Asymptotic Spectrum and Matrix Multiplication

Volker Strassen

Affiliation: University of Konstanz
Residence: Dresden, Germany

ISSAC 2012, Grenoble

Ladies and gentlemen, my talk will be on the complexity of bilinear maps, the motivating example being matrix multiplication.

$$C\begin{pmatrix}\text{multiplication of}\\ m \times m \text{ matrices}\end{pmatrix}$$

Complexity

By its *complexity* we mean the minimal number of arithmetic operations sufficient to multiply two such (dense) matrices by an arithmetical circuit, or, if we allow computation trees, the minimal generic number.

I'll use the field of complex numbers as *ground field*, since its symbolic notation is conspicuous and well known. The main results, however, are valid over any infinite field or any infinite class of finite fields of a given characteristic. An exception is the discussion around group algebras, where we have to restrict the characteristic of the ground fields.

Our main concern is asymptotics. The complexity of the multiplication of large matrices is controlled by the so-called *exponent* ω of matrix multiplication:

$$C\begin{pmatrix}\text{multiplication of}\\ m\times m\text{ matrices}\end{pmatrix} = m^{\omega} + o(1)$$

Exponent

Complexity

We have the following trivial estimates of ω.

Exponent

$$C\begin{pmatrix}\text{multiplication of}\\ m \times m \text{ matrices}\end{pmatrix} = m^{\omega} + o(1)$$

Complexity

$$2 \leq \omega \leq 3$$

No better lower bound than 2 is known, but the story of upper bounds for ω is worth telling.

The exponent is an extremely important quantity: For example, it controls the asymptotic complexity of almost all significant computational tasks of linear algebra.

There is a standard strategy for reducing such a task to matrix multiplication: You rewrite some classical algorithm as a recursive procedure, where you recur to half size using a constant number of matrix multiplications. This turns out to be sufficient.

The crucial problem for the other direction is to reduce matrix multiplication to evaluating the determinant, i.e. m^2 functions to just one. It is solved by the following inequality:

Baur - S.:

$$C\left(f, \partial f/\partial x_1, \ldots, \partial f/\partial x_n\right) \leq 4C(f)$$

The inequality has originally been designed to enlarge the scope of the so-called *geometric degree bound*, which gives lower bounds of order $n \log n$ to the complexity of many classical algebraic problems of size n, but which yields only trivial results for the evaluation of a single rational function.

Now, in order to obtain a lower bound for $C(f)$, you simply apply the geometric degree bound to the left-hand side of the inequality, and that usually works.

What are the first order derivatives of the determinant? 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 to a single evaluation of the determinant. Reducing matrix multiplication to matrix inversion is a relatively easy matter.

After having discussed matrix multiplication, the principle character of our story, let me now introduce the whole cast. It consists of all bilinear maps between finite dimensional complex vector spaces.

$b: \ U \times V \longrightarrow W$ bilinear, concise

Conciseness is a harmless technical condition. It means that the maps travel in economy class, i.e. they don't have any unnecessary space. Any bilinear map has a unique concise version.

Interesting bilinear maps are as numerous as sand on the beach, just think of associative algebras or Lie-algebras or the structure maps of their modules. Here are a few examples, which will accompany us in the talk.

$$b: \; U \times V \longrightarrow W \text{ bilinear, concise}$$

Diagonal Map:

$$\langle n \rangle : \quad u, v \mapsto (u_1 v_1, \ldots, u_n v_n)$$

$$p, q \in \mathbb{C}^n$$

The diagonal map $\langle n \rangle$ is as good as n independent complex number multiplications.

$b: \; U \times V \longrightarrow W$ bilinear, concise

Polynomial Multiplication Modulo F:

$$\mathbb{C}[T]/F : \quad p, q \mapsto p \cdot q \bmod F$$

$$p, q \in \mathbb{C}[T]/F$$

$b: \ U \times V \longrightarrow W$ bilinear, concise

Group Algebra:

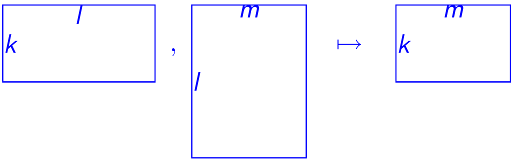
$\mathbb{C}[G]:$ Linearize $g, h \mapsto g \cdot h$

$g, h \in G$

$b: \; U \times V \longrightarrow W$ bilinear, concise

General Matrix Multiplication:

$\langle k, l, m\rangle$:



Let me draw your attention to the *Schönhage notation* of this map. In the following, when I will talk about matrix multiplication without further comment, I will mean the multiplication of square matrices.

Next, let us dicuss a few relations and operations among bilinear maps.

Isomorphism:

$$a \simeq b$$

An isomorphism is of course given by three linear isomorphisms compatible with the maps.

Isomorphism:

$$a \simeq b$$

Direct Sum:

$$a \oplus b$$

Computing the direct sum of a and b means computing a and b concurrently.

The importance of this operation is illustrated on the one hand by the Chinese Remainder Theorem, on the other by Wedderburn's Theorem, which identifies the class of direct sums of matrix multiplications with the class of semisimple algebras, up to isomorphism. In particular, by the Theorem of Maschke, any group algebra is isomorphic to a direct sum of matrix multiplications.

Isomorphism:

$$a \simeq b$$

Direct Sum:

$$a \oplus b$$

Tensor Product:

$$a \otimes b$$

The tensor product of a and b is a map with a block decomposition, such that each block is a scalar multiple of b, while the superstructure of scalars is modeled after a.

