $\begin{pmatrix} n \\ k \end{pmatrix} = \prod_{i=1}^d \begin{pmatrix} n_i \\ k_i \end{pmatrix}$ and $\sum_{k=0}^n = \sum_{k_1=0}^{n_1} \cdots \sum_{k_d=0}^{n_d}$. We say moments $\mu_n$ are **Hausdorff bounded** if there exists a constant $C$ such that $\sum_{k=0}^n | \begin{pmatrix} n \\ k \end{pmatrix} (\Delta^k \mu)_n| \le C$ for all $n \in \mathbb{N}^d$. The **theorem of Hausdorff-Hildebrandt-Schoenberg** is

**Theorem:** Hausdorff bounded moments $\mu_n$ belong to a measure $\mu$.

The above result is due to Theophil Henry Hildebrandt and Isaac Jacob Schoenberg from 1933. [164]. Moments also allow to compare measures: a measure $\mu$ is called **uniformly absolutely continuous** with respect to $\nu$ if there exists $f \in L^\infty(\nu)$ such that $\mu = f\nu$. A positive probability measure $\mu$ is uniformly absolutely continuous with respect to a second probability measure $\nu$ if and only if there exists a constant $C$ such that $(\Delta^k \mu)_n \le C \cdot (\Delta^k \nu)_n$ for all $k, n \in \mathbb{N}^d$. In particular it gives a generalization of a result of Felix Hausdorff from 1921 [159] assuring that $\mu$ is positive if and only if $(\Delta^k \mu)_n \ge 0$ for all $k, n \in \mathbb{N}^d$. An other special case is that $\mu$ is uniformly absolutely continuous with respect to Lebesgue measure $\nu$ on $I^d$ if and only if

$|\Delta^k \mu_n| \leq \left( \begin{array}{c} n \\ k \end{array} \right) (n+1)^d$ for all $k$ and $n$. Moments play an important role in statistics, when looking at **moment generating functions** $\sum_n \mu_n t^n$ of random variables $X$, where $\mu_n = \mathrm{E}[X^n]$ as well as in **multivariate statistics**, when looking at random vectors $(X_1, \dots, X_d)$, where $\mu_n = \mathrm{E}[X_1^{n_1} \cdots X_d^{n_d}]$ are **multivariate moments**. See [199]

## 116. Martingales

A sequence of random variables $X_1, X_2, \dots$ on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ is called a **discrete time stochastic process**. We assume the $X_k$ to be in $L^2$ meaning that the expectation $\mathrm{E}[X_k^2] < \infty$ for all $k$. Given a sub-$\sigma$ algebra $\mathcal{B}$ of $\mathcal{A}$, the **conditional expectation** $\mathrm{E}[X|\mathcal{B}]$ is the projection of $L^2(\Omega, \mathcal{A}, P)$ to $L^2(\Omega, \mathcal{B}, P)$. Extreme cases are $\mathrm{E}[X|\mathcal{A}] = X$ and $\mathrm{E}[X|\{\emptyset, \Omega\}] = \mathrm{E}[X]$. A finite set $Y_1, \dots, Y_n$ of random variables generates a sub- $\sigma$-algebra $\mathcal{B}$ of $\mathcal{A}$, the smallest $\sigma$-algebra for which all $Y_j$ are still measurable. Write $\mathrm{E}[X|Y_1, \cdots, Y_n] = \mathrm{E}[X|\mathcal{B}]$, where $\mathcal{B}$ is the $\sigma$-algebra generated by $Y_1, \cdots Y_n$. A discrete time stochastic process is called a **martingale** $\mathrm{E}[X_{n+1}|X_1, \cdots, X_n] = \mathrm{E}[X_n]$ for all $n$. If the equal sign is replaced with $\leq$ then the process is called a **super-martingale**, if $\geq$ it is a **sub-martingale**. The **random walk** $X_n = \sum_{k=1}^n Y_k$ defined by a sequence of independent $L^2$ random variables $Y_k$ is an example of a martingale because independence implies $\mathrm{E}[X_{n+1}|X_1, \cdots, X_n] = \mathrm{E}[X_{n+1}]$ which is $\mathrm{E}[X_n]$ by the identical distribution assumption. If $X$ and $M$ are two discrete time stochastic processes, define the **martingale transform** (=discrete Ito integral) $X \cdot M$ as the process $(X \cdot M)_n = \sum_{k=1}^n X_k (M_k - M_{k-1})$. If the process $X$ is **bounded** meaning that there exists a constant $C$ such that $\mathrm{E}[|X_k|] \leq C$ for all $k$, then if $M$ is a martingale, also $X \cdot M$ is a martingale. The **Doob martingale convergence theorem** is

**Theorem:** For a bounded super martingale $X$, then $X_n$ converges in $L^1$.

The convergence theorem can be used to prove the **optimal stopping time theorem** which tells that the expected value of a **stopping time** is the initial expected value. In finance it is known as the **fundamental theorem of asset pricing**. If $\tau$ is a stopping time adapted to the martingale $X_k$, it defines the random variable $X_\tau$ and $\mathrm{E}[X_\tau] = \mathrm{E}[X_0]$. For a super-martingale one has $\geq$ and for a sub-martingale $\leq$. The proof is obtained by defining the **stopped process** $X_n^\tau = X_0 + \sum_{k=0}^{\min(\tau,n)-1} (X_{k+1} - X_k)$ which is a martingale transform and so a martingale. The martingale convergence theorem gives a limiting random variable $X_\tau$ and because $\mathrm{E}[X_n^\tau] = \mathrm{E}[X_0]$ for all $n$, $\mathrm{E}[X_\tau] = \mathrm{E}[X_0]$. This is rephrased as "you can not beat the system" [342]. A trivial implication is that one can not for example design a strategy allowing to win in a fair game by designing a "clever stopping time" like betting on "red" in roulette if 6 times "black" in a row has occurred. Or to follow the strategy to always to stop the game, if one has a first positive total win, which one can always do by doubling the bet in case of losing a game. Martingales were introduced by Paul Lévy in 1934, the name "martingale" (referring to the just mentioned doubling betting strategy) was added in a 1939 probability book of Jean Ville. The theory was developed by Joseph Leo Doob in his book of 1953. [98]. See [342].

## 117. Theorema Egregium

A Riemannian metric on a two-dimensional manifold $S$ defines the quadratic form $I = Edu^2 + 2Fdudv + Gdv^2$ called **first fundamental form** on the surface. If $r(u,v)$ is a parameterization of $S$, then $E = r_u \cdot r_u, F = r_u \cdot r_v$ and $G = r_v \cdot r_v$. The **second fundamental form** of $S$ is $II = Ldu^2 + 2Mdudv + Ndv^2$, where $L = r_{uu} \cdot n, M = r_{uv} \cdot n, N = r_{vv} \cdot n$, written using the normal vector $n = (r_u \times r_v)/|r_u \times r_v|$. The **Gaussian curvature** $K = \det(II)/\det(I) = (LN - M^2)/(EG - F^2)$. depends on the embedding $r : R \to S$ in space $\mathbb{R}^3$, but it actually only depends on the intrinsic metric, the first fundamental form. This is the **Theorema egregium** of Gauss:

**Theorem:** The Gaussian curvature only depends on the Riemannian metric.

Gauss himself already gave explicit formulas, but a formula of **Brioschi** gives the curvature $K$ explicitly as a ratio of determinants involving $E, F, G$ as well as and first and second derivatives of them. In the case when the surface is given as a graph $z = f(x, y)$, one can give $K = D/(1 + |\nabla f|^2)^2$, where $D = (f_{xx}f_{yy} - f_{xy}^2)$ is the **discriminant** and $(1 + |\nabla f|^2)^2 = \det(II)$. If the surface is rotated in space so that $(u, v)$ is a critical point for $f$, then the **discriminant** $D$ is equal to the curvature. One can see the independence of the embedding also from the **Puiseux formula** $K = 3(|S_0(r)| - S(r))/(\pi r^3)$, where $|S_0(r)| = 2\pi r$ is the circumference of the circle $S_0(r)$ in the flat case and $|S(r)|$ is the circumference of the **geodesic circle** of radius $r$ on $S$. The theorem Egregium also follows from Gauss-Bonnet as the later allows to write the curvature in terms of the angle sum of a geodesic infinitesimal triangle with the angle sum $\pi$ of a flat triangle. As the angle sums are entirely defined intrinsically, the curvature is intrinsic. The Theorema Egregium was found by Karl-Friedrich Gauss in 1827 and published in 1828 in "Disquisitiones generales circa superficies curvas". It is not an accident, that Gauss was occupied with concrete geodesic triangulation problems too.

## 118. Entropy

Given a random variable $X$ on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$ which is **discrete** in the sense that takes only finitely many values, the **entropy** is defined as $S(X) = -\sum_x p_x \log(p_x)$, where $p_x = \mathrm{P}[X = x]$. To compare, for a random variable $X$ with cumulative distribution function $F(x) = \mathrm{P}[X < x]$ has a continuous derivative $F' = f$, the entropy is defined as $S(X) = -\int f(x)\log(f(x))\,dx$ allowing the value $-\infty$ if the integral does not converge. (We always read $p\log(p) = 0$ if $p = 0$.) In the continuous case, one also calls this the **differential entropy**. Two discrete random variables $X, Y$ are called **independent** if one can realize them on a product probability space $\Omega = A \times B$ so that $X(a, b) = X(a)$ and $Y(a, b) = Y(b)$ for some functions $X : A \to \mathbb{R}, Y : B \to \mathbb{R}$. Independence implies that the random variables are uncorrelated $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$ and that the **entropy adds up** $S(XY) = S(X) + S(Y)$. We can write $S(X) = \mathrm{E}[\log(W(x))]$, where $W$ is the "Wahrscheinlichkeits" random variable assigning to $\omega \in \Omega$ the value $W(\omega) = 1/p_x$ if $X(\omega) = x$. Let us say, a functional on discrete random variables is **additive** if it is of the form $H(X) = \sum_x f(p_x(X))$ for some continuous function $f$ for which $f(t)/t$ is monotone. We say it is **multiplicative** if $H(XY) = H(X) + H(Y)$ for independent random variables. The functional is **normalized** if $H(X) = \log(4)$ if $X$ is the random variable taking two values $\{0, 1\}$ with probability $p_0 = p_1 = 1/2$. Shannon's theorem is:

**Theorem:** Any normalized, additive and multiplicative $H$ is entropy $S$.

The word "entropy" was introduced by Rudolf Clausius in 1850 [281]. Ludwig Bolzmann saw the importance of $d/dtS \geq 0$ in the context of heat and wrote in 1872 $S = k_B \log(W)$, where $W(x) = 1/p_x$ is the inverse "Wahrscheinlichkeit" that a state has the value $x$. His equation is understood as the expectation $S = k_B\mathrm{E}[\log(W)] = \sum_x p_x \log(W(x))$ which is the **Shannon entropy**, introduced in 1948 by Claude Shannon in the context of information theory. (Shannon characterized functionals $H$ with the property that if $H$ is continuous in $p$, that for random variables $H_n$ with $p_x(H_n) = 1/n$, one has $H(X_n)/n \leq H(X_m)/m$ if $n \leq m$ and that if $X, Y$ are two random variables so that the finite $\sigma$-algebras $\mathcal{A}$ defined by $X$ is a sub-$\sigma$-algebra $\mathcal{B}$ defined by $Y$ then $H(Y) = H(X) + \sum_x p_x H(Y_x)$, where $Y_x(\omega) = Y(\omega)$ for $\omega \in \{X = x\}$. One can show that these Shannon conditions are equivalent to the combination of being additive and multiplicative.) In statistical thermodynamics, where $p_x$ is the probability of a **micro-state**, then $k_B S$ is also called the **Gibbs entropy**, where $k_B$ is the **Boltzmann constant**. For general random variables $X$ on $(\Omega, \mathcal{A}, \mathrm{P})$ and a finite $\sigma$-sub-algebra $\mathcal{B}$, Gibbs looked in 1902 at **course grained entropy**, which is the entropy of the conditional expectation $Y = \mathrm{E}[X|\mathcal{B}|$, which is now a random variable $Y$ taking only finitely many values so that entropy is defined. See [292].

## 119. Mountain Pass

Let $H$ be a **Hilbert space**, and $f$ is a twice Fréchet differentiable function from $H$ to $\mathbb{R}$. The **Fréchet derivative** $A = f'$ at a point $x \in H$ is a linear operator satisfying $f(x+h) - f(x) - Ah = o(h)$ for all $h \to 0$. A point $x \in H$ is called a **critical point** of $f$ if $f'(x) = 0$. The functional satisfies the **Palais-Smale condition**, if every sequence $x_k$ in $H$ for which $\{f(x_k)\}$ is bounded and $f'(x_k) \to 0$, has a convergent subsequence in the closure of $\{x_k\}_{k \in \mathbb{N}}$. A pair of points $a, b \in H$ defines a **mountain pass**, if there exists $\epsilon > 0$ and $r > 0$ such that $f(x) \geq f(a) + \epsilon$ on $S_r(a) = \{x \in H \mid ||x - a|| = r\}$, $f$ is not constant on $S_r(a)$ and $f(b) \leq f(a)$. A critical point is called a **saddle** if it is neither a maximum nor a minimum of $f$.

**Theorem:** If a Palais-Smale $f$ has a mountain pass, it features a saddle.

The idea is to look at all continuous paths $\gamma$ from $a$ to $b$ parametrized by $t \in [0,1]$. For each path $\gamma$, the value $c_\gamma = f(\gamma(t))$ has to be maximal for some time $t \in [0,1]$. The infimum over all these critical values $c_\gamma$ is a critical value of $f$. The mountain pass condition leads to a "mountain ridge" and the critical point is a "mountain pass", hence the name. The example $(2\exp(-x^2 - y^2) - 1)(x^2 + y^2)$ with $a = (0,0), b = (1,0)$ shows that the non-constant condition is necessary for a saddle point on $S_r(a)$ with $r = 1/2$. The reason for sticking with a Hilbert space is that it is easier to realize the compactness condition due to weak star compactness of the unit ball. But it is possible to weaken the conditions and work with a Banach manifolds $X$ continuous Gâteaux derivatives: $f' : X \to X^*$ if $X$ has the strong and $X^*$ the weak-$*$ topology. It is difficult to pinpoint historically the first use of the mountain pass principle as it must have been known intuitively since antiquity. The crucial Palais-Smale **compactness condition** which makes the theorem work in infinite dimensions appeared in 1964. [21] call it condition (C), a notion which already appeared in the original paper [262].

## 120. Exponential sums

Given a smooth function $f : \mathbf{R} \to \mathbf{R}$ which maps integers to integers, one can look at **exponential sums** $\sum_{x=a}^{b} \exp(i\pi f(x))$ An example is the **Gaussian sum** $\sum_{x=0}^{n-1} \exp(i\alpha x^2)$. There are lots of interesting relations and estimates. One of the magical formulas are the **Landsberg-Schaar relations** for the finite sums $S(q,p) = \frac{1}{\sqrt{p}} \sum_{x=0}^{p-1} \exp(i\pi x^2 q/p)$.

**Theorem:** If $p, q$ are odd, then $S(2q, p) = e^{i\pi/4} S(-p, 2q)$.

One has $S(1,p) = (1/\sqrt{p})\sum_{x=0}^{p-1}\exp(ix^2/p) = 1$ for all positive integers $p$ and $S(2,p) = (e^{i\pi/4}/\sqrt{p})\sum_{x=0}^{p-1}\exp(2ix^2/p) = 1$ if $p = 4k+1$ and $i$ if $p = 4k-1$. The method of exponential sums has been expanded especially by Vinogradov's papers [329] and used for number theory like for quadratic reciprocity [251]. The topic is of interest also outside of number theory. Like in dynamical systems theory as Fürstenberg has demonstrated. An ergodic theorist would look at the dynamical system $T(x,y) = (x+2y+1, y+1)$ on the 2-torus $\mathbb{T}^2 = \mathbb{R}^2/(\pi\mathbb{Z})^2$ and define $g_\alpha(x,y) = \exp(i\pi x\alpha)$. Since the orbit of this toral map is $T^n(1,1) = (n^2,n)$, the exponential sum can be written as a **Birkhoff sum** $\sum_{k=0}^{p-1} g_{q/p}(T^k(1,1))$ which is a particular orbit of a stochastic process or deterministic random walk. Results as mentioned above show that the random walk grows like $\sqrt{p}$, similarly as in a random setting. Now, since the dynamical system is minimal, the growth rate should not depend on the initial point and $\pi q/p$ should be replaceable by any irrational $\alpha$ and no more be linked to the length of the orbit. The problem is then to study the growth rate of the **stochastic process** $S^t(x,y) = \sum_{k=0}^{p-1} g(T^k(x,y))$ (= sequence of random variables) for any continuous $g$ with zero expectation which by Fourier boils down to look at exponential sums. Of course $S^t(x,y)/t \to 0$ by Birkhoff's ergodic theorem, but as in the law of iterated log one is interested in precise growth rates. This can be subtle. Already in the simpler case of an integrable $T(x) = x + \alpha$ on the 1-torus, there is Denjoy-Koskma theory which shows that the growth rate depends on Diophantine properties of $\pi\alpha$. Unlike for irrational rotations, the Fürstenberg type skew systems $T$ leading to the theta functions are not integrable: it is not conjugated to a group translation (there is some randomness, even-so weak as Kolmogorov-Sinai entropy is zero). The dichotomy between structure and randomness and especially the similarities between dynamical and number theoretical set-ups has been discussed in [320].

## 121. Sphere theorem

A compact **Riemannian manifold** $M$ is said to have **positive curvature**, if all **sectional curvatures** are positive. The **sectional curvature** at a point $x \in M$ in the direction of the 2-dimensional plane $\Sigma \subset T_xM$ is defined as the Gaussian curvature of the surface $\exp_x(\Sigma) \subset M$ at the point. In terms of the **Riemannian curvature tensor** $R : T_xM^4 \to \mathbb{R}$ and an orthonormal basis $\{u,v\}$ spanning $\Sigma$, this is $R(u,v,u,v)$. The curvature is called **quarter pinched**, if it the sectional curvature is in the interval $(1,4]$ at all points $x \in M$. In particular, a quarter pinched manifold is a manifold with positive curvature. We say here, a compact Riemannian manifold **is a sphere** if it is homeomorphic to a sphere. The **sphere theorem** is:

**Theorem:** A simply-connected quarter pinched manifold is a sphere

The theorem was proven by Marcel Berger and Wilhelm Klingenberg in 1960. That a pinching condition would imply a manifold to be a sphere had been conjectured already by Heinz Hopf. Hopf himself proved in 1926 that constant sectional curvature implies that $M$ is even isometric to a sphere. Harry Rauch, after visiting Hopf in Zürich in the 1940's proved that a 3/4-pinched simply connected manifold is a sphere. In 2007, Simon Brendle and Richard Schoen proved that the theorem even holds if the statement $M$ **is a sphere** means $M$ **is diffeomorphic to a sphere**. This is the **differentiable sphere theorem**. Since John Milnor had given in 1956 examples of spheres which are homeomorphic but not diffeomorphic to the standard sphere (so called **exotic spheres**, spheres which carry a smooth maximal atlas different from the standard one), the differentiable sphere theorem is a substantial improvement on the topological sphere theorem. It needed completely new techniques, especially the **Ricci flow** $\dot{g} = -2\text{Ric}(g)$ of

Richard Hamilton which is a weakly parabolic partial differential equation deforming the metric $g$ and uses the **Ricci curvature** Ric of $g$. See [32, 45].

## 122. Word problem

The **word problem** in a **finitely presented group** $G = (g|r)$ with **generators** $g$ and **relations** $r$ is the problem to decide, whether a given set of two words $v, w$ represent the same group element in $G$ or not. The word problem is not solvable in general. There are concrete finitely presented groups in which it is not. The following theorem of Boone and Higman relates the solvability to algebra. A group is **simple** if its only **normal subgroup** is the trivial group or the group itself.

**Theorem:** Finitely presented simple groups have a solvable word problem.

More generally, if $G \subset H \subset K$ where $H$ is simple and $K$ is finitely presented, then $G$ has a solvable word problem. Max Dehn proposed the word problem in 1911. Pyotr Novikov in 1955 proved that the word problem is undecidable for finitely presented groups. William W. Boone and Graham Higman proved the theorem in 1974 [38]. Higman would in the same year also find an example of an infinite finitely presented simple group. The non-solvability of the word problem implies the non-solvability of the homeomorphism problem for $n$-manifolds with $n \geq 4$. See [348].

## 123. Finite simple groups

A **finite group** $(G, *, 1)$ is a finite set $G$ with an operation $* : G \times G \to G$ and 1 **element**, such that the operation is **associative** $(a*b)*c = a*(b*c)$, for all $a, b, c$, such that $a*1 = 1*a = a$ for every $a$ and such that every $a$ has an inverse $a^{-1}$ satisfying $a*a^{-1} = 1$. A group $G$ is **simple** if the only **normal subgroups** of $G$ are the **trivial group** $\{1\}$ or the group itself. A subgroup $H$ of $G$ is called **normal** if $gH = Hg$ for all $g$. Simple groups play the role of the primes in the integers. A theorem of Jordan-Hölder is that a decomposition of $G$ into simple groups is essentially unique up to permutations and isomorphisms. The **classification theorem of finite simple groups** is

**Theorem:** Every finite simple group is cyclic, alternating, Lie or sporadic.

There are 18 so called **regular families** of finite simple groups made of **cyclic**, **alternating** and 16 **Lie type** groups. Then there are 26 so called **sporadic groups**, in which 20 are **happy groups** as they are subgroups or sub-quotients of the **monster** and 6 are **pariahs**, outcasts which are not under the spell of the monster. The classification was a huge collaborative effort with more than 100 authors covering 500 journal articles. According to Daniel Gorenstein, the classification was completed 1981 and fixes were applied until 2004 (Michael Aschbacher and Stephen Smith resolving the last problems which lasted several years) leading to a full proof of 1300 pages. A second generation cleaned-out proof written with more details is under way and currently has 5000 pages. Some history is given in [302].

## 124. God number

Given a finite finitely presented group $G = (g|r)$ like for example the Rubik group. It defines the **Cayley graph** $\Gamma$ in which the group elements are the nodes and where two nodes $a, b$ are connected if there is a generator $x$ in in $g$ such that $xa = b$. The **diameter** of a graph is the largest geodesic distance between two nodes in $\Gamma$. It is also called **God number** of the puzzle. The **Rubik cube** is an example of a finitely presented group. The original $3 \times 3 \times 3$ cube allows to permute the 26 boundary cubes using the 18 possible rotations of the 6 faces as generators. From the $X = 8!12!3^8 2^{12}$ possible ways to physically build the cube, only $|G| = X/12 = 43252003274489856000$ are present in the Rubik group $G$. Some of the positions "quarks" [136] can not be realized but combinations of them "mesons" or "baryons" can.

**Theorem:** The God number of the Rubik cube is 20.

This means that from any position, one could, in principle solve the puzzle in 20 moves. Note that one has to specify clearly the generators of the group as this defines the Cayley graph and so a metric on the group. The lower bound 18 had already been known in 1980 as counting the possible moves with 17 moves produces less elements. The lower bound 20 came in 1995 when Michael Reid proved that the **superflip position** (where the edges are all flipped but corners are correct) needs 20 moves. In July 2010, using about 35 CPU years, a team around Tomas Rokicki established that the God number is 20. They partitioned the possible group positions into roughly 2 billion sets of 20 billions positions each. Using symmetry the reduced it to 55 million positions, then found solutions for any of the positions in these sets. [114] It appears silly to put a God number computation as a fundamental theorem, but the status of the Rubik cube is enormous as it has been one of the most popular puzzles for decades and is a **prototype** for many other similar puzzles, the choice can be defended. [1] One can ask to compute the god number of any finitely presented finite group. Interesting in general is the complexity of evaluating that functional. Something easier: the simplest nontrivial **Rubik cuboid** is the $2 \times 2 \times 1$ one. It has 6 positions and 2 generators $a, b$. The finitely presented group is $\{a, b | a^2 = b^2 = (ab)^3 = 1\}$ which is the **dihedral group** $D_3$. Its group elements are $G = \{1, a = babab, ab = baba, aba = bab, abab = ba, ababa = b\}$. The group is isomorphic to the **symmetry group of the equilateral triangle**, generated by the two reflections $a, b$ at two altitude lines. The God number of that group can be seen easily to be 3 because the Cayley graph $\Gamma$ is the cyclic graph $C_6$. It is funny that the puzzle solver has here "no other choice than solving the puzzle" than to make non-trivial move in each step. See [180] or [26] for general combinatorial group theory.

## 125. Sard Theorem

Let $f : M \to N$ be a smooth map between smooth manifolds $M, N$ of dimension $\dim(M) = m$ and $\dim(N) = n$. A point $x \in M$ is called a **critical point** of $f$, if the Jacobian $n \times m$ matrix $df(x)$ has rank both smaller than $m$ and $n$. If $C$ is the set of critical points, then $f(C) \subset N$ is called the **critical set** of $f$. The **volume measure** on $N$ is a choice of a volume form, obtained for example after introducing a Riemannian metric. **Sard**'s theorem is

**Theorem:** The critical set of $f : M \to N$ has zero volume measure in $N$.

The theorem applied to smooth map $f : M \to \mathbb{R}$ tells that for almost all $c$, the set $f^{-1}(c)$ is a smooth hypersurface of $M$ or then empty. The later can happen if $f$ is constant. We assumed $C^\infty$ but one can relax the smoothness assumption of $f$. If $n \geq m$, then $f$ needs only to be continuously differentiable. If $n < m$, then $f$ needs to be in $C^{m-n+1}$. The case when $N$

[1] I presented the God number problem in the 80ies as an undergraduate in a logic seminar of Ernst Specker and the choice of topic had been objected to by Specker himself as a too "narrow problem". But for me, the Rubik cube and its group theoretical properties have "cult status" and was one of the triggers to study math.

is one-dimensional has been covered by Antony Morse (who is unrelated to Marston Morse) in 1939 and by Arthur Sard in general in 1942. A bit confusing is that Marston Morse (not Antony) covered the case $m = 1, 2, 3$ and Sard in the case $m = 4, 5, 6$ in unpublished papers before as mentioned in a footnote to [287]. Sard also notes already that examples of Hassler Whitney show that the smoothness condition can not be relaxed. Sard formulated the results for $M = \mathbb{R}^m$ and $N = \mathbb{R}^n$ (by the way with the same choice $f : M \to N$ as done here and not as in many other places). The manifold case appears for example in [308].

## 126. Elliptic curves

An **elliptic curve** is a plane algebraic curve defined by the points satisfying the **Weierstrass equation** $y^2 = x^3 + ax + b = f(x)$. One assumes the curve to be **non-singular**, meaning that the **discriminant** $\Delta = -16(4a^3 + 27b^2)$ is not zero. This assures that there are no cusps nor multiple roots for the simple reason that the explicit solution formulas for roots of $f(x) = 0$ involves only square roots of $\Delta$. A curve is an **Abelian variety**, if it carries an Abelian algebraic group structure, meaning that the addition of a point defines a morphism of the variety.

**Theorem:** Elliptic curves are Abelian varieties.

The theorem seems first have been realized by Henry Poincaré in 1901. Weierstrass before had used the Weierstrass $\mathcal{P}$ function earlier in the case of elliptic curves over the complex plane. To define the group multiplication, one uses the **chord-tangent construction**: first add point $O$ called the **point at infinity** which serves as the **zero** in the group. Then define $-P$ as the point obtained by reflecting at the $x$-axes. The **group multiplication** between two different points $P, Q$ on the curve is defined to be $-R$ if $R$ is the point of intersection of the line through $P, Q$ with the curve. If $P = Q$, then $R$ is defined to be the intersection of the tangent with the curve. If there is no intersection that is if $P = Q$ is an inflection point, then one defines $P + P = -P$. Finally, define $P + O = O + P = P$ and $P + (-P) = 0$. This recipe can be explicitly given in coordinates allowing to define the multiplication in any field of characteristic different from 2 or 3. The group structure on elliptic curves over finite fields provides a rich source of **finite Abelian groups** which can be used for cryptological purposes, the so called **elliptic curve cryptograph** ECC. Any procedure, like public key, Diffie-Hellman or factorization attacks on integers can be done using groups given by elliptic curves. [331].

## 127. Billiards

Billiards are the geodesic flow on a smooth compact $n$-manifold $M$ with boundary. The dynamics is extended through the boundary by applying the law of reflection. While the flow of the geodesic $X^t$ is Hamiltonian on the unit tangent bundle $SM$, the billiard flow is only piecewise smooth and also the return map to the boundary is not continuous in general but it is a map preserving a natural volume so that one can look at ergodic theory. Already difficult are flat 2-manifolds $M$ homeomorphic to a disc having convex boundary homeomorphic to a circle. For smooth convex tables this leads to a return map $T$ on the annulus $X = \mathbb{T} \times [-1, 1]$ which is $C^{r-1}$ smooth if the boundary is $C^r$ [99]. It defines a **monotone twist map**: in the sense that it preserves the boundary, is area and orientation preserving and satisfies the **twist condition** that $y \to T(x, y)$ is strictly monotone. A **Bunimovich stadium** is the 2-manifold with boundary obtained by taking the convex hull of two discs of equal radius in $\mathbb{R}$ with different center. The billiard map is called **chaotic**, if it is ergodic and the **Kolmogorov-Sinai**

**entropy** is positive. By Pesin theory, this metric entropy is the **Lyapunov exponent** which is the exponential growth rate of the Jacobian $dT^n$ (and constant almost everywhere due to ergodicity). There are coordinates in the tangent bundle of the annulus $X$ in which $dT$ is the composition of a horizontal shear with strength $L(x, y)$, where $L$ is the trajectory length before the impact with a vertical shear with strength $-2\kappa/\sin(\theta)$ where $\kappa(x)$ is the curvature of the curve at the impact $x$ and $y = \cos(\theta)$, with **impact angle** $\theta \in [0, \pi]$ between the tangent and the trajectory.

**Theorem:** The Bunimovich stadium billiard is chaotic.

Jacques Hadmard in 1898 and Emile Artin in 1924 already looked at the geodesic flow on a surface of constant negative curvature. Yakov Sinai constructed in 1970 the first chaotic billiards, the Lorentz gas or Sinai billiard. An example, where Sinai's result applies is the hypocycloid $x^{1/3} + y^{1/3} = 1$. The Bernoulli property was established by Giovanni Gallavotti and Donald Ornstein in 1974. In 1973, Vladimir Lazutkin proved that a generic smooth convex two-dimensional billiard can not be ergodic due to the presence of KAM whisper galleries using Moser's twist map theorem. These galleries are absent in the presence of flat points (by a theorem of John Mather) or points, where the curvature is unbounded (by a theorem of Andrea Hubacher). Leonid Bunimovich constructed in 1979 the first convex chaotic billiard. No smooth convex billiard table with positive Kolmogorov-Sinai entropy are known. A candidate is the real analytic $x^4 + y^4 = 1$. Various generalizations have been considered like in [343]. A detailed proof that the Bunimovich stadium is measure theoretically conjugated to a Bernoulli system (the shift on a product space) is surprisingly difficult: one has to show positive Lyapunov exponents on a set of positive measure. Applying Pesin theory with singularities (Katok-Strelcyn theory) gives a Markov process. One needs then to establish ergodicity using a method of Eberhard Hopf of 1936 which requires to understand stable and unstable manifolds [66]. See [318, 330, 248, 135, 184, 66] for sources on billiards.

