Probability, Algorithms and Complexity[1]

Volker Strassen

Universität Konstanz
Residence: Dresden

Jo60, May 27-29, 2010

[1]This is a revised version of a talk I gave at the conference Jo60 celebrating Joachim von zur Gathen on the occasion of his 60th birthday. There is a large overlap with my talk given at the SODA 2009 in New York.

Dear Joachim, dear Dorothea and Rafaela, dear friends, colleagues and students,

We all like to talk about our new results. Unfortunately, I don't have any new results. So I am going to talk about some old things, also with the hope, that the students and the students of students of Joachim may learn something about the academic background of their teacher.

Although - the mother comes into my mind, who said to her 15 years old daughter: "It is time to talk about the sexual things", and got the answer: "Yes, of course, mum, what do you want to know?" In fact, most of what I have to say is contained in one of these two fundamental monographs:

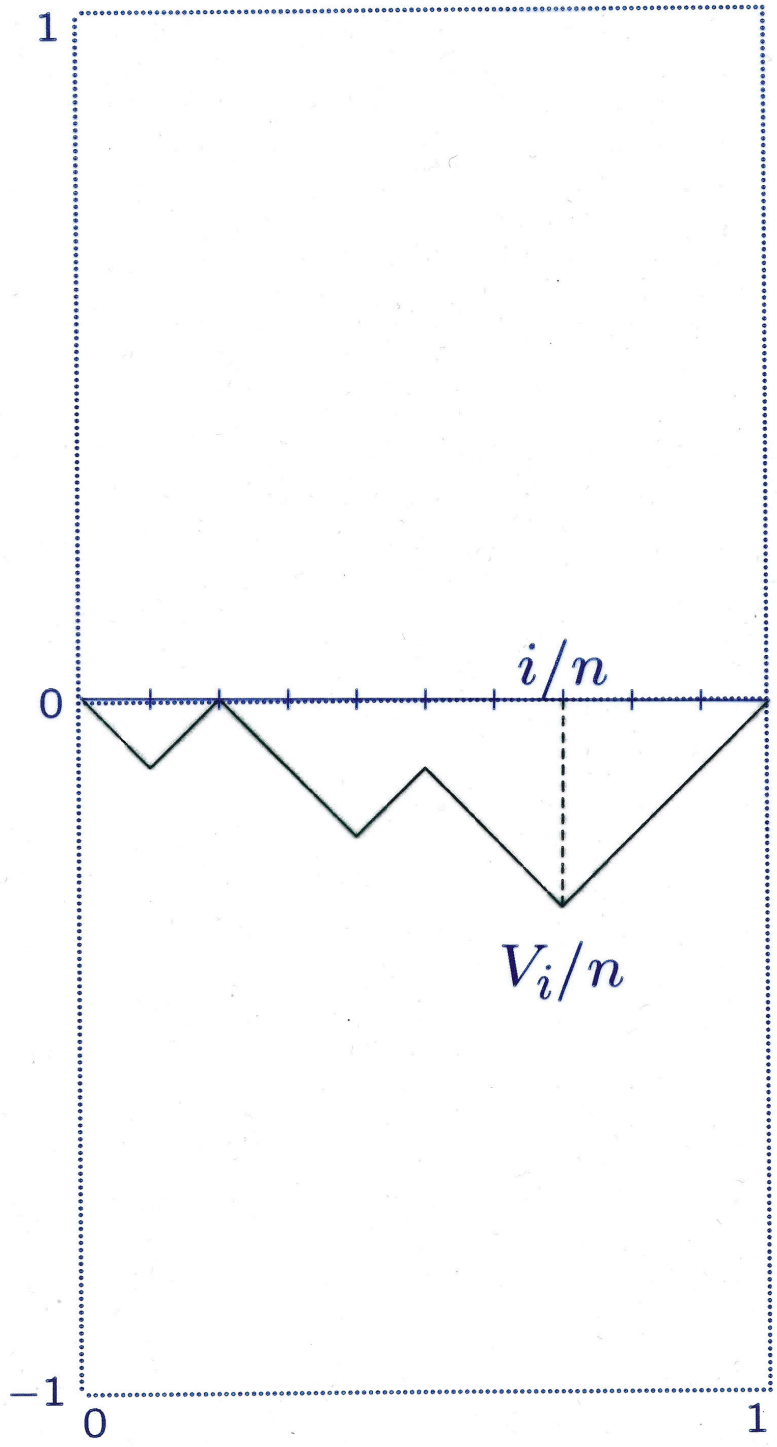
Joachim von zur Gathen, Jürgen Gerhard:
Modern Computer Algebra