Let us look at an example. The fact that a multiplication of matrices of order $2m$ may be viewed as a 2 by 2 matrix multiplication with coefficients that are not numbers but matrices of order m, may be elegantly expressed in the language of tensorproducts:

$$\langle 2m, 2m, 2m \rangle \simeq \langle 2, 2, 2 \rangle \otimes \langle m, m, m \rangle$$

$$\langle 2^N, 2^N, 2^N \rangle \simeq \langle 2, 2, 2 \rangle^{\otimes N}$$

Disregarding gaps, large matrix multiplications may therefore be viewed as high tensor powers of 2 by 2 matrix multiplication. This fact has been decisive for the theory from its very beginning.

It also allows us to generalize the asymptotic point of view from matrix multiplication, which is given to us as a sequence, to arbitrary bilinear maps: Any such map is the first member of the sequence of its tensor powers.

Next we come to the basic computational concept of bilinear complexity.

Restriction:

$$a \leq b$$

$\exists\, \alpha, \beta, \gamma$ linear $\quad a(u,v) = \gamma b(\alpha u, \beta v)$

GASTINEL

Roughly speaking, this means that a is a restriction of b, when a can be computed by one call of b together with some linear work.

This concept is due to the French numerical analyst Noël Gastinel. His definition is equivalent to a more symmetric notion of restriction, that we use in research. His definition, however, is easier to motivate, and I shall use it here.

If "restriction" is to have a computational meaning, why don't we count the linear work? One reason is that the linear work depends on a choice of bases, and we want a notion that is independent of such a choice. The second, and most important, reason is that when we form high tensor powers of the maps for an a asymptotic study, the linear work magically becomes negligible.

Restriction also has pleasant algebraic properties:

Restriction:

$$a \leq b$$

$$\exists\, \alpha, \beta, \gamma \text{ linear} \quad a(u, v) = \gamma b(\alpha u, \beta v)$$

GASTINEL

$\leq$ is a preorder

compatible with $\oplus$ and $\otimes$

Rank:

$$R(b) := \min\{r : b \leq \langle r \rangle\}$$

While “restriction” is a relative notion, “rank” is an absolute one. The rank of a bilinear map b is the minimal number of complex number multiplications, embodied by the diagonal map $\langle r \rangle$, that suffices to compute b by restriction.

Rank inherits the virtues of restriction: It is an invariant of the map, it is subadditive with respect to direct sum and submultiplicative with respect to tensor product, and, asymptotically, it is equivalent to complexity.

In particular we may replace complexity by rank in the definition of the matrix exponent.

Rank:

$$R(b) := \min\{r : b \leq \langle r\rangle\}$$

$$R\big(\langle m, m, m\rangle\big) = m^{\omega + o(1)}$$

Suppose, we can prove that the rank of 2 by 2 matrix multiplication is less or equal than 7. Using the submultiplicativity and what we know about tensor powers of 2 by 2 matrix multiplication, we conclude

Rank:

$$R(b) := \min\{r : b \leq \langle r \rangle\}$$

$$R\big(\langle m, m, m \rangle\big) = m^{\omega + o(1)}$$

$$R\big(\langle 2, 2, 2 \rangle\big) \leq 7 \implies R\big(\langle 2^N, 2^N, 2^N \rangle\big) \leq 7^N$$

$$\implies 2^{\omega} \leq 7$$

Of course, there is nothing special about 2 and 7, thus by the same reasoning we obtain

$$R\big(\langle m, m, m\rangle\big) \leq r \Longrightarrow m^{\omega} \leq r$$

In this way a rank inequality for small maps leads to an inequality for the asymptotic quantity ω.

Here the proper story about upper bounds for ω begins, and it may be the right moment to show you in a little trailer how it has evolved. Since it is a bit abstract I'll illustrate it by another story: the invention and technical development of *bicycles*.

I have taken that 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.

$\omega \le 3.00$

$\omega < 2.81$ 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.

There is no new inequality yet for the group approach, so I can put the recumbent bike into the lower left corner.

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

Last year the world record of Coppersmith-Winograd has been broken by Williams (and also by Stothers, but I haven't seen his paper yet). Unfortunately I have no more bicycles, so I will use a different parable.

The wren is a little bird living in bushes and hedges. In German the female bird is called *Zaunkönigin*, which in literal translation means hedge queen.

Once upon a time the birds wanted to choose a king or queen and decided that whoever could fly higher than all the other birds should be their king or queen. As had been expected, the eagle won the contest.

Or, did he really? When the eagle, quite exhausted from surpassing the other birds, began to descend again, there was this little bird, who had hidden herself in the feathers of the eagle, and who started with fresh strength to fly a bit higher than the eagle had. “I'm the queen”, she said, and she was right. And that's why she is called Zaunkönigin in German.

WILLIAMS $\omega < 2.373$

CW

$\omega < 2.376$

Here you see the eagle with his two wings Coppersmith and Winograd, and the Zaunkönigin Vassilevska Williams, who is now the queen of our story.

Let us return to the technical part of the talk. Of the major inventions I'll skip border rank completely, although it is very important for obtaining initial constructions, and for balance I'll not discuss the laser method either.

