Asymptotic spectra: Theory, applications and extensions

Avi Wigderson and Jeroen Zuiddam

October 24, 2023

Dedicated to Volker Strassen, who blazed for us beautiful trails to explore in the forests of computation

Abstract

In 1969, Strassen shocked the computational world with his subcubic algorithm for multiplying matrices. Attempting to understand the best possible algorithm for this problem, Strassen went on to develop his magnificent theory of *asymptotic spectra* in three papers between 1986–1991. Expressed in the great generality of partially ordered semirings, the centerpiece of this theory is a *duality* theorem between the asymptotic "rank" of elements, and a topological space which is called asymptotic spectrum. This duality theorem is a vast generalization of linear programming duality (in which we have a semigroup rather than a semiring), and indeed also of certain versions of the Positivstellensatz, the duality theorem of polynomial inequalities over the reals. Focusing on understanding the structure of the asymptotic spectrum of matrix multiplication, the theory has provided surprising *connectivity* and *convexity* theorems for it.

Strassen's theory has led to many subsequent results, especially new algorithmic, structural and barrier results on matrix multiplication, and more generally for the semiring of tensors (which includes the matrix multiplication tensors). Perhaps even more impressively, the generality of Strassen's theory has been applied recently to the study of a variety of very different settings and parameters, in diverse fields including communication theory, graph theory, probability theory, quantum information theory and computational complexity.

We feel that these developments call for an exposition of this growing field. This paper gives a comprehensive, self-contained, modern survey of Strassen's theory of asymptotic spectra and its various old and new application areas. For accessibility we provide many examples and high-level discussions of definitions and techniques.

The paper contains some new ingredients. We disentangle some proofs to make them more modular, and each part as general as possible. We introduce some new notions, which sometimes lead to simpler, more intuitive proofs, as well as to some stronger or more general theorems. One such consequence is our connectivity theorem for the asymptotic spectrum of *any* tensor network, greatly generalizing Strassen's connectivity theorem for the special case of matrix multiplication. Another consequence is progress on a conjecture of Strassen which generalizes Schönhage's tau theorem.

Contents

1. Introduction

In his seminal 1969 paper "Gaussian elimination is not optimal", Strassen sent a clear message to the (very young at the time) community of algorithm designers and complexity theorists, which has become part of its ethos since: *natural, obvious and centuries-old methods for solving natural, important computational problems may be far from the fastest.* For the problem of multiplying $n \times n$ matrices[1], he proved that the obvious $\mathcal{O}(n^3)$ algorithm known to Gauss (and in some form, already to the Egyptians and Chinese millennia earlier) is far from optimal, by designing a new one which takes only $\mathcal{O}(n^{\log_2 7}) \approx \mathcal{O}(n^{2.8})$ arithmetic operations.[2] The possibility of obtaining even faster algorithms for this central problem set Strassen and many other computer scientists on a quest to obtain them (with the current record below $\mathcal{O}(n^{2.4})$). The quest to understand the *matrix multiplication exponent*, namely the smallest[3] real number ω such that multiplying $n \times n$ matrices can be performed in n^ω arithmetic operations, is still raging on.

Then in the years 1986–1991 Strassen published three magnificent papers elaborating his theory of *asymptotic spectra*.[4] While chiefly motivated by trying to understand the complexity of matrix multiplication, Strassen's theory is far more general, putting the quest for proving upper and lower bounds on ω in a broader framework that suits other problems and settings. This theory sent another clear message to the same community, which too became part of its ethos: *studying the computational complexity of natural problems may both require and generate deep and sophisticated mathematics.* And indeed, today researchers from many different disciplines use and develop this theory.

This paper is devoted to a modern, self-contained exposition of Strassen's theory. In this introduction we begin by motivating it. Naturally, this introduction will be described in high-level, intuitive terms, which will be formalized in the technical sections.

1.1. Amortization and asymptotics

The following question arises in numerous parts of mathematics, physics, economics and computer science. Let $\mathcal{R}$ be a universe, and $f : \mathcal{R} \to \mathbb{N}$ be an $\mathbb{N}$-valued function on $\mathcal{R}$. In different contexts these objects may have a variety of meanings. In complexity theory, $\mathcal{R}$ may be a set of computational tasks, and for each task $r \in \mathcal{R}$ the number $f(r)$ may be the minimal cost (in some resource, e.g., time, space, communication, etc.) which is required to perform r. In economics, $\mathcal{R}$ may be a set of commodities or services, and $f(r)$ may denote the monetary value of r. In graph theory, $\mathcal{R}$ may be the set of undirected graphs, and $f(r)$ may be some graph parameter, for example the size of the largest clique or matching in r. And so on. To fix terminology, we will use "task" for elements of $\mathcal{R}$ and "cost" for the function f, but of course any of the many other interpretations are valid for what follows. In general, the cost $f(r)$ of a task r may be very complex to understand and compute.

[1] A problem whose computational complexity dominates all linear algebra calculations (e.g., solving systems of linear equations, computing the determinant, inversion of a matrix, etc.) and thus central to numerous other mathematical computations that use it. While this complexity may depend on the underlying field, almost all our discussions hold for every field.

[2] This algorithm works for any field and so do all the subsequent matrix multiplication algorithms and results that we discuss.

[3] More precisely, the infimum of such numbers.

[4] The theory was first presented at the conference FOCS in 1986 [Str86]. The work in [Str87, Str88, Str91] was accompanied by the PhD theses of Strassen's students Bürgisser [Bür90], Tobler [Tob91], and (later) Mauch [Mau98], and was followed up by Strassen's paper [Str05] which surveys the theory and presents new ideas.

Assume now that for every task $r \in \mathcal{R}$, and every integer n, also the task denoted r^n is in $\mathcal{R}$. Here r^n intuitively means "n copies of r", namely the task of performing r n times. Then the asymptotic behavior of $f(r^n)$ for any fixed r as n gets larger, and the limit[5] $\tilde{f}(r)$ of $(f(r^n))^{1/n}$ as n tends to infinity, captures some notion of "bulk" or "amortized" cost of performing r.

Remark 1.1. The choice of *multiplicative* notation for asymptotic behavior is standard and natural in Strassen's theory, as we shall see. Moreover, in many settings the "cost" tends to accumulate multiplicatively with additional tasks. Of course, in many other settings "cost" tends to add up with additional tasks, and it may be more natural to work *additively*. Namely, denoting nr for "n copies of r", and assuming $nr \in \mathcal{R}$ for all n, one can study the asymptotics of $f(nr)/n$ for large n. Needless to say, in the generality above one can switch between the two by replacing such "additive" cost function f by (say) 2^f to move to the multiplicative notation. Moreover, as in Strassen's theory the universe $\mathcal{R}$ will be a (semi)ring under addition and multiplication, both asymptotic notions can be (and are) studied simultaneously. Nonetheless, the focus is on the multiplicative notion, and we proceed with it.

The main object of study of Strassen's theory is this amortized cost $\tilde{f}(r)$. There are multiple reasons for studying it, some very general, and some depending on the setting and application area. The most obvious general reason is that in many computational and economic settings problems do indeed come in batches (e.g., computing the same function on multiple inputs, simultaneously communicating many messages through parallel channels, buying or selling many identical items, etc.). In others, multiple copies are generated artificially to achieve some "amplification", making hard problems harder (e.g., for cryptographic purposes or in optimization settings.) In many areas of mathematics and physics, "direct sum" and "tensor product" apply to different types of objects, and it is natural to study how various parameters of these objects are affected by sum and product.[6] Finally, in many cases the amortized cost $\tilde{f}$ is an analytically easier function to study or compute than the cost f, and thus, for example, $\tilde{f}(r)$ can be used to bound $f(r)$. Moreover, as we shall see in the next subsection on matrix multiplication, sometimes $\tilde{f}(r_0)$ for a single, fixed $r_0 \in \mathcal{R}$ can approximate $\tilde{f}(r)$ extremely well for infinitely many $r \in \mathcal{R}$!

A particularly basic question is understanding the settings in which amortization does not help, namely $\tilde{f}(r)$ is equal, or very close to $f(r)$ for every r. In such settings one often says that a "direct sum" or "direct-product" theorem holds. When direct sum does not hold, one often says that "economy of scale" is achieved. Such an understanding is essential for many of the applications above. We conclude this subsection with a list of different settings, and some (sample of) references to what is known about this basic question and amortization in general. This list is not meant to be exhaustive (nor is the list of references complete), but rather to impress the reader with the variety of areas in which it arises and is studied. We will expand on some of these and others in Section 2.3.

- **Communication:** In the setting of one-sided, zero-error communication, surprising economy of scale can be achieved. This led Shannon to the asymptotic notion of Shannon capacity of graphs, conjecturing a direct sum for this notion which was refuted by Alon [Alo98b]. For (interactive) communication complexity, such economy of scale was characterized for every problem via the information-theoretic notion of information complexity [BR14].

[5] Assuming a limit exists, which will be the case throughout the paper.

[6] Such investigations are often carried out not only for many copies of identical objects, but also for many copies of objects selected from a collection of few, different objects.

- **Optimization:** Direct sum and direct product theorems were proved for linear programming (under natural notions of sum and product), both for linear and convex objective functions [JM64, Mon65]. A direct product theorem for the SDP (semidefinite program) defining the Lovász theta function ϑ of a graph plays a central role in establishing the Shannon capacity of the pentagon [Lov79].

- **Amplification:** A large family of (approximate) direct product theorems (often named by the related concept of "XOR lemmas" or "parallel repetition") were proved for a variety of computational models, mainly for hardness amplification in cryptographic, pseudorandomness and optimization applications. These models include Boolean circuits (see the survey [GNW11]), multiparty communication and $\mathbb{F}_2$-polynomials [VW07], decision trees [NRS98], (classical and quantum) 2-prover proof systems [Raz98, Hol07] and many others.

- **Computation:** Perhaps the most famous economy-of-scale result is the following. Fix a field $\mathbb{F}$ and consider the number of operations (additions and multiplications of numbers in $\mathbb{F}$) required to compute the linear transformation Ax for a fixed $n \times n$ matrix A and an input n-vector x. It is known (and quite easy) to see that for every matrix this complexity is at most $\mathcal{O}(n^2/\log n)$, and that this is tight up to constant factors for most matrices A. Strikingly, for any matrix A, the total complexity of this problem for n input vectors $x_1, \ldots, x_n$ is below $n^{2.4}$, namely the amortized complexity per vector is only $n^{1.4}$! This is a simple consequence of the best-known algorithm for the matrix multiplication problem.[7]

The last example segues us to the next subsection on matrix multiplication, explaining how its arithmetic complexity begs an asymptotic study of the nature discussed above.

1.2. From matrix multiplication to preordered semirings and asymptotic rank

The following natural sequence of ideas interpolates between Strassen's algorithm for the very specific problem of matrix multiplication, to the general framework of preordered semirings, in which he develops the theory of asymptotic spectra.

Idea 0. Matrix multiplication is a bilinear function of its two input matrices, thus requiring additions and multiplications. In analyzing algorithms we will focus on counting multiplications only. As it happens (and is not hard to see), for this problem and the approach below, there is no loss of generality, as the number of additions will be at most a constant factor larger. Below we denote the problem of multiplying two $n \times n$ matrices by MM_n, and by $f(\mathrm{MM}_n)$ the multiplicative complexity of this problem.

Idea 1. Matrix multiplication is a "self-reducible" problem, so one can recurse, by viewing an $n \times n$ matrix as a 2×2 block-matrix, with each block being an $n/2 \times n/2$ matrix. Thus, any algorithm for MM_2 (in which we think of entries as belonging to a non-commutative ring) using k multiplications, allows reducing one instance of MM_n to k instances of $M_{n/2}$. This leads to a bound of $n^{\log_2 k}$ multiplications.[8] Clearly, 2×2 matrix multiplication can be performed using $k = 8$

[7] Simply, regard all input vectors as one matrix X, and compute AX. As A is fixed, only linear operations in X are performed.

[8] Up to constants, one can assume that n is a power of 2 for this argument without losing generality.

multiplications (namely, $f(\mathrm{MM}_2) \leq 8$). Strassen's seminal paper [Str69], which introduced the idea above, does it using $k = 7$ multiplications (proving $f(\mathrm{MM}_2) \leq 7$), leading to the first subcubic algorithm for MM_n, with $\mathcal{O}(n^{\log_2 7}) \approx \mathcal{O}(n^{2.8})$ multiplications.

Idea 2. For such recursive algorithms, it suffices to have an *asymptotic* bound on the complexity of the 2×2 matrix multiplication problem. Namely, assume we had an upper bound of $k^{m+o(m)}$ multiplications on computing MM_{2^m} for large m. Here and in the rest of the paper $o(m)$ denotes some function $g(m)$ for which $g(m)/m \to 0$. (Note that this may possibly be achievable with a much smaller value of k than 7.) Then, we could use recursion on size 2^m instead of 2, and again obtain $n^{\log_2 k + o(1)}$ complexity in general. This suggests studying

$$\lim_{m\to\infty} (f(\mathrm{MM}_{2^m}))^{1/m}.$$

We shall presently see that this expression is precisely $\tilde{f}(\mathrm{MM}_2)$ in the asymptotic notation of the previous subsection.

Idea 3. Matrix multiplication, like any bilinear map, may be naturally viewed as a (3-dimensional) *tensor* $T_n = T_{\mathrm{MM}_n}$ of size n^2 in each dimension (for entries of each of the two input matrices as well as the output matrix).[9] With this representation it is easy to see that the multiplicative complexity $f(\mathrm{MM}_n)$ equals $\mathrm{R}(T_n)$, with the usual notion of tensor rank.[10] Observe that taking Kronecker products of two such matrix multiplication tensors yields another matrix multiplication tensor; more precisely, for any s, r, $T_s \otimes T_r = T_{sr}$. With the idea above, we are led to study the *asymptotic* rank of *any* tensor T,

$$\widetilde{\mathrm{R}}(T) = \lim_{m\to\infty} \mathrm{R}(T^{\otimes m})^{1/m},$$

and see immediately that $\tilde{f}(\mathrm{MM}_n) = \widetilde{\mathrm{R}}(T_n)$ for every n. Indeed, we see that understanding $\widetilde{\mathrm{R}}(T_2)$, the asymptotic rank of a *single* tensor, determines the complexity of matrix multiplication in general (up to constant factors).

Idea 4. Tensors[11] form a *semiring* (under direct sum and Kronecker product). We have already seen the role of multiplication in this semiring. The importance of addition arises from the additive definition of tensor rank. Namely, one can express any upper bound $\mathrm{R}(T) \leq k$ (k positive integer) as an inequality $T \leq kI$ with I a unit (diagonal) tensor, and $\leq$ denotes a "reduction" that allows applying linear transformations[12] on each of the tensor "legs" (or fiber directions) and transform kI to T. Moreover, the relation $T \leq T'$ can be extended using this notion of reduction to any pair of tensors T, T'. To summarize, this semiring of tensors naturally comes with a *preorder*[13] that captures naturally the rank (and asymptotic rank) parameter, and is easily seen to be consistent with the addition and multiplication in this semiring.[14]

[9] Simply, $T((i,j),(j,k),(k,i)) = 1$ and all its other entries are 0.

[10] Namely, $\mathrm{R}(T)$ is the smallest number of rank-1 tensors which add up to T, where a tensor has rank 1 if it is the outer product of vectors.

[11] We think of d-dimensional tensors for fixed d, for example $d = 3$. This d is sometimes called the *order* of the tensor.

[12] Not necessarily invertible.

[13] Throughout we use preorder rather than the closely related notion of partial order, which is a preorder that is antisymmetric.

[14] Explicitly, if $S \leq T$ and $S' \leq T'$, then $S \oplus S' \leq T \oplus T'$ and $S \otimes S' \leq T \otimes T'$.

Idea 5. The above preordered semiring of tensors is just one of many (as we will see in Section 2.3) natural semirings $\mathcal{R}$ with preorder, in which the positive integers $\mathbb{N}_{\geq 1}$ are embedded in a similar natural way, as a totally ordered subsemiring. It is precisely in this vast generality of preordered semirings (with some added simple technical conditions) that Strassen develops his asymptotic theory. In each such preordered semiring one can define a notion of *rank* of an element $r \in \mathcal{R}$ in exactly the same way: $\mathrm{R}(r)$ is the smallest[15] positive integer $k \in \mathbb{N}_{\geq 1}$ such that $r \leq k$, using the given preorder $\leq$ on $\mathcal{R}$. Moreover, it is equally natural, and in some semirings more interesting[16] to consider the *subrank* $\mathrm{Q}(r)$ of elements $r \in \mathcal{R}$, defined as the largest integer $k \in \mathbb{N}_{\geq 1}$ such that $k \leq r$. For each of these we can define the analogous amortized, asymptotic notions, $\widetilde{\mathrm{R}}(r), \widetilde{\mathrm{Q}}(r)$, which can now be real-valued.

Idea 6. The final generalization is that there is no reason to consider asymptotic relations of semiring elements to integers only. Indeed, from the given preorder $\leq$, one can define a new, asymptotic preorder $\lesssim$ as follows: for any two elements $a, b \in \mathcal{R}$ let $a \lesssim b$, iff a has an "amortized reduction" to b, namely, if for large n we have $a^n \leq b^{n+o(n)}$.[17] In particular, this means that the amortized complexity of a is at most that of b. (It turns out that this asymptotic preorder essentially generalizes the notions of asymptotic rank (resp. asymptotic subrank), in which we essentially take b (resp. a) to be an integer.) Thus, the goal of the theory is to understand this asymptotic preorder $\lesssim$ for general preordered (and bounded) semirings.

1.3. The asymptotic spectrum and duality

In this very general, abstract set-up, how does one prove inequalities like $a \lesssim b$? Let us see that we at least understand the answer in a very concrete set-up, namely when the semiring consists of all continuous non-negative real functions on a compact space. Let $\mathcal{X}$ be a compact topological space (for concreteness, the reader can imagine simply a compact subset of some Euclidean space $\mathbb{R}^d$). Let $\mathcal{R}_{\mathcal{X}} = C(\mathcal{X}, \mathbb{R}_{\geq 1})$ be the set of continuous $\mathbb{R}_{\geq 1}$-valued functions on $\mathcal{X}$. Clearly, $\mathcal{R}_{\mathcal{X}}$ is a semiring under pointwise addition and multiplication, namely $(f+g)(x) = f(x)+g(x)$ and $(fg)(x) = f(x)g(x)$ for all $f, g \in \mathcal{R}_{\mathcal{X}}$ and $x \in \mathcal{X}$. Moreover, $\mathcal{R}_{\mathcal{X}}$ comes with a natural preorder $\leq$ of pointwise ordering, namely $f \leq g$ iff for *all* $x \in \mathcal{R}_{\mathcal{X}}$ we have $f(x) \leq g(x)$ (where here $\leq$ is the standard ordering of real numbers).