Peter Bürgisser, Michael Clausen,
Amin Shokrollahi:
Algebraic Complexity Theory

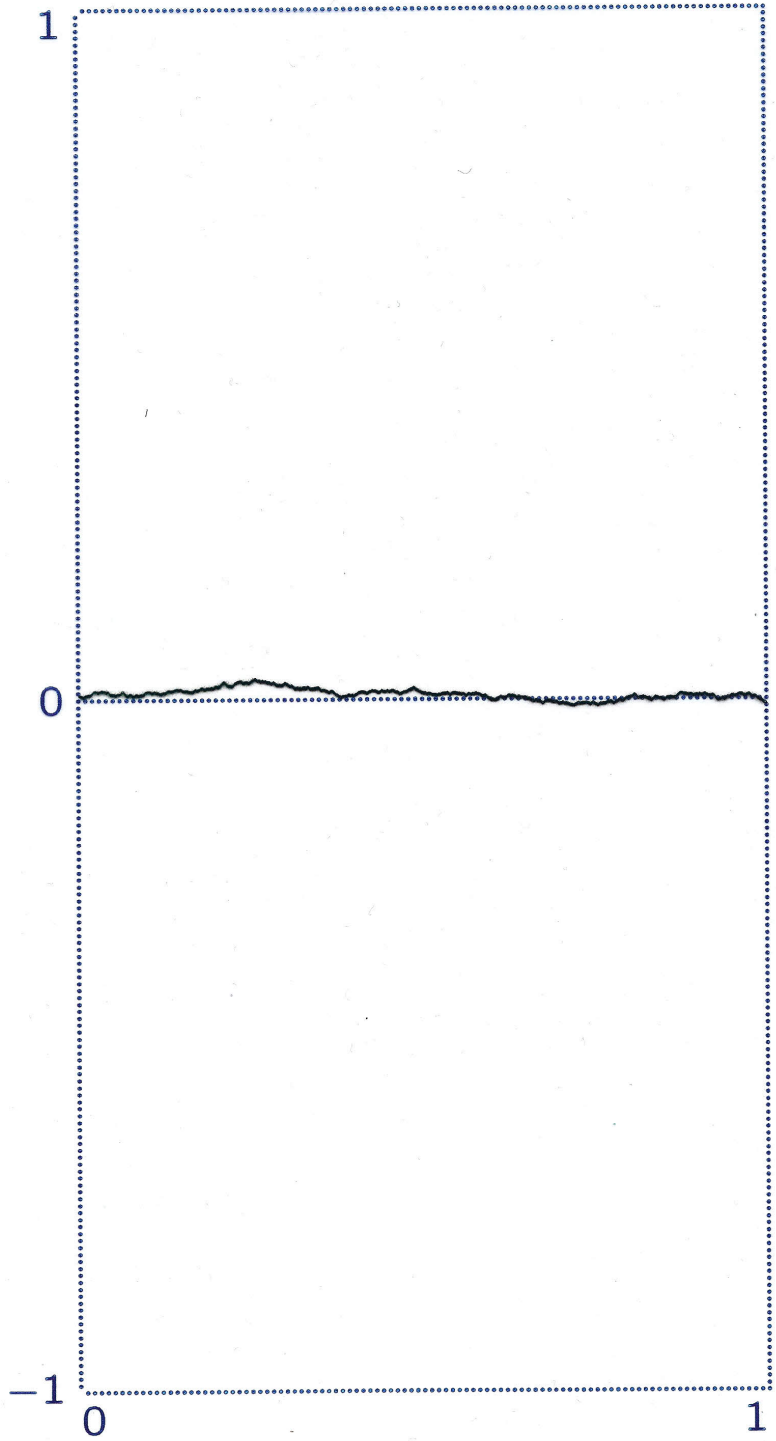
My first subject, probability, is an exception. Suppose you play with a friend for money. You toss a fair coin many, many times - let's pretend, infinitely often. Whenever head is up, you get 1 cent from your friend, when tail is up, he gets 1 cent from you.

