Internationale Mathematische Nachrichten

International Mathematical News

Nouvelles Mathématiques Internationales

Die IMN wurden 1947 von R. Inzinger als „Nachrichten der Mathematischen Gesellschaft in Wien“ gegründet. 1952 wurde die Zeitschrift in „Internationale Mathematische Nachrichten“ umbenannt und war bis 1971 offizielles Publikationsorgan der „Internationalen Mathematischen Union“.

Von 1953 bis 1977 betreute W. Wunderlich, der bereits seit der Gründung als Redakteur mitwirkte, als Herausgeber die IMN. Die weiteren Herausgeber waren H. Vogler (1978–79), U. Dieter (1980–81, 1984–85), L. Reich (1982–83), P. Flor (1986–99), M. Drmota (2000–2007) und J. Wallner (2008–2017).

Herausgeber:

Österreichische Mathematische Gesellschaft, Wiedner Hauptstraße 8–10/104, A-1040 Wien. email *imn@oemg.ac.at*, *http://www.oemg.ac.at/*

Bezug:

Die IMN erscheinen dreimal jährlich und werden von den Mitgliedern der Österreichischen Mathematischen Gesellschaft bezogen.

Jahresbeitrag: € 35,–

Bankverbindung:
IBAN AT83-1200-0229-1038-9200 bei der Bank Austria-Creditanstalt (BIC-Code BKAUATWW).

Eigentümer, Herausgeber und Verleger: Österr. Math. Gesellschaft. Satz: Österr. Math. Gesellschaft. Druck: Weinitzendruck, 8044 Weinitzen.

ISSN 0020-7926

Österreichische Mathematische Gesellschaft

Gegründet 1903
http://www.oemg.ac.at/
email: *oemg@oemg.ac.at*

Sekretariat:

Alpen-Adria-Universität Klagenfurt,
Institut für Mathematik
Universitätsstraße 65-67
A-9020 Klagenfurt
email: *oemg@oemg.ac.at*

Mitgliedsbeitrag:

Jahresbeitrag: € 35,–

Bankverbindung: IBAN AT83-1200-0229-1038-9200

Internationale Mathematische Nachrichten

International Mathematical News

Nouvelles Mathématiques Internationales

Nr. 244 (74. Jahrgang) **August 2020**

Inhalt

Die Titelseite zeigt den Graphen der Funktion $f(x) = \sin(9x) + \sin(10x)$ im Intervall $[-5, 5]$ geplottet mit Wolfram MATHEMATICA 12. Bekanntermaßen werden solche Kurven verwendet, um Wellen mathematisch zu beschreiben. 2020 ist das “International Year of Sound”, eine globale Initiative, um die Bedeutung des Schalls zu betonen. Mehr Informationen zu dieser Initiative findet man unter https://sound2020.org. Als Beispiel für mathematische Aspekte der Schallforschung sei an dieser Stelle auf den START-Preisträger von 2011 und jetzigen Direktor des Instituts für Schallforschung der Österreichischen Akademie der Wissenschaften, Doz. Dr. Peter Balazs, verwiesen, der Zeit-Frequenz-Analyse, Gabor-Analysen, Numerik, Frame-Theorie, Signalverarbeitung, Akustik und Psychoakustik zu seinen Forschungsinteressen zählt, sowie auf den START-Preisträger von 2019, José Luis Romero von der Universität Wien, der zu den Themen Zeit-Frequenz-Analyse, Zufälligkeit und Abtastung forscht.

Internat. Math. Nachrichten
Nr. 244 (2020), 1–10

Average bounds for the ℓ-torsion in class groups of number fields

Christopher Frei
University of Manchester

1 Introduction

This is a slightly extended account of my plenary lecture given at the ÖMG Conference 2019 in Dornbirn. It is intended as a brief and somewhat informal introduction to an area of mathematics which I find highly interesting, and by no means as a comprehensive survey. I will start by introducing and motivating basic notions of algebraic number theory, in particular those that appear in the title, and then bridge the gap to current questions and some of my own joint work with Martin Widmer (Royal Holloway, University of London).

2 Number fields and factorisation

An *algebraic number field* K is an extension field of $\mathbb{Q}$ that has finite dimension as a $\mathbb{Q}$-vector space. This dimension is called the *degree* of K and written as $[K:\mathbb{Q}]$. The arithmetical structure of the simplest number field, the field $\mathbb{Q}$ itself, is largely governed by its subring $\mathbb{Z}$, the ring of rational integers. For example, one can extend the prime factorisation of integers to $\mathbb{Q}$ by factoring numerator and denominator separately, writing the factors of the denominator with negative exponent. A similar rôle is played in an arbitrary number field K by its *ring of integers*, often denoted by O_K and defined as the set of all elements of K that satisfy a monic polynomial equation with coefficients in $\mathbb{Z}$,

$$O_K = \{\alpha \in K \;:\; \alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0 \text{ for some } n \in \mathbb{N},\ a_i \in \mathbb{Z}\}.$$

Such elements α are called *integral* over the ring $\mathbb{Z}$, a notion that generalises and shares many features with *algebraic* elements over a field. As a simple concrete

ISSN 0020-7926

example, take the number field

$$K = \mathbb{Q}[\sqrt{-5}] = \{a + b\sqrt{-5} \ : \ a, b \in \mathbb{Q}\} \subseteq \mathbb{C}.$$

It has degree $[K : \mathbb{Q}] = 2$, and its ring of integers is the ring

$$O_K = \mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \ : \ a, b \in \mathbb{Z}\}.$$

In general, the ring of integers of a number field is not as simple to write down as suggested by this example, but, additively, it is always a free abelian group of rank $[K : \mathbb{Q}]$.

Number fields and their rings of integers appear naturally in the study of *Diophantine equations*, by which I mean polynomial equations to be solved in integers. One of the most popular examples is *Fermat's last theorem*, which was conjectured by Pierre de Fermat in 1637, and whose proof has eluded mathematicians for more than 350 years, unless it was finally completed by Andrew Wiles in 1994. One way to state it is that the Diophantine equation

$$x^p + y^p = z^p,$$

where p is an odd prime, has no solution with $x, y, z \in \mathbb{Z}$ and $xyz \neq 0$. The quest for its proof has led to many important advances in number theory, in particular the development of algebraic number theory and the modularity of elliptic curves.

An approach followed by nineteenth century mathematicians was as follows: given a solution x, y, z, factor the left-hand side and obtain, with $\zeta_p = e^{2\pi i/p}$ a primitive p-th root of unity, the equation

$$\prod_{k=0}^{p-1} (x + \zeta_p^k y) = z^p.$$

Then one would hope to show, essentially, that the factors $x + \zeta_p^k y$ are coprime in pairs, so their product can only be a p-th power when each of the factors is a p-th power, leading to a contradiction further down the road. This approach has the obvious issue that it tries to apply the fundamental theorem of arithmetic to numbers of the form $x + \zeta_p^k y$, which are clearly not integers. Instead, they are elements of the ring $\mathbb{Z}[\zeta_p]$, and this turns out to be the ring of integers O_K of the number field $K = \mathbb{Q}[\zeta_p]$, the *p-th cyclotomic field*. Unfortunately, an analogue of the fundamental theorem of arithmetic does not always hold in rings of integers of number fields. In more algebraic terms, many of these rings are not *unique factorisation domains* (UFDs). The first counterexample among cyclotomic fields occurs for $p = 23$, but let us look at a simpler example.

In the ring of integers $O_K = \mathbb{Z}[\sqrt{-5}]$ of the number field $K = \mathbb{Q}[\sqrt{-5}]$, we can write

$$2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}), \tag{1}$$

and it is not hard to verify that these are two inequivalent factorisations of 6 into irreducibles. Hence, $\mathbb{Z}[\sqrt{-5}]$ is not a UFD. These issues were already well known to some 19th century mathematicians. To deal with them, Ernst Eduard Kummer introduced the concept of *ideal numbers*, extending the integers of a number field in such a way that every integer has a unique factorisation into *ideal primes*. In the above example, there are *ideal primes* P, Q, R, such that $2 = P^2$, $3 = QR$, $1+\sqrt{-5} = PQ$ and $1-\sqrt{-5} = PR$. Hence, upon further factorisation into ideal primes, the equation in (1) would yield $P^2QR = PQPR$, thus recovering uniqueness.

Richard Dedekind introduced the concept of ideals, one of the fundamental notions of ring theory, to implement Kummer's ideal numbers intrinsically in O_K. In our example, one can take $P = \langle 2, 1+\sqrt{-5}\rangle$, the ideal of $\mathbb{Z}[\sqrt{-5}]$ generated by 2 and $1+\sqrt{-5}$, and, similarly, $Q = \langle 3, 1+\sqrt{-5}\rangle$, $R = \langle 3, 1-\sqrt{-5}\rangle$. Then P, Q, R are prime ideals, and one can factor the principal ideals generated by the numbers in equation (1) as $\langle 2\rangle = P^2$, $\langle 3\rangle = QR$, $\langle 1+\sqrt{-5}\rangle = PQ$, $\langle 1-\sqrt{-5}\rangle = PR$.

In general, the fundamental theorem of arithmetic is replaced by the following central fact of algebraic number theory: all non-zero ideals of the ring of integers O_K of any number field K factor uniquely as products of prime ideals. Integral domains with this property are now called *Dedekind rings*.

3 The class group

Let us try to measure how severely we have changed the problem by passing from factorisations of elements into irreducibles to factorisations of ideals into prime ideals.

To this end, we turn the multiplicative monoid of non-zero ideals of O_K into a group by adding inverses for all ideals. The resulting (abelian) group is the group $\mathcal{J}_K$ of non-zero *fractional ideals* of K. One can define fractional ideals rigorously as finitely generated O_K-submodules of K, but this is of no further relevance for this note. Similarly as with principal ideals, we can also look at the subgroup $P_K = \{\alpha O_K \ : \ \alpha \in K^\times\}$ of *principal fractional ideals*. With these definitions in place, there is a classical exact sequence of abelian groups as follows:

$$1 \to O_K^\times \to K^\times \to \mathcal{J}_K \to \mathcal{J}_K/P_K \to 1, \tag{2}$$

with the map $K^\times \to \mathcal{J}_K$ in the middle given by $\alpha \mapsto \alpha O_K$, sending every element to its principal fractional ideal.

Factorisations into irreducibles of O_K take place in $K^\times$, whereas factorisations into ideals of O_K take place in $\mathcal{J}_K$. The change that occurs when we pass from $K^\times$ to $\mathcal{J}_K$ is described by the kernel $O_K^\times$ and the cokernel $\mathcal{J}_K/P_K$ in (2). The units $O_K^\times$ play only a minor rôle in the context of factorisations, which leaves the cokernel

$\mathrm{Cl}_K = \mathcal{J}_K/P_K$ as the main object of interest. This group is called the *class group* of the number field K.

One of the main theorems of classical algebraic number theory states that Cl_K is always a finite abelian group. Its relation to factorisation in O_K is emphasised by the fact that Cl_K is the trivial group if and only if O_K is a UFD.

Using in addition some facts about units, Kummer was able to prove in 1850 Fermat's last theorem for all exponents p that are odd *regular primes*. These are primes p that do not divide $|\mathrm{Cl}_{\mathbb{Q}[\zeta_p]}|$, the cardinality of the class group of the p-th cyclotomic field. Out of the primes up to 100, all are regular except for $37, 59$ and 67. Kummer claimed that there are infinitely many regular primes, but a proof of this claim has yet to be found. Heuristics suggest that about 61% of all primes are regular.

For more rather easy reading on algebraic number theory, its historical developments, unique and non-unique factorisation in number fields, regular primes and Fermat's last theorem, I recommend the book [21].

4 The ℓ-torsion conjecture

As the class group Cl_K is known to be finite, it is of interest to bound its size. This is usually done in terms of the degree $d = [K : \mathbb{Q}]$ and another numerical invariant $D_K = |\Delta_K|$ of K, defined as the absolute value of its discriminant Δ_K. Instead of defining the discriminant Δ_K rigorously, let me just say that this quantity is an integer that measures, in a certain sense, the "arithmetical complexity" of the number field K, so number fields with larger discriminant can be thought of as being more complex. For a quadratic number field $K = \mathbb{Q}(\sqrt{n})$, with squarefree $n \in \mathbb{Z} \smallsetminus \{0, 1\}$, we have

$$\Delta_K = \begin{cases} n & \text{if } n \equiv 1 \bmod 4, \\ 4n & \text{if } n \equiv 2, 3 \bmod 4. \end{cases} \tag{3}$$

With these notions in place, Edmund Landau [15] has proved that the class group of every number field K of degree d satisfies the asymptotic bound

$$|\mathrm{Cl}_K| \ll_{d,\varepsilon} D_K^{1/2+\varepsilon}, \tag{4}$$

for any $\varepsilon > 0$. The notation $\ll_{d,\varepsilon}$ is ubiquitous in analytic number theory and means that there is a constant $C(d, \varepsilon)$, depending only on d and ε, such that $|\mathrm{Cl}_K| \leq C(d, \varepsilon) D_K^{1/2+\varepsilon}$.

Landau's bound (4) is essentially best possible. However, in applications it is often not necessary to bound the size of the full class group; instead it is enough to consider its ℓ-torsion part

$$\mathrm{Cl}_K[\ell] = \{C \in \mathrm{Cl}_K \;:\; C^\ell = P_K\}$$

for some $\ell \in \mathbb{N}$. The *ℓ-torsion conjecture* states that there is a much better bound than Landau's for the ℓ-torsion part, namely

$$|\,\mathrm{Cl}_K[\ell]| \ll_{d,\ell,\varepsilon} D_K^{\varepsilon}. \tag{5}$$

This conjecture, which plays an important rôle in modern number theory, was to my best knowledge mentioned for the first time by Armand Brumer and Joseph Silverman [5] in the context of counting elliptic curves, and by William Duke [9] in connection with discriminant multiplicities.

It might seem somewhat counterintuitive that one expects every ℓ-torsion to be so small when it is known that the full group can be essentially as large as in Landau's bound (4). However, this expectation fits well together with heuristic models for the distribution of class groups. In fact, very recent work of Lillian Pierce, Caroline Turnage-Butterbaugh and Melanie Matchett Wood [19] demonstrates that the ℓ-torsion conjecture is implied by other standard conjectures in number theory; it would follow (individually) from the Cohen-Lenstra and Cohen-Martinet heuristics on the distribution of class groups, from the discriminant multiplicity conjecture, and from a modified version of Malle's conjecture on the distribution of number fields (see later).

Class groups of quadratic fields are intimately connected to binary quadratic forms over the integers, and Carl Friedrich Gauß's genus theory for these quadratic forms immediately implies the ℓ-torsion conjecture in case $d = 2$, $\ell = 2^e$ for $e \in \mathbb{N}$. These are, however, the only cases in which the full conjecture is known to hold. Generalisations of Gauß's genus theory can handle other cases as well, but under additional restrictions on K.

With the full strength of the conjecture known in only very few cases, it is of interest to obtain nontrivial bounds towards it. Here, "nontrivial" refers to the observation that, appealing to Landau's result (4), one always has

$$|\,\mathrm{Cl}_K[\ell]| \leq |\,\mathrm{Cl}_K\,| \ll_{d,\varepsilon} D_K^{1/2+\varepsilon}. \tag{6}$$

Much effort was spent on the case $d = 2, \ell = 3$, initiated by Lillian Pierce [18] and, independently, Harald Helfgott and Akshay Venkatesh [13]. The strongest result available today in this case is

$$|\,\mathrm{Cl}_K[3]| \ll_{\varepsilon} D_K^{1/3+\varepsilon}, \tag{7}$$

due to Jordan Ellenberg and Venkatesh [11]. In the same paper, Ellenberg and Venkatesh also proved a *conditional* result valid for all values of d and ℓ, but assuming a version of the *Generalised Riemann Hypothesis*, more precisely the Riemann Hypothesis for Dedekind zeta functions of number fields. Under this strong assumption, they proved that

$$|\,\mathrm{Cl}_K[\ell]| \ll_{d,\ell,\varepsilon} D_K^{1/2-1/2\ell(d-1)+\varepsilon}, \tag{8}$$

which is often referred to as the *GRH-bound*. Note that for large values of ℓ or d, it is unfortunately much closer to the trivial bound (6) than to the conjectured bound (5).

5 Counting number fields

Given a conditional result such as the GRH-bound (8), analytic number theorists often try to obtain unconditional weaker versions, for example averaged results. To make clear what is meant by averaging over number fields, consider the function

$$N_d(X) = \#\{K/\mathbb{Q}\ :\ [K:\mathbb{Q}] = d,\ D_K \le X\},$$

counting number fields of fixed degree d and discriminant bounded by X in absolute value. It was already shown by Charles Hermite that $N_d(X) < \infty$ for all degrees d and $X > 0$, and a folklore conjecture predicts that this function has linear asymptotics for large X,

$$N_d(X) \sim c_d X, \qquad \text{as } X \to \infty, \tag{9}$$

with a constant c_d. This conjecture is widely open for all degrees $d \ge 6$, but it was shown to hold for small degrees $d \le 5$, using explicit parameterisations of number fields by arithmetical objects represented by lattice points, and using the geometry of numbers to count these lattice points.

In case $d = 2$, this is simple; every quadratic field is generated by the square-root of its discriminant, $K = \mathbb{Q}(\sqrt{\Delta_K})$. Hence, quadratic fields are parameterised by their discriminants, squarefree integers that satisfy certain congruence conditions coming from (3).

As a first idea to extend this to higher degrees, one could try to associate to a number field $K = \mathbb{Q}(\alpha)$ of degree d the minimal polynomial $f = a_d x^d + \cdots + a_0$ of its generator α, and thus the $(d+1)$-tuple $(a_0, \ldots, a_d) \in \mathbb{Z}^{d+1}$. However, while this approach gives good parameterisations for the algebraic numbers α, it falls short of achieving what we hope for, as many algebraic numbers α will generate the same fields $\mathbb{Q}(\alpha)$.

For $d = 3$, one can instead use a beautiful one-to-one correspondence, due to Boris Delone and Dmitriĭ Faddeev [8] between cubic rings (up to isomorphism) and equivalence classes of integral binary cubic forms,

$$\{\text{cubic rings}\}/\simeq \quad \longleftrightarrow \quad \mathrm{Sym}^3\,\mathbb{Z}^2/\,\mathrm{GL}_2(\mathbb{Z}). \tag{10}$$

Cubic rings are unitary commutative rings whose additive groups are free abelian of rank three, so this includes in particular the rings of integers O_K of cubic number fields K. In fact, in the correspondence (10), rings of integers O_K are identified with irreducible cubic forms that satisfy certain congruence conditions, and, moreover, the discriminant Δ_K of the number field K is equal to the discriminant of the corresponding form. The verification of (9) in case $d = 3$ is a seminal result of Harold Davenport and Hans Heilbronn [7].

For degrees 4 and 5, Manjul Bhargava [1, 2, 3, 4] found more intricate parameterisations, leading to the verification of the conjecture (9) in these cases. Unfortunately, these techniques do not seem to extend to degrees 6 and higher.

6 Average bounds for ℓ-torsion

With the basic quantitative behaviour of number fields of bounded discriminant established, at least for degrees $d \leq 5$, one can now study quantities like $|\mathrm{Cl}_K[\ell]|$ on average. Jordan Ellenberg, Lillian Pierce and Melanie Matchett Wood [10] succeeded in proving an unconditional average bound of the same quality as the GRH-bound (8). Their result implies that, for $d \leq 5$ and sufficiently large values of ℓ,

$$\frac{1}{N_d(X)} \sum_{\substack{[K:\mathbb{Q}]=d \\ D_K \leq X}}^{*} |\mathrm{Cl}_K[\ell]| \ll_{\ell,\varepsilon} X^{1/2-1/(2\ell(d-1))+\varepsilon}. \tag{11}$$

The asterisk above the sum indicates that, in case $d = 4$, there is an additional technical condition on the number fields under consideration, which I will not describe here. Average bounds, conditional and unconditional, for many more families of number fields were recently obtained in groundbreaking work by Lillian Pierce, Caroline Turnage-Butterbaugh and Melanie Matchett Wood [20].

A key ingredient used to prove many of the available bounds for the ℓ-torsion in class groups, including the bound (7), the GRH-bound (8), the average version (11), and the many new results from [20], is the following lemma from [11]. We say that a prime p *splits completely* in a number field K, if the principal ideal of O_K generated by p factors into $d = [K : \mathbb{Q}]$ distinct prime ideals, $pO_K = P_1 \cdots P_d$, which is the maximum possible.