The family of preordered semirings $\mathcal{R}_{\mathcal{X}}$ is extremely friendly. First, compactness guarantees that it is bounded, as each function $f \in \mathcal{R}$ attains a maximum $\max(f)$ (and a minimum $\min(f)$) over $\mathcal{X}$. Second, the asymptotic preorder here is the same as the non-asymptotic one, namely defined by inequalities of real numbers, since for any x in $\mathcal{X}$, $f(x)^n \leq g(x)^{n+o(n)}$ if and only if $f(x) \leq g(x)$.

Now comes the main message of Strassen's theory: *every preordered semiring has such a simple form!* More precisely, Strassen's *duality theorem* states that for *every* preordered semiring $\mathcal{R}$ there exists a compact space $\mathcal{X}$ such that $\mathcal{R}$ is isomorphic to a subsemiring of $\mathcal{R}_{\mathcal{X}}$. Namely, there is a semiring homomorphism $\Phi : \mathcal{R} \to \mathcal{R}_{\mathcal{X}}$ which preserves the asymptotic preorder $\lesssim$. The pair $(\mathcal{X}, \Phi)$ is called the *asymptotic spectrum*[18] of $\mathcal{R}$. Strassen proves further that the asymptotic spectrum of

[15] In some rings it may be that no such finite bound exists; we will demand that our semirings are bounded (or *Archimedean*), namely that *rank* is defined for all elements in $\mathcal{R}$.

[16] As is the case in the semiring of graphs, where subrank corresponds to the Shannon capacity — an example we'll discuss in detail.

[17] The general definition will be slightly different, as we will see when we formally define this in Section 2.

[18] In commutative algebra, the spectrum of a ring R (the set of prime ideals) is a topological space X such that elements of R can be viewed as functions on X. In real algebra, there is an analogous object called spectrum (the set

every preordered semiring is essentially unique!

Let us denote $f_a = \Phi(a)$ the image of $a \in \mathcal{R}$ in $\mathcal{R}_{\mathcal{X}}$. By definition, $f_a \le f_b$ if for all $x \in \mathcal{X}$ we have $f_a(x) \le f_b(x)$.

The impact of this theorem is astounding. Let us start from the most direct consequence, which is also the most important. The theorem characterizes the asymptotic preorder on $\mathcal{R}$: for any two elements a, b, $a \lesssim b$ in $\mathcal{R}$ iff $f_a \le f_b$. This leads to a characterization of asymptotic rank by $\widetilde{\mathrm{R}}(a) = \max(f_a)$, converting a minimization problem into a maximization one. Similarly, the asymptotic subrank satisfies $\widetilde{\mathrm{Q}}(a) = \min(f_a)$, turning a maximization problem into a minimization one.[19] These are strong duality theorems in a very general context. Indeed, in the next subsection we will see how they vastly generalize linear programming duality, and indeed even some versions of the Positivstellensatz, the duality theorem for systems of polynomial inequalities in real algebraic geometry. We will also explain a less direct consequence of the duality theorem to Schönhage's tau theorem, a central tool for developing fast matrix multiplication algorithms. But before these, let us see how much simpler the duality theory becomes when we specialize it to *finitely generated* rings.

The fact that the mapping Φ is a semiring homomorphism means that it commutes with any polynomial. More precisely, if $p \in \mathbb{N}[Z_1, Z_2, \ldots, Z_d]$ is any multivariate polynomial with non-negative integer coefficients, then $\Phi(p(a_1, a_2, \ldots, a_d)) = p(\Phi(a_1), \Phi(a_2), \ldots, \Phi(a_d))$. Thus, if our semiring $\mathcal{R}$ is finitely generated, say by some d elements $a_1, a_2, \ldots, a_d$, then every element $b \in \mathcal{R}$ *is* such a polynomial, say p_b. Thus $\Phi(b)$ is uniquely determined (via p_b) by the (coordinate) functions $\Phi(a_1), \Phi(a_2), \ldots, \Phi(a_d)$. In other words, the asymptotic spectrum $\mathcal{X}$ may be identified with a compact set in $\mathbb{R}^d_{\ge 1}$, a much friendlier object than the abstract set in the general duality theorem. Moreover, as elements in such semirings are in 1-1 correspondence with polynomials, the asymptotic order $b \lesssim c$ between any pair of elements holds iff $p_c - p_b \ge 0$ holds on $\mathcal{X}$.

The study of finitely generated semirings is extremely natural and useful. Indeed, in many cases we care really about the asymptotic rank or subrank of *specific* elements in a ring (e.g., the Shannon capacity of the 7-cycle C_7, or the asymptotic complexity of the 2×2 matrix multiplication tensor). In such cases we can restrict ourselves to the subsemiring generated by that single element, whose asymptotic spectrum is simply a subset of $\mathbb{R}_{\ge 1}$, namely a union of closed intervals. Let us use this understanding.

Schönhage's tau theorem (a.k.a. asymptotic-sum inequality) One of the most important advances in the design of fast matrix multiplication algorithms was Schönhage's tau theorem [Sch81] (see also [Blä13]). To state it, recall that we denoted by MM_n the matrix multiplication problem (equivalently, tensor), and we capture its asymptotic complexity by ω, the smallest real number such that $\widetilde{\mathrm{R}}(\mathrm{MM}_n) \le n^\omega$. Thus trivially, an upper bound of the form $\widetilde{\mathrm{R}}(\mathrm{MM}_n) \le n^\tau$ (presumably via a new algorithm, as Strassen's first breakthrough gave $\omega \le 2.8$) immediately implies $\omega \le \tau$. Schönhage's tau theorem greatly generalizes this to upper bounds on the asymptotic rank of the direct sum of any number of matrix multiplication tensors (e.g., to algorithms which compute many matrix multiplication problems on disjoint variables). It states[20] for any $k, n_i \in \mathbb{N}_{\ge 1}$ and $\tau \in \mathbb{R}_{\ge 0}$ that $\widetilde{\mathrm{R}}(\sum_{i=1}^k \mathrm{MM}_{n_i}) \le \sum_{i=1}^k n_i^\tau$ also implies $\omega \le \tau$. Giving an algorithm proving such an inequality for a direct sum of two specific (rectangular) matrix multiplication tensors, Schönhage improved Strassen's initial bound to $\omega \le 2.55$. This was a great advance, which led to the development of

of "pre-primes") [Mar08], which is the origin of the name asymptotic spectrum.

[19] The precise conditions under which this holds are discussed in Section 3.5.

[20] He has a more general version for rectangular matrix multiplication that we discuss further in Section 7.31 — we state it here only for square matrices.

even better algorithms. Schönhage's proof is a very clever exercise in asymptotic analysis! Now let us see that the tau theorem (and more) directly follows from the duality theorem specialized to univariate rings.

Let MM be the semiring generated by MM_2. While this ring contains MM_n not for every n, but only powers of 2, this need not bother us as we are doing asymptotic analysis. Let us denote the asymptotic spectrum of MM by $\Delta \subseteq \mathbb{R}_{\geq 1}$. By Strassen's duality, as $\widetilde{\mathrm{R}}(\mathrm{MM}_2) = \max \Phi(\mathrm{MM}_2) = \max \Delta$ we see that[21] 2^ω is an element of Δ, indeed the largest one! Recall the assumption of the tau theorem was $\widetilde{\mathrm{R}}(\sum_{i=1}^k \mathrm{MM}_{n_i}) \leq \sum_{i=1}^k n_i^\tau$, that is, $\sum_{i=1}^k \mathrm{MM}_{n_i} \lesssim \sum_{i=1}^k n_i^\tau$. Assuming (wlog) that the sizes n_i appearing in the assumption of the tau theorem to be powers of 2, the LHS is a polynomial in MM_2. Applying Φ to it we get $\sum_{i=1}^k n_i^\omega \leq \sum_{i=1}^k n_i^\tau$, and can conclude $\omega \leq \tau$. This argument followed essentially just from the membership of 2^ω in Δ. Moreover, as 2^ω is in fact the largest element in the spectrum, we have also proved that Schönhage's theorem is in fact tight! It is not clear how one can prove this without the duality theorem. This argument implies a direct sum theorem for asymptotic rank, namely that

$$\widetilde{\mathrm{R}}\Big(\sum_i \mathrm{MM}_{n_i}\Big) = \Omega\Big(\sum_i \widetilde{\mathrm{R}}(\mathrm{MM}_{n_i})\Big).$$ [22][23]

It is easy to see now how to generalize the tau theorem to any semiring with an identical proof. Let $a \in \mathcal{R}$ denote any element whose asymptotic rank we are after. Then the argument above shows bounding the rank of *any* element generated by a immediately bounds $\widetilde{\mathrm{R}}(a)$. More precisely, for *any* polynomial (again, univariate with non-negative integer coefficients) p, the assumption $p(a) \lesssim p(r)$ for any real number r implies $\widetilde{\mathrm{R}}(a) \leq r$.

In Section 9.2 we will discuss even more general statements than Schönhage's tau theorem that rely on Strassen's duality.

1.4. Optimization: Linear programming duality and Positivstellensatz

Understanding systems of polynomial inequalities (and the special case of linear inequalities) is a central quest of real algebraic geometry and optimization. Let us see how these connect to Strassen's theory, in its special case of finitely generated semirings. We will do so informally, and send the interested readers to a formal and detailed exposition in [Fri21, Fri17, RZ22]. Strassen starts with a preordered semiring, and finds its dual, the asymptotic spectrum, which in the finitely generated case turns out to be a set cut out by polynomial inequalities. We will reverse the order now — start from such a semi-algebraic set, and generate a preordered semiring from it.

Consider an arbitrary semi-algebraic set, say $\mathcal{X}$ in $\mathbb{R}^d$, cut out by a system of k integer[24] polynomial inequalities in d variables $\{p_i \leq 0\}$. Moving negative coefficients in each p_i to the RHS, we obtain an equivalent system $\{p_i' \leq p_i''\}$ where $p_i', p_i'' \in \mathbb{N}[Z_1, Z_2, \ldots, Z_d]$ have non-negative integer coefficients. We can define an abstract preordered semiring $\mathcal{R}_\mathcal{X}$ as follows. It is generated by d

[21] The appearance of the exponential function is natural, as we ω is the *exponent* of matrix multiplication. This observation is more general, and indeed in any setting it is natural to consider $\log \Phi(b)$ as the exponent governing the rate of growth in $\widetilde{\mathrm{R}}(b)$.

[22] In fact, this equality is true even without the Ω.

[23] We do not know if a similar result holds for rank. One result in this direction is that direct-sum holds for MM_2, that is, $\mathrm{R}(\sum_i \mathrm{MM}_2) = \sum_i \mathrm{R}(\mathrm{MM}_2)$ [Bür90, Rem 17.13 (6) and Eq. 14.8]. It is not known whether this direct-sum statement is true for larger square matrix multiplication tensors MM_n, $n > 2$. Generally, Shitov proved that there are tensors T_1, T_2 such that $\mathrm{R}(T_1 \oplus T_2) < \mathrm{R}(T_1) + \mathrm{R}(T_2)$ [Shi19].

[24] We could allow rational coefficients, and then clear denominators to get integer ones.

abstract elements $Z_1, Z_2, \ldots, Z_d$, and its elements are *all* non-negative integer polynomials in these generators. The preorder is generated by the given system of inequalities. More explicitly, for any non-negative integer polynomial in k variables $q \in \mathbb{N}[X_1, X_2, \ldots, X_k]$, we have the inequality $q(p'_1, p'_2, \ldots, p'_k) \leq q(p''_1, p''_2, \ldots, p''_k)$ between these two elements of our semiring. Finally, define the asymptotic preorder $\lesssim$ extending $\leq$. Note that the constant polynomials provide the embedding of the integers in $\mathcal{R}_{\mathcal{X}}$. In this case it turns out that $\mathcal{X}$ is the asymptotic spectrum of $\mathcal{R}_{\mathcal{X}}$. So, by Strassen's duality, the maximum value of any non-negative integer polynomial r over $\mathcal{X}$ is, say, α precisely if the asymptotic rank of that polynomial in the semiring $\mathcal{R}_{\mathcal{X}}$ is α.

Here the reader might wonder what is going on. By the celebrated Positivstellensatz of real algebraic geometry (there are several versions [Kri64, Ste74, Put93, Sch91] which we will not distinguish for this level of exposition) the duality should have a completely different form. Positivstellensatz essentially states that the maximum value of any polynomial r is α on $\mathcal{X}$ precisely if the polynomial $r - \alpha$ is in the cone generated by the system $\{p_i \geq 0\}$ with sum-of-square coefficients. This looks extremely different from Strassen's characterization. While one can try to find the differences in the exact technical assumptions in each duality theorem[25] the difference is too striking and demands explanation. This is done in some detail in the paper of Fritz [Fri21]. But at a high level, these seemingly very different duality theorems are practically equivalent. Let us say a few words on each of the two directions.

In one direction, Positivstellensatz implies Strassen's duality for rings $\mathcal{R}_{\mathcal{X}}$ as above (namely, when both the elements and the preorder are finitely generated). We remark that this is non-trivial, and crucially requires the transition from the preorder $\leq$ to the asymptotic preorder $\lesssim$. The key point is that in *every* preordered semiring, asymptotic inequalities of the form $2ab \lesssim a^2 + b^2$ are true for every a, b in the semiring.[26] In our case this leads to the equivalent non-negativity of any square of any[27] integer polynomial (via $0 \lesssim (p - q)^2$), and this can be used within Strassen's framework to generate the coefficients in the sum-of-squares combination of the given inequalities defining $\mathcal{X}$.

In the other direction Strassen's duality seems far more general, as it pertains to arbitrary semirings where elements and preorder may not be finitely generated, and thus having asymptotic spectra $\mathcal{X}$ which may be infinite-dimensional and may require an infinite number of polynomial inequalities to define. Here the connection is that Strassen's original proof actually uses a vast generalization of Positivstellensatz of Kadison and Dubois (more precisely, the version of Becker and Schwartz [BS83]) to such settings. The proof of Strassen's duality theorem we will give in this paper (based on Zuiddam [Zui18]) will be however completely self-contained, without appeal to these general results.

Finally, we note that if we demand that polynomials p_i in the definition of $\mathcal{X}$ above are actually *affine*, everything becomes much simpler, and at the same time very interesting. Here full details are provided in [RZ22], and we only give the gist of the argument. Carrying out the same construction as for polynomials, one obtains a finitely generated *semigroup* (under addition) instead of a semiring, again with a finitely generated preorder via non-negative affine functions q. It is thus natural to define amortization, and thus the asymptotic rank additively. In this linear setting Strassen's duality theorem becomes *equivalent* to the duality theorem of linear programming! Here a key point in using linear programming duality to prove Strassen's duality is showing that the former implies a certain "catalytic" inequality, in the defined preorder $\leq$. One then uses the asymptotic preorder

[25] These are important, and are related to different "boundedness" conditions in both Positivstellensatz and in Strassen's framework.

[26] Proving this is a really good exercise for understanding the power of asymptotics.

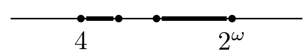
[27] Namely, allowing negative coefficients.

$\lesssim$ to remove the catalyst and show that the asymptotic rank of any non-negative integral affine function r over $\mathcal{X}$ is certified by the asymptotic rank of r in this semigroup.

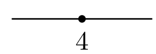
1.5. Connectivity and convexity of the asymptotic spectrum

Let us return to Strassen's duality theorem. As we have seen, if the spectrum of a semiring of interest is known, everything becomes simple, and we can in principle compute asymptotic rank and subrank of any element. But even in extremely simple semirings generated by a single element, where the asymptotic spectrum is just a union of intervals in the real line, (so the maximum and minimum points respectively determine the asymptotic rank and asymptotic subrank), it is hard to find these points, or sometimes indeed any point in the spectrum.

Strassen's next insight is that even if we cannot determine the asymptotic spectrum, proving additional structural results can lead to further understanding of the basic semiring parameters it encodes. Natural questions to explore beyond compactness can be: Is it connected? Is it convex? Strassen's major result answers all of these in the affirmative for matrix multiplication, namely the spectrum Δ of the semiring generated by MM_2 which we discussed above. We have already observed that $\Delta \subseteq \mathbb{R}_{\geq 1}$, and compactness guarantees that it is a *union of closed intervals.* We know that its largest point is 2^ω, where ω is the (yet unknown) matrix multiplication exponent. It turns out that the asymptotic subrank of MM_2 is known to be 4, and so this is the smallest point of Δ. For example, it is a priori possible that Δ looks like a union of two disjoint closed intervals:



Of course, if $\omega = 2$, then the minimum and maximum points in Δ coincide and so Δ consists of *one single point*:



Thus, if we were able to prove by any argument whatsoever that Δ is *not* connected, we would prove the lower bound $\omega > 2$. Strassen's result rules out this line of attack on a lower bound, by proving the theorem that Δ is connected. Thus, it must be one closed interval:

In other words, the theorem says that the asymptotic spectrum of matrix multiplication is given by the closed interval $\Delta = [4, 2^\omega]$.

Let us comment first that connectivity of the spectrum is a special and possibly rare situation; It is not hard to see that even for univariate semirings, the spectrum can be an *arbitrary* compact subset of the non-negative reals. Indeed as we have discussed before, one can pick any compact set $\mathcal{X} \subseteq \mathbb{R}$, and take the semiring of all non-negative integer polynomials under pointwise addition, multiplication and ordering.

Thus Strassen's proof of connectivity of the matrix multiplication spectrum must make use of very special properties of this specific problem. Indeed, the proof is ingenious, and we will soon return to discuss it. But before, let us demonstrate its power by showing how it significantly extends Schönhage's tau theorem.

Schönhage's tau theorem revisited and generalized Here is an equivalent phrasing of the original tau theorem: If $\sum_i \mathrm{MM}_{n_i} \lesssim r$ then $\sum_i n_i^\tau \le r$ for every $\tau \in \Delta$. Now since the real function $r - \sum_i n_i^\tau$ is decreasing in τ, we can solve the equation $r - \sum_i n_i^\tau = 0$ and the resulting value of τ is in fact an upper bound on ω, the largest point in Δ. The assumption of the tau theorem, $\sum_i \mathrm{MM}_{n_i} \lesssim r$, can naturally arise from a clever algorithm for such a direct sum of matrix multiplication problems (as indeed Schönhage demonstrated in his original paper). But other approaches to matrix multiplication algorithms, in particular the group-theoretic approach of Cohn and Umans [CU03, CKSU05, CU13], give rise to more general inequalities, in which the RHS is not a diagonal tensor, but another direct sum of matrix multiplication tensors. Namely, inequalities of the form $\sum_i \mathrm{MM}_{n_i} \lesssim \sum_j \mathrm{MM}_{m_j}$. We can again conclude from the duality theorem that for every $\tau \in \Delta$ we have $\sum_i n_i^\tau \le \sum_j m_j^\tau$. Can we again use this numeric inequality to derive an upper bound on ω? Notice that now the real function $\sum_j m_j^\tau - \sum_i n_i^\tau$ is not necessarily decreasing, and so potentially can be negative at some $\alpha < \omega$ as in Fig. 1.