Since you are friends, the money is not really paid out: You are content with seeing a chart on a screen that shows, at any moment of the game, the development of your account up to that moment. Your friend is also content, since his chart is just the negative of yours. When the chart after n tosses of the coin is scaled by a factor of $1/n$, a screen of width 1 and height 2 will suffice.

Here is such a chart for $n = 10$. V_i is the value of your account at time i.


1
0
−1
0
1
i/n
V_i/n

What happens if n gets larger and larger? Here you see the result of a random experiment for $n = 1000$.


1
0
−1
0
1

The problem is: You don't see much, and when n becomes much larger still, you will see nothing, just zero. This is explained by the strong law of large numbers, that is usually and legitimately attributed to Kolmogoroff in its natural generality, but that is due to Emile Borel in the case at hand.

Borel 1909:

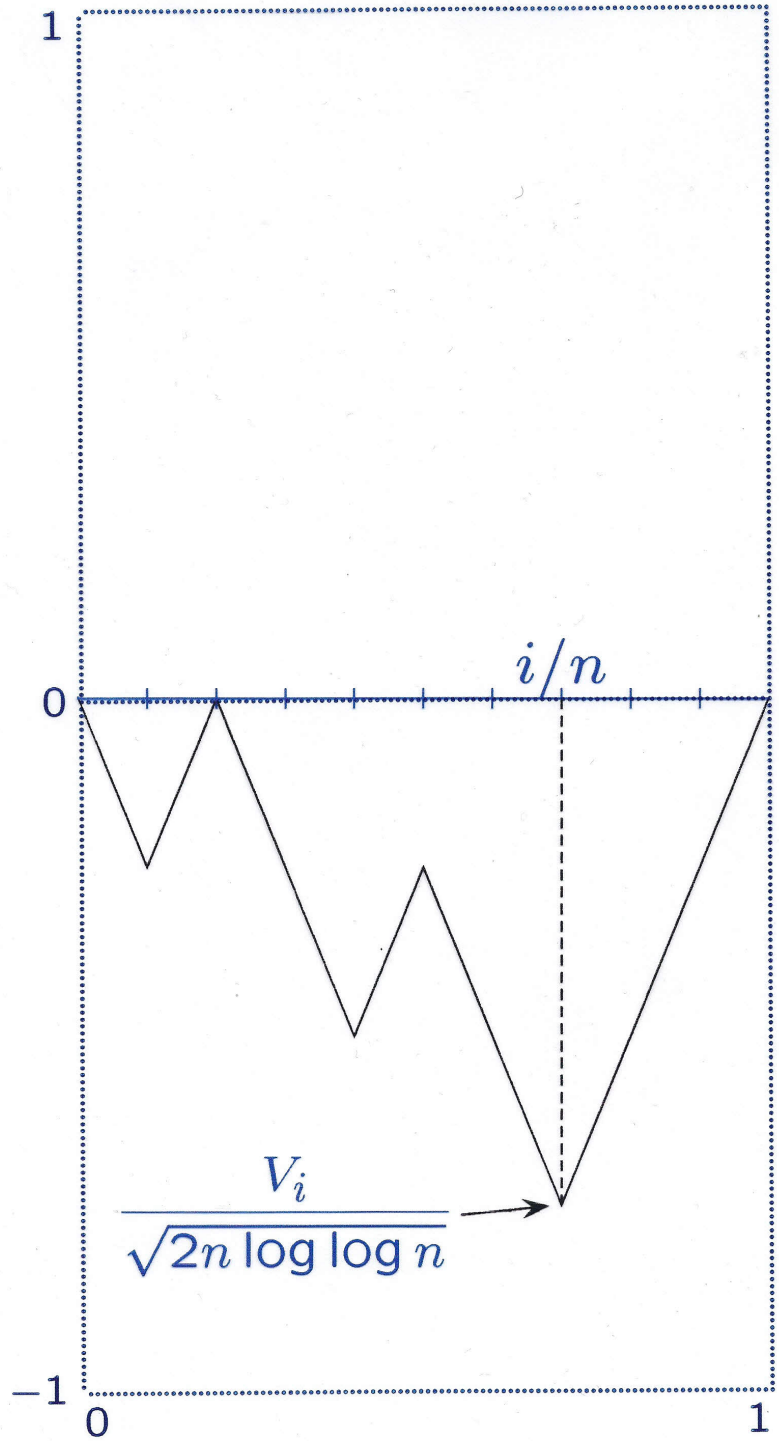
$$\lim V_n/n = 0$$

This and subsequent results hold with probability one.