Lemma 1 (Key lemma)**.** *Let K be a number field and $d = [K : \mathbb{Q}]$. If M primes $p_1, \ldots, p_M \leq D_K^{\delta/2(d-1)}$ split completely in K, for some $\delta < 1/\ell$, then*

$$|\mathrm{Cl}_K[\ell]| \ll_{d,\delta,\varepsilon} D_K^{1/2+\varepsilon} M^{-1}.$$

Hence, one can get savings over the trivial bound (6) on $|\mathrm{Cl}_K[\ell]|$ by showing the existence of many small primes that split completely. Quite a lot is known about the distribution of split primes in a given number field K, in particular they have positive density amongst all primes by the *Chebotarev density theorem*. A standard conditional version of this theorem is enough to deduce the GRH-bound (8) from Lemma 1, and a strong unconditional version in families of number fields is the main innovation of [20].

In joint work with Martin Widmer [12], we have found a refined version of the key lemma, in which the bound $D_K^{\delta/2(d-1)}$ for the primes is replaced by $\eta_\ell(K)^\delta$, with a new specialised invariant $\eta_\ell(K)$ that is defined as follows:

$$\eta_\ell(K) = \inf\left\{H_K(\alpha) : K = \mathbb{Q}[\alpha],\ \alpha O_K = (P_1/P_2)^\ell, P_1 \neq P_2 \text{ linear prime ideals}\right\}.$$

Here, $H_K(\alpha)$ is the Weil height relative to K, a standard measure for the "arithmetical complexity" of a number in K. Hence, $\eta_\ell(K)$ is the minimal height of a generator α of K for which the factorisation of the principal fractional ideal αO_K into prime ideals has a specified shape. With this invariant, we get the following version of the key lemma.

Lemma 2 (Key lemma)**.** *Let K be a number field and $d = [K : \mathbb{Q}]$. If M primes $p_1, \dots, p_M \leq \eta_\ell(K)^\delta$ split completely in K, for some $\delta < 1/\ell$, then*

$$|\mathrm{Cl}_K[\ell]| \ll_{d,\delta,\varepsilon} D_K^{1/2+\varepsilon} M^{-1}.$$

Here one should note that one always has $\eta_\ell(K) \geq D_K^{1/2(d-1)}$, so the bound in Lemma 2 is never worse than the bound in the original Lemma 1. However, we could show that for many families of number fields, $\eta_\ell(K)$ is much bigger than $D_K^{1/2(d-1)}$ on average. This comes essentially from comparing asymptotics for number fields of bounded discriminant with asymptotics for algebraic numbers of bounded height, hence from playing the two approaches to counting number fields mentioned in the previous section against each other.

Lemma 2 leads to stronger average bounds for many families of number fields. For example we have obtained the following result improving upon (11): for $d \leq 5$ and large enough ℓ, we have

$$\frac{1}{N_d(X)} \sideset{}{^*}\sum_{\substack{[K:\mathbb{Q}]=d \\ D_K \leq X}} |\mathrm{Cl}_K[\ell]| \ll_{\ell,\varepsilon} X^{1/2-1/(\ell(d-1)+3)+\varepsilon}. \tag{12}$$

Our work has an immediate application to certain refined questions concerning the distribution of number fields. Instead of looking at asymptotics for all number fields K of degree d and bounded discriminant, as in (9), one can restrict to subfamilies in which the normal closure K^{nc} of K has a prescribed Galois group $G \subset S_d$, by studying the counting function

$$N_d(G;X) := \#\{K/\mathbb{Q} \ :\ [K : \mathbb{Q}] = d,\ \mathrm{Gal}(K^{\mathrm{nc}}/\mathbb{Q}) \simeq G,\ D_K \leq X\}.$$

Gunter Malle [16, 17] has formulated conjectures about the asymptotic behaviour of this counting function, which in their weaker form state that

$$X^{a(G)} \ll_d N_d(G;X) \ll_{d,\varepsilon} X^{a(G)+\varepsilon}. \tag{13}$$

Here, $\varepsilon > 0$ is arbitrarily small and $a(G)$ is an invariant of the permutation group $G \subset S_d$, defined as $a(G) = \min\{\mathrm{ind}(g) \ : \ g \in G \smallsetminus \{\mathrm{id}\}\}^{-1}$, where $\mathrm{ind}(g)$ is the *index* of the permutation g, that is d minus the number of orbits of g on $\{1,\ldots,d\}$. The most compelling evidence for Malle's conjectures is provided by the case where G is an abelian group. In this case, they are known to be true by a result of David Wright [22], preceding their formulation. In most other cases, Malle's conjecture is open. It is believed to be extremely hard, as in particular it predicts the existence of number fields with any given Galois group, and thus a positive answer to the inverse Galois problem.

In the case where $d = p$ is prime and $G = D_p$, the dihedral group of order $2p$, one has $a(D_p) = 2/(p-1)$. Here, Jürgen Klüners [14] expressed the counting function $N_p(D_p;X)$ in terms of averages of $|\mathrm{Cl}_K[p]|$ over quadratic fields K, and in particular succeeded in proving the lower bound in (13). Average bounds for the p-torsion in quadratic fields translate, via Klüners's method, directly to upper bounds for $N_p(D_p;X)$; in particular, the ℓ-torsion conjecture (5), for $d = 2$ and $\ell = p$, would imply the upper bound in (13).

As an application of the case $d = 2$ of my work with Martin Widmer, we have obtained the currently best known upper bound

$$N_p(D_p,X) \ll_{\varepsilon,p} X^{3/(p-1)-2/(p+2)(p+1)+\varepsilon},$$

improving upon the earlier record

$$N_p(D_p,X) \ll_{\varepsilon,p} X^{3/(p-1)-1/p(p-1)+\varepsilon},$$

which Henri Cohen and Frank Thorne [6] had obtained as a consequence of (11).

References

[1] M. Bhargava, *Higher composition laws. III. The parametrization of quartic rings*, Ann. of Math. (2) **159** (2004), no. 3, 1329–1360.

[2] M. Bhargava, *The density of discriminants of quartic rings and fields*, Ann. of Math. **162** (2005), 1031–1063.

[3] M. Bhargava, *Higher composition laws. IV. The parametrization of quintic rings*, Ann. of Math. (2) **167** (2008), no. 1, 53–94.

[4] M. Bhargava, *The density of discriminants of quintic rings and fields*, Ann. of Math. **172** (2010), 1559–1591.

[5] A. Brumer and J. H. Silverman, *The number of elliptic curves over* $\mathbf{Q}$ *with conductor* N, Manuscripta Math. **91** (1996), no. 1, 95–102.

[6] H. Cohen and F. Thorne, *On* D_ℓ*-extensions of odd prime degree* ℓ, arXiv:1609.09153 (2017).

[7] H. Davenport and H. Heilbronn, *On the density of discriminants of cubic field extensions. II*, Proc. London. Math. Soc. **322** (1971), 405–420.

[8] B. N. Delone and D. K. Faddeev, *Theory of Irrationalities of Third Degree*, Acad. Sci. URSS. Trav. Inst. Math. Stekloff, **11** (1940), 340.

[9] W. Duke, *Bounds for arithmetic multiplicities*, Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998), no. Extra Vol. II, 1998, pp. 163–172.

[10] J. Ellenberg, L. B. Pierce, and M. M. Wood, *On* ℓ*-torsion in class groups of number fields*, Algebra and Number Theory **11-8** (2017), 1739–1778.

[11] J. Ellenberg and A. Venkatesh, *Reflection principles and bounds for class group torsion*, Int. Math. Res. Not. no.1, Art. ID rnm002 (2007).

[12] C. Frei and M. Widmer, *Averages and higher moments for the* ℓ*-torsion in class groups*, arXiv:1810.04732 (2018).

[13] H. A. Helfgott and A. Venkatesh, *Integral points on elliptic curves and 3-torsion in class groups*, J. Amer. Math. Soc. **19** (2006), 527–550.

[14] J. Klüners, *Asymptotics of number fields and the Cohen-Lenstra heuristics*, J. Théor. Nombres Bordeaux **18** (2006), 607–615.

[15] E. Landau, *Abschätzungen von Charaktersummen, Einheiten und Klassenzahlen*, Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl. Jg. 1918, 79–97.

[16] G. Malle, *On the distribution of Galois groups*, J. Number Theory **92** (2002), 315–329.

[17] G. Malle, *On the distribution of Galois groups. II*, Experiment. Math. **13** (2004), no. 2, 129–135.

[18] L. B. Pierce, *3-part of class numbers of quadratic fields*, J. London Math. Soc. **71** (2005), 579–598.

[19] L. B. Pierce, C. Turnage-Butterbaugh, and M. M. Wood, *On a conjecture for* ℓ*-torsion in class groups of number fields: from the perspective of moments*, arXiv:1902.02008 (2019).

[20] ———, *An effective Chebotarev density theorem for families of number fields, with an application to* ℓ*-torsion in class groups*, Invent. Math. (to appear).

[21] I. Stewart and D. Tall, *Algebraic number theory and Fermat's last theorem*, fourth ed., CRC Press, Boca Raton, FL, 2016.

[22] D. J. Wright, *Distribution of discriminants of abelian extensions*, Proc. London Math. Soc. (3) **58**, no. 1 (1989), 17–50.

Author's address:
Department of Mathematics, The University of Manchester, Oxford Road, Manchester M13 9PL
email christopher.frei@manchester.ac.uk

Internat. Math. Nachrichten
Nr. 244 (2020), 11–26

Self-avoiding walks and their languages

Christian Lindorfer
TU Graz

This is an introductory review of recent work on self-avoiding walks on thin-ended quasi-transitive graphs. The goal is not to give precise proofs, but instead show the main ideas by giving simple examples.
Our main interest lies in two different kinds of formal languages induced by the set of all self-avoiding walks on the graph starting at a given origin. The language of self-avoiding walks on a deterministically edge-labelled graph consists of all words obtained by reading along self-avoiding walks. We discuss that this language is regular or context-free, if and only if the given graph has ends of size at most 1 or 2, respectively.
Using tree-decomposition, thin ended graphs can be decomposed into finite parts. Configurations are appearances of self-avoiding walks on single parts, and compatible configurations on the composition tree correspond to self-avoiding walks on the original graph. By encoding compatible configurations as words we obtain the language of configurations. This language is context-free and thus the generating function of self-avoiding walks is algebraic.

1 Introduction

Consider the following discrete process: Start at a given vertex o of a simple (undirected) graph, possibly containing infinitely many vertices. In each step move along an edge incident to the current vertex and arrive at its second endpoint. The only rule is that no vertex of the graph may be visited twice. Such a walk is often called path, here we will use the notion of *self-avoiding walk*, abbreviated with *SAW*. One natural question arising from this process is,

"For a given number of steps n, how many different SAWs of length n are there?"

ISSN 0020-7926

We will denote this number of SAWs of length n starting at o by $c_n(o)$. Obviously this question only makes sense if G is *locally finite*, i.e., every vertex has finitely many neighbours. Additionally we may assume that G is connected, otherwise we just consider the component of G containing o. In some graphs $c_n(o)$ can be easily calculated for every n. The most basic example is the infinite k-regular tree, where every vertex has exactly k neighbours. Independently of the choice of the starting vertex o there are k possible directions for the first step. Backtracking is not allowed, so there are at most $k-1$ choices for every consecutive step and all of them are valid because trees are cycle-free. Thus we obtain $c_n(o) = k(k-1)^{n-1}$ for any choice of o. While the calculations in this example are simple, they seem to be very difficult for most graphs.

In 1953 the famous chemist and Nobel laureate PAUL J. FLORY [9] introduced SAWs as a model for long-chain polymer molecules. Although these chains live in the continuum, in many cases a lattice approximation is good enough. The self avoidance of the walk models the excluded volume effect, namely that no two monomers can occupy the same position in space. The most important lattices for practical applications are the d-dimensional integer lattices $\mathbb{Z}^d$ for $d \geq 2$, where every vertex is connected to the $2d$ vertices at Euclidean distance 1. Many interesting results about SAWs on these lattices can be found in the book *The Self-avoiding Walk* by MADRAS and SLADE [20].

Our main interest lies in studying SAWs not only on lattices but on the more general class of quasi-transitive graphs, defined as follows: An *automorphism* of a graph G is a bijection on the vertex set of G mapping edges onto edges and non-edges onto non-edges. The group of graph automorphisms of G is denoted by $\mathsf{AUT}(G)$. A graph is *quasi-transitive* if there is a finite subset U of vertices of G such that for any vertex v of G there are a vertex $u \in U$ and a graph automorphism $\gamma \in \mathsf{AUT}(G)$ such that $\gamma(u) = v$. For a given vertex v of G and any subgroup $\Gamma \leq \mathsf{AUT}(G)$ the set $\{\gamma(v) \mid \gamma \in \mathsf{AUT}(G)\}$ is called the orbit of v under the action of Γ. Using this notation, quasi-transitivity of G means that $\mathsf{AUT}(G)$ acts with finitely many orbits on G. The graph G is called *transitive* if there is only a single orbit, so a subset U of V consisting of a single vertex can be chosen.

It seems natural to ask about the asymptotic behaviour of the number of SAWs of length n for n going to infinity, leading to the connective constant of a graph. The following theorem from 1957 is due to HAMMERSLEY [16].

Theorem 1. *Let G be an infinite quasi-transitive graph. Then there is a constant $\mu(G) \in [1, \infty)$, called* connective constant *of G, such that for every vertex o of G*

$$\lim_{n \to \infty} c_n^{1/n}(o) = \mu(G).$$

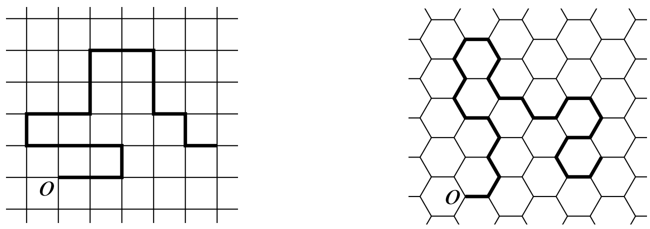
While its proof is technical in the quasi-transitive case, it is rather simple for transitive graphs because $c_n(o)$ does not depend on the choice of o. Any SAW of length $m+n$ can be decomposed into a SAW of length m and a SAW of length n. For transitive graphs this implies that the sequence $(c_n(o))_{n\in\mathbb{N}_0}$ is submultiplicative, i.e. $c_{n+m}(o) \leq c_n(o)c_m(o)$. Application of Fekete's Subadditive Lemma to the sequence $(\log c_n(o))_{n\in\mathbb{N}_0}$ yields the desired result.

As usual in enumerative combinatorics, we use generating functions for counting walks. The ordinary generating function of SAWs starting at o is

$$F_{SAW,o}(t) = \sum_{n\in\mathbb{N}_0} c_n(o)t^n.$$

Note that by Cauchy-Hadamard's formula the connective constant $\mu(G)$ is the reciprocal of the radius of convergence of $F_{SAW,o}(t)$.

While physicists provided good estimates for the connective constant of many lattices, their precise values remain unknown for almost all lattices of dimension at least 2, including the integer lattices $\mathbb{Z}^d$. A highlight in the theory of SAWs is the remarkable paper by DUMINIL-COPIN and SMIRNOV [6] in 2012 containing the first rigorous calculation of the connective constant of the honeycomb lattice $\mathbb{H}$ (see Figure 1).



$\mu(\mathbb{Z}^2) \approx 2.6381585$ $\quad$ $\mu(\mathbb{H}) = \sqrt{2+\sqrt{2}}$

Figure 1: SAWs and connective constants of the lattices $\mathbb{Z}^2$ and $\mathbb{H}$.

2 Regular and context-free languages

Our goal is to relate SAWs with formal language theory, so we give a brief introduction into the necessary theory of regular and context-free languages. These two classes of languages build the first two layers of the Chomsky-Hierarchy, an important containment hierarchy in formal language theory. A good source for additional information is the book *Introduction to Formal Language Theory* by HARRISON [17].

An *alphabet* Σ is a finite set of elements called *letters* and we write

$$\Sigma^* = \{a_1a_2\ldots a_n : n \geq 0, a_i \in \Sigma\}$$

for the set of finite *words* over the alphabet Σ. Let $|w|$ denote the length (number of letters) of a word w and ε stand for the empty word of length 0. A *(formal) language* over Σ is a subset of Σ^*.

A *context-free grammar* is a quadruple $\mathcal{C} = (N, \Sigma, P, S)$ consisting of a finite set of *non-terminals* N (with $N \cap \Sigma = \emptyset$), an alphabet Σ, a finite set of *(production) rules* $P \subseteq V \times (V \cup \Sigma)^*$ and a *start symbol* $S \in N$. We use the notation $A \vdash \alpha$ to indicate that $(A, \alpha) \in P$ and $A \vdash \alpha_1 \mid \cdots \mid \alpha_n$ to indicate the existence of n rules $A \vdash \alpha_1, \ldots, A \vdash \alpha_n$.

Production rules should be seen as valid ways of replacing non-terminals in strings. The rule $A \vdash \alpha$ means that a non-terminal A in any string $\beta = \beta_1 A \beta_2 \in (N \cup \Sigma)^*$ can be replaced by α to obtain a new string $\gamma = \beta_1 \alpha \beta_2$. We write $\beta \Rightarrow \gamma$ and say β *generates* γ if γ can be obtained from β by replacing the leftmost non-terminal according to some rule in P. A *leftmost derivation* is a sequence $(\alpha = \alpha_0, \alpha_1, \ldots, \alpha_k = \beta)$ such that $\alpha_{i-1} \Rightarrow \alpha_i$ for every i; we denote the existence of such a sequence by $\alpha \overset{*}{\Rightarrow} \beta$. Then each non-terminal $A \in N$ generates a language $L_A = \{w \in \Sigma^* : A \overset{*}{\Rightarrow} w\}$. In particular, the language generated by the grammar $\mathcal{C}$ is the set L_S of all words which can be derived from the start symbol S.

A context-free grammar is called *regular*, if every production rule has the form

$$A \vdash \alpha B \quad \text{or} \quad A \vdash \alpha, \quad \text{where} \quad A, B \in N,\ \alpha \in \Sigma^*.$$

Moreover, it is *unambiguous*, if there is at most one leftmost derivation $(S, \alpha_1, \ldots, \alpha_{k-1}, w)$ for every word $w \in \Sigma^*$. A language L is called *regular* or *unambiguous*, if there is a context-free grammar of the respective type generating L. It is not hard to show that any regular language is unambiguous.

It is standard to associate a language L with its ordinary generating function $F_L(t) = \sum_{w \in L} t^{|w|}$. A famous result of CHOMSKY and SCHÜTZENBERGER [4] states that the generating function of any language generated by an unambiguous context-free grammar is algebraic over $\mathbb{Q}$. In other words, for every unambiguous context-free language L there is an irreducible polynomial $P(x, y)$ in two variables with coefficients in $\mathbb{Q}$ such that $P(F_L(t), t) = 0$. The main idea in the proof of this statement is to translate the production rules to obtain a single polynomial equation for every non-terminal in the grammar and reduce this system to a single polynomial equation using elimination theory. In the case where L is regular the obtained system of equations is linear, so the resulting generating function $F_L(t)$ is rational.

3 Edge-labelled graphs and the language of SAWs

In combinatorial and geometric group theory Cayley graphs are an important tool for studying the structure of groups. All groups considered here are finitely generated, and the inverse of a group element x is denoted $\bar{x}$. Let S be a finite symmetric (i.e. $S = \bar{S}$) generating set of a group Γ not containing the neutral element 1_Γ. Then the *Cayley graph* $G(\Gamma, S)$ of Γ with respect to S has vertex set Γ and two vertices x and y are connected by an edge whenever $\bar{x}y \in S$. Note that Cayley graphs do not contain loops ($1_\Gamma \notin S$) or multiple edges, are connected (S generates Γ), locally finite (S is finite) and moreover Γ acts transitively on $G(\Gamma, S)$. Connective constants of transitive graphs and in particular Cayley graphs have been intensively studied by various authors including GRIMMETT and LI [11].

Cayley graphs come with a natural edge-labelling. It is a common concept to treat every edge $\{x, y\}$ as a pair of oriented edges (x, y), (y, x) with opposite orientation. Using the set of generators S as an alphabet, every oriented edge (x, xs) of $G(\Gamma, S)$ is labelled with its corresponding generator $s \in S$. Let us extend this concept to obtain a larger class of graphs. An *edge-labelled* graph (G, ℓ) is a graph G together with a label function ℓ attaching to every oriented edge a label of some alphabet Σ. We call such an edge-labelling *deterministic*, if different edges starting or ending at a common vertex have different labels attached. Let us denote by $\mathsf{AUT}(G, \ell)$ the set of *label preserving graph automorphisms* on (G, ℓ), that is the set of graph automorphisms τ such that $\ell(x, y) = \ell(\tau x, \tau y)$ for every oriented edge (x, y) of G. As before we call a deterministically edge-labelled graph *quasi-transitive*, if $\mathsf{AUT}(G, \ell)$ acts with finitely many orbits on G and *transitive*, it there is exactly one orbit.