The last technical slide has been this one:

$$R\big(\langle m,m,m\rangle\big) \le r \Longrightarrow m^{\omega} \le r$$

By the definition of rank, this is equivalent to

$$\langle m, m, m \rangle \leq \langle r \rangle \Longrightarrow m^{\omega} \leq r$$

$$\langle m, m, m \rangle \leq \langle r \rangle \Longrightarrow m^{\omega} \leq r$$

Schönhage's τ-Theorem:

$$\bigoplus_i \langle m_i, m_i, m_i \rangle \leq \langle r \rangle \Longrightarrow \sum_i m_i^{\omega} \leq r$$

Schönhage's τ-Theorem differs from this simple implication by taking into account direct sums, but unlike its predecessor the τ-Theorem is not trivial and Schönhage had to introduce a novel recursion technique to prove it.

Now let us simplify our notation and our thoughts and shift the focus from individual bilinear maps to their isomorphism classes.

The isomorphism classes of bilinear maps

form a commutative semiring:

$$a + b := a \oplus b$$

$$a \cdot b := a \otimes b$$

A commutative semiring is an algebraic structure modeled after the natural numbers $\mathbb{N}$, so that the rules of calculation you are used from there are valid here too.

In the following I will continue to speak of bilinear maps even though I will usually mean their isomorphism classes.

Most of the time we will be discussing interesting subsemirings of the semiring of all bilinear maps, such as the semiring of semisimple algebras, i.e. of all direct sums of matrix multiplications.

The smallest semiring of bilinear maps consists of the diagonal maps. It is isomorphic to the semiring $\mathbb{N}$ of natural numbers, the number n corresponding to the class of the diagonal map of length n.

The isomorphism classes of bilinear maps

form a commutative semiring:

$$a + b := a \oplus b$$

$$a \cdot b := a \otimes b$$

$\leq$ is a partial order

compatible with $+$ and $\cdot$

Restriction induces a partial order in the semiring of bilinear maps, compatible with addition and multiplication.

Since we are interested in asymptotics, we need an asymptotic analogue of restriction. Since we interpret asymptotics in terms of high tensor powers, we are led to this definition.

Asymptotic Restriction:

$$a \lesssim b \Longleftrightarrow a^N \leq b^{N+o(N)}$$

Thus, a is an asymptotic restriction of b, when high tensor powers of a are restrictions of slightly higher tensor powers of b. Clearly restriction implies asymptotic restriction.

Asymptotic restriction has many useful properties, the most important being the validity of the following central result.

Spectral Theorem:

For any semiring $\mathcal{S}$ of bilinear maps there are

a compact space Δ

a homomorphism $\Phi : \mathcal{S} \longrightarrow \mathcal{C}^+(\Delta)$

such that $\Phi(\mathcal{S})$ separates points and such that

$$a \lesssim b \Longleftrightarrow \Phi(a) \leq \Phi(b)$$

$\mathcal{C}^+(\Delta)$ denotes the semiring of nonnegative continuous real valued functions on the compact space Δ. Thus, the last line of the Spectral Theorem says that the asymptotic restriction relation of bilinear maps is faithfully reflected by the pointwise order relation of the corresponding spectral functions.

Given the semiring $\mathcal{S}$ of bilinear maps, the pair (Δ, Φ) is essentially unique and will be called the *Asymptotic Spectrum* of $\mathcal{S}$.

The proof reduces the Spectral Theorem to a classical representation theorem for a certain class of ordered rings, due essentially to Kadison, but the reduction is not trivial.

As Peter Bürgisser has shown, no such theorem holds when *asymptotic restriction* is replaced by *restriction*.

Given a semiring of bilinear maps, by the Spectral Theorem it suffices to compute its asymptotic spectrum (Δ, Φ) in order to understand asymptotic restriction. But what about asymptotic complexity, for example the exponent of matrix multiplication?

$$\begin{aligned} \max \Phi(b) &= \min \left\{ r : \Phi(b) \leq r \right\} \\ &= \min \left\{ r : b \lesssim r \right\} \\ &= \min \left\{ r : b^N \leq r^{N+o(N)} \right\} \end{aligned}$$

Here we have used the Spectral Theorem and the definition of asymptotic restriction. (The difficulty that this reasoning presumes that r is a natural number is easily removed by replacing $\Phi(b) \leq r$ by the equivalent $\Phi(b)^N \leq r^N$ and rounding the right hand side.)

Thus, the maximum of the spectral function of b tells us, how many complex number multiplications are needed to compute high tensor powers of b by restriction. This last qualification may even be omitted. Let us call $\max \Phi(b)$ the *price* of b.

$$\text{price of } b := \max \Phi(b)$$

$$\text{value of } b := \min \Phi(b)$$

As a counterpart to the price we have the *value* of b, which tells us, how many complex number multiplications may be computed from high tensor powers of b by restriction.

Value is always less or equal than price and, as in real life, the ratio may be arbitrarily small. (Look at $\langle m, 1, 1\rangle$.)

Exponent of a bilinear map:

$$\omega(b) := \log \max \Phi(b)$$

Taking the (binary) logarithm of the price, we obtain a quantity that has the same relation to the map b as the classical matrix exponent has to matrix multiplication of order two. Naturally, we call it the *exponent of* b.

In particular we have

Exponent of a bilinear map:

$$\omega(b) := \log \max \Phi(b)$$

$$\omega\big(\langle 2, 2, 2\rangle\big) = \omega$$

It is time to discuss some concrete asymptotic spectra.