So the scaling is too severe and the question arises as to the correct scaling. The answer is given by another classical result of probability theory, Khinchine's law of the iterated logarithm:

Khinchine 1924:

$$\limsup \; V_n / \sqrt{2n \log \log n} \; = \; 1$$


1
0
−1
0
1
i/n
$\frac{V_i}{\sqrt{2n \log \log n}}$

If you take the Khinchine scaling, the charts will fit eventually into the screen, at least if we enlarge the screen vertically by an arbitrarily small amount, and the charts will come close to the upper and lower edges of the screen infinitely often.

Now you will see something. But what? Which kind of curves are you going to see infinitely often under the Khinchine scaling? Let's call such curves *recurrent*.

The question leads into physics. Interpret a curve $x(t)$ as the one-dimensional movement of a particle of mass 1 from time 0 to time 1. Then:

Which curves are recurrent?

Those with mean kinetic energy $\leq \frac{1}{2}$:

$$\int_0^1 \frac{1}{2}\dot{x}(t)^2 dt \leq \frac{1}{2}$$

More precisely, given a curve $x(t)$ that represents the movement of a particle of mean kinetic energy at most $1/2$ (in particular, it is the indefinite integral of a square integrable function $\dot{x}(t)$), you will see that curve on your screen again and again, even if you look through a (fixed) microscope of arbitrary magnification.

On the other hand, if the curve is not of this special kind, there is a microscope under which eventually your charts will all look different from the curve.

The result has numerous applications, one of which I should like to tell you, since it might cheer you up in this time of financial crisis.

Call n a boom when

$$V_n / \sqrt{2n \log \log n} \geq \frac{1}{2}$$

The fraction of the past consisting of booms, how large can it become infinitely often?

Answer:

$$0.999993...$$

Thus, infinitely often you will find out that practically all your life so far has consisted of booms.

This is good news. The bad news is that the same is true for times of depression, as you can see by trading places with your friend.

The double logarithm appearing in Khinchine's law reminds me of another story, where this function had played a role for many years, until Martin Fürer has got rid of it: the multiplication of large integers.

Karatsuba 1962
Toom 1963, Cook 1966
Schönhage 1966
Schönhage, Strassen 1971
Fürer 2007:

$$d \cdot \log d \cdot 2^{O(\log^* d)}$$

This is the running time of Fürer's multiplication algorithm; d denotes the number of digits of the integers to be multiplied and $\log^*$ is the least number of times you have to apply the logarithm function in order to get a number < 1. The third factor is practically a constant.

What a beautiful result! In my view, the asymptotic complexity of integer multiplication is one of the challenges to our culture. If a bottle imp would offer me to answer 3 mathematical questions correctly, integer multiplication would be one of them.

Returning to the kinetic energy result, you may wonder what physics has to do with gambling. The answer is that the result is not as frivolous as it seems in its gambling context. It applies to a large class of stochastic processes, including Brownian Motion.

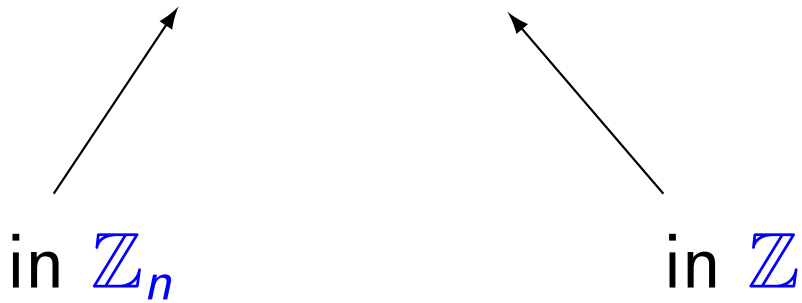
You may also wonder, where the inherent randomness of the coin tossing game has gone. This leads to simple and fundamental principles of probability theory, the so-called zero-one laws. Instead of discussing these, let me turn to a subject, in which probability doesn't miraculously disappear, but where it miraculously pops up in a landscape of eternal truths - in number theory.

Let me convince you that a simplified version of the first two probabilistic primality tests can be taught to high school students, which is what I have done a number of times many years ago. I have seen others doing similar things and I'm not claiming any originality.