Starting with an edge-labelled graph (G, ℓ) with label alphabet Σ, we extend the label function ℓ to walks. To a walk $\pi = (e_1, e_2, \ldots, e_n)$ on G given by its sequence of edges e_i we define

$$\ell(\pi) = \ell(e_1)\ell(e_2)\ldots\ell(e_n) \in \Sigma^*.$$

With this definition any set Π of walks on (G, ℓ) corresponds to some language $\ell(\Pi) = \{\ell(\pi) : \pi \in \Pi\}$. Note that if all walks in Π start at a common vertex o, a deterministic labelling guarantees that ℓ bijectively maps Π to its corresponding language $\ell(\Pi)$. This is true because any walk π can be uniquely reconstructed from the sequence of letters of its word $\ell(\pi)$ by starting at o and following the edges labelled with these letters.

The *language of self-avoiding walks* on (G, ℓ) is defined as $L_{SAW,o} = \ell(\Pi_{SAW,o})$, where $\Pi_{SAW,o}$ denotes the set of SAWs starting at the vertex o. As the length of every walk π coincides with the length of the word $\ell(\pi)$, the corresponding generating functions coincide,

$$F_{SAW,o}(t) = F_{L_{SAW,o}}(t).$$

Let us visualise the previous discussion by looking at an easy example. Consider the direct product $\Gamma_L = \mathbb{Z} \times \mathbb{Z}_2$ of the infinite cyclic group and the cyclic group of order 2. A generic symmetric generating set is $S = \{a, \bar{a}, b\}$, where a and its inverse $\bar{a}$ have infinite order, b has order 2 and the relation $ab = ba$ holds.

Figure 2 illustrates the SAW π corresponding to the word $\ell(\pi) = \bar{a}baabaaab\bar{a}$ on the edge-labelled graph (G_L, ℓ_L), where $G_L = G(\Gamma_L, S)$ is the Cayley graph of Γ_L with respect to S and ℓ_L is its inherited edge-labelling.

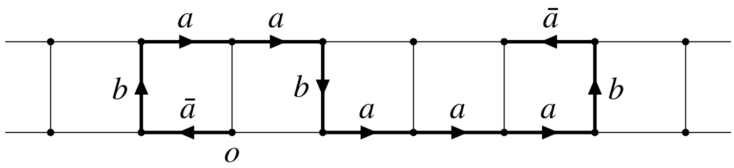


Figure 2: The SAW π on the edge-labelled Cayley graph (G_L, ℓ_L).

The interplay between walks on edge-labelled graphs and their respective languages has been studied in many variations. For any fixed vertex o of an edge-labelled Cayley graph $(G(\Gamma, S), \ell)$, the language $\ell(\Pi_{o,o})$ corresponding to the set of all closed walks $\Pi_{o,o}$ starting and ending at o is called the *word problem* of the group Γ. ANISIMOV [2] showed that this language is regular if and only if the group Γ is finite and this result can be extended to quasi-transitive deterministically edge-labelled graphs. Furthermore in a celebrated work MULLER and SCHUPP [21] proved that the word problem is context-free if and only if the group Γ is virtually free, i.e., it contains a free subgroup of finite index. This is equivalent to the corresponding Cayley graph being tree-like, as defined in the upcoming section.

4 From one-dimensional lattices to tree-like graphs

SAWs are well understood on the one-dimensional lattices. For our purpose a one dimensional lattice is a connected infinite graph G such that there is a subgroup $\Gamma \leq \mathsf{AUT}(G)$ isomorphic to $\mathbb{Z}$ acting quasi-transitively on G. A very good reference for the behaviour of SAWs on one-dimensional lattices is ALM and JANSON [1]. Their ideas can be seen as the basis for our own work, so we want to briefly describe them here. First observe that it is possible to write the vertex set of G as $\mathbb{Z} \times \{1, \ldots, k\}$, such that each of the k orbits of the action of Γ on G has the form $\mathbb{Z} \times \{i\}$ for some $i \in \{1, \ldots, k\}$. A vertex (m, a) is said to have *longitude m.* By grouping successive longitudes into a single new longitude we can always assume that edges of G occur only between vertices of successive longitudes.

Every lattice G represented as above can be decomposed into a sequence of *segments* of identical shape, where each segment consists of all vertices of the same longitude, edges between these vertices and also edges linking the segment to the previous and the subsequent segment. Any SAW π can be seen as a sequence $(c_0, c_1, \ldots, c_m)$ of *configurations*, which are appearances of π on the segments of G. By smartly defining the notion of *compatibility* of two configurations, the authors also obtain the converse statement and thus a bijection between SAWs and compatible sequences of configurations. The sequence of configurations corresponding to the walk π in Figure 2 is shown in Figure 3, numbers above walk-fragments correspond to their order in π.

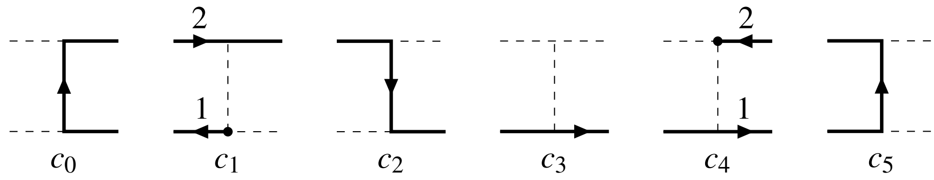


Figure 3: Sequence of configurations corresponding to the SAW π on G_L.

Sequences of configurations can be constructed from left to right; the configuration c_i appended in step i to the sequence $(c_0,\ldots,c_{i-1})$ just needs to be compatible with c_{i-1}. In the notion of formal language theory this means that the language of compatible configurations obtained by treating sequences of configurations as words over the alphabet of configurations is regular. Using standard methods from enumerative combinatorics, the generating function of SAWs on G can be obtained by solving a system of linear equations, whose coefficients are polynomials in the single variable t. As a conclusion the generating function of SAWs is rational and the connective constant of G is an algebraic number.

The foundation of this result is the idea of decomposing one-dimensional lattices along finite sets of edges to obtain finite segments, such that translations map segments onto each other. While such a thing cannot be achieved for lattice graphs of dimension at least 2, quasi-transitive infinite trees can be decomposed in a straight-forward way by taking single vertices and all their incident edges as segments. The main reason for this difference lies in the end-structures of these classes of graphs.

The notion of ends of infinite graphs was introduced by HALIN [14] and - without graph terminology - earlier by FREUDENTHAL [10] and can be summarised as follows: A *ray* in a graph G is a one-way infinite SAW. For a finite set K of vertices of G denote by $G-K$ the graph obtained by removing K and all edges incident to K from G. Two rays in G are called equivalent, if for any finite set K of vertices all but finitely many vertices of the rays are contained in the same connective component of $G-K$, or in other words, if G contains infinitely many disjoint paths connecting the two rays. An *end* of the graph G is an equivalence class of this equivalence relation on the set of rays. Informally, the different ends of a graph can be seen as different directions of moving to infinity. Ends can be classified by their *size*, that is the maximal number of pairwise vertex disjoint rays contained in the end. Note that this number can also be infinite; in this case the end is called *thick*, whereas an end of finite size is called *thin*.

It is well-known that infinite quasi-transitive graphs have either one, two or infinitely many ends, see FREUDENTHAL [10]. Lattice graphs of dimension at least two are standard examples for one-ended graphs. Let us consider the lattice $\mathbb{Z}^2$ as an example. Removing any finite set K of vertices, the resulting graph $\mathbb{Z}^2 - K$ has only one infinite component, so all but finitely many vertices of any given ray must lie in this component. Thus $\mathbb{Z}^2$ has only a single end. Obviously there are infinite sets of disjoint rays, so the single end has infinite size. In fact, the single end of any quasi-transitive one-ended graph is thick.

On the other hand, any lattice of dimension one has two ends. Representing the vertex set of a one-dimensional lattice as $\mathbb{Z} \times \{1, \dots, k\}$ as above, rays can be distinguished by whether the sequence of longitudes converges to $+\infty$ or $-\infty$, so the two ends are inherited from $\mathbb{Z}$. As a specific example let us mention the graph G_L (see Figure 2), where both ends have size two. In fact, the two ends of any quasi-transitive two-ended graph are thin and have the same size.

Finally, graphs with infinitely many ends always have thin ends. Easy examples are infinite trees, all of their ends have size one.

As mentioned before, thick ends seem difficult to handle, so one might get the idea to first try finding more general results for graphs having only thin ends. WOESS [25] showed that these are exactly those graphs which are quasi-isometric to a tree and called them *tree-like* graphs. Tree-like Cayley graphs are precisely those whose underlying groups are virtually free.

5 The language of self-avoiding walks on tree-like graphs with small end-size

The goal of this section is to develop a characterisation result for languages of SAWs on different types of tree-like graphs. While the rigorous proofs are long and technical, the main ideas behind them can be seen by studying some simple examples. Most of the upcoming theory was developed in a paper by LINDORFER and WOESS [19].

Let us consider the edge-labelled graph $(G_\triangle, \ell_\triangle)$ from Figure 4. One way to obtain this graph is to study the free product $\Gamma_\triangle = \mathbb{Z}_3 * \mathbb{Z}_2$ of the cyclic groups of order 3 and 2, respectively. A generic symmetric generating set of this group is $S = \{a, \bar{a}, b\}$, where a and its inverse $\bar{a}$ have order 3 and b has order 2. Then $G_\triangle = G(\Gamma_\triangle, S)$ is the Cayley graph of $\Gamma_\triangle$ with respect to S and $\ell_\triangle$ its inherited edge-labelling. Our goal is to calculate the generating function of SAWs and the connective constant of $G_\triangle$ by decomposing $G_\triangle$ into finite parts. A simple way of achieving this is the block cut-vertex decomposition of $G_\triangle$; the corresponding decomposition tree $D_\triangle$ is shown in Figure 4.

The automorphism group $AUT(G_\triangle, \ell_\triangle) \simeq \Gamma_\triangle$ of the edge-labelled graph $(G_\triangle, \ell_\triangle)$ acts quasi-transitively on $D_\triangle$ by mapping blocks onto blocks and cut-

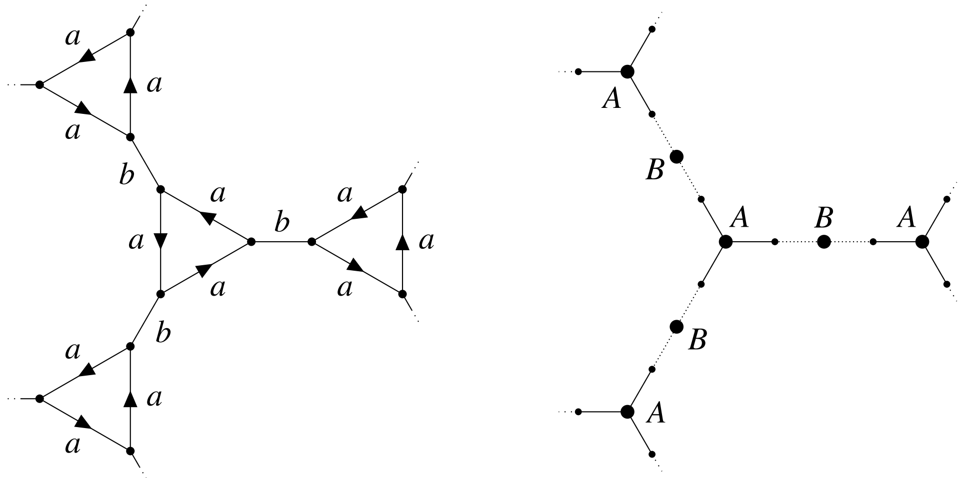


Figure 4: The labelled graph $G_\triangle$ and its block cut-vertex decomposition tree $D_\triangle$.

vertices onto cut-vertices. In particular the blocks can be partitioned into two types; we say that a block has type A if it is isomorphic to the complete graph on 3 vertices K_3, and it has type B if it is isomorphic to K_2. The possible configurations of walks on blocks can be seen in Figure 5; for ambiguity reasons we only list configurations of walks containing at least one edge of the block.

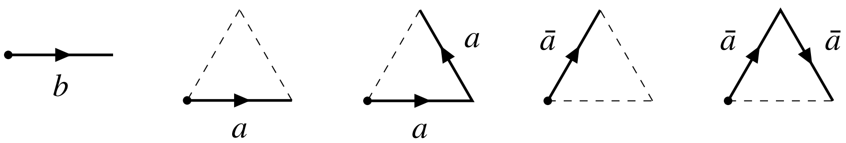


Figure 5: Possible non-empty configurations and their corresponding words.

Let us fix some vertex o of $G_\triangle$ and formulate a grammar $\mathcal{C}$ generating the language $L_{SAW,o}$ of SAWs on $G_\triangle$ starting at o. The group $\Gamma_\triangle$ acts with two orbits on the set of edges of $D_\triangle$; two edges are in the same orbit if they are incident to blocks of the same type. We introduce two non-terminals per edge orbit, call them A_C, C_A, B_C and C_B. The idea behind this notation is that the non-terminal A_C corresponds to a step from a block of type A to a cut-vertex and is paired with the non-terminal C_A corresponding to a step in the opposite direction. The non-terminal C_A is able to generate words corresponding to non-empty configurations on blocks of type A, possibly followed by a non-terminal A_C indicating that the walk has not ended yet. On the other hand A_C can only generate the non-terminal

C_B. We end up with the following set of production rules:

$$\begin{aligned} S &\vdash \varepsilon \mid C_A \mid C_B\,, \\ C_A &\vdash a \mid aA_C \mid a^2 \mid a^2A_C \mid \bar{a} \mid \bar{a}A_C \mid \bar{a}^2 \mid \bar{a}^2A_C\,, \\ C_B &\vdash b \mid bB_C\,, \\ A_C &\vdash C_B\,, \\ B_C &\vdash C_A\,. \end{aligned}$$

While the non-terminals A_C and B_C could be easily eliminated in these productions, they are required in more complicate examples and also help to understand the underlying procedure. Note that the sequence of involved non-terminals in the sequence of derivations of a word corresponding to a walk on $G_\triangle$ is very much related to the 'projected shape' of the walk on the derivation tree $D_\triangle$.

Using the theory of CHOMSKY and SCHÜTZENBERGER [4] the productions can be translated into a system of polynomial equations. One of its solutions is

$$F_{SAW,o}(t) = \frac{(1+t)(1+2t+2t^2)}{1-2t^2-2t^3},$$

the generating function of SAWs on $G_\triangle$ starting at o. The connective constant is the reciprocal of the smallest positive root of the denominator,

$$\mu(G_\triangle) \approx 1.7693.$$

Finally we want to mention two other methods to obtain these results. First note that $G_\triangle$ is the free product $K_3 * K_2$ of the complete graphs on 3 and 2 vertices, respectively. GILCH and MÜLLER [13] derived a formula to calculate the generating function of SAWs on free products $G = G_1 * G_2 * \cdots * G_r$ of graphs in terms of the generating functions of SAWs on the factors G_i. In particular, if all G_i are finite graphs, the resulting graph G has only ends of size 1. In this case the generating functions of SAWs on the factors are polynomials and a rational generating function will be obtained for G.

$G_\triangle$ also arises as the Fisher transformation of the 3-regular tree, so it is possible to apply a formula by GRIMMETT and LI [12] to obtain the connective constant $\mu(G_\triangle)$ from the connective constant of the 3-regular tree. However using their approach, it is not possible to derive the generating function of SAWs.

This simple example serves as the motivation for the general case. Start with any quasi-transitive edge-labelled graph (G,ℓ) having only ends of size 1. Then the block cut-vertex decomposition of G contains only finitely many different types of finite blocks, one for each orbit of blocks provided by the action of $\mathsf{AUT}(G,\ell)$ on the decomposition tree D of G. By introducing two non-terminals for every orbit of the action of $\mathsf{AUT}(G,\ell)$ on the edges of D, a grammar as in the example above can be constructed. This grammar is regular, thus the generating function of SAWs is a rational function and the connective constant is an algebraic number.

Our next step is to allow ends of size 2 in (G,ℓ). While it is still possible to use the block cut-vertex decomposition to decompose G, the obtained blocks may contain infinitely many vertices. However, infinite blocks do not contain cut-vertices, so they are 2-connected graphs. The theory of 3-block decompositions introduced by TUTTE [24] only works for finite 2-connected graphs, but a generalisation by DROMS, SERVATIUS and SERVATIUS [5] treat the infinite case and can be utilised here. The basic idea is to cleverly decompose the graph into so-called 3-blocks, where any two 3-blocks are either disjoint or their intersection is a separating set of size 2 in G. In particular the 3-blocks obtained from this decomposition are finite, if all ends of the underlying graph have size at most 2. As before the automorphism group $\mathsf{AUT}(G,\ell)$ acts quasi-transitively on the corresponding 3-block decomposition tree. Thus we end up with finitely many types of finite 3-blocks.

The main difference occurs when constructing the grammar corresponding to the language of SAWs. As adjacent 3-blocks intersect in sets of size 2, SAWs may now take 'detours'. This means that SAWs can leave a 3-block via one of these separating sets of size 2 and later enter this 3-block again via the same separating set. This problem can be solved by adding additional non-terminals generating the language of these detours and modifying our production rules to possibly contain several 'detour' non-terminals. For this reason the constructed grammar is (unambiguous) context-free and the generating function of SAWs and the connective constant of G are algebraic.

A combination of block cut-vertex decomposition and 3-block decomposition of all occurring infinite blocks yields the following result:

Theorem 2. *For every vertex o of a quasi-transitive deterministically edge-labelled graph X the following holds:*

(a) If all ends of X have size 1, then $L_{SAW,o}$ is regular.

(b) If all ends of X have size at most 2, then $L_{SAW,o}$ is unambiguous context-free.

In Section 4 we have seen that SAW-generating functions of one-dimensional lattices are rational, so one might be inclined to think that the language of their SAWs could be regular. In general this is not the case, as the following example shows.

Consider again the deterministically edge-labelled graph (G_L,ℓ_L) in Figure 6, which is a Cayley graph of the group $\mathbb{Z}\times\mathbb{Z}_2$. Fix some vertex o and assume that the language $L_{SAW,o}$ of SAWs starting at o is regular. Then by the closure properties of regular languages the intersection $L = L_{SAW,o} \cap \{ba^m b\bar{a}^n \mid m,n\in\mathbb{N}_0\}$ of two regular languages is regular. It consists of all words

$$w = ba^m b\bar{a}^n, \quad m > n \geq 0$$

because $n \geq m$ implies that the spiral-shaped walk corresponding to w has a self-intersection at vertex o, which contradicts $w \in L_{SAW,o}$. Figure 6 shows the walk corresponding to the word w for $m = 4$ and $n = 3$. It is a standard exercise to show that the language $\{a^m \bar{a}^n \mid m > n \geq 0\}$ is not context-free by using the pumping lemma for regular languages. The language L can be treated in the same way, contradicting our assumption.

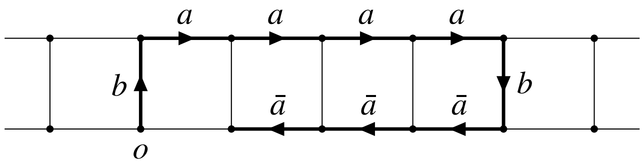


Figure 6: Spiral shaped walk on G_L.

Let us look at a slight generalisation. Consider the graph G_{DL} in Figure 7 consisting of 3 bi-infinite paths connected by edges. The subset of spiral shaped SAWs as shown in the figure behaves similar to the language $\{a^l b^m c^n \mid l > m > n \geq 0\}$. By the pumping lemma this language is not context-free, so the language of SAWs on the graph G_{DL} cannot be context-free.

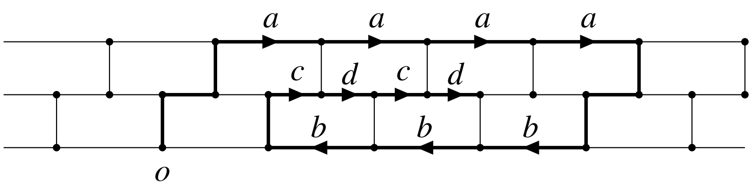


Figure 7: Spiral-shaped walk on G_{DL}.

These examples are of great interest in the general case. One can show that whenever a graph G contains a thin end of size at least 2, there is a sub-graph H of G, such that H is quasi-transitive and isomorphic to a subdivision of G_L, i.e., it arises by adding new vertices of degree 2 on edges of G_L. Using closure properties of the class of regular languages, it can be shown that the non-regularity of the language of SAWs on G_L extends to its subdivision H and in a second step also to the graph G. An identical approach shows that ends of size at least 3 in a graph G imply the existence of a subgraph isomorphic to a subdivision of G_{DL} and thus that the language of SAWs on G cannot be context-free. Furthermore similar ideas can be used to show that any graph G containing a thick end cannot have a context-free language of SAWs; the rigorous proof however is more difficult and technical. Together with the previous theorem, we obtain the following result:

Theorem 3. *For any vertex o of a quasi-transitive deterministically edge-labelled graph G the following holds:*

(a) $L_{SAW,o}$ *is regular if and only if all ends of G have size 1.*

(b) $L_{SAW,o}$ *is context-free if and only if all ends of G have size at most 2. In this case* $L_{SAW,o}$ *is unambiguous context-free.*

In a current work in progress we study the general case where G is tree-like and may contain ends of arbitrary finite size. A promising generalisation of context-free languages are the multiple context-free languages introduced by SEKI et al. [23]. These languages deal with tuples of strings and can model cross-serial dependencies in natural languages. We show that the language of SAWs on a graph G is m-multiple context-free if and only if all ends of the underlying quasi-transitive deterministically edge-labelled graph have size at most $2m$.