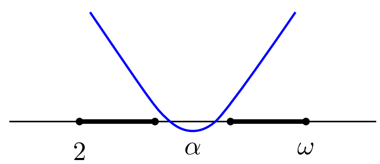


Figure 1: Can there be a non-monotone curve negative at some α with $2 < \alpha < \omega$? No, because Δ is connected.

Of course, such α will not be part of the spectrum, but how in this situation can we get an upper bound on ω? Clearly, the connectivity of Δ rules out this possibility, and allows us again to upper bound ω by the smallest root $\tau \ge 2$ of the function above.

Characterizing connectivity: type theory of polynomials and anchors We have seen that proving connectivity of spectra can be extremely powerful both for obtaining upper bounds on asymptotic rank, and ruling out lower bound approaches (and the opposite for asymptotic subrank). Now the question remains, how to prove connectivity for *any* preordered semiring? Indeed, how does Strassen prove his connectivity result for the spectrum of matrix multiplication? Our main impetus for this write-up was our attempt to understand Strassen's ingenious (and mysterious) proof, and to write it up in as much clarity and generality as possible. At a very high level, there are four main steps to his argument. We state them rather informally, and then try to explain and abstract the contents of each below, but caution that this subsection may be too complex to understand to readers without any background in this area; in this case please see the relevant technical sections, in which precise definitions and far more intuition are provided.

1. A "compression" or "self-correcting" theorem. Any algorithm which multiplies $n \times n$ matrices and errs only on matrices outside some large dimensional subspace (say of co-dimension $< n^2/4$), can be converted into an *errorless* algorithm of essentially the same complexity for multiplying $(n/2) \times (n/2)$ matrices.

2. A "cancellation" theorem: Assume that we manage to prove an asymptotic inequality $\mathrm{MM}_n \lesssim T + R$, where the RHS is the direct sum of two arbitrary tensors (not necessarily matrix

multiplication ones). Assume further that the dimension of T as a subspace of $n \times n$ matrices[28] is at most $d \leq n^2/4$ as above. Then, using the compression theorem above, we can remove T from the RHS at the cost of a small reduction in the size of matrices we multiply: $\mathrm{MM}_{n/2} \lesssim R$.

3. A "monomial ordering" theorem: any inequality between a pair of arbitrary direct *sums* of matrix multiplication tensors,[29] can be converted into an asymptotic inequality between (specially chosen) Kronecker *products* of these tensors, each a monomial of a high power of the original inequality.

4. A "connectivity" theorem: If the given preorder of polynomial inequalities can always be converted to an asymptotic "monomial ordering" between high powers, then the spectrum must be connected.

The description above already disentangles some parts of Strassen's presentation, and we try to clarify more by saying a few words about each part, and how we further separate, abstract and generalize each. Step (1) is a non-trivial linear algebra exercise, which the reader may attempt. We will give a new proof of it, which generalizes way beyond matrix multiplication tensors to all tensor networks. Step (2) is a rather simple use of linear projection, and generalizes as well to tensor networks as above. Step (3) is the most subtle and complex. Notice first the interplay between working in the general semiring of tensors, while being interested in the spectrum of the subsemiring of matrix multiplication. Moreover (and this cannot be clear without the spelling out of appropriate order of quantifiers in the statement), the following magic occurs. *The value of a single point in the asymptotic spectrum (here, it is the dimension of the matrix subspace above) implies bounds on the whole (unknown) spectrum.* We capture this magic in our notion of an *anchor*, which can be defined in the context of any semiring. Step (4) as well holds in the generality of any semiring, and is rather simple given (3). The proofs of steps (3) and (4) require several tools of asymptotic analysis of polynomial inequalities, which is developed in this generality.

Rectangular matrix multiplication and convexity Let us briefly conclude with the issue of convexity of the spectrum, a far stronger property (if it holds) than simply connectivity. We note that in the reals, every connected set is obviously convex as well. So, we trivially know that the (1-dimensional) spectrum of *square* matrix multiplication is convex. However, Strassen's theorem is much more general, and applies to the spectrum of *rectangular* matrix multiplication as well. The semiring which captures the complexity of rectangular matrix multiplication is generated by three tensors, and so its asymptotic spectrum is a subset of $\mathbb{R}^3_{\geq 1}$. Here connectivity certainly does not imply convexity. While the question of convexity remains open, Strassen's proof comes close to that: it shows that the spectrum is not only connected, but in fact it is "star-convex" with respect to a subset of the spectrum: for every point in this subset the line segments from it to *every* other point are fully contained in the spectrum.

We note that we do not know of applications (in practice) of convexity of the spectrum, either for the special case of matrix multiplication, or to any other semiring.

[28] Observe that each of the three "sides" of T is indexed by $n \times n$ matrices, and we can think of the fibers along any fixed side we choose as a set of matrices. It is the dimension of their span we refer to here as the "dimension" of T (this is sometimes called a "flattening rank").

[29] As in the generalized tau theorem above.

1.6. Organization of this paper

While it is fair to say that almost all ideas presented here appear in Strassen's three celebrated papers on the subject, our treatment of many of them, and their evolution, is often quite different, hopefully more accessible, general and current. We provide many modern examples and applications of the theory which did not exist when the original papers were written. We introduce new notions, interesting in their own right, which sometimes enable more intuitive, streamlined proofs, stronger theorems, and clearer connections to newly studied settings.

The paper may be viewed as having three separate (but obviously related) parts. We briefly describe each, and the contents of each section, highlighting (if any) new ingredients added here.

- Part I (Sections 2, 3 and 4): Here we develop the main aspects of the theory of asymptotic spectra. In Section 2 we will give a self-contained introduction to the general framework of preordered semirings, amortized complexity and the key parameters of interest. We conclude this section with numerous (some very recent) concrete examples from diverse fields of mathematics, information theory, optimization and computer science which fit into this framework. In Section 3 we introduce the asymptotic spectrum, and prove the main duality theorem of the theory, namely Strassen's duality theorem. Our proof is completely self-contained, without any reference to the Kadison–Dubois duality theory for semirings (which Strassen uses in his proof). In this we mainly follow the treatment in [Zui18]. As a new element, we define the *fractional* rank and subrank, which are relaxations of the integral ones. These help simplify the construction of the asymptotic spectrum.

- Part II (Sections 5, 6, 7, 8 and 9): Here we develop tools for asymptotic analysis of polynomial inequalities, with the central aim of understanding and characterizing connectivity and convexity of the asymptotic spectrum. We choose to focus on univariate polynomials only, as this situation is much simpler to describe, while it still captures all ideas (namely, it can be easily extended to multivariate polynomials in the same way). In Section 5 we give basic characterizations of boundedness and connectedness of the spectrum. In Section 6, the most technical in this part, we abstract and generalize, using new terminology, many of the technical manipulations of asymptotic polynomial inequalities in Strassen's (and later) papers as a "type theory" for polynomials, somewhat analogous to (and motivated by) the "typical sequences" used in Shannon's information theory. Section 7 then uses these to give new necessary and sufficient conditions for the connectedness of the spectrum. This is done using a new asymptotic partial order, called a *monomial order*, which uses the type decomposition above. The necessary condition is based on an interesting use of von Neumann's minimax theorem which uses the above information-theoretic representation in a natural way. In Section 8 we give a sequence of weaker and weaker sufficient conditions for connectivity, culminating in the existence of what we call an *anchor* of the semiring. We feel that this new notion clarifies a mysterious element in Strassen's connectivity proof for matrix multiplication, and makes it more modular. It separates an ingredient definable in any semiring, from the complementary ingredient of *compression* essential to the proof (which is dealt with in Part III). Finally, in Section 9 we discuss the consequences of connectivity of asymptotic spectra.

- Part III (Sections 11 and 12): In this part we return to the concrete semiring of tensors. It is devoted to shifting and compression theorems, and deriving from them connectivity proofs of spectra. In Section 11 we explain the fundamental notion of basis shifting and compression, and give a new proof of Strassen's basis shifting theorem for linear spaces of *matrices*. Our new proof easily extends by induction to basis shifting linear spaces of tensors of arbitrary dimension. Moreover, it is far simpler as our base case for the induction is 1-dimensional, namely basis shifting linear spaces of *vectors*, which is particularly simple (in contrast to Strassen's proof which deals directly with the 2-dimensional case). We also relate shifting to bounds on dimension of matrix and tensor subspaces, which were studied independently with different motivations. In Section 12 we define tensor networks, a special family of tensors important in quantum information theory and computation. Our generalization above of shifting and compression in all dimensions results in a connectivity theorem for the spectra of all tensor networks (of which matrix multiplication is the simplest non-trivial example).

- In the final Section 13 we discuss open problems and research directions (some very concrete, and some very broad) related to all the material in the paper.

Part I

Asymptotic spectra and duality

Contents

In Part I we develop the main aspects of Strassen's theory of asymptotic spectra. In Section 2 we introduce the basic definitions of this theory. In Section 3 we prove the main duality theorem. Finally, in Section 4 we discuss applications, variations and extensions of the theory.

Looking ahead to the other parts, in Part II we will discuss methods to gain structural insights in the theory of asymptotic spectra. In Part III we will apply these to a specific setting, namely matrix multiplication and tensor networks. While Part II and Part III are for a large part independent of Part I, their motivation comes directly from the material in Part I.

2. Semirings, Strassen preorders, rank and asymptotic rank

In this section we lay out the playing field by defining the kind of objects (elements of a semiring) and the kinds of reductions between them (Strassen preorders) that the theory of asymptotic spectra deals with. We introduce the main parameters of interest: rank, subrank, and their asymptotic versions. Then we discuss interesting fundamental examples.

2.1. Basic objects of study

Understanding the behavior of objects under natural addition and multiplication operations is a central theme in mathematics, computer science and physics. Examples of such basic objects include communication channels in information theory, computational problems in algebraic complexity theory, entangled states in quantum computing, and graphs in discrete mathematics. The addition and multiplication operations correspond to natural ways of combining the objects to form new ones. For example, for communication channels the addition and multiplication operations will correspond to combining channels *in series* or *in parallel*. Sometimes the existence and construction of a natural algebraic structure on such objects is far from obvious and requires some work.

In instances of interest, the addition and multiplication operations behave nicely with each other, so that the objects form a structure known as a *commutative semiring*. A commutative semiring is essentially a commutative ring without the requirement for every element to have an additive inverse. The formal definition is as follows.

Definition 2.1 (Commutative semiring)**.** For any set $\mathcal{R}$ and binary operations $+$ and $\cdot$ on $\mathcal{R}$, the triple $(\mathcal{R}, +, \cdot)$ is called a *commutative semiring* or *semiring*, for short, if the following properties hold:

1. The addition $+$ is associative and commutative.
2. The multiplication $\cdot$ is associative and commutative.
3. The multiplication distributes over the addition: $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.
4. There is a multiplicative unit 1.

We abbreviate $a \cdot b$ to ab. We write $2 = 1 + 1$, $3 = 1 + 1 + 1$, etcetera, for the elements in $\mathcal{R}$ obtained by adding 1 to itself. In this way in any commutative semiring we have the positive integers $\mathbb{N}_{\geq 1} = \{1, 2, \ldots\}$.[30] We do not require these elements in $\mathcal{R}$ to be distinct from each other at this point, but this will be a requirement that we impose later. We point out that we can artificially add or remove a zero element from the semiring without changing anything else. We tried to choose the definition that makes notation and proofs simplest and, therefore, we will not have a zero in our semiring.

Simple examples of commutative semirings are the natural numbers $\mathbb{N}_{\geq 1}$, the rationals ≥ 1 and the reals ≥ 1, with their usual addition and multiplication. Another example, which will play an important role later as a complete example, is the set of continuous functions from some topological space to the reals ≥ 1 under pointwise addition and multiplication. We will see much more complicated and interesting examples.

Besides addition and multiplication operations, we are interested in objects that naturally come with a (partial) ordering. This ordering will often correspond to some notion of reduction as is common in complexity theory, so that the larger the object in the ordering, the more expensive it is in terms of computational resources or communication resources, but also the more useful it is. The formal notion that we use — which captures the most basic and obvious properties that any kind of reduction should have — is that of a preorder relation.

[30] In this paper we use $\mathbb{N} = \{0, 1, 2, \ldots\}$.

Definition 2.2 (Preorder relation)**.** A relation $\leq$ on a set $\mathcal{R}$ is called a *preorder relation*, or preorder, if it satisfies the following properties:[31]

1. Reflexive: for every $a \in \mathcal{R}$ we have $a \leq a$.
2. Transitive: for every $a, b, c \in \mathcal{R}$, if $a \leq b$ and $b \leq c$, then $a \leq c$.

When we are manipulating preorders, it will often be convenient to identify our preorder $\leq$ on $\mathcal{R}$ with the set of pairs $P = \{(a, b) \in \mathcal{R}^2 : a \leq b\}$. Vice versa, given such a set P we will denote the corresponding inequality symbol by $\leq_P$. That is, for any $a, b \in \mathcal{R}$ we have $a \leq_P b$ if and only if $(a, b) \in P$. As usual, we write $a \not\leq_P b$ if and only if $(a, b) \notin P$. Naturally, we call any set $P \subseteq \mathcal{R}^2$ a preorder if the relation $\leq_P$ is a preorder.

We are interested in the interaction between the semiring operations $+$ and $\cdot$ and the preorder $\leq$. For this we first require the semiring operations and the preorder to be compatible. Namely, thinking of the preorder as a notion of reduction, we require reductions to *compose* under the semiring operations, in the same way that they do for the simple examples above. This is formalized as follows.

Definition 2.3 (Semiring preorder)**.** A preorder $\leq$ on a semiring $\mathcal{R}$ is called a *semiring preorder* if for every $a, b, c, d \in \mathcal{R}$, if $a \leq b$ and $c \leq d$, then also $a + c \leq b + d$ and $ac \leq bd$.

Note that a preorder $\leq$ is a semiring preorder if and only if for every $a, b, c \in \mathcal{R}$, if $a \leq b$, then also $a + c \leq b + c$ and $ac \leq bc$. This more compact characterization is often easier to deal with.

Finally, we will assume our preorder has three more properties. First, we assume that the simplest example of a semiring, the semiring of natural numbers $\mathbb{N}_{\geq 1}$, is embedded in our semiring in their natural order. In applications these embedded natural numbers will always correspond to the objects that we understand best. Secondly, we assume that elements are "non-negative". Third, we assume that the elements of our semiring are "not too large" and "not too small".[32][33] The formal notion that we use is as follows.

Definition 2.4 (Strassen preorder)**.** Let $\mathcal{R}$ be a semiring. (Then $\mathcal{R}$ by definition contains the positive integers $\mathbb{N}_{\geq 1}$ by adding the multiplicative unit 1 to itself.) A semiring preorder P on $\mathcal{R}$ is called a *Strassen preorder* if it satisfies the following properties:

1. Embedding of natural numbers: for every $n, m \in \mathbb{N}_{\geq 1}$ we have $n \leq m$ in $\mathbb{N}_{\geq 1}$ if and only if $n \leq_P m$ in $\mathcal{R}$.
2. For every $a, b \in \mathcal{R}$ it holds that $a \leq_P a + b$.[34]

[31] One could go one step further and require the relation to satisfy the antisymmetry property: for any $a, b \in \mathcal{R}$, if $a \leq b$ and $b \leq a$, then $a = b$. A preorder satisfying antisymmetry is called a partial order. We will not use partial orders.

[32] Compared to Strassen's treatment, we simplify the whole discussion by using a simplified version of his "not too small" condition. Namely we will simply require that all elements are at least 1 (as this serves all applications that we discuss), whereas Strassen's original condition asks that every (nonzero) element is at least 1 after being multiplied by some natural number. The condition that we use is a stronger version of what is commonly called the *Archimedean property*, so we will refer to it as the *strong Archimedean property*.

[33] It is easy to find semirings that do not satisfy the (strong) Archimedean property, for example the semiring of functions on a non-compact space.

[34] When there is a zero in the semiring, this condition can equivalently be replaced by $0 \leq_P 1$, since then $0 = 0 \cdot a \leq_P 1 \cdot a = a$ and so $a \leq_P a + b$, using that P is a semiring preorder. However, in our exposition we have no zero in our semiring.

3. Strong Archimedean property: for every $a \in \mathcal{R}$ it holds that $1 \leq_P a \leq_P n$ for some $n \in \mathbb{N}_{\geq 1}$.[35]

Semirings $\mathcal{R}$ with a Strassen preorder P (with inequality symbol denoted by $\leq_P$) will be our object of study for the rest of the section.

2.2. Rank, subrank and their asymptotic versions

Our main goal is to understand how the elements in the semiring relate to the natural numbers in the semiring in an asymptotic fashion. In order to discuss this, we first introduce two functions that relate the semiring elements to the natural numbers:

Definition 2.5 (Rank)**.** For every $a \in \mathcal{R}$ the *rank* $\mathrm{R}_P(a)$ is defined as the smallest number $n \in \mathbb{N}_{\geq 1}$ such that $a \leq_P n$.

Definition 2.6 (Subrank)**.** For every $a \in \mathcal{R}$ the *subrank* $\mathrm{Q}_P(a)$ is defined as the largest number $n \in \mathbb{N}_{\geq 1}$ such that $n \leq_P a$.

We may think of the rank $\mathrm{R}_P(a)$ as the *cost* of the element a in terms of our *currency*, the natural numbers. Similarly, the subrank $\mathrm{Q}_P(a)$ is the *value* of the element a in terms of the natural numbers. Of course value is at most cost, and they need not be equal: $\mathrm{Q}_P(a) \leq \mathrm{R}_P(a)$. Note that both rank and subrank require optimization over the integers. It is thus not surprising that they capture NP-hard parameters (like the rank of tensors and the independence number of graphs).

We will now discuss some basic properties of rank and subrank. To do this we will use the following terminology for any function $\phi : \mathcal{R} \to \mathbb{R}_{\geq 1}$. We say that ϕ is *sub-additive* if for every $a, b \in \mathcal{R}$ we have $\phi(a + b) \leq \phi(a) + \phi(b)$. We say that ϕ is *sub-multiplicative* if for every $a, b \in \mathcal{R}$ we have $\phi(ab) \leq \phi(a)\phi(b)$. The terms *super-additive* and *super-multiplicative* are defined similarly. We say that ϕ is *normalized* if $\phi(1) = 1$. Finally, we say that ϕ is *P-monotone* if for every $a, b \in \mathcal{R}$, if $a \leq_P b$, then $\phi(a) \leq \phi(b)$. The basic properties of rank and subrank are as follows; note that these are the properties we would intuitively expect from a cost function and a value function.