## 128. Uniformization

A **Riemann surface** is a one-dimensional **complex manifold**. It is **simply connected** if its fundamental group is trivial meaning that its genus $b_1$ is zero. Two Riemann surfaces are **conformally equivalent** or simply **equivalent** if they are equivalent as complex manifolds, that if a bijective morphism $f$ between them is a diffeomorphism which respects the complex structure (meaning that $f$ and $f^{-1}$ are analytic). The **curvature** is the Gaussian curvature of the surface. The **uniformization theorem** is:

**Theorem:** A Riemann surface is equivalent to one with constant curvature.

This is a "geometrization statement" and means that every Riemann surface is conformally equivalent either to the **Riemann sphere** (positive curvature), the **complex plane** (zero curvature) or the **unit disk** (negative curvature). It implies that any region $G \subset \mathbb{C}$ whose complement contains 2 or more points has a universal cover which is the disk which especially implies the **Riemann mapping theorem** assuring that and region $U$ homeomorphic to a disk is conformally equivalent to the unit disk. (see [59]). It also follows that all **Riemann surfaces** (without restriction of genus) can be obtained as quotients of these three spaces: for the sphere one does not have to take any quotient, the genus 1 surfaces = **elliptic curves** can be obtained as quotients of the complex plane and any genus $g > 1$ surface can be obtained as quotients of the unit disk. Since every closed 2-dimensional orientable surface is characterized by their genus $g$, the uniformization theorem implies that any such surface admits a metric of constant curvature. Teichmüller theory parametrizes the possible metrics, and there are $3g-3$ dimensional parameters for $g \geq 2$, whereas for $g = 0$ there is one and for $g = 1$ a moduli space $\mathbb{H}/SL_2(\mathbb{Z})$. In higher dimensions, close to the uniformization theorem comes the **Killing-Hopf theorem** telling that every connected complete Riemannian manifold of **constant sectional curvature** and dimension $n$ is isometric to the quotient of a sphere $\mathbb{S}^n$, Euclidean space $\mathbb{R}^n$ or Hyperbolic $n$-space $\mathbb{H}^n$. Constant curvature geometry is either Elliptic, Parabolic=Euclidean or Hyperbolic geometry. Complex analysis has rich applications in complex dynamics [27, 242, 59] and relates to much more geometry [240].

## 129. Control Theory

A **Kalman filter** is an optional estimates algorithm of a linear dynamic system from a series of possibly noisy measurements. The idea is similar as in a **dynamic Bayesian network** or **hidden Markov model**. It applies both to **differential equations** $\dot{x}(t) = Ax(t) + Bu(t) + Gz(t)$ as well as **discrete dynamical system** $x(t+1) = Ax(t) + Bu(t) + Gz(t)$, where $u(t)$ is **external input** and $z(t)$ **input noise** given by independent identically distributed usually Gaussian **random variables**. Kalman calls this a **Wiener problem**. One does not see the **state** $x(t)$ of the system but some **output** $y(t) = Cx(t) + Du(t)$. The filter then "filters out" or "learns" the best estimate $x^*(t)$ from the observed data $y(t)$. The linear space $X$ is defined as the vector space spanned by the already observed vectors. The optimal solution is given by a sophisticated dynamical data fitting.

**Theorem:** The optimal estimate $x^*$ is the projection of $y$ onto the $X$.

This is the informal 1-sentence description which can be found already in Kalman's article. Kalman then gives explicit formulas which generate from the **stochastic difference equation** a concrete **deterministic linear system**. For a modern exposition, see [232]. This is the **Kalman filter**. It is named after Rudolf Kalman who wrote [181] in 1960. Kalman's paper is one of the most cited papers in applied mathematics. The ideas were used in the Apollo and Space Shuttle program. Similar ideas have been introduced in statistics by the Danish astronomer Thorvald Thiele and the radar theoreticians Peter Swerling. There are also nonlinear version of the Kalman filter which is used in nonlinear state estimation like navigation systems and GPS. The nonlinear version uses a multi-variate Taylor series expansion to linearize about a working point. See [115, 232].

## 130. Zariski main theorem

A **variety** is called **normal** if it can be covered by open affine varieties whose rings of functions are normal. A commutative ring is called **normal** if it has no non-zero nilpotent elements and is integrally closed in its complete ring of fractions. For a curve, a one-dimensional variety, normality is equivalent to being non-singular but in higher dimensions, a normal variety still can have singularities. The normal complex variety is called **unibranch at a point** $x \in X$ if there are arbitrary small neighborhoods $U$ of $x$ such that the set of non-singular points of $U$ is connected. **Zariski's main theorem** can be stated as:

**Theorem:** Any closed point of a normal complex variety is unibranch.

Oscar Zariski proved the theorem in 1943. To cite [249], *"it was the final result in a foundational analysis of birational maps between varieties. The 'main Theorem' asserts in a strong sense that the normalization (the integral closure) of a variety $X$ is the maximal variety $X'$ birational over $X$, such that the fibres of the map $X' \to X$ are finite. A generalization of this fact became Alexandre Grothendieck's concept of the 'Stein factorization' of a map.* The result has been generalized to schemes $X$, which is called **unibranch** at a point $x$ if the local ring at $x$ is unibranch. A generalization is the **Zariski connectedness theorem** from 1957: if $f : X \to Y$ is a birational projective morphism between Noetherian integral schemes, then the inverse image of every normal point of $Y$ is connected. Put more colloquially, the fibres of a birational morphism from a projective variety $X$ to a normal variety $Y$ are connected. It implies that a birational morphism $f : X \to Y$ of algebraic varieties $X, Y$ is an open embedding into a neighbourhood of a normal point $y$ if $f^{-1}(y)$ is a finite set. Especially, a birational morphism between normal varieties which is bijective near points is an isomorphism. [158, 249]

## 131. Poincaré's last theorem

A **homeomorphism** $T$ of an annulus $X = \mathbb{T} \times [0,1]$ is **measure preserving** if it preserves the Lebesgue (area) measure and preserves the orientation of $X$. As a homeomorphism it induces also homeomorphisms on each of the two boundary circles. It is called **twist homeomorphism**, if it rotates the boundaries into different directions.

**Theorem:** A twist map on an annulus has at least two fixed points.

This is called the **Poincaré-Birkhoff theorem** or Poincaré's last theorem. It was stated by Henry Poincaré in 1912 in the context of the three body problem. Poincaré already gave an index argument for the existence of one fixed point gives a second. The existence of the first was proven by George Birkhoff in 1913 and in 1925, he added the precise argument for the existence of the second. The twist condition is necessary as the rotation of the annulus $(r, \theta) \to (r, \theta + 1)$ has no fixed point. Also area preservation is necessary as $(r, \theta) \to (r(2 - r), \theta + 2r - 1)$ shows. [36, 49]

## 132. Geometrization

A **closed manifold** $M$ is a smooth compact manifold without boundary. A closed manifold is **simply connected** if it is connected and the fundamental group is trivial meaning that every closed loop in $M$ can be pulled together to a point within $M$: (if $r : \mathbb{T} \to M$ is a parametrization of a closed path in $M$, then there exists a continuous map $R : \mathbb{T} \times [0,1] \to M$ such that $R(0,t) = r(t)$ and $R(1,t) = r(0)$. We say that **M is 3-sphere** if $M$ is homeomorphic to a 3-dimensional sphere $\{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1^2 + x_2^2 + x_3^2 + x_3^2 = 1\}$.

**Theorem:** A closed simply connected 3-manifold is a 3-sphere.

Henry Poincaré conjectured this in 1904. It remained the **Poincaré conjecture** until its proof of Grigori Perelman in 2006 [245]. In higher dimensions the statement was known as the **generalized Poincaré conjecture**, the case $n > 4$ had been proven by Stephen Smale in 1961 and the case $n = 4$ by Michael Freedman in 1982. A $d$**-homotopy sphere** is a closed d-manifold that is homotopic to a $d$-sphere. (A manifold $M$ is **homotopic** to a manifold $N$ if there exists a continuous map $f : M \to N$ and a continuous map $g : N \to M$ such that the composition $g \circ f : M \to M$ is homotopic to the identity map on $M$ (meaning that there exists a continuous map $F : M \times [0,1] \to M$ such that $F(x,0) = g(f(x))$ and $F(x,1) = x$) and the map $f \circ g : N \to N$ is homotopic to the identity on $N$.) The Poincaré conjecture itself, the case $d = 3$, was proven using a theory built by **Richard Hamilton** who suggested to use the **Ricci flow** to solve the conjecture and more generally the **geometrization conjecture** of William Thurston: every closed 3-manifold can be decomposed into **prime manifolds** which are of 8 types, the so called **Thurston geometries** $S^3, E^3, H^3, S^2 \times R, H^2 \times R, \tilde{SL}(2, R), \text{Nil}, \text{Solv}$. If the statement **M is a sphere** is replaced by $M$ **is diffeomorphic to a sphere**, one has the **smooth Poincaré conjecture**. Perelman's proof verifies this also in dimension $d = 3$. The smooth Poincaré conjecture is false in dimension $d \geq 7$ as $d$-spheres then admit non-standard smooth structures, so called exotic spheres constructed first by John Milnor. For $d = 5$ it is true following result of Dennis Barden from 1964. It is also true for $d = 6$. For $d = 4$, it is open, and called "the last man standing among all great problems of classical geometric topology" [228]. See [246] for details on Perelman's proof.

## 133. Steinitz theorem

A non-empty finite simple connected graph $G$ is called **planar** if it can be embedded in the plane $\mathbb{R}^2$ without self crossings. The abstract edges of the graph are then realized as actual curves in the plane connecting two vertices. The embedding of $G$ in the plane subdivides the plane into a finite collection $F$ of simply connected regions called **faces**. Let $v = |V|$ is the number of vertices, $e = |E|$ the number of edges and $f = |F|$ is the number of faces. A planar graph is called **polyhedral** if it can be realized as a **convex polyhedron**, a convex hull of finitely many points in $\mathbb{R}^3$. A graph is called 3**-connected**, if it remains connected also after removing one or two of its vertices. A connected, planar 3-connected graph is also called a 3-polyhedral graph. The **Polyhedral formula of Euler** combined with **Steinitz's theorem** means:

**Theorem:** $G$ planar $\Rightarrow v - e + f = 2$. Planar 3-connected $\Leftrightarrow$ polyhedral.

The Euler polyhedron formula which was first seen in examples by René Descartes [3] but seen by Euler in 1750 to work for general planar graphs. Euler already gave an induction proof (also in 1752) but the first complete proof appears having been given first by Legendre in 1794. The Steinitz theorem was proven by Ernst Steinitz in 1922, evenso he obtained the result already in 1916. In general, a planar graph always defines a finite generalized CW complex in which the faces are the 2-cells, the edges are the 1-cells and the vertices are the 0-cells. The embedding in the plane defines then a **geometric realization** of this combinatorial structure as a topological 2-sphere. But the realization is not necessarily be achievable in the form of a convex polyhedron. Take a **tree graph** for example, a connected graph without triangles and without closed loops. It is planar but it is not even 2-connected. The number of vertices $v$ and the number of edges $e$ satisfy $v - e = 1$. After embedding the tree in the plane, we have one face so that $f = 1$. The Euler polyhedron formula $v - e + f = 2$ is verified but the graph is not polyhedral. Even in the extreme case where $G$ is a one-point graph, the Euler formula holds: in that case there is $v = 1$ vertex, $e = 0$ edges and $f = 1$ faces (the complement of the point in the plane) and still $v - e + f = 2$. The 3-connectedness assures that the realization can be done using convex polyhedra. It is even possible to have force the vertices of the polyhedron to be on the integer lattice. See [146, 350]. In [146], it is stated that the Steinitz theorem is "the most important and deepest known result for 3-polytopes".

## 134. Hilbert-Einstein action

Let $(M, g)$ be a smooth 4-dmensional **Lorentzian manifold** which is symptotically flat. A Lorentzian manifold is a 4 dimensional pseudo Riemannian manifold with signature $(1, 3)$ which in the flat case is $dx^2 + dy^2 + dz^2 - dt^2$. We assume that the volume form $d\mu$ has the property that the **scalar curvature** $R$ is in $L^1(M, d\mu)$. One can now look at the variational problem to find extrema of the functional $g \to \int_M R d\mu$. More generally, one can add a **Lagrangian** $L$ one uses the Hilbert-Einstein functional $\int_M R/\kappa + L d\mu$, where $\kappa = 8\pi G/c^4$ is the **Einstein constant**. Let $R_{ij}$ be the Ricci tensor, a symmetric tensor, and $T_{ij}$ the energy-momentum tensor. The **Einstein field equations** are

**Theorem:** $G_{ij} = R_{ij} - g_{ij}R/2 = \kappa T_{ij}$

These are the Euler-Lagrange equation of an infinite dimensional extremization problem. The variational problem was proposed by David Hilbert in 1915. Einstein published in the same year the **general theory of relativity**. In the case of a **vacuum**: $T = 0$, solutions $g$ define **Einstein manifolds** $(M, g)$. An example of a solution to the vaccuum Einstein equations different from the flat space solution is the **Schwarzschild solution**, which was found also in 1915 and published in 1916. It is the metric given in spherical coordinates as $-(1 - r/\rho)c^2dt^2 + (1-r/\rho)^{-1}d\rho^2 + \rho^2 d\phi^2 + \rho^2 \sin^2\phi d\theta^2$, where $r$ is the **Schwarzschild radius**, $\rho$ the distance to the singularity, $\theta, \phi$ are the standard **Euler angles** (**longitude** and **colatitude**) in calculus. The metric solves the Einstein equations for $\rho > r$. The flat metric $-c^2dt^2 + d\rho^2 + \rho^2 d\theta^2 + \rho^2 \sin^2\theta d\phi^2$ describes the vacuum and the Schwarzschild solution describes the gravitational field near a **massive body**. Intuitively, the metric tensor $g$ is determined by $g(v, v)$, and the Ricci tensor by $R(v, v)$ which is 3 times the average sectional curvature over all planes passing through a plane through $v$. The scalar curvature is 6 times the average over all sectional curvatures passing through a point. See [86, 69].

## 135. Hall stable marriage

Let $X$ be a finite set and $\mathcal{A}$ be a family of finite subset $A$ of $X$. A **transversal** of $\mathcal{A}$ is an injective function $f : \mathcal{A} \to X$ such that $f(A) \in A$ for all $A \in \mathcal{A}$. The set $\mathcal{A}$ satisfies the **marriage condition** if for every finite subset $\mathcal{B}$ of $\mathcal{A}$, one has $|\mathcal{B}| \leq |\bigcup_{A \in \mathcal{B}} A|$. The **Hall marriage theorem** is

**Theorem:** $\mathcal{A}$ has a transversal $\Leftrightarrow$ $\mathcal{A}$ satisfies marriage condition.

The theorem was proven by Philip Hall in 1935. It implies for example that if a deck of cards with 52 cards is partitioned into 13 equal sized piles, one can chose from each deck a card so that the 13 cards have exactly one card of each rank. The theorem can be deduced from a result in graph geometry: if $G = (V, E) = (X, \emptyset) + (Y, \emptyset)$ is a bipartite graph, then a **matching** in $G$ is a collection of edges which pairwise have no common vertex. For a subset $W$ of $V$, let $S(W)$ denote the set of all vertices adjacent to some element in $W$. The theorem assures that there is an **X-saturating matching** (a matching covering $X$) if and only if $|W| \leq |S(W)|$ for every $W \subset X$. The reason for the name "marriage" is the situation that $X$ is a set of men and $Y$ a set of women and that all men are eager to marry. Let $A_i$ be the set of women which could make a spouse for the $i$'th man, then marrying everybody off is an $X$-saturating matching. The condition is that any set of $k$ men has a combined list of at least $k$ women who would make suitable spouses. See [50].

## Epilogue: Value

Which mathematical theorems are the most important ones? This is a complicated variational problem because it is a general and fundamental problem in economics to define **"value"**. The difficulty with the concept is that "value" is often a matter of taste or fashion or social influence and so an **equilibrium of a complex social system**. Value can change rapidly, sometimes triggered by small things. The reason is that the notion of value like in game theory depends on how it is valued by others. A fundamental principle of **catastrophe theory** is that maxima of a functional can depend discontinuously on parameter. As value is often a social concept, this can be especially brutal or lead to unexpected **viral effects**. The value of a company depends on what "investors think" or what analysts see for potential gain in the future. Social media try to measure value using "likes" or "number of followers". A majority vote is a measure but how well can it predict correctly what be valuable in the future? Majority votes taken over longer times would give a more reliable value functional. Assume one could persuade every mathematician to give a list of the two dozen most fundamental theorems and do that every couple of years, and reflect the "wisdom of an educated crowd", one could get a good value functional. Ranking theorems and results in mathematics are a mathematical optimization problem itself. One could use techniques known in the "search industry". One idea is to look at the finite graph in which the theorems are the nodes and where two theorems are related with each other if one can be deduced from the other (or alternatively connect them if one influences the other strongly). One can then run a **page rank algorithm** [220] to see which ones are important. Running this in each of the major mathematical fields could give an algorithm to determine which theorems deserve the name to be "fundamental".

## Opinions

Teaching a course called "Math from a historical perspective" at the Harvard extension school led me to write up the present document. This course Math E 320 (now planned to be taught for the 10th time) is a rather pedestrian but pretty comprehensive stroll through all of mathematics. At the end of the course, students are asked as part of a project to write some short stories about theorems or mathematical fields or mathematical persons. The present document benefits already from these writings and also serves a bit as preparation for the course. It is interesting to see what others consider important. Sometimes, seeing what others feel can change your own view. I was definitely influenced by students, teachers, colleagues and literature as well of course by the limitations of my own understanding. And my point of view has already changed while writing the actual theorems down. Value is more like an equilibrium of many different factors. In mathematics, values have changed rapidly over time. And mathematics can describe the rate of change of value [271]. Major changes in appreciation for mathematical topics came definitely at the time of Euclid, then at the time when calculus was developed by Newton and Leibniz. Also the development of more abstract algebraic constructs or topological notions, the start of set theory changed things considerably. In more modern time, the categorization of mathematics and the development of rather general and abstract new objects, (for example with new approaches taken by Grothendieck) changed the landscape. In most of the new development, I remain the puzzled tourist wondering how large the world of mathematics is. It has become so large that continents have emerged: we have **applied mathematics**, **mathematical physics**, **statistics**, **computer science** and **economics** which have diverged to independent subjects and departments. Classical mathematicians like Euler would now be called applied mathematicians, de Moivre would maybe be a statistician, **Newton** a mathematical physicist and **Turing** a computer scientist and **von Neuman** an economist.

## Search

A couple of months ago, when looking for "George Green", the first hit in a search engine would be a 22 year old soccer player. (This was not a search bubble thing [263] as it was tested with cleared browser cache and via anonymous VPN from other locations, where the search engine can not determine the identity of the user). Now, I love soccer, played it myself a lot as a kid and also like to watch it on screen, but is the English soccer player George William Athelston Green really more "relevant" than the British mathematician George Green, who made fundamental break through discoveries which are used in mathematics and physics? Shortly after I had tweeted about this strange ranking on December 27, 2017, the page rank algorithm must have been adapted, because already on January 4th, 2018, the Mathematician George Green appeared first (again not a search bubble phenomenon, where the search engine adapts to the users taste and adjusts the search to their preferences). It is not impossible that my tweet has reached, meandering through social media, some search engine engineer who was able to rectify the injustice done to the miller and mathematician George Green. The theory of networks shows "small world phenomena" [334, 25, 333] can explain that such influences or synchronizations are not that impossible [314]. But coincidences can also be deceiving. Humans just tend to observe coincidences even so there might be a perfectly mathematical explanation prototyped by the **birthday paradox**. See [236]. But one must also understand that search needs to serve the majority. For a general public, a particular subject like mathematics is not that important. When searching for "Hardy" for example, it is not Godfrey Hardy who is mentioned first as a person belonging to that keyword but Tom Hardy, an English actor. This obviously serves most of the searches better. As this might infuriate particular groups (here mathematicians), search engines have started to adapt the searches to the user, giving the search some **context** which is an important ingredient in artificial intelligence. The problem is the search bubble phenomenon which runs hard against objectivity. Textbooks of the future might adapt its language, difficulty and even citation or the history on who reads it. Novels might adapt the language to the age of the user, the country where the user lives, and the ending might depends on personal preferences or even the medical history of the user. Many computer games are already customizable as such. A person flagged as sensitive or a young child might be served a happy ending rather than ending the novel in an ambivalent limbo or even a disaster. [263] explains the difficulty. The issues has become manifested even more since that book came out as it might even influence elections.

## Beauty

In order to determine what is a "fundamental theorem", also aesthetic values matter. But the question of "what is beautiful" is even trickier but many have tried to define and investigate the mechanisms of beauty: [154, 338, 339, 280, 299, 6, 244]. In the context of mathematical formulas, the question has been investigated in **neuro-aesthetics**. Psychologists, in collaboration with mathematicians have measured the brain activity of 16 mathematicians with the goal to determine what they consider beautiful [286]. The **Euler identity** $e^{i\pi} + 1 = 0$ was rated high with a value 0.8667 while a formula for $1/\pi$ due to Ramanujan was rated low with an average rating of -9.7333. Obviously, what mattered was not only the complexity of the formula but also how much **insight** the participants got when looking at the equation. The authors of that paper cite Plato who said **it "nothing without understanding would ever be more beauteous than with understanding"**. Obviously, the formula of Ramanujan is much deeper but it requires some background knowledge for being appreciated. But the authors acknowledge in the discussion that that correlating "beauty and understanding" can be tricky. Rota [280] notes that the appreciation of mathematical beauty in some statement requires the ability to understand it. And [244] notices that "even professional mathematicians specialized in a certain field might find results or proofs in other fields obscure" but that this is not much different from say music, where "knowledge bout technical details such as the differences between things like cadences, progressions or chords changes the way we appreciate music" and that "the symmetry of a fugue or a sonata are simply invisible without a certain technical knowledge". As history has shown, there were also always "artistic connections" [130, 52] as well as "religious influences" [227, 300]. The book [130] cites Einstein who defines "mathematics as the poetry of logical ideas". It also provides many examples and illustrations and quotations. And there are various opinions like Rota who argues that beauty is a rather objective property which depends on historic-social contexts. And then there is taste: what is more appealing, the element of **surprise** like the Birthday paradox or Petersburg paradox in probability theory Banach-Tarski paradox in measure theory or the element of **enlightenment** and **understanding**, which is obviously absent if one hears the first time that one can disassemble a sphere into 5 pieces, rotate and translate them to build two spheres or that the infinite sum $1+2+3+4+5+\dots$ is naturally equal to $-1/12$ as it is $\zeta(-1)$ (defined by analytic continuation). The role of aesthetic in mathematics is especially important in education, where mathematical models [125] or 3D printing [290, 206] can help to make mathematics more approachable.

## The fate of fame

Aesthetics is a fragile subject however. If something beautiful has become too popular and so entered **pop-culture**, a natural aversion against it might develop. It is in danger to become a clishé or even become **kitsch** (which is a word used to tear down popular stuff). The Mandelbrot set for example is just marvelous, but it does not excite anymore because it is so commonly known. The Monty-Hall problem which became famous by Gardner columns in the nineties (see [301, 279]) was cool to teach in 1994 three years after the infamous "parade column" of 1991 by Marilyn vos Savant. Especially after a cameo in the movie "21", the theorem has become part of **mathematical kitsch**. I myself love mathematical kitsch. A topic gaining that status must have been nice and innovative to obtain that label. Kitsch becomes only tiresome if it is not presented in a new and original form. The book [264] for example, in the context of complex dynamics, remains a master piece still today, even-so the picture have become only too familiar. In that context, it appears strange that mathematicians do not jump on the "mandelbulb set" $M$ which is one of the most beautiful mathematical objects there is but the reason is maybe just that it is a "youtube star" and so not worthy yet. (More likely is that the object is just too difficult to study as we lack the mathematical analytic tools which for example would just to answer the basic question whether $M$ is connected.) A second example is **catastrophe theory** [271, 327] a beautiful part of **singularity theory** which started with Hassler Whitney and was developed by René Thom [323], which was hyped to much that it fell into a fall from which it has not yet fully recovered. And this despite the fact that Thom pointed out the limits as well as the controversies of the theory already. It had to pay a prize for its fame and appears to be forgotten. Chaos theory from the 60ies which started to peak with Edward Lorenz, Mandelbrot and strange attractors etc started to become a **clishé** latest after that infamous scene featuring the character Ian Malcolm in the 1993 movie Jurassic park. It was already laughed at within the same movie franchise, in the third Jurassic Park installment of 2001, where the kid Erik Kirby snuffs on Malcolm's preachiness and quotes his statement "everything is chaos" in a despective way. In art, architecture, music, fashion or design also, if something has become too popular, it is despised by the "connaisseurs". Hardly anybody would consider a "lava lamp" (invented in 1963) a object of taste nowadays, even so, the fluid dynamics and motion is objectively rich and interesting. The piano piece "Für Elise" by Ludwig van Beethoven became so popular that it can not even be played any more as background music in a supermarket. There is something which prevents a "music connaisseur" to admit that the piece is great. Such examples suggest that it might be better for an achievement (or theorem in mathematics) not to enter pop-culture. The lack of "deepness" is despised by the elite. The principle of having fame torn down to disgrace is common also outside of mathematics. Famous actors, entrepreneurs or politicians are not all admired but also hated to the guts, or torn to pieces and certainly can not live normal lives any more. The phenomenon of accumulated critique got amplified with mob type phenomena in social media. There must be something fulfilling to trash achievements. Film critics are often harsh and judge negatively because this elevates their own status as they appear to have a "high standard". Similarly morale judgement is expressed often just to elevate the status of the judge even so experience has shown that often judges are offenders themselves. Maybe it is also human Schadenfreude, or greed which makes so many to voice critique. History shows however that social value systems do not matter much in the long term. A good and rich theory will show its value if it is appreciated also in hundreds of years, where fashion and social influence will have less impact. The theorem of Pythagoras will be important independent of fame and even as a cliché, it is too important to be labeled as such. It has not only earned the status of kitsch, it is also a prototype as well as a useful tool.

## Media

There is no question that the **Pythagoras theorem**, the **Euler polyhedron formula** $\chi = v - e + f$ the **Euler identity** $e^{i\pi} + 1 = 0$, or the **Basel problem formula** $1 + 1/4 + 1/9 + 1/16 + \cdots = \pi/6$ will always rank highly in any list of beautiful formulas. Most mathematicians agree that they are elegant and beautiful. These results will also in the future keep top spots in any ranking. On social networks, one can find lists of favorite formulas. On "Quora", one can find the arithmetic mean-geometric mean inequality $\sqrt{ab} \leq (a+b)/2$ or the **geometric summation formula** $1 + a + a^2 + \cdots = 1/(1-a)$ high up. One can also find strange contributions in social media like the identity $1 = 0.99999\dots$ (which is used by Piaget inspired educators to probe mathematical maturity of kids. Similarly as in Piaget's experiments, there is time of mathematical maturity where a student starts to understand that this is an identity. A very young student thinks 1 is larger than 0.9999... even if told to point out a number in between). At the moment, searching for the "most beautiful formula in mathematics" gives the Euler identity and search engines agree. But the concept of taste in a time of social media can be confusing. We live in a time, when a 17 year old "social influencer" can in a few days gather more "followers" and become more widely known than **Sophie Kovalewskaya** who made fundamental beautiful and lasting contributions in mathematics and physics like the Cauchy-Kovalevskaya theorem mentioned above. This theorem is definitely more lasting than a few "selfie shots" of a pretty face, but measured by a "majority vote", it would not only lose, but disappear. But time rectifies this. Kovalewskaya will also be ranked highly in 50 years, while the pretty face has faded. Hardy put this even more extreme by comparing a mathematician with a literary heavy weight: *Archimedes will be remembered when Aeschylus is forgotten, because languages die and mathematical ideas do not.* [154] There is no doubt that film and TV (and now internet like "Youtube" and blogs) has a great short-term influence on value or exposure of a mathematical field. Examples of movies with influence are **Good will hunting** (1997) which featured some graph theory and Fourier theory, **21** from (2008) which has a scene in which the Monty Hall problem appears, **The man who knew infinity** featuring the work of Ramanujan and promotes some combinatorics. There are lots of movies featuring cryptology like **Sneakers** (1992), **Breaking the code** (1996), **Enigma** (2001) or **The imitation game** (2014). For TV, mathematics was promoted nicely in **Numb3rs** (2005-2010). For more, see [269] or my own online math in movies collection.

## Professional opinions

Interviews with professional mathematicians can also probe the waters. In [210], Natasha Kondratieva had asked a number of mathematicians: "What three mathematical formulas are the most beautiful to you". The **formulas of Euler** or the **Pythagoras theorem** naturally were ranked high. Interestingly, Michael Atiyah included even a formula **"Beauty = Simplicity + Depth"**. Also other results, like the **Leibniz series** $\pi/4 = 1-1/3+1/5-1/7+1/9-\dots$, the **Maxwell equations** $dF = 0, d^*F = J$ or the **Schrödinger equation** $i\hbar u' = (i\hbar\nabla + eA)^2 u + Vu$, the **Einstein formula** $E = mc^2$ or the **Euler's golden key** $\sum_{n=1}^\infty 1/n^s = \prod_p (1-1/p^s)^{-1}$ or the **Gauss identity** $\int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}$ or the **volume of the unit ball** in $R^{2n}$ given as $\pi^n/n!$ appeared. Gregory Margulis mentioned an application of the **Poisson summation formula** $\sum_n f(n) = \sum_n \hat{f}(n)$ which is $\sqrt{2}\sum_n e^{-n^2} = \sum_n e^{-n^2/4}$ or the **quadratic reciprocity law** $(p|q) = (-1)^{(p-1)/2(q-1)/2}$, where $(p|q) = 1$ if $q$ is a **quadratic residue** modulo $p$ and $-1$ else. Robert Minlos gave the **Gibbs formula**, a **Feynman-Kac formula** or the **Stirling formula**. Yakov Sinai mentioned the **Gelfand-Naimark realization** of an Abelian $C^*$ algebra as an algebra of continuous function or the **second law of thermodynamics**. Anatoly Vershik gave the generating function $\prod_{k=0}^\infty (1+x^k) = \sum_{n=0}^\infty p(n)x^n$ for the **partition function** and the **generalized Cauchy inequality** between arithmetic and geometric mean. An interesting statement of David Ruelle appears in that article who quoted Grothendieck by " my life's ambition as a mathematician, or rather my joy and passion, have constantly been to discover obvious things ...". Combining Grothendieck's and Atiyah's quote, fundamental theorems should be "obvious, beautiful, simple and still deep".