You first introduce the ring $\mathbb{Z}_n$ of integers modulo n. You say that the rules they know from high school for manipulating integers are valid here too, which is what we mean by calling $\mathbb{Z}_n$ a (commutative) ring.

$$\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$$

$$ab := ab \bmod n$$



$\mathbb{Z}_7$ is a convenient example, since the week has seven days and you can multiply Thursday by Friday for practice.

At the University of Konstanz we used to have a student, who had made himself a T-shirt with this inscription:


3 · 5 = 1

When he would walk towards you, your heart would beat faster in view of this provocation. But as with the Doppler effect your heart would slow down immediately after he had passed by, since from behind you could see the wisdom of his shirt.


$3 \cdot 5 = 1$


in $\mathbb{Z}_7$

Next you ask, why people have invented rational numbers. You may get the answer: In order to torment high school children. You explain that rings which allow unrestricted division by nonzero elements are particularly comfortable to work with and that such comfortable rings are called fields and that the rational numbers form the smallest field containing the integers in a natural way.

So the question arises: Which of the rings $\mathbb{Z}_n$ are fields. The answer is given by Fermat's Theorem (as you know, but they don't):

FERMAT 1640:

$$n \text{ prime} \iff \mathbb{Z}_n \text{ field}$$

$$\iff a^{n-1} = 1$$

for all $a \neq 0$ in $\mathbb{Z}_n$

Prove this (counterclockwise) by Ivory's classical argument.

Use Fermat's Theorem to motivate your high school students for primality testing. Discuss computing in $\mathbb{Z}_n$, where n is a 1000-digit number. Tell them that they have to turn their school wisdom upside down: Multiplication is easy, at least for a modern computer, counting impossible.

Thus the standard method for primality testing won't work, and Fermat's Theorem won't work, although computing a^{n-1} for a single a is easy. (You can make this plausible by discussing the case when $n-1$ is a power of 2.)

Then you tell them that a slight refinement of Fermat's Theorem does work, at least in a certain sense.

We have

$$n \text{ prime} \implies a^{\frac{n-1}{2}} = \pm 1$$

$$\text{for all } a \neq 0$$

since

$$\left(a^{\frac{n-1}{2}} - 1\right) \cdot \left(a^{\frac{n-1}{2}} + 1\right) = a^{n-1} - 1 = 0$$

Assume

$$n = 4k + 3$$

Then

$$n \text{ composite} \implies a^{\frac{n-1}{2}} = \pm 1$$

for at most half

of the $a \neq 0$

Tell your high school students that the value of this statement doesn't lie in its proof, which is very simple by the standards of modern number theory, but in its application: A statistical test of primality. You just take a hundred random numbers a and check them.

After explaining this I usually had enough time to discuss Rabin's degenerate case of RSA, where you encode the plain text by just squaring it in $\mathbb{Z}_n$.

What I like most about probabilistic tests of primality is that an eternal truth - the primality of some huge number - is proved beyond any reasonable doubt by coin tossing. So I view such tests not so much as fast algorithms but as a new paradigm of proof. In fact, my dream is that one day a mathematical theorem more interesting than the primality of a particular number will be proved by coin tossing. (Here number theory would have to be replaced by proof theory.)

When I gave talks about primality testing many years ago, I could add that no classical, i.e. deterministic, proof method was known for the primality of huge numbers. This punch line has of course been spoiled by Agrawal, Kayal and Saxena in 2002.

Let me tell you a little bit of history: Inspired by a paper of Elwyn Berlekamp on "Factoring Polynomials over Large Finite Fields" I had found, in the Spring of 1974, what I called "A Fast Monte Carlo Test for Primality", and had submitted it to SIAM Journal on Computing.

In the fall of the same year I attended an Oberwolfach Conference on Computational Complexity, where the test caused a lot of discussion. I said that I believed that within 10 years a fast deterministic test would have been found, more precisely, a proof would have been published that the set of primes is polynomially decidable. My friend Ernst Specker disagreed. So we made a bet, whose terms were written down in the Oberwolfach conference book.

A few months after the conference Bob Solovay, who had heard about the result, found a sharper estimate for the error probability of my test with a much simpler proof and became a coauthor. From him I also learned about Gary Miller's fast deterministic test of primality under the extended Riemann Hypothesis. Gary and I had been working simultaneously, without knowing of each other's work.

Finally, in 1976, Michael Rabin combined Gary Miller's test with the idea of randomness and created what is now called the Miller-Rabin test, the fastest test known.