Lemma 2.7. *The rank R_P is sub-additive, sub-multiplicative, normalized and P-monotone. The subrank Q_P is super-additive, super-multiplicative, normalized and P-monotone.*

Proof. The properties follow directly from the definition of a Strassen preorder and the definition of rank and subrank. □

We have introduced the rank and subrank as cost and value functions with respect to the natural numbers. We now introduce the main parameters we want to understand, which are *amortized cost* and *amortized value*. More precisely, we want to understand how the rank and the subrank behave under taking large powers. For example, it is possible that $\mathrm{R}_P(a) = 2$ while $\mathrm{R}_P(a^2) = 3 < 2^2$, that is, large powers of the element a are relatively cheap compared to the cost of a. The following two parameters capture the amortized cost and amortized value:

Definition 2.8 (Asymptotic rank)**.** For every $a \in \mathcal{R}$ the *asymptotic rank* $\widetilde{\mathrm{R}}_P(a)$ is defined as the infimum of $\mathrm{R}_P(a^n)^{1/n}$ over all $n \in \mathbb{N}$.

[35] As remarked, Strassen's definition of a "good" preorder uses the Archimedean property rather than the strong Archimedean property. The Archimedean property says that for every $a, b \in \mathcal{R}$ with $b \neq 0$ there is an $n \in \mathbb{N}$ such that $a \leq_P nb$. Although the notion of good preorder is more general than our notion of Strassen preorder, we use the latter notion since it is a lot simpler and suffices for the applications that we discuss.

Definition 2.9 (Asymptotic subrank)**.** For every $a \in \mathcal{R}$ the *asymptotic subrank* $\widetilde{\mathrm{Q}}_P(a)$ is defined as the supremum of $\mathrm{Q}_P(a^n)^{1/n}$ over all $n \in \mathbb{N}$.

The following standard lemma, which is known as Fekete's Lemma [PS98, no. 98] (see also [Sch03, Theorem 2.2]), allows us to replace the infimum and supremum in the definition of asymptotic subrank and asymptotic rank by *limits*, highlighting their asymptotic quality.

Lemma 2.10 (Fekete's lemma)**.** *Let $r_1, r_2, \ldots$ be a sequence of non-negative real numbers that satisfy the sub-additivity property $r_{n+m} \leq r_n + r_m$ for all $n, m \in \mathbb{N}$. Then $\inf_n r_n/n = \lim_{n\to\infty} r_n/n$.*

Corollary 2.11 (Consequence of Fekete's Lemma)**.** *For every $a \in \mathcal{R}$ the asymptotic rank is given by the limit $\widetilde{\mathrm{R}}(a) = \lim_{n\to\infty} \mathrm{R}(a^n)^{1/n}$.*

Proof. Apply Fekete's Lemma to the sequence $r_n = \log \mathrm{R}(a^n)^{1/n}$ while using the fact that rank is sub-multiplicative (Lemma 2.7). □

Corollary 2.12 (Consequence of Fekete's Lemma)**.** *For every $a \in \mathcal{R}$ the asymptotic subrank is given by the limit $\widetilde{\mathrm{Q}}_P(a) = \lim_{n\to\infty} \mathrm{Q}_P(a^n)^{1/n}$.*

Proof. This proof is similar to the proof of Corollary 2.11. □

In several applications it is meaningful to know whether the limits in the expression for asymptotic rank and asymptotic subrank are attained at a finite power $n \in \mathbb{N}$. There is a natural abstract example (semiring $\mathcal{R}$ and a Strassen preorder P) that we will discuss in the next section (Example 2.17) which shows that generally:

- for every $n \in \mathbb{N}$ there is an element $a \in \mathcal{R}$ for which the asymptotic rank (or asymptotic subrank) is attained at the nth power but not at any smaller power
- there is an element for which the asymptotic rank (or asymptotic subrank) is not attained at any finite power.

Besides the aforementioned abstract example, there are also concrete examples where asymptotic rank and asymptotic subrank are not attained at any finite power, namely for the semirings of graphs (Example 4.4) and tensors (Example 4.5).

We conclude by discussing how the asymptotic rank and asymptotic subrank share some of the properties of the rank and the subrank (Lemma 2.7 and Lemma 2.7). Besides these properties being fundamental, these proofs serve as a warm-up to more abstract proofs later that use similar ingredients.

Lemma 2.13. *The asymptotic subrank $\widetilde{\mathrm{Q}}_P$ is super-additive, super-multiplicative, normalized and P-monotone.*

The proof of Lemma 2.13 is simple except for the proof of super-additivity, which we will give here. We will use the following basic lemma.

Lemma 2.14 (AM-GM inequality)**.** *For any non-negative real numbers x_1 and x_2 it holds that*

$$\max_{0\leq p\leq 1} 2^{h(p)} x_1^p x_2^{1-p} = x_1 + x_2$$

where $h(p) \coloneqq -p\log p - (1-p)\log(1-p)$ is the binary entropy function in which $0 \log 0$ is defined to be 0.

Proof. The AM-GM inequality states that for any non-negative real numbers y_1 and y_2 the p-weighted geometric average is at most the p-weighted arithmetic average, $y_1^p y_2^{1-p} \leq py_1 + (1-p)y_2$, with equality if and only if $y_1 = y_2$. Expanding the definition of $h(p)$ in the claim of the lemma, we need to prove that

$$\max_{0\leq p\leq 1} (\tfrac{x_1}{p})^p\, (\tfrac{x_2}{1-p})^{1-p} = x_1 + x_2.$$

We may assume that both x_1 and x_2 are nonzero, since the case that either is zero can be dealt with directly. It then suffices to prove that

$$\max_{0< p< 1} (\tfrac{x_1}{p})^p\, (\tfrac{x_2}{1-p})^{1-p} = x_1 + x_2.$$

Letting $y_1 = x_1/p$ and $y_2 = x_2/(1-p)$, we thus need to prove that

$$y_1^p\, y_2^{1-p} \leq py_1 + (1-p)y_2$$

holds for all p, with equality for some choice of p. By the AM-GM inequality we have that $y_1^p\, y_2^{1-p} \leq py_1 + (1-p)y_2$ for every p. We now choose p such that $y_1 = y_2$. One may verify that this is true for $p = x_1/(x_1 + x_2)$. The tightness condition of the AM-GM inequality then gives the equality $y_1^p\, y_2^{1-p} = py_1 + (1-p)y_2$. □

Proof of Lemma 2.13. Super-multiplicativity, normalization and monotonicity are not hard to prove. We give the proof of super-additivity. For any fixed $n \in \mathbb{N}$, we consider the binomial expansion $(a+b)^n = \sum_p \binom{n}{pn} a^{pn} b^{(1-p)n}$, where the sum goes over all $0 \leq p \leq 1$ such that pn is integral. We will use the important relation between binomial coefficients and the binary entropy function, which says that $\binom{n}{pn} = 2^{nh(p)-o(n)}$. Then by the super-additivity and super-multiplicativity of the subrank, we find that

$$\mathrm{Q}((a+b)^n) \geq \sum_p \binom{n}{pn} \mathrm{Q}(a^{pn})\, \mathrm{Q}(b^{(1-p)n}),$$

where the sum goes over all $0 \leq p \leq 1$ such that pn is integral. We may lower bound the sum on the right-hand side by lower bounding its largest summand as:

$$\sum_p \binom{n}{pn} \mathrm{Q}(a^{pn})\, \mathrm{Q}(b^{(1-p)n}) \geq \max_p\, 2^{nh(p)-o(n)}\, \mathrm{Q}(a^{pn})\, \mathrm{Q}(b^{(1-p)n}),$$

where the maximization goes over all $0 \leq p \leq 1$ such that pn is integral. Taking the nth root of the right-hand side and letting n go to infinity gives

$$\lim_{n\to\infty} \Big(\max_{\substack{0\leq p\leq 1:\\ pn\in\mathbb{N}}} 2^{nh(p)-o(n)}\, \mathrm{Q}(a^{pn})\, \mathrm{Q}(b^{(1-p)n}) \Big)^{1/n} = \max_{0\leq p\leq 1} 2^{h(p)}\, \widetilde{\mathrm{Q}}(a)^p\, \widetilde{\mathrm{Q}}(b)^{1-p}.$$

Combining the above three steps gives

$$\widetilde{\mathrm{Q}}(a+b) \geq \max_p 2^{h(p)}\, \widetilde{\mathrm{Q}}(a)^p\, \widetilde{\mathrm{Q}}(b)^{1-p}$$

where the maximization goes over all $0 \leq p \leq 1$. From Lemma 2.14 we know that

$$\max_p 2^{h(p)}\, \widetilde{\mathrm{Q}}(a)^p\, \widetilde{\mathrm{Q}}(b)^{1-p} = \widetilde{\mathrm{Q}}(a) + \widetilde{\mathrm{Q}}(b)$$

which finishes the proof. □

Lemma 2.15. *The asymptotic rank $\widetilde{\mathrm{R}}_P$ is sub-additive, sub-multiplicative, normalized and P-monotone.*

Proof. Again the sub-multiplicativity, normalization and monotonicity are not hard to prove. The sub-additivity is proven similarly as for Lemma 2.13 with one additional observation, namely that the number of summands in the binomial expansion $(a+b)^n = \sum_p \binom{n}{pn} a^{pn} b^{(1-p)n}$ grows polynomially in n. This extra polynomial factor disappears when taking the nth root and letting n go to infinity. $\square$

2.3. Examples

We will now discuss examples of semirings and Strassen preorders, building from the simplest examples to the most complicated. In some of these semirings, even understanding the asymptotic rank or subrank of one particular element is not only difficult (e.g., the graph C_7 in Example 2.20), but actually captures a major problem in a field (e.g., for $\mathrm{MM}_{2,2,2}$ in Example 2.22 it captures the whole complexity of matrix multiplication, and for the trivial hypergraph in Example 2.24 it captures the full cap set problem.)

Example 2.16 (Natural numbers)**.** We begin with the absolutely simplest example, and build up from there. Let $\mathcal{R}$ be the semiring $\mathbb{N}_{\geq 1} = \{1, 2, 3, \ldots\}$ of positive integers, with the usual addition and multiplication operations, and let P be the usual preorder on $\mathbb{N}_{\geq 1}$.

This preorder P is directly verified to be a Strassen preorder. The rank, subrank, asymptotic rank and asymptotic subrank all coincide, namely for every $n \in \mathbb{N}$ they take value n. Two very special properties of this simplest example that are very much unlike the typical examples, are that (asymptotic) subrank and rank coincide and that the preorder P is a total order (i.e., linear order). In the following examples, we will be moving away from these special properties towards preorders that have a more intricate structure, and that are in particular not total orders. Moving away from total orders is really necessary to get any interesting examples: every totally ordered Archimedean semiring is a subsemiring of the positive real numbers.

Example 2.17 (Reals)**.** This example slightly extends the previous example (Example 2.16). The preorder will still be total, but the rank and subrank will no longer be equal. Let $\mathcal{R}$ be the semiring $\mathbb{R}_{\geq 1}$ of all real numbers ≥ 1, with the usual addition and multiplication operations, and let P be the usual preorder on $\mathbb{R}_{\geq 1}$. Again, P is trivially a Strassen preorder. The natural numbers are obviously embedded in $\mathcal{R}$. The rank is the ceiling function and the subrank is the floor function. The asymptotic rank and asymptotic subrank simply have value r for every $r \in \mathbb{R}_{\geq 1}$. Indeed, $\inf_n \lceil r^n \rceil^{1/n} = \sup_n \lfloor r^n \rfloor^{1/n} = r$. Thus subrank and rank no longer coincide with each other and with their asymptotic versions, although the gaps are at most 1. However, we are still in the atypical situation that the preorder is total.

Example 2.18 (Matrices)**.** Here is another simple example, but in disguise. Indeed, this example will turn out to be essentially equivalent to the example of the natural numbers (Example 2.16), but only after some work. We will later extend this example to tensors (Example 2.21), where it becomes highly non-trivial. Another related non-trivial example that we will discuss later is the example of graphs (Example 2.20), which can be thought of as a symmetric and combinatorial variation on this example.

Let $\mathbb{F}$ be any field and let $\mathcal{R}$ be the set of all nonzero matrices of any dimensions with coefficients in $\mathbb{F}$. So S contains all the nonzero $n \times m$ matrices for any $n, m \in \mathbb{N}_{\geq 1}$. We define addition and multiplication on $\mathcal{R}$ as the direct sum $\oplus$, which puts matrices as blocks on the main diagonal,

$$A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

and the tensor product $\otimes$ (a.k.a. the Kronecker product), which is the matrix consisting of all pairwise products of the coefficients of A and B,

$$A \otimes B = (A_{ij} B_{k\ell})_{(i,k),(j,\ell)}.$$

To make $\mathcal{R}$ a commutative semiring we identify any two matrices that are equal up to permuting rows and permuting columns. We define the preorder P on $\mathcal{R}$ by saying that for any matrices X and Y we have $X \leq_P Y$ if and only if there are matrices A and B such that $X = AYB$, where A and B must have the appropriate shape such that the matrix product AYB makes sense. In other words, $X \leq_P Y$ if and only if X can be obtained from Y by taking linear combinations of the rows and the columns. The natural numbers are embedded in $\mathcal{R}$ as the 1×1 identity matrix I_1, the 2×2 identity matrix I_2, and so on. With these definitions it is not hard to see that P is a Strassen preorder. The most fundamental matrix parameter is the matrix rank. One verifies directly that the matrix rank $\mathrm{rk}(A)$ of a matrix A equals the smallest number n such that $A \leq_P I_n$. On the other hand, it follows from Gaussian elimination that the matrix rank of A also equals the largest n such that $I_n \leq_P A$. Thus rank and subrank are equal to matrix rank, and it then follows from the general relation $\mathrm{Q}(A) \leq \widetilde{\mathrm{Q}}(A) \leq \widetilde{\mathrm{R}}(A) \leq \mathrm{R}(A)$ that also their asymptotic versions are equal to the matrix rank. To put it differently, for every matrices A and B we have $A \leq_P B$ if and only if $\mathrm{rk}(A) \leq \mathrm{rk}(B)$. Thus, identifying matrices with their matrix rank (and noting that matrix rank is additive under direct sum and multiplicative under Kronecker product) shows that this example is equivalent to the example of natural numbers Example 2.16.

Example 2.19 (Continuous functions)**.** We have seen only examples of total preorders so far. We now go from the simplest examples straight to the richest example. We will in fact see that this example is *complete*, in the sense that all possible behavior of any semiring with a Strassen preorder can be simulated by this semiring (Section 4.4). Fortunately, although this is the richest example, we do fully understand it.

We start with a subcase that is more concrete than the general case, but already shows the new behavior. Let $k \geq 1$ be some integer and let $\mathcal{R}$ be the set of k-tuples $\mathbb{R}^k_{\geq 1}$, with addition and multiplication defined pointwise. Thus for any two k-tuples $f = (f_1, \ldots, f_k), g = (g_1, \ldots, g_k) \in \mathcal{R}$ we have $f \cdot g = (f_1 g_1, \ldots, f_k g_k)$ and $f + g = (f_1 + g_1, \ldots, f_k + g_k)$. Let P be the pointwise preorder on $\mathcal{R}$, that is, for every $f, g \in \mathcal{R}$ we have $f \leq_P g$ if and only if for every $i = 1, \ldots, k$ we have $f_i \leq g_i$. The natural numbers in this semiring are the tuples $(1, \ldots, 1)$, $(2, \ldots, 2)$, $(3, \ldots, 3)$, and so on. It is clear that with these definitions P is a Strassen preorder. Note that if $k = 1$, then $\mathcal{R}$ is simply $\mathbb{R}_{\geq 1}$ and P is the total preorder as discussed in Example 2.17. However, when $k \geq 2$ the preorder P becomes manifestly *non*-total. We see that the subrank $\mathrm{Q}_P(f)$ is the floor of $\min_i f_i$ and the rank $\mathrm{R}_P(f)$ is the ceiling of $\max_i f_i$. The asymptotic subrank equals $\min_i f_i$ and the asymptotic rank equals $\max_i f_i$.

For the general case, instead of considering k-tuples of elements from $\mathbb{R}_{\geq 1}$, we consider certain collections of elements from $\mathbb{R}_{\geq 1}$ indexed by topological spaces. More precisely, let $\mathcal{X}$ be any nonempty compact space — we will use compactness to apply the extreme value theorem. Let $\mathcal{R}$ be the semiring $C(\mathcal{X}, \mathbb{R}_{\geq 1})$ of all continuous functions from $\mathcal{X}$ to $\mathbb{R}_{\geq 1}$ with addition and multiplication defined pointwise. In other words, for every $f, g \in \mathcal{R}$ and $x \in \mathcal{X}$ we have $(f \cdot g)(x) = f(x)g(x)$

and $(f+g)(x) = f(x) + g(x)$. Let P be the pointwise preorder on $\mathcal{R}$, that is, for every $f, g \in \mathcal{R}$ we have $f \leq_P g$ if and only if for every $x \in \mathcal{X}$ it holds that $f(x) \leq g(x)$. One verifies that this is a Strassen preorder; indeed the strong Archimedean property follows from the extreme value theorem which says that every continuous function on a compact space attains a maximum. We see that the earlier subcase corresponds to taking $\mathcal{X}$ to be a discrete space of k points. From the extreme value theorem, it is easy to see, as a natural extension of the situation in the simpler subcase, that $\min_{x\in\mathcal{X}} f(x) = \underset{\sim}{\mathrm{Q}}_P(f)$ and $\underset{\sim}{\mathrm{R}}_P(f) = \max_{x\in\mathcal{X}} f(x)$. We conclude, imagining that $\mathcal{X}$ is explicitly given to us, that we know everything there is to know about the asymptotic subrank and the asymptotic rank, as they are simply the pointwise minimum and pointwise maximum over $\mathcal{X}$, respectively. The next examples will be much more interesting since we do not have this level of understanding.

Example 2.20 (Graphs and the Shannon capacity)**.** Now we are ready for the first example that we *do not* understand, and for which progress is an open research problem in discrete mathematics and information theory. This example is about a combinatorial parameter called the Shannon capacity of a graph, which arises from basic combinatorial objects (graphs) and operations, and models efficient communication. This combinatorial parameter was introduced by Shannon in his seminal paper from 1956 [Sha56]. One way to think about this example is as a symmetric and combinatorial variation on the example of matrices (Example 2.18). We will consider an extension of this example (to k-uniform hypergraphs) later (Example 2.23).

For any finite set V we denote by $\binom{V}{2}$ the set of all subsets of V of size two. For any subset $E \subseteq \binom{V}{2}$ we call the pair $G = (V, E)$ a *graph*. The elements of V are called the *vertices* of G and the element of E are called the *edges* of G, and thus V is called the *vertex set* of G and E is called the *edge set* of G. For any two vertices $u, v \in V$ we say that u and v are *adjacent* in G if $\{u, v\} \in E$, and we say that u and v are *non-adjacent* in G if $\{u, v\} \notin E$. Given a graph G we will denote its vertex set by $V(G)$ and its edge set by $E(G)$.