6 Counting self-avoiding walks on tree-like graphs

Let us forget about edge-labels and get back to counting SAWs on quasi-transitive tree-like graphs with ends of arbitrary (finite) size. Block cut-vertex decomposition and 3-block decomposition are not sufficient to decompose such graphs, so we need a more general approach. Tree decompositions of arbitrary finite graphs were first introduced by HALIN [15] in 1976 and later rediscovered by ROBERTSON and SEYMOUR [22]. A version for infinite graphs was described by DUNWOODY [7], who used edges as separating sets, and later by DUNWOODY and KRÖN [8], who decomposed graphs at their vertices. We also want to mention a recent paper of CARMESIN, HAMANN and MIRAFTAB [3], as their results are essential for our work.

For our purpose, a *tree decomposition* of a graph G consists of a *decomposition tree* D and a family $(V_s)_{s\in D}$ associating every vertex s of D with a subset V_s of the vertex set of G, such that conditions (T1) to (T3) are satisfied.

(T1) The vertex set of G is the union of all sets V_s.

(T2) For every edge of G there is some V_s containing both of its endpoints.

(T3) $V_s \cap V_t \subseteq V_r$ for every vertex r on the unique s–t-path in D.

The set V_s is called the *part* of s and for any neighbour t of s in D, the set $V_s \cap V_t$ is called the *adhesion set* of s and t.

Block cut-vertex decompositions are natural examples of tree decompositions: The set of parts consists of all vertex sets of blocks and sets of size 1 for every cut-vertex of G. Additionally all adhesion sets are sets of size 1 containing only a single cut-vertex of G.

Let us again motivate and explain the general approach for obtaining generating functions by using our toy example $G_\triangle$ and its decomposition tree $D_\triangle$. Most of the upcoming ideas are part of a joint work with Florian LEHNER [18] currently in progress.

The basic idea is to study the language of configurations on $D_\triangle$ constructed in the following way: Fix some vertex r of $D_\triangle$ such that the part V_r contains the vertex o of $G_\triangle$ and think of $D_\triangle$ as an ordered rooted tree with root r, i.e., a rooted tree with an order on the children of every vertex. In our example we pick r such that $V_r = \{o\}$ is the part corresponding to the cut-vertex o.

The alphabet Σ consists of all possible configurations on parts of the tree decomposition and an additional letter e. Configurations and their corresponding letters in our specific example can be seen in Figure 8.

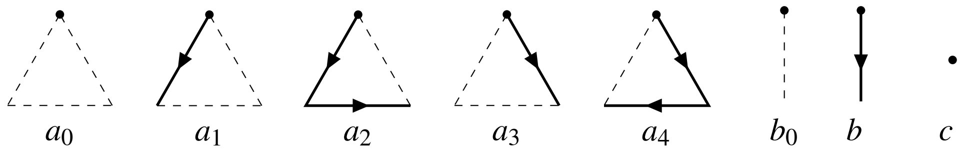


Figure 8: Possible configurations and their corresponding letters.

Every SAW π is related to some unique word $w(\pi)$ over Σ: Let D' be the minimal sub-tree of $D_\triangle$ such that all edges of π are contained in parts V_s corresponding to vertices s of D'. Every vertex s of D' is labelled with the configuration of π on V_s and every vertex of $D_\triangle$, which has a neighbour in D' but is not in D' itself is labelled with e, indicating that π does not continue further in this direction. We associate π with the word $w(\pi)$ constructed by writing down the labels of the labelled part of $D_\triangle$ in depth-first search order. Figure 9 shows the walk π associated with $w(\pi) = ca_2ecbca_3ecbee$ and its decomposition tree $D_\triangle$ with the corresponding vertex-labels.

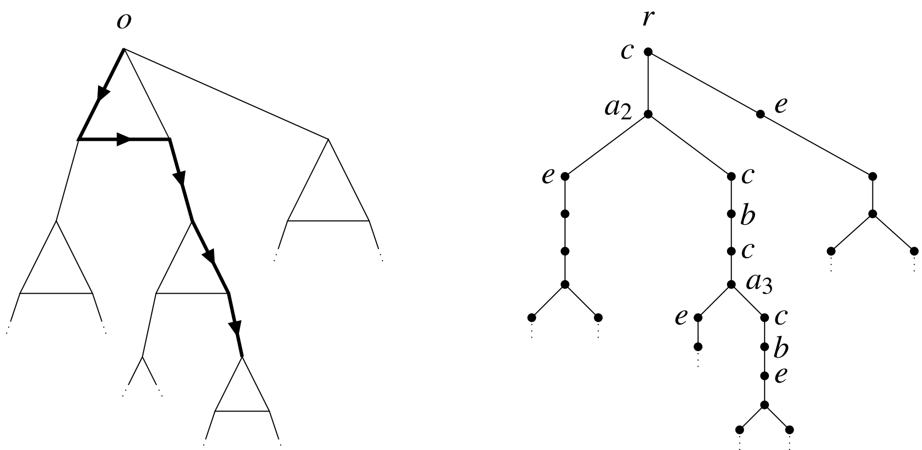


Figure 9: A SAW on $G_\triangle$ and its configurations on the ordered rooted tree $D_\triangle$.

For general tree-like graphs the construction of a grammar generating the language of configurations can be complicated and a large set of non-terminals may be necessary. However for our simple example it is easy to find the following production rules, which are already sufficient:

$$\begin{aligned} S &\vdash e \mid ceC_Be \mid cC_Ae\,, \\ C_A &\vdash a_1ee \mid a_2ee \mid a_3ee \mid a_4ee \mid a_1A_Ce \mid a_2eA_C \mid a_3eA_C \mid a_4A_Ce\,, \\ C_B &\vdash be \mid bB_C\,, \\ B_C &\vdash cC_A\,, \\ A_C &\vdash cC_B\,. \end{aligned}$$

The ordinary generating function of this language counts non-trivial configurations of the underlying SAWs instead of their edges. However it is not hard to obtain the generating function of SAWs. We assign to every letter of Σ as a weight the number of edges contained in its corresponding configuration. Then the theory of CHOMSKY and SCHÜTZENBERGER [4] can be applied to obtain a weighted language generating function, which is precisely the generating function $F_{SAW,o}(t)$ of SAWs on $G_\triangle$.

Finally, let us move to the general case of quasi-transitive tree-like graphs having ends of arbitrary size. Any such graph G admits a tree decomposition consisting of finite parts, such that $\mathsf{AUT}(G)$ acts quasi-transitively on the decomposition tree D. Following the steps in the example above it is possible to construct a context-free grammar generating the language of all configurations on the decomposition tree D corresponding to SAWs on G. We finish with the following nice result:

Theorem 4. *Let o be a vertex of a quasi-transitive tree-like graph G. Then the generating function $F_{SAW,o}(t)$ of SAWs on G starting at o is algebraic. In particular the connective constant $\mu(G)$ is an algebraic number.*

References

[1] Alm, S. E., and Janson, S.: *Random self-avoiding walks on one-dimensional lattices,* Comm. Statist. Stochastic Models **6** (1990), 169–212.

[2] Anisimov, A. V.: *Group languages,* Kibernetika **4** (1971), 18–24.

[3] Carmesin, J., Hamann, M. and Miraftab, B.: *Canonical trees of tree-decompositions,* preprint: `arXiv:2002.12030`, (2020).

[4] Chomsky, N. and Schützenberger, M.-P.: *The algebraic theory of context-free languages,* Computer Programming and Formal Systems **26** (P. Braffort abd D. Hirschberg, eds.) North-Holland, Amsterdam, 1963, pp. 118–161.

[5] Droms, C., Servatius, B., and Servatius, H.: *The structure of locally finite two-connected graphs,* Electron. J. Combin. **2** (1995), Research Paper 17, 10 pp.

[6] Duminil-Copin, H., and Smirnov, St.: *The connective constant of the honeycomb lattice equals $\sqrt{2+\sqrt{2}}$,* Ann. of Math. **175** (2012), 1653–1665.

[7] Dunwoody, M. J.: *Cutting up graphs,* Combinatorica **2** (1982), 15–23.

[8] Dunwoody, M. J., and Krön, B.: *Vertex Cuts,* J. Graph Th. **80 (2)** (2014), 136–171.
[9] Flory, P. J.: *The configuration of a real polymer chain,* J. Chem. Phys. **17** (1949), 303–310.
[10] Freudenthal, H.: *Über die Enden diskreter Räume und Gruppen,* Comment. Math. Helv. **17** (1945), 1–38.
[11] Grimmett, G., and Li, Z.: *Self-Avoiding Walks and Connective Constants,* Sojourns in Prob. Th. and Stat. Physics **III** (2019), 215–241.
[12] Grimmett, G., and Li, Z.: *Self-avoiding walks and the Fisher transformation,* Electron. J. Combin. **20** (2013) Paper 47, 14.
[13] Gilch, L. A., and Müller, S.: *Counting self-avoiding walks on free products of graphs,* Discrete Math. **340** (2017), 325–332.
[14] Halin, R.: *Über unendliche Wege in Graphen,* Math. Ann. **157** (1964), 125–137.
[15] Halin, R.: *S-functions for graphs,* Journal of Geometry **8** (1976), 171–186.
[16] Hammersley, J. M.: *Percolation processes. II. The connective constant*, Proc. Cambridge Philos. Soc. **53** (1957), 642–645.
[17] Harrison, M. A.: *Introduction to Formal Language Theory*, Addison-Wesley, Reading, MA, 1978.
[18] Lehner, F., and Lindorfer, C.: *Self-avoiding walks and multiple context-free languages*, in preparation.
[19] Lindorfer, C., and Woess, W.: *The language of self-avoiding walks*, Combinatorica, in print, preprint: `arXiv:1903.02368`, (2019).
[20] Madras, N., and Slade, G.: *The Self-avoiding Walk*, Probability and its Applications, Birkhäuser, Boston, MA, 1993.
[21] Muller, D. E., and Schupp, P. E.: *Groups, the theory of ends and context-free languages*, J. Comput. System Sc. **26** (1983), 295–310.
[22] Robertson, N., and Seymour, P. D.: *Graph minors III: Planar tree-width*, J. Combinatorial Th., Ser. B, **36 (1)** (1984), 49–64.
[23] Seki, H., Matsumura, T., Fujii, M., and Kasami, T.: *On multiple context-free grammars*, Theor. Comput. Sci. **88** (1991), 191–229.
[24] Tutte, W. T.: *Graph Theory.* With a foreword by C. St. J. A. Nash-Williams. Encyclopedia of Mathematics and its Applications **21**, Addison-Wesley, Reading, MA, 1984.
[25] Woess, W.: *Graphs and groups with tree-like properties*, J. Combinatorial Th., Ser. B, **47** (1989), 361–371.

Author's address:
Institute of Discrete Mathematics
Graz University of Technology
Steyrergasse 30/III
A-8010 Graz, Austria
email lindorfer@math.tugraz.at

Internat. Math. Nachrichten
Nr. 244 (2020), 27–45

Universal structures in Mathematical Finance

Christa Cuchiero
Universität Wien

Universal structures in the sense of this article and the START project with the same title pertain literally to both, mathematics and finance. On the financial side we observe robust empirical features that hold universally across different markets, different asset classes and in particular over time. On the mathematical side it concerns universally appearing model classes and probabilistic properties, inherent in many at first sight unrelated phenomena.

It is the purpose of the START project to explore these mathematical and financial universalities and to provide a unifying stochastic framework, building on so-called affine and polynomial processes, which appear to have, appropriately seen, universal approximation properties, otherwise also well known from the theory of machine learning.

In this article we shall in particular shed some light on the mathematical universality of affine and polynomial processes.

1 Introduction and Motivation

Recognizing and exploring universal structures to learn more about the world which surrounds us is one major driving force of science. Karl Raimund Popper [34, 33] states this as follows:

"If it is the aim of science to explain, it will also be its aim to explain what so far has been accepted as an explicans such as a law of nature. Thus the task of science constantly renews itself. We could go on forever, proceeding to explanations of a higher and higher level of universality..."

What are these explicantia or laws of nature in finance and economics? Indeed, in this context universality might sound surprising as (financial) markets certainly do *not* obey a "law of nature" as it is the case for instance in physics. However,

ISSN 0020-7926

even if finance rather appears as a phenomenon of social interaction, universal market features *do* exist. Such "laws of social interaction" can be made observable through statistically measurable structures that hold universally over time and across markets. However these observables are often enigmatic, hard to detect and difficult to describe dynamically.

Let us illustrate this by means of two examples: *first*, the *stability of capital distribution curves* over time, as shown in Figure 1. Each of these curves depicts the

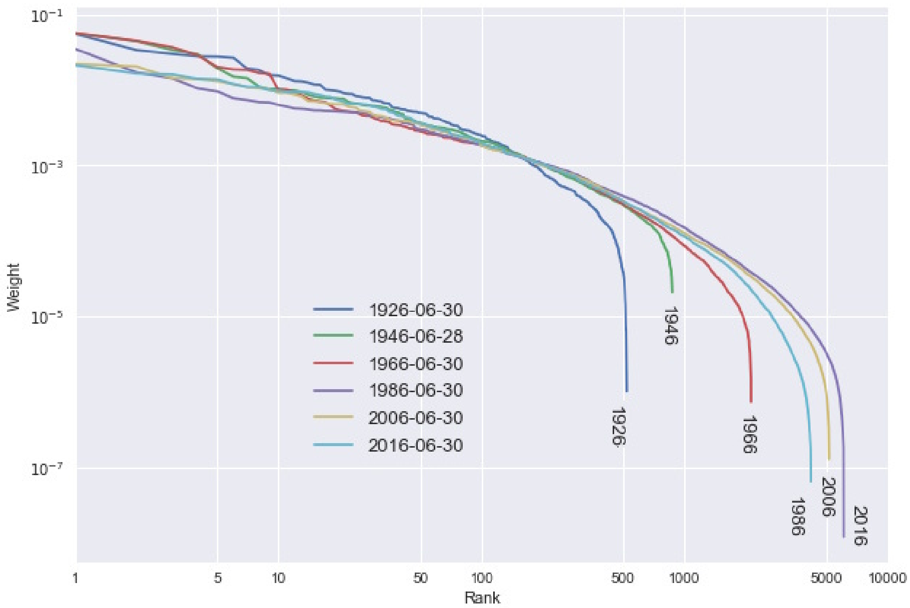


Figure 1: Capital distribution curves: 1926-2016, source [36].

relative market capitalization in ranked order of the major US markets' stocks on a log-log scale from 1926 to 2016. The relative market capitalization or market weight is defined as the percentage of the market capitalization of a fixed company, i.e., the number of outstanding shares times the current price of one share, with respect to the capitalization of the whole market. The striking feature of these curves is their remarkably stable shape over the past 90 years. Although the market weights of each company fluctuate stochastically, the shape of the capital distribution curves differs (in first order) over the years only by the number of stocks present in the considered market. This fundamental observation was the starting point for R. Fernholz to develop *stochastic portfolio theory* about 20 years ago, see the monograph [23]. Since then this stylized fact has been detected in many circumstances, most recently on the new market of crypto-currencies by J. Ruf [36].

The *second* universal phenomenon that we consider here is *rough volatility*. The rough volatility paradigm asserts that the trajectories of assets' volatility are rougher than Brownian motion, as shown in Figure 2. This revolutionary per-

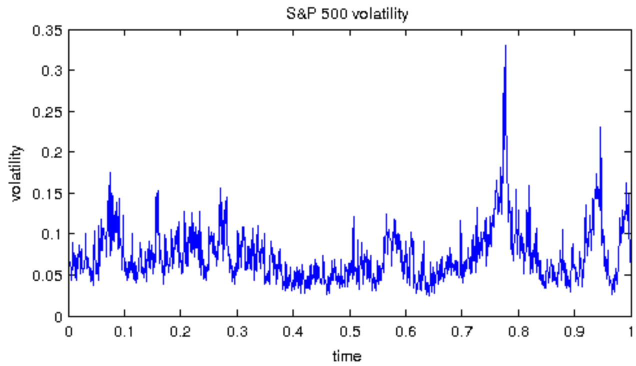


Figure 2: Trajectory of the S&P 500 volatility over one year.

spective was pioneered by J. Gatheral, T. Jaisson and M. Rosenbaum in [27]. In contrast to the first phenomenon, it is a rather recent empirical observation, also due to the availability of high frequency time series data. Even if the choice of the estimation procedure has a certain impact on the resulting quantities (see, e.g., [26]), *roughness as universal property* of various kinds of assets *is* convincing, in particular because the time series conclusions are reinforced by option price data. Concerning the evidence from time series data, there is general agreement that roughness as a feature cannot be rejected, even though it is hard to measure its degree (the so-called Hurst parameter) exactly. Microstructural foundations explaining rough volatility from simple but expressive tick price models further complement these empirical observations.

Beside providing paragons of universality in financial markets, both phenomena have – not quite coincidentally – (co)-founded two very active research areas in mathematical finance, namely *stochastic portfolio theory (SPT)* and what we call *contemporary stochastic volatility modeling (CSV)*.

Even though both fields have emerged from the analysis of robust and universal market properties, their mathematical similarity is not obvious at all. Still they share one common crucial feature, namely a generic *non-Markovanity* of at first sight naturally low dimensional objects, be it relative market capitalization or volatility. In most cases, this non-Markovanity can only be dissolved by passing to *infinite dimensional stochastic processes.* Intriguingly such infinite dimensional models stem almost always from the class of *affine and polynomial processes:* essentially all examples of rough volatility models, such as Hawkes processes [28], rough Heston [20], rough Wishart [17] or rough Bergomi [3] can be viewed as infinite dimensional affine or polynomial processes as shown in [1, 15, 17, 18]. This is similar for SPT, where the most flexible and tractable models appear to be (measure-valued) polynomial [8, 14].

Also in other not finance related areas, like population genetics, we encounter a plethora of affine and polynomial models: all well-known measure-valued diffusions such as the Fleming–Viot process, the Super–Brownian motion, and the Dawson–Watanabe superprocess (see, e.g., [21] and the references therein) are affine or polynomial processes.

The deeper reason behind this predominance can be interpreted as a *universal approximation property* of affine or polynomial dynamics in the space of *all* stochastic dynamics, driven by say, Brownian motion (or many other stochastic processes). This suggests an inherent universality of these model classes, which in the finite dimensional setting were first systematically treated in [19] (affine case) and [12, 24] (polynomial case).

It is the purpose of this article to elaborate on the *mathematical universality of affine and polynomial processes.* This is a topic that has accompanied me since a long time and the crucial and open question is how to make this universality mathematically precise. In several recent articles [8, 13, 14, 15, 18] we could make progress by showing how many models which are not at all recognized as affine or polynomial can nevertheless be embedded in this framework. The current article is based on [16], where we show that generic classes of diffusion models with possibly path-dependent characteristics (as defined in Equation (21) below) are projections of infinite dimensional affine processes (which in this setup coincide with polynomial processes). A key ingredient to establish this result is the so-called *signature process,* well known from K. Chen's work [5, 6] as well as from rough paths theory and regularity structures [25]. Indeed, the signature process of all these generic models turns out to be an infinite dimensional affine process. This then allows to transfer the powerful affine technology to these much more general stochastic models. In particular, their characteristic function can be represented in terms of solutions of (infinite dimensional) Riccati ordinary differential equations (ODEs). Viewing this universality property from a slightly different angle also opens up the way to see affine processes as a family of processes that is dense (with respect to an appropriate topology) in the space of all stochastic dynamics exhibiting trajectories of a certain regularity. This is in analogy with universal model classes, such as artificial neural networks, used in machine learning. Establishing a result in this direction is one of the goals of this project.

The remainder of the article is organized as follows: in Section 2 we introduce affine and polynomial processes in the simplest one-dimensional setting and state their key properties. In Section 3 we review the concept of signature, while Section 4 shows that any generic diffusion model given by Equation (21) is a projection of an affine process. Section 5 explains how this can be exploited for universal modeling in finance.

2 Definition and key properties of affine and polynomial processes

Let us start by introducing affine and polynomial processes in the simplest setting, namely as diffusion processes in *one* dimension. They here appear as narrow class of processes whose universal character is at this stage by no means visible. Consider on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, an Itô diffusion process of the form

$$dX_t = b(X_t)dt + \sqrt{a(X_t)}dB_t, \tag{14}$$

with $a : \mathbb{R} \to \mathbb{R}_+$ and $b : \mathbb{R} \to \mathbb{R}$ continuous functions and B a standard Brownian motion. The initial condition X_0 is deterministic and lies in the state space S which is supposed to be some (bounded or unbounded) interval of $\mathbb{R}$.