But what happened to our bet? Well, Agrawal, Kayal and Saxena were 18 years late to rescue me. I lost the bet and invited Ernst Specker and his family to a balloon ride along the Zürich lake. Because of an unexpected gust the landing turned out to be hazardous and the balloon was partially wrecked.

According to the terms of the bet I could have given Ernst Specker one ounce of gold instead of the balloon ride. In the end, Ernst got both. For as a conscientious Swiss he had bought one ounce of gold immediately after the Oberwolfach conference as an insurance for being able to pay his potential debt to me - and between 1974 and 1984 the price of gold more than doubled.

In the early seventies I also met Joachim as an auditor in some of my courses. In 1973 he obtained his master's degree from the ETH in Zurich and he soon became the star of his generation of doctoral students at the University. From the beginning he had this wonderful sense of humour. I have been exposed to him for many years, but unfortunately I don't remember a single of his jokes. Except scratches like "unless it is an immovable object". I know that we almost fell off our chairs with laughing, but I can't recall, why. It's almost like with jokes in dreams.

He has always been hilarious, and he has always been clever and independent. For example, during the summer semester, when the sun was shining, you wouldn't find him in his office, you would find him on the beach of Lake Zurich, sitting in the shadow of a large tree, learning Chow rings and Chern classes amidst all those young students, dressed in the topless fashion of the day.

He was supposed to work on secant spaces to curves (this is also the title of his later publication), a subject neither of us knew anything about initially. So we had decided to learn it together, which meant that I had to go to the beach, too.

In the first picture you see Joachim in the year 1980 of his doctoral promotion, on the second you see him together with Dorothea and Rafaela (in front), obviously after he had just made a good joke. Joachim and Dorothea haven't changed much, but Rafaela has.

Joachim and I wrote one joint paper. It is about the complexity of certain univariate polynomials and it stands on the shoulders of previous work. The first half of the cited papers is concerned with the complexity of almost all polynomials of degree N, which is of the order $\sqrt{N}$ if linear operations are free; the second half deals with lower complexity bounds for specific polynomials.

Ostrowski 1954
Motzkin 1955
Belaga 1961
Paterson, Stockmeyer 1973
Strassen 1974
Schnorr 1978
Heintz, Sieveking 1980
von zur Gathen, Strassen 1980:

For $s \in \mathbb{Q} \setminus \mathbb{Z}$

$$L\Big(\sum_{n=1}^{N} \frac{z^n}{n^s}\Big) \;\geq\; c\,\sqrt{N}/\log N$$

where c is a positive constant that may depend on s.

This is our best result, and I think it is a particularly beautiful application of the above theory. It says that these polynomials have almost maximal complexity, when s is rational, but not an integer, whereas at least for nonpositive integers s it is easy to see that they have the minimal complexity $O(\log N)$. Moreover, the polynomials are the initial segments of a power series that defines an important analytic function, the *polylogarithm*:

Polylogarithm:

$$Li_s(z) := \sum_{n=1}^{\infty} \frac{z^n}{n^s}$$

The power series converges on the open unit circle of the complex plane, but the domain of the polylogarithm can be extended by analytic continuation. The function turns up in such diverse fields as number theory (take $z = 1$ to obtain the Riemann zeta-function as a function of s), K-theory and statistical physics. In our library in Dresden there are at least 3 monographs on the polylogarithm.

The initial segments of the power series approximate the polylogarithm, say on a closed disk contained in the open unit disk. For rational nonintegral values of the parameter s they are hard to compute; that's what Joachim and I proved. But of course there are other good approximations to the polylogarithm. Perhaps some of them are easy to compute. The following conjecture about the approximate complexity of the polylogarithm rules this out:

Let $0 < r < 1$ and

$$L_N(Li_s) := \min\{L(f) : |f(z) - Li_s(z)| \leq 2^{-N} \text{ for } |z| \leq r\}$$

Conjecture: Let $s \in \mathbb{Q} \setminus \mathbb{Z}$. Then

$$\exists c > 0 \, \forall N \quad L_N(Li_s) \geq c\sqrt{N}/\log N$$

View this with a grain of salt. The conjecture wouldn't take offence, I think, if it should turn out that $\log N$ had to be replaced by some power of $\log N$, for example.