For every $n \in \mathbb{N}$ we denote by K_n the graph with vertex set $V(K_n) = \{1, 2, \ldots, n\}$ and all possible edges, $E(K_n) = \binom{V(K_n)}{2}$. This graph is called the *complete graph* on n vertices. In other words, K_n is the graph on n vertices in which any two vertices are adjacent. For any graph G we denote by $\overline{G}$ the graph with vertex set $V(\overline{G}) = V(G)$ and edge set $E(\overline{G}) = \binom{V}{2} \setminus E(G)$. This is called the *complement graph* of G. In other words, we go from G to the complement graph $\overline{G}$ by making adjacent vertices non-adjacent and vice versa. For example, the empty graph on n vertices $\overline{K_n}$ is the graph with vertex set $V(\overline{K_n}) = \{1, 2, \ldots, n\}$ and no edges, $E(\overline{K_n}) = \emptyset$.

For two graphs G and H we say $\phi : G \to H$ is a *homomorphism* if ϕ is a map $V(G) \to V(H)$ such that for every $u, v \in V(G)$, if $\{u, v\} \in E(G)$, then $\{\phi(u), \phi(v)\} \in E(H)$. Two graphs G and H are called *isomorphic* if $|V(G)| = |V(H)|$ and there is a bijective map $\pi : V(G) \to V(H)$ such that for every $u, v \in V(G)$ we have that $\{u, v\} \in E(G)$ if and only if $\{\pi(u), \pi(v)\} \in E(H)$. In other words, G and H are isomorphic if they are the same up to relabeling on their vertex names. Since we are only interested in properties of graphs that are invariant under such relabelings, we may tacitly identify any two isomorphic graphs, that is, formally we are not working with graphs but with isomorphism classes of graphs.[36] We need this technicality in the following.

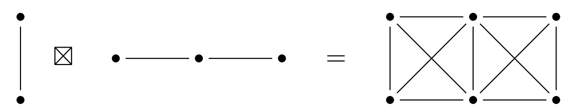
The objects of our semiring $\mathcal{R}$ will be all (non-empty) graphs. To turn this into a commutative semiring we need an addition and a multiplication operation. The natural choice that is motivated

[36] This is exactly the same with what we did with matrices in Example 2.18, allowing permutations of rows and columns. Indeed, it is often useful to identify graphs with their adjacency matrix. Then graphs are isomorphic precisely when their adjacency matrices are equal up to simultaneously permuting rows and columns.

by communication channels (more about this later) is to define the addition via the disjoint union and the multiplication via the strong product (also called and-product). Let G and H be two graphs. The *disjoint union* $G \sqcup H$ is the graph with vertex set given by the disjoint union of the vertex sets of G and H and edge set given by the disjoint union of the edge sets of G and H. The *strong product* $G \boxtimes H$ is the graph with vertex set given by the cartesian product $V(G) \times V(H)$ and edge set given by all pairs $\{(u_1, v_1), (u_2, v_2)\}$ such that $(u_1, v_1) \neq (u_2, v_2)$ and

$$(\{u_1, u_2\} \in E(G) \text{ or } u_1 = u_2) \text{ and } (\{v_1, v_2\} \in E(H) \text{ or } v_1 = v_2).$$

For example, the strong product of the graph with one edge K_2 and the path with three vertices P_3 looks like



The strong product is essentially the same as the Kronecker product $\otimes$ for matrices. Namely, letting $A(G)$ be the adjacency matrix of G with 1's on the main diagonal, we have $A(G \boxtimes H) = A(G) \otimes A(H)$. Recall that we are implicitly working with isomorphism classes of graphs and here is why: The disjoint union and the strong product are commutative operations, since $G \sqcup H$ and $H \sqcup G$ are isomorphic, and $G \boxtimes H$ and $H \boxtimes G$ are isomorphic. We will write $G \cdot H$ for $G \boxtimes H$, and $G + H$ for $G \sqcup H$. Under these operations the set of graphs becomes a semiring.

The role of natural numbers in the semiring of graphs is played by the graphs $\overline{K_n}$ for $n \in \mathbb{N}$, which have n vertices and no edges. Indeed, $\overline{K_n} = \overline{K_1} \sqcup \cdots \sqcup \overline{K_1}$ (n summands) and $\overline{K_1}$ acts as a multiplicative unit in the semiring.

To discuss the natural preorder on graphs we start by discussing two important graph parameters, namely the independence number $\alpha(G)$ and the clique cover number $\overline{\chi}(G)$. For any graph G, a subset $S \subseteq V(G)$ is called an *independent set* in G if every two vertices $u, v \in S$ are non-adjacent. The independence number $\alpha(G)$ is the largest number $n \in \mathbb{N}$ such that there is an independent set in G of size n. A subset $S \subseteq V(G)$ is called a *clique* in G if every two distinct vertices $u, v \in S$ are adjacent. A *clique cover* of G is a partition of the vertices of G into cliques, that is, it is a collection of (disjoint) subsets $S_1, \ldots, S_n \subseteq V(G)$ such that $\cup_i S_i = V(G)$ and each S_i is a clique. The number of cliques in the clique cover is called its size. The *clique cover number* $\overline{\chi}(G)$ is the smallest number n such that there is a clique cover of G of size n.

There is a natural preorder P on the set of graphs such that the corresponding notion of rank R_P equals the clique cover number while the notion of subrank Q_P equals the independence number. Namely, we define the *cohomomorphism preorder* P on graphs by saying that $G \leq_P H$ if and only if there is a graph homomorphism $\overline{G} \to \overline{H}$.[37] It is then readily verified that indeed $\mathrm{R}_P(G) = \overline{\chi}(G)$ and $\mathrm{Q}_P(G) = \alpha(G)$. It is also readily verified that P is a Strassen preorder. This preorder is not total. We refer to the book [HN04] for more information on the cohomomorphism preorder.

The asymptotic subrank $\widetilde{\mathrm{Q}}_P(G) = \lim_{n\to\infty} \alpha(G^n)^{1/n}$, which we will denote as is usual by $\Theta(G)$, is known as the *Shannon capacity* in graph theory, and this is the main graph parameter that we want to understand. So far, the value of $\Theta(G)$ is known precisely for very few types of graphs. For a subclass of graphs called *perfect* graphs, $\alpha(G) = \Theta(G) = \overline{\chi}(G)$. A non-perfect graph for which the Shannon capacity is known is the cycle of length 5 [Lov79], but the Shannon capacity of any larger odd cycle is not known (see, e.g., [PS19]). Other basic questions include: Can the Shannon capacity

[37] Thus $G \leq_P H$ if and only if there is a (not necessarily injective) map from the vertices of G to the vertices of H which preserves non-edges.

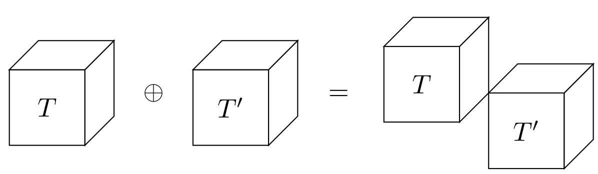
be computed? [AL06] What tools are there to bound the Shannon capacity? [Hae79, Lov79, BC19] How does the Shannon capacity behave under the semiring operations? [Hae79, Alo98b] How does the Shannon capacity behave on random graphs? [Juh82] We will get partial answers to and deeper insight into some of these questions using the theory of asymptotic spectra.

To conclude this example, we say something about Shannon's [Sha56] motivation behind the definition of the Shannon capacity of a graph. A graph G models a communication channel from one party to another, in which $V(G)$ is the alphabet of symbols that may be transmitted, and $E(G)$ captures the pairs of symbols which the channel my confuse during transmission. Shannon asked the mathematically clean question of how much information can be sent over the channel if we are allowed to use the channel many times and if no errors are allowed. For one usage of the channel the answer is the independence number of G. For k usages of the channel the answer is the independence number of the kth power $G^{\boxtimes k}$. In this way the Shannon capacity of a graph measures the amount of information that can be sent over the channel, per usage of the channel, in the limit.

Example 2.21 (Tensors)**.** Understanding tensors (high-dimensional matrices) is deeply connected to several hard mathematical and computational problems, including problems in additive combinatorics [Tao16], the P versus NP problem [Raz09] and the theory of quantum entanglement [DVC00]. Strassen introduced the theory of asymptotic spectra in [Str88] in order to understand the computational complexity of matrix multiplication. We will discuss matrix multiplication in Example 2.22. Tensors are higher-dimensional matrices, so this example will obviously extend our matrix example Example 2.18. Recall that in Example 2.18 we considered the semiring of matrices under direct sum and Kronecker product and a preorder corresponding to taking linear combinations of rows and columns. Via Gaussian elimination we observed that this preordered semiring is essentially equivalent to the simplest semiring $\mathbb{N}$. The natural generalization to tensors that we are about to consider turns out to be much more complex. In particular, we do not understand it well and its study is an active research area, with links to quantum information theory, additive combinatorics (Example 2.24) and complexity theory (Example 2.22) that we will see later.

Let us define tensors and their semiring operations. Let $k \geq 2$ and let $\mathbb{F}$ be a field. For positive integers $n_1, \ldots, n_k$ we call any k-dimensional array $T = (T_{i_1,\ldots,i_k})_{i_1 \in [n_1],\ldots,i_k \in [n_k]}$ of field elements $T_{i_1,\ldots,i_k} \in \mathbb{F}$ an $n_1 \times \cdots \times n_k$ *tensor over* $\mathbb{F}$ or a k*-tensor over* $\mathbb{F}$. Thus a 2-tensor is simply a matrix. We let $\mathbb{F}^{n_1 \times \cdots \times n_k}$ denote the set of $n_1 \times \cdots \times n_k$ tensors.

To simplify the notation we will now focus on $k = 3$. Let $\mathcal{R}$ be the set of all nonzero $n_1 \times n_2 \times n_3$ tensors going over all possible values of n_1, n_2, n_3. We tacitly identify any two tensors that are equal up to a permutation among each of the three indices, like we did for matrices. We also define direct sum and Kronecker product analogously. Namely, *the direct sum* $T \oplus T' \in \mathbb{F}^{(n_1+n_1')\times(n_2+n_2')\times(n_3+n_3')}$ of $T \in \mathbb{F}^{n_1\times n_2\times n_3}$ and $T' \in \mathbb{F}^{n_1'\times n_2'\times n_3'}$ is defined by setting $(T\oplus T')_{i_1,i_2,i_3} = T_{i_1,i_2,i_3}$ for $i_1 \in [n_1], i_2 \in [n_2], i_3 \in [n_3]$ and setting $(T \oplus T')_{n_1+j_1,n_2+j_2,n_3+j_3} = T'_{j_1,j_2,j_3}$ for $j_1 \in [n_1'], j_2 \in [n_2'], j_3 \in [n_3']$; the other coefficients are set to zero. Pictorially, the direct sum thus looks as follows:



The *Kronecker product* $T \otimes T' \in \mathbb{F}^{(n_1n_1')\times(n_2n_2')\times(n_3n_3')}$ of $T \in \mathbb{F}^{n_1\times n_2\times n_3}$ and $T' \in \mathbb{F}^{n_1'\times n_2'\times n_3'}$ is defined by setting $(T \otimes T')_{(i_1,j_1),(i_2,j_2),(i_3,j_3)} = T_{i_1,i_2,i_3}T'_{j_1,j_2,j_3}$. The direct sum and Kronecker product are commutative operations on isomorphism classes of tensors and turn $\mathcal{R}$ into a semiring. Also the embedding of the integers into this semiring is analogous to matrices. Namely, for every $n \in \mathbb{N}$ we let I_n denote the $n \times n \times n$ tensor with ones on the main diagonal and zeros elsewhere, that is, $(I_n)_{i_1,i_2,i_3}$ equals 1 if $i_1 = i_2 = i_3$ and 0 otherwise. These are called *unit tensors*. The natural numbers in $\mathcal{R}$ are then given by the tensors I_n for $n \in \mathbb{N}$.

The preorder that we will define not only naturally extends the preorder that we defined on matrices, but also models natural notions of reduction in applications (e.g., the notion of reduction between arithmetic circuits by linear projections by Valiant [Val79]). We define the *restriction preorder* P on $\mathcal{R}$ by saying for any tensors $T \in \mathbb{F}^{n_1 \times n_2 \times n_3}$ and $T' \in \mathbb{F}^{n_1' \times n_2' \times n_3'}$ that $T \leq_P T'$ if T' can be transformed to T by applying, for each of the three directions, a linear transformation to all fibers of T' in that direction. Formally, $T \leq_P T'$ if there are matrices A_1, A_2, A_3 of the appropriate size such that $T = (A_1, A_2, A_3) \cdot T'$ where $((A_1, A_2, A_3) \cdot T')_{u,v,w} = \sum_{i,j,k} (A_1)_{u,i} (A_2)_{v,j} (A_3)_{w,k} T'_{i,j,k}$. It is not hard to verify that this is a Strassen preorder.

What is the meaning of rank R_P and subrank Q_P for the chosen preorder P? An important and natural extension of matrix rank to tensors is the notion of *tensor rank*. The tensor rank is defined as follows. We say that a tensor $S \in \mathbb{F}^{n_1 \times n_2 \times n_3}$ has *tensor rank one* if and only if there are nonzero vectors $u \in \mathbb{F}^{n_1}, v \in \mathbb{F}^{n_2}, w \in \mathbb{F}^{n_3}$ such that $T_{ijk} = u_i v_j w_k$. For every tensor T we define the *tensor rank* $\mathrm{rk}(T)$ as the smallest number n such that $T = \sum_{i=1}^{n} S_i$ where the S_i are tensors of rank one. (Here the sum is the entry-wise sum, not the direct sum.) Recall that the *rank* $\mathrm{R}_P(T)$ is defined as the minimum number n such that $T \leq_P I_n$. It is not hard to prove that the tensor rank $\mathrm{rk}(T)$ equals the rank $\mathrm{R}_P(T)$. Recall that for matrices it followed from Gaussian elimination that subrank equals rank. For tensors this is not the case; the subrank $\mathrm{Q}_P(T)$, which is defined as the largest number n such that $I_n \leq_P T$, can in fact be strictly smaller than the rank. It is known that the tensor rank is NP-hard to compute (and NP-complete over finite fields) [Hås90]. The same is likely true for the subrank.

We are interested in the asymptotic rank $\widetilde{\mathrm{R}}_P(T)$ and the asymptotic subrank $\widetilde{\mathrm{Q}}_P(T)$. The asymptotic rank $\widetilde{\mathrm{R}}_P(T)$ is the amortized tensor rank of $T^{\otimes n}$ when we let n go to infinity, and similarly the asymptotic subrank $\widetilde{\mathrm{Q}}_P(T)$ is the amortized subrank of $T^{\otimes n}$ when we let n go to infinity. Several applications require the study of asymptotic rank and subrank of tensors. For example such parameters are central in the theory of matrix multiplication and in problems in additive combinatorics and quantum information theory.

Example 2.22 (Matrix multiplication)**.** We have already discussed matrix multiplication in the introduction, but will now discuss it more carefully and rigorously. Matrix multiplication is an important subroutine in many algorithms. It is an open problem what the computational complexity of this problem is. To multiply two $n \times n$ matrices, the standard matrix multiplication algorithm uses $\mathcal{O}(n^3)$ arithmetic operations (additions and multiplications of scalars). Strassen in 1969 designed an algorithm that uses only $\mathcal{O}(n^{2.81})$ operations [Str69]. Strassen's algorithm was only the beginning of the search for ever faster algorithms, powered by ever more ingenious techniques. The best algorithm currently [AW21] uses $\mathcal{O}(n^{2.37\cdots})$ operations. On the other hand, we know that at least n^2 operations are required, leaving open the question whether the exponent $2.37\ldots$ can be improved all the way down to 2. The *exponent of matrix multiplication*, denoted by ω, is defined as the infimum over all real numbers such that the complexity of multiplying two $n \times n$ matrices is $\mathcal{O}(n^\omega)$. Thus we know that $2 \leq \omega \leq 2.37$.

It was realized early on [Str72, Pan72, Pan78] that *tensors* provide the right framework to study the complexity of matrix multiplication via the general correspondence between bilinear maps

$\mathbb{F}^{n_1} \times \mathbb{F}^{n_2} \to \mathbb{F}^{n_3}$, trilinear maps $\mathbb{F}^{n_1} \times \mathbb{F}^{n_2} \times \mathbb{F}^{n_3} \to \mathbb{F}$ and tensors in $\mathbb{F}^{n_1 \times n_2 \times n_3}$. We will now discuss how the matrix multiplication problem is phrased in terms of tensors and thus connects to Example 2.21, and in particular to asymptotic tensor rank. Multiplication of $a \times b$ and $b \times c$ matrices is a bilinear map $\mathbb{F}^{ab} \times \mathbb{F}^{bc} \to \mathbb{F}^{ca}$, where we think of $\mathbb{F}^{ab}$ as the space of $a \times b$ matrices, etcetera. Via the aforementioned general correspondence, the corresponding trilinear map $\mathbb{F}^{ab} \times \mathbb{F}^{bc} \times \mathbb{F}^{ca} \to \mathbb{F}$ is the function $(X_1, X_2, X_3) \mapsto \mathrm{Tr}(X_1 X_2 X_3)$ that takes the trace of the product of three matrices. The corresponding tensor in $\mathbb{F}^{(ab)\times(bc)\times(ac)}$, denoted by $\mathrm{MM}_{a,b,c}$ is defined as follows. Denote by $e_{(i,j),(k,\ell),(m,n)}$ the standard basis elements of the space $\mathbb{F}^{(ab)\times(bc)\times(ac)}$. Then

$$\mathrm{MM}_{a,b,c} \coloneqq \sum_{i\in[a]} \sum_{j\in[b]} \sum_{k\in[c]} e_{(i,j),(j,k),(k,i)} \in \mathbb{F}^{(ab)\times(bc)\times(ca)}$$

where the sum is the usual element-wise sum. We will denote $\mathrm{MM}_{a,a,a}$ by MM_a. Crucially, first of all, it is true (and the reader is challenged to prove) that if $\mathrm{R}(\mathrm{MM}_n) \leq r$, then there is an arithmetic algorithm to multiply any two $n \times n$ matrices using r scalar multiplications. The implication in the opposite direction is also true, up to a constant factor. Second of all, the matrix multiplication tensors have the property that the product of matrix multiplication tensors $\mathrm{MM}_{a,b,c} \in \mathbb{F}^{(ab)\times(bc)\times(ca)}$ and $\mathrm{MM}_{d,e,f} \in \mathbb{F}^{(de)\times(ef)\times(fd)}$ is again a matrix multiplication tensor: $\mathrm{MM}_{a,b,c} \otimes \mathrm{MM}_{d,e,f} = \mathrm{MM}_{ad,be,cf} \in \mathbb{F}^{(adbe)\times(becf)\times(cfad)}$. This is essentially for the same reason that multiplying block matrices can be done block-wise. From these ingredients it can be shown that $\widetilde{\mathrm{R}}(\mathrm{MM}_2) = 2^\omega$ [Str88], that is, the matrix multiplication exponent ω is characterized by the asymptotic rank of the matrix multiplication tensor MM_2.