Logarithmic Imbedding:

$$\Delta(b_1, \ldots, b_q) \subset \mathbb{R}^q$$

$$\Phi(b_i) = 2^{\tau_i}$$

As a straightforward consequence of the Spectral Theorem we have a natural imbedding of $\Delta(b_1, \ldots, b_q)$, the asymptotic spectrum of the semiring generated by $b_1, \ldots, b_q$, into real q-space, such that the spectral function of b_i is the exponential function 2^{τ_i}, where τ_i denotes the i-th coordinate function of $\mathbb{R}^q$ restricted to the asymptotic spectrum.

In particular, for 2 by 2 matrix multiplication we have this:

Logarithmic Imbedding:

$$\Delta(b_1, \ldots, b_q) \subset \mathbb{R}^q$$

$$\Phi(b_i) = 2^{\tau_i}$$

$$\Delta\big(\langle 2,2,2\rangle\big) \subset \mathbb{R}$$

$$\Phi\big(\langle 2,2,2\rangle\big) = 2^{\tau}$$

Since large matrix multiplications are wedged between high tensor powers of $\langle 2, 2, 2 \rangle$, it is easy to see that $\Delta\big(\langle 2, 2, 2 \rangle\big)$ actually serves as an asymptotic spectrum for all semisimple algebras in such a way that the spectral function of matrix multiplication of order m is the exponential function m^{τ}, restricted to the hitherto unknown $\Delta\big(\langle 2, 2, 2 \rangle\big)$.

Logarithmic Imbedding:

$$\Delta(b_1, \ldots, b_q) \subset \mathbb{R}^q$$

$$\Phi(b_i) = 2^{\tau_i}$$

Square Matrices:

$$\Delta_{\mathrm{m}} = \Delta\big(\langle 2, 2, 2\rangle\big) \subset \mathbb{R}$$

$$\Phi\big(\langle m, m, m\rangle\big) = m^{\tau}$$

Therefore we will call $\Delta\langle 2,2,2\rangle)$ the *matrix spectrum* and will denote it by Δ_{m}.

But how does it look like?

Logarithmic Imbedding:

$$\Delta(b_1, \ldots, b_q) \subset \mathbb{R}^q$$

$$\Phi(b_i) = 2^{\tau_i}$$

Square Matrices:

$$\Delta_{\mathsf{m}} = \Delta\big(\langle 2, 2, 2\rangle\big) \subset \mathbb{R}$$

$$\Phi\big(\langle m, m, m\rangle\big) = m^{\tau}$$

$$\omega = \log \max \Phi\big(\langle 2, 2, 2\rangle\big) = \max_{\Delta_{\mathsf{m}}} \tau = \max \Delta_{\mathsf{m}} \in \Delta_{\mathsf{m}}$$

This line of reasoning shows that ω is a point of the matrix spectrum. Already such a small amount of information yields a two-line-proof of the τ-Theorem.

$$\sum_i \langle m_i, m_i, m_i \rangle \lesssim r \implies \sum_i m_i^{\tau} \le r \text{ on } \Delta_m$$

$$\implies \sum_i m_i^{\omega} \le r$$

Schönhage's τ-Theorem:

$$\bigoplus_i \langle m_i, m_i, m_i \rangle \lesssim \langle r \rangle \implies \sum_i m_i^{\omega} \le r$$

Since $m_i^{\mathcal{T}}$ is the spectral function of $\langle m_i, m_i, m_i \rangle$, the first line is actually an equivalence by the Spectral Theorem.

We seem to have generalized the original τ-Theorem by weakening its hypothesis, but actually this is what Schönhage proves.

So far we have only used that ω is a point of the matrix spectrum. In fact, we know more, namely that it is the largest point. This immediately yields the converse of the τ-Theorem.

$$\sum_i \langle m_i, m_i, m_i \rangle \lesssim r \iff \sum_i m_i^\tau \leq r \text{ on } \Delta_m$$

$$\iff \sum_i m_i^\omega \leq r$$

Schönhage's τ-Theorem with converse:

$$\bigoplus_i \langle m_i, m_i, m_i \rangle \lesssim \langle r \rangle \iff \sum_i m_i^\omega \leq r$$

I can see no proof of this converse without using the Spectral Theorem.

In recent years group theoretic methods have lead to asymptotic restrictions, whose right hand sides aren't diagonals anymore, but group algebras. So they look like the left hand sides, although they are mor special.

The reasoning that leads to Schönhage's τ-Theorem generalizes immediately.

$$\bigoplus_i \langle m_i, m_i, m_i \rangle \lesssim \bigoplus_j \langle n_j, n_j, n_j \rangle$$

$$\Longrightarrow \quad \sum_j n_j^{\mathcal{T}} - \sum_i m_i^{\mathcal{T}} \geq 0 \quad \text{on} \quad \Delta_{\mathrm{m}}$$

$$\Longrightarrow \quad \sum_j n_j^{\omega} - \sum_i m_i^{\omega} \geq 0$$

But unlike $r - \sum_i m_i^{\mathcal{T}}$ the function $\sum_j n_j^{\mathcal{T}} - \sum_i m_i^{\mathcal{T}}$ is not necessarily decreasing, so we don't get a converse. The situation is even worse.

$$\bigoplus_i \langle m_i, m_i, m_i \rangle \lesssim \bigoplus_j \langle n_j, n_j, n_j \rangle$$

$$\Longrightarrow \quad \sum_j n_j^{\mathcal{T}} - \sum_i m_i^{\mathcal{T}} \geq 0 \quad \text{on} \quad \Delta_{\mathrm{m}}$$