Definition 1. *A solution*[1] *X of* (14) *is called* polynomial process *if*

- *b is a linear function, i.e. $b(x) = b + \beta x$ for some constants b and β and*
- *a is a quadratic function, i.e. $a(x) = a + \alpha x + Ax^2$ for some constants a, α and A.*

If additionally $A = 0$, then the process is called affine.[2]

This simple definition has remarkable implications, namely that all marginal moments of a polynomial process, i.e. $\mathbb{E}[X_t^n]$ can be computed by solving a system of linear ODEs. In the case of an affine process, we additionally get that exponential moments $\mathbb{E}[e^{uX_t}]$ for $u \in \mathbb{C}$ can be expressed in terms of solutions of Riccati ODEs whenever $\mathbb{E}[|e^{uX_t}|] < \infty$. This is stated in Theorem 5 and Theorem 6, which summarize results of [12] and [19] in the one dimensional diffusion case.

We here present the theory from the point of view of dual processes (see e.g. Chapter 4 in [22] and Remark 3), which has not been considered in this form in the original papers [12] and [19].

Let us first introduce the so-called *extended generator* of (14), which is in the current one-dimensional diffusion setting the following second order differential operator $\mathcal{A}$ acting on functions $f \in C^2$ and given by

$$\mathcal{A}f(x) = f'(x)b(x) + \frac{1}{2}f''(x)a(x). \tag{15}$$

[1] We here do not specify the precise solution concept but just mention that weak solutions are sufficient to develop the whole theory.

[2] Note that in this diffusion setting all affine processes are polynomial (in general this only holds true under moment conditions on the jump measures).

In contrast to the functional analytic notion of the infinitesimal generator of a Markov process (where the domain is usually restricted to fewer functions), the defining property of the extended generator is that

$$f(X_t) - f(X_0) - \int_0^t \mathcal{A} f(X_s) ds$$

is a *local martingale.*

For certain classes of functions (depending on the form of a and b) it will be possible to associate a so-called dual operator to $\mathcal{A}$.

Definition 2. *Let $U \subset \mathbb{C}^n$ and define the following class of functions*

$$\mathcal{C}^U = \{f : \mathbb{R} \times U \to \mathbb{R} : (x,u) \mapsto f(x,u) \,|\, f(\cdot,u) \in C^2 \textit{ for all } u \in U\}.$$

A dual operator *associated to $\mathcal{A}$, denoted by $\mathcal{B}$, is a linear operator such that*

$$\mathcal{A} f(\cdot,u)|_x = \mathcal{B} f(x,\cdot)|_u \quad \textit{for all } x \in S,\, u \in U,\, f \in \mathcal{C}^U.$$

In the sequel we shall apply the extended generator and its dual operator to particular classes of functions, either to exponentials or polynomials. Let us here start with spaces of polynomials up to degree $k \in \mathbb{N}$, which we denote by $\mathcal{P}_k$, i.e.,

$$\mathcal{P}_k = \left\{ x \mapsto \sum_{i=0}^{k} c_i x^i \,|\, c_i \in \mathbb{R} \right\}.$$

For a polynomial in $\mathcal{P}_k$ we denote by $c = (c_0, \dots, c_k)^\top \in \mathbb{R}^{k+1}$ its coefficients vector and write $\overline{x}$ for $(1, x, \dots, x^k)^\top$ (without indicating the dependence on k). We then abbreviate the polynomial $\sum_{i=0}^k c_i x^i$ by $\langle c, \overline{x} \rangle_k$ and write

$$p(x,c) := \langle c, \overline{x} \rangle_k = \sum_{i=0}^{k} c_i x^i.$$

Note that in the case of a polynomial process, $\mathcal{A}$ maps $\mathcal{P}_k$ to $\mathcal{P}_k$, i.e. *polynomials to polynomials of same or smaller degree*, as b is a linear and a a quadratic function. Leaving $\mathcal{P}_k$ invariant for every $k \in \mathbb{N}$, is the defining property of a *polynomial operator,* a notion which can also be applied to more general linear operators of so-called Lévy-Khintchine type (beyond second order differential operators).

To such polynomial operators we can – in spirit of Definition 2 – associate a family of *dual operators* which are applied to the coefficients vectors and defined as follows.

Definition 3. *Let $\mathcal{A}$ be a polynomial operator mapping $\mathcal{P}_k$ to $\mathcal{P}_k$ for every $k \in \mathbb{N}$. Fix now k and consider a dual operator $\mathcal{B}$ as of Definition 2 with $C^U = \mathcal{P}_k$ acting on $c \mapsto p(x,c)$ such that*

$$\mathcal{A}p(\cdot,c)|_x = \mathcal{B}p(x,\cdot)|_c.$$

By the linearity in $\overline{x}$ we can identify $\mathcal{B}$ on $\mathcal{P}_k$ with a linear map L_k from $\mathbb{R}^{k+1}$ to $\mathbb{R}^{k+1}$ such that

$$\mathcal{A}p(\cdot,c)|_x = \langle L_k c, \overline{x}\rangle_k = p(x, L_k c), \quad \text{for all } x \in S. \tag{16}$$

*We call this map the k-*th dual operator *(associated to $\mathcal{A}$).*

To illustrate this notion let us consider the well-known *Wright-Fisher diffusion* from population genetics which takes values in $[0,1]$.

Example 1. *Let $S = [0,1]$, $b(x) = 0$ and $a(x) = x(1-x)$. Thus the diffusion that we consider is given by*

$$dX_t = \sqrt{X_t(1-X_t)}dB_t, \quad X_0 \in [0,1],$$

and its extended generator is of the following form

$$\mathcal{A}f(x) = \frac{1}{2}f''(x)x(1-x).$$

Computing $\mathcal{A}p(x,c)$ for $p(x,c) = \langle c, \overline{x}\rangle_k$ yields

$$\mathcal{A}p(x,c)| = c_2 x + \cdots + \frac{1}{2}(-j(j-1)c_j + j(j+1)c_{j+1})x^j + \cdots + \frac{1}{2}(-k(k-1)c_k)x^k.$$

From this together with (16) *we see that the k-th dual operator L_k is given by the following $\mathbb{R}^{(k+1)\times(k+1)}$ matrix*

$$L_k = \begin{pmatrix} 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & -1 & 3 & \cdots & \cdots & \cdots & 0 \\ & & \ddots & & & & & \vdots \\ 0 & \cdots & & -\frac{j(j-1)}{2} & \frac{j(j+1)}{2} & \cdots & & 0 \\ & & & & \ddots & & & \\ & & & & & & & -\frac{k(k-1)}{2} \end{pmatrix}.$$

With these preparations we are now ready to state the announced theorem.

Theorem 5. *Let $T > 0$ be fixed and let X be a polynomial process as of Definition 1. Moreover, denote by $c(t) = (c_0(t), \dots, c_k(t))^\top$ the solution of the following linear ODE*

$$\partial_t c(t) = L_k c(t), \quad c(0) = c \in \mathbb{R}^{k+1},$$

where L_k is the k-th dual operator defined in (16)*. Then $\mathbb{E}[|X_T^n|] < \infty$ for all $n \in \mathbb{N}$ and its conditional moments for $0 \leq t < T$ are given by*

$$\mathbb{E}\left[\sum_{i=0}^{k} c_i X_T^i \mid \mathcal{F}_t\right] = \sum_{i=0}^{k} c_i(T-t) X_t^i, \quad a.s. \tag{17}$$

As $c(t) = \exp(L_k t)c$, (17) *can thus be written as*

$$\mathbb{E}[\langle c, \overline{X}_T \rangle_k | \mathcal{F}_t] = \langle \exp(L_k(T-t))c, \overline{X}_t \rangle_k.$$

Proof. The proof follows from Theorem 2.7 and Theorem 2.15 in [12]. We refer also to Remark 3 below to get an intuition for the proof. □

Remark 1. *The deterministic "process" $c(t)$ is the previously mentioned dual process here.*

The above theorem asserts that not only the (extended) generator of a polynomial process, but also the associated semigroup maps $\mathcal{P}_k$ to $\mathcal{P}_k$. As this is a finite dimensional space, computing conditional moments of polynomial processes thus just amounts to computing matrix exponentials. This is considerably simpler than the usual situation where one has to solve the associated Kolmogorov partial differential equation (PDE).

Theorem 5 extends to much more general settings than the simple one dimensional case presented here. In [15] we present the theory for Banach space valued polynomial processes (compare for instance Theorem 3.4 in [15] for an analogous statement in the Banach space framework).

Having treated the polynomial case, let us now come to affine processes. This process class exhibits an even higher tractability in the sense that we can go beyond expectations of polynomials and additionally compute exponential moments, and thus in particular the characteristic function, of the process' marginals. The role of the linear ODEs of Theorem 5 is now taken by Riccati (i.e. quadratic) ODEs and we obtain an exponential affine form of the moment generating/characteristic function.

To deduce this let us introduce the dual operator for affine processes.

Definition 4. *Let $\mathcal{A}$ be the extended generator of an affine process, meaning that the functions a and b in* (15) *are affine. Let C^U as of Definition 2 be given by*

$$C^U = \{(x, (u_0, u)) \mapsto \exp(u_0 + u_1 x) \,|\, (u_0, u) \in \mathbb{C}^2\}$$

and consider a dual operator $\mathcal{B}$ acting on $(u_0, u) \mapsto \exp(u_0 + ux)$ such that

$$\mathcal{A}\exp(u_0 + u\cdot)|_x = \mathcal{B}\exp(\cdot + \cdot x)|_{(u_0,u)}, \quad x \in S.$$

We call $\mathcal{B}$ affine dual operator.

Remark 2. *To explicitely compute the form of $\mathcal{B}$, define the functions*

$$\begin{aligned} R(u) &:= \frac{1}{2}\alpha u^2 + \beta u, \\ F(u) &:= \frac{1}{2}au^2 + b. \end{aligned} \tag{18}$$

Then, by definition

$$\begin{aligned} \mathcal{B}\exp(u_0 + ux) &= \mathcal{A}\exp(u_0 + ux) \\ &= ((b + \beta x)u + \frac{1}{2}(a + \alpha x)u^2)\exp(u_0 + ux) \\ &= (F(u) + R(u)x)\exp(u_0 + ux). \end{aligned}$$

Therefrom we can guess that $\mathcal{B}$ is the restriction of the following transport operator applied to function $g \in C^1(\mathbb{C}^2, \mathbb{C})$:

$$\mathcal{B}g(u_0, u) = \partial_{u_0} g(u_0, u)F(u) + \partial_u g(u_0, u)R(u).$$

With these notions at hand we can now state the announced theorem.

Theorem 6. *Let $T > 0$ be fixed and let X be an affine process as of Definition 1. Let $u \in \mathbb{C}$ such that $\mathbb{E}[\exp(\Re(u)X_T)] < \infty$ where $\Re(u)$ denotes the real part of u. Then, for all $0 \leq t \leq T$,*

$$\mathbb{E}[\exp(uX_T)|\mathcal{F}_t] = \exp(\phi(T - t) + \psi(T - t)X_t), \quad a.s.$$

where

$$\begin{aligned} \partial_t \exp(\phi(t) + \psi(t)x) &= \mathcal{B}\exp(\phi(t) + \psi(t)x) \\ &= (F(\psi(t)) + R(\psi(t))x)\exp(\phi(t) + \psi(t)x) \end{aligned}$$

with $\mathcal{B}$ the affine dual operator of Definition 4 and F and R defined in (18)*. Hence, ϕ and ψ (the components of the dual process here) solve the following Riccati differential equations*

$$\begin{aligned} \partial_t \psi(t) &= R(\psi(t)), \qquad \psi(0) = u, \\ \partial_t \phi(t) &= F(\psi(t)), \qquad \phi(0) = 0. \end{aligned}$$

Proof. For $u \in \mathbb{C}$ such that $x \mapsto \exp(ux)$ is bounded, the result follows from Theorem 2.7 in [19]. For all other $u \in \mathbb{C}$ it is a consequence of Theorem 3.7 in [38] (see also Theorem 2.26 in [29]). In particular, the explosion time of the Riccati equations coincides with the time when the exponential moment ceases to exist. □

Remark 3. *The philosophy behind Theorem 5 and Theorem 6 relies on the following notion of duality for stochastic processes (see Theorem 4.4.11 in [22]), which we state here without specifying all technical conditions: let $(X_t)_{t\geq 0}$ and $(U_t)_{t\geq 0}$ be independent processes and f, g measurable functions from $\mathbb{R} \times U$ to $\mathbb{R}$. If*

$$f(X_t, u) - \int_0^t g(X_s, u)ds$$

is a martingale for each u and if

$$f(x, U_t) - \int_0^t g(x, U_s)ds$$

is a martingale for each x, then for almost every $t \geq 0$, we have

$$\mathbb{E}[f(X_t, U_0)] = \mathbb{E}[f(X_0, U_t)]. \tag{19}$$

Applying this result to the polynomial case, means choosing f to be a polynomial, say with coefficients vector c so that the role of u is taken by c and $f(x,c) = p(x,c)$. Since all moments of X exist, one can prove that

$$p(X_t, c) - \int_0^t \mathcal{A}p(X_s, c)ds$$

is a martingale (and not only a local martingale). Moreover, since $\partial_t c(t) = L_k c(t)$, we have

$$p(x, c(t)) - \int_0^t \langle L_k c(s), \overline{x}\rangle_k ds = p(x, c(0)),$$

for all $t \geq 0$, thus a constant "martingale". By the definition of the k-th dual operator $\langle L_k c, \overline{x}\rangle_k = \mathcal{A}p(x,c)$ for all $x \in S$ and $c \in \mathbb{R}^{k+1}$, so that the duality relation (19) *implies*

$$\mathbb{E}[p(X_t, c)] = p(X_0, c(t)).$$

Hence, the dual process is the deterministic evolution of the coefficients vector.

In the affine case, the function f is given by $\exp(u_0 + ux)$. *For appropriate $u \in \mathbb{C}$ and $T \geq 0$ (possibly depending on u)*

$$\exp(u_0 + uX_t) - \int_0^t \underbrace{\mathcal{A}\exp(u_0 + uX_s)}_{(F(u)+R(u)X_s)\exp(u_0+uX_s)} \, ds, \quad 0 \leq t \leq T,$$

is actually a martingale. Due to the specification of ϕ *and* ψ *(here with* $\phi(0) = u_0$*), we also have*

$$\exp(\phi(t) + \psi(t)x) - \int_0^t \mathcal{B} \exp(\phi(s) + \psi(s)x) ds = \exp(u_0 + ux)$$

for all $0 \leq t \leq T$*, thus a constant "martingale". By the definition of the affine dual operator we have* $\mathcal{A}\exp(u_0 + ux) = \mathcal{B}\exp(u_0 + ux)$*, so that all the assumptions for duality are satisfied. By inserting* $u_0 = 0$ *we thus obtain the affine transform formula as asserted in Theorem 6 , i.e.*

$$\mathbb{E}[\exp(uX_t)] = \exp(\phi(t) + \psi(t)X_0)$$

and the deterministic "processes" $(\phi(t), \psi(t))$ *are the components of the dual process* $(U_t)_{t \geq 0}$.

We have here treated the one-dimensional diffusion setting, mainly to ease notation and technicalities. But both affine and polynomial processes have been analyzed in more general situations:

- on different state spaces (e.g. positive definite matrices or more generally symmetric cones [9, 11]);
- in finite and infinite dimensions (on subsets of Hilbert and Banach spaces, e.g. [37, 15]; for measure valued processes, see [21] and the references therein);
- beyond continuous trajectories, e.g. [19, 12];
- in the realm of fractional stochastic Volterra processes, e.g. [2, 18].

The fascination exuded by this broad research area is twofold: first it opens up unforeseen connections to different areas of mathematics, e.g. to non-linear PDEs[3] or dynamic versions of universal approximation theorems. Second, a plethora of stochastic processes used in diverse fields of application turn out to be affine and/or polynomial. This becomes in particular apparent by lifting low dimensional processes to higher dimensions. Aspects of this inherent universality are explained in Section 4. But now something completely different ...

[3] They appear as generalized Riccati equations in infinite dimensions with non-linearities coming from the jump behavior of the process.

3 Towards universality: the signature of a path

One perspective how to understand this universality, is based on the so-called *signature process,* which plays a prominent role in *rough path theory* introduced in [31]. In a nutshell rough paths theory can be viewed as an extension of the classical theory of differential equations which is robust enough to allow for a pathwise treatment of stochastic differential equations driven by rough signals.[4] The signature of a path, first studied by K. Chen [6, 5], is a particular important object in this context and owes its relevance to the following three key facts:

- The signature of a path (of bounded p-variation) uniquely determines the path up to the so-called tree-like equivalences (which includes e.g., reparametrizations; see [4]).
- Under certain regularity conditions, the expected signature of a stochastic process determines the law of the signature (see Proposition 6.1 and Theorem 6.3 in [7]).
- Every continuous path functional can be approximated by a linear function of the signature arbitrarily well (on compacts of non tree-like paths).

Due to the last property signature can be thought of as a feature map (i.e. a map that captures the specific characteristics of the path), which serves as a *linear regression basis* for the space of continuous functions on paths. This linearization via signature can be made precise: indeed, Theorem 8 below states a universal approximation result in the current dynamic context.

To fix notation and to introduce the relevant notions we follow here [30]. By a d-dimensional path on some finite time horizon $T > 0$, we mean a continuous mapping $X : [0,T] \to \mathbb{R}^d$. The signature of a path X is then defined as a formal power series with d non-commutative indeterminates whose coefficients are iterated integrals of the path.

Definition 5. *Let X be a path of finite p-variation such that the following integration makes sense. The signature $\mathbb{X}_T$ of X over the time interval $[0,T]$ is defined as follows*

$$\mathbb{X}_T = (1, X_T^{(1)}, \dots, X_T^{(n)}, \dots),$$

where for each integer $n \geq 1$,

$$X_T^{(n)} := \int_{0<t_1<\cdots<t_n<T} dX_{t_1} \otimes \cdots \otimes dX_{t_n} \in (\mathbb{R}^d)^{\otimes n}, \quad n \geq 1.$$

The truncated signature of X of order n over the time interval $[0,T]$ is denoted by $\mathbb{X}_T^{(n)}$, i.e. $\mathbb{X}_T^{(n)} = (1, X_T^{(1)}, \dots, X_T^{(n)})$ for every integer $n \geq 1$.

[4]Rough means here in particular rougher than typical trajectories of Brownian motion and also semimartingales, where the latter is the class of "good integrators" in classical stochastic analysis.

Remark 4. *When the path X is of finite variation, the integration can be understood in the sense of the Stieltjes integral. When X is a path of a semimartingale we define it always in the sense of the Stratonovich integral (which is a first order calculus).*

The signature is an element of the *tensor algebra space* $T((\mathbb{R}^d))$ given by

$$T((\mathbb{R}^d)) := \{(a_0, a_1, \dots, a_n, \dots) \mid \text{ for all } n \geq 0,\, a_n \in (\mathbb{R}^d)^{\otimes n}\},$$

where by convention $(\mathbb{R}^d)^{\otimes 0} = \mathbb{R}$. The truncated signature takes values in the *n-th truncated tensor algebra space*

$$T^n(\mathbb{R}^d) := \bigoplus_{i=0}^{n} (\mathbb{R}^d)^{\otimes i}.$$

To define the coordinate signature of X, let us introduce the notation I_d , which denotes the set of *multi-indexes* with entries in $\{1, \dots, d\}$. The length of an index I is denoted by $|I|$.

Definition 6. *Let $(e_1, \dots, e_d)$ be the canoncial basis of $\mathbb{R}^d$. For any positive integer n, the space $(\mathbb{R}^d)^{\otimes n}$ is isomorphic to the free vector space generated by all the words of length n in I_d and $(e_{i_1} \otimes \cdots \otimes e_{i_n})_{(i_1,\dots,i_n)\in\{1,\dots,d\}^n}$ form a basis of $(\mathbb{R}^d)^{\otimes n}$. The coordinate signature of X indexed by $I = (i_1, \dots, i_n)$ denoted by $C_{I,T}(X)$ is defined to be*

$$C_{I,T}(X) := \int_{0<t_1<\cdots<t_n<T} \circ dX_{t_1}^{i_1} \cdots \circ dX_{t_n}^{i_n},$$

where $\circ$ stands here for a first order calculus, in particular to indicate the Stratonovic integral in the case of a semimartingale. Thus it follows that

$$\mathbb{X}_T = 1 + \sum_{n=1}^{\infty} \sum_{|I|=n} C_{I,T}(X) e_{i_1} \otimes \cdots \otimes e_{i_n} \in T((\mathbb{R}^d)).$$

Example 2. *Let X be a one-dimensional path of finite variation. Then, for every $n \geq 1$, the iterated integrals are given by*

$$C_{(\underbrace{1,\dots,1}_{n\ times}),T}(X) = \frac{(X_T - X_0)^n}{n!}$$

and thus correspond to polynomials. This form translates one to one to semimartingales in one-dimension since we deal with the Stratonovich integral.