Example 2.23 (Hypergraphs)**.** We have seen in Example 2.20 how to define a semiring and preorder on graphs in order to define the Shannon capacity as an asymptotic subrank. In this example we will consider a generalization of graphs, namely k-uniform hypergraphs. The generalization is fairly direct. There is one technical issue to take care of in the definition of the preorder so that it remains a Strassen preorder. Uniform hypergraphs are basic objects in combinatorics and we will see in Example 2.24 how they play a role in problems like the cap set problem and the sunflower problem.

Let $k \geq 2$. For any finite set V we denote by $\binom{V}{k}$ the set of all subsets of V of size k. For any subset $E \subseteq \binom{V}{k}$ we call the pair $G = (V, E)$ a *k-uniform hypergraph*, or simply k-graph. We will use the same terminology and notation of vertices and edges as for graphs.

For every $n \in \mathbb{N}_{\geq 1}$ we denote by K_n the k-graph with vertex set $V(K_n) = \{1, 2, \ldots, n\}$ and all possible edges, $E(K_n) = \binom{V(K_n)}{k}$. This is the *complete k-graph.* For any k-graph G we denote by $\overline{G}$ the k-graph with vertex set $V(\overline{G}) = V(G)$ and edge set $E(\overline{G}) = \binom{V}{k} \setminus E(G)$. This we call the *complement k-graph* of G. Thus $\overline{K_n}$ is the k-graph with vertex set $V(\overline{K_n}) = \{1, 2, \ldots, n\}$ and no edges, $E(\overline{K_n}) = \emptyset$.

We say that two k-graphs G and H are *isomorphic* if $|V(G)| = |V(H)|$ and there is a bijective map $\pi : V(G) \to V(H)$ such that for every $v_1, \ldots, v_k \in V(G)$ we have that $\{v_1, \ldots, v_k\} \in E(G)$ if and only if $\{\pi(v_1), \ldots, \pi(v_k)\} \in E(H)$. Like for graphs, we need this technicality to obtain a commutative semiring, as follows.

We define addition and multiplication operations on the set $\mathcal{R}$ of (non-empty) k-graphs. They are natural in the light of problems like the cap set problem and the sunflower problem, and directly extend the disjoint union and strong product of graphs. Let G and H be k-graphs. The *disjoint union* $G \sqcup H$ is the k-graph with vertex set given by the disjoint union of the vertex sets of G and H and edge set given by the disjoint union of the edge sets of G and H. The *strong product* $G \boxtimes H$

is the k-graph with vertex set given by the cartesian product $V(G) \times V(H)$ and edge set given by all sets $\{(u_1, v_1), (u_2, v_2), \ldots, (u_k, v_k)\} \in \binom{V(G)\times V(H)}{k}$ such that

$$(\{u_1, u_2, \ldots, u_k\} \in E(G) \vee u_1 = u_2 = \cdots = u_k) \wedge (\{v_1, v_2, \ldots, v_k\} \in E(H) \vee v_1 = v_2 = \cdots = v_k).$$

The k-graphs $G \sqcup H$ and $H \sqcup G$ are isomorphic, and the k-graphs $G \boxtimes H$ and $H \boxtimes G$ are isomorphic. We will write $G \cdot H$ for $G \boxtimes H$, and $G + H$ for $G \sqcup H$. Under these operations the set of k-graphs becomes a commutative semiring.

For any k-graph G, a subset $S \subseteq V(G)$ is called an *independent set* in G if no k vertices in S form an edge. The independence number $\alpha(G)$ is the largest number $n \in \mathbb{N}$ such that there is an independent set in G of size n.

At this point we will diverge from the way we treated graphs in Example 2.20. We need to extend a modified notion of a preorder for graphs in order to obtain a Strassen preorder on k-graphs. The modified preorder on graphs still has the independence number as the corresponding subrank, but it has the number of vertices as its rank.

We define the preorder P on $\mathcal{R}$ by saying that $G \leq_P H$ if and only if there is an *injective* map π from $V(G)$ to $V(H)$ such that for every set $\{v_1, v_2, \ldots, v_k\} \in \binom{V}{k} \setminus E(G)$ we have that $\{\pi(v_1), \pi(v_2), \ldots, \pi(v_k)\} \in \binom{V}{k} \setminus E(H)$. It is then readily verified that indeed $\mathrm{R}_P(G) = |V(G)|$ and $\mathrm{Q}_P(G) = \alpha(G)$. It is also readily verified that P is a Strassen preorder.

The asymptotic rank $\widetilde{\mathrm{R}}_P(G)$ simply equals the number of vertices $|V(G)|$ and is thus not of much interest. The asymptotic subrank $\widetilde{\mathrm{Q}}_P(G)$, however, is highly non-trivial and equals the amortized independence number of strong powers of the k-graph G, that is, $\widetilde{\mathrm{Q}}_P(G) = \lim_{n\to\infty} \alpha(G^n)^{1/n}$.

Example 2.24 (Additive combinatorics)**.** We will give one example of a problem in additive combinatorics that corresponds to computing the asymptotic subrank $\widetilde{\mathrm{Q}}_P(G)$ from Example 2.23 for a specific k-graph G. This is the *cap set problem*.

The cap set problem started in arithmetic combinatorics, in the study of 3-term arithmetic progressions in the integers, and seeking bounds on the size of sets which don't have any. The problem in $\mathbb{Z}/3\mathbb{Z}$ (and over other groups) was devised as an analog to gain intuition and for a long time only moderate bounds were known even for this "simpler" problem (see [AD93, Mes95, Tao08]). The problem asks: what is the largest subset $S \subseteq (\mathbb{Z}/3\mathbb{Z})^n$ such that no three different elements $x, y, z \in S$ lie on a line, that is, what is the largest subset $S \subseteq (\mathbb{Z}/3\mathbb{Z})^n$ such that for every three distinct elements $x, y, z \in S$ we have $x + y + z = 0$.[38] Such a set S is called a *cap set*. Let G be the 3-graph with vertex set $V = \{0, 1, 2\}$ and a single edge, $E = \{\{0, 1, 2\}\}$. Then any cap set $S \subseteq (\mathbb{Z}/3\mathbb{Z})^n$ corresponds precisely to an independent set in G^n. Thus $\widetilde{\mathrm{Q}}_P(G)$ determines the asymptotic rate of growth of cap sets and models the cap set problem. The breakthrough of Ellenberg and Gijswijt [EG17] (following the techniques of the closely related result by Croot, Lev and Pach [CLP17]) gave the upper bound $\widetilde{\mathrm{Q}}_P(G) \leq 2.756 < 3$. The best lower bound $\widetilde{\mathrm{Q}}_P(G) \geq 2.216$ was obtained by Edel [Ede04] by lower bounding the subrank of a finite power of G.

2.4. Summary

In this section, we have introduced the kind of objects that the theory of asymptotic spectra is about. The main components are:

- a *commutative semiring* $(\mathcal{R}, +, \cdot)$ with a multiplicative unit 1

[38] It is easily seen that, in $\mathbb{Z}/3\mathbb{Z}$, $x+y+z = 0$ is equivalent to there being u, v such that $(x, y, z) = (u, u+v, u+2v)$.

- a *Strassen preorder* P on $\mathcal{R}$, also denoted by $\leq_P$
- the *rank* function R_P and the *subrank* function Q_P on $\mathcal{R}$ with respect to P
- the *asymptotic rank* $\widetilde{\mathrm{R}}$ and *asymptotic subrank* $\widetilde{\mathrm{Q}}$.

The rank and asymptotic rank are sub-additive, sub-multiplicative, normalized and P-monotone, and the subrank and asymptotic subrank have the same properties but with sub-additive and sub-multiplicative replaced by super-additive and super-multiplicative.

A prototypical example is the semiring of *tensors* under the direct sum and the Kronecker product (Example 2.21). The Strassen preorder is the restriction preorder and the rank function is the usual tensor rank. The asymptotic rank of the matrix multiplication tensor exactly characterizes the asymptotic arithmetic complexity of matrix multiplication.

Another prototypical example is the semiring of *graphs* under the disjoint union and the strong graph product (Example 2.20). The Strassen preorder is the cohomomorphism preorder, the subrank is the independence number, the rank is the clique cover number, and the asymptotic subrank is the Shannon capacity (and the asymptotic rank is the asymptotic clique cover number, which turns out to be equal to the fractional clique cover number).

In the next section we introduce and prove the duality theorem for the setting that we have set up in this section. We will then also return to the main examples that we introduced in this section.

3. Strassen's duality theorem

We will in this section state and prove Strassen's duality theorem. In Section 3.1 we initiate a crucial shift of focus from the class of Strassen preorders to the subclass of *closed* Strassen preorders, and we prove that the latter has very special properties. In Section 3.2 we characterize the closed Strassen preorders as an intersection of *total* closed Strassen preorders. This characterization is a non-trivial extension of the well-known analogous statement for ordinary preorders, and constitutes the main content of the duality theory, albeit in very abstract form. The remainder of the section is dedicated to giving a meaningful description of the duality theory in terms of the more concrete monotone homomorphisms. In Section 3.3 we introduce *relaxations* of rank and subrank, which we call respectively fractional rank and fractional subrank. We then prove their special properties. In Section 3.4 we prove that the fractional rank and fractional subrank defined by total Strassen preorders are equal and give rise to monotone homomorphisms. Using these we define the *asymptotic spectrum*. We then prove Strassen's duality theorem for the asymptotic preorder in terms of this asymptotic spectrum. In Section 3.5 we use the duality theorem for the asymptotic preorder to prove the duality theorems for asymptotic rank and asymptotic subrank. The proof of the aforementioned duality theorems uses a compactness property of the asymptotic spectrum. We defer the proof of this to Section 3.6. (In Part II we will prove and use further, stronger properties than compactness of the asymptotic spectrum, for specific important preorders.) In Section 3.7 we revisit the examples of semirings and preorders that we introduced in Section 2.3 and discuss their asymptotic spectra. Finally, in Section 3.8 we give a summary of the important concepts, notation and results from this section.

3.1. Closed Strassen preorders

Starting with any Strassen preorder P on a semiring $\mathcal{R}$, we defined the rank and subrank, and we have set ourselves the goal of understanding the *asymptotic* rank and *asymptotic* subrank. We will do this by first defining and studying the *asymptotic* preorder associated to P. This is a preorder that contains P and that itself has a natural interpretation in applications. We will prove that the asymptotic preorder is again a Strassen preorder and that taking the asymptotic preorder *again* will not make a difference. Taking the asymptotic preorder should thus be thought of — and this is a simple but helpful change of terminology — as a *closure* operation, and the asymptotic preorder associated to any Strassen preorder as a *closed* Strassen preorder. We will prove that the closed Strassen preorders have strong properties that resemble the properties of the usual ordering of the non-negative reals.

Definition 3.1 (Closure of a Strassen preorder)**.** For any preorder P, and any $a, b \in \mathcal{R}$ we will write $a^n \leq_P b^n 2^{o(n)}$ when there is a function $f : \mathbb{N} \to \mathbb{N}$ such that $f(n) = o(n)$[39] and for all $n \in \mathbb{N}$ it holds that $a^n \leq_P b^n 2^{f(n)}$. We will then also write $(a^n, b^n 2^{o(n)}) \in P$. We define the *closure* $\widetilde{P}$ of P as

$$\widetilde{P} \coloneqq \{(a, b) : (a^n, b^n 2^{o(n)}) \in P\}.$$

In other words, for every $a, b \in \mathcal{R}$ we have $a \leq_{\widetilde{P}} b$ if and only if $a^n \leq_P b^n 2^{o(n)}$. The preorder $\widetilde{P}$ is also called the *asymptotic preorder* associated to P.

In many applications (in particular for continuous functions (Example 2.19), tensors (Example 2.21) and graphs (Example 2.20)), the above Definition 3.1 is equivalent to a slightly different definition in which the term $b^n 2^{o(n)}$ is replaced by $b^{n+o(n)}$. That is, we write $a^n \leq_P b^{n+o(n)}$ when there is a function $f : \mathbb{N} \to \mathbb{N}$ such that $f(n) = o(n)$ and for all $n \in \mathbb{N}$ it holds that $a^n \leq_P b^{n+f(n)}$. We will state and prove this equivalence later for the aforementioned applications.[40]

Remark 3.2. The $o(n)$ slack in the definition of the asymptotic preorder will in particular allow us to swallow subexponential $c^{o(n)}$ or $\mathrm{poly}(n)$ factors on the right-hand side.

Given any two relations P and Q on $\mathcal{R}$ we say that Q is an *extension* of P if and only if $P \subseteq Q$. In other words, Q *extends* P if and only if for every $a, b \in \mathcal{R}$, if $a \leq_P b$, then $a \leq_Q b$.

Lemma 3.3. *The closure $\widetilde{P}$ of a Strassen preorder P is a Strassen preorder that extends P.*

Proof. The only property that is non-trivial to prove is that $a \leq_{\widetilde{P}} b$ implies $a + c \leq_{\widetilde{P}} b + c$. Suppose that $a^n \leq_P b^n 2^{f(n)}$ where $f(n) = o(n)$. We may assume that $f(n)$ is non-decreasing by replacing it

[39] We recall that this means that $f(n)/n \to 0$.

[40] In our proofs of the abstract theory we will use the expression $a^n \leq_P b^n 2^{o(n)}$, since this allows us to have a transparent treatment of the theory and a duality theorem that covers our applications.

with $\max_{m \leq n} f(m)$. Then

$$\begin{aligned}(a+c)^n &= \sum_{i=0}^{n} \binom{n}{i} a^i c^{n-i} \\ &\leq_P \sum_{i=0}^{n} \binom{n}{i} b^i 2^{f(i)} c^{n-i} \\ &\leq_P \sum_{i=0}^{n} \binom{n}{i} b^i c^{n-i} 2^{f(n)} \\ &= (b+c)^n 2^{f(n)}.\end{aligned}$$

This proves the claim. □

Recall that we used Fekete's Lemma (Lemma 2.10) to prove that the asymptotic rank can be understood in two equivalent ways: as an infimum or as a limit (Corollary 2.11), and similarly for the asymptotic subrank. We will now state the analogous statement for the asymptotic preorder, which also follows from Fekete's Lemma (which we leave to the reader). This will give us useful flexibility in the upcoming proofs.

Lemma 3.4 (Consequence of Fekete's lemma)**.** *Let $a, b \in \mathcal{R}$. Then $a^n \leq b^n 2^{o(n)}$ if and only if there is a function $g : \mathbb{N} \to \mathbb{N}$ such that $\inf_n g(n)/n = 0$ and for all $n \in \mathbb{N}$[41] it holds that $a^n \leq b^n 2^{g(n)}$.*

The closure of a Strassen preorder has several special properties that we will discuss now. First we justify the term *closure* by proving that taking the closure is an idempotent operation:

Lemma 3.5. *For any Strassen preorder it holds that $\widetilde{\widetilde{P}} = \widetilde{P}$.*

Proof. We already know that $\widetilde{P} \subseteq \widetilde{\widetilde{P}}$. We need to prove that $\widetilde{\widetilde{P}} \subseteq \widetilde{P}$. Suppose that $(a, b) \in \widetilde{\widetilde{P}}$. Then $a^n \leq_{\widetilde{P}} b^n 2^{f_1(n)}$ where $f_1(n) = o(n)$. This means that

$$a^{nm} \leq_P (b^n 2^{f_1(n)})^m 2^{f_2(n,m)} = b^{nm} 2^{f_1(n)m + f_2(n,m)}$$

where $f_2(n, m) = o(m)$ for every n. We may then select elements $m(n)$ so that

$$\inf_n (f_1(n)m + f_2(n, m(n)))/(nm(n)) = 0.$$

This implies $(a, b) \in \widetilde{P}$ (Lemma 3.4). □

Definition 3.6. We say that P is *closed* if it coincides with its closure: $\widetilde{P} = P$.

In particular, for every Strassen preorder P we have that the closure $\widetilde{P}$ is closed (Lemma 3.5).

Lemma 3.7 (Key properties of a closed Strassen preorder)**.** *Let P be a closed Strassen preorder. Then the following properties hold.*

1. *Multiplicative cancellation property: for every $a, b, c \in \mathcal{R}$ if $ac \leq_P bc$, then $a \leq_P b$.*

[41] For us $\mathbb{N}$ does not contain 0, so $g(n)/n$ cannot cause division by zero.

2. *Additive cancellation property: for every $a, b, c \in \mathcal{R}$, if $a + c \leq_P b + c$, then $a \leq_P b$.*

3. *Gap property: If $a \not\leq_P b$, then for every $c \in \mathcal{R}$ there is an $m \in \mathbb{N}_{\geq 1}$ such that $ma \not\leq_P mb + c$.*

The gap property says that if $a \not\leq_P b$, then in particular, taking $c = 1$, there is an $m \in \mathbb{N}_{\geq 1}$ such that also $ma \not\leq_P mb + 1$. We may intuitively think of the latter as saying that $a \not\leq_P b + \frac{1}{m}$. In other words, the inequality $a \leq_P b$ remains false even if we add a "small fraction" to the right-hand side. That is, we can create a small "gap" in which the inequality remains false. Contrapositively, the gap property says that for any $a, b, c \in \mathcal{R}$, if for all $m \in \mathbb{N}_{\geq 1}$ we have $ma \leq_P mb + c$ (which we may think of as: for all $m \in \mathbb{N}_{\geq 1}$ we have $a \leq_P b + \frac{c}{m}$), then also $a \leq_P b$. That is, a closed Strassen preorder can "absorb infinitesimals".

Proof. Obviously, in order to prove this lemma, we need to make heavy use of the fact that P is closed. Let $a, b, c \in \mathcal{R}$. By the strong Archimedean property (Definition 2.4) we have $1 \leq_P c$, and $c \leq_P r$ for some $r \in \mathbb{N}_{\geq 1}$, and in particular $c \leq_P 2^{o(n)}$. For the multiplicative cancellation property, suppose that $ac \leq_P bc$. Then $a^n \leq_P a^n c \leq_P b^n c \leq_P b^n 2^{o(n)}$.This means that $a \leq_P b$ since P is closed.

The gap property we prove by proving its contrapositive. Suppose that for every $m \in \mathbb{N}_{\geq 1}$ we have $ma \leq_P mb + c$. By the strong Archimedean property, there is a $k \in \mathbb{N}_{\geq 1}$ such that $c \leq_P kb$. Then for every $m \in \mathbb{N}_{\geq 1}$ we have $(km)a \leq_P (km)b + c \leq_P k(m+1)b$ and by multiplicative cancellation $ma \leq_P (m+1)b$. By an inductive argument, we find for every $n \in \mathbb{N}_{\geq 1}$ that

$$a^n \leq_P a^{n-1}b2 \leq_P a^{n-2}b^2 3 \leq_P \cdots \leq_P b^n(n+1).$$

Since $(n+1) \leq 2^{o(n)}$ and P is closed we get $a \leq_P b$.