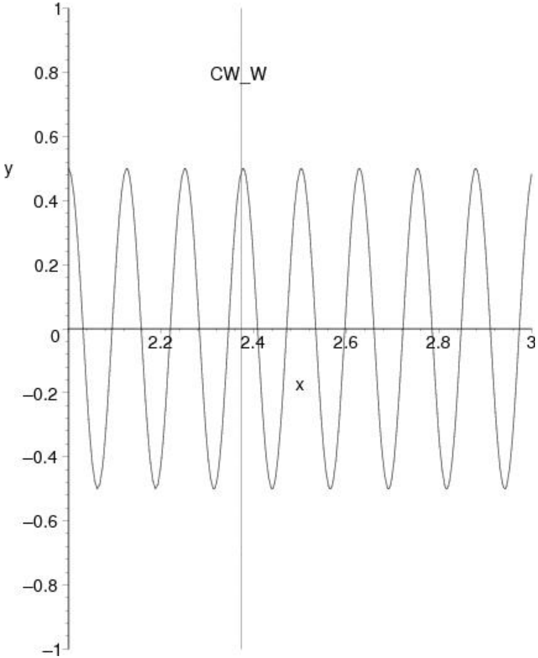
$$\Longrightarrow \quad \sum_j n_j^{\omega} - \sum_i m_i^{\omega} \geq 0$$

The set of all

$$\frac{1}{N}\Big(\sum_j n_j^{\mathcal{T}} - \sum_i m_i^{\mathcal{T}}\Big)$$

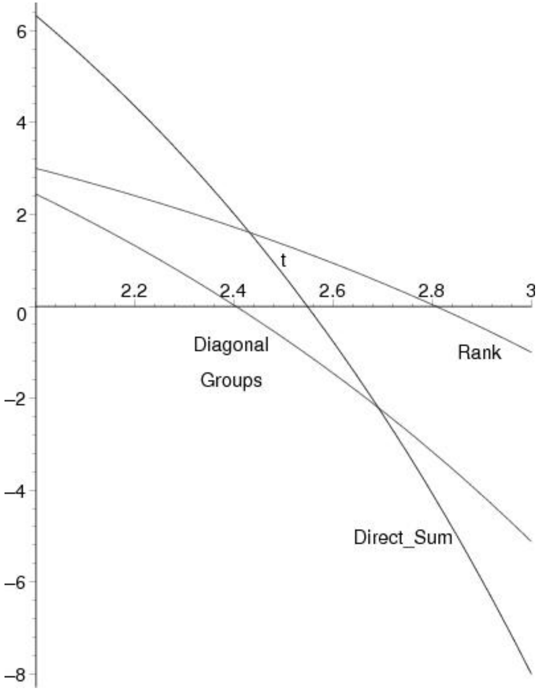
is dense in $\mathcal{C}(\Delta_{\mathrm{m}})$

Let's call a *spectral difference* any difference of two spectral functions, not necessarily one that comes from an asymptotic restriction. Then this says that you can approximate arbitrary continuous functions on the matrix spectrum by rescaled spectral differences. Actually this is true for any asymptotic spectrum as a consequence of the Stone-Weierstrass-Theorem.


CW_W
y
x
1
0.8
0.6
0.4
0.2
0
–0.2
–0.4
–0.6
–0.8
–1
2.2
2.4
2.6
2.8
3

So there are spectral differences that look like this after rescaling. For all we know such a spectral difference might come from an asymptotic restriction, which would mean that it is nonnegative on the matrix spectrum. This fact would give a lot of information on the matrix spectrum, but no upper bound for ω, since at the Williams bound the function happens to be nonnegative, so ω could coincide with this bound, and if you had no previous knowledge, ω could even be equal to 3.

Let us visualize in contrast the *traditional situation*, where the spectral differences are decreasing and the matrix spectrum, in particular ω, has to lie to the left of their zero.


6
4
2
0
–2
–4
–6
–8
2.2
2.4
2.6
2.8
3
t
Diagonal
Groups
Rank
Direct_Sum

In the general case we obviously need more information on the matrix spectrum. We cannot expect to know this spectrum completely, since we don't know its right endpoint ω. The following theorem makes the best of this situation.

Theorem:

$$\Delta_{\mathsf{m}} = [2, \omega]$$

First look at the endpoints: We know that m^τ is the spectral function of $\langle m, m, m \rangle$. So its price is m^ω, its value m^2. Unlike the cynic, who according to Oscar Wilde knows the price of everything and the value of nothing, we know the value of $\langle m, m, m \rangle$ exactly, while our knowledge of its price is handicapped by our ignorance about ω.

The hardest part of the proof is to show that the whole interval belongs to the matrix spectrum. But this fact has important consequences.