A recent column "Roots of unity" in the Scientific American asks mathematicians for their favorite theorem: examples are **Noether's theorem**, **the uniformization theorem**, the **Ham Sandwich theorem**, the **fundamental theorem of calculus**, **the circumference of the circle**, the **classification of compact 2-surfaces**, **Fermat's little theorem**, the **Gromov non-squeezing theorem**, a theorem about Betti numbers, the **Pythagorean theorem**, the **classification of Platonic solids**, the **Birkhoff ergodic theorem**, the **Burnside lemma**, the **Gauss-Bonnet theorem**, Conways rational tangle theorem, **Varignon's theorem**, an upper bound on **Reidemeister moves in knot theory**, the **asymptotic number of relative prime pairs**, the **Mittag Leffler theorem**, a theorem about **spectral sparsifiers**, the **Yoneda lemma** and the **Brouwer fixed point theorem**. These interviews illustrate also that the choices are different if asked for "personal favorite theorem" or "objectively favorite theorem".

## Fundamental versus important

Asking for fundamental theorems is different than asking for "deep theorems" or "important theorems". Examples of deep theorems are the **Atiyah-Singer** or **Atiyah-Bott theorems** in differential topology, the **KAM theorem** related to the strong implicit function theorem, or the **Nash embedding theorem** in Riemannian geometry. An other example is the **Gauss-Bonnet-Chern theorem** in Riemannian geometry or the **Pesin theorem** in partially hyperbolic dynamical systems. Maybe the **shadowing lemma** in hyperbolic dynamics is more fundamental than the much deeper **Pesin theorem** (which is still too complex to be proven with full details in any classroom). One can also argue, whether the **"theorema egregium"** of Gauss, stating that the curvature of a surface is intrinsic and not dependent on an embedding is more "fundamental" than the **"Gauss-Bonnet"** result, which is definitely deeper. In number theory, one can argue that the **quadratic reciprocity formula** is deeper than the **little Theorem of Fermat** or the **Wilson theorem**. (The later gives an if and only criterion for primality but still is far less important than the little theorem of Fermat which as the later is used in many applications.) The **last theorem of Fermat** is an example of an important theorem as it is deep and related to other fields and culture, but it is not yet so much a "fundamental theorem". Similarly, the **Perelman theorem** fixing the **Poincaré conjecture** is important, but it is not (yet) a fundamental theorem. It is still a mountain peak and not a sediment in a rock. Important theorems are not much used by other theorems as they are located at the end of a development. Also the solution to the **Kepler problem** on sphere packings or the proof of the **4-color theorem** [65] or the proof of the **Feigenbaum conjectures** [89, 169] are important results but not so much used by other results. Important theorems build **the roof** of the building, while fundamental theorems form the **foundation** on which a building can be constructed. But this can depend on time as what is the roof today, might be in the foundation later on, once more floors have been added.

## Open problems

The importance of a result is also related to related to **open problems** attached to the theorem. Open problems fuel new research and new concepts. Of course this is a moving target but any "value functional" for "fundamental theorems" is time dependent and a bit also a matter of **fashion**, **entertainment** (TV series like "Numbers" or Hollywood movies like "good will hunting" changed the value) and under the influence of **big shot mathematicians** which serve as "influencers". Some of the problems have **prizes** attached like the **23 problems of Hilbert**, the **15 problems of Simon** [294], the **18 problems of Smale**, or the **10 Millenium problems**. There are beautiful open problems in any major field and building a ranking would be as difficult as ranking theorems. There appears to be wide consensus that the **Riemann hypothesis** is the most important open problem in mathematics. It states that the roots of the Riemann zeta function are all located on the axes $\text{Re}(z) = 1/2$. In number theory, the **prime twin problem** or the **Goldbach problem** have a high exposure as they can be explained to a general audience without mathematics background. For some reason, an equally simple problem, the **Landau problem** asking whether there are infinitely many primes of the form $n^2 + 1$ is less well known. In recent years, due to an alleged proof by Shinichi Mochizuki of the ABC conjecture using a new theory called **Inner-Universal Teichmuller Theory** (IUT) which so far is not accepted by the main mathematical community despite strong efforts. But it has put the ABC conjecture from 1985 in the spot light like [344]. It has been described in [134] as the most important problem in Diophantine equations. It can be expressed using the **quality** $Q(a,b,c)$ of three integers $a,b,c$ which is $Q(a,b,c) = \log(c)/\log(\text{rad}(abc))$, where the **radical** $\text{rad}(n)$ of a number $n$ is the product of the distinct prime factors of $n$. The ABC conjecture is that for any real number $q > 1$ there exist only finitely many triples $(a,b,c)$ of positive relatively prime integers with $a + b = c$ for which $Q(a,b,c) > q$. The triple with the highest quality so far is $(a,b,c) = (2, 3^{10}109, 23^5)$; its quality is $Q = 1.6299$. And then there are entire collections of conjectures, one being the **Langlands program** which relates different parts of mathematics like number theory, algebra, representation theory or algebraic geometry. I myself can not appreciate this program because I would need first to understand it. My personal favorite problem is the **entropy problem** in smooth dynamical systems theory [183]. The Kolmogorov-Sinai entropy of a smooth dynamical system can be described using Lyapunov exponents. For many systems like smooth convex billiards, one measures positive entropy but can not prove it. An example is the table $x^4 + y^4 = 1$ [176]. For ergodic theory see [83, 90, 128, 298]

## Classification results

One can also see classification theorems like the above mentioned Gelfand-Naimark realization as mountain peaks in the landscape of mathematics. Examples of **classification results** are the classification of regular or semi-regular polytopes, the classification of discrete subgroups of a Lie group, the classification of "Lie algebras", the classification of "von Neumann algebras", the "classification of finite simple groups", the **classification of Abelian groups**, or the classification of associative **division algebras** which by Frobenius is given either by the real or complex or quaternion numbers. Not only in algebra, also in differential topology, one would like to have classifications like the classification of $d$-dimensional manifolds. In topology, an example result is that every Polish space is homeomorphic to some subspace of the **Hilbert cube**. Related to physics is the question what "functionals" are important. Uniqueness results help to render a functional important and fundamental. The classification of **valuations** of fields is classified by **Ostrowski's theorem** classifying valuations over the rational numbers either being the absolute value or the $p$-adic norm. The **Euler characteristic** for example can be characterized as the unique **valuation** on simplicial complexes which assumes the value 1 on simplices or functional which is invariant under Barycentric refinements. A theorem of Claude Shannon [292] identifies the **Shannon entropy** is the unique functional on probability spaces being compatible with additive and multiplicative operations on probability spaces and satisfying some normalization condition.

## Bounds and inequalities

An other class of important theorems are **best bounds** like the **Hurwitz estimate** stating that there are infinitely many $p/q$ for which $|x - p/q| < 1/(\sqrt{5}q^2)$. In packing problems, one wants to find the best packing density, like for **sphere packing problems**. In complex analysis, one has the **maximum principle**, which assures that a harmonic function $f$ can not have a local maximum in its domain of definition. One can argue for including this as a fundamental theorem as it is used by other theorems like the **Schwarz lemma** or calculus of variations. In probability theory or statistical mechanics, one often has thresholds, where some **phase transition** appears. Computing these values is often important. The concept of **maximizing entropy** explains many things like why the Gaussian distribution is fundamental as it maximizes entropy. Measures maximizing entropy are often special and often **equilibrium measures**. This is a central topic in statistical mechanics [282, 283]. In combinatorial topology, the **upper bound theorem** was a milestone. It was long a conjecture of Peter McMullen and then proven by Richard Stanley that **cyclic polytopes** maximize the volume in the class of polytopes with a given number of vertices. Fundamental area also some **inequalities** [131] like the **Cauchy-Schwarz inequality** $|a \cdot b| \leq |a||b|$, the **Chebyshev inequality** $\mathrm{P}[|X - [\mathrm{E}[X]| \geq |a|] \leq \mathrm{Var}[X]/a^2$. In complex analysis, the **Hadamard three circle theorem** is important as gives bounds between the maximum of $|f|$ for a holomorphic function $f$ defined on an annulus given by two concentric circles. Often inequalities are more fundamental and powerful than equalities because they are more widely used. Related to inequalities are **embedding theorems** like **Sobolev embedding theorems**. For more inequalities, see [53]. Apropos embedding, there are the important Whitney or Nash embedding theorems which are appealing.

## Big ideas

Classifying and valuing **big ideas** is even more difficult than ranking individual theorems. Examples of big ideas are the idea of **axiomatisation** which stated with planar geometry and number theory as described by Euclid and the concept of **proof** or later the concept of **models**. Archimedes idea of **comparison**, leading to ideas like the **Cavalieri principle**, integral geometry or measure theory. René Descartes idea of **coordinates** which allowed to work on geometry using algebraic tools, the use of **infinitesimals and limits** leading to calculus, allowing to merge concepts of rate of change and accumulation, the idea of **extrema** leading to the calculus of variations or Lagrangian and Hamiltonian dynamics or descriptions of fundamental forces. **Cantor's set theory** allowed for a universal simple language to cover all of mathematics, the **Klein Erlangen program** of “classifying and characterizing geometries through symmetry”. The abstract idea of a group or more general mathematical structures like monoids. The concept of extending **number systems** like completing the real numbers or extending it to the **quaternions** and **octonions** or then producing **p-adic number** or **hyperreal numbers**. The concept of **complex numbers** or more generally the idea of **completion** of a field. The idea of **logarithms** [306]. The idea of **Galois** to relate problems about solving equations with **field extensions** and **symmetries**. The **Grothendieck program** of “geometry without points” or “locales” as topologies without points in order to overcome shortcomings of set theory. This lead to new objects like **schemes** or **topoi**. An other basic big idea is the concept of **duality**, which appears in many places like in projective geometry, in polyhedra, **Poincaré duality** or **Pontryagin duality** or **Langlands duality** for reductive algebraic groups. The idea of **dimension** to measure topological spaces numerically leading to **fractal geometry**. The idea of **almost periodicity** is an important generalization of periodicity. Crossing the boundary of integrability leads to the important paradigm of stability and randomness [247] and the interplay of structure and randomness [320]. These themes are related to **harmonic analysis** and **integrability** as integrability means that for every invariant measure one has almost periodicity. It is also related to spectral properties in solid state physics or via **Koopman theory** in ergodic theory or then to fundamental new number systems like the **p-adic numbers**: the **p-adic integers** form a compact topological group on which the translation is almost periodic. It also leads to problems in **Diophantine approximation**. The concept of **algorithm** and building the foundation of computation using precise mathematical notions. The use of algebra to track problems in topology starting with Betti and Poincaré. An other important principle is to reduce a problem to a **fixed point problem**. The **categorical approach** is not only a unifying language but also allows for generalizations of concepts allowing to solve problems. Examples are generalizations of Lie groups in the form of **group schemes**. Then there is the **deformation idea** which was used for example in the Perelman proof of the **Poincaré conjecture**. Deformation often comes in the form of **partial differential equations** and in particular heat type equations. Deformations can be abstract in the form of **homotopies** or more concrete by analyzing concrete partial differential equations like the mean curvature flow or **Ricci flow**. An other important line of ideas is to use **probability theory** to prove results, even in combinatorics. A probabilistic argument can often give existence of objects which one can not even construct. Examples are graphs with $n$ nodes for which the Euler characteristic of the defining **Whitney complex** is exponentially large in $n$. The idea of **non-commutative geometry** generalizing geometry through functional analysis or the idea of **discretization** which leads to numerical methods or computational geometry. The power of coordinates allows to solve geometric problems more easily. The above mentioned examples have all proven their use. Grothendieck's ideas lead to the solution of the **Weyl conjectures**, fixed point theorems were used in Game theory, to prove uniqueness of solutions of differential equations or justify perturbation theory like the **KAM theorem** about the persistence of quasi-periodic motion leading to **hard implicit function theorems**. In the end, what really counts is whether the big idea can solve problems or to prove theorems. The history of mathematics clearly shows that abstraction for the sake of abstraction or for the sake of generalization rarely could convince the mathematical community. At least not initially. But it can also happen that the break through of a new theory or generalization only pays off much later. A big idea might have to age like a good wine.

## Taxonomies

When looking at mathematics overall, **taxonomies** are important. They not only help to navigate the landscape and are also interesting from a pedagogical as well as historical point of view. I borrow here some material from my course Math E 320 which is so global that a taxonomy is helpful. Organizing a field using markers is also important when teaching intelligent machines, a field which be seen as the **pedagogy for AI**. The big bulk of work in [204] was to teach a bot mathematics, which means to fill in thousands of entries of knowledge. It can appear a bit mind numbing as it is a similar task than writing a dictionary. But writing things down for a machine actually is even tougher than writing things down for a student. We can not assume the machine to know anything it is not told. This document by the way could relatively easily be adapted into a database of "important theorems" and actually one my aims is it to feed it eventually to the Sofia bot. If the machine is asked about "important theorem in mathematics", it would be surprisingly well informed, even so it is just stupid encyclopedic data entry. Historically, when knowledge was still sparse, one has classified teaching material using the **liberal arts of sciences**, the **trivium**: grammar, logic and rhetoric, as well as the **quadrivium**: arithmetic, geometry, music, and astronomy. More specifically, one has built the eight ancient roots of mathematics which are tied to activities: counting and sorting (arithmetic), spacing and distancing (geometry), positioning and locating (topology), surveying and angulating (trigonometry), balancing and weighing (statics), moving and hitting (dynamics), guessing and judging (probability) and collecting and ordering (algorithms). This led then to the 12 topics taught in that course: Arithmetic, Geometry, Number Theory, Algebra, Calculus, Set theory, Probability, Topology, Analysis, Numerics, Dynamics and Algorithms. The **AMS classification** is much more refined and distinguishes 64 fields. The Bourbaki point of view is given in [93]: it partitions mathematics into algebraic and differential topology, differential geometry, ordinary differential equations, Ergodic theory, partial differential equations, non-commutative harmonic analysis, automorphic forms, analytic geometry, algebraic geometry, number theory, homological algebra, Lie groups, abstract goups, commutative harmonic analysis, logic, probability theory, categories and sheaves, commutative algebra and spectral theory. What are **hot spots in mathematics**? Michael Atiyah [20] distinguished parameters like **local - global**, **low and high dimensional**, **commutative - non-commutative**, **linear - nonlinear**, **geometry - algebra**, **physics and mathematics**.

## Key examples

The concept of **experiment** came even earlier and has always been part of mathematics. Experiments allow to get good examples and set the stage for a theorem. Obviously the theorem can not contradict any of the examples. But examples are more than just a tool to falsify statements; a good example can be the **seed** for a new theory or for an entire subject. Here are a few examples: in **smooth dynamical systems** the **Smale horse shoe** comes to mind, in **differential topology** the **exotic spheres** of Milnor, in one-dimensional dynamics the **logistic map**, or **Henon map**, in perturbation theory of Hamiltonian systems the **Standard map** featuring KAM tori or Mather sets, in homotopy theory the **dunce hat** or **Bing house**, in combinatorial topology the **Rudin sphere**, the **Nash-Kuiper non-smooth embedding** of a torus into Euclidean space. in topology the **Alexander horned sphere** or the **Antoine necklace**, or the **busy beaver** in Turing computation which is an illustration with how small machines one can achive great things, in fractal analysis the **Cantor set**, the **Menger sponge**, in Fourier theory the series of $f(x) = x \bmod 1$, in Diophantine approximation the **golden ratio**, in the calculus of sums the **zeta function**, in dimension theory the **Banach Tarski paradox**. In harmonic analysis the **Weierstrass function** as an example of a nowhere differentiable function. The case of **Peano curves** giving concrete examples of a continuous bijection from an interval to a square or cube. In **complex dynamics** not only the **Mandelbrot set** plays an important role, but also individual, specific Julia sets can be interesting. Examples like the **Mandelbulb** have not even been started to be investigated. There seem to be no theorems known about this object. In mathematical physics, the **almost Matthieu operator** [87] produced a rich theory related to spectral theory, Diophantine approximation, fractal geometry and functional analysis.

## Physics

One can also make a list of great ideas in physics [104] and see the relations with the fundamental theorems in mathematics. A high applicability should then contribute to a **value functional** in the list of theorems. Great ideas in physics are **the concept of space and time** meaning to describe physical events using **differential equations**. In cosmology, one of the insights is to look at space-time and realize the expansion of the universe or that the idea of a **big bang**. More general is the Platonic idea that **physics is geometry**. Or calculus: Lagrange developed his **calculus of variations** to find laws of physics. Then there is the idea of **Lorentz invariance** which leads to **special relativity**, there is the idea of **general relativity** which allows to describe gravity through geometry and a larger symmetry seen through the **equivalence principle**. There is the idea of see elementary particles using **Lie groups**. There is the **Noether theorem** which is the idea that any **symmetry** is tied to a **conservation law**: translational symmetry leads to momentum conservation, rotation symmetry to angular momentum conservation for example. Symmetries also play a role when **spontaneous broken symmetry** or **phase transitions**. There is the idea of quantum mechanics which mathematically means replacing differential equations with **partial differential equations** or replacing commutative algebras of observables with **non-commutative algebras**. An important idea is the concept of **perturbation theory** and in particular the notion of **linearization**. Many laws are simplifications of more complicated laws and described in the simplest cases through linear laws like Ohms law or Hooks law. **Quantization processes** allow to go from commutative to non-commutative structures. **Perturbation theory** allows then to extrapolate from a simple law to a more complicated law. Some is easy application of the **implicit function theorem**, some is harder like KAM theory. There is the idea of using **discrete mathematics** to describe complicated processes. An example is the language of **Feynman graphs** or the language of graph theory in general to describe physics as in loop quantum gravity or then the language of **cellular automata** which can be seen as partial difference equations where also the function space is quantized. The idea of **quantization**, a formal transition from an ordinary differential equation like a Hamiltonian system to a partial differential equation or to replace single particle systems with infinite particle systems (Fock). There are other quantization approaches through **deformation of algebras** which is related to **non-commutative geometry**. There is the idea of using **smooth functions** to describe discrete particle processes. An example is the **Vlasov dynamical system** or **Boltzmann's equation** to describe a plasma, or thermodynamic notions to describe large sets of particles like a gas or fluid. Dual to this is the use of **discretization** to describe a smooth system by discrete processes. An example is **numerical approximation**, like using the Runge-Kutta scheme to compute the trajectory of a differential equation. There is the realization that we have a whole spectrum of dynamical systems, **integrability** and **chaos** and that some of the transitions are **universal**. An other example is the **tight binding approximation** in which a continuum Schrödinger equation is replaced with a bounded **discrete Jacobi operator**. There is the general idea of finding the **building blocks** or **elementary particles**. Starting with Demokrit in ancient Greece, the idea got refined again and again. Once, atoms were found and charges found to be quantized (Millikan), the structure of the atom was explored (Rutherford), and then the atom got split (Meitner, Hahn). The structure of the nuclei with protons and neutrons was then refined again using quarks. There is furthermore the idea to use statistical methods for complex systems. An example is the use of stochastic differential equations like diffusion processes to describe actually deterministic particle systems. There is the insight that complicated systems can form **patterns** through interplay between symmetry, conservation laws and **synchronization**. Large scale patterns can be formed from systems with local laws. Finally, there is the idea of solving **inverse problems** using mathematical tools like Fourier theory or basic geometry (Erathostenes could compute the radius of the earth by comparing the lengths of shadows at different places of the earth.) An example is **tomography**, where the structure of some object is explored using **resonance**. Then there is the idea of **scale**

**invariance** which allows to describe objects which have **fractal nature**.

## Computer science

As in physics, it is harder to pinpoint "big ideas" in computer science as they are in general not theorems. The initial steps of mathematics was to build a **language**, where **numbers** represent quantities [80]. Physical tools which assist in manipulating numbers can already been seen as a **computing device**. Marks on a bone, pebbles in a clay bag, knots in a Quipu, marks on a Clay tablet were the first step. Papyri, paper, magnetic, optical and electric storage, the tools to build **memory** were refined over the millenniums. The mathematical language allowed us to get further than the finite. Using a finite number of symbols we can represent and count infinite sets, have notions of **cardinality**, have **various number systems** and more general **algebraic structures**. Numbers can even be seen as **games** [79, 207]. A major idea is the concept of an **algorithm**. Adding or multiplying on an **abacus** already was an algorithm. The concept was refined in geometry, where **ruler and compass** were used as **computing devices**, like the construction of points in a triangle. To measure the effectiveness of an algorithm, one can use notions of **complexity**. This has been made precise by computing pioneers like Turing as one has to formulate first what a computation is. In the last century one has seen that computations and proofs are very similar and that they have similar general restrictions. There are some tasks which can not be computed with a Turing machine and there are theorems which can not be proven in a specific axiom system. As mathematics is a language, we have to deal with concepts of **syntax**, **grammar**, **notation**, **context**, **parsing**, **validation**, **verification**. As Mathematics is a **human activity** which is done in our **brains**, it is related to psychology and **computer architecture**. Computer science aspects are also important also in **pedagogy** and **education** how can an idea be communicated **clearly**? How do we **motivate**? How do we **convince** peers that a result is true? Examples from history show that this is often done by **authority** and that the validity of some proofs turned out to be wrong or incomplete, even in the case of fundamental theorems or when treated by great mathematicians. (Examples are the fundamental theorem of arithmetic, the fundamental theorem of algebra or the wrong published proof of Kempe of the 4 color theorem). How come we trust a human brain more than an electronic one? We have to make some fundamental assumptions for example to be made like that if we do a logical step "if A and B then "A and B" holds. This assumes for example that **our memory is faithful**: after having put A and B in the memory and making the conclusion, we have to assume that we do not forget A or B! Why do we trust this more than the memory of a machine? As we are also assisted more and more by electronic devices, the question of the validity of **computer assisted proofs** comes up. The **4-color theorem** of Kenneth Appel and Wolfgang Haken based on previous work of many others like Heinrich Heesch or the proof of the **Feigenbaum conjecture** of Mitchell Feigenbaum first proven by Oscar Lanford III or the proof of the Kepler problem by Thomas Hales are examples. A great general idea is related to the representation of **data**. This can be done using matrices like in a **relational database** or using other structures like **graphs** leading to **graph databases**. The ability to use computers allows mathematicians to do **experiments**. A branch of mathematics called **experimental mathematics** [172] relies heavily on experiments to find new theorems or relations. Experiments are related to **simulations**. We are able, within a computer to build and explore new worlds, like in **computer games**, we can enhance the physical world using **virtual reality** or **augmented reality** or then **capturing a world** by

**3D scanning** and **realize** a world by **printing the objects** [206]. A major theme is **artificial intelligence** [285, 173]. It is related to optimization problems like optimal transport, neural nets as well as **inverse problems** like **structure from motion problems**. An intelligent entity must be able to take information, build a model and then find an optimal strategy to solve a given task. A self-driving car for example has to be able to translate pictures from a camera and build a map, then determine where to drive. Such tasks are usually considered part of **applied mathematics** but they are very much related with pure mathematics because computers also start to learn how to read mathematics, how to **verify proofs** and to **find new theorems**. Artificial intelligent agents [337] were first developed in the 1960ies learned also some mathematics. I myself learned about it when incorporated computer algebra systems into a chatbots in [204]. AI has now become a big business as **Alexa**, **Siri**, **Google Home**, **IBM Watson** or **Cortana** demonstrate. But these information systems must be taught, they must be able to rank alternative answers, even inject some humor or opinions. Soon, they will be able to learn themselves and answer questions like "what are the 10 most important theorems in mathematics?"

## Brevity

We live in a instagram, snapchat, twitter, microblog, vine, watch-mojo, petcha-kutcha time. Many of us multi task, read news on smart phones, watch faster paced movies, read shorter novels and feel that a million word Marcel Proust's masterpiece "a la recherche du temps perdu" is "temps perdu". Even classrooms and seminars have become more aphoristic. Micro blogging tools are only the latest incarnation of "miniature stories". They continue the tradition of older formats like "mural art" by Romans to modern graffiti or "aphorisms" [213, 214]), poetry, cartoons, Unix fortune cookies [16]. Shortness has appeal: aphorisms, poems, ferry tales, quotes, words of wisdom, life hacker lists, and tabloid top 10 lists illustrate this. And then there are books like "Math in 5 minutes", "30 second math", "math in minutes" [28, 133, 107], which are great coffee table additions. Also short proofs are appealing like "Let epsilon smaller than zero" which is the shortest known math joke, or "There are three type of mathematicians, the ones who can count, and the ones who can't.". Also short open problems are attractive, like the **twin prime problem** "there are infinitely many twin primes" or the **Landau problem** "there are infinitely many primes of the form $n^2+1$, or the **Goldbach** problem "every $n > 2$ is the sum of two primes". For the larger public in mathematics shortness has appeal: according to a poll of the Mathematical Intelligencer from 1988, the most favorite theorems are short [339]. Results with longer proofs can make it to graduate courses or specialized textbooks but still then, the results are often short enough so that they can be tweeted without proof. Why is shortness attractive? Paul Erdös expressed short elegant proofs as "proofs from the book" [8]. Shortness reduces the possibility of error as complexity is always a stumbling block for understanding. But is beauty equivalent to brevity? Buckminster Fuller once said: "If the solution is not beautiful, I know it is wrong." [6]. Much about the aesthetics in mathematics is investigated in [244]. According to [280], the beauty of a piece of mathematics is frequently associated with the shortness of statement or of proof: *beautiful theories are also thought of as short, self-contained chapters fitting within broader theories. There are examples of complex and extensive theories which every mathematician agrees to be beautiful, but these examples are not the one which come to mind.* Also psychologists and educators know that simplicity appeals to children: From [299] *For now, I want simply to draw attention to the fact that even for a young, mathematically naive child, aesthetic sensibilities and values (a penchant for* *simplicity, for finding the building blocks of more complex ideas, and a preference for shortcuts and "liberating" tricks rather than cumbersome recipes) animates mathematical experience.* It is hard to exhaust them all, as the more than googool$^2 = 10^{200}$ texts of length 140 can hardly all ever be written down. But there are even short story collections. Berry's paradox tells in this context that the shortest non-tweetable text in 140 characters can be tweeted: "The shortest non-tweetable text". Since we insist on giving proofs, we have to cut corners. Books containing lots of elegant examples are [13, 8].

## Twitter math

The following 42 tweets were written in 2014, when twitter had still a 140 character limit. Some of them were actually tweeted. The experiment was to see which theorems are short enough so that one can tweet both the theorem as well as the proof in 140 characters. Of course, that often requires a bit of cheating. See [8] for proofs from the books, where the proofs have full details.

**Euclid:** The set of primes is infinite. Proof: let $p$ be largest prime, then $p!+1$ has a larger prime factor than $p$. Contradiction.

**Euclid:** $2^p-1$ prime then $2^{p-1}(2^p-1)$ is perfect. Proof. $\sigma(n) =$ sum of factors of $n$, $\sigma(2^n-1)2^{n-1}) = \sigma(2^n-1)\sigma(2^{n-1}) = 2^n(2^n-1) = 2 \cdot 2^n(2^n-1)$ shows $\sigma(k) = 2k$.

**Hippasus:** $\sqrt{2}$ is irrational. Proof. If $\sqrt{2} = p/q$, then $2a^2 = p^2$. To the left is an odd number of factors 2, to the right it is even one. Contradiction.

**Pythagorean triples:** all $x^2+y^2=z^2$ are of form $(x,y,z) = (2st, s^2-t^2, s^2+t^2)$. Proof: $x$ or $y$ is even (both odd gives $x^2+y^2=w^k$ with odd $k$). Say $x^2$ is even: write $x^2 = z^2-y^2 = (z-y)(z+y)$. This is $4s^2t^2$. Therefore $2s^2 = z-y, 2t^2 = z+y$. Solve for $z, y$.

**Pigeon principle:** if $n+1$ pigeons live in $n$ boxes, there is a box with 2 or more pigeons. Proof: place a pigeon in each box until every box is filled. The pigeon left must have a roommate.

**Angle sum in triangle:** $\alpha+\beta+\gamma = KA+\pi$ if $K$ is curvature, $A$ triangle area. Proof: Gauss-Bonnet for surface with boundary. $\alpha, \beta, \gamma$ are Dirac measures on the boundary.

**Chinese remainder theorem:** a(i) x = b(i) mod n(i) has a solution if gcd(a(i),n(i))=0 and gcd(n(i),n(j))=0 Proof: solve eq(1), then increment x by n(1) to solve eq(2), then increment x by n(1) n(2) until second is ok. etc.

**Nullstellensatz:** algebraic sets in $K^n$ are 1:1 to radical ideals in $K[x_1...x_n]$. Proof: An algebra over K which is a field is finite field extension of K.

**Fundamental theorem algebra:** a polynomial of degree $n$ has exactly $n$ roots. Proof: the metric $g = |f|^{-2/n}|dz|^2$ on the Riemann sphere has curvature $K = n^{-1}\Delta \log |f|$. Without root, K=0 everywhere contradicting Gauss-Bonnet. [11]:

**Fermat:** p prime $(a, p) = 1$, then $p|a^p - a$ Proof: induction with respect to $a$. Case $a = 1$ is trivial $(a + 1)^p - (a + 1)$ is congruent to $a^p - a$ modulo $p$ because Binomial coefficients $B(p, k)$ are divisible by $p$ for $k = 1, \dots p - 1$.

**Wilson:** $p$ is prime iff $p|(p - 1)! + 1$ Proof. Group $2, \dots p - 2$ into pairs $(a, a^-1)$ whose product is 1 modulo $p$. Now $(p-1)! = (p-1) = -1$ modulo $p$. If $p = ab$ is not prime, then $(p-1)! = 0$ modulo $p$ and $p$ does not divide $(p - 1)! + 1$.

**Bayes:** $A, B$ are events and $A^c$ is the complement. $P[A|B] = P[B|A]P[A]/(P[B|A]P[A] + P[B|A^c]P[A^c]$ Proof: By definition $P[A|B]P[B] = P[A \cap B]$. Also $P[B] = (P[B|A]P[A] + P[B|A^c]P[A^c]$.

**Archimedes:** Volume of sphere $S(r)$ is $4\pi r^3/3$ Proof: the complement of the cone inside the cylinder has at height $z$ the cross section area $r^2 - z^2$, the same as the cross section area of the sphere at height $z$.

**Archimedes:** the area of the sphere $S(r)$ is $4\pi r^2$ Proof: differentiate the volume formula with respect to $r$ or project the sphere onto a cylinder of height 2 and circumference $2\pi$ and not that this is area preserving.

**Cauchy-Schwarz:** $|v \cdot w| <= |v||w|$. Proof: scale to get $|w| = 1$, define $a = v.w$, so that $0 \leq (v-aw) \dots (v-aw) = |v|^2 - a^2 = |v|^2|w|^2 - (v \cdot w)^2$.

**Angle formula:** Cauchy-Schwarz defines the angle between two vectors as $\cos(A) = v.w/|v||w|$. If $v, w$ are centered random variables, then $v \cdot w$ is the covariance, $|v|, |w|$ are standard deviations and $\cos(A)$ is the correlation.

**Cos formula:** $c^2 = a^2 + b^2 - ab\cos(A)$ in a triangle ABC is also called Al-Kashi theorem. Proof: $v = AB, w = AC$ has length $a = |v|, b = |w|, |c| = |v - w|$. Now: $(v - w).(v - w) = |v|^2 + |w|^2 - 2|v||w|\cos(A)$.