I have always been more intrigued by lower complexity bounds than by algorithms. This is in line with my enthusiasm, as a student, for Gödel's incompleteness theorem and for undecidability. In the sixties, when the new field of computational complexity was emerging, the fundamental results of Cook, Karp, Meyer-Stockmeyer, Fischer-Rabin, Valiant and others had not yet been published. Complexity in those early days meant low level complexity, or "practical complexity", as I prefer to call it, since it is concerned with problems that can be solved in practice.

Here I was inspired by Ostrowski's pioneering paper on Horner's rule and by Cook and Anderaa's brilliant $n \log n$ lower bound for the complexity of online multiplication of integers. Ostrowski's paper introduced the algebraic model (straightline programs), which is particularly adequate when your algorithms are supposed to run over large classes of ground fields, such as all fields or all finite fields of a given characteristic. The algebraic model is well suited for proving lower complexity bounds. Here is one:

Degree Bound:

Let $f_1, \ldots, f_P$ be polynomials in $x_1, \ldots, x_N$.
Then

$$L(f_1, \ldots, f_P) \geq \log_2 \deg(f_1, \ldots, f_P)$$

This means the following: $f_1, \ldots, f_P$ define a map from N-space to P-space over a suitable algebraically closed field, usually the complex numbers. Intersect its graph with an affine space of dimension P. If you get a finite number M of points (the typical situation), then $\log_2 M$ is a lower bound for the complexity of $f_1, \ldots, f_P$, even if linear operations are not counted. The degree bound also works for rational functions. It has numerous applications, for example on the sequence of elementary symmetric functions:

Computing the coefficients of a univariate polynomial of degree N from its roots,

$$(T - x_1) \cdot \ldots \cdot (T - x_N) = T^N + a_1 T^{N-1} + \ldots + a_N,$$

has complexity

$$L(a_1, \ldots, a_N) \sim N \log N$$

The degree bound can also be used to show that the Knuth-Schönhage algorithm for expanding a rational function into a continued fraction has the optimal order of magnitude, uniformly in the inputs. (Any algorithm for this problem has to contain branching instructions, so its running time will be a function of the inputs.)

Although it gives "nonlinear" lower complexity bounds for a number of problems, the degree bound has a shortcoming: For the evaluation of a single polynomial it only yields the trivial binary logarithm of the degree of that polynomial. To remedy this, the following inequality can be used:

Baur, Strassen 1983:

$$L\left(f, \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_N}\right) \leq 3L(f)$$

Thus, in order to obtain a lower bound for $L(f)$, you simply apply the degree bound to the lefthand side of the inequality. In this way you can show, for example, that an elementary symmetric function in the middle range has about the same complexity as the whole sequence.

The first order derivatives of the determinant function are the minors of the matrix. Thus, if you apply the inequality to the determinant and use Cramer's rule, you get a reduction of matrix inversion (and hence of matrix multiplication) to a single evaluation of the determinant. This is a central ingredient of showing that almost all nontrivial computational tasks of linear algebra have the same asymptotic exponent ω as matrix multiplication.

Exponent:

$$\omega := \inf\left\{\tau : L\begin{pmatrix}\text{multiplication of}\\ n \text{ by } n \text{ matrices}\end{pmatrix} = O(n^{\tau})\right\}$$

It is clear that ω lies between 2 and 3. No better lower bound than 2 is known, but the story of upper bounds for ω is worth telling. Since I don't have the time for a detailed discussion I shall illustrate this story by another one that is easier to grasp: The invention and technical development of bicycles.

Here it is. I have taken the story and some of the bike pictures from the English Wikipedia.

$\omega \leq 3.00$

We begin with a well known pedestrian: Carl Friedrich Gauss, who lived from 1777 to 1855.

$\omega \leq 3.00$

$\omega < 2.81$ rank

Here on the left is the first form of a bicycle, the Draisine, invented by Karl von Drais in 1817. It has got a handlebar, but you have to push off the ground in order to move.

There is a rumor that von Drais got into an argument with Gauss about the best way to move and that he wrote a paper with the title “Gaussian locomotion is not optimal”.

$\omega \leq 3.00$

$\omega < 2.81$ rank

$\omega < 2.78$ P, BCLR border rank

In this Scottish invention on the right the driver's feet don't touch the ground anymore; their movement is transmitted to the rear wheel by some mechanism, which, to be sure, is not yet fully developed.