Theorem:

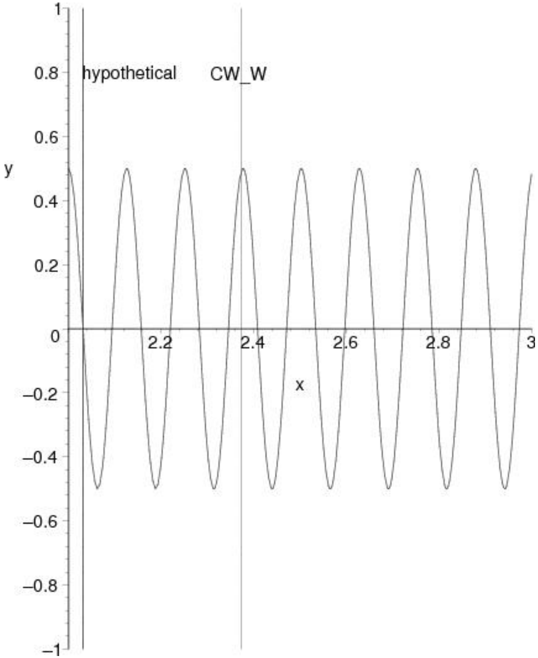
$$\Delta_{\mathrm{m}} = [2, \omega]$$

General τ-Theorem:

$$\bigoplus_i \langle m_i, m_i, m_i \rangle \lesssim \bigoplus_j \langle n_j, n_j, n_j \rangle$$

$$\iff \forall \tau \in [2, \omega] \quad \sum_j n_j^{\tau} - \sum_i m_i^{\tau} \geq 0$$

These two theorems are in fact equivalent, by the Spectral Theorem.



1
0.8
0.6
0.4
0.2
0
–0.2
–0.4
–0.6
–0.8
–1
y
hypothetical
CW_W
2.2
2.4
2.6
2.8
3
x

Looking again at our artificial example, we would suddenly get an extremely good bound on ω from the assumption that this spectral difference comes from an asymptotic restriction and is therefore nonnegative on the matrix spectrum $[2, \omega]$.

Now let us turn to the group theoretic methods in the light of the previous discussion.

Quotient of $S \subset G$:

$$Q(S) := \{st^{-1} : s, t \in S\}$$

Cohn, Umans:

$$\langle m, m, m \rangle \quad \leq_{\mathrm{cu}} \quad \mathbb{C}[G]$$

if and only if

there are $S_1, S_2, S_3 \subset G$ of size m

such that for any $q_\kappa \in Q(S_\kappa)$

$$q_1 q_2 q_3 = 1 \quad \Longrightarrow \quad q_1 = q_2 = q_3 = 1.$$

A few years ago Cohn and Umans obtained this lovely result and thus started the group theoretic approach to ω.

The condition is easy to understand when the subsets are in fact subgroups. Then it simply says that a product of three elements, one from each subgroup, is never equal to the unit element except in the trivial case.

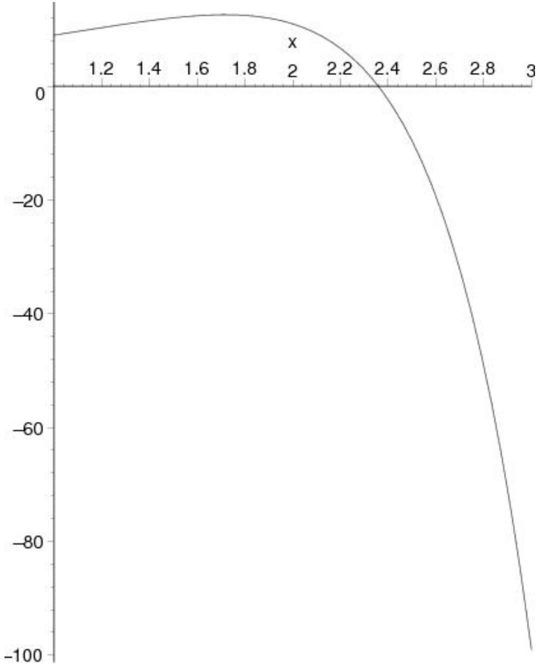
When this condition is satisfied then $\langle m, m, m \rangle$ is a restriction of the group algebra $\mathbb{C}[G]$ in a particularly *simple, combinatorial* way, which I indicate by the Cohn-Umans-subscript.

Unfortunately, the corresponding spectral differences are still traditional in the sense that they have at most one zero and are positive to the left and negative to the right of it, so that the immediate generalisation of Schönhage's τ-Theorem suffices.

COHN-UMANS spectral differences

$$\sum_j n_j^T - m^T$$

are traditional


x
1.2
1.4
1.6
1.8
2
2.2
2.4
2.6
2.8
3
0
–20
–40
–60
–80
–100

Of course this does not mean that the method is not suitable for proving, say, $\omega = 2$. It just means that the method does not make use of the additional degrees of freedom offered by a dense supply of spectral differences.

There is a second paper by Cohn and Umans in collaboration with Kleinberg and Szegedy, that allows you to locate *direct sums* of matrix multiplications in a group algebra.

Cohn, Kleinberg, Szegedy, Umans

characterize

$$\bigoplus_i \langle m_i, m_i, m_i \rangle \quad \leq_{\mathrm{cu}} \quad \mathbb{C}[G]$$

The above characterization is straightforward if you start from Cohn-Umans.

But the four authors succeed in reproving classical results including Coppersmith-Winograd using wreath products of groups in a very clever way.

Nevertheless, their *use* of the characterization is still traditional. Perhaps this is the reason why they do not surpass the classical bounds.

Actually, spectral differences of the form encountered here may be nontraditional.