**Pythagoras:** $A = \pi/2$, then $c^2 = a^2 + b^2$. Proof: Let $v = AB$, $w = AC$, $v - w = BC$ be the sides of the triangle. Multiply out $(v - w) \cdot (v - w) = |v|^2 + |w|^2$ and use $v \cdot w = 0$.

**Euler formula:** $\exp(ix) = \cos(x) + i\sin(x)$. Proof: $\exp(ix) = 1 + (ix) + (ix)^2/2! - ...$ Pair real and imaginary parts and use definition $\cos(x) = 1 - x^2/2! + x^4/4!...$ and $\sin(x) = x - x^3/3! + x^5/5! - .....$

**Discrete Gauss-Bonnet** $\sum_x K(x) = \chi(G)$ with $K(x) = 1 - V_0(x)/2 + V_1(x)/3 + V_2(x)/4...$ curvature $\chi(G) = v_0 - v_1 + v_2 - v_3...$ Euler characteristic Proof: Use handshake $\sum_x V_k(x) = v_{k+1}/(k + 2)$.

**Poincaré-Hopf:** let $f$ be a coloring, $i_f(x) = 1 - \chi(S_f^-(x))$, where $S_f^-(x) = y \in S(x)|f(y) < f(x)$ $\sum i_f(x) = \chi(G)$. Proof by induction. Removing local maximum of $f$ reduces Euler characteristic by $\chi(B_f(x)) - \chi(S^- f(x)) = i_f(x)$.

**Lefschetz:** $\sum_x i_T(x) = \text{str}(T|H(G))$. Proof: LHS is $\text{str}(\exp(-0L)U_T)$ and RHS is $\text{str}(exp(-tL)U_T)$ for $t \to \infty$. The super trace does not depend on $t$.

**Stokes:** orient edges $E$ of graph $G$. $F : E \to R$ function, $S$ surface in G with boundary C. $d(F)(ijk) = F(ij) + F(jk) - F(ki)$ is the curl. The sum of the curls over all triangles is the line integral of $F$ along $C$.

**Plato:** there are exactly 5 platonic solids. Proof: number $f$ of $n$-gon satisfies $f = 2e/n$, $v$ vertices of degree $m$ satisfy $v = 2e/m$ $v - e + f - 2$ means $2e/m - e + 2e/n = 2$ or $1/m + 1/n = 1/e + 1/2$ with solutions: $(m = 4, n = 3), (m = 3, n = 5), (n = m = 3), (n = 3, m = 5), (m = 3, n = 4)$.

**Poincaré recurrence:** $T$ area-preserving map of probability space $(X, m)$. If $m(A) > 0$ and $n > 1/m(A)$ we have $m(T^k(A) \cap A) > 0$ for some $1 \leq k \leq n$ Proof. Otherwise $A, T(A), ..., T^n(A)$ are all disjoint and the union has measure $n \cdot m(A) > 1$.

**Turing:** there is no Turing machine which halts if input is Turing machine which halts: Proof: otherwise build an other one which halts if the input is a non-halting one and does not halt if input is a halting one.

**Cantor:** the set of reals in [0,1] is uncountable. Proof: if there is an enumeration $x(k)$, let $x(k, l)$ be the $l$'th digit of $x(k)$ in binary form. The number with binary expansion $y(k) = x(k, k) + 1$ mod 2 is not in the list.

**Niven:** $\pi \notin Q$: Proof: $\pi = a/b$, $f(x) = x^n(a - bx)^n/n!$ satisfies $f(pi - x) = f(x)$ and $0 < f(x) < \pi^n a^n/n^n$ $f^(j)(x) = 0$ at 0 and $\pi$ for $0 \leq j \leq n$ shows $F(x) = f(x) - f^{(2)}(x) + f^{(4)}(x) \cdots + (-1)^n f^{(2n)}(x)$ has $F(0), F(\pi) \in Z$ and $F + F'' = f$. Now $(F'(x)\sin(x) - F(x)\cos(x)) = f\sin(x)$, so $\int_0^\pi f(x)\sin(x)dx \in Z$.

**Fundamental theorem calculus:** With differentiation $Df(x) = f(x+1) - f(x)$ and integration $Sf(x) = f(0) + f(1) + ... + f(n-1)$ have $SDf(x) = f(x) - f(0), DSf(x) = f(x)$.

**Taylor:** $f(x+t) = \sum_k f^{(k)}(x)t^k/k!$. Proof: $f(x+t)$ satisfies transport equation $f_t = f_x = Df$ an ODE for the differential operator $D$. Solve $f(x+t) = \exp(Dt)f(x)$.

**Cauchy-Binet:** $\det(1 + F^T G) = \sum_P \det(F_P)\det(G_P)$ Proof: $A = F^T G$. Coefficients of $\det(x - A)$ is $\sum_P = k\det(F_P)\det(G_P)$.

**Intermediate:** $f$ continuous $f(0) < 0, f(1) > 0$, then there exists $0 < x < 1, f(x) = 0$. Proof. If $f(1/2) < 0$ do proof with $(1/2, 1)$ If $f(1/2) > 0$ redo proof with $(0, 1/2)$.

**Ergodicity:** $T(x) = x + a \bmod 1$ with irrational $a$ is ergodic. Proof. $f = \sum_n a(n) \exp(inx)$ $Tf = \sum_n a(n) \exp(ina) \exp(inx) = f$ implies $a(n) = 0$.

**Benford:** first digit $k$ of $2^n$ appears with probability $\log(1 - 1/k)$ Proof: $T : x \to x + \log(2) \bmod 1$ is ergodic. $\log(2^n) \mathrm{mod}\ 1 = k$ if $\log(k) \leq T^n(0) < \log(k+1)$. The probability of hitting this interval is $\log(k+1)/\log(k)$.

**Rank-Nullity:** $\dim(\ker(A)) + \dim(\mathrm{im}(A)) = n$ for $m \times n$ matrix $A$. Proof: a column has a leading 1 in $rref(A)$ or no leading 1. In the first case it contributes to the image, in the second to a free variable parametrizing the kernel.

**Column-Row picture:** $A : R^m \to R^n$. The $k$'th column of $A$ is the image $Ae_k$. If all rows of $A$ are perpendicular to $x$ then $x$ is in the kernel of $A$.

**Picard:** $x' = f(x), x(0) = x_0$ has locally a unique solution if $f \in C^1$. Proof: the map $T(y) = \int_0^t f(y(s))\ ds$ is a contraction on $C([0, a])$ for small enough $a > 0$. Banach fixed point theorem.

**Banach:** a contraction $d(T(x), T(y)) \leq a d(x, y)$ on complete $(X, d)$ has a unique fixed point. Proof: $d(x_k, x_n) \leq a^k/(1 - a)$ using triangle inequality and geometric series. Have Cauchy sequence.

**Liouville:** every prime p=4k+1 is the sum of two squares. Proof: there is an involution on $S = (x, y, z)|x^2 + 4yz = p$ with exactly one fixed point showing —S— is odd implying $(x, y, z) -> (x, z, y)$ has a fixed point. [349]

**Banach-Tarski:** The unit ball in $R^3$ can be cut into 5 pieces, re-assembled using rotation and translation to get two spheres. Proof: cut cleverly using axiom of choice.

## Math areas

We add here the core handouts of Math E320 which aimed to give for each of the 12 mathematical subjects an overview on two pages. For that course, I had recommended books like [116, 140, 34, 311, 312].

**E-320: Teaching Math with a Historical Perspective** **O. Knill, 2010-2018**

# Lecture 1: Mathematical roots

Similarly, as one has distinguished the **canons of rhetorics**: memory, invention, delivery, style, and arrangement, or combined the **trivium**: grammar, logic and rhetorics, with the **quadrivium**: arithmetic, geometry, music, and astronomy, to obtain the seven **liberal arts and sciences**, one has tried to **organize all mathematical activities**.

Historically, one has distinguished **eight ancient roots of mathematics**. Each of these 8 activities in turn suggest a key area in mathematics:

| | |
|---|---|
| counting and sorting | **arithmetic** |
| spacing and distancing | **geometry** |
| positioning and locating | **topology** |
| surveying and angulating | **trigonometry** |
| balancing and weighing | **statics** |
| moving and hitting | **dynamics** |
| guessing and judging | **probability** |
| collecting and ordering | **algorithms** |

To morph these 8 roots to the 12 mathematical areas covered in this class, we complemented the ancient roots with calculus, numerics and computer science, merge trigonometry with geometry, separate arithmetic into number theory, algebra and arithmetic and turn statics into analysis.

Let us call this modern adaptation the

**12 modern roots of Mathematics**:

| | |
|---|---|
| counting and sorting | **arithmetic** |
| spacing and distancing | **geometry** |
| positioning and locating | **topology** |
| dividing and comparing | **number theory** |
| balancing and weighing | **analysis** |
| moving and hitting | **dynamics** |
| guessing and judging | **probability** |
| collecting and ordering | **algorithms** |
| slicing and stacking | **calculus** |
| operating and memorizing | **computer science** |
| optimizing and planning | **numerics** |
| manipulating and solving | **algebra** |

While relating **mathematical areas** with **human activities** is useful, it makes sense to select specific topics in each of this area. These 12 topics will be the 12 lectures of this course.

| | |
|---|---|
| Arithmetic | numbers and number systems |
| Geometry | invariance, symmetries, measurement, maps |
| Number theory | Diophantine equations, factorizations |
| Algebra | algebraic and discrete structures |
| Calculus | limits, derivatives, integrals |
| Set Theory | set theory, foundations and formalisms |
| Probability | combinatorics, measure theory and statistics |
| Topology | polyhedra, topological spaces, manifolds |
| Analysis | extrema, estimates, variation, measure |
| Numerics | numerical schemes, codes, cryptology |
| Dynamics | differential equations, maps |
| Algorithms | computer science, artificial intelligence |

Like any classification, this chosen division is rather arbitrary and a matter of personal preferences. The **2010 AMS classification** distinguishes 64 areas of mathematics. Many of the just defined main areas are broken off into even finer pieces. Additionally, there are fields which relate with other areas of science, like economics, biology or physics:a

00 General
01 History and biography
03 Mathematical logic and foundations
05 Combinatorics
06 Lattices, ordered algebraic structures
08 General algebraic systems
11 Number theory
12 Field theory and polynomials
13 Commutative rings and algebras
14 Algebraic geometry
15 Linear/multi-linear algebra; matrix theory
16 Associative rings and algebras
17 Non-associative rings and algebras
18 Category theory, homological algebra
19 K-theory
20 Group theory and generalizations

45 Integral equations
46 Functional analysis
47 Operator theory
49 Calculus of variations, optimization
51 Geometry
52 Convex and discrete geometry
53 Differential geometry
54 General topology
55 Algebraic topology
57 Manifolds and cell complexes
58 Global analysis, analysis on manifolds
60 Probability theory and stochastic processes
62 Statistics
65 Numerical analysis
68 Computer science
70 Mechanics of particles and systems

22 Topological groups, Lie groups
26 Real functions
28 Measure and integration
30 Functions of a complex variable
31 Potential theory
32 Several complex variables, analytic spaces
33 Special functions
34 Ordinary differential equations
35 Partial differential equations
37 Dynamical systems and ergodic theory
39 Difference and functional equations
40 Sequences, series, summability
41 Approximations and expansions
42 Fourier analysis
43 Abstract harmonic analysis
44 Integral transforms, operational calculus

74 Mechanics of deformable solids
76 Fluid mechanics
78 Optics, electromagnetic theory
80 Classical thermodynamics, heat transfer
81 Quantum theory
82 Statistical mechanics, structure of matter
83 Relativity and gravitational theory
85 Astronomy and astrophysics
86 Geophysics
90 Operations research, math. programming
91 Game theory, Economics Social and Behavioral Sciences
92 Biology and other natural sciences
93 Systems theory and control
94 Information and communication, circuits
97 Mathematics education

What are

**fancy developments**

in mathematics today? Michael Atiyah [20] identified in the year 2000 the following **six hot spots**:

| | | |
|---|---|---|
| local | and | global |
| low | and | high dimension |
| commutative | and | non-commutative |
| linear | and | nonlinear |
| geometry | and | algebra |
| physics | and | mathematics |

Also this choice is of course highly personal. One can easily add 12 other **polarizing** quantities which help to distinguish or parametrize different parts of mathematical areas, especially the ambivalent pairs which produce a captivating gradient:

| | | |
|---|---|---|
| regularity | and | randomness |
| integrable | and | non-integrable |
| invariants | and | perturbations |
| experimental | and | deductive |
| polynomial | and | exponential |
| applied | and | abstract |

| | | |
|---|---|---|
| discrete | and | continuous |
| existence | and | construction |
| finite dim | and | infinite dimensional |
| topological | and | differential geometric |
| practical | and | theoretical |
| axiomatic | and | case based |

The goal is to illustrate some of these structures from a historical point of view and show that "Mathematics is the science of structure".

# Lecture 2: Arithmetic

The oldest mathematical discipline is **arithmetic**. It is the theory of the construction and manipulation of numbers. The earliest steps were done by **Babylonian**, **Egyptian**, **Chinese**, **Indian** and **Greek** thinkers. Building up the number system starts with the **natural numbers** $1, 2, 3, 4...$ which can be added and multiplied. Addition is natural: join 3 sticks to 5 sticks to get 8 sticks. Multiplication $*$ is more subtle: $3*4$ means to take 3 copies of 4 and get $4+4+4 = 12$ while $4*3$ means to take 4 copies of 3 to get $3+3+3+3 = 12$. The first factor counts the number of operations while the second factor counts the objects. To motivate $3*4 = 4*3$, spacial insight motivates to arrange the 12 objects in a rectangle. This commutativity axiom will be carried over to larger number systems. Realizing an addition and multiplicative structure on the natural numbers requires to define 0 and 1. It leads naturally to more general numbers. There are two major motivations to **to build new numbers**: we want to

1. **invert operations** and still get results. 2. **solve equations**.

To find an additive inverse of 3 means solving $x + 3 = 0$. The answer is a negative number. To solve $x * 3 = 1$, we get to a rational number $x = 1/3$. To solve $x^2 = 2$ one need to escape to real numbers. To solve $x^2 = -2$ requires complex numbers.

| Numbers | Operation to complete | Examples of equations to solve |
|---|---|---|
| Natural numbers | addition and multiplication | $5 + x = 9$ |
| Positive fractions | addition and division | $5x = 8$ |
| Integers | subtraction | $5 + x = 3$ |
| Rational numbers | division | $3x = 5$ |
| Algebraic numbers | taking positive roots | $x^2 = 2$ , $2x + x^2 - x^3 = 2$ |
| Real numbers | taking limits | $x = 1 - 1/3 + 1/5 - +...$,$\cos(x) = x$ |
| Complex numbers | take any roots | $x^2 = -2$ |
| Surreal numbers | transfinite limits | $x^2 = \omega$, $1/x = \omega$ |
| Surreal complex | any operation | $x^2 + 1 = -\omega$ |

The development and history of arithmetic can be summarized as follows: humans started with natural numbers, dealt with positive fractions, reluctantly introduced negative numbers and zero to get the integers, struggled to "realize" real numbers, were scared to introduce complex numbers, hardly accepted surreal numbers and most do not even know about surreal complex numbers. Ironically, as simple but impossibly difficult questions in number theory show, the modern point of view is the opposite to Kronecker's **"God made the integers; all else is the work of man"**:

The **surreal complex** numbers are the most **natural** numbers;
The **natural** numbers are the most **complex, surreal** numbers.

**Natural numbers**. Counting can be realized by sticks, bones, quipu knots, pebbles or wampum knots. The **tally stick** concept is still used when playing card games: where bundles of fives are formed, maybe by crossing 4 "sticks" with a fifth. There is a "log counting" method in which graphs are used and vertices and edges count. An old stone age tally stick, the **wolf radius bone** contains 55 notches, with 5 groups of 5. It is probably more than 30'000 years old. [303] The most famous paleolithic tally stick is the **Ishango bone**, the fibula of a baboon. It could be 20'000 - 30'000 years old. [116] Earlier counting could have been done by assembling **pebbles**, tying **knots** in a string, making **scratches** in dirt or bark but no such traces have survived the thousands of years. The **Roman system** improved the tally stick concept by introducing new symbols for larger numbers like $V = 5, X = 10, L = 40, C = 100, D = 500, M = 1000$. in order to avoid bundling too many single sticks.

The system is unfit for computations as simple calculations $VIII + VII = XV$ show. **Clay tablets**, some as early as 2000 BC and others from 600 - 300 BC are known. They feature **Akkadian arithmetic** using the base 60. The hexadecimal system with base 60 is convenient because of many factors. It survived: we use 60 minutes per hour. **The Egyptians** used the base 10. The most important source on Egyptian mathematics is the **Rhind Papyrus** of 1650 BC. It was found in 1858 [188, 303]. Hieratic numerals were used to write on papyrus from 2500 BC on. **Egyptian numerals** are hieroglyphics. Found in carvings on tombs and monuments they are 5000 years old. The modern way to write numbers like 2018 is the **Hindu-Arab system** which diffused to the West only during the late Middle ages. It replaced the more primitive **Roman system**. [303] Greek arithmetic used a number system with no place values: 9 Greek letters for $1, 2, \dots 9$, nine for $10, 20, \dots, 90$ and nine for $100, 200, \dots, 900$.

**Integers**. **Indian Mathematics** morphed the place-value system into a modern method of writing numbers. Hindu astronomers used words to represent digits, but the numbers would be written in the opposite order. Independently, also the Mayans developed the concept of 0 in a number system using base 20. Sometimes after 500, the Hindus changed to a digital notation which included the symbol 0. Negative numbers were introduced around 100 BC in the **Chinese** text "Nine Chapters on the Mathematical art". Also the **Bakshali manuscript**, written around 300 AD subtracts numbers carried out additions with negative numbers, where + was used to indicate a negative sign. [267] In Europe, negative numbers were avoided until the 15'th century.

**Fractions**: **Babylonians** could handle fractions. The **Egyptians** also used fractions, but wrote every fraction a as a sum of fractions with unit numerator and distinct denominators, like $4/5 = 1/2 + 1/4 + 1/20$ or $5/6 = 1/2 + 1/3$. Maybe because of such cumbersome computation techniques, Egyptian mathematics failed to progress beyond a primitive stage. [303]. The modern decimal fractions used nowadays for numerical calculations were adopted only in 1595 in Europe.

**Real numbers:** As noted by the Greeks already, the diagonal of the square is not a fraction. It first produced a crisis until it became clear that "most" numbers are not rational. **Georg Cantor** saw first that the cardinality of all real numbers is much larger than the cardinality of the integers: while one can count all rational numbers but not enumerate all real numbers. One consequence is that most real numbers are transcendental: they do not occur as solutions of polynomial equations with integer coefficients. The number $\pi$ is an example. The concept of real numbers is related to the **concept of limit**. Sums like $1 + 1/4 + 1/9 + 1/16 + 1/25 + \dots$ are not rational.

**Complex numbers:** some polynomials have no real root. To solve $x^2 = -1$ for example, we need new numbers. One idea is to use pairs of numbers $(a, b)$ where $(a, 0) = a$ are the usual numbers and extend addition and multiplication $(a, b) + (c, d) = (a + c, b + d)$ and $(a, b) \cdot (c, d) = (ac - bd, ad + bc)$. With this multiplication, the number $(0, 1)$ has the property that $(0, 1) \cdot (0, 1) = (-1, 0) = -1$. It is more convenient to write $a + ib$ where $i = (0, 1)$ satisfies $i^2 = -1$. One can now use the common rules of addition and multiplication.

**Surreal numbers:** Similarly as real numbers fill in the gaps between the integers, the surreal numbers fill in the gaps between Cantors ordinal numbers. They are written as $(a, b, c, ...|d, e, f, ...)$ meaning that the "simplest" number is larger than $a, b, c...$ and smaller than $d, e, f, ...$ We have $(|) = 0, (0|) = 1, (1|) = 2$ and $(0|1) = 1/2$ or $(|0) = -1$. Surreals contain already transfinite numbers like $(0, 1, 2, 3...|)$ or infinitesimal numbers like $(0|1/2, 1/3, 1/4, 1/5, ...)$. They were introduced in the 1970'ies by John Conway. The late appearance confirms the pedagogical principle: **late human discovery manifests in increased difficulty to teach it**.

# Lecture 3: Geometry

Geometry is the science of **shape, size and symmetry**. While arithmetic deals with numerical structures, geometry handles metric structures. Geometry is one of the oldest mathematical disciplines. Early geometry has relations with arithmetic: the multiplication of two numbers $n \times m$ as an area of a **shape** that is invariant under rotational **symmetry**. Identities like the **Pythagorean triples** $3^2 + 4^2 = 5^2$ were interpreted and drawn geometrically. The **right angle** is the most "symmetric" angle apart from 0. Symmetry manifests itself in quantities which are **invariant**. Invariants are one the most central aspects of geometry. Felix Klein's **Erlangen program** uses symmetry to classify geometries depending on how large the symmetries of the shapes are. In this lecture, we look at a few results which can all be stated in terms of invariants. In the presentation as well as the worksheet part of this lecture, we will work us through smaller miracles like **special points in triangles** as well as a couple of gems: **Pythagoras**, **Thales**,**Hippocrates**, **Feuerbach**, **Pappus**, **Morley**, **Butterfly** which illustrate the importance of symmetry.

Much of geometry is based on our ability to measure **length**, the **distance** between two points. Having a distance $d(A, B)$ between any two points $A, B$, we can look at the next more complicated object, which is a set $A, B, C$ of 3 points, a **triangle**. Given an arbitrary triangle ABC, are there relations between the 3 possible distances $a = d(B, C), b = d(A, C), c = d(A, B)$? If we fix the scale by $c = 1$, then $a + b \geq 1, a + 1 \geq b, b + 1 \geq a$. For any pair of $(a, b)$ in this region, there is a triangle. After an identification, we get an abstract space, which represent all triangles uniquely up to similarity. Mathematicians call this an example of a **moduli space**.

A **sphere** $S_r(x)$ is the set of points which have distance $r$ from a given point $x$. In the plane, the sphere is called a **circle**. A natural problem is to find the circumference $L = 2\pi$ of a unit circle, or the area $A = \pi$ of a unit disc, the area $F = 4\pi$ of a unit sphere and the volume $V = 4 = \pi/3$ of a unit sphere. Measuring the length of segments on the circle leads to new concepts like **angle** or **curvature**. Because the circumference of the unit circle in the plane is $L = 2\pi$, angle questions are tied to the number $\pi$, which Archimedes already approximated by fractions.

Also **volumes** were among the first quantities, Mathematicians wanted to measure and compute. A problem on **Moscow papyrus** dating back to 1850 BC explains the general formula $h(a^2 + ab + b^2)/3$ for a truncated pyramid with base length $a$, roof length $b$ and height $h$. Archimedes achieved to compute the **volume of the sphere**: place a cone inside a cylinder. The complement of the cone inside the cylinder has on each height $h$ the area $\pi - \pi h^2$. The half sphere cut at height $h$ is a disc of radius $(1 - h^2)$ which has area $\pi(1 - h^2)$ too. Since the slices at each height have the same area, the volume must be the same. The complement of the cone inside the cylinder has volume $\pi - \pi/3 = 2\pi/3$, half the volume of the sphere.

The first geometric playground was **planimetry**, the geometry in the flat two dimensional space. Highlights are **Pythagoras theorem**, **Thales theorem**, **Hippocrates theorem**, and **Pappus theorem**. Discoveries in planimetry have been made later on: an example is the Feuerbach 9 point theorem from the 19th century. Ancient Greek Mathematics is closely related to history. It starts with **Thales** goes over Euclid's era at 500 BC and ends with the threefold destruction of Alexandria 47 BC by the Romans, 392 by the Christians and 640 by the Muslims. Geometry was also a place, where the **axiomatic method** was brought to mathematics: theorems are proved from a few statements which are called axioms like the 5 axioms of Euclid:

1. Any two distinct points $A, B$ determines a line through $A$ and $B$.
2. A line segment $[A, B]$ can be extended to a straight line containing the segment.
3. A line segment $[A, B]$ determines a circle containing $B$ and center $A$.
4. All right angles are congruent.
5. If lines $L, M$ intersect with a third so that inner angles add up to $< \pi$, then $L, M$ intersect.

**Euclid** wondered whether the fifth postulate can be derived from the first four and called theorems derived from the first four the "absolute geometry". Only much later, with **Karl-Friedrich Gauss** and **Janos Bolyai** and **Nicolai Lobachevsky** in the 19'th century in **hyperbolic space** the 5'th axiom does not hold. Indeed, geometry can be generalized to non-flat, or even much more abstract situations. Basic examples are geometry on a sphere leading to **spherical geometry** or geometry on the Poincare disc, a **hyperbolic space**. Both of these geometries are non-Euclidean. **Riemannian geometry**, which is essential for **general relativity theory** generalizes both concepts to a great extent. An example is the geometry on an arbitrary surface. Curvatures of such spaces can be computed by measuring length alone, which is how long light needs to go from one point to the next.

An important moment in mathematics was the **merge of geometry with algebra**: this giant step is often attributed to **René Descartes**. Together with algebra, the subject leads to algebraic geometry which can be tackled with computers: here are some examples of geometries which are determined from the amount of symmetry which is allowed:

| | |
|---|---|
| Euclidean geometry | Properties invariant under a group of rotations and translations |
| Affine geometry | Properties invariant under a group of affine transformations |
| Projective geometry | Properties invariant under a group of projective transformations |
| Spherical geometry | Properties invariant under a group of rotations |
| Conformal geometry | Properties invariant under angle preserving transformations |
| Hyperbolic geometry | Properties invariant under a group of Möbius transformations |

Here are four pictures about the 4 special points in a triangle and with which we will begin the lecture. We will see why in each of these cases, the 3 lines intersect in a common point. It is a manifestation of a **symmetry** present on the space of all triangles. **size** of the distance of intersection points is constant 0 if we move on the space of all triangular **shapes**. It's Geometry!

**E-320: Teaching Math with a Historical Perspective** **Oliver Knill, 2010-2018**

# Lecture 4: Number Theory

Number theory studies the structure of integers like prime numbers and solutions to Diophantine equations. Gauss called it the "Queen of Mathematics". Here are a few theorems and open problems.

An integer larger than 1 which is divisible by 1 and itself only is called a **prime number**. The number $2^{57885161}-1$ is the largest known prime number. It has 17425170 digits. **Euclid** proved that there are infinitely many primes: [Proof. Assume there are only finitely many primes $p_1 < p_2 < \cdots < p_n$. Then $n = p_1 p_2 \cdots p_n + 1$ is not divisible by any $p_1, \dots, p_n$. Therefore, it is a prime or divisible by a prime larger than $p_n$.] Primes become more sparse as larger as they get. An important result is the **prime number theorem** which states that the $n$'th prime number has approximately the size $n\log(n)$. For example the $n = 10^{12}$'th prime is $p(n) = 29996224275833$ and $n\log(n) = 27631021115928.545...$ and $p(n)/(n\log(n)) = 1.0856...$ Many questions about prime numbers are unsettled: Here are four problems: the third uses the notation $(\Delta a)_n = |a_{n+1} - a_n|$ to get the absolute difference. For example: $\Delta^2(1,4,9,16,25...) = \Delta(3,5,7,9,11,...) = (2,2,2,2,...)$. Progress on prime gaps has been done in 2013: $p_{n+1} - p_n$ is smaller than 100'000'000 eventually (Yitang Zhang). $p_{n+1} - p_n$ is smaller than 600 eventually (Maynard). The largest known gap is 1476 which occurs after $p = 1425172824437699411$.

| | |
|---|---|
| **Landau** | there are infinitely many primes of the form $n^2+1$. |
| **Twin prime** | there are infinitely many primes $p$ such that $p+2$ is prime. |
| **Goldbach** | every even integer $n > 2$ is a sum of two primes. |
| **Gilbreath** | If $p_n$ enumerates the primes, then $(\Delta^k p)_1 = 1$ for all $k > 0$. |
| **Andrica** | The prime gap estimate $\sqrt{p_{n+1}} - \sqrt{p_n} < 1$ holds for all $n$. |

If the sum of the proper divisors of a $n$ is equal to $n$, then $n$ is called a **perfect number**. For example, 6 is perfect as its proper divisors $1,2,3$ sum up to 6. All currently known perfect numbers are even. The question whether odd perfect numbers exist is probably the oldest open problem in mathematics and not settled. Perfect numbers were familiar to Pythagoras and his followers already. Calendar coincidences like that we have 6 work days and the moon needs "perfect" 28 days to circle the earth could have helped to promote the "mystery" of perfect number. **Euclid of Alexandria** (300-275 BC) was the first to realize that if $2^p - 1$ is prime then $k = 2^{p-1}(2^p-1)$ is a perfect number: [Proof: let $\sigma(n)$ be the sum of **all** factors of $n$, including $n$. Now $\sigma(2^n-1)2^{n-1}) = \sigma(2^n-1)\sigma(2^{n-1}) = 2^n(2^n-1) = 2 \cdot 2^n(2^n-1)$ shows $\sigma(k) = 2k$ and verifies that $k$ is perfect.] Around 100 AD, **Nicomachus of Gerasa** (60-120) classified in his work "Introduction to Arithmetic" numbers on the concept of perfect numbers and lists four perfect numbers. Only much later it became clear that Euclid got all the even perfect numbers: Euler showed that all even perfect numbers are of the form $(2^n-1)2^{n-1}$, where $2^n - 1$ is prime. The factor $2^n - 1$ is called a **Mersenne prime**. [Proof: Assume $N = 2^k m$ is perfect where $m$ is odd and $k > 0$. Then $2^{k+1}m = 2N = \sigma(N) = (2^{k+1}-1)\sigma(m)$. This gives $\sigma(m) = 2^{k+1}m/(2^{k+1}-1) = m(1+1/(2^{k+1}-1)) = m + m/(2^{k+1}-1)$. Because $\sigma(m)$ and $m$ are integers, also $m/(2^{k+1}-1)$ is an integer. It must also be a factor of $m$. The only way that $\sigma(m)$ can be the sum of only two of its factors is that $m$ is prime and so $2^{k+1} - 1 = m$.] The first 39 **known Mersenne primes** are of the form $2^n - 1$ with n = 2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941, 11213, 19937, 21701, 23209, 44497, 86243, 110503, 132049, 216091, 756839, 859433, 1257787, 1398269, 2976221, 3021377, 6972593, 13466917. There are 11 more known from which one does not know the rank of the corresponding Mersenne prime: n = 20996011, 24036583, 25964951, 30402457, 32582657, 37156667, 42643801,43112609,57885161, 74207281,77232917. The last was found in December 2017 only. It is unknown whether there are infinitely many.