In higher dimension these expressions become more involved. For example, if we consider the two dimensional path $t \mapsto (t, B_t)$ for B a standard Brownian motion, *we have*

$$
\begin{aligned}
&C_{(1),T} = T,\ C_{(2),T} = B_T,\\
&C_{(1,1),T} = \frac{T^2}{2},\ C_{(1,2),T} = TB_T - \int_0^T B_t dt,\ C_{(2,1),T} = \int_0^T B_t dt,\ C_{(2,2),T} = \frac{B_T^2}{2}\\
&\dots,
\end{aligned}
$$

so that we get expressions that depend on the whole path of the Brownian motion.

As signature should serve as linear regression basis, we need to introduce *linear functionals* on the signature space. By identifying the dual of $\mathbb{R}^d$ with itself we still denote the corresponding basis by $(e_1,\dots,e_d)$ and write $e_I := e_{i_1}\otimes\cdots\otimes e_{i_n}$ with multi-index $I=(i_1,\dots,i_n)$ for the basis elements of $(\mathbb{R}^d)^{\otimes n}$ (seen here as space of linear functionals). We denote by $T((\mathbb{R}^d))^*$ the space of linear functionals on $T((\mathbb{R}^d))$ induced by linear combinations of $(e_I)_{I\in I_d}$. For any multi-index $I\in I_d$, we therefore define $\ell_I \in T((\mathbb{R}^d))^*$ via

$$
\ell_I(\mathbb{X}_T) = e_I(\mathbb{X}_T) = e_I(X_T^{(|I|)}) = C_{I,T}(X).
$$

The crucial and remarkable property is now that *the pointwise product of two linear functionals ℓ_I and ℓ_J* (which is clearly a quadratic functional) *is still a linear functional*. In other words every polynomial on signatures may be realized as a linear functional, which is a consequence of the following theorem going back to [35], where suddenly the shuffle Hopf Algebra shows up.

Theorem 7. *Fix two multi-indices $I=(i_1,\dots,i_n)$ and $J=(j_1,\dots,j_m)$. Then*

$$
\ell_I(\mathbb{X}_T)\ell_J(\mathbb{X}_T) = (\ell_I \sqcup\!\sqcup\, \ell_J)(\mathbb{X}_T), \tag{20}
$$

where the shuffle product $\sqcup\!\sqcup$ is recursively defined as

$$
e_I \sqcup\!\sqcup\, e_J = e_{i_1}\otimes((e_{i_2}\otimes\cdots\otimes e_{i_n})\sqcup\!\sqcup\, e_J) + e_{j_1}\otimes(e_I \sqcup\!\sqcup\,(e_{j_2}\otimes\cdots\otimes e_{j_m})),
$$

with $e_i \sqcup\!\sqcup\, 1 := e_i$ and $1 \sqcup\!\sqcup\, e_i := e_i$.

Property (20) is the crucial ingredient that guarantees that linear functionals form an algebra. Moreover, this algebra also separates points on sets of paths that are not “tree-like”, since there the signature uniquely determines the path (see [4], in particular for the definition of tree-like). Hence an application of the Stone-Weierstrass Theorem yields the announced universal approximation theorem and justifies the use of signature as a *linear* regression basis.

Theorem 8. *Let $p\geq 1$ and let K be a compact set of finite p-variation paths up to tree-like equivalences. Suppose $f: K\to\mathbb{R}$ is continuous with respect to* *the corresponding variation norm. Then for every* $\varepsilon > 0$, *there exists a linear functional* $\ell \in T((\mathbb{R}^d))^*$ *such that*

$$\sup_{(X_t)_{t\in[0,T]}\in K} |f((X_t)_{t\in[0,T]}) - \ell(\mathbb{X}_T)| < \varepsilon.$$

Remark 5. *Note that adding time as extra coordinate, that is working with* $(t,X_t)_{t\in[0,T]}$ *instead of* $(X_t)_{t\in[0,T]}$, *allows to avoid tree-like equivalences, which we shall do subsequently.*

4 Aspects of universality of affine and polynomial processes

We are now ready to connect the two previous sections and to show how signature relates to affine and polynomial processes.

Let us consider the following generic class of diffusion models (here for simplicity in one dimension) driven by some Brownian motion B and given by

$$dX_t = b(\widehat{\mathbb{X}}_t)dt + \sigma(\widehat{\mathbb{X}}_t)dB_t, \tag{21}$$

where $(\widehat{\mathbb{X}}_t)_{t\geq 0}$ denotes the signature of $t \mapsto (t,X_t)$ and where $b,\sigma \in T((\mathbb{R}^2))^*$. Note that as a consequence of Theorem 8 choosing b and σ appropriately allows to approximate any continuous path functional arbitrarily well so that we deal here with a *truly general class of diffusions whose coefficients can depend on the whole path.* We here do not deal with the question of existence and uniqueness of solutions to this equations but simply suppose that X exists and solves (21) in an appropriate sense on an appropriate state space S. On non compact state spaces this can be a crucial question.

Even though the dynamics of (21) have at first sight nothing in common with an affine or polynomial process as of Definition 1, they are surprisingly always the projection of an affine process, namely their signature process $(\widehat{\mathbb{X}}_t)_{t\geq 0}$. This is expressed in the following theorem proved in [16], which we here state in an informal manner without proof. In particular, we do not give the precise definition of an affine process taking values in $T((\mathbb{R}_+ \times S))$, but analogously as in Definition 1 it corresponds to a process where the drift and the squared diffusion coefficient depend in an affine (here actually, linear) way on the state $\widehat{\mathbb{X}}_t$.

Theorem 9. *Let X be a solution of* (21) *on some state space S. Then the signature process* $(\widehat{\mathbb{X}}_t)_{t\geq 0}$ *of* $t \mapsto (t,X_t)$ *is an affine (actual linear) process taking values in* $T((\mathbb{R}_+ \times S))$. *Hence X is the projection of an affine process and its characteristic function can therefore be computed via*

$$\mathbb{E}[\exp(iu(X_t - X_0))] = \exp(\psi_0(t)), \quad u \in \mathbb{R},$$

where $\psi_0(t) \in \mathbb{C}$ is here the first component of the solution of an (potentially infinite dimensional) Riccati ODE with initial value $\psi(0) = iue_2$.

Moreover, by exploiting the polynomial property, linear functionals (corresponding to a multi-index I) of the expected signature can be computed via

$$\ell_I(\mathbb{E}[\widehat{\mathbb{X}}_t]) = \mathbb{E}[\ell_I(\widehat{\mathbb{X}}_t)] = c_0(t), \tag{22}$$

where $c_0(t)$ is the first component of the solution of the (potentially infinite dimensional) linear ODE

$$\partial_t c(t) = L_1 c(t), \quad c(0) = e_I,$$

with L_1 denoting the first dual operator in this setting (analogous to (16) *in Definition 3).*

To explain the significance of this result some remarks are in order.

Remark 6. *The above theorem asserts that the methodology of affine and polynomial processes that we saw in Section 2 can be transferred to generic classes of diffusions with path dependent coefficients as defined in* (21). *Note here that the initial value of $(\widehat{\mathbb{X}}_t)_{t\geq 0}$ is $(1,0,0\ldots)$. Therefore only the first component of the solutions of the corresponding ODEs enters in the above formulas.*

The reason why $(\widehat{\mathbb{X}}_t)_{t\geq 0}$ is an affine process is a consequence of Theorem 7. Indeed, the characteristics of $(\widehat{\mathbb{X}}_t)_{t\geq 0}$ are actually polynomials, but the shuffle product allows to linearize them. The prize to pay is that one has to increase the dimension and consider the infinite dimensional signature process. Sometimes the analysis can be restricted to the truncated signature which is for some models a polynomial (but non affine) process. This is for instance the case for Brownian motion (see [15]).

The fundamental property, namely that every polynomial on signatures may be realized as a linear functional, implies also that the law of the signature when it is compactly supported is determined by its expected signature. This result has been extended in [7] by introducing exponential moment conditions. All these conditions correspond to the situation when moments determine the law of a random variable. Theorem 9 now gives a systematic way based on polynomial processes how to compute this expected signature for generic models given by (21).

5 Applications in Finance

Stochastic differential equations of type (21) have been recently considered in Mathematical Finance for *asset price modeling,* see e.g. [32][5]. They are likely to become a widely applied model class as they distinguish themselves in

[5] There the coefficients depend on the signature of the driving Brownian motion and not directly on the signature of the process itself.

- universality, since the dynamics of all classical models can be arbitrarily well approximated (due to Theorem 8);
- efficient simulation techniques (see e.g. [32]);
- tractability with respect to efficient calibration, pricing and hedging.

The last point builds again on Theorem 8 which allows to approximate any continuous option payoff by a linear functional of the signature. Pricing and calibration therefore crucially hinges on computing expected signature. Formula (22) gives a systematic methodology which amounts to solving a linear ODE.

Another application in finance is *learning dynamics of time series*, also called *market generation.* In the current framework this can be done by viewing the observed path as a linear map of the signature of an underlying driving signal, which should – in view of prediction and scenario simulation – be easy to sample, e.g. Brownian motion. The task then consists in finding the linear map by regressing on the signature of the driving signal. Surprisingly, it is possible to replace the classical signature which involves the rather cumbersome computation of iterated integrals in high dimensions by a finite dimensional randomized signature process of polynomial type. Indeed, as proved in [10], this also serves as an approximative linear regression basis, which gives another indication of the universality of the polynomial class. In a similar spirit we would like to analyze when affine processes taking values in $T((\mathbb{R}^d))$ (or truncations thereof) actually correspond to affine transformations of the (truncated) signature of some $\mathbb{R}^d$ valued process. This would provide a theoretical justification to actually start directly with an affine model for the signature, thus leading potentially to a universal approximation theorem via affine processes.

6 Acknowledgments

The author would like to thank the Austrian Science Fund (FWF) for the generous support through grant Y 1235 of the START-program.

References

[1] E. Abi Jaber and O. El Euch. Markovian structure of the Volterra Heston model. *Statistics & Probability Letters*, 149:63–72, 2019.

[2] E. Abi Jaber, M. Larsson, and S. Pulido. Affine Volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200, 2019.

[3] C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[4] H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. The signature of a rough path: uniqueness. *Advances in Mathematics*, 293:720–737, 2016.
[5] K. Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Annals of Mathematics*, pages 163–178, 1957.
[6] K. Chen. Iterated path integrals. *Bulletin of the American Mathematical Society*, pages 831–879, 1977.
[7] I. Chevyrev and T. Lyons. Characteristic functions of measures on geometric rough paths. *The Annals of Probability*, 44(6):4049–4082, 2016.
[8] C. Cuchiero. Polynomial processes in stochastic portfolio theory. *Stochastic processes and their applications*, 129(5):1829–1872, 2019.
[9] C. Cuchiero, D. Filipović, E. Mayerhofer, and J. Teichmann. Affine processes on positive semidefinite matrices. *Annals of Applied Probability*, 21(2):397–463, 2011.
[10] C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega, and J. Teichmann. Approximation of dynamics by randomized signature. *Working paper*, 2019.
[11] C. Cuchiero, M. Keller-Ressel, E. Mayerhofer, and J. Teichmann. Affine processes on symmetric cones. *Journal of theoretical probability*, 29(2):359–422, 2016.
[12] C. Cuchiero, M. Keller-Ressel, and J. Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16(4):711–740, 2012.
[13] C. Cuchiero, M. Larsson, and S. Svaluto-Ferro. Polynomial jump-diffusions on the unit simplex. *The Annals of Applied Probability*, 28(4):2451–2500, 2018.
[14] C. Cuchiero, M. Larsson, and S. Svaluto-Ferro. Probability measure-valued polynomial diffusions. *Electronic Journal of Probability*, 24, 2019.
[15] C. Cuchiero and S. Svaluto-Ferro. Infinite dimensional polynomial processes. *Preprint available at arXiv:1911.02614*, 2019.
[16] C. Cuchiero, S. Svaluto-Ferro, and J. Teichmann. Affine processes are universal. *Working paper*, 2020.
[17] C. Cuchiero and J. Teichmann. Markovian lifts of positive semidefinite affine Volterra-type processes. *Decisions in Economics and Finance*, 42(2):407–448, 2019.
[18] C. Cuchiero and J. Teichmann. Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case. *Journal of Evolution Equations*, pages 1–48, 2020.
[19] D. Duffie, D. Filipović, and W. Schachermayer. Affine processes and applications in finance. *Annals of Applied Probability*, 13:984–1053, 2003.
[20] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29(1):3–38, 2019.
[21] A. Etheridge. *An Introduction to Superprocesses*. University lecture series. American Mathematical Society Volume 20, 2000.
[22] S. N. Ethier and T. G. Kurtz. Fleming–Viot processes in population genetics. *SIAM Journal on Control and Optimization*, 31(2):345–386, 1993.
[23] R. Fernholz. *Stochastic Portfolio Theory*. Applications of Mathematics. Springer-Verlag, New York, 2002.
[24] D. Filipović and M. Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4):931–972, 2016.

[25] P. K. Friz and M. Hairer. *A course on rough paths*. Springer, 2014.
[26] M. Fukasawa, T. Takabatake, and R. Westphal. Is volatility rough? *Preprint available at arXiv:1905.04852*, 2019.
[27] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6):933–949, 2018.
[28] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.
[29] M. Keller-Ressel and E. Mayerhofer. Exponential moments of affine processes. *The Annals of Applied Probability*, 25(2):714–752, 2015.
[30] D. Levin, T. Lyons, and H. Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *Preprint available at arXiv:1309.0260*, 2013.
[31] T. J. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998.
[32] I. Perez-Arribas, C. Salvi, and L. Szpruch. Sig-SDEs model for quantitative finance. *Preprint available at arXiv:2006.00218*, 2020.
[33] K. R. Popper. *Realism and the aim of science: From the postscript to the logic of scientific discovery*. Routledge, 2013.
[34] K. R. Popper. The aim of science. *Ratio*, 1, 1957.
[35] R. Ree. Lie elements and an algebra associated with shuffles. *Annals of Mathematics*, pages 210–220, 1958.
[36] J. Ruf. Personal communication. 2018.
[37] T. Schmidt, S. Tappe, and W. Yu. Infinite dimensional affine processes. *Preprint available at arXiv:1907.10337*, 2019.
[38] P. Spreij and E. Veerman. The affine transform formula for affine jump-diffusions with a general closed convex state space. *Preprint available at arXiv:1005.1099*, 2010.

Author's address:
Faculty of Business, Economics and Statistics
Department of Statistics and Operations Research
Oskar-Morgenstern-Platz 1 A-1090 Wien, Austria
email christa.cuchiero@univie.ac.at

Internat. Math. Nachrichten
Nr. 244 (2020), 47–60

Roman Schnabl zum 80. Geburtstag

Karl Sigmund, Wolfgang Woess, Gerhard Murtinger, Harald Skarke, Barbara Stöckl, Peter Grabner, Jörg Markowitsch, Emil Simeonov, Annemarie Luger, Clemens Fuchs, Paul Surer, Wolfgang Trutschnig, Janice Goodenough

Universität Wien, TU Graz, Siemens,
TU Wien, KIWI-TV Filmprod.g.m.b.H., TU Graz,
3s Unternehmensberatung, FH Technikum Wien und minimath,
Universität Stockholm, Universität Salzburg, Universität für Bodenkultur,
Universität Salzburg, Hydrogrid

1 Vorwort von Karl Sigmund

Damals, als die Professoren noch Pfeife rauchten und ihre Vorlesungen frei vortrugen, waren wir beide junge Mathematiker, die unseren Vorbildern nacheiferten. Schon bevor ich Roman Schnabl zum ersten Mal sah, hörte ich von Hlawka und Schmetterer nur die besten Dinge über ihn. Besonders Schmetterer erwähnte immer wieder Schnabls Arbeit zu den symmetrischen Maßen über kartesischen Produkten, die in der *Zeitschrift für Wahrscheinlichkeitstheorie* erschienen war, und seine in den *Mathematischen Annalen* veröffentlichten Untersuchungen über Bernsteinpolynome. Ich lernte Schnabl bald persönlich kennen, zwar nicht näher, aber dafür dauerhaft. Noch heute sehe ich ihn gelegentlich, meist an der Speisinger Schnellbahnstation. Da trägt er einen zünftigen Rucksack und Wanderschuhe und hat offenbar einen Ausflug in den Wienerwald vor. Er lächelt freundlich und still und ist mit der Welt im Reinen.

Nie bin ich in der Lage gewesen, ihm eine Episode in Erinnerung zu rufen, die sich vor einem halben Menschenalter ereignet hat. Er hat sie höchstwahrscheinlich vergessen – obwohl, wer weiß, vielleicht nicht, sein Gedächtnis ist legendär. Für mein Leben war die Episode jedenfalls eine Weichenstellung.

ISSN 0020-7926

Das kam so: Der Chemiker Peter Schuster – auch er ein junger Professor in unserem Alter – hatte gemeinsam mit dem Göttinger Nobelpreisträger Manfred Eigen ein Modell aufgestellt, das eine Reaktion der präbiotischen Evolution beschreiben sollte: den sogenannten n-gliedrigen Hyperzyklus. Die entsprechende chemische Kinetik liefert eine nichtlineare Differentialgleichung, die eine sonderbare Abhängigkeit von n aufweist: stabiles Gleichgewicht für $n = 2, 3, 4$ und einen stabilen Grenzzyklus für $n > 4$. Von Differentialgleichungen wusste ich damals so gut wie gar nichts: Die Vorlesung, die ich gehört hatte, war die einzige gewesen, die mich regelrecht abgestoßen hatte, eine öde Ansammlung von Rezepten. Jetzt wollte ich aber doch verstehen, was da beim Hyperzyklus vor sich ging. Nach vielen Fehlschlägen und fruchtlosen Grübeleien – ich erinnere mich an ein Modell, das ich aus den Drähten von Pfeifenputzern gebastelt hatte – ging ich an die TU, um mir ein paar Plots auszudrucken. Der große Stolz dort war ein Analog-Computer. Ich erinnere mich, wie mir die Mitarbeiter (Hörtlehner, Vogl und Rattay, alte Freunde aus Studienzeiten) vorführten, wie man durch Umstecken von ein paar bunten Kabeln das kuriose Gerät, das einen halben Raum einnahm, dazu bringen konnte, neue Rechenaufgaben zu lösen.

Roman Schnabl driftete in den Raum, angekündigt vom Duft seiner Pfeife, lächelte freundlich und still und hörte zu, wie wir ihm erklärten, was wir da gerade trieben. Am nächsten Tag, als ich wieder vorbeikam, zeigte er mir, gütig schmunzelnd wie stets, auf einem halben Blatt Papier seine Lösung: eine Ljapunov-Funktion. Es war die erste Ljapunov Funktion meines Lebens – und übrigens mein letzter Blick auf den Analogrechner seligen Angedenkens.

Natürlich bedankte ich mich in der Publikation, die 1978 im *Bulletin of Mathematical Biology* erschien, bei Professor Schnabl für den wertvollen Hinweis. Aber genau genommen hätte ich das bei allen meinen weiteren Publikationen auch tun müssen: Denn durch das Erfolgserlebnis, zu dem mir Schnabl verholfen hatte, wurde aus meiner Stippvisite in der Biomathematik mein wissenschaftliches Lebensthema. Die Biomathematik und evolutionäre Spieltheorie ließen mich nicht mehr los.

Lang habe ich nach einer Gelegenheit gesucht, ihm das zu sagen. Eine Zufallsbegegnung am Bahnsteig war dazu wohl nicht der rechte Platz. Aber jetzt, in diesem Heft der IMN, ist es wirklich höchste Zeit, zu bekunden, wie sehr ich in Schnabls Dankesschuld stehe. Die folgenden Beiträge seiner ehemaligen Schülerinnen und Schüler zeichnen das Bild eines hervorragenden Mathematikers, dessen breites Wissen ebenso beeindruckend ist wie seine Gabe, es zu vermitteln. Sie alle hatten das Privileg, bei Roman Schnabl Vorlesungen zu hören. Ich bin zu alt für einen Schnabl-Jahrgang. Ich kann nur von einer gemütlich-entspannten Lektion berichten, die kaum zehn Minuten dauerte. Aber sie hat mich geprägt.

Ich wünsche dem Jubilar noch viele schöne Ausflüge, sowohl in die Mathematik als auch in den Wienerwald!