CKSU spectral differences

$$\underbrace{\sum_j n_j^T}_{\text{group algebra}} \quad - \textstyle\sum_i m_i^T$$

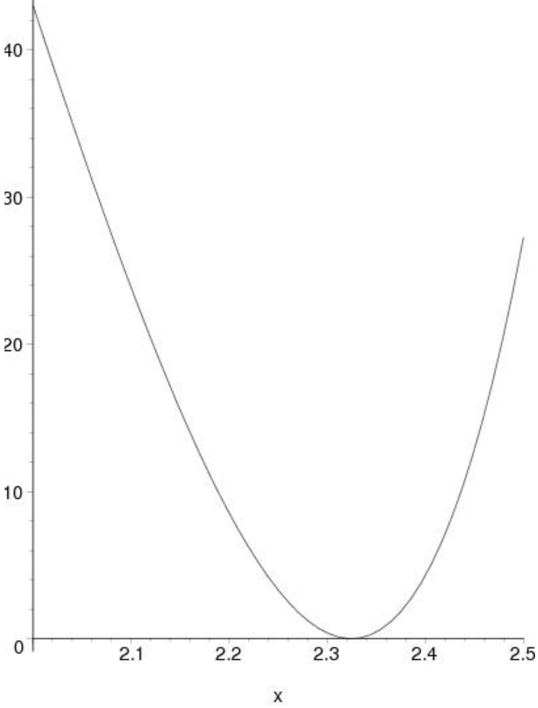
may be non-traditional!

Here is an example of such a spectral difference

$$\underbrace{\left(2^{4\mathcal{T}} + 2^8\right)}_{\mathbb{C}[2^{1+8}]} - \left(6 \cdot 2^{3\mathcal{T}} + 5 \cdot 2^{2\mathcal{T}} + 2^{\mathcal{T}} + 1\right)$$

extra special group

which looks like a *probe*:


40
30
20
10
0
2.1
2.2
2.3
2.4
2.5
x

Suppose, you could prove, for example by a combinatorial restriction of the Cohn-Umans type, that this function is nonnegative on the matrix spectrum. Since it is negative just on a tiny interval, the direct information on the matrix spectrum would be minuscule. Nevertheless, the general τ-Theorem would give you a new bound on ω.

$$\underbrace{\left(2^{4\mathcal{T}} + 2^{8}\right)}_{\mathbb{C}[2^{1+8}]} - \left(6 \cdot 2^{3\mathcal{T}} + 5 \cdot 2^{2\mathcal{T}} + 2^{\mathcal{T}} + 1\right) \geq 0$$

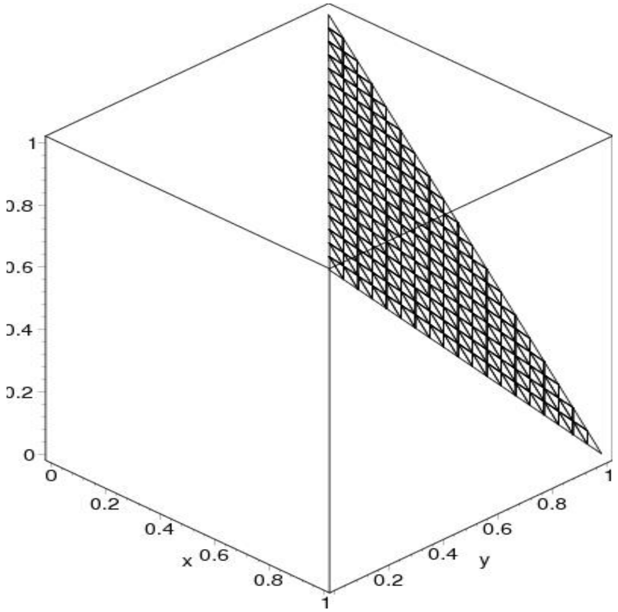
on $[2, \omega]$ would imply

$$\omega < 2.323$$

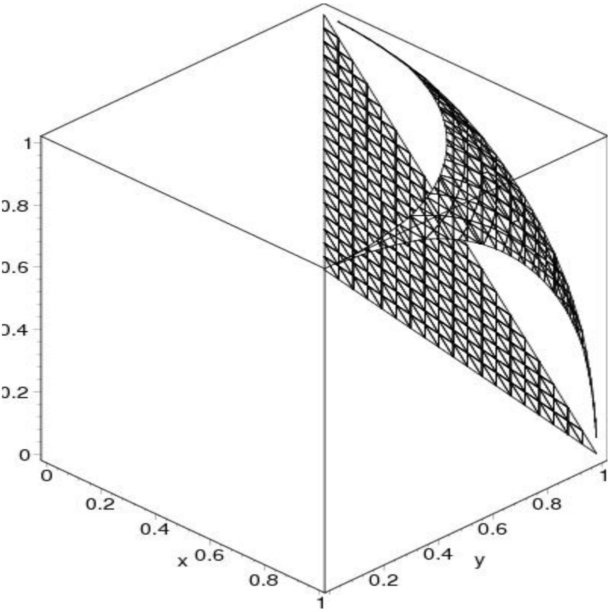
A rather modest improvement of the Williams bound, you may say. But it is easy to produce much sharper hypothetical bounds by similar probes using more complicated groups.

An advice for those of you who may want to investigate this: Don't get stuck with this particular example, which I made up for the talk. A computer search using the known character tables of groups may lead to many promising spectral differences. The rest is combinatorics.