For the additive cancellation property, suppose that $a + c \leq_P b + c$. Then by induction for every $m \in \mathbb{N}_{\geq 1}$ we have that $ma \leq_P ma + c \leq_P mb + c$. By the gap property we get $a \leq_P b$. □

3.2. Total closed Strassen preorders

Any relation is called *total* if every pair of elements is comparable, that is, a relation P on $\mathcal{R}$ is total if for every $a, b \in \mathcal{R}$ it holds that $(a, b) \in P$ or $(b, a) \in P$ or both. Recall that for any relations P and Q we say that Q *extends* P if $P \subseteq Q$. For for any Strassen preorder P we say that P is *closed* if $\widetilde{P} = P$ with the closure $\widetilde{P}$ defined as in Definition 3.1.

The main theorem of this section (Theorem 3.11) is a characterization of any closed Strassen preorder as the intersection of all its extensions that are total closed Strassen preorders.

This theorem is very similar to a well-known theorem about ordinary preorders, and only slightly harder to prove (given the work that we have already done on Strassen preorders). Let us state this theorem about ordinary preorders and its simple proof first, as to illuminate the proof structure for our theorem about Strassen preorders later. (The experienced reader may safely skip ahead to Theorem 3.11.)

Theorem 3.8. *Every preorder is the intersection of all total preorders that extend it.*

The proof has two parts: a construction of an extension and an argument for the existence of a total extension. To make the second part work, we need to use the concept of the *symmetric part* of a preorder P, which is defined as the set $\{(x, y) \in \mathcal{R}^2 : (x, y) \in P \text{ and } (y, x) \in P\}$. Keeping track of the symmetric part will make sure that we do not add *too many* new relations in our extension, which would render our extensions useless. (Alternatively, in the following discussion, instead of keeping track of the symmetric part, one can turn the preorder into a partial order by working with equivalence classes. We are not doing this to stay close to the later discussion of closed Strassen preorders.)

Lemma 3.9 (One-step extension lemma)**.** *Let P be a preorder on a set $\mathcal{R}$ and let $a, b \in \mathcal{R}$ such that $(a, b) \notin P$. Then there is a preorder Q that extends P, satisfies $(b, a) \in Q$ and has the same symmetric part as P.*

Proof. We start by defining the relation $R = P \cup \{(b, a)\}$. This relation satisfies $(b, a) \in R$ but it may not be transitive. Thus, we let Q be the transitive closure of R, that is, for every $x, y \in \mathcal{R}$ we have $(x, y) \in Q$ if and only if $(x, y) \in P$ or $[(x, b) \in P$ and $(a, y) \in P]$. It remains to verify that the symmetric part of Q equals that of P, but this follows directly from the construction of Q. □

Lemma 3.10 (Total extension lemma)**.** *For every preorder P there is a preorder Q that extends P, is total and has the same symmetric part as P.*

Proof. Let $\mathcal{P}$ be the poset of all preorders Q that extend P and have the same symmetric part as P, with partial order given by the inclusion partial order. Let $\mathcal{C} \subseteq \mathcal{P}$ be a chain. Then the union of (all elements of) $\mathcal{C}$ is again in $\mathcal{P}$ and obviously upper bounds every element in $\mathcal{C}$ in the inclusion partial order. Thus, by Zorn's Lemma, $\mathcal{P}$ contains a maximal element Q. If Q were not total, then we could extend it using Lemma 3.9 and thus Q would not be maximal. Therefore, Q is total and satisfies the claim. □

Proof of Theorem 3.8. Let P be a preorder. We need to prove that P is the intersection of all total preorders that extend it. Clearly P is contained in that intersection. For the other direction, suppose that $(a, b) \notin P$. Then there is a preorder Q extending P such that $(b, a) \in Q$ and with the same symmetric part as P (Lemma 3.9). For this preorder Q we can find a total extension R such that $(b, a) \in R$ and with the same symmetric part as Q (Lemma 3.10). In particular, $(a, b) \notin R$. □

We will now prove the analogous statement for closed Strassen preorders. The proof follows the same pattern as above, but requires adjustments to make sure that all extensions that we obtain are Strassen preorders.

Theorem 3.11 (Representation of a closed Strassen preorder by total closed Strassen extensions)**.** *Let P be a closed Strassen preorder. Then $P = \bigcap_{P \subseteq Q} Q$ where the intersection is over all total closed Strassen preorders Q extending P.*

The proof has two parts again: a construction of an extension and an argument for the existence of a total extension.

Lemma 3.12 (One-step extension lemma)**.** *If P is a closed Strassen preorder such that $(a, b) \notin P$, then there is a Strassen preorder Q that extends P and satisfies $(b, a) \in Q$ and $(a, b) \notin Q$.*

Here we already see a difference with the ordinary one-step extension lemma (Lemma 3.9). Namely, the extension Q that we construct in Lemma 3.12 does not necessarily have the same symmetric part as P, that is, we allow new equivalences to be introduced in the extension. However, it is true that the preorders P and Q have the same symmetric part when restricted to the natural numbers, since P and Q are both Strassen preorders.

Proof of Lemma 3.12. We construct Q in such a manner that most of the required properties hold by construction, namely we set

$$Q \coloneqq \{(x, y) \in \mathcal{R}^2 : \exists s \in \mathcal{R},\ (x + sa, y + sb) \in P\}.$$

This is a semiring preorder by construction, and it clearly extends P. It remains to prove that it is a Strassen preorder. The Archimedean property follows from P being Archimedean: any extension of an Archimedean preorder is also Archimedean. It remains to prove that for every $n, m \in \mathbb{N}_{\geq 1}$ if $n \leq_Q m$, then $n \leq m$ in $\mathbb{N}_{\geq 1}$. Suppose $n \leq_Q m$. Then there exists an element $s \in \mathcal{R}$ such that $n + sa \leq_P m + sb$. Suppose that $n \geq m + 1$. Since P is closed, we can use additive cancellation to get $1 + sa \leq_P sb$, which implies $sa \leq_P sb$. Then we can apply multiplicative cancellation to $sa \leq_P sb$ to get $a \leq_P b$, which contradicts $(a, b) \notin P$. We conclude that $n \leq m$. □

Lemma 3.13 (Total extension lemma)**.** *For every closed Strassen preorder there is a Strassen preorder Q that extends P, is maximal, and thus total and closed.*

The same comment as above applies to Lemma 3.13: contrary to the ordinary total extension lemma (Lemma 3.10), the extension Q that we find in Lemma 3.13 does not necessarily have the same symmetric part as P.

Proof of Lemma 3.13. Let P be a Strassen preorder. Let $\mathcal{P}$ be the poset of all Strassen preorders that extend P, ordered by inclusion. We will apply Zorn's lemma to $\mathcal{P}$. If there is a chain $\mathcal{C} \subseteq \mathcal{P}$, then one verifies that the union of all elements of $\mathcal{C}$ is again in $\mathcal{P}$ and contains all elements of $\mathcal{C}$. Therefore, by Zorn's lemma, $\mathcal{P}$ contains a maximal element, Q. If Q were not total, then we could extend it (Lemma 3.12) and thus Q would not be maximal. If Q were not closed, then we could take the closure (Lemma 3.3) and thus Q would not be maximal. □

We will now prove Theorem 3.11. Since the one-step extension lemma (Lemma 3.12) and total extension lemma (Lemma 3.13) may introduce new equivalences, the proof of Theorem 3.11 needs an extra trick compared to the proof of the ordinary Theorem 3.8 to ensure that if $(x, y) \notin P$, then we can find a Strassen preorder Q extending P that is total and closed and *that still satisfies* $(x, y) \notin Q$. This extra trick is the gap property (Lemma 3.7) that closed Strassen preorders have. Namely, the gap property allows us to boost the assumption $(x, y) \notin P$ to $(nx, ny + 1) \notin P$ before embarking on extending P.

Proof of Theorem 3.11. By Lemma 3.13 there exists at least one total closed Strassen preorder that extends P, so the intersection over all total closed Strassen preorders that extend P is a sensible object. Clearly P is contained in this intersection. For the other direction, suppose that $(x, y) \notin P$. Use the gap property (Lemma 3.7) to make a gap, so $(nx, ny + 1) \notin P$. Extend P to a Strassen preorder Q_1 such that $(ny + 1, nx) \in Q_1$ (Lemma 3.12). Extend Q_1 to a total and closed Strassen preorder Q_2 (Lemma 3.13). Then still $(ny + 1, nx) \in Q_2$ and thus $(x, y) \notin Q_2$, for otherwise $(x, y) \in Q_2$ would imply the statement $ny + 1 \leq_{Q_2} ny$ and thus $ny + 2 \leq_{Q_2} ny + 1$. By additive cancellation this implies the contradictory statement $2 \leq_{Q_2} 1$. This proves the claim. □

We may rephrase Theorem 3.11 as a characterization of the *closure* of any Strassen preorder P in the following straightforward way:

Corollary 3.14. *Let P be a Strassen preorder. Then $\widetilde{P} = \bigcap_{P \subseteq Q} Q$ where the intersection is over all total closed Strassen preorders Q extending P.*

3.3. Fractional rank and fractional subrank

We now return to the study of rank and subrank. We will follow an approach that is common and successful in combinatorial optimization. Namely, we consider the *fractional* relaxation of the (integral) rank and subrank. It turns out that these functions are as well-behaved as rank and subrank. Later (Section 3.4) we will see how they are related to total Strassen preorders, and then use them in the proof of Strassen's duality theorem.

Let $\mathcal{R}$ be a semiring and let P be a Strassen preorder on $\mathcal{R}$. Recall that we have defined, for every $a \in \mathcal{R}$, the rank $\mathrm{R}_P(a)$ as the minimum over all $n \in \mathbb{N}$ such that $a \leq_P n$, and the subrank $\mathrm{Q}_P(a)$ as the maximum over all $n \in \mathbb{N}$ such that $n \leq_P a$. The obvious way to make these parameters fractional is as follows:

Definition 3.15. For every $a \in \mathcal{R}$ the *fractional rank* $\rho_P(a)$ is defined as the infimum over all rational numbers n/m such that $ma \leq_P n$.

Definition 3.16. For every $a \in \mathcal{R}$ the *fractional subrank* $\kappa_P(a)$ is defined as the supremum over all rational numbers n/m such that $n \leq_P ma$.

The definitions of fractional rank and fractional subrank given above are the down-to-earth definitions. However, our intuition (and readability of proofs) is aided a great deal by writing the inequality $ma \leq_P n$ appearing in Definition 3.15 as a *fractional* upper bound $a \leq_P n/m$. We have in fact appealed to this intuition before when discussing the gap property of a closed Strassen preorder (Lemma 3.7): if $a \not\leq_P b$, then for every c there is an $m \in \mathbb{N}$ such that $a \not\leq_P b + \frac{c}{m}$. We will use this fractional notation in some of our proofs.

How do the fractional rank and fractional subrank compare to rank and subrank? We naturally have that $\mathrm{Q}_P(a) \leq \kappa_P(a)$ and $\rho_P(a) \leq \mathrm{R}_P(a)$. Moreover, it follows from transitivity and embedding of the rational numbers that $\kappa_P(a) \leq \rho_P(a)$. Therefore, it holds that

$$\mathrm{Q}_P(a) \leq \kappa_P(a) \leq \rho_P(a) \leq \mathrm{R}_P(a).$$

We will develop a precise understanding of when the middle inequality is an equality in Section 3.4.

It turns out that fractional rank and fractional subrank have the same nice properties as rank and subrank (Lemma 2.7 and Lemma 2.7). Recall that for any function $\phi : \mathcal{R} \to \mathbb{R}_{\geq 1}$ we use the following terminology. We say that ϕ is *sub-additive* if for every $a, b \in \mathcal{R}$ we have $\phi(a + b) \leq \phi(a) + \phi(b)$. We say that ϕ is *sub-multiplicative* if for every $a, b \in \mathcal{R}$ we have $\phi(ab) \leq \phi(a)\phi(b)$. The terms *super-additive* and *super-multiplicative* are defined similarly. We say that ϕ is *normalized* if $\phi(1) = 1$. Finally, we say that ϕ is *P-monotone* if for every $a, b \in \mathcal{R}$, if $a \leq_P b$, then $\phi(a) \leq \phi(b)$.

Lemma 3.17. *The fractional rank ρ_P is sub-additive, sub-multiplicative, normalized and P-monotone.*

Proof. The proof is fairly direct. Let $a, b \in \mathcal{R}$. For sub-multiplicativity and sub-additivity, suppose that $a \leq_P n_a/m_a$ and $b \leq_P n_b/m_b$ where n_a, n_b, m_a, m_b are positive integers. Then $ab \leq_P \frac{n_a}{m_a}\frac{n_b}{m_b}$ as well as $a + b \leq_P \frac{n_a}{m_a} + \frac{n_b}{m_b}$. For P-monotonicity, if $a \leq_P b$ and $b \leq_P n/m$, then also $a \leq_P n/m$. For ρ_P being normalized, one verifies directly that the infimum over n/m such that $1 \leq n/m$ equals 1. $\square$

Lemma 3.18. *The fractional subrank κ_P is super-additive, super-multiplicative, normalized and P-monotone.*

Proof. A similar proof. $\square$

3.4. Monotone homomorphisms, the asymptotic spectrum and duality theorem for asymptotic preorder

We proved that every closed Strassen preorder can be represented as the intersection of its maximal (and thus total and closed) Strassen extensions (Theorem 3.11). We will now discuss the special properties of these maximal Strassen preorders. We will see that each corresponds precisely to a monotone *homomorphism*. To prove this we will make use of our knowledge of the fractional rank and subrank from the previous section. This will be used to define the central notion of the *asymptotic spectrum* (Definition 3.20) of a Strassen preorder. Finally, in the main duality theorem Theorem 3.22 we will see that $\widetilde{P}$ is characterized by its spectrum $\mathcal{X}_P$. As a first application, we will see how to use this duality theorem to characterize "rates of transformation" between semiring elements.

Let P be a Strassen preorder on $\mathcal{R}$. Recall that ρ_P denotes the fractional rank corresponding to P, and that κ_P denotes the fractional subrank corresponding to P. In general, for every $a \in \mathcal{R}$ it holds that the fractional subrank is at most the fractional rank, $\kappa_P(a) \leq \rho_P(a)$.

For any function $\phi : \mathcal{R} \to \mathbb{R}_{\geq 1}$ we will use the following terminology. We say ϕ is *additive* if for every $a, b \in \mathcal{R}$ we have $\phi(a + b) = \phi(a) + \phi(b)$. We say ϕ is *multiplicative* if for every $a, b \in \mathcal{R}$ we have $\phi(ab) = \phi(a)\phi(b)$. We say that ϕ is a *homomorphism* if ϕ is additive, multiplicative and normalized.

Lemma 3.19. *If P is total, then the fractional subrank and fractional rank coincide, that is, for every $a \in \mathcal{R}$ we have $\kappa_P(a) = \rho_P(a)$, and thus ρ_P is a P-monotone homomorphism.*

Proof. Suppose that there is an element $a \in \mathcal{R}$ such that $\kappa_P(a) < \rho_P(a)$. Then there is a rational number in between: $\kappa_P(a) < n/m < \rho_P(a)$. Then by the definition of ρ_P we have that *not* $a \leq_P n/m$. Thus, since P is total, we have $n/m \leq_P a$. However, this implies by the definition of κ_P that $n/m \leq \kappa_P(a)$, which contradicts our assumption. Finally, from $\kappa_P = \rho_P$ it follows using Lemma 3.17 and Lemma 3.18 that ρ_P is a homomorphism. ☐

We are about to define the central object of the theory, the asymptotic spectrum. This object will serve as the dual space in all our duality theorems. Its definition is simple, but explicit construction will in many cases be hard.

Definition 3.20. Let $\mathcal{R}$ be a semiring with a Strassen preorder P. The *asymptotic spectrum* is defined as the set of all P-monotone homomorphisms $\phi : \mathcal{R} \to \mathbb{R}_{\geq 1}$. We will denote the asymptotic spectrum by $\mathcal{X}_P$. We refer to the elements of $\mathcal{X}_p$ as *spectral points.*[42]

The asymptotic spectrum is in principle an infinite-dimensional object. Often it is useful to consider the following finite-dimensional rendering.

Definition 3.21. Let $a_1, \ldots, a_k \in \mathcal{R}$. The *asymptotic spectrum of* $a_1, \ldots, a_k$ we define as the set of simultaneous evaluations of the spectral points at the given semiring elements:

$$\mathcal{X}_P(a_1, \ldots, a_k) \coloneqq \{(\phi(a_1), \ldots, \phi(a_k)) : \phi \in \mathcal{X}_P\} \subseteq \mathbb{R}^k_{\geq 1}.$$

[42]The terminology "spectrum" and "spectral" goes back to Strassen [Str88] and to older terminology used in real algebra for objects that play a similar role as the spectrum of a ring (the set of all prime ideals of the ring) in commutative algebra [Mar08].

Studying the finite-dimensional $\mathcal{X}(a_1, \dots, a_k)$ is in particular natural when trying to understand specific semiring elements a_i or when the semiring of interest is finitely generated (by the a_i). We note that $\mathcal{X}(a_1, \dots, a_k)$ and $\mathcal{X}(b_1, \dots, b_k)$ may be different even if the semiring generated by the a_i and the semiring generated by the b_i are the same. The sets $\mathcal{X}(a_1, \dots, a_k)$ play an important role in Part II and Part III.

Using Lemma 3.19 we can now present the duality theorem for closed Strassen preorders in terms of the asymptotic spectrum.

Theorem 3.22 (Duality theorem for the asymptotic preorder). *Let P be a Strassen preorder on $\mathcal{R}$. Let $a, b \in \mathcal{R}$. We have $a \leq_{\widetilde{P}} b$ if and only if for every $\phi \in \mathcal{X}_P$ it holds that $\phi(a) \leq \phi(b)$.*

Proof. For the non-trivial direction, suppose that not $a \leq_{\widetilde{P}} b$. Then there is a maximal Strassen extension Q of $\widetilde{P}$ such that $a \not\leq_Q b$ (Theorem 3.11). The preorder Q is closed (by maximality) and thus by the gap property, there is an $n \in \mathbb{N}$ such that $a \not\leq_Q b + \frac{1}{n}$ (Lemma 3.7). Then $b + \frac{1}{n} \leq_Q a$, since Q is total (by maximality). The function ρ_Q is a Q-monotone homomorphism (Lemma 3.19), and thus P-monotone, and thus ρ_Q is an element of $\mathcal{X}_P$. We find that $\rho_Q(b) < \rho_Q(b) + \frac{1}{n} \leq \rho_Q(a)$. □

As a corollary of Theorem 3.22 we prove a duality theorem for the "asymptotic relative cost" of one semiring element relative to another semiring element. Here the goal is to understand what is the smallest real number r such that $a^n \leq_P b^{\lceil rn \rceil} 2^{o(n)}$, as n goes to infinity, so that r indicates how much larger a power of b we need to take to "create" a large power of a. In other words, r describes the "asymptotic relative cost" of a in terms of b. (One can also naturally define the "asymptotic relative value" of b in terms of a as the supremum of all r so that $a^{\lfloor rn \rfloor} \leq_P b^n 2^{o(n)}$. This turns out to give the reciprocal of the asymptotic relative value, which we leave to the reader.)