2 Wolfgang Woess, Schnabl-Jahrgang 73

Vor mir liegt ein schlankes Skriptum (ja, ich habe es bis heute aufgehoben!):

```
MATHEMATIK C1
Wintersemester 1971/72
Definitionen und Sätze
o.Prof. Dr. E. BUKOVICS
Doz. Dr. R. Schnabl
```

Als ich im Herbst 1973 an der TU Wien Technische Mathematik zu studieren begann, war Roman Schnabl – Firnberg sei Dank – bereits ao.Professor.

Aus meinem Diplomstudium 1973-1978 sind mir zwei hervorragende Vortragende im Gedächtnis geblieben. Der *zweite* war Edmund Hlawka: damals waren die Mathematikstudien an TU und Uni Wien ziemlich streng getrennt, aber im 2. Studienabschnitt durfte man Hlawka-Vorlesungen in der Strudlhofgasse als Wahlfächer belegen. Für mich war das zunächst die elementare Zahlentheorie (dabei lernte ich auch die Studienkollegen Lettl und Tichy kennen; später sind wir alle drei in Graz gelandet).

Der *erste* der zwei hervorragenden Vortragenden war Roman Schnabl, und dieser erste Eindruck war ein nachhaltiger. Rückblickend ist es nach so vielen (47 !) Jahren gar nicht leicht, dies an spezifischen Erinnerungsstücken festzumachen. Vor mir liegt auch meine Mitschrift, oder besser deren Reinschrift, der Vorlesungen Mathematik C_1, C_2 und C_3. (Auch diese habe ich bis heute aufgehoben, und sie auch herangezogen, als ich zu Beginn des Jahrtausends an der TU Graz einige Male die Analysis 1 für Mathematik halten durfte.) Die Inhalte sind jene, die man von einem guten Analysis-Zyklus für Mathematik-Studierende erwartet. Was den bleibenden Eindruck ausmacht, war vor allem der Vorlesungs*stil.*

Professor Schnabls Vortrag war selbstverständlich immer frei, und der freie Vorlesungsstil war dezidiert sein eigener. Die Vorlesungen waren keineswegs immer glattpoliert; kleine Nachdenkpausen, Selbstkorrekturen – auch nachträglich – waren durchaus dabei: Mathematik-Vermittlung als permanenter intellektueller Prozess. Was mir Roman Schnabl durch sein Auftreten bleibend mitgegeben hat, ist die Faszination der Mathematik als Denkarbeit, die Faszination und Kreativität des mathematischen Denkens und Wissens. Sein Fachwissen ist enorm breit, tief und umfangreich. Vor diesem verblasst für mich mein eigenes Streben nach Publikationen und sonstigen bescheidenen Erfolgen in meinem kleinen Arbeitsgebiet.

Wenn ich jetzt spezifische Erinnerungen aus meinem Gedächtnis hervorkramen soll, dann fällt mir z.B. meine Prüfung aus Mathematik C_2 ein. Der mündliche Prüfungsteil fand in den typischen Pfeifenrauchschwaden im Büro von Professor Schnabl statt. In meiner Mitschrift gab (gibt) es eine Lücke im Paragraphen über Funktionen mit beschränkter Variation auf dem Einheitsintervall – das Kompaktheitsargument. Und genau nach diesem wurde ich gleich als Erstes gefragt.

Ich hatte mir meinen eigenen Beweis zurechtgelegt, und war tapfer genug, diesen anzudeuten (mit Intervallhalbierung). “So haben wir es aber in der Vorlesung nicht gemacht” war der Kommentar des Professors. “Ja, aber so geht es auch” war meine trotzige Antwort. Daraufhin wurde ich sofort mit einem “Sehr gut” entlassen und war geradezu entäuscht, mein sonstiges Wissen gar nicht vorbringen zu können.

Im Folgenden besuchte ich noch Seminare und Spezialvorlesungen bei Roman Schnabl. Insbesondere eine “Potentialtheorie” wurde später für mich nützlich. Und in späteren Jahren, selbst nicht mehr in Wien, besuchte ich bei Wien-Besuchen immer wieder die Kollegen an der TU, auch meinen Lehrer Roman Schnabl. Vor 20 Jahren, gerade nach Österreich zurückgekehrt, hatte ich die große Ehre, beim Kolloquium zum 60. Geburtstag von Roman Schnabl einen Vortrag halten zu dürfen. Danke!

3 Gerhard Murtinger, Schnabl-Jahrgang 81

Ich habe von 1981–1986 Technische Mathematik an der TU-Wien studiert und anschließend an der Fakultät für Maschinenbau bei Prof. Troger promoviert. Ich bin seit 1991 bei Siemens Mobility beschäftigt, vormals Simmering Graz Pauker, und arbeite im Bereich Computersimulation für Schienenfahrzeuge: Finite Elemente, CFD, Crash. Ich habe in den Jahren 1981–1983 Analysis I-IV bei Prof. Schnabl gehört. Noch heute denke ich mit großer Freude an diese vier Semester zurück. Ich erinnere mich an den brillanten Vortragsstil von Prof. Schnabl, an sein umfassendes mathematisches Wissen, seine Begeisterung für tiefgehende mathematische Zusammenhänge, die er zu transportieren verstand und nicht zuletzt an den herrlichen Pfeifenduft bei den mündlichen Prüfungen. Neben dem Grundlagenwissen der Analysis hat Prof. Schnabl vor allem selbstständiges mathematisches Denken vermittelt. Diese Fähigkeiten sind bei meiner beruflichen Tätigkeit, insbesondere beim Verständnis der mathematischen Grundlagen vieler für den Schienenfahrzeugbau relevanter technisch-physikalischer Gebiete, von unschätzbarem Wert. Mit lieben Grüßen und Wünschen zum 80. Geburtstag!

4 Harald Skarke, Schnabl-Jahrgang 81

In Professor Schnabls Vorlesungen hatte man nie den Eindruck, es würde einfach nur wiedergegeben, was andere sich früher ausgedacht hatten – immer war es so, als ob man mittendrin in der Entwicklung des Stoffes sei. Dabei kam aber die Einordnung in größere Zusammenhänge nicht zu kurz: Unvergesslich sind mir etwa wiederholte Hinweise der Art “Das werden wir erst im Rahmen der Funktionentheorie richtig verstehen”, die mich schon extrem gespannt sein ließen.

Ich hatte dann das Privileg, auch die Funktionentheorie bei Roman Schnabl hören zu dürfen; und ja, die äußerst hohen Erwartungen wurden tatsächlich zur Gänze erfüllt.

5 Barbara Stöckl, Schnabl-Jahrgang 81

Ich habe meinen beruflichen Weg nicht in der Wissenschaft gemacht, denke aber als TV-Journalistin immer noch mit großer Freude und Respekt an meine Jahre an der TU Wien. Und an die Persönlichkeiten, die mir dort begegnet sind, die merkt man sich nämlich! Und Prof. Roman Schnabl war und ist so eine Persönlichkeit! Außerdem einer der in der mathematischen Analysis umfassend gebildetsten und herausragendsten akademischen Lehrer der TU! Er hat uns sein Wissen vermittelt, als würden wir schon selbst alles wissen, freundlich, genial, ein Lehrer im allerbesten Sinn. Und eine wissenschaftliche Kapazität von außerordentlicher Bescheidenheit – Ruhm, Publikationswahn, das war nicht seine Sache. Es ging ihm ums Eigentliche, wie selten begegnet man solchen Menschen! Danke für Erkenntnisse und Begegnungen und alles Gute zum 80. Geburtstag!

6 Peter Grabner, Schnabl-Jahrgang 85

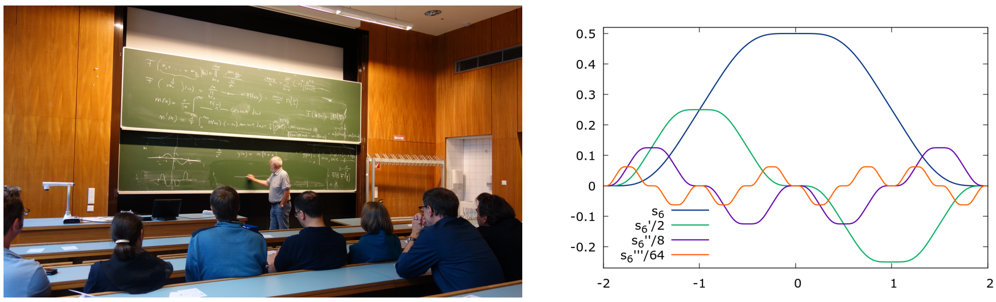
“Es soll eine Eins geben, die von der Null verschieden ist.” war einer der ersten Sätze der Analysis 1 Vorlesung, die ich bei Roman Schnabl im Wintersemester 1985 hören durfte. In der ersten Vorlesungsstunde diskutierte er die Axiome des archimedisch angeordneten Körpers, und sofort packte mich damals sein nonchalanter Vortragsstil. Anscheinend mühelos entwickelte er vor uns Erstsemestrigen die Analysis, immer ohne Unterlagen, dafür mit großem Engagement und mitreißender Begeisterung. Gelegentlich passierten auch kleine Schnitzer. Ich erinnere mich etwa an den Beweis, dass das Cauchy-Produkt einer absolut konvergenten Reihe und einer konvergenten Reihe konvergiert. Schließlich gelang der Beweis nur, indem auch noch die zweite Reihe absolut konvergent sein musste, was er mit dem trockenen Kommentar abschloss “Jetzt stimmt zwar der Beweis, dafür gefällt mir der Satz nicht mehr”. Jedenfalls hatten wir dabei gelernt, wie Mathematik funktioniert, und gesehen, wie ein Mathematiker denkt. Insgesamt habe ich bei ihm den gesamten Analysis-Zyklus (1-3) und die Funktionentheorie absolviert, alle diese Lehrveranstaltungen fand ich äußerst inspirierend und zugleich auch fordernd. Danach suchten einige Kollegen und ich an jedem Semesterbeginn die Schaukästen des Instituts ab, was Roman Schnabl denn an Spezialvorlesungen anbot. So hörte ich dann bei ihm noch eine Vorlesung über Differentialformen.

Für meine eigene Lehrtätigkeit, vor allem für die Analysis, habe ich mitgenommen, dass eine Analysis-Vorlesung darin besteht, dass man “Analysis macht”, also analytische Sätze beweist, fast egal, welche. Man lernt ein Bündel von Methoden kennen, die man nur dann richtig verstanden hat, wenn man sie verwenden kann. Und genau das hat mir Roman Schnabl beigebracht. Ich blicke immer wieder gerne auf den Beginn meines Studiums zurück, dessen erste Semester er maßgeblich geprägt hat. Zum 80. Geburtstag wünsche ich Roman Schnabl alles Gute, Gesundheit und *ad multos annos*.

7 Schnabl-Jahrgang 89

Am 1. Oktober 2019 versammelten sich etwa 30 ehemalige Studierende des Schnabl-Jahrgangs 89, um 30 Jahre nach ihrem Studienbeginn dieses Jubiläum so zu feiern, wie es damals begann: Mit einer Schnabl-Vorlesung!

Prof. Schnabl ist jung geblieben! Wir fühlten uns in die Studienzeit zurückversetzt, seinen Ausführungen lauschend, die er auf Dinge aufbaute, die er uns seinerzeit erzählte: in gewohnter Intonation, mit den typischen, seine tiefen Einsichten zeigenden Bonmots und den nichttriviale Querverbindungen herstellenden Exkursen.



Links: Die Vorlesung vom 1.10.2019; rechts: Die Funktion s_6 und ihre ersten 3 Ableitungen (geeignet skaliert).

Im Zuge der Vorlesung hatte er, aufbauend auf einer einfachen parametrischen, reellwertigen Funktionenfamilie, mithilfe der Faltung eine Funktionenfolge $(s_n)_{n\in\mathbb{N}}$ konstruiert, die sehr schnell und gleichmäßig gegen eine Grenzfunktion konvergiert.

Die Funktionen erben aus den Basisfunktionen über die Faltung ein paar Eigenschaften, die sie der Grenzfunktion weitervererben. O-Ton Schnabl: "Gute Eigenschaften vererben sich über die Faltung. Überhaupt erkennt man gute Eigenschaften daran, dass sie sich über die Faltung vererben."

Die Grenzfunktion hat ein paar zusätzliche schöne Eigenschaften, darunter die fraktale Struktur in den Ableitungen, die in der obigen Abbildung beispielhaft an s_6 illustriert wird.

Wir, die ehemaligen Studierenden des Schnabl-Jahrgangs 89, gratulieren Prof. Schnabl recht herzlich zu seinem Jubiläum!

8 Jörg Markowitsch, Schnabl-Jahrgang 89

“Wenn sich die Funktion nicht differenzieren lässt, ist sie entweder nicht differenzierbar, oder Sie haben nicht genug geübt!” (Roman Schnabl, Vorlesung, Analysis III, 1996).

Roman Schnabl vermochte es in einer lapidaren Bemerkung, die vermeintlich logischen und philosophischen Grundlagen der Mathematik sowie die Grenzen der Mathematikdidaktik gleichermaßen auf den Punkt zu bringen und dabei auch noch seinen speziellen Humor aufblitzen zu lassen. Im Zuge meiner Dissertation über implizites Wissen und Verstehen in der Mathematik hatte ich das Vergnügen, Roman Schnabl und andere große Mathematiker, wie etwa Edmund Hlawka oder Friedrich Hirzebruch, zu interviewen. Die folgenden Passagen sind einem Interview, das ich leicht benebelt von passivem Pfeifenrauchen mit Roman Schnabl in seinem Büro an der TU Wien im Jahre 1995 führte, entnommen. Ich denke, sie geben einen guten Einblick sowohl in die Praxis der Mathematik als auch sein persönliches Verständnis von Mathematik und Didaktik.

Ich: Wie prüfen Sie bei einer mündlichen Prüfung, ob der Student einen Beweis oder einen Satz verstanden hat?
Schnabl: “Die erste Frage ist einmal, ob man das selbst verstanden hat. Da ist sicher ein Gefühl der Evidenz da. Man hat das Gefühl, dass es so ist. Dass man das Gefühl selbst kennt, das ist sicher irgendwie in einem gewachsen. Diese Bekanntschaft mit diesem Gefühl, dass man etwas verstanden hat. Man hat zuerst einmal im Laufe seiner Entwicklung gewisse Sachen mathematischer Natur verstanden – gewisse Rechenmethoden. Das fängt bei einfachen Sachen an: Wenn Sie eine quadratische Gleichung lösen, sehen Sie, wie man durch Umformung zu den Lösungen kommt. Dann kommen kleinere Sachen: Beweis – unter Anführungszeichen – des Pythagoras in der Mittelschule. Solche Einsichten. Und da entwickelt man ein gewisses Gefühl dafür, dass man was verstanden hat.”
“Wenn ich den Beweis führe und verstehen will, um was es geht, muss ich zwei Sachen haben. Erstens einmal muss ich das Gefühl haben, dass es logisch, formal korrekt ist. Ich sehe aber mehr. Ich sehe die Zusammenhänge, wie die Begriffe ineinanderarbeiten, wie weit sie tragen. Das macht man auch. Wenn man den Beweis geführt hat, stellt man sich nachher die Frage: Wo sind die Voraussetzungen eingegangen? Habe ich sie wirklich verwendet? Das ist eine wichtige Angelegenheit: Habe ich das ökonomisch durchgeführt? Habe ich nicht unnötige Voraussetzungen mitgeschleppt? Dann wird man sehen: die Voraussetzung ist zwar eingegangen, vielleicht könnte ich sie eliminieren. Kann ich die Voraussetzungen abschwächen – das wird immer gemacht. Dann hat man also das Gefühl, dass

es formallogisch richtig ist, und man sieht, wie die Begriffe ineinander zusammenhängen. Und beides zusammen gibt das Gefühl des Verstehens. Woher das Gefühl kommt? Ich möchte sagen: das ist ein Elementarereignis!"
"Beweisidee? Das sind keine präzisierten Begriffe. Man sagt: 'Die Idee von dem Beweis ist das!' und 'Der hat eine Beweisidee.' und 'Die arbeitet nicht, das reicht nicht.' und 'Ich brauche eine Beweisidee.' Der eine glaubt, es ist ihm etwas Neues eingefallen, dabei hat er es schon woanders gelesen, hat es aber vergessen sozusagen und übernimmt diese Methode. [...] Also die neuen Ideen sind rar. [...]."
"Und es ist ein Können des Autors, ob er einen Beweis verständlich führen kann. Je mehr ich sehe in dem Beweis, desto mehr habe ich von dem Beweis. Da gibt es lange Beweise, da kann man Schritt für Schritt folgen: "Aha, da folgt das; aha das, und da zieht er das heran, das passt...", aber die Beweisstruktur hat man nicht verstanden. Ein guter Autor gliedert den Beweis in Zwischenschritte, führt einmal die Beweisidee vor – das sind ja ganz andere Sachen, die plötzlich dazukommen. Da sieht man ja, was ein gut geführter Beweis ist, der einem viel Verständnis bringt, wenn man ihn liest."

"Wie kommt man zu einer Vermutung? Man hat eine gewisse Erfahrung. [...] Dann ist das erste Problem: Ist das überhaupt gut vermutet? Schlechte Mathematiker machen blöde Vermutungen. Die sind weit weg, um realistisch zu sein. Gute Mathematiker vermuten eher das Richtige."

Alle Zitate stammen von Roman Schnabl und sind aus: Markowitsch, Jörg (1997): Metaphysik und Mathematik. Über implizites Wissen, Verstehen und die Praxis in der Mathematik, Universität Wien, unveröffentlichte Dissertation.

9 Emil Simeonov, Schnabl-Jahrgang 89

Wenn man etwas lernen will, sucht man, ob bewusst oder unbewusst, nach Lehrern. Gute Lehrer zu finden, ist meist Glückssache.

Nachdem ich als Jugendlicher "Wind, Sand und Sterne" von Antoine de Saint-Exupéry gelesen hatte, war ich der Meinung, dass alles, was nicht direkt mit einer bestimmten physischen Konstitution verbunden ist, trainierbar, also erlernbar bzw. verbesserbar ist – dazu gehört auch das Denken. Dieser Überzeugung bin ich immer noch. Training allein reicht dazu aber nicht aus. Viele meinen, es brauche darüber hinaus noch "Talent", ohne irgendeine Evidenz dafür zu haben. Inzwischen habe ich auch dafür den richtigen Begriff gefunden. Mein Sohn hat mich vor zwei Jahren darauf hingewiesen: Es ginge um Hingabe, nicht um Talent!

1989 begann ich an der TU Wien Technische Physik zu studieren. Im zweiten Semester, also im Frühjahr 1990, stieg ich dann zusätzlich in das Studium der Technischen Mathematik ein und begegnete erstmals Professor Schnabl, der gerade Analysis II unterrichtete. Professor Schnabl war dann letztlich der Grund, weshalb ich bei der Mathematik hängenblieb und mich später vom formalen Physikstudium zurückzog. Ich kannte dort niemanden, der auch nur entfernt eine ähnliche Auseinandersetzung mit "der Sache" verkörperte wie Professor Schnabl.

Roman Schnabl hat nicht einfach "vorgetragen". Er hat "vorgedacht"! Wir im Auditorium mussten währenddessen "mit-denken" und danach "nach-denken". Es ging dabei nicht nur um "mathematische Gedanken", sondern um den Prozess des Denkens, das "Tun" an sich. Es war die Gelegenheit, an der Verbesserung meines Denkens zu arbeiten. Das Schnablsche Tempo mitzuhalten war schwer und sehr oft gar nicht möglich. Aber allein die Anstrengung, "mitzukommen" und "mitzudenken", war, im Nachhinein betrachtet, ein unglaubliches mathematisches Training. Ich versuche heute selbst, soweit es die Situation erlaubt, diesem Vorbild in meiner Tätigkeit als Hochschullehrer nachzueifern. Die Herausforderungen und Zwänge sind aber gänzlich andere geworden, was mich zu meinem nächsten Punkt führt.

Professor Schnabl publizierte nur sehr wenig. Im heutigen akademischen Betrieb des "publish or perish" undenkbar. Der Grund dafür war sein unglaublich hoher Anspruch an die Mathematik. Entweder sagte er mir das in einem Gespräch selbst oder es war eine der vielen Schnabl-Geschichten, die im Umlauf waren: Für Roman Schnabl war der Großteil der publizierten mathematischen Forschung (er bezog sich auf 80 oder 90 Prozent aller Publikationen) nicht würdig, publiziert zu werden.

Gerücht oder nicht, es wurde mir später von anderer Stelle bestätigt. Im Rahmen meiner Arbeit zur Philosophie der Mathematik war ich im Jahr 1999 bei der Konferenz "Visions in Mathematics" in Tel Aviv. Ich wollte dort einige Interviews machen. Zum ersten Mal begegnete ich vielen führenden Mathematikern, also der "Top-Liga"[1]. Es war ein ziemlich einmaliges Erlebnis, zu sehen (auch wenn ich das meiste mathematisch nicht verstand), mit welcher Hingabe diese Leute Mathematik betrieben. Im Übrigen, der einzige Mathematiker, den ich kannte und der dort sofort hineinpassen würde, war Roman Schnabl. Das dachte ich mir jedenfalls damals und so sehe ich das heute immer noch. Jedenfalls sprach ich bei meinem Interview mit Timothy Gowers die Sache mit den Publikationen an und siehe da, Gowers meinte sogar, dass dieser Schnablsche Prozentsatz eher noch zu niedrig sei!