Now let us take a quick look at the asymptotic spectrum Δ_{gm} of *general matrix multiplication*. It has a natural (logarithmic) imbedding into the unit cube of real 3-space, where it contains the full triangle displayed here:

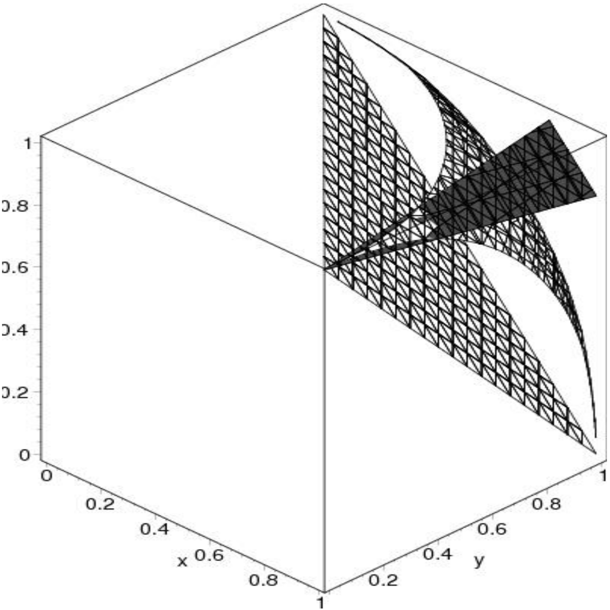


1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
x
0.6
0.8
1
0.2
0.4
y
0.6
0.8
1


1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
x
0.6
0.8
1
0.2
0.4
y
0.6
0.8
1

The first use of the laser method implies that the general matrix spectrum is wedged between the triangle and the umbrella (approximately).

Coppersmith-Winograd would give a better umbrella, but the calculations are more difficult.



1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
x
0.6
0.8
1
0.2
0.4
y
0.6
0.8
1

The affine linear bound shown here comes from Coppersmith's "Rapid Multiplication of Rectangular Matrices". Near the corners it is better than any other known bound.

Finally, we have the following

Theorem:

Δ_{gm} is star shaped with respect to the triangle

$$\mathrm{conv}\Big\{(0,1,1),(1,0,1),(1,1,0)\Big\}$$

This means that if the general matrix spectrum is the world you live in, you can see the whole triangle from everywhere.

At the end of the talk let us return to the discussion of more general asymptotic spectra $\Delta(b_1, \ldots, b_q)$.

In the same way as a linear map is represented by a matrix, when bases have been chosen for the vector spaces involved, a bilinear map is represented by a 3-dimensional array. A map is called *oblique*, when for a suitable choice of bases the support of the array (consisting of the places with nonzero entry) is an antichain with respect to the product order of the array.

The oblique maps form a semiring, containing all the examples of maps we consider in this talk.

Standard 2-simplex:

$$\Theta = \{\theta = (\theta_1, \theta_2, \theta_3) : \theta_i \geq 0, \sum_i \theta_i = 1\}$$

Logarithmmic Support Function of $b \neq 0$:

$$\sigma_b : \Theta \longrightarrow \mathbb{R}$$

$$\sigma_b(\theta) := \max \left\{ \sum_{i=1}^{3} \theta_i H(P_i) : P \text{ prob. on } \mathrm{supp}(b) \right\}$$

$$\begin{aligned} P_i &= i\text{-th marginal of } P \\ H &= \text{entropy function} \end{aligned}$$

The logarithmic support function of a bilinear map is a real valued function on the standard 2-simplex, that in general depends on the chosen bases. It can be proved that for oblique maps (and antichain supports) it does not.

Given oblique bilinear maps $b_1, \dots, b_q$, their logarithmic support functions combine to form a well defined map from the standard 2-simplex to real q-space:

Minimal Simplex of $b_1, \ldots, b_q$:

$$\sigma : \Theta \longrightarrow \mathbb{R}^q$$

$$\sigma(\theta) := \big(\sigma_{b_1}(\theta), \ldots, \sigma_{b_q}(\theta)\big)$$

Thus the image of the minimal simplex of oblique maps $b_1, \ldots, b_q$ is a compact subset of $\mathbb{R}^q$, as is the asymptoric spectrum of $b_1, \ldots, b_q$ under the logarithmic imbedding. How are these two sets related to each other?

Theorem:

Let $b_1, \ldots, b_q$ be oblique. Then

$$\sigma(\Theta) \subset \Delta(b_1, \ldots, b_q)$$

It may well be that we always have equality here. This would imply $\omega = 2$.

An indication in this direction follows from the *Diagonal Theorem* of Coppersmith-Winograd. When the bilinear maps $b_1, \ldots, b_q$ satisfy a somewhat sharper condition than just obliqueness (as do all the examples of bilinear maps we have seen in this talk), then $\sigma(\Theta)$ contains all minimal points of $\Delta(b_1, \ldots, b_q)$ with respect to the product order of $\mathbb{R}^q$.

These results have many applications. Here is one.

Theorem:

If $F = \prod_i (T - \epsilon_i)^{n_i}$ then

$$\Delta\left(\mathbb{C}[T]/F\right) = \log\left[\sum_i z(n_i), \sum_i n_i\right]$$

where

$$z(n) = \frac{\zeta^n - 1}{\zeta - 1} \cdot \zeta^{-2(n-1)/3}$$

$$\frac{1}{\zeta - 1} - \frac{n}{\zeta^n - 1} = \frac{n-1}{3}$$

Research Projects that may be feasable:

Prove $\omega < 3$ by a *nontraditional* restriction

Prove that $\Delta(b)$ is an *interval* for any oblique b

Prove that $\sigma(\Theta)$ contains all minimal points
of $\Delta(b_1, \ldots, b_q)$ for *arbitrary* oblique $b_1, \ldots, b_q$