A polynomial equations for which all coefficients and variables are integers is called a **Diophantine equation**. The first Diophantine equation studied already by Babylonians is $x^2 + y^2 = z^2$. A solution $(x, y, z)$ of this equation in positive integers is called a **Pythagorean triple**. For example, $(3, 4, 5)$ is a Pythagorean triple. Since 1600 BC, it is known that all solutions to this equation are of the form $(x, y, z) = (2st, s^2 - t^2, s^2 + t^2)$ or $(x, y, z) = (s^2 - t^2, 2st, s^2 + t^2)$, where $s, t$ are different integers. [Proof. Either $x$ or $y$ has to be even because if both are odd, then the sum $x^2 + y^2$ is even but not divisible by 4 but the right hand side is either odd or divisible by 4. Move the even one, say $x^2$ to the left and write $x^2 = z^2 - y^2 = (z - y)(z + y)$, then the right hand side contains a factor 4 and is of the form $4s^2t^2$. Therefore $2s^2 = z - y, 2t^2 = z + y$. Solving for $z, y$ gives $z = s^2 + t^2, y = s^2 - t^2$, $x = 2st$.]
Analyzing Diophantine equations can be difficult. Only 10 years ago, one has established that the **Fermat equation** $x^n + y^n = z^n$ has no solutions with $xyz \neq 0$ if $n > 2$. Here are some **open problems** for Diophantine equations. Are there nontrivial solutions to the following Diophantine equations?

| | |
|---|---|
| $x^6 + y^6 + z^6 + u^6 + v^6 = w^6$ | $x, y, z, u, v, w > 0$ |
| $x^5 + y^5 + z^5 = w^5$ | $x, y, z, w > 0$ |
| $x^k + y^k = n!z^k$ | $k \geq 2, n > 1$ |
| $x^a + y^b = z^c, a, b, c > 2$ | $\gcd(a, b, c) = 1$ |

The last equation is called **Super Fermat**. A Texan banker **Andrew Beals** once sponsored a prize of $100'000$ dollars for a proof or counter example to the statement: "If $x^p + y^q = z^r$ with $p, q, r > 2$, then $\gcd(x, y, z) > 1$."
Given a prime like 7 and a number $n$ we can add or subtract multiples of 7 from $n$ to get a number in $\{0, 1, 2, 3, 4, 5, 6\}$. We write for example $19 = 12$ mod 7 because 12 and 19 both leave the rest 5 when dividing by 7. Or $5 * 6 = 2$ mod 7 because 30 leaves the rest 2 when dividing by 7. The most important theorem in elementary number theory is **Fermat's little theorem** which tells that if $a$ is an integer and $p$ is prime then $a^p - a$ is divisible by $p$. For example $2^7 - 2 = 126$ is divisible by 7. [Proof: use induction. For $a = 0$ it is clear. The binomial expansion shows that $(a+1)^p - a^p - 1$ is divisible by $p$. This means $(a+1)^p - (a+1) = (a^p - a) + mp$ for some $m$. By induction, $a^p - a$ is divisible by $p$ and so $(a + 1)^p - (a + 1)$.] An other beautiful theorem is **Wilson's theorem** which allows to characterize primes: It tells that $(n - 1)! + 1$ is divisible by $n$ if and only if $n$ is a prime number. For example, for $n = 5$, we verify that $4! + 1 = 25$ is divisible by 5. [Proof: assume $n$ is prime. There are then exactly two numbers $1, -1$ for which $x^2 - 1$ is divisible by $n$. The other numbers in $1, \dots, n - 1$ can be paired as $(a, b)$ with $ab = 1$. Rearranging the product shows $(n - 1)! = -1$ modulo $n$. Conversely, if $n$ is not prime, then $n = km$ with $k, m < n$ and $(n - 1)! = ...km$ is divisible by $n = km$. ]
The solution to systems of linear equations like $x = 3 \pmod 5, x = 2 \pmod 7$ is given by the **Chinese remainder theorem**. To solve it, continue adding 5 to 3 until we reach a number which leaves rest 2 to 7: on the list $3, 8, 13, 18, 23, 28, 33, 38$, the number 23 is the solution. Since 5 and 7 have no common divisor, the system of linear equations has a solution.
For a given $n$, how do we solve $x^2 - yn = 1$ for the unknowns $y, x$? A solution produces a square root $x$ of 1 modulo $n$. For prime $n$, only $x = 1, x = -1$ are the solutions. For composite $n = pq$, more solutions $x = r \cdot s$ where $r^2 = -1$ mod $p$ and $s^2 = -1$ mod $q$ appear. Finding $x$ is equivalent to factor $n$, because the greatest common divisor of $x^2 - 1$ and $n$ is a factor of $n$. **Factoring is difficult** if the numbers are large. It assures that **encryption algorithms** work and that bank accounts and communications stay safe. Number theory, once the least applied discipline of mathematics has become one of the most applied one in mathematics.

**E-320: Teaching Math with a Historical Perspective** **Oliver Knill, 2010-2018**

# Lecture 5: Algebra

Algebra studies **algebraic structures** like "groups" and "rings". The theory allows to solve polynomial equations, characterize objects by its symmetries and is the heart and soul of many puzzles. Lagrange claims **Diophantus** to be the inventor of Algebra, others argue that the subject started with solutions of **quadratic equation** by **Mohammed ben Musa Al-Khwarizmi** in the book Al-jabr w'al muqabala of 830 AD. Solutions to equation like $x^2 + 10x = 39$ are solved there by **completing the squares**: add 25 on both sides go get $x^2 + 10x + 25 = 64$ and so $(x + 5) = 8$ so that $x = 3$.
The use of **variables** introduced in school in **elementary algebra** were introduced later. Ancient texts only dealt with particular examples and calculations were done with concrete numbers in the realm of **arithmetic**. **Francois Viete** (1540-1603) used first letters like $A, B, C, X$ for variables.

The search for formulas for polynomial equations of degree 3 and 4 lasted 700 years. In the 16'th century, the cubic equation and quartic equations were solved. **Niccolo Tartaglia** and **Gerolamo Cardano** reduced the cubic to the quadratic: [first remove the quadratic part with $X = x - a/3$ so that $X^3 + aX^2 + bX + c$ becomes the **depressed cubic** $x^3 + px + q$. Now substitute $x = u - p/(3u)$ to get a quadratic equation $(u^6 + qu^3 - p^3/27)/u^3 = 0$ for $u^3$.] **Lodovico Ferrari** shows that the quartic equation can be reduced to the cubic. For the **quintic** however no formulas could be found. It was **Paolo Ruffini**, **Niels Abel** and **Évariste Galois** who independently realized that there are no formulas in terms of roots which allow to "solve" equations $p(x) = 0$ for polynomials $p$ of degree larger than 4. This was an amazing achievement and the birth of "group theory".

Two important algebraic structures are **groups** and **rings**.

In a **group** $G$ one has an operation $*$, an inverse $a^{-1}$ and a one-element 1 such that $a*(b*c) = (a*b)*c, a*1 = 1*a = a, a*a^{-1} = a^{-1}*a = 1$. For example, the set $Q^*$ of nonzero fractions $p/q$ with multiplication operation $*$ and inverse $1/a$ form a group. The integers with addition and inverse $a^{-1} = -a$ and "1"-element 0 form a group too. A **ring** $R$ has two compositions $+$ and $*$, where the plus operation is a group satisfying $a+b = b+a$ in which the one element is called 0. The multiplication operation $*$ has all group properties on $R^*$ except the existence of an inverse. The two operations $+$ and $*$ are glued together by the **distributive law** $a*(b+c) = a*b+a*c$. An example of a ring are the **integers** or the **rational numbers** or the **real numbers**. The later two are actually **fields**, rings for which the multiplication on nonzero elements is a group too. The ring of integers are no field because an integer like 5 has no multiplicative inverse. The ring of rational numbers however form a field.

Why is the theory of groups and rings not part of arithmetic? First of all, a crucial ingredient of algebra is the appearance of **variables** and computations with these algebras without using concrete numbers. Second, the algebraic structures are not restricted to "numbers". Groups and rings are general structures and extend for example to objects like the set of all possible symmetries of a geometric object. The set of all **similarity operations** on the plane for example form a group. An important example of a ring is the **polynomial ring** of all polynomials. Given any ring $R$ and a variable $x$, the set $R[x]$ consists of all polynomials with coefficients in $R$. The addition and multiplication is done like in $(x^2 + 3x + 1) + (x - 7) = x^2 + 4x - 7$. The problem to factor a given polynomial with integer coefficients into polynomials of smaller degree: $x^2 - x + 2$ for example can be written as $(x + 1)(x - 2)$ have a number theoretical flavor. Because symmetries of some structure form a group, we also have intimate connections with geometry. But this is not the only connection with geometry. Geometry also enters through the polynomial rings with several variables. Solutions to $f(x, y) = 0$ leads to geometric objects with shape and symmetry which sometimes even have their own algebraic structure. They are called **varieties**, a central object in **algebraic geometry**, objects which in turn have been generalized further to **schemes**, **algebraic spaces** or **stacks**.

Arithmetic introduces addition and multiplication of numbers. Both form a group. The operations can be written additively or multiplicatively. Lets look at this a bit closer: for integers, fractions and reals and the addition +, the 1 element 0 and inverse $-g$, we have a group. Many groups are written multiplicatively where the 1 element is 1. In the case of fractions or reals, 0 is not part of the multiplicative group because it is not possible to divide by 0. The nonzero fractions or the nonzero reals form a group. In all these examples the groups satisfy the commutative law $g * h = h * g$.
Here is a group which is not commutative: let $G$ be the set of all rotations in space, which leave the unit cube invariant. There are 3*3=9 rotations around each major coordinate axes, then 6 rotations around axes connecting midpoints of opposite edges, then 2*4 rotations around diagonals. Together with the identity rotation $e$, these are 24 rotations. The group operation is the composition of these transformations.
An other example of a group is $S_4$, the set of all permutations of four numbers $(1,2,3,4)$. If $g : (1,2,3,4) \to (2,3,4,1)$ is a permutation and $h : (1,2,3,4) \to (3,1,2,4)$ is an other permutation, then we can combine the two and define $h * g$ as the permutation which does first $g$ and then $h$. We end up with the permutation $(1,2,3,4) \to (1,2,4,3)$. The rotational symmetry group of the cube happens to be the same than the group $S_4$. To see this "isomorphism", label the 4 space diagonals in the cube by $1,2,3,4$. Given a rotation, we can look at the induced permutation of the diagonals and every rotation corresponds to exactly one permutation. The symmetry group can be introduced for any geometric object. For shapes like the triangle, the cube, the octahedron or tilings in the plane.

Symmetry groups describe geometric shapes by algebra.

Many **puzzles** are groups. A popular puzzle, the **15-puzzle** was invented in 1874 by **Noyes Palmer Chapman** in the state of New York. If the hole is given the number 0, then the task of the puzzle is to order a given random start permutation of the 16 pieces. To do so, the user is allowed to transposes 0 with a neighboring piece. Since every step changes the signature $s$ of the permutation and changes the taxi-metric distance $d$ of 0 to the end position by 1, only situations with even $s + d$ can be reached. It was **Sam Loyd** who suggested to start with an impossible solution and as an evil plot to offer 1000 dollars for a solution. The 15 puzzle group has 16!/2 elements and the "god number" is between 152 and 208. The **Rubik cube** is an other famous puzzle, which is a group. Exactly 100 years after the invention of the 15 puzzle, the Rubik puzzle was introduced in 1974. Its still popular and the world record is to have it solved in 5.55 seconds. All Cubes 2x2x2 to 7x7x7 in a row have been solved in a total time of 6 minutes. For the 3x3x3 cube, the **God number** is now known to be 20: one can always solve it in 20 or less moves.

Many puzzles are groups.

A small Rubik type game is the "floppy", which is a third of the Rubik and which has only 192 elements. An other example is the **Meffert's great challenge**. Probably the simplest example of a Rubik type puzzle is the **pyramorphix**. It is a puzzle based on the tetrahedron. Its group has only 24 elements. It is the group of all possible permutations of the 4 elements. It is the same group as the group of all reflection and rotation symmetries of the cube in three dimensions and also is relevant when understanding the solutions to the quartic equation discussed at the beginning. The circle is closed.

**E-320: Teaching Math with a Historical Perspective** **Oliver Knill, 2010-2018**

# Lecture 6: Calculus

Calculus generalizes the process of **taking differences** and **taking sums**. Differences measure **change**, sums explore how quantities **accumulate**. The procedure of taking differences has a limit called **derivative**. The activity of taking sums leads to the **integral**. Sum and difference are dual to each other and related in an intimate way. In this lecture, we look first at a simple set-up, where functions are evaluated on integers and where we do not take any limits.

Several dozen thousand years ago, numbers were represented by units like $1, 1, 1, 1, 1, 1, \dots$. The units were carved into sticks or bones like the **Ishango bone** It took thousands of years until numbers were represented with symbols like $0, 1, 2, 3, 4, \dots$. Using the modern concept of function, we can say $f(0) = 0, f(1) = 1, f(2) = 2, f(3) = 3$ and mean that the **function** $f$ assigns to an input like 1001 an output like $f(1001) = 1001$. Now look at $Df(n) = f(n+1) - f(n)$, the **difference**. We see that $Df(n) = 1$ for all $n$. We can also formalize the summation process. If $g(n) = 1$ is the constant 1 function, then then $Sg(n) = g(0) + g(1) + \cdots + g(n-1) = 1 + 1 + \cdots + 1 = n$. We see that $Df = g$ and $Sg = f$. If we start with $f(n) = n$ and apply **summation** on that function Then $Sf(n) = f(0) + f(1) + f(2) + \cdots + f(n-1)$ leading to the values $0, 1, 3, 6, 10, 15, 21, \dots$. The new function $g = Sf$ satisfies $g(1) = 1, g(2) = 3, g(2) = 6$, etc. The values are called the **triangular numbers**. From $g$ we can get back $f$ by taking difference: $Dg(n) = g(n+1) - g(n) = f(n)$. For example $Dg(5) = g(6) - g(5) = 15 - 10 = 5$ which indeed is $f(5)$. Finding a formula for the sum $Sf(n)$ is not so easy. Can you do it? When **Karl-Friedrich Gauss** was a 9 year old school kid, his teacher, a Mr. Büttner gave him the task to sum up the first 100 numbers $1 + 2 + \cdots + 100$. Gauss found the answer immediately by pairing things up: to add up $1 + 2 + 3 + \cdots + 100$ he would write this as $(1 + 100) + (2 + 99) + \cdots + (50 + 51)$ leading to 50 terms of 101 to get for $n = 101$ the value $g(n) = n(n-1)/2 = 5050$. Taking differences again is easier $Dg(n) = n(n+1)/2 - n(n-1)/2 = n = f(n)$. If we add up he triangular numbers we compute $h = Sg$ which has the first values $0, 1, 4, 10, 20, 35, \dots..$ These are the **tetrahedral numbers** because $h(n)$ balls are needed to build a tetrahedron of side length $n$. For example, $h(4) = 20$ golf balls are needed to build a tetrahedron of side length 4. The formula which holds for $h$ is $\boxed{h(n) = n(n-1)(n-2)/6}$. Here is the fundamental theorem of calculus, which is the core of calculus:

$$Df(n) = f(n) - f(0), \qquad DSf(n) = f(n) \; .$$

Proof.

$$SDf(n) = \sum_{k=0}^{n-1} [f(k+1) - f(k)] = f(n) - f(0) \; ,$$

$$DSf(n) = [\sum_{k=0}^{n-1} f(k+1) - \sum_{k=0}^{n-1} f(k)] = f(n) \; .$$

The process of adding up numbers will lead to the **integral** $\boxed{\int_0^x f(x) \, dx}$. The process of taking differences will lead to the **derivative** $\boxed{\frac{d}{dx} f(x)}$.

The familiar notation is

$$\boxed{\int_0^x \frac{d}{dt} f(t) \, dt = f(x) - f(0), \qquad \frac{d}{dx} \int_0^x f(t) \, dt = f(x)}$$

If we define $[n]^0 = 1, [n]^1 = n, [n]^2 = n(n-1)/2, [n]^3 = n(n-1)(n-2)/6$ then $D[n] = [1], D[n]^2 = 2[n], D[n]^3 = 3[n]^2$ and in general

$$\boxed{\frac{d}{dx} [x]^n = n[x]^{n-1}}$$

The calculus you have just seen, contains the essence of single variable calculus. This core idea will become more powerful and natural if we use it together with the concept of limit.

**Problem:** The Fibonnacci sequence $1, 1, 2, 3, 5, 8, 13, 21, \dots$ satisfies the rule $f(x) = f(x-1) + f(x-2)$. For example, $f(6) = 8$. What is the function $g = Df$, if we assume $f(0) = 0$? We take the difference between successive numbers and get the sequence of numbers $0, 1, 1, 2, 3, 5, 8, ...$ which is the same sequence again. We see that $\boxed{Df(x) = f(x-1)}$.
If we take the same function $f$ but now but now compute the function $h(n) = Sf(n)$, we get the sequence $1, 2, 4, 7, 12, 20, 33, ...$. What sequence is that? **Solution:** Because $Df(x) = f(x-1)$ we have $f(x) - f(0) = SDf(x) = Sf(x-1)$ so that $Sf(x) = f(x+1) - f(1)$. Summing the Fibonnacci sequence produces the Fibonnacci sequence shifted to the left with $f(2) = 1$ is subtracted. It has been relatively easy to find the sum, because we knew what the difference operation did. This example shows: we can study differences to understand sums.
**Problem:** The function $f(n) = 2^n$ is called the **exponential function**. We have for example $f(0) = 1, f(1) = 2, f(2) = 4, \dots$. It leads to the sequence of numbers

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| n= | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | ... |
| f(n)= | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | ... |

We can verify that $f$ satisfies the equation $\boxed{Df(x) = f(x)}$. because $Df(x) = 2^{x+1} - 2^x = (2-1)2^x = 2^x$. This is an important special case of the fact that

$$\boxed{\text{The derivative of the exponential function is the exponential function itself.}}$$

The function $2^x$ is a special case of the exponential function when the Planck constant is equal to 1. We will see that the relation will hold for any $h > 0$ and also in the limit $h \to 0$, where it becomes the classical exponential function $e^x$ which plays an important role in science.
Calculus has many applications: computing areas, volumes, solving differential equations. It even has applications in arithmetic. Here is an example for illustration. It is a proof that $\pi$ is irrational The theorem is due to Johann Heinrich Lambert (1728-1777): We show here the proof by Ivan Niven is given in a book of Niven-Zuckerman-Montgomery. It originally appeared in 1947 (Ivan Niven, Bull.Amer.Math.Soc. 53 (1947),509). The proof illustrates how calculus can help to get results in arithmetic.
**Proof**. Assume $\pi = a/b$ with positive integers $a$ and $b$. For any positive integer $n$ define

$$f(x) = x^n (a - bx)^n / n! \, .$$

We have $f(x) = f(\pi - x)$ and

$$0 \leq f(x) \leq \pi^n a^n / n! (*)$$

for $0 \leq x \leq \pi$. For all $0 \leq j \leq n$, the j-th derivative of $f$ is zero at 0 and $\pi$ and for $n <= j$, the j-th derivative of $f$ is an integer at 0 and $\pi$.
The function $F(x) = f(x) - f^{(2)}(x) + f^{(4)}(x) - ... + (-1)^n f^{(2n)}(x)$ has the property that $F(0)$ and $F(\pi)$ are integers and $F + F'' = f$. Therefore, $(F'(x)\sin(x) - F(x)\cos(x))' = f \sin(x)$. By the fundamental theorem of calculus, $\int_0^\pi f(x)\sin(x)\, dx$ is an integer. Inequality (*) implies however that this integral is between 0 and 1 for large enough $n$. For such an $n$ we get a contradiction.

# Lecture 7: Set Theory and Logic

**Set theory** studies sets, the fundamental building blocks of mathematics. While **logic** describes the language of all mathematics, set theory provides the framework for additional structures like category theory. In **Cantorian set theory**, one can compute with subsets of a given set $X$ like with numbers. There are two basic operations: the **addition** $A+B$ of two sets is defined as the set of all points which are in exactly one of the sets. The **multiplication** $A \cdot B$ of two sets contains all the points which are in both sets. With the symmetric difference as addition and the intersection as multiplication, the subsets of a given set $X$ become a **ring**. This **Boolean ring** has the property $A+A=0$ and $A \cdot A = A$ for all sets. The zero element is the empty set $\emptyset = \{\}$. The additive inverse of $A$ is the complement $-A$ of $A$ in $X$. The multiplicative 1-element is the set $X$ because $X \cdot A = A$. As in the ring $\mathbb{Z}$ of integers, the addition and multiplication on sets is commutative. Multiplication does not have an inverse in general. Two sets $A, B$ have the **same cardinality**, if there exists a one-to-one map from $A$ to $B$. For finite sets, this means that they have the same number of elements. Sets which do not have finitely many elements are called **infinite**. Do all sets with infinitely many elements have the same cardinality? The integers $\mathbb{Z}$ and the natural numbers $\mathbb{N}$ for example are infinite sets which have the same cardinality: the map $f(2n) = n, f(2n+1) = -n$ establishes a bijection between $\mathbb{N}$ and $\mathbb{Z}$. Also the rational numbers $\mathbb{Q}$ have the same cardinality than $\mathbb{N}$. Associate a fraction $p/q$ with a point $(p,q)$ in the plane. Now cut out the column $q=0$ and run the **Ulam spiral** on the modified plane. This provides a numbering of the rationals. Sets which can be counted are called of cardinality $\aleph_0$. Does an interval have the same cardinality than the reals? Even so an interval like $I = (-\pi/2, \pi/2)$ has finite length, one can bijectively map it to $\mathbb{R}$ with the tan function as $\tan : I \to \mathbb{R}$ is bijective. Similarly, one can see that any two intervals of positive length have the same cardinality. It was a great moment of mathematics, when **Georg Cantor** realized in 1874 that the interval $(0,1)$ does not have the same cardinality than the natural numbers. His argument is ingenious: assume, we could count the points $a_1, a_2, \dots$. If $0.a_{i1}a_{i2}a_{i3}...$ is the **decimal expansion** of $a_i$, define the real number $b = 0.b_1b_2b_3...$, where $b_i = a_{ii} + 1 \bmod 10$. Because this number $b$ does not agree at the first decimal place with $a_1$, at the second place with $a_2$ and so on, the number $b$ does not appear in that enumeration of all reals. It has positive distance at least $10^{-i}$ from the $i$'th number (and any representation of the number by a decimal expansion which is equivalent). This is a contradiction. The new cardinality, the **continuum** is also denoted $\aleph_1$. The reals are **uncountable**. This gives elegant proofs like the existence of **transcendental number**, numbers which are not algebraic, meaning that they are not the root of any polynomial with integer coefficients: algebraic numbers can be counted. Similarly as one can establish a bijection between the natural numbers $\mathbb{N}$ and the integers $\mathbb{Z}$, there is a bijection $f$ between the interval $I$ and the unit square: if $x = 0.x_1x_2x_3\dots$ is the decimal expansion of $x$ then $f(x) = (0.x_1x_3x_5\dots, 0.x_2x_4x_6\dots)$ is the bijection. Are there cardinalities larger than $\aleph_1$? Cantor answered also this question. He showed that for an infinite set, the set of all subsets has a larger cardinality than the set itself. How does one see this? Assume there is a bijection $x \to A(x)$ which maps each point to a set $A(x)$. Now look at the set $B = \{x \mid x \notin A(x)\ \}$ and let $b$ be the point in $X$ which corresponds to $B$. If $y \in B$, then $y \notin B(x)$. On the other hand, if $y \notin B$, then $y \in B$. The set $B$ does appear in the "enumeration" $x \to A(x)$ of all sets. The set of all subsets of $N$ has the same cardinality than the continuum: $A \to \sum_{j \in A} 1/2^j$ provides a map from $P(N)$ to $[0,1]$. The set of all **finite subsets** of $N$ however can be counted. The set of all subsets of the real numbers has cardinality $\aleph_2$, etc. Is there a cardinality between $\aleph_0$ and $\aleph_1$? In other words, is there a set which can not be counted and which is strictly smaller than the continuum in the sense that one can not find a bijection between it and $R$? This was the first of the 23 problems posed by Hilbert in 1900. The answer is surprising: one has a choice. One can accept either the "yes" or the "no" as a new axiom. In both cases, Mathematics is still fine. The nonexistence of a cardinality between $\aleph_0$ and $\aleph_1$ is called the **continuum hypothesis** and is usually abbreviated CH. It is independent of the other axioms making up mathematics. This was the work of **Kurt Gödel** in 1940 and **Paul Cohen** in 1963. The story of exploring the consistency and completeness of axiom systems of all of mathematics is exciting. Euclid axiomatized geometry, Hilbert's program was more ambitious. He aimed at a set of axiom systems for all of mathematics. The challenge to prove Euclid's 5'th postulate is paralleled by the quest to prove the CH. But the later is much more fundamental because it deals with **all of mathematics** and not only with some geometric space. Here are the **Zermelo-Frenkel Axioms** (ZFC) including the Axiom of choice (C) as established by **Ernst Zermelo** in 1908 and **Adolf Fraenkel** and **Thoral Skolem** in 1922.

| | |
|---|---|
| **Extension** | If two sets have the same elements, they are the same. |
| **Image** | Given a function and a set, then the image of the function is a set too. |
| **Pairing** | For any two sets, there exists a set which contains both sets. |
| **Property** | For any property, there exists a set for which each element has the property. |
| **Union** | Given a set of sets, there exists a set which is the union of these sets. |
| **Power** | Given a set, there exists the set of all subsets of this set. |
| **Infinity** | There exists an infinite set. |
| **Regularity** | Every nonempty set has an element which has no intersection with the set. |
| **Choice** | Any set of nonempty sets leads to a set which contains an element from each. |

There are other systems like ETCS, which is the **elementary theory of the category of sets**. In category theory, not the sets but the categories are the building blocks. Categories do not form a set in general. It elegantly avoids the Russel paradox too. The **axiom of choice (C)** has a nonconstructive nature which can lead to seemingly paradoxical results like the **Banach Tarski paradox**: one can cut the unit ball into 5 pieces, rotate and translate the pieces to assemble two identical balls of the same size than the original ball. Gödel and Cohen showed that the axiom of choice is logically independent of the other axioms ZF. Other axioms in ZF have been shown to be independent, like the **axiom of infinity**. A **finitist** would refute this axiom and work without it. It is surprising what one can do with finite sets. The **axiom of regularity** excludes Russellian sets like the set $X$ of all sets which do not contain themselves. The **Russell paradox** is: Does $X$ contain $X$? It is popularized as the **Barber riddle**: a barber in a town only shaves the people who do not shave themselves. Does the barber shave himself? **Gödels theorems** of 1931 deal with **mathematical theories** which are strong enough to do basic arithmetic in them.

**First incompleteness theorem:**
In any theory there are true statements which can not be proved within the theory.

**Second incompleteness theorem:**
In any theory, the consistency of the theory can not be proven within the theory.

The proof uses an encoding of mathematical sentences which allows to state liar paradoxical statement "this sentence can not be proved". While the later is an odd recreational entertainment gag, it is the core for a theorem which makes striking statements about mathematics. These theorems are not limitations of mathematics; they illustrate its infiniteness. How awful if one could build axiom system and enumerate mechanically all possible truths from it.

**E-320: Teaching Math with a Historical Perspective** **Oliver Knill, 2010-2018**

# Lecture 8: Probability theory

**Probability theory** is the science of chance. It starts with **combinatorics** and leads to a theory of **stochastic processes**. Historically, probability theory initiated from gambling problems as in **Girolamo Cardano's** gamblers manual in the 16th century. A great moment of mathematics occurred, when **Blaise Pascal** and **Pierre Fermat** jointly laid a foundation of mathematical probability theory.
It took a while to formalize "randomness" precisely. Here is the setup as which it had been put forward by **Andrey Kolmogorov**: all possible experiments of a situation are modeled by a set $\Omega$, the "laboratory". A measurable subset of experiments is called an "event". Measurements are done by real-valued functions $X$. These functions are called **random variables** and are used to **observe the laboratory**.
As an example, let us model the process of throwing a coin 5 times. An experiment is a word like *httht*, where $h$ stands for "head" and $t$ represents "tail". The laboratory consists of all such 32 words. We could look for example at the event $A$ that the first two coin tosses are tail. It is the set $A = \{ttttt, tttth, tttht, ttthh, tthtt, tthth, tthht, tthhh\}$. We could look at the random variable which assigns to a word the number of heads. For every experiment, we get a value, like for example, $X[tthht] = 2$.
In order to make statements about randomness, the concept of a **probability measure** is needed. This is a function $P$ from the set of all events to the interval $[0, 1]$. It should have the property that $P[\Omega] = 1$ and $P[A_1 \cup A_2 \cup \cdots] = P[A_1] + P[A_2] + \cdots$, if $A_i$ is a sequence of disjoint events.
The most natural probability measure on a finite set $\Omega$ is $P[A] = \|A\|/\|\Omega\|$, where $\|A\|$ stands for the number of elements in $A$. It is the "number of good cases" divided by the "number of all cases". For example, to count the probability of the event $A$ that we throw 3 heads during the 5 coin tosses, we have $|A| = 10$ possibilities. Since the entire laboratory has $|\Omega| = 32$ possibilities, the probability of the event is $10/32$. In order to study these probabilities, one needs **combinatorics**:

| **How many ways are there to:** | **The answer is:** |
|---|---|
| rearrange or permute $n$ elements | $n! = n(n-1)...2 \cdot 1$ |
| choose $k$ from $n$ with repetitions | $n^k$ |
| pick $k$ from $n$ if order matters | $\frac{n!}{(n-k)!}$ |
| pick $k$ from $n$ with order irrelevant | $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ |

The **expectation** of a random variable $E[X]$ is defined as the sum $m = \sum_{\omega \in \Omega} X(\omega) P[\{\omega\}]$. In our coin toss experiment, this is $5/2$. The **variance** of $X$ is the expectation of $(X - m)^2$. In our coin experiments, it is $5/4$. The square root of the variance is the **standard deviation**. This is the expected deviation from the mean. An event happens **almost surely** if the event has probability 1.
An important case of a random variable is $X(\omega) = \omega$ on $\Omega = R$ equipped with probability $P[A] = \int_A \frac{1}{\sqrt{\pi}} e^{-x^2} \, dx$, the **standard normal distribution**. Analyzed first by **Abraham de Moivre** in 1733, it was studied by **Carl Friedrich Gauss** in 1807 and therefore also called **Gaussian distribution**.
Two random variables $X, Y$ are called **uncorrelated**, if $E[XY] = E[X] \cdot E[Y]$. If for any functions $f, g$ also $f(X)$ and $g(Y)$ are uncorrelated, then $X, Y$ are called **independent**. Two random variables are said to have the same distribution, if for any $a < b$, the events $\{a \leq X \leq b\}$ and $\{a \leq Y \leq b\}$ are independent. If $X, Y$ are uncorrelated, then the relation $\text{Var}[X] + \text{Var}[Y] = \text{Var}[X + Y]$ holds which is just **Pythagoras theorem**, because uncorrelated can be understood geometrically: $X - E[X]$ and $Y - E[Y]$ are orthogonal. A common problem is to study the sum of independent random variables $X_n$ with identical distribution. One abbreviates this IID. Here are the three most important theorems which we formulate in the case, where all random variables are assumed to have expectatation 0 and standard deviation 1. Let $S_n = X_1 + ... + X_n$ be the $n$'th sum of the

IID random variables. It is also called a **random walk**.