Diese kompromisslose Haltung, was Qualität betrifft, beeindruckt mich immer noch sehr. Wenn man keine ausreichend guten Ideen hat, soll man auch nichts publizieren! Aus, Schluss, basta! Das Wichtigste dabei ist aber, und dies konnte man von Roman Schnabl lernen, diese Forderung an sich selbst zu richten und nicht an andere.

Einmal darauf angesprochen, wie er denn "Verstehen" prüfe, meinte Professor Schnabl, dass das sehr schwer sei, aber es ginge unter anderem um "eine gewisse Geläufigkeit im Umgang mit dem Kalkül". Diese Einstellung spiegelt sich auch im

[1] http://www.math.tau.ac.il/ rudnick/conf2000.html

folgenden Schnablschen Bonmot wider: "Wenn sich die Funktion nicht integrieren lässt, so ist sie entweder nicht integrierbar oder Sie haben nicht genug geübt."

Wichtig ist also der menschliche Faktor, das habe ich beim Schnablschen Unterricht so erlebt – der persönliche Zugang zur Mathematik und die Zur-Schau-Stellung dieses Zugangs als *einziges* Unterrichtsprinzip. Dadurch wird der Unterricht "subjektiv", wobei aber das Subjekt, in diesem Fall Schnabl, nach der tiefsten Durchdringung des "Objektiven", des "mathematischen Spiels" strebt. Eine fast schon buddhistische Verschmelzung zwischen dem Menschen und seiner Handhabung des Abstrakten. Egal, ob man den Details folgen konnte oder nicht, diese zur Schau gestellte und von uns erlebte Verschmelzung war viel wichtiger ...und der Humor! Wenn ich über die Dinge lachen kann, so beherrsche ich sie und nicht umgekehrt.

Professor Schnabl hatte immer Zeit für Fragen – gefühlt "unendlich" viel. Ich kann mich an einen von vielen Fällen erinnern, den ich wohl nie vergessen werde. Es war eine kalte Jahreszeit und ich hatte eine Frage, die ich Professor Schnabl stellen wollte. Ich ging zu seinem Zimmer, aber er war nicht da. Schließlich traf ich ihn am Gang am Weg zum Aufzug, im Mantel – offensichtlich am Weggehen. Ich stellte ihm die Frage und er kehrte in aller Ruhe um, ging mit mir zurück in sein Zimmer, zog den Mantel aus und erläuterte mir die Sache ungefähr eine halbe Stunde lang. Als er fertig war, zog er mit derselben Ruhe den Mantel wieder an und ging wieder zum Aufzug. Inzwischen habe ich die Frage und Schnabls Ausführungen nicht mehr parat, aber diese Einstellung, jede Frage ernst zu nehmen und sich für sie Zeit zu nehmen, werde ich nie vergessen. Die Beantwortung von Fragen ging auch weit über eine einfache konkrete Antwort hinaus und als Fragender lernte man sehr viel aus Schnabls Antworten. Wiegt all das das Nicht-Publizieren auf?

In seinem Artikel "On Proof and Progress in Mathematics"[2] weist William Thurston darauf hin, dass das Beweisen von neuen Sätzen nicht die einzige mathematische Tätigkeit ist, die "credits" verdient: "I think that our strong communal emphasis on theorem-credits has a negative effect on mathematical progress. If what we are accomplishing is advancing human understanding of mathematics, then we would be much better off recognizing and valuing a far broader range of activity." Gefühlt ist die Leistung von Roman Schnabl in Bezug auf "advancing human understanding of mathematics" mindestens so bedeutend wie etliche andere mathematischen Leistungen, die vielleicht zu berühmten Preisen geführt haben.

Professor Schnabl hat Mathematik also nicht nur vorgedacht, sondern auch vorgelebt. Meine Studienkolleginnen und -kollegen und ich konnten das im Oktober letzten Jahres wieder erleben, als wir das 30-Jahre Jubiläum unseres Studienbeginns im selben Hörsaal wie damals mit einer Vorlesung von Professor Schnabl

[2] https://arxiv.org/abs/math/9404236

feierten. Ich konnte somit den Beginn des Mathematik-Studiums “nachholen”. Das Wunderbare dabei war, dass Roman Schnabl nicht nur unsere Einladung annahm, sondern zu unserem Erstaunen und unserer Freude genauso war wie eh und je. Als hätte jemand die Uhr um 30 Jahre zurückgedreht. Es war wieder dieser mathematisch tiefe, aber auch humorvolle und manchmal selbstironische Zugang. Bei der Vitalität, die Professor Schnabl ausstrahlte, rechne ich fest damit, dass wir ihn zu unserem 35. und auch zum 40. Jubiläum wieder einladen werden. Für das Glück, dass ich Professor Schnabl begegnen und von ihm lernen durfte, bin ich unendlich dankbar!

10 Annemarie Luger, Schnabl-Jahrgang 91

Im Oktober 1991 begann ich mein Mathematikstudium an der TU Wien, und da mit Analysis bei Herrn Prof. Schnabl. Ich kann mich noch genau an die allererste Vorlesung im alten EI erinnern, Peano-Axiome. Vom ersten Tag an hatte ich das Gefühl von Glück: hier darf man “richtige Mathematik” erleben! Erst viel später habe ich doch erst richtig verstanden, und dann auch immer wieder bestätigt gesehen, was für ein großes Privileg es war, Analysis bei Herrn Prof. Schnabl lernen zu dürfen.

Wenn er unterrichtet, wird Mathematik lebendig! Man konnte sehen, wie etwas ensteht und bekam schon sehr früh eine Idee davon, wie ein Mathematiker denkt. Ein einzigartiger – und vermutlich hier öfter erwähnter – Aspekt in Analysis 1-3 war ja, dass Herr Prof. Schnabl ohne schriftliche Vorbereitung an der Tafel stand. Und manchmal, sehr selten, aber doch, gab es da die Momente, in denen im ersten Anlauf etwas nicht ganz richtig wurde, wo vielleicht eine Aussage im Nachhinein umformuliert werden musste, ein Beweis neu begonnen werden musste oder ein Beispiel anders angegangen werden musste. Das waren – aus meiner Sicht – die richtig großen Momente, in denen Gedanken sichtbar wurden!

Kleine Erinnerungen an diese Zeit gibt es viele, wie zum Beispiel die anschauliche Erklärung, dass man die reelle Achse nimmt, verbiegt und zu einem Kreis zusammenfügt, um die reellen Zahlen zu kompaktifizieren, die eine staunende Übungsgruppe in Analysis 2 zu hören und sehen bekam. Oder die geduldige Wiederholung der Ableitungen für cos und sin in der Übung Analysis 3 für einen Studenten, der das nicht parat hatte. Oder der ermutigende Kommentar auf eine Frage zu einem schwierigen alten Prüfungsbeispiel in Funktionentheorie, die jedoch nicht seine war: “Gut, dass ich die Prüfung jetzt nicht machen muss!”. Oder die Frage an uns Studenten, die gerade eben die Topologieprüfung mit “Sehr gut” bestanden hatten: “Haben Sie noch Fragen?” Und wir hatten! Und bekamen eine ausführliche Antwort!

Die Vorlesungen und auch Übungen bei Prof. Schnabl hatten eine Vielschichtigkeit, die nur schwer zu erklären ist. Nach Analysis 3 hatten ein Kollege und ich als Studenten das Gefühl, dass in dieser Vorlesung noch so viel mehr enthalten war, das wir nicht mitbekommen hatten (obwohl wir beide die Prüfung mit "Sehr gut" abgeschlossen hatten), dass wir uns den Luxus gönnten und beim nächsten Durchgang zwei Jahre später alles noch einmal anhörten. Was für eine gute Entscheidung! Diese zweite Analysis 3 hat mir noch einmal viel Wissen gebracht, für das ich zwei Jahre zuvor nicht bereit, inbesondere aber auch das Erlebnis einer offenbar völlig überarbeiteten Vorlesung mit anderer Struktur und neuen Beispielen.

Prof. Schnabls Student gewesen zu sein, hat mich sicherlich stark geprägt, mathematisch und pädagogisch. Wenn ich heute selber im Hörsaal stehe, hoffe ich, diese Erfahrungen färben zumindest ein wenig auf meine Vorlesungen ab.

Vielen Dank und alles Gute!

11 Clemens Fuchs, Schnabl-Jahrgang 95

Die Analysis-Vorlesung von Prof. Schnabl war ein tägliches Highlight, 9-10 Uhr, direkt nach der Lineare Algebra-Vorlesung bei Prof. Havlicek, im Pötzlhörsaal in der Gußhausstraße 27-29 im 4. Wiener Gemeindebezirk unweit von Freihaus und Hauptgebäude der TU Wien. Die Lehrveranstaltung von Prof. Schnabl war dabei geprägt von einem Vortragenden, der scheinbar ohne Vorbereitung und mühelos die Analysis von Grund auf entwickeln konnte. Auch später noch war und ist die Mitschrift aus dieser Vorlesung eine Quelle von erstaunlichen Hinweisen, die wir damals als Studierende gar nicht alle richtig begreifen, geschweige denn würdigen konnten. Meine Mitschrift gehört daher nach wie vor zu einem gut behüteten Schatz. Die zur Vorlesung gehörenden Übungen waren durchwegs nicht unangenehm. In meinem Fall in den ersten beiden Semestern noch bei Doz. Taschner, im dritten Teil dann beim Meister selbst (vermutlich aufgrund der bereits stark zurückgegangenen Studierendenzahl). Besonders erinnere ich mich daran, dass ich in Analysis III zu Aufgabe 33 an der Tafel war, die lautete: *Sei $f(x,y) = x^3 - 3xy^2$ (Affensattel). Skizziere die Niveaulinien. Bestimme in (a,b) die Tangentialebene an $z = f(x,y)$ und diskutiere lokal um (a,b) den Schnitt der Fläche $z = f(x,y)$ mit der Tangentialebene in (a,b).* Prof. Schnabl hat mich mit Begeisterung gleich weitere Schnitte berechnen lassen, um so den Graphen des Affensattels gut verstehen zu können. Meine Begeisterung vom Vortragsstil und Enthusiasmus von Prof. Schnabl war so groß, dass ich als Draufgabe noch eine Spezialvorlesung über Topologie bei ihm hörte. Die Prüfungen waren in schriftlichen und mündlichen Teil gegliedert. Nach der schriftlichen Prüfung, betreut durch Assistenten, konnte man die Ergebnisse dieser sowie den Termin der mündlichen Prüfung per Aushang an der Bürotür von Prof. Schnabl in Erfahrung bringen. Beim mündlichen Teil bot sich die Gelegenheit, eingehüllt in dichten Pfeifentabakschwaden (im Freihaus wurde zwar kurz vorher ein Rauchverbot verhängt, das Rauchen in den eigenen Büroräumlichkeiten wurde aber noch toleriert), dann sein Wissen über Analysis zum Besten zu geben. Auch heute noch, 25 Jahre später, versuche ich, den von Prof. Schnabl vorzelebrierten Unterrichtsstil, d.h. ohne Unterlagen sowie durch spontantes Abweichen von den geplanten Inhalten, z.B. Vorrechnen von weiteren Beispielen, im Laufe der Unterrichtseinheit in meinen eigenen Vorlesungen umzusetzen. Ich hoffe, dass ich bei meinen eigenen Studierenden einen ähnlich kompetenten und nachhaltigen Eindruck erzeuge, wie Prof. Schnabl ihn bei mir hinterlassen hat. Dem Jubilar wünsche ich auf diesem Wege Gesundheit und alles Gute!

12 Paul Surer, Schnabl-Jahrgang 95

Es ist jetzt schon eine ganze Weile her, dass ich als Studienanfänger im Hörsaal gesessen bin. Trotzdem kommt es mir manchmal vor, als wäre die Analysis-Vorlesung bei Prof. Schnabl erst gestern gewesen. Aber warum? Natürlich gibt es viele Anekdoten und heitere Erlebnisse aus dieser Zeit, aber daran liegt es nicht, dass meine Erinnerungen an genau seinen Unterricht noch derart präsent sind. Es sind seine Liebe und Leidenschaft für die Mathematik, die sehr ansteckend auf mich gewirkt und einen tiefen und bleibenden Eindruck bei mir hinterlassen haben. Es ist mir daher eine besondere Freude, Prof. Schnabl auf diesem Wege alles Gute zum 80er zu wünschen.

13 Wolfgang Trutschnig, Schnabl-Jahrgang 95

Auf die Studienzeit an der TU Wien zurückzublicken, heißt in meinem Fall zurückzublicken auf eine solide Grundausbildung, auf zahlreiche interessante Vorlesungen und motivierte Vortragende. Von allen Professoren am besten in Erinnerung geblieben ist mir, obwohl ich nach dem Studium keinen Kontakt mit mehr ihm hatte, Prof. Schnabl, bei dem ich Analysis I–III und Topologie hören durfte. Nicht nur, weil seine Vorlesungen inhaltlich spannend waren, sondern insbesondere aufgrund seines einzigartigen Vortragsstils: Prof. Schnabl trug ohne Verwendung jeglicher Unterlagen frei vor, wodurch es passieren konnte, dass ab und zu ein Beweis nicht im ersten Anlauf funktionierte. Aus genau diesem Scheitern im ersten Schritt, dem darauffolgenden spontanen Entwickeln neuer Ideen und der sofortigen Umsetzung selbiger habe ich mehr für das Leben und insbesondere für die wissenschaftliche Karriere gelernt als in allen anderen Grundkursen zusammen. Ich wünsche Prof. Schnabl auf diesem Wege alles Gute zum 80er und nutze die Gelegenheit, um mich spät, aber doch für wunderbare und nachhaltige Vorlesungen zu bedanken.

14 Janice Goodenough, Schnabl-Jahrgang 01

“Sehr geehrte Damen und Herren, wir befinden uns in einem archimedisch angeordneten Körper.” Mit diesem Satz eröffnete Prof. Schnabl die allererste Analysis-Vorlesung des Jahrgangs 2001 und er ist mir bis heute in Erinnerung geblieben. Diese Vorlesung – wie auch Prof. Schnabl selbst – war für mich und meine Studienkollegen eine Institution, die uns vieles gelehrt hat und den Grundstein für das weitere Studium legte. Legendär auch die Examen, während derer Prof. Schnabel (manchmal) an seiner Pfeife pfaffte und streng, aber wohlwollend fachlich auf Herz & Nieren prüfte. Prof. Schnabl war immer ein Unikat und war von den Studenten vielleicht gerade deswegen immer hoch geschätzt und respektiert. Vielen Dank für alles, was wir von Ihnen lernen durften!

15 Liste der Autoren

Die obigen Beiträge wurden beigesteuert von:

- Karl Sigmund, Universität Wien
- Wolfgang Woess, TU Graz
- Gerhard Murtinger, Siemens
- Harald Skarke, TU Wien
- Barbara Stöckl, KIWI-TV Filmprod.g.m.b.H.
- Peter Grabner, TU Graz
- Jörg Markowitsch, 3s Unternehmensberatung
- Emil Simeonov, FH Technikum Wien und minimath
- Annemarie Luger, Universität Stockholm
- Clemens Fuchs, Universität Salzburg
- Paul Surer, BOKU
- Wolfgang Trutschnig, Universität Salzburg
- Janice Goodenough, Hydrogrid

Buchbesprechungen

H. Derksen, J. Weyman: An Introduction to Quiver Representations. (Graduate Studies in Mathematics, Vol. 184.) American Mathematical Society, Providence (USA), 2017, 344 S. ISBN 978-1-4704-2556-2 H/b $ 83.

Die Autoren schreiben im Vorwort, dass sich ihr Buch an nicht-Spezialisten richtet, die eine Einführung in die Darstellungstheorie von Köchern (quiver) suchen und zwar *just starting from basic linear algebra*. In der Tat versuchen die Autoren (mit gutem Erfolg), die Theorie von Grund auf zu entwickeln. Neben einem soliden Grundverständnis von linearer Algebra sollte der Leser aber definitiv auch ein Lehrbuch über Kategorientheorie zur Hand haben sowie fundiertes Wissen in abstrakter Algebra mitbringen.

Bereits in der Einleitung wird die Wegealgebra (path algebra) eines Köchers eingeführt und mit der Kategorie der Köcherdarstellungen verknüpft sowie der Satz von Krull-Remak-Schmidt bewiesen. Die erste Hälfte des Buchs widmet sich den beiden fundamentalen Sätzen von Gabriel und Kac, welche die Grundlage für das große Interesse an der Darstellungstheorie von Köchern sind. In Kapitel 4 wird der Satz von Gabriel bewiesen, bevor in den nachfolgenden Kapiteln die Auslander-Reiten Theorie entwickelt wird. In Kapitel 8 folgt der Beweis des Satzes von Kac. An dieser Stelle sollte der Klappentext des Buchs als Warnung dienen, heißt es dort doch über den bisherigen Inhalt: *Once this basic material is established, the book goes on with (...)*. In der zweiten Hälfte entwickeln die Autoren in Kapitel 9 die geometrische Invarianten Theorie (geometric invariant theory), deren Methoden wichtig für die Untersuchung wilder Algebren sind und somit die Möglichkeiten der Auslander-Reiten Theorie entscheidend erweitern. Kapitel 10 präsentiert die Theorie der Semi-Invarianten von Köcherdarstellungen und insbesondere die Charakterisierung des Darstellungstyps mithilfe von Semi-Invarianten, an der die Autoren mit ihren eigenen wissenschaftlichen Arbeiten entscheidend mitgewirkt haben. Die beiden letzten Kapitel widmen sich neueren Entwicklungen und beschäftigen sich mit orthogonalen Kategorien, *exceptional sequences* und *cluster categories*.

Dieses herausfordernde Buch enthält eine Fülle von schönen und tiefen Resultaten, von denen einige zum ersten Mal in Buchform erscheinen – es wird der im Klappentext definierten Zielgruppe von graduate students und Experten in Darstellungstheorie, Invariantenthorie und algebraischer Geometrie sicherlich gute Dienste erweisen.

F. Pausinger (Queen’s University Belfast)

M. Grandis: Category Theory and Applications. A Textbook for Beginners. World Scientific Publishing Co., Singapur, 2018, 304 S. ISBN 978-981-3231-06-1 H/b £ 86.

This is very enjoyable introduction into category theory. It addresses students, is very well suited for beginners with some knowledge of algebra, lattice theory or topology, and can definitely be used as a textbook. It starts from examples taken from elementary mathematical theories and develops category theory with numerous examples and many exercises, which are either solved, partially solved or provided with hints.

The first three chapters cover categories and functors, limits, and adjunctions, with examples from free algebraic structures to Stone-Čech compactification and metric completion. The following three chapters treat applications in algebra, topology and algebraic topology, and homological algebra. The last chapter pertains to higher dimensional category theory.

W. Imrich (Leoben)

Nachrichten der Österreichischen Mathematischen Gesellschaft

Franz Kappel 1940–2020

Am 3. April 2020 ist Franz Kappel, em.o.Univ. Prof. der Karl-Franzens-Universität Graz, im 81. Lebensjahr verstorben. Franz Kappel beschäftigte sich mit gewöhnlichen Differentialgleichungen und mit deren Anwendungen in den Naturwissenschaften. Er war Dekan der Naturwissenschaftlichen Fakultät sowie Vizerektor für Finanzen an der KFU sowie seit 1968 Mitglied der ÖMG.

Persönliches

Prof. Adrian Constantin (Universität Wien) hat den Wittgenstein-Preis 2020 erhalten. Der Preis wurde aufgrund der bahnbrechenden Beiträge zur Mathematik der Wellenausbreitung vergeben, die Adrian Constantin in den vergangenen Jahren geleistet hat. Der jährlich vergebene Wittgenstein-Preis des Wissenschaftsfonds FWF richtet sich an exzellente Forscher/innen aller Fachdiziplinen. Er ist mit 1,5 Millionen Euro dotiert und wird von einer mit internationalen Expert/innen besetzten Jury vergeben. Die Redaktion der IMN gratuliert herzlich. Ein ausführlicher Bericht zum wissenschaftlichen Schaffen des Preisträgers ist für das nächste Heft der IMN geplant.

Neue Mitglieder

Hobel Benedikt, Mag. – Oswaldgasse 23-25/3/12, 1120 Wien. geb. 1991. Abschluss des Lehramtsstudiums in den Fächern Mathematik und Informatik im Jahr 2018. Derzeit Lehrtätigkeit im BHS-Bereich. email *benedikt.hobel@gmail.com*