This corresponds to the momentous notion of border rank, introduced by Bini, Capovani, Lotti, Romani, who didn't fully develop the tools for handling their concept either and were able to obtain just a slight improvement of the previous bound. This improvement is even smaller than shown on the graph, since Pan had obtained $\omega < 2.79$ by a different method somewhat earlier.

$\omega \leq 3.00$

$\omega < 2.81$ rank

$\omega < 2.78$ P, BCLR border rank

$\omega < 2.55$ SCH direct sum

A much more efficient realization of the transmission from the feet to the wheel is the *pedal*, invented by Michaux and Lallement in France.

On the matrix side Schönhage's analysis of the relation between border rank and direct sum, formulated in his τ-Theorem, produced a quantum jump in estimating ω.

By the way, the blue inequality signs in our graph are located approximately true to scale, when time goes from left to right and the height of the bounds from top to bottom, except that the classical bound $\omega \leq 3$ has to be shifted to the left.

$\omega \leq 3.00$

$\omega < 2.81$ rank

$\omega < 2.78$ P, BCLR border rank

$\omega < 2.55$ Sch direct sum

$\omega < 2.50$ CW

Once you have a bike with pedals at the front wheel, you can increase its speed by *expanding* the wheel.

This is quite analogous to what Coppersmith and Winograd did in their first joint paper: Whenever you have an algorithm of the kind considered before, you can speed it up somewhat. (Here I have omitted bounds by Pan and by Romani, that lie between those of Schönhage and of Coppersmith-Winograd.)

$\omega \leq 3.00$

$\omega < 2.81$ rank

$\omega < 2.78$ P, BCLR border rank

$\omega < 2.55$ Sch direct sum

$\omega < 2.50$ CW

$\omega < 2.48$ laser

This speeding up was superseded by the next progress, the *chain transmission* for bicycles, an English invention, and what I call the laser method for matrix multiplication.

There are similarities here too: In both constructions you gain speed by some kind of focussing, and both are parts of larger technologies: Gear transmission on the one hand and the asymptotic spectrum on the other.

$\omega \leq 3.00$

$\omega < 2.81$ rank

$\omega < 2.78$ P, BCLR border rank

$\omega < 2.55$ Sch direct sum

$\omega < 2.50$ CW

$\omega < 2.48$ laser

$\omega < 2.38$ CW diagonal

The next invention brings us close to the present state of the art: On the bicycle side we have the perfection of the chain transmission by *gearshift*, a French development, on the matrix side the perfection of the laser method by the Diagonal Theorem of Coppersmith-Winograd.

A similarity here is that both inventions are technically brilliant.

There is a more recent approach to matrix multiplication, started by Cohn and Umans and developed by them in collaboration with Kleinberg and Szegedy, which uses groups and their representations in a systematic way.

They draw level with Coppersmith-Winograd, and it is possible that their approach will yield even better bounds in the future. In any case it adds a new perspective to the old concepts.

$\omega \leq 3.00$

$\omega < 2.81$ rank

$\omega < 2.78$ P, BCLR border rank

$\omega < 2.55$ Sch direct sum

$\omega < 2.50$ CW

$\omega < 2.48$ laser

CUKS groups

$\omega < 2.38$ CW diagonal

Our bicycle counterpart of their construction is the *recumbent byke*, which, as I understand, competes with the classical byke for speed records. This very recent development doesn't fit into the screen, but if I bend the screen so that it becomes a tube, I can put the recumbent bike in the lower left corner.

You see: The overall construction of the recumbent byke is quite different from that of the traditional byke, but many of the details are similar. The same is true on the matrix side.

Perhaps you think that I should end my talk with a conjecture about ω. But this is dangerous. A few years ago I gave a talk in Prague on bilinear complexity and I stated a conjecture about ω. Then, in the evening, I had a vision, which I wrote down as a limerick in the visitor's book. Here it is:

An E.T. residing on Vega
Determined the size of ω.
In Praha, at night,
He told me that quite
Positively ω was nega....

If the E.T. tells the truth, my conjecture was utterly wrong. Perhaps ω isn't a universal constant. At any rate I don't want to make another conjecture. But I don't want to end this talk with a limerick about ω either. It should be a limerick about somebody else:

A jolly professor at b-it,
Brilliant scientist, teacher and wit,
Wrote his thesis out of reach
Studying curves on the beach;
The fresh air, said he, kept himself fit.

Thank you for your attention. I wish all of us a pleasant birthday conference for Joachim.