LLN **Law of Large Numbers** assures that $S_n/n$ converges to 0.

CLT **Central Limit Theorem**:$S_n/\sqrt{n}$ approaches the Gaussian distribution.

LIL **Law of Iterated Logarithm:** $S_n/\sqrt{2n \log\log(n)}$ accumulates in $[-1,1]$.

The LLN shows that one can find out about the expectation by averaging experiments. The CLT explains why one sees the standard normal distribution so often. The LIL finally gives us a precise estimate how fast $S_n$ grows. Things become interesting if the random variables are no more independent. Generalizing LLN,CLT,LIL to such situations is part of ongoing research.

Here are two open questions in probability theory:

Are numbers like $\pi, e, \sqrt{2}$ **normal**: do all digits appear with the same frequency?
What growth rates $\Lambda_n$ can occur in $S_n/\Lambda_n$ having limsup 1 and liminf $-1$?

For the second question, there are examples for $\Lambda_n = 1, \lambda_n = \log(n)$ and of course $\lambda_n = \sqrt{n \log\log(n)}$ from LIL if the random variables are independent. Examples of random variables which are not independent are $X_n = \cos(n\sqrt{2})$.

**Statistics** is the science of modeling random events in a probabilistic setup. Given data points, we want to find a **model** which fits the data best. This allows to **understand the past**, **predict the future** or **discover laws of nature**. The most common task is to find the **mean** and the **standard deviation** of some data. The mean is also called the **average** and given by $m = \frac{1}{n}\sum_{k=1}^n x_k$. The variance is $\sigma^2 = \frac{1}{n}\sum_{k=1}^n (x_k - m)^2$ with standard deviation $\sigma$.

A sequence of random variables $X_n$ define a so called **stochastic process**. Continuous versions of such processes are where $X_t$ is a curve of random random variables. An important example is **Brownian motion**, which is a model of a random particles.

Besides gambling and analyzing data, also **physics** was an important motivator to develop probability theory. An example is statistical mechanics, where the laws of nature are studied with probabilistic methods. A famous physical law is **Ludwig Boltzmann's** relation $S = k\log(W)$ for entropy, a formula which decorates Boltzmann's tombstone. The **entropy** of a probability measure $P[\{k\}] = p_k$ on a finite set $\{1,...,n\}$ is defined as $S = -\sum_{i=1}^n p_i \log(p_i)$. Today, we would reformulate Boltzmann's law and say that it is the expectation $S = E[\log(W)]$ of the logarithm of the "Wahrscheinlichkeit" random variable $W(i) = 1/p_i$ on $\Omega = \{1,...,n\}$. Entropy is important because nature tries to maximize it

# Lecture 9: Topology

**Topology** studies properties of geometric objects which do not change under continuous reversible deformations. In topology, a coffee cup with a single handle is the same as a doughnut. One can deform one into the other without punching any holes in it or ripping it apart. Similarly, a plate and a croissant are the same. But a croissant is not equivalent to a doughnut. On a doughnut, there are closed curves which can not be pulled together to a point. For a topologist the letters $O$ and $P$ are the equivalent but different from the letter $B$. The mathematical setup is beautiful: a **topological space** is a set $X$ with a set $O$ of subsets of $X$ containing both $\emptyset$ and $X$ such that finite intersections and arbitrary unions in $O$ are in $O$. Sets in $O$ are called **open sets** and $O$ is called a **topology**. The complement of an open set is called **closed**. Examples of topologies are the **trivial topology** $O = \{\emptyset, X\}$, where no open sets besides the empty set and $X$ exist or the **discrete topology** $O = \{A \mid A \subset X\}$, where every subset is open. But these are in general not interesting. An important example on the plane $X$ is the collection $O$ of sets $U$ in the plane $X$ for which every point is the center of a small disc still contained in $U$. A special class of topological spaces are **metric spaces**, where a set $X$ is equipped with a **distance function** $d(x, y) = d(y, x) \geq 0$ which satisfies the **triangle inequality** $d(x, y) + d(y, z) \geq d(x, z)$ and for which $d(x, y) = 0$ if and only if $x = y$. A set $U$ in a metric space is open if to every $x$ in $U$, there is a **ball** $B_r(x) = \{y|d(x, y) < r\}$ of positive radius $r$ contained in $U$. Metric spaces are topological spaces but not vice versa: the trivial topology for example is not in general. For doing **calculus** on a topological space $X$, each point has a neighborhood called **chart** which is topologically equivalent to a disc in Euclidean space. Finitely many neighborhoods covering $X$ form an **atlas** of $X$. If the charts are glued together with identification maps on the intersection one obtains a **manifold**. Two dimensional examples are the **sphere**, the **torus**, the projective plane or the **Klein bottle**. Topological spaces $X, Y$ are called **homeomorphic** meaning "topologically equivalent" if there is an invertible map from $X$ to $Y$ such that this map induces an invertible map on the corresponding topologies. How can one decide whether two spaces are equivalent in this sense? The surface of the coffee cup for example is equivalent in this sense to the surface of a doughnut but it is not equivalent to the surface of a sphere. Many properties of geometric spaces can be understood by discretizing it like with a graph. A graph is a finite collection of vertices $V$ together with a finite set of edges $E$, where each edge connects two points in $V$. For example, the set $V$ of cities in the US where the edges are pairs of cities connected by a street is a graph. The **Königsberg bridge problem** was a trigger puzzle for the study of graph theory. **Polyhedra** were an other start in graph theory. It study is loosely related to the analysis of surfaces. The reason is that one can see polyhedra as discrete versions of surfaces. In computer graphics for example, surfaces are rendered as finite graphs, using triangularizations. The **Euler characteristic** of a convex polyhedron is a remarkable topological invariant. It is $V - E + F = 2$, where $V$ is the number of vertices, $E$ the number of edges and $F$ the number of **faces**. This number is equal to 2 for connected polyhedra in which every closed loop can be pulled together to a point. This formula for the Euler characteristic is also called **Euler's gem**. It comes with a rich history. **René Descartes** stumbled upon it and written it down in a secret notebook. It was Leonard Euler in 1752 was the first to proved the formula for convex polyhedra. A convex polyhedron is called a **Platonic solid**, if all vertices are on the unit sphere, all edges have the same length and all faces are congruent polygons. A theorem of **Theaetetus** states that there are only five Platonic solids: [Proof: Assume the faces are regular $n$-gons and $m$ of them meet at each vertex. Beside the Euler relation $V + E + F = 2$, a polyhedron also satisfies the relations $nF = 2E$ and $mV = 2E$ which come from counting vertices or edges in different ways. This gives $2E/m - E + 2E/n = 2$ or $1/n + 1/m = 1/E + 1/2$. From $n \geq 3$ and $m \geq 3$ we see that it is impossible that both $m$ and $n$ are larger than 3. There are now nly two possibilities: either $n = 3$ or $m = 3$. In the case $n = 3$ we have $m = 3, 4, 5$ in the case $m = 3$ we have $n = 3, 4, 5$. The five possibilities $(3, 3), (3, 4), (3, 5), (4, 3), (5, 3)$

represent the five Platonic solids.] The pairs $(n, m)$ are called the **Schläfly symbol** of the polyhedron:

| Name | V | E | F | V-E+F | Schläfli |
|---|---|---|---|---|---|
| tetrahedron | 4 | 6 | 4 | 2 | $\{3,3\}$ |
| hexahedron | 8 | 12 | 6 | 2 | $\{4,3\}$ |
| octahedron | 6 | 12 | 8 | 2 | $\{3,4\}$ |

| Name | V | E | F | V-E+F | Schläfli |
|---|---|---|---|---|---|
| dodecahedron | 20 | 30 | 12 | 2 | $\{5,3\}$ |
| icosahedron | 12 | 30 | 20 | 2 | $\{3,5\}$ |

The Greeks proceeded geometrically: Euclid showed in the "Elements" that each vertex can have either 3,4 or 5 equilateral triangles attached, 3 squares or 3 regular pentagons. (6 triangles, 4 squares or 4 pentagons would lead to a total angle which is too large because each corner must have at least 3 different edges). **Simon Antoine-Jean L'Huilier** refined in 1813 Euler's formula to situations with holes: $V - E + F = 2 - 2g$ , where $g$ is the number of holes. For a doughnut it is $V - E + F = 0$. Cauchy first proved that there are 4 non-convex regular **Kepler-Poinsot** polyhedra.

| Name | V | E | F | V-E+F | Schläfli |
|---|---|---|---|---|---|
| small stellated dodecahedron | 12 | 30 | 12 | -6 | $\{5/2,5\}$ |
| great dodecahedron | 12 | 30 | 12 | -6 | $\{5,5/2\}$ |
| great stellated dodecahedron | 20 | 30 | 12 | 2 | $\{5/2,3\}$ |
| great icosahedron | 12 | 30 | 20 | 2 | $\{3,5/2\}$ |

If two different face types are allowed but each vertex still look the same, one obtains 13 **semi-regular polyhedra.** They were first studied by **Archimedes** in 287 BC. Since his work is lost, **Johannes Kepler** is considered the first since antiquity to describe all of them them in his "Harmonices Mundi". The **Euler characteristic** for surfaces is $\chi = 2-2g$ where $g$ is the number of holes. The computation can be done by triangulating the surface. The Euler characteristic characterizes smooth compact surfaces if they are orientable. A non-orientable surface, the **Klein bottle** can be obtained by gluing ends of the Möbius strip. Classifying higher dimensional manifolds is more difficult and finding good invariants is part of modern research. Higher analogues of polyhedra are called **polytopes** (Alicia Boole Stott). **Regular polytopes** are the analogue of the Platonic solids in higher dimensions. Examples:

| dimension | name | Schläfli symbols |
|---|---|---|
| 2: | Regular polygons | $\{3\},\{4\},\{5\},...$ |
| 3: | Platonic solids | $\{3,3\},\{3,4\},\{3,5\},\{4,3\},\{5,3\}$ |
| 4: | Regular 4D polytopes | $\{3,3,3\},\{4,3,3\},\{3,3,4\},\{3,4,3\},\{5,3,3\},\{3,3,5\}$ |
| $\geq 5$: | Regular polytopes | $\{3,3,3,\dots,3\},\{4,3,3,\dots,3\},\{3,3,3,\dots,3,4\}$ |

**Ludwig SchlÏafly** saw in 1852 exactly six convex regular convex 4-polytopes or **polychora**, where "Choros" is Greek for "space". Schlaefli's polyhedral formula is $V - E + F - C = 0$ holds, where $C$ is the number of 3-dimensional **chambers**. In dimensions 5 and higher, there are only 3 types of polytopes: the higher dimensional analogues of the tetrahedron, octahedron and the cube. A general formula $\sum_{k=0}^{d-1}(-1)^k v_k = 1-(-1)^d$ gives the Euler characteristic of a convex polytop in $d$ dimensions with $k$-dimensional parts $v_k$.

# Lecture 10: Analysis

**Analysis** is a science of measure and optimization. As a rather diverse collection of mathematical fields, it contains **real and complex analysis**, **functional analysis**, **harmonic analysis** and **calculus of variations**. Analysis has relations to calculus, geometry, topology, probability theory and dynamical systems. We focus here mostly on "the geometry of fractals" which can be seen as part of dimension theory. Examples are Julia sets which belong to the subfield of "complex analysis" of "dynamical systems". "Calculus of variations" is illustrated by the Kakeya needle set in "geometric measure theory", "Fourier analysis" appears when looking at functions which have fractal graphs, "spectral theory" as part of functional analysis is represented by the "Hofstadter butterfly". We somehow describe the topic using "pop icons".

A **fractal** is a set with non-integer dimension. An example is the **Cantor set**, as discovered in 1875 by Henry Smith. Start with the unit interval. Cut the middle third, then cut the middle third from both parts then the middle parts of the four parts etc. The limiting set is the Cantor set. The mathematical theory of fractals belongs to **measure theory** and can also be thought of a playground for real analysis or topology. The term **fractal** had been introduced by Benoit Mandelbrot in 1975. Dimension can be defined in different ways. The simplest is the **box counting definition** which works for most household fractals: if we need $n$ squares of length $r$ to cover a set, then $d = -\log(n)/\log(r)$ converges to the dimension of the set with $r \to 0$. A curve of length $L$ for example needs $L/r$ squares of length $r$ so that its dimension is 1. A region of area $A$ needs $A/r^2$ squares of length $r$ to be covered and its dimension is 2. The Cantor set needs to be covered with $n = 2^m$ squares of length $r = 1/3^m$. Its dimension is $-\log(n)/\log(r) = -m\log(2)/(m\log(1/3)) = \log(2)/\log(3)$. Examples of fractals are the graph of the Weierstrass function 1872, the Koch snowflak (1904), the Sierpinski carpet (1915) or the Menger sponge (1926).
**Complex analysis** extends calculus to the complex. It deals with functions $f(z)$ defined in the complex plane. Integration is done along paths. Complex analysis completes the understanding about functions. It also provides more examples of fractals by iterating functions like the **quadratic map** $f(z) = z^2 + c$:
One has already iterated functions before like the Newton method (1879). The Julia sets were introduced in 1918, the Mandelbrot set in 1978 and the Mandelbar set in 1989. Particularly famous are the **Douady rabbit** and the **dragon**, the **dendrite**, the **airplane**. **Calculus of variations** is calculus in infinite dimensions. Taking derivatives is called taking "variations". Historically, it started with the problem to find the curve of fastest fall leading to the **Brachistochrone** curve $\vec{r}(t) = (t - \sin(t), 1 - \cos(t))$. In calculus, we find maxima and minima of functions. In calculus of variations, we extremize on much larger spaces. Here are examples of problems:

| | |
|---|---|
| Brachistochrone | 1696 |
| Minimal surface | 1760 |
| Geodesics | 1830 |
| Isoperimetric problem | 1838 |
| Kakeya Needle problem | 1917 |

**Fourier theory** decomposes a function into basic components of various frequencies $f(x) = a_1 \sin(x) + a_2 \sin(2x) + a_3 \sin(3x) + \cdots$. The numbers $a_i$ are called the **Fourier coefficients**. Our ear does such a decomposition, when we listen to music. By distinguish different frequencies, our ear produces a Fourier analysis.

| | |
|---|---|
| Fourier series | 1729 |
| Fourier transform (FT) | 1811 |
| Discrete FT | Gauss? |
| Wavelet transform | 1930 |

The Weierstrass function mentioned above is given as a series $\sum_n a^n \cos(\pi b^n x)$ with $0 < a < 1, ab > 1 + 3\pi/2$. The dimension of its graph is believed to be $2 + \log(a)/\log(b)$ but no rigorous computation of the dimension was done yet. **Spectral theory** analyzes linear maps $L$. The **spectrum** are the real numbers $E$ such that $L - E$ is not invertible. A Hollywood celebrity among all linear maps is the **almost Matthieu operator** $L(x)_n = x_{n+1} + x_{n-1} + (2 - 2\cos(cn))x_n$: if we draw the spectrum for for each $c$, we see the **Hofstadter butterfly**. For fixed $c$ the map describes the behavior of an electron in an almost periodic crystal. An other famous system is the **quantum harmonic oscillator**, $L(f) = f''(x) + f(x)$, the **vibrating drum** $L(f) = f_{xx} + f_{yy}$, where $f$ is the amplitude of the drum and $f = 0$ on the boundary of the drum.

| | |
|---|---|
| Hydrogen atom | 1914 |
| Hofstadter butterfly | 1976 |
| Harmonic oscillator | 1900 |
| Vibrating drum | 1680 |

All these examples in analysis look unrelated at first. Fractal geometry ties many of them together: spectra are often fractals, minimal configurations have fractal nature, like in solid state physics or in **diffusion limited aggregation** or in other critical phenomena like **percolation** phenomena, **cracks** in solids or the formation of **lighting bolts** In Hamiltonian mechanics, minimal energy configurations are often fractals like **Mather theory**. And solutions to minimizing problems lead to fractals in a natural way like when you have the task to turn around a needle on a table by 180 degrees and minimize the area swept out by the needle. The minimal turn leads to a Kakaya set, which is a fractal. Finally, lets mention some unsolved problems in analysis: does the **Riemann zeta function** $f(z) = \sum_{n=1}^{\infty} 1/n^z$ have all nontrivial roots on the axis $Re(z) = 1/2$? This question is called the **Riemann hypothesis** and is the most important open problem in mathematics. It is an example of a question in **analytic number theory** which also illustrates how analysis has entered into number theory. Some mathematicians think that spectral theory might solve it. Also the Mandelbrot set $M$ is not understood yet: the "holy grail" in the field of complex dynamics is the problem whether it $M$ is locally connected. From the Hofstadter butterfly one knows that it has measure zero. What is its dimension? An other open question in spectral theory is the "can one hear the sound of a drum" problem which asks whether there are two convex drums which are not congruent but which have the same spectrum. In the area of calculus of variations, just one problem: how long is the shortest curve in space such that its convex hull (the union of all possible connections between two points on the curve) contains the unit ball.

**E-320: Teaching Math with a Historical Perspective** **Oliver Knill, 2010-2018**

# Lecture 11: Cryptography

**Cryptography** is the theory of **codes**. Two important aspects of the field are the **encryption** rsp. **decryption** of information and **error correction**. Both are crucial in daily life. When getting access to a computer, viewing a bank statement or when taking money from the ATM, encryption algorithms are used. When phoning, surfing the web, accessing data on a computer or listening to music, error correction algorithms are used. Since our lives have become more and more digital: music, movies, books, journals, finance, transportation, medicine, and communication have become digital, we rely on strong error correction to avoid errors and encryption to assure things can not be tempered with. Without error correction, airplanes would crash: small errors in the memory of a computer would produce glitches in the navigation and control program. In a computer memory every hour a couple of bits are altered, for example by cosmic rays. Error correction assures that this gets fixed. Without error correction music would sound like a 1920 gramophone record. Without encryption, everybody could intrude electronic banks and transfer money. Medical history shared with your doctor would all be public. Before the digital age, error correction was assured by extremely redundant information storage. Writing a letter on a piece of paper displaces billions of billions of molecules in ink. Now, changing any single bit could give a letter a different meaning. Before the digital age, information was kept in well guarded safes which were physically difficult to penetrate. Now, information is locked up in computers which are connected to other computers. Vaults, money or voting ballots are secured by mathematical algorithms which assure that information can only be accessed by authorized users. Also life needs error correction: information in the genome is stored in a **genetic code**, where a error correction makes sure that life can survive. A cosmic ray hitting the skin changes the DNA of a cell, but in general this is harmless. Only a larger amount of radiation can render cells cancerous.

How can an encryption algorithm be safe? One possibility is to invent a new method and keep it secret. An other is to use a well known encryption method and rely on the **difficulty of mathematical computation tasks** to assure that the method is safe. History has shown that the first method is unreliable. Systems which rely on "security through obfuscation" usually do not last. The reason is that it is tough to keep a method secret if the encryption tool is distributed. Reverse engineering of the method is often possible, for example using plain text attacks. Given a map $T$, a third party can compute pairs $x, T(x)$ and by choosing specific texts figure out what happens.

The **Caesar cypher** permutes the letters of the alphabet. We can for example replace every letter $A$ with $B$, every letter $B$ with $C$ and so on until finally $Z$ is replaced with $A$. The word "Mathematics" becomes so encrypted as "Nbuifnbujdt". Caesar would shift the letters by 3. The right shift just discussed was used by his Nephew Augustus. **Rot13** shifts by 13, and **Atbash cypher** reflects the alphabet, switch $A$ with $Z$, $B$ with $Y$ etc. The last two examples are involutive: encryption is decryption. More general cyphers are obtained by permuting the alphabet. Because of $26! = 403291461126605635584000000 \sim 10^{27}$ permutations, it appears first that a brute force attack is not possible. But Cesar cyphers can be cracked very quickly using statistical analysis. If we know the frequency with which letters appear and match the frequency of a text we can figure out which letter was replaced with which. The **Trithemius cypher** prevents this simple analysis by changing the permutation in each step. It is called a polyalphabetic substitution cypher. Instead of a simple permutation, there are many permutations. After transcoding a letter, we also change the key. Lets take a simple example. Rotate for the first letter the alphabet by 1, for the second letter, the alphabet by 2, for the third letter, the alphabet by 3 etc. The word "Mathematics" becomes now "Ncwljshbrmd". Note that the second "a" has been translated to something different than $a$. A frequency analysis is now more difficult. The **Viginaire cypher** adds even more complexity: instead of shifting the alphabet by 1, we can take a key like "BCNZ", then shift the first letter by 1, the second letter by 3 the third letter by 13, the fourth letter by 25 the shift the 5th letter by

1 again. While this cypher remained unbroken for long, a more sophisticated frequency analysis which involves first finding the length of the key makes the cypher breakable. With the emergence of computers, even more sophisticated versions like the German **enigma** had no chance.

**Diffie-Hellman key exchange** allows Ana and Bob want to agree on a secret key over a public channel. The two palindromic friends agree on a prime number $p$ and a base $a$. This information can be exchanged over an open channel. Ana chooses now a secret number $x$ and sends $X = a^x$ modulo $p$ to Bob over the channel. Bob chooses a secret number $y$ and sends $Y = a^y$ modulo $p$ to Ana. Ana can compute $Y^x$ and Bob can compute $X^y$ but both are equal to $a^{xy}$. This number is their common secret. The key point is that eves dropper Eve, can not compute this number. The only information available to Eve are $X$ and $Y$, as well as the base $a$ and $p$. Eve knows that $X = a^x$ but can not determine $x$. The key difficulty in this code is the **discrete log problem**: getting $x$ from $a^x$ modulo $p$ is believed to be difficult for large $p$.

The **Rivest-Shamir-Adleman public key system** uses a **RSA public key** $(n, a)$ with an integer $n = pq$ and $a < (p-1)(q-1)$, where $p, q$ are prime. Also here, $n$ and $a$ are public. Only the factorization of $n$ is kept secret. Ana publishes this pair. Bob who wants to email Ana a message $x$, sends her $y = x^a$ mod $n$. Ana, who has computed $b$ with $ab = 1$ mod $(p-1)(q-1)$ can read the secrete email $y$ because $y^b = x^{ab} = x^{(p-1)(q-1)} = x$ mod$n$. But Eve, has no chance because the only thing Eve knows is $y$ and $(n, a)$. It is believed that without the **factorization** of $n$, it is not possible to determine $x$. The message has been transmitted securely. The core difficulty is that **taking roots** in the ring $Z_n = \{0, \dots, n-1\}$ is difficult without knowing the factorization of $n$. With a factorization, we can quickly take arbitrary roots. If we can take square roots, then we can also factor: assume we have a product $n = pq$ and we know how to take square roots of 1. If $x$ solves $x^2 = 1$ mod $n$ and $x$ is different from 1, then $x^2 - 1 = (x-1)(x+1)$ is zero modulo $n$. This means that $p$ divides $(x-1)$ or $(x+1)$. To find a factor, we can take the greatest common divisor of $n, x-1$. Take $n = 77$ for example. We are given the root 34 of 1. ( $34^2 = 1156$ has reminder 1 when divided by 34). The greatest common divisor of $34 - 1$ and 77 is 11 is a factor of 77. Similarly, the greatest common divisor of $34 + 1$ and 77 is 7 divides 77. Finding roots modulo a composite number and factoring the number is equally difficult.

| Cipher | Used for | Difficulty | Attack |
|---|---|---|---|
| Cesar | transmitting messages | many permutations | Statistics |
| Viginere | transmitting messages | many permutations | Statistics |
| Enigma | transmitting messages | no frequency analysis | Plain text |
| Diffie-Helleman | agreeing on secret key | discrete log mod p | Unsafe primes |
| RSA | electronic commerce | factoring integers | Factoring |

The simplest **error correcting code** uses 3 copies of the same information so single error can be corrected. With 3 watches for example, one watch can fail. But this basic error correcting code is not efficient. It can correct single errors by tripling the size. Its efficiency is 33 percent.

**E-320: Teaching Math with a Historical Perspective** **Oliver Knill, 2010-2018**

# Lecture 12: Dynamical systems

**Dynamical systems theory** is the science of time evolution. If time is **continuous** the evolution is defined by a **differential equation** $\dot{x} = f(x)$. If time is **discrete** then we look at the **iteration of a map** $x \to T(x)$.

The goal of the theory is to **predict the future** of the system when the present state is known. A **differential equation** is an equation of the form $d/dtx(t) = f(x(t))$, where the unknown quantity is a path $x(t)$ in some "phase space". We know the **velocity** $d/dtx(t) = \dot{x}(t)$ at all times and the initial configuration $x(0)$), we can to compute the **trajectory** $x(t)$. What happens at a future time? Does $x(t)$ stay in a bounded region or escape to infinity? Which areas of the phase space are visited and how often? Can we reach a certain part of the space when starting at a given point and if yes, when. An example of such a question is to predict, whether an asteroid located at a specific location will hit the earth or not. An other example is to predict the weather of the next week.

An examples of a dynamical systems in one dimension is the differential equation

$$x'(t) = x(t)(2 - x(t)), x(0) = 1$$

It is called the **logistic system** and describes population growth. This system has the solution $x(t) = 2e^t/(1 + e^{2t})$ as you can see by computing the left and right hand side.

A **map** is a rule which assigns to a quantity $x(t)$ a new quantity $x(t+1) = T(x(t))$. The state $x(t)$ of the system determines the situation $x(t+1)$ at time $t+1$. An example is is the **Ulam map** $T(x) = 4x(1-x)$ on the interval $[0, 1]$. This is an example, where we have no idea what happens after a few hundred iterates even if we would know the initial position with the accuracy of the Planck scale.

Dynamical system theory has applications all fields of mathematics. It can be used to find roots of equations like for

$$T(x) = x - f(x)/f'(x) \ .$$

A system of number theoretical nature is the **Collatz map**

$$T(x) = \frac{x}{2} \text{ (even x)}, 3x + 1 \text{ else} \ .$$

A system of geometric nature is the **Pedal map** which assigns to a triangle **the pedal triangle**.

About 100 years ago, **Henry Poincaré** was able to deal with **chaos** of low dimensional systems. While **statistical mechanics** had formalized the evolution of large systems with probabilistic methods already, the new insight was that simple systems like a **three body problem** or a **billiard map** can produce very complicated motion. It was Poincaré who saw that even for such low dimensional and completely deterministic systems, random motion can emerge. While physisists have dealt with chaos earlier by assuming it or artificially feeding it into equations like the **Boltzmann equation**, the occurrence of stochastic motion in geodesic flows or billiards or restricted three body problems was a surprise. These findings needed half a century to sink in and only with the emergence of computers in the 1960ies, the awakening happened. Icons like Lorentz helped to popularize the findings and we owe them the **"butterfly effect"** picture: a wing of a butterfly can produce a tornado in Texas in a few weeks. The reason for this statement is that the complicated equations to simulate the weather reduce under extreme simplifications and truncations to a simple differential equation $\dot{x} = \sigma(y - x), \dot{y} = rx - y - xz, \dot{z} = xy - bz$, the **Lorenz system**. For $\sigma = 10, r = 28, b = 8/3$, Ed Lorenz discovered in 1963 an interesting long time behavior and an aperiodic "attractor". Ruelle-Takens called it a

**strange attractor**. It is a **great moment** in mathematics to realize that attractors of simple systems can become fractals on which the motion is chaotic. It suggests that such behavior is abundant. What is chaos? If a dynamical system shows **sensitive dependence on initial conditions**, we talk about **chaos**. We will experiment with the two maps $T(x) = 4x(1-x)$ and $S(x) = 4x - 4x^2$ which starting with the same initial conditions will produce different outcomes after a couple of iterations.
The sensitive dependence on initial conditions is measured by how fast the derivative $dT^n$ of the $n$'th iterate grows. The exponential growth rate $\gamma$ is called the **Lyapunov exponent**. A small error of the size $h$ will be amplified to $he^{\gamma n}$ after $n$ iterates. In the case of the Logistic map with $c = 4$, the Lyapunov exponent is $\log(2)$ and an error of $10^{-16}$ is amplified to $2^n \cdot 10^{-16}$. For time $n = 53$ already the error is of the order 1. This explains the above experiment with the different maps. The maps $T(x)$ and $S(x)$ round differently on the level $10^{-16}$. After 53 iterations, these initial fluctuation errors have grown to a macroscopic size.
Here is a famous open problem which has resisted many attempts to solve it: Show that the map $T(x, y) = (c\sin(2\pi x) + 2x - y, x)$ with $T^n(x, y) = (f_n(x, y), g_n(x, y))$ has sensitive dependence on initial conditions on a set of positive area. More precisely, verify that for $c > 2$ and all $n$ $\frac{1}{n}\int_0^1\int_0^1 \log|\partial_x f_n(x, y)|\, dxdy \geq \log(\frac{c}{2})$. The left hand side converges to the average of the Lyapunov exponents which is in this case also the **entropy** of the map. For some systems, one can compute the entropy. The logistic map with $c = 4$ for example, which is also called the **Ulam map**, has entropy $\log(2)$. The **cat map**

$$T(x, y) = (2x + y, x + y) \bmod 1$$

has positive entropy $\log|(\sqrt{5} + 3)/2|$. This is the logarithm of the larger eigenvalue of the matrix implementing $T$.
While questions about simple maps look artificial at first, the mechanisms prevail in other systems: in astronomy, when studying planetary motion or electrons in the van Allen belt, in mechanics when studying coupled pendulum or nonlinear oscillators, in fluid dynamics when studying vortex motion or turbulence, in geometry, when studying the evolution of light on a surface, the change of weather or tsunamis in the ocean. Dynamical systems theory started historically with the problem to understand the **motion of planets**. Newton realized that this is governed by a differential equation, the **n-body problem**

$$x_j''(t) = \sum_{i=1}^{n} \frac{c_{ij}(x_i - x_j)}{|x_i - x_j|^3} ,$$

where $c_{ij}$ depends on the masses and the gravitational constant. If one body is the sun and no interaction of the planets is assumed and using the common center of gravity as the origin, this reduces to the **Kepler problem** $x''(t) = -Cx/|x|^3$, where planets move on **ellipses**, the radius vector sweeps equal area in each time and the period squared is proportional to the semi-major axes cubed. A great moment in astronomy was when Kepler derived these laws empirically. An other great moment in mathematics is Newton's theoretically derivation from the differential equations.

**E-320: Teaching Math with a Historical Perspective** **Oliver Knill, 2010-2018**

# Lecture 13: Computing

**Computing** deals with algorithms and the art of programming. While the subject intersects with computer science, information technology, the theory is by nature very mathematical. But there are new aspects: computers have opened the field of **experimental mathematics** and serve now as the **laboratory** for new mathematics. Computers are not only able to **simulate** more and more of our physical world, they allow us to **explore** new worlds.

A mathematician pioneering new grounds with computer experiments does similar work than an experimental physicist. Computers have smeared the boundaries between physics and mathematics. According to Borwein and Bailey, experimental mathematics consists of:

| | |
|---|---|
| Gain insight and intuition. | Explore possible new results |
| Find patterns and relations | Suggest approaches for proofs |
| Display mathematical principles | Automate lengthy hand derivations |
| Test and falsify conjectures | Confirm already existing proofs |

When using computers to prove things, reading and verifying the computer program is part of the proof. If Goldbach's conjecture would be known to be true for all $n > 10^{18}$, the conjecture should be accepted because numerical verifications have been done until $2 \cdot 10^{18}$ until today. The first famous theorem proven with the help of a computer was the "4 color theorem" in 1976. Here are some pointers in the history of computing:

| | | | | | |
|---|---|---|---|---|---|
| 2700BC | Sumerian Abacus | 1935 | Zuse 1 programmable | 1973 | Windowed OS |
| 200BC | Chinese Abacus | 1941 | Zuse 3 | 1975 | Altair 8800 |
| 150BC | Astrolabe | 1943 | Harvard Mark I | 1976 | Cray I |
| 125BC | Antikythera | 1944 | Colossus | 1977 | Apple II |
| 1300 | Modern Abacus | 1946 | ENIAC | 1981 | Windows I |
| 1400 | Yupana | 1947 | Transistor | 1983 | IBM PC |
| 1600 | Slide rule | 1948 | Curta Gear Calculator | 1984 | Macintosh |
| 1623 | Schickard computer | 1952 | IBM 701 | 1985 | Atari |
| 1642 | Pascal Calculator | 1958 | Integrated circuit | 1988 | Next |
| 1672 | Leibniz multiplier | 1969 | Arpanet | 1989 | HTTP |
| 1801 | Punch cards | 1971 | Microchip | 1993 | Webbrowser, PDA |
| 1822 | Difference Engine | 1972 | Email | 1998 | Google |
| 1876 | Mechanical integrator | 1972 | HP-35 calculator | 2007 | iPhone |

We live in a time where technology explodes exponentially.**Moore's law** from 1965 predicted that semiconductor technology doubles in capacity and overall performance every 2 years. This has happened since. Futurologists like Ray Kurzweil conclude from this technological singularity in which artificial intelligence might take over. An important question is how to decide whether a computation is "easy" or "hard". In 1937, **Alan Turing** introduced the idea of a **Turing machine**, a theoretical model of a computer which allows to quantify complexity. It has finitely many states $S = \{s_1, ..., s_n, h\}$ and works on an tape of $0-1$ sequences. The state $h$ is the "halt" state. If it is reached, the machine stops. The machine has rules which tells what it does if it is in state $s$ and reads a letter $a$. Depending on $s$ and $a$, it writes 1 or 0 or moves the tape to the left or right and moves into a new state. Turing showed that anything we know to compute today can be computed with Turing machines. For any known machine, there is a polynomial $p$ so that a computation done in $k$ steps with that computer can be done in $p(k)$ steps on a Turing machine. What can actually be computed? Church's thesis of 1934 states that everything which can be computed can be computed with Turing machines. Similarly as in mathematics itself, there are limitations of computing. Turing's setup allowed him to enumerate all possible Turing machine and use them as input of an other machine. Denote by $TM$ the set of all pairs $(T, x)$, where $T$

is a Turing machine and $x$ is a finite input. Let $H \subset TM$ denote the set of Turing machines $(T, x)$ which halt with the tape $x$ as input. Turing looked at the decision problem: is there a machine which decides whether a given machine $(T, x)$ is in $H$ or not. An ingenious Diagonal argument of Turing shows that the answer is "no". [Proof: assume there is a machine $HALT$ which returns from the input $(T, x)$ the output $\mathrm{HALT(T,x)} = \mathrm{true}$, if $T$ halts with the input $x$ and otherwise returns $\mathrm{HALT(T,x)} = \mathrm{false}$. Turing constructs a Turing machine DIAGONAL, which does the following: 1) Read x. 2) Define Stop=HALT(x,x) 3) While Stop=True repeat Stop:=True; 4) Stop.
Now, DIAGONAL is either in H or not. If DIAGONAL is in H, then the variable Stop is true which means that the machine DIAGONAL runs for ever and DIAGONAL is not in H. But if DIAGONAL is not in H, then the variable Stop is false which means that the loop 3) is never entered and the machine stops. The machine is in H.]
Lets go back to the problem of distinguishing "easy" and "hard" problems: One calls **P** the class of decision problems that are solvable in polynomial time and **NP** the class of decision problems which can efficiently be tested if the solution is given. These categories do not depend on the computing model used. The question "N=NP?" is the most important open problem in theoretical computer science. It is one of the seven **millenium problems** and it is widely believed that $P \neq NP$. If a problem is such that every other NP problem can be reduced to it, it is called **NP-complete**. Popular games like Minesweeper or Tetris are NP-complete. If $P \neq NP$, then there is no efficient algorithm to beat the game. The intersection of NP-hard and NP is the class of NP-complete problems. An example of an NP-complete problem is the **balanced number partitioning problem**: given $n$ positive integers, divide them into two subsets $A, B$, so that the sum in $A$ and the sum in $B$ are as close as possible. A first shot: chose the largest remaining number and distribute it to alternatively to the two sets.
We all feel that it is harder to **find a solution to a problem** rather than to **verify a solution**. If $N \neq NP$ there are one way functions, functions which are easy to compute but hard to verify. For some important problems, we do not even know whether they are in NP. Examples are the **the integer factoring problem**. An efficient algorithm for the first one would have enormous consequences. Finally, lets look at some mathematical problems in artificial intelligence AI:

| | |
|---|---|
| problem solving | playing games like chess, performing algorithms, solving puzzles |
| pattern matching | speech, music, image, face, handwriting, plagiarism detection, spam |
| reconstruction | tomography, city reconstruction, body scanning |
| research | computer assisted proofs, discovering theorems, verifying proofs |
| data mining | knowledge acquisition, knowledge organization, learning |
| translation | language translation, porting applications to programming languages |
| creativity | writing poems, jokes, novels, music pieces, painting, sculpture |
| simulation | physics engines, evolution of bots, game development, aircraft design |
| inverse problems | earth quake location, oil depository, tomography |
| prediction | weather prediction, climate change, warming, epidemics, supplies |

## About this document

It should have become obvious that I'm reporting on many of these theorems as a **tourist** and not as a **local**. In some few areas I could qualify as a tour guide but hardly as a local. The references contain only parts consulted but it does not imply that I know all of that source. My own background was in dynamical systems theory and mathematical physics. Both of these subjects by nature have many connections with other branches of mathematics.
The motivation to try such a project came through teaching a course called **Math E 320** at the Harvard extension school. This math-multi-disciplinary course is part of the "math for teaching program", and tries to map out the major parts of mathematics and visit some selected placed on 12 continents.

It is wonderful to visit other places and see connections. One can learn new things, relearn old ones and marvel again about how large and diverse mathematics is but still to notice how many similarities there are between seemingly remote areas. A goal of this project is also to get back up to speed up to the level of a first year grad student (one forgets a lot of things over the years) and maybe pass the quals (with some luck).

This summer 2018 project also illustrates the challenges when trying to tour the most important mountain peaks in the mathematical landscape with limited time. Already the identification of major peaks and attaching a "height" can be challenging. Which theorems are the most important? Which are the most fundamental? Which theorems provide fertile seeds for new theorems? I recently got asked by some students what I consider the most important theorem in mathematics (my answer had been the "Atiyah-Singer theorem").

Theorems are the entities which build up mathematics. Mathematical ideas show their merit only through theorems. Theorems not only help to bring ideas to live, they in turn allow to solve problems and justify the language or theory. But not only the results alone, also the history and the connections with the mathematicians who created the results are fascinating.

The first version of this document got started in May 2018 and was posted in July 2018. Comments, suggestions or corrections are welcome. I hope to be able to extend, update and clarify it and explore also still neglected continents in the future if time permits.

It should be pretty obvious that one can hardly do justice to all mathematical fields and that much more would be needed to cover the essentials. A more serious project would be to identify a dozen theorems in each of the major MSC 2010 classification fields. This would roughly lead to a "thousand theorem" list. In some sense, this exists already: on Wikipedia, there are currently about 1000 theorems discussed. The one-document project getting closest to this project is maybe the beautiful book [255].

## 136. Document history

The first draft was posted on July 22 [202]. On July 23, a short list of theorems was made available on [203]. This document history section got started July 25-27, 2018.

- July 28: Entry 36 had been a repeated prime number theorem entry. Its alternative is now the Fredholm alternative. Also added are the Sturm theorem and Smith normal form.
- July 29: The two entries about Lidskii theorem and Radon transform are added.
- July 30: An entry about linear programming.
- July 31: An entry about random matrices.
- August 2: An entry about entropy of diffeomorphisms
- August 4: 104-108 entries: linearization, law of small numbers, Ramsey, Fractals and Poincare duality.
- August 5: 109-111 entries: Rokhlin and Lax approximation, Sobolev embedding
- August 6: 112: Whitney embedding.
- August 8: 113-114: AI and Stokes entries
- August 12: 115 and 116: Moment entry and martingale theorem
- August 13: 117 and 118: theorema egregium and Shannon theorem
- August 14: 119 mountain pass
- August 15: 120, 121,122,123 exponential sums, sphere theorem, word problem and finite simple groups
- August 16: 124, 125, 126, Rubik, Sard and Elliptic curves,
- August 17: 127, 128, 129 billiards, uniformization, Kalman filter
- August 18: 130,131 Zarisky and Poincare's last theorem
- August 19: 132, 133 Geometrization, Steinitz
- August 21: 134, 135 Hilbert-Einstein, Hall marriage

## 137. Top choice

The short list of 10 theorems mentioned in the youtube clip were:

- Fundamental theorem of arithmetic (prime factorization)
- Fundamental theorem of geometry (Pythagoras theorem)
- Fundamental theorem of logic (incompleteness theorem)
- Fundamental theorem of topology (rule of product)
- Fundamental theorem of computability (Turing computability)
- Fundamental theorem of calculus (Stokes theorem)
- Fundamental theorem of combinatorics, (pigeon hole principle)
- Fundamental theorem of analysis (spectral theorem)
- Fundamental theorem of algebra (polynomial factorization)
- Fundamental theorem of probability (central limit theorem)

Let us try to justify this shortlist. It should go without saying that similar arguments could be found for any other choice except maybe for the five classical fundamental theorems: Arithmetic, Geometry (which is undisputed Pythagoras), Calculus and Algebra, where one can hardly must argue much: except for Pythagoras, their given name already suggests that they are considered fundamental.

- **Analysis**. Why chose the spectral theorem and not say the more general **Jordan normal form theorem**? This is not an easy call but the **Jordan normal form theorem** is less simple to state and furthermore, that it does not stress the importance of **normality** giving the possibility for a **functional calculus**. Also, the spectral theorem holds in infinite dimensions for operators on Hilbert spaces. If one looks at mathematical physics for example, then it is the **functional calculus of operators** which is really made use of; the Jordan normal form theorem appears rarely in comparison. In infinite dimensions, a Jordan normal form theorem would be much more difficult as the operator $Au(n) = u(n+1)$ on $l^2(\mathbb{Z})$ is both unitary as well as a "Jordan form matrix". The spectral theorem however sails through smoothly to infinite dimensions and even applies with adaptations to **unbounded self-adjoint operators** which are important in physics. And as it is a core part of **analysis**, it is also fine to see the theorem as part of analysis. The main reason of course is that the fundamental theorem of algebra is already occupied by a theorem. One could object that "analysis" is already represented by the fundamental theorem of calculus but calculus is so important that it can represent its own field. The idea of the fundamental theorem of calculus goes beyond calculus. It is essentially a **cancellation property**, a **telescopic sum** or **Pauli principle** ($d^2 = 0$ for exterior derivatives) which makes the principle work. Calculus is the idea of an exterior derivative, the idea of cohomology, a link between algebra and geometry. One can see calculus also as a theory of "time". In some sense, the fundamental theorem of calculus also represents the field of **differential equations** and this is what "time is all about".
- **Probability**. One can ask also why to pick the **central limit theorem** and not say the **Bayes formula** or then the deeper **law of iterated logarithm**. One objection against the Bayes formula is that it is essentially a definition, like the basic arithmetic properties "commutativity, distributivity or associativity" in an algebraic structure like a ring. One does not present the identity $a + b = b + a$ for example as a fundamental theorem. Yes, the Bayes theorem has an unusual high appeal to scientists as it appears like a **magic bullet**, but for a mathematician, the statement just does not have enough beef: it is a definition, not a theorem. Not to belittle the Bayes theorem, like the notion of **entropy** or the notion of **logarithm**, it is a **genius concept**. But it is not an actual theorem, as the cleverness of the statement of Bayes lies in the **definition** and so the clarification of conditional probability theory. For the central limit theorem, it is pretty clear that it should be high up on any list of theorems, as the name suggests: it is central. But also, it actually is **stronger** than some versions of the law of large numbers. The strong law is also super seeded by Birkhoff's ergodic theorem which is much more general. One could argue to pick the **law of iterated logarithm** or some **Martingale theorem** instead but there is something appealing in the central limit theorem which goes over to other set-ups. One can formulate the central limit theorem also for random variables taking values in a compact topological group like when doing statistics with spherical data [256]. An other pitch for the central limit theorem is that it is a **fixed**

**point of a renormalization map** $X \to \overline{X+X}$ (where the right hand side is the sum of two independent copies of $X$) in the space of random variables. This map **increases entropy** and the fixed point is is a random variable whose distribution function $f$ has the **maximal entropy** $-\int_{\mathbb{R}} f(x)\log(f(x))\,dx$ among all probability density functions. The entropy principle justifies essentially all known probability density functions. Nature just likes to maximize entropy and minimize energy or more generally - in the presence of energy - to minimize the free energy.

- **Topology**. Topology is about geometric properties which do not change under continuous deformation or more generally under homotopies. Quantities which are invariant under homeomorphisms are interesting. Such quantities should add up under disjoint unions of geometries and multiply under products. The Euler characteristic is **the** prototype. Taking products is fundamental for building up Euclidean spaces (also over other fields, not only the real numbers) which locally patch up more complicated spaces. It is the essence of vector spaces that after building a basis, one has a product of Euclidean spaces. Field extensions can be seen therefore as product spaces. How does the counting principle come in? As stated, it actually is quite strong and calling it a "fundamental principle of topology" can be justified if the product of topological spaces is defined properly: if 1 is the one-point space, one can see the statement $G \times 1 = G_1$ as the **Barycentric refinement** of $G$, implying that the Euler characteristic is a Barycentric invariant and so that it is a "counting tool" which can be pushed to the continuum, to manifolds or varieties. And the compatibility with the product is the key to make it work. Counting in the form of Euler characteristic goes throughout mathematics, combinatorics, differential geometry or algebraic geometry. Riemann-Roch or Atiyah-Singer and even dynamical versions like the Lefschetz fixed point theorem (which generalizes the Brouwer fixed point theorem) or the even more general Atiyah-Bott theorem can be seen as **extending the basic counting principle**: the **Lefschetz number** $\chi(X,T)$ is a dynamical Euler characteristic which in the static case $T = Id$ reduces to the Euler characteristic $\chi(X)$. In "school mathematics", one calls the principle the "fundamental principle of counting" or "rule of product". It is put in the following way: "If we have $k$ ways to do one thing and $m$ ways to do an other thing, then we have $n * m$ ways to do both". It is so simple that one can argue that it is over represented in teaching but it is indeed important. [35] makes the point that it should be considered a **founding stone of combinatorics**.

Why is the multiplicative property more fundamental than the **additive counting principle**. It is again that the additive property is essentially placed in as a definition of what a **valuation** is. It is in the **in-out-formula** $\chi(A \cup B) + \chi(A \cap B) = \chi(A) + \chi(B)$. Now, this inclusion-exclusion formula is also important in combinatorics but it is already in the **definition** of what we call counting or "adding things up". The multiplicative property on the other hand is not a definition; it actually is quite non-trivial. It characterizes classical mathematics as **quantum mechanics** or **non-commutative flavors of mathematics** have shown that one can extend things. So, if the "rule of product" (which is taught in elementary school) is beefed up to be more geometric and interpreted to Euler characteristic, it becomes fundamental.

- **Combinatorics**. The pigeon hole principle stresses the importance of **order structure**, partially ordered sets (posets) and cardinality or comparisons of cardinality. The use of injective functions to express cardinality is a key part of Cantor. Like some of the ideas of Grothendieck it is of "infantile simplicity" (quote Grothendieck about schemes) but powerful. It allowed for the stunning result that there are different infinities. One of the reason for the success of Cantor's set theory is the immediate applicability. For any new theory, one has to ask: "does it tell me something I did not know?" In "set theory" the larger cardinality of the reals (uncountable) than the cardinality of the algebraic numbers (countable) gave immediately the existence of **transcendental numbers**. This is very elegant. The pigeon hole principle similarly gives combinatorial results which are non trivial and elegant. Currently, searching for "the fundamental theorem of combinatorics" gives the **"rule of product"**. As explained above, we gave it a geometric spin and placed it into topology. Now, combinatorics and topology have always been very hard to distinguish. Euler, who somehow booted up topology by reducing the **Königsberg problem** to a problem in graph theory did that already. Combinatorial topology is essentially part of topology. Today, some very geometric topics like algebraic geometry have been placed within pure **commutative algebra** (this is how I myself was exposed to algebraic geometry) On the other hand, some very hard core combinatorial problems like the upper bound conjecture have been proven with algebro-geometric methods like toric varieties which are geometric. In any case, order structures are important everywhere and the pigeon principle justifies the importance of order structures.
- **Computation**. There is no official "fundamental theorem of computer science" but the **Turing completeness theorem** comes up as a top candidate when searching on engines. Turing formalized using Turing machines in a precise way, what computing is, and even what a proof is. It nails down **mathematical activity** of running an algorithm or argument in a mathematical way. It is also pure as it is **not hardware dependent**. One can also only appreciate Turing's definition if one sees how different programming languages can look like and also in logic, what type of different frame works have been invented. Turing breaks all this complexity with a machine which can be itself part of mathematics leading to the **Halte problem** illustrating the basic limitations of computation. **Quantum computing** would add a hardware component and might break through the **Turing-Church thesis** that everything we can compute can be computed with Turing machines in the same complexity class. Goedel and Turing are related and the Turing incompleteness theorem has a similar flavor than the Goedel incompleteness theorems. There is an other angle to it and that is the question of **complexity**. I would predict that most mathematicians would currently favor the Platonic view of the Church thesis and predict that also new paradigms like quantum computing will never go beyond Turing computability or even not break through complexity barriers like P-NP thresholds. It is just that the Turing completeness theorem is too beautiful to be spoiled by a different type of complexity tied to a physical world. The point of view is that anything we see in the physical world can in principle be computed with a machine **without changing the complexity class**. But that picture could

be as naive as Hilbert's dream one hundred years ago. Still, whatever happens in the future, the Turing completeness theorem remains a theorem. Theorems stay true.

- **Logic**. One can certainly argue whether it would be justified to have Goedel's theorem replaced by a theorem in category theory like the Yoneda lemma. The Yoneda result is not easy to state and it does not produce yet an "Aha moment" like Goedel's theorem does (the liars paradox explains the core of Goedel's theorem). Maybe in the future, when all mathematics has been naturally and pedagogically well expressed in categorical language. I'm personally not sure whether this will ever happen: not everything which is nice also had been penetrating large parts of mathematics: an example is given by **non-standard analysis**, which makes calculus orders of magnitudes easier and which is related also to **surreal numbers**, which are the most "natural" numbers. Both concepts have not entered calculus or algebra textbooks and there are reasons: the subjects need mathematical maturity and one can easily make mistakes. (I myself use non-standard analysis on an intuitive level as presented by Nelson [254, 277] and think of a compact set as a finite set for example which for example, where basic theorems almost require no proof like the Bolzano theorem telling that a continuous function on a compact set takes a maximum). Much of category theory is still also a conglomerate of definitions. Also the language of set theory have been overkill. The work of Russel and Whitehead demonstrates, how clumsy things can become if boiled down to the small pieces. We humans like to think and programming in higher order structures, rather than doing assembly coding, we like to work in object oriented languages which give more insight. But we like and make use of that higher order codes can be boiled down to assembly closer to what the basic instructions are. This is similar in mathematics and also in future, a topologist working in 4 manifold theory will hardly think about all the definitions in terms of sets. Category theory has a chance to change the landscape because it is close to computer science and natural data structures. It is more pictorial and flexible than set theory alone. It definitely has been very successful to find new structures and see connections within different fields like computer science [268].

## Index

# Bibliography

## References

[1] P. Abad and J. Abad. The hundred greatest theorems. http://pirate.shu.edu/~kahlnath/Top100.html, 1999.
[2] R. Abraham, J.E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis and Applications*. Applied Mathematical Sciences, 75. Springer Verlag, New York etc., second edition, 1988.
[3] A. Aczel. *Descartes's secret notebook, a true tale of Mathematics, Mysticism and the Quest to Understand the Universe*. Broadway Books, 2005.
[4] C.C. Adams. *The Knot Book*. Freeman and Company, 1994.
[5] D. Adams. *The Hitchhiker's guide to the galaxy*. Pan Books, 1979.
[6] R. Aharoni. *Mathematics, Poetry and Beauty*. World Scientific, 2015.
[7] L. Ahlfors. *Complex Analysis*. McGraw-Hill Education, 1979.
[8] M. Aigner and G.M. Ziegler. *Proofs from the book*. Springer Verlag, Berlin, 2 edition, 2010. Chapter 29.
[9] A. Alexander. *Duel at Dawn*. Harvard University Press, 2010.
[10] P. Alexandroff. *Combinatorial topology*. Dover books on Mathematics. Dover Publications, Inc, 1960. Three volumes bound as one.
[11] J.M. Almira and A. Romero. Yet another application of the Gauss-Bonnet Theorem for the sphere. *Bull. Belg. Math. Soc.*, 14:341–342, 2007.
[12] S. Alpern and V.S. Prasad. Combinatorial proofs of the Conley-Zehnder-Franks theorem on a fixed point for torus homeomorphisms. *Advances in Mathematics*, 99:238–247, 1993.
[13] C. Alsina and R.B. Nelsen. *Charming Proofs. A journey into Elegant Mathematics*, volume 42 of *Dolciani Mathematical Expositions*. MAA, 2010.
[14] J.W. Anderson. *Hyperbolic Geometry*. Springer, 2 edition, 2005.
[15] G.E. Andrews. *The theory of partitions*. Cambridge Mathematical Library. Cambridge University Press, 1976.
[16] K. Arnold and J. Peyton, editors. *A C User's Guide to ANSI C*. Prentice-Hall, Inc., 1992.
[17] V.I. Arnold. *Mathematical Methods of classical mechanics*. Springer Verlag, New York, 2 edition, 1980.
[18] V.I. Arnold. *Lectures on Partial Differential Equations*. Springer Verlag, 2004.
[19] M. Atiyah. *K-Theory*. W.A. Benjamin, Inc, 1967.
[20] M. Atiyah. Mathematics in the 20th century. *American Mathematical Monthly*, 108, 2001.
[21] J-P. Aubin and I. Ekeland. *Applied nonlinear Analysis*. John Wiley and Sons, 1984.
[22] J. Bach. The lebowski theorem. https://twitter.com/plinz/status/985249543582355458, Apr 14, 2018.
[23] J.C. Baez. The octonions. *Bull. Amer. Math. Soc. (N.S.)*, 39(2):145–205, 2002.
[24] R. Balakrishnan and K. Ranganathan. *A textbook of Graph Theory*. Springer, 2012.
[25] A-L. Barabasi. *Linked, The New Science of Networks*. Perseus Books Group, 2002.
[26] G. Baumslag. *Topics in combinatorial group theory*. Birkhäuser Verlag, 1993.
[27] A. Beardon. *Iteration of rational functions*. Graduate Texts in Mathematics. Springer-Verlag, New York, 1990.
[28] E. Behrends. *Fünf Minuten Mathematik*. Vieweg + Teubner, 2006.
[29] E.T. Bell. *The development of Mathematics*. McGraw Hill Book Company, 1945.
[30] E.T. Bell. *Men of Mathematics*. Penguin books, 1953 (originally 1937).
[31] B.Engquist and W. Schmid (Editors). *Mathematics Unlimited - 2001 and Beyond*. Springer, 2001.
[32] M. Berger. *A Panoramic View of Riemannian Geometry*. Springer Verlag, Berlin, 2003.
[33] M. Berger and B. Gostiaux. *Differential geometry: manifolds, curves, and surfaces*, volume 115 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1988.
[34] W. Berlinghoff and F. Gouvea. *Math through the ages*. Mathematical Association of America Textbooks, 2004.
[35] N.L. Biggs. The roots of combinatorics. *Historia Mathematica*, 6:109–136, 1979.
[36] G. D. Birkhoff. An extension of poincareés last theorem. *Acta Math.*, 47:297–311, 1925.
[37] J. Bondy and U. Murty. *Graph theory*, volume 244 of *Graduate Texts in Mathematics*. Springer, New York, 2008.

[38] W.W. Boone and G. Higman. An algebraic characterization of the solvability of the word problem. *J. Austral. Math. Soc.*, 18, 1974.
[39] K.C. Border. *Fixed point theorems with applications to economics and game theory*. Cambridge University Press, 1985.
[40] N. Bourbaki. *Elements d'histoire des Mathematiques*. Springer, 1984.
[41] N. Bourbaki. *Elements de Mathematique*. Springer, 2006.
[42] J. Bourgain. *Green's function estimates for lattice Schrödinger operators and applications*, volume 158 of *Annals of Mathematics Studies*. Princeton Univ. Press, Princeton, NJ, 2005.
[43] C. Boyer. *A History of Mathematics*. John Wiley and Sons, Inc, 2. edition edition, 1991.
[44] G.E. Bredon. *Topology and Geometry*, volume 139 of *Graduate Texts in Mathematics*. Springer Verlag, 1993.
[45] S. Brendle. *Ricci Flow and the Sphere theorem*. Graduate Studies in Mathematics. AMS, 2010.
[46] D. Bressoud. Historical reflections on the fundamental theorem of calculus. MAA Talk of May 11, 2011.
[47] O. Bretscher. Calculus I. Lecture Notes Harvard, 2006.
[48] H. Brezis. *Functional Anslysis, Sobolev Spaces and Partial Differential Equations*. University text. Springer, 2011.
[49] M. Brown and W.D. Neuman. Proof of the Poincaré-Birkhoff fixed point theorem. *Michigan Mathematical Journal*, 24:21–31, 1977.
[50] R.A. Brualdi. *Introductory Combinatorics*. Pearson Prantice Hall, forth edition, 2004.
[51] G. Van Brummelen. *Heavenly Mathematics*. Princeton University Press, 2013.
[52] C. Bruter. *Mathematics and Modern Art*, volume 18 of *Springer Proceedings in mathematics*. Springer, 2012.
[53] P. Bullen. *Dictionary of Inequalities*. CRC Press, 2 edition, 2015.
[54] E.B. Burger. *Exploring the Number Jungle*. AMS, 2000.
[55] W. Burnside. *Theory of Groups of Finite Order*. Cambridge at the University Press, 1897.
[56] W. Byers. *How Mathematicians Think*. Princeton University Press, 2007.
[57] W.T. Haight II C. Cavagnaro. *Classical and Theoretical mathematics*. CRC Press, 2001.
[58] O. Caramello. *Theories, Sites, Toposes*. Oxford University Press, 2018.
[59] L. Carleson and T.W. Gamelin. *Complex Dynamics*. Springer-Verlag, New York, 1993.
[60] A-L. Cauchy. *Cours d'Analyse*. L'Imprimerie Royale, 1821.
[61] C.Clapman and J. Nicholson. *Oxford Concise Dictionary of Mathematics*. Oxford Paperback Reference. Oxford University Press, 1990.
[62] A. Lax C.D. Olds and G. Davidoff. *The Geometry of Numbers*, volume 41 of *Anneli Lax New Mathematical Library*. AMS, 2000.
[63] P. E. Ceruzzi. *A History of Modern Computing*. MIT Press, second edition, 2003.
[64] K. Chandrasekharan. *Introduction to Analytic Number Theory*, volume 148 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1968.
[65] G. Chartrand and P. Zhang. *Chromatic Graph Theory*. CRC Press, 2009.
[66] N. Chernof and R. Markarian. *Chaotic billiards*. AMS, 2006.
[67] C. Chevalley. *Theory of Lie Groups*. Princeton University Press, 1946.
[68] J.R. Choksi and M.G.Nadkarni. Baire category in spaces of measures, unitary operators and transformations. In *Invariant Subspaces and Allied Topics*, pages 147–163. Narosa Publ. Co., New Delhi, 1990.
[69] I. Ciufolini and J.A. Wheeler. *Gravitation and inertia*. Princeton Series in Physics, 1995.
[70] R. Cochrane. *The secret Life of equations: the 50 greatest equations and how they work*. Octopus Company, 2016.
[71] R. Cochrane. *Math Hacks*. Cassell Illustrated, 2018.
[72] E.A. Coddington and N. Levinson. *Theory of Ordinary Differential Equations*. McGraw-Hill, New York, 1955.
[73] P.R. Comwell. *Polyhedra*. Cambridge University Press, 1997.
[74] A. Connes. *Noncommutative geometry*. Academic Press, 1994.
[75] J. Conway. Universal quadratic forms and the fifteen theorem. *Contemporary Mathematics*, 272:23–26, 1999.
[76] J.B. Conway. *Functions of One Complex Variable*. Springer Verlag, 2. edition, 1978.
[77] J.B. Conway. *A course in functional analysis*. Springer Verlag, 1990.

[78] J.B. Conway. *Mathematical Connections: A Capstone Course*. American Mathematical Society, 2010.
[79] J.H. Conway. *On Numbers and Games*. A K Peters, Ltd, 2001.
[80] J.H. Conway and R.K. Guy. *The book of numbers*. Copernicus, 1996.
[81] J.H. Conway and N.J.A.Sloane. What are all the best sphere packings in low dimensions. *Discr. Comp. Geom.*, 13:383–403, 1995.
[82] J.H. Conway and N.J.A. Sloane. *Sphere packings, Lattices and Groups*, volume 290 of *A series of Comprehensive Studies in Mathematics*. Springer Verlag, New York, 2.nd edition edition, 1993.
[83] I.P. Cornfeld, S.V.Fomin, and Ya.G.Sinai. *Ergodic Theory*, volume 115 of *Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen*. Springer Verlag, 1982.
[84] R. Courant and H. Robbins. *Was ist Mathematik*. Springer, fifth edition, 1941.
[85] T. Crilly. *50 mathematical ideas you really need to know*. Quercus, 2007.
[86] K.S. Thorne C.W. Misner and J.A. Wheeler. *Gravitation*. Freeman, San Francisco, 1973.
[87] H.L. Cycon, R.G.Froese, W.Kirsch, and B.Simon. *Schrödinger Operators—with Application to Quantum Mechanics and Global Geometry*. Springer-Verlag, 1987.
[88] D.Downing. *Dictionary of Mathematical Terms*. Barron's Educational Series, 1995.
[89] W. de Melo and S. van Strien. *One dimensional dynamics*, volume 25 of *Series of modern surveys in mathematics*. Springer Verlag, 1993.
[90] M. Denker, C. Grillenberger, and K. Sigmund. *Ergodic Theory on Compact Spaces*. Lecture Notes in Mathematics 527. Springer, 1976.
[91] L.E. Dickson. *History of the theory of numbers.Vol. I:Divisibility and primality*. Chelsea Publishing Co., New York, 1966.
[92] L.E. Dickson. *History of the theory of numbers.Vol.II:Diophantine analysis*. Chelsea Publishing Co., New York, 1966.
[93] J. Dieudonné. *A Panorama of Pure Mathematics*. Academic Press, 1982.
[94] J. Dieudonné. *Grundzüge der modernen Analysis*. Vieweg, dritte edition, 1985.
[95] C. Ding, D. Pei, and A. Salomaa. *Chinese Remainder Theorem, Applications in Computing, Coding Cryptography*. World Scientific, 1996.
[96] M.P. do Carmo. *Differential Forms and Applications*. Springer Verlag, 1994.
[97] A. Dold. *Lectures on algebraic topology*. Springer, 1980.
[98] J. Doob. *Stochastic processes*. Wiley series in probability and mathematical statistics. Wiley, New York, 1953.
[99] R. Douady. Application du théorème des tores invariantes. These 3 ème cycle, Université Paris VII, 1982.
[100] B.A. Dubrovin, A.T.Fomenko, and S.P.Novikov. *Modern Geometry-Methods and Applications Part I,II,III*. Graduate Texts in Mathematics. Springer Verlag, New York, 1985.
[101] W. Dunham. *Journey through Genius*. Wiley Science Editions, 1990.
[102] J.-P. Eckmann, H. Koch, and P. Wittwer. A computer-assisted proof of universality for area-preserving maps. *Memoirs of the AMS*, 47:1–122, 1984.
[103] J-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57:617–656, 1985.
[104] B.G. Sidharth (Editor). *A century of Ideas*. Springer, 2008.
[105] K. Ito (Editor). *Encyclopedic Dictionary of Mathematics (2 volumes)*. MIT Press, second edition edition, 1993.
[106] N.J. Higham (Editor). *The Princeton Companion to Applied Mathematics*. Princeton University Press, 2015.
[107] R. Brown (Editor). *30-Second Maths*. Ivy Press, 2012.
[108] S.G. Krantz (editor). *Comprehensive Dictionary of Mathematics*. CRC Press, 2001.
[109] S.G. Krantz (Editor). *Dictionary of Algebra, Arithmetic and Trigonometry*. CRC Press, 2001.
[110] T. Gowers (Editor). *The Princeton Companion to Mathematics*. Princeton University Press, 2008.
[111] C.H. Edwards. *The historical Development of the Calculus*. Springer Verlag, 1979.
[112] R. L. Eisenman. Classroom Notes: Spoof of the Fundamental Theorem of Calculus. *Amer. Math. Monthly*, 68(4):371, 1961.
[113] R. Elwes. *Math in 100 Key Breakthroughs*. Quercus, 2013.
[114] T. Rokicki et al. God's number is 20. http://www.cube20.org/, 2010.
[115] R.L. Eubank. *A Kalman Filter Primer*. Chapman and Hall, CRC, 2006.

[116] H. Eves. *Great moments in mathematics (I and II.* The Dolciani Mathematical Expositions. Mathematical Association of America, Washington, D.C., 1981.
[117] K.J. Falconer. *Fractal Geometry.* Wiley, second edition edition, 2003.
[118] B. Farb and R.K. Dennis. *Noncommutative Algebra.* Springer, 1993.
[119] B. Farb and J. Wolfson. Resolvent degree, hilbert's 13th problem and geometry. https://www.math.uchicago.edu/ farb/papers/RD.pdf, 2018.
[120] O. Faugeras. *Three-dimensional computer vision: a geometric viewpoint.* Cambridge MA: MIT Press, Cambridge, MA, USA, second edition edition, 1996.
[121] H. Federer. *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.
[122] E.A. Fellmann. *Leonhard Euler.* Birkhäuser, Basel, Boston, Berlin, 2007.
[123] B. Fine and G. Rosenberger. *The Fundamental Theorem of Algebra.* Undergraduate Texts in Mathematics. Springer, 1997.
[124] K. Fink. *A brief History of Mathematics.* OPen Court Publishing Co, 1900.
[125] G. Fischer. *Mathematical Models.* Springer Spektrum, 2 edition, 2017.
[126] J-P. Francoise, G.L. Naber, and T.S. Tsun. *Encylopedia of Mathematical Physics.* Elsevier, 2006.
[127] T. Franzen. *Gödel's Theorem.* A.K. Peters, 2005.
[128] N.A. Friedman. *Introduction to Ergodic Theory.* Van Nostrand-Reinhold, Princeton, New York, 1970.
[129] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory.* Princeton University Press, Princeton, N.J., 1981. M. B. Porter Lectures.
[130] L. Gamwell. *Mathematics + Art, a cultural history.* Princeton University Press, 2016.
[131] J.E. Littlewood G.H. Hardy and G. Polya. *Inequalities.* Cambridge at the University Press, 1959.
[132] M. Giaquinta and S. Hildebrandt. *Calculus of variations. I,II*, volume 310 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1996.
[133] P. Glendinning. *Math In Minutes.* Quercus, 2012.
[134] D. Goldfeld. Beyond the last theorem. *The Sciences*, 3/4:43–40, 1996.
[135] C. Golé. *Symplectic Twist Maps, Global Variational Techniques.* World Scientific, 2001.
[136] S.W. Golomb. Rubik's cube and quarks: Twists on the eight corner cells of rubik's cube provide a model for many aspects of quark behavior. *American Scientist*, 70:257–259, 1982.
[137] T. Gowers. *Mathematics, a very short introduction.* Oxford University Press, 2002.
[138] R. Graham. *Rudiments of Ramsey Theory*, volume 45 of *Regional conference series in Mathematics.* AMS, 1980.
[139] R. Graham. Some of my favorite problems in ramsey theory. *Integers: electronic journal of combinatorial number theory*, 7, 2007.
[140] I. Grattan-Guinness. *The Rainbow of Mathematics.* W. W. Norton, Company, 2000.
[141] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, 167:481–547, 2008.
[142] P. Griffiths and J. Harris. *Principles of Algebraic Geometry.* Pure and Applied Mathematics. John Wiley and Sons, 1978.
[143] H. Groemer. Existenzsätze für Lagerungen im Euklidischen Raum. *Mathematische Zeitschrift*, 81:260–278, 1963.
[144] J. Gross and J. Yellen, editors. *Handbook of graph theory.* Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, 2004.
[145] B. Grünbaum. Are your polyhedra the same as my polyhedra? In *Discrete and computational geometry*, volume 25 of *Algorithms Combin.*, pages 461–488. Springer, Berlin, 2003.
[146] B. Grünbaum. *Convex Polytopes.* Springer, 2003.
[147] B. Grünbaum and G.C. Shephard. *Tilings and Patterns.* Dover Publications, 2013.
[148] R. Guy. The strong law of small numbers. *Amer. Math. Monthly*, 95:697–712, 1988.
[149] T.C. Hales. Jordan's proof of the jordan curve theorem. *Studies in logic, grammar and rhetorik*, 10, 2007.
[150] G.R. Hall. Some examples of permutations modelling area preserving monotone twist maps. *Physica D*, 28:393–400, 1987.
[151] P. Halmos. *Lectures on ergodic theory.* The mathematical society of Japan, 1956.
[152] P.R. Halmos. *Naive set theory.* Van Nostrand Reinhold Company, 1960.
[153] G.H. Hardy. *Divergent Series.* AMS Chelsea Publishing, 1991.

[154] G.H. Hardy. *A mathematician's Apology*. Cambridge University Press, 1994.
[155] G.H. Hardy and M. Riesz. *The general theory of Dirichlet's series*. Hafner Publishing Company, 1972.
[156] G.H. Hardy and E.M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 1980.
[157] R. Hartley and A. Zissermann. *Multiple View Geometry in computer Vision*. Cambridge UK: Cambridge University Press, 2003. Second edition.
[158] R. Hartshorne. *Algebraic Geometry*. Springer-Verlag, New York, 1977. Graduate Texts in Mathematics, No. 52.
[159] F. Hausdorff. Summationsmethoden und Momentfolgen, I,II. *Mathematische Zeitschrift*, 9:74–109, 280–299, 1921.
[160] G.A. Hedlund. Endomorphisms and automorphisms of the shift dynamical system. *Math. Syst. Theor.*, 3:320–375, 1969.
[161] S. Helgason. *The Radon transform*, volume 5 of *Progress in mathematics*. Birkhaeuser, 1980.
[162] M. Herman. Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. *IHES*, 49:5–233, 1979.
[163] R. Hersh. *What is Mathematics, Really?* Oxford University Press, 1997.
[164] T.H. Hildebrandt and I.J. Schoenberg. On linear functional operations and the moment problem for a finite interval in one or several dimensions. *Annals of Mathematics*, 34:317–328, 1933.
[165] H. Hofer and E. Zehnder. *Symplectic invariants and Hamiltonian dynamics*. Birkhäuser advanced texts. Birkhäuser Verlag, Basel, 1994.
[166] A.J. Hoffman and C.W. Wu. A simple proof of a generalized cauchy-binet theorem. *American Mathematical Monthly*, 123:928–930, 2016.
[167] L.K. Hua. *Introduction to Number theory*. Springer Verlag, Berlin, 1982.
[168] O. Lanford III. A computer-assisted proof of the feigenbaum conjectures. *Bull. Amer. Math. Soc.*, 6:427–434, 1982.
[169] O.E. Lanford III. A shorter proof of the existence of the Feigenbaum fixed point. *Commun. Math. Phys*, 96:521–538, 1984.
[170] L. Illusie. What is a topos? *Notices of the AMS*, 51, 2004.
[171] I.R.Shafarevich and A.O. Remizov. *Linear Algebra and Geometry*. Springer, 2009.
[172] R. Girgensohn J. Borwein, D.Bailey. *Experimentation in Mathematics*. A.K. Peters, 2004. Computational Paths to Discovery.
[173] P.C. Jackson. *Introduction to artificial intelligence*. Dover publications, 1985. Second, Enlarged Edition.
[174] T. Jackson. *An illustrated History of Numbers*. Shelter Harbor Press, 2012.
[175] T. Jech. *The axiom of choice*. Dover, 2008.
[176] M. Jeng and O. Knill. Billiards in the $l^p$ unit balls of the plane. *Chaos, Fractals, Solitons*, 7:543–545, 1996.
[177] C. Goodman-Strauss J.H. Conway, H.Burgiel. *The Symmetries of Things*. A.K. Peterse, Ltd., 2008.
[178] M.C. Jordan. *Cours d'Analyse*, volume Tome Troisieme. Gauthier-Villards,Imprimeur-Libraire, 1887.
[179] J. Jost. *Riemannian Geometry and Geometric Analysis*. Springer Verlag, 2005.
[180] D. Joyner. *Adventures in Group Theory, Rubik's Cube, Merlin's Machine and Other Mathematical Toys*. Johns Hopkins University Press, second edition, 2008.
[181] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.
[182] L.A. Kaluzhnin. *The Fundamental Theorem of Arithmetic*. Little Mathematics Library. Mir Publishers, Moscow, 1979.
[183] A. Katok. Fifty years of entropy in dynamics: 1958-2007. *Journal of Modern Dynamics*, 1:545–596, 2007.
[184] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its applications*. Cambridge University Press, 1995.
[185] A.B. Katok and A.M. Stepin. Approximations in ergodic theory. *Russ. Math. Surveys*, 22:77–102, 1967.
[186] A.B. Katok and A.M. Stepin. Metric properties of measure preserving homemorphisms. *Russ. Math. Surveys*, 25:191–220, 1970.
[187] V. Katz. The history of Stokes theorem. *Mathematics Magazine*, 52, 1979.
[188] V. Katz. *Mathematics of Egypt, Mesopotamia, China, India and Islam*. Princeton University Press, 2007.
[189] V.J. Katz. The history of stokes theorem. *Mathematics Magazine*, 52:146–156, 1979.
[190] Y. Katznelson. *An introduction to harmonic analysis*. John Wiley and Sons, Inc, New York, 1968.

[191] A.S. Kechris. *Classical Descriptive Set Theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 1994.
[192] D.A. Klain and G-C. Rota. *Introduction to geometric probability*. Lezioni Lincee. Accademia nazionale dei lincei, 1997.
[193] F. Klein. *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert*. Springer, 1979 (originally 1926).
[194] J.R. Kline. What is the Jordan curve theorem? *American Mathematical Monthly*, 49:281–286, 1942.
[195] M. Kline. *Mathematical Thought from Ancient to Modern Time*. Oxford University Press, 1972.
[196] O. Knill. Singular continuous spectrum and quantitative rates of weakly mixing. *Discrete and continuous dynamical systems*, 4:33–42, 1998.
[197] O. Knill. Weakly mixing invariant tori of hamiltonian systems. *Commun. Math. Phys.*, 205:85–88, 1999.
[198] O. Knill. A multivariable chinese remainder theorem and diophantine approximation. http://people.brandeis.edu/∼kleinboc/EP0405/knill.html, 2005.
[199] O. Knill. *Probability Theory and Stochastic Processes with Applications*. Overseas Press, 2009.
[200] O. Knill. A multivariable chinese remainder theorem. https://arxiv.org/abs/1206.5114, 2012.
[201] O. Knill. A Cauchy-Binet theorem for Pseudo determinants. *Linear Algebra and its Applications*, 459:522–547, 2014.
[202] O. Knill. Some fundamental theorems in mathematics. https://arxiv.org/abs/1807.08416, 2018.
[203] O. Knill. Top 10 fundamental theorems. https://www.youtube.com/watch?v=Dvj0gfajdNI, 2018.
[204] O. Knill, J. Carlsson, A. Chi, and M. Lezama. An artificial intelligence experiment in college math education. http://www.math.harvard.edu/knill/preprints/sofia.pdf, 2003.
[205] O. Knill and O. Ramirez-Herran. On ullman's theorem in computer vision. Retrieved February 2, 2009, from http://arxiv.org/abs/0708.2442, 2007.
[206] O. Knill and E. Slavkovsky. Visualizing mathematics using 3d printers. In C. Fonda E. Canessa and M. Zennaro, editors, *Low-Cost 3D Printing for science, education and Sustainable Development*. ICTP, 2013. ISBN-92-95003-48-9.
[207] D. E. Knuth. *Surreal numbers*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1974.
[208] T.W. Koerner. *Fourier Analysis*. Cambridge University Press, 1988.
[209] N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin, 1933. English: Foundations of Probability Theory, Chelsea, New York, 1950., 1933.
[210] N. Kondratieva. Three most beautiful mathematical formulas. http://fy.chalmers.se/˜tfkhj/BeautifulFormulas.pdf, 2006.
[211] J. Kottke. The lebowski theorem of machine superintelligence. https://kottke.org/18/04/the-lebowski-theorem-of-machine-superintelligence, Apr 16, 2018.
[212] S. Krantz. *Function Theory of Serveral Complex Variables*. AMS Chelsea publishing, 2 edition, 2001.
[213] S.G. Krantz. *Mathematical Apocrypha*. AMS, 2002.
[214] S.G. Krantz. *Mathematical Apocrypha Redux*. AMS, 2005.
[215] S.G. Krantz and H. Parks. *A mathematical Odyssey: journal from the real to the complex*. Springer Verlag, 2014.
[216] R. Krömer. *Tool and Object: A history and Philosophy of Category Theory*. Birkhäuser, 2007.
[217] I. Lakatos. *Proofs and Refutations*. Cambridge University Press, 1976.
[218] S. Mac Lane. *Categories for the Working Mathematician*. Springer, 1998.
[219] O. Lanford. Lecture notes in dynamical systems. ETH Zuerich, 1987.
[220] A.N. Langville and C.D. Meyer. *Google's page rank and beyond: the science of search engine rankings*. Princeton University Press, 2006.
[221] P.D. Lax. Approximation of measure preserving transformations. *Commun. Pure Appl. Math.*, 24:133–135, 1971.
[222] T. Leinster. Rethinking set theory. https://arxiv.org/abs/1212.6543, 2012.
[223] S. Lem. *The futurological congress*. A Harvest book, 1985.
[224] H.W. Lenstra. Solving the pell equation. *Algorithmic Number Theory*, 44, 2008.

[225] E.H. Lieb and M. Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, 1996.
[226] M. Livio. *The equation that couldn't be solved*. Simon and Schuster, 2005.
[227] M. Livio. *Is God a mathematician*. Simon and Schuster Paperbacks, 2009.
[228] S. Morrison M. Feedman, R. Gompf and K. Walker. Man and machine thinking about the smooth 4-dimensional poincaré conjecture. https://arxiv.org/pdf/0906.5177.pdf, 2009.
[229] J. MacCormick. *9 Algorithms that changed the future*. Princeton University Press, 2012.
[230] B.B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freeman and Company, 1982.
[231] E. Maor. *The Pythagorean Theorem: A 4000 year history*. Princeton University Press, 2007.
[232] R. Marchthaler and S. Dingler. *Kalman-Filter*. Springer Verlag, 2017.
[233] R. Martinez. Proofs that every mathematician should know. https://math.stackexchange.com/questions/178940/proofs-that-every-mathematician-should-know, 2018.
[234] J-C. Martzloff. *A history of Chinese Mathematics*. Springer Verlag, second edition, 1997.
[235] B. Mazur. *Imagining Numbers*. Penguin Books, 2003.
[236] J. Mazur. *Fluke, the math and myth of coincidence*. Basic Books, 2016.
[237] C. McLarty. The uses and abuses of the history of topos theory. *Brit. J. Phil. Sci*, pages 351–375, 1990.
[238] C. McLarty. *Elementary Categories, Elementary Toposes*. Clarendon Press, 1992.
[239] C. McLarty. The rising sea: Grothendieck on simplicity and generality. In *Episodes in the history of modern algebra (1800–1950)*, volume 32 of *Hist. Math.*, pages 301–325. Amer. Math. Soc., Providence, RI, 2007.
[240] C. McMullen. Riemann surfaces, dynamics and geometry. Course Notes, Harvard University, 2014.
[241] R. Merris. *Graph theory*. Interscience Series in discrete mathematics and optimization. Wiley, 2001.
[242] J. Milnor. Dynamics in one complex variable. Introductory Lectures, SUNY, 1991.
[243] M. Minsky. *The Society of Mind*. A Touchstone Book, published by Simon @ Shuster Inc, New York, London, Toronto, Sydney and Tokyo, 1988.
[244] U. Montano. *Explaining Beauty in Mathematics: An Aesthetic Theory of Mathematics*, volume 370 of *Synthese Library*. Springer, 2013.
[245] J.W. Morgan. The poincaré conjecture. In *Proceedings of the ICM, 2006*, pages 713–736, 2007.
[246] J.W. Morgan and G. Tian. *Ricci Flow and the Poincaré Conjecture*, volume 3 of *Clay Mathematics Monographs*. AMS, Clay Math Institute, 2007.
[247] J. Moser. *Stable and random Motion in dynamical systems*. Princeton University Press, Princeton, 1973.
[248] J. Moser. *Selected chapters in the calculus of variations*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003. Lecture notes by Oliver Knill.
[249] D. Mumford. O. zariski. *National Academy of Sciences*, 2013.
[250] K.G. Murty. *Linear Programming*. Wiley, 1983.
[251] M.Ram Murty and A Pacelli. Quadratic reciprocity via theta functions. *Proc. Int. Conf. Number Theory*, 1:107–116, 2004.
[252] E. Nagel and J.R. Newman. *Godel's proof*. New York University Press, 2001.
[253] I.P. Natanson. *Constructive Function theory*. Frederick Ungar Publishing, 1965.
[254] E. Nelson. Internal set theory: A new approach to nonstandard analysis. *Bull. Amer. Math. Soc*, 83:1165–1198, 1977.
[255] J. Neunhäuserer. *Schöne Sätze der Mathematik*. Springer Spektrum, 2017.
[256] T.Lewis N.I. Fisher and B.J. Embleton. *Statistical analysis of spherical data*. Cambridge University Press, 1987.
[257] M. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, 2006.
[258] J. Owen. *The grammar of Ornament*. Quaritch, 1910.
[259] J.C. Oxtoby. *Measure and Category*. Springer Verlag, New York, 1971.
[260] R.S. Palais. *Seminar on the Atiyah-Singer Index Theorem*, volume 57 of *Annals of Mathematics Studies*. Princeton University Press, 1965.
[261] R.S. Palais. A simple proof of the banach contraction principle. *Journal Fixed point theory and Applications*, 2:221–223, 2007.
[262] R.S. Palais and S. Smale. A generalized morse theory. *Bull. Amer. Math. Soc.*, 70:165–172, 1964.
[263] E. Pariser. *The Filter Bubble*. Penguin Books, 2011.
[264] H-O. Peitgen and P.H.Richter. *The beauty of fractals*. Springer-Verlag, New York, 1986.

[265] C. Petzold. *The Annotated Turing: A Guided Tour Through Alan Turing's Historic Paper on Computability and the Turing Machine*. Wiley Publishing, 2008.
[266] R. R. Phelps. *Lectures on Choquet's theorem*. D.Van Nostrand Company, Princeton, New Jersey, 1966.
[267] C.A. Pickover. *The Math Book*. Sterling, New York, 2012.
[268] B.C. Pierce. *Basic Category Theory for Computer Scientists*. MIT Press, 1991.
[269] B. Polster and M. Ross. *Math Goes to the Movies*. Johns Hopkins University Press, 2012.
[270] A.S. Posamentier. *The Pythagorean Theorem: The Story of its Power and Beauty*. Prometheus Books, 2010.
[271] T. Poston and I. Stewart. *Catastrophe theory and its applications*. Pitman, 1978.
[272] Q.I. Rahman and G. Schmeisser. *Analytic theory of polynomials*, volume 26 of *London Mathematical Society Monographs*. Oxford University Press, new series edition, 2002.
[273] F. Rannou. Numerical study of discrete area-preserving mappings. *Acta Arithm*, 31:289–301, 1974.
[274] M. Reed and B. Simon. *Methods of modern mathematical physics*. Academic Press, Orlando, 1980.
[275] D.S. Richeson. *Euler's Gem*. Princeton University Press, Princeton, NJ, 2008. The polyhedron formula and the birth of topology.
[276] H. Riesel. *Prime numbers and computer methods for factorization*, volume 57 of *Progress in Mathematics*. Birkhäuser Boston Inc., 1985.
[277] A. Robert. *Analyse non standard*. Presses polytechniques romandes, 1985.
[278] J. Rognes. On the Atiyah-Singer index theorem. http://www.abelprize.no, 2004.
[279] J. Rosenhouse. *The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brain Teaser*. Oxford University Press, 2009.
[280] G-C. Rota. The phenomenology of mathematical beauty. *Synthese*, 111(2):171–182, 1997. Proof and progress in mathematics (Boston, MA, 1996).
[281] C. Rovelli. *The order of Time*. Riverhead Books, 2018.
[282] D. Ruelle. *Statistical Mechanics*. Mathematics Physics Monograph Series. W.A. Benjamin, Inc, 1968.
[283] D. Ruelle. *Thermodynamic Formalism. The Mathematical Structures of Classical Equilibrium Statistical Mechanics*, volume 5 of *Encyclopedia of mathematics and its applications*. Addison-Wesley Publishing Company, London, 1978.
[284] D. Ruelle. *The mathematician's brain*. Princeton University Press, Princeton, NJ, 2007.
[285] S. Russell and P. Norvig. *Artificial Intelligence: A modern approach*. Prentice Hall, 1995, 2003. second edition.
[286] D.M.T.Benincasa S. Zeiki, J.P. Romaya and M.F.Atiyah. The experience of mathematical beauty and its neural correlates. *Front. Hum. Neurosci*, 13, 2014.
[287] A. Sard. The measure of the critical values of differentiable maps. *Bull. Amer. Math. Soc.*, 48:883–890, 1942.
[288] H. Schenk. *Computationsl algebraic Geometry*, volume 58 of *Student texts*. Cambridge University Press, 2003.
[289] L. Schläfli. *Theorie der Vielfachen Kontinuität*. Cornell University Library Digital Collections, 1901.
[290] H. Segerman. *Visualizing Mathematics with 3D Printing*. John Hopkins University Press, 2016.
[291] G. Shafer and V. Vovk. The sources of komogorov's grundbegriffe. *Statistical Science*, pages 70–98, 2006.
[292] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423,623–656, 1948.
[293] T. Siegfried. *A beautiful math*. Joseph Henry Press, 2006.
[294] B. Simon. Fifteen problems in mathematical physics. In *Anniversary of Oberwolfach*, Perspectives in Mathematics. Birkhäuser Verlag, Basel, 1984.
[295] B. Simon. Operators with singular continuous spectrum: I. General operators. *Annals of Mathematics*, 141:131–145, 1995.
[296] B. Simon. *Trace Ideals and their Applications*. AMS, 2. edition, 2010.
[297] B. Simon. *A comprehensive course in Analysis*. AMS, 2017.
[298] Ya.G. Sinai. *Introduction to ergodic theory*. PrincetonUniversity press, Princeton, 1976.
[299] N. Sinclair. *Mathematics and Beauty*. Teachers College Press, 2006.
[300] S. Skinner. *Sacred Geometry*. Sterling, 2006.
[301] J.L. Snell and R. Vanderbei. Three bewitching paradoxes. In *Topics in Contemporary Probability and its Applications*, Probability and Stochastics Series, pages 355–370. CRC Press, Boca Raton, 1995.

[302] R. Solomon. A brief history of the classification of the finite simple groups. *Bulletin (new series) of the AMS*, 38:315–352, 2001.

[303] E. Sondheimer and A. Rogerson. *Numbers and infinity*. Dover Publications Inc., Mineola, NY, 2006. A historical account of mathematical concepts, Reprint of the 1981 original.

[304] M. Spivak. *A comprehensive Introduction to Differential Geometry I-V*. Publish or Perish, Inc, Berkeley, third edition, 1999.

[305] H. Spohn. *Large scale dynamics of interacting particles*. Texts and monographs in physics. Springer-Verlag, New York, 1991.

[306] F. Staudacher. *Jost Bürgi*. Verlag NZZ, 2016.

[307] S. Sternberg. Local contractions and a theorem of Poincaré. *Amer. J. Math.*, 79:809–824, 1957.

[308] S. Sternberg. *Lectures on Differential Geometry*. Prentice Hall, 1964.

[309] I. Stewart. *In Pursuit of the Unknown: 17 equations that changed the world*. Basic Books, 2012.

[310] I. Stewart and D. Tall. *Algebraic Number Theory and Fermat's last Theorem*. A.K. Peters, third edition, 2002.

[311] J. Stillwell. *Mathematics and its history*. Springer, 2010.

[312] J. Stillwell. *Elements of Mathematics: From Euclid to Goedel*. Princeton University Press, 2016.

[313] G. Strang. The fundamental theorem of linear algebra. *Amer. Math. Monthly*, 100(9):848–855, 1993.

[314] S. H. Strogatz. *Sync: The Ermerging Science of Spontaneous Order*. Hyperion, 2003.

[315] D.W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, 1993.

[316] D.J. Struik. *A concise History of Mathematics*. Dover, 1948.

[317] F.J. Swetz and T.I. Kao. *Was Pythagoras Chinese?* Pennsylvania State University Press, 1988.

[318] S. Tabachnikov. *Billiards*. Panoramas et synthèses. Société Mathématique de France, 1995.

[319] J. Tanton. *Encyclopedia of Mathematics*. Facts on File, Inc, 2005.

[320] T. Tao. *Structure and Randomness*. AMS, 2008.

[321] T. Tao. *Topics in Random Matrix Theory*. Graduate Studies in Mathematics. AMS, 2012.

[322] C. Tapp. *An den Grenzen des Endlichen: Das Hilbertprogramm im Kontext von Formalismus und Finitismus*. Springer Verlag, 2013.

[323] R.F. Thom. *Mathematical Models of Morphogenesis*. Mathematics and Its Applications. Horwood Ltd, 1984.

[324] J. Todd. *Constructive Theory of Functions*. Birkhäuser Verlag, 1963.

[325] E. Trucco and A. Verri. *Introductory techniques for 3-D computer vision*. Upper Saddle River NJ: Prentice Hall, 1998.

[326] S. Ullman. *The interpretation of visual motion*. Cambridge MA : MIT Press, 1979.

[327] V.Arnold. *Catastrophe theory*. Springer, 1981.

[328] C. Villani. *Optimal transport, old and new*. Springer Verlag, 2008.

[329] I.M. Vinogradov. *The Method of Trigonometric Sums in the Theory of Numbers*. Dover Publications, 1954.

[330] D.V. Treshchev V.V. Kozlov. *Billiards*, volume 89 of *Translations of mathematical monographs*. AMS, 1991.

[331] Fulton W. *Algebraic Curves. An Introduction to Algebraic Geometry*. Addison Wesley, 1969.

[332] F. Warner. *Foundations of differentiable manifolds and Lie groups*, volume 94 of *Graduate texts in mathematics*. Springer, New York, 1983.

[333] D. J. Watts. *Small Worlds*. Princeton University Press, 1999.

[334] D. J. Watts. *Six Degrees*. W. W. Norton and Company, 2003.

[335] J.N. Webb. *Game Theory*. Springer, 2007.

[336] E.W. Weinstein. *CRC Concise Encyclopedia of Mathematics*. CRC Press, 1999.

[337] J. Weizenbaum. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9:36–45, 1965.

[338] D. Wells. Which is the most beautiful? *Mathematical Intelligencer*, 10:30–31, 1988.

[339] D. Wells. Are these the most beautiful? *The Mathematical Intelligencer*, 12:9, 1990.

[340] W.Fulton and J. Harris. *Representation theory*. Graduate Texts in Mathematics. Springer, 2004.

[341] H. Whitney. *Collected Works*. Birkhaueser, 1992.

[342] D. Williams. *Probability with Martingales*. Cambridge mathematical Texbooks, 1991.

[343] M. Wojtkowski. Principles for the disign of billiards with nonvanishing lyapunov exponents. *Communications in Mathematical Physics*, 105:391–414, 1986.

[344] N. Wolchover. Proven - 'most important unsolved problem' in numbers. *NBC news, 9/11/2012*, 2012.
[345] S. Wolfram. *Theory and Applications of Cellular Automata*. World Scientific, 1986.
[346] S. Wolfram. *A new kind of Science*. Wolfram Media, 2002.
[347] H. Wussing. *6000 Jahre Mathematik*. Springer, 2009.
[348] F.B. Cannonito W.W. Boone and R.C. Lyndon. *Word problems*. North-Holland Publishing Company, 1973.
[349] D. Zagier. A one-sentence proof that every prime p = 1 mod 4 is a sum of two squares. *Amer. Math. Monthly*, 97:144, 1990.
[350] G.M. Ziegler. *Lectures on Polytopes*. Springer Verlag, 1995.
[351] J. D. Zund. George David Birkhoff and John von Neumann: A question of priority and the ergodic theorems, 1931-1932. *Historia Mathematica*, 29:138–156, 2002.
[352] D. Zwillinger. *CRC Standard Mathematical Tables and Formulas*. CRC Press, 33 edition, 2017.

Department of Mathematics, Harvard University, Cambridge, MA, 02138