Corollary 3.23. *Let $a, b \in \mathcal{R}$. Then the asymptotic relative cost of a in terms of b,*

$$\inf\{r : a^n \leq_P b^{\lceil rn \rceil} 2^{o(n)}\},$$

equals the maximization

$$\max_{\phi \in \mathcal{X}_P} \frac{\log \phi(a)}{\log \phi(b)}.$$

For the proof of Corollary 3.23 we will use the following compactness property of the asymptotic spectrum, which we will greatly extend and prove in Section 3.6.[43]

Lemma 3.24 (Lemma 3.33). *For every $a \in \mathcal{R}$ the set $\{\phi(a) : \phi \in \mathcal{X}_P\}$ is a compact subset of $\mathbb{R}_{\geq 1}$ with the Euclidean topology.*

Proof of Corollary 3.23. Suppose that $a^n \leq_P b^{\lceil rn \rceil} 2^{o(n)}$. Let $\phi \in \mathcal{X}_P$. Applying ϕ to both sides of the inequality and using multiplicativity gives $\phi(a)^n \leq \phi(b)^{\lceil rn \rceil} 2^{o(n)}$. After taking the nth root and letting n go to infinity, we obtain $\phi(a) \leq \phi(b)^r$, and so $\log \phi(a) / \log \phi(b) \leq r$. We conclude that

$$\sup_{\phi \in \mathcal{X}_P} \frac{\log \phi(a)}{\log \phi(b)} \leq \inf\{r : a^n \leq_P b^{\lceil rn \rceil} 2^{o(n)}\}.$$

To prove the other direction, if for all $\phi \in \mathcal{X}_P$ we have $\log \phi(a) / \log \phi(b) \leq r$, then $\phi(a^n) \leq \phi(b^{\lceil rn \rceil})$ for all $n \in \mathbb{N}$ and thus $a^n \leq_{\widetilde{P}} b^{\lceil rn \rceil}$ (Theorem 3.22). From this follows $a^{nm} \leq_P (b^{\lceil rn \rceil})^m 2^{o(m)}$ and so

[43] Without the compactness property we can already obtain Corollary 3.23 with max replaced by sup.

$a^n \leq_P b^{\lceil rn \rceil} 2^{o(n)}$. We conclude that

$$\sup_{\phi \in \mathcal{X}_P} \frac{\log \phi(a)}{\log \phi(b)} \geq \inf\{r : a^n \leq_P b^{\lceil rn \rceil} 2^{o(n)}\}.$$

Finally, since $\{(\phi(a), \phi(b)) : \phi \in \mathcal{X}_P\}$ is compact (Lemma 3.24), the supremum is attained. □

3.5. Duality theorem for asymptotic rank and asymptotic subrank

Recall that for $a \in \mathcal{R}$ the asymptotic rank is defined as $\widetilde{\mathrm{R}}(a) = \lim_{n\to\infty} \mathrm{R}(a^n)^{1/n}$ and the asymptotic subrank as $\widetilde{\mathrm{Q}}(a) = \lim_{n\to\infty} \mathrm{Q}(a^n)^{1/n}$. We have seen in Section 2.3 that these parameters correspond to important problems in many applications. In this section we will use the duality theorem Theorem 3.22 to prove duality theorems for asymptotic rank and asymptotic subrank.

We know from Fekete's lemma (Lemma 2.10) that the limits in the definition of asymptotic rank and asymptotic subrank can be replaced by the infimum and the supremum, respectively, so $\widetilde{\mathrm{R}}(a) = \inf_n \mathrm{R}(a^n)^{1/n}$ and $\widetilde{\mathrm{Q}}(a) = \sup_n \mathrm{Q}(a^n)^{1/n}$. We may thus think of the asymptotic rank as a minimization problem and the asymptotic subrank as a maximization problem. The duality theorems we will prove characterize asymptotic rank as a maximization, and the asymptotic subrank as a minimization. For the latter we need to assume a mild condition on the semiring element, which we will call "gapped". This condition is satisfied by all our examples.

Let us deal with the asymptotic rank first. We will deal with the asymptotic subrank afterwards, which will require a bit more work. Before moving on we note this simple rephrasing in terms of the preorder P: the asymptotic rank $\widetilde{\mathrm{R}}(a)$ equals the infimum over $r \in \mathbb{R}_{\geq 1}$ such that there is a sequence of numbers $f(n) \in \mathbb{N}$ with $\inf_n f(n)^{1/n} = r$ and $a^n \leq_P f(n)$. In the following theorem we characterize this number in terms of the asymptotic spectrum. The first two expressions we have already discussed here. The third is a characterization as a fractional expression, which we include to show the fractional nature of asymptotic rank, but which is not crucial late. The fourth is the important characterization in terms of the asymptotic spectrum $\mathcal{X}$.

Theorem 3.25. *For any $a \in \mathcal{R}$, the following values are equal:*

(i) $\widetilde{\mathrm{R}}(a)$

(ii) $\inf\{r \in \mathbb{R}_{\geq 1} : a^n \leq_P f(n),\ f(n) \in \mathbb{N},\ \inf_n f(n)^{1/n} = r\}$

(iii) $\inf\{\frac{n}{m} \in \mathbb{Q}_{\geq 1} : ma \leq_{\widetilde{P}} n\}$

(iv) $\sup\{\phi(a) : \phi \in \mathcal{X}\}$.

Proof. We have already observed that (i) = (ii) holds directly from the definition.

(iii) $\geq$ (iv). Suppose $ma \leq_{\widetilde{P}} n$. Let $\phi \in \mathcal{X}$. Then applying ϕ to both sides and rearranging gives the inequality $\phi(a) \leq \frac{n}{m}$.

(iv) $\geq$ (iii). Suppose (iii) $>$ (iv). Let $\frac{n}{m}$ be a value that is strictly in between. Then for every $\phi \in \mathcal{X}$ we have $\phi(a) < \frac{n}{m}$ so $m\phi(a) < n$, and thus $ma \leq_{\widetilde{P}} n$ by Theorem 3.22, which is a contradiction with our choice of r.

(ii) $\geq$ (iv). Let $r \in \mathbb{R}_{\geq 1}$ be any value satisfying the conditions in (ii), so that $a^n \leq_P f(n)$ for some numbers $f(n) \in \mathbb{N}$ with $\inf_n f(n)^{1/n} = r$. Let $\phi \in \mathcal{X}$. Apply it to both sides, take nth roots and take the infimum over n to get $\phi(a) \leq r$.

(iv) $\geq$ (ii). Suppose (ii) $>$ (iv). Let $r \in \mathbb{R}_{\geq 1}$ be a value that is strictly in between. Then for every $\phi \in \mathcal{X}$ we have $\phi(a) < r$ so $\phi(a^n) \leq \lceil r^n \rceil$ and so $a^n \leq_{\widetilde{P}} \lceil r^n \rceil$ (Theorem 3.22). Then $a^{nm} \leq_P \lceil r^n \rceil^m 2^{o(m)} = \lceil r^n \rceil^{m+o(m)}$, so we find $a^\ell \leq_P f(\ell)$ for some function $f(\ell) \in \mathbb{N}$ with $\inf_\ell f(\ell)^{1/\ell} = r$, which is a contradiction with our choice of r. □

Now we deal with the asymptotic subrank. From Fekete's lemma (Lemma 2.10) we know that the asymptotic subrank $\widetilde{\mathrm{Q}}(a)$ equals the supremum over $r \in \mathbb{R}_{\geq 1}$ such that there is a sequence of numbers $f(n) \in \mathbb{N}$ with $\sup_n f(n)^{1/n} = r$ and $f(n) \leq_P a^n$. Note that replacing $f(n) \leq_P a^n$ by $f(n) \leq_P a^{n+o(n)}$ does not change this number. We want to characterize this number in terms of the asymptotic spectrum, just like we characterized the asymptotic rank above. However, as we will see, the natural approach will at first lead to the characterization of a slightly different parameter, namely the supremum over $r \in \mathbb{R}_{\geq 1}$ such that there are sequences of numbers $f(n), g(n) \in \mathbb{N}$ with $\sup_n (f(n)/g(n))^{1/n} = r$ and $f(n) \leq_P a^n g(n)$. (Here we imagine $g(n) = 2^{o(n)}$ and $\lim_{n\to\infty} f(n)^{1/n} = r$.) That is, a potentially more powerful slack of $g(n) \in \mathbb{N}$ is allowed on the right-hand side rather than a slack of the form $a^{o(n)}$. (Note that if there is a number $k \in \mathbb{N}$ such that $a^k \geq_P 2$, then the two notions coincide, however there may be elements a for which no such k exists.) We will prove this characterization first, and then provide a sufficient condition for this number to be equal to the asymptotic subrank $\widetilde{\mathrm{Q}}(a)$.

Theorem 3.26. *For any $a \in \mathcal{R}$, the following values are equal:*

(i) $\sup\{r \in \mathbb{R}_{\geq 1} : f(n), g(n) \in \mathbb{N},\ \sup_n (f(n)/g(n))^{1/n} = r,\ f(n) \leq_P a^n g(n)\}$

(ii) $\sup\{\frac{n}{m} \in \mathbb{Q}_{\geq 1} : n \leq_{\widetilde{P}} ma\}$

(iii) $\inf\{\phi(a) : \phi \in \mathcal{X}\}$.

Proof. (ii) $\geq$ (iii). Suppose $n \leq_{\widetilde{P}} ma$. Let $\phi \in \mathcal{X}$. Then applying ϕ to both sides and rearranging gives the inequality $\frac{n}{m} \leq \phi$.

(iii) $\geq$ (ii). Suppose (ii) $>$ (iii). Let $\frac{n}{m}$ be a value that is strictly in between. Then for every $\phi \in \mathcal{X}$ we have $\phi(a) > \frac{n}{m}$ so $m\phi(a) > n$, and thus $ma \geq_{\widetilde{P}} n$ by Theorem 3.22, which is a contradiction with our choice of r.

(i) $\geq$ (iii). Let $r \in \mathbb{R}_{\geq 1}$ be any value satisfying the conditions in (i), so that $f(n) \leq_P a^n g(n)$ for some numbers $f(n), g(n) \in \mathbb{N}$ with $\sup_n (f(n)/g(n))^{1/n} = r$. Let $\phi \in \mathcal{X}$. Apply it to both sides, take nth roots and take the supremum over n to get $r \leq \phi(a)$.

(iii) $\geq$ (i). Suppose (i) $>$ (iii). Let $r \in \mathbb{R}_{\geq 1}$ be a value that is strictly in between. Then for every $\phi \in \mathcal{X}$ we have $r < \phi(a)$ so $\lfloor r^n \rfloor \leq \phi(a^n)$ and so $\lfloor r^n \rfloor \leq_{\widetilde{P}} a^n$ (Theorem 3.22). Then $\lfloor r^n \rfloor^m \leq_P a^{nm} 2^{o(m)}$, so we find $f(\ell) \leq_P a^\ell g(\ell)$ for some functions $f(\ell), g(\ell) \in \mathbb{N}$ with $\sup_\ell f(\ell)^{1/\ell} = r$ and $\inf_n g(\ell)^{1/\ell} = 1$, so that $\inf_\ell (f(\ell)/g(\ell))^{1/\ell} = r$, which is a contradiction with our choice of r. □

As we mentioned, the value described in Theorem 3.26 may be larger than the asymptotic subrank. To link it to the asymptotic subrank, we introduce the following sufficient condition.

Definition 3.27. We say an element $a \in \mathcal{R}$ is *gapped* if there is a number $k \in \mathbb{N}$ such that $a^k \geq_P 2$, or there is a map $\phi \in \mathcal{X}$ such that $\phi(a) = 1$. We call $(\mathcal{R}, P)$ gapped if all elements of $\mathcal{R}$ are gapped.

Note that if $a \in \mathcal{R}$ is gapped, and if there is no number $k \in \mathbb{N}$ such that $a^k \geq_P 2$, then we are only requiring that *there is* a map $\phi \in \mathcal{X}$ such that $\phi(a) = 1$. In particular, we are not requiring that for all $\phi \in \mathcal{X}$ we have $\phi(a) = 1$ (that is, that a is equivalent to 1 under $\widetilde{P}$), or that $a = 1$. Indeed, requiring that either $a^k \geq_P 2$ or $a = 1$, will be too strong for our applications.[44]

Lemma 3.28. *Let $a \in \mathcal{R}$ be gapped. Then the following are equal:*

(i) $\sup\{r \in \mathbb{R}_{\geq 1} : f(n), g(n) \in \mathbb{N}, \sup_{n\geq 1}(f(n)/g(n))^{1/n} = r,\ f(n) \leq_P a^n g(n)\}$

(ii) $\sup\{r \in \mathbb{R}_{\geq 1} : f(n) \in \mathbb{N}, \sup_{n\geq 1} f(n)^{1/n} = r,\ f(n) \leq_P a^n\}$

Proof. The inequality (ii) $\leq$ (i) is clear. Suppose $f(n) \leq_P a^n g(n)$. If there is a number $k \in \mathbb{N}$ such that $2 \leq_P a^k$, then $f(n) \leq_P a^{n+k}$. Since $\sup_n f(n)^{1/(n+k)} = r$ we have (i) $\leq$ (ii). Otherwise, there is a map $\phi \in \mathcal{X}$ such that $\phi(a) = 1$. Then $f(n) \leq g(n)$ so $r = 1$. Then also (i) $\leq$ (ii). $\square$

Combining Theorem 3.26 and Lemma 3.28 we directly get:

Corollary 3.29. *Let $a \in \mathcal{R}$ be gapped. Then $\widetilde{\mathrm{Q}}(a)$ is equal to the (equal) values (i), (ii) and (iii) in Theorem 3.26, and in particular $\widetilde{\mathrm{Q}}(a) = \inf_{\phi\in\mathcal{X}} \phi(a)$.*

Finally, using the compactness lemma Lemma 3.24 (which we will prove in the next section), we obtain as a corollary of the above Theorem 3.26 and Corollary 3.29 the following characterizations of asymptotic rank and asymptotic subrank (with the only difference from before that infimum is replaced by minimum and supremum is replaced by maximum):

Corollary 3.30. *Let $a \in \mathcal{R}$. Then $\widetilde{\mathrm{R}}(a) = \max_{\phi\in\mathcal{X}} \phi(a)$.*

Corollary 3.31. *Let $a \in \mathcal{R}$ be gapped. Then $\widetilde{\mathrm{Q}}(a) = \min_{\phi\in\mathcal{X}} \phi(a)$.*

3.6. Compactness

In the previous subsections we have seen how the asymptotic spectrum characterizes the asymptotic preorder. We then used this characterization to characterize asymptotic relative cost, asymptotic rank and asymptotic subrank. These proofs relied on a compactness property of the asymptotic spectrum to argue that the optimizations were attained. Compactness will play an important role in applications of the duality theory that we will discuss in Section 4. We will now discuss this compactness property in detail (and in particular the topology that we put on the asymptotic spectrum) and give its proof. After that we briefly discuss another basic topological property of the asymptotic spectrum that we will need later, namely that it is Hausdorff. Let P be a Strassen preorder on the semiring $\mathcal{R}$. Let $\mathcal{X}$ be the asymptotic spectrum.

The simplest version of the compactness property that we will prove is that for every element $a \in \mathcal{R}$ the set $\{\phi(a) : \phi \in \mathcal{X}\}$ is a compact subset of $\mathbb{R}_{\geq 1}$ with the Euclidean topology. The set is clearly bounded since $1 \leq \phi(a) \leq \mathrm{R}(a)$ for every $\phi \in \mathcal{X}$, where $\mathrm{R}(a)$ is the rank, so it remains to prove that it is closed. Indeed we have been using this property in the previous sections to argue that optimizations over $\mathcal{X}$ are attained, for instance $\inf_{\phi\in\mathcal{X}} \phi(a) = \min_{\phi\in\mathcal{X}} \phi(a)$. A higher-dimensional version of the compactness property that is true is that for any elements $a_1, \ldots, a_k \in \mathcal{R}$ the set $\mathcal{X}(a_1, \ldots, a_k) \coloneqq \{(\phi(a_1), \ldots, \phi(a_k)) : \phi \in \mathcal{X}\}$ is a compact subset of $\mathbb{R}^k_{\geq 1}$. Again, boundedness is easy to see using rank, and it remains to prove closedness.

[44]The reason for needing a map $\phi \in \mathcal{X}$ such that $\phi(a) = 1$ is to prevent a situation where say $4 \leq_P 2a^\ell$ for some $\ell \in \mathbb{N}$ while not $2 \leq_P a^k$ for any $k \in \mathbb{N}$, as will get clear in the proof of the following lemma. Note that we are not assuming P to be closed here.

We will prove the above as a special case of the following infinite-dimensional version of the compactness property. First we specify the topology on the asymptotic spectrum that we use here and in the rest of the paper. For every semiring element $a \in \mathcal{R}$ there is an *evaluation function*

$$\hat{a} : \mathcal{X} \to \mathbb{R}_{\geq 1} : \phi \mapsto \phi(a).$$

We endow $\mathbb{R}_{\geq 1}$ with the Euclidean topology. We endow the asymptotic spectrum $\mathcal{X}$ with the weakest topology such that for every $a \in \mathcal{R}$ the function $\hat{a}$ is continuous, which means precisely that we give $\mathcal{X}$ the topology generated by the sets $\{\hat{a}^{-1}(U) : a \in \mathcal{R},\, U \subseteq \mathbb{R}_{\geq 1} \text{ open}\}$. This is the same as the topology induced by the embedding $\mathcal{X} \subseteq \mathbb{R}_{\geq 1}^{\mathcal{R}}$ with the product topology on $\mathbb{R}_{\geq 1}^{\mathcal{R}}$. We will prove that $\mathcal{X}$ with this topology is compact.

Lemma 3.32. *The asymptotic spectrum $\mathcal{X}$ is compact.*

Proof. Define the set $Y = \prod_{a \in \mathcal{R}} [1, \mathrm{R}_P(a)]$ with $\mathrm{R}_P(a)$ the rank of a. Then $\mathcal{X} \subseteq Y \subseteq \mathbb{R}_{\geq 1}^{\mathcal{R}}$ because for every $\phi \in \mathcal{X}$ and $a \in \mathcal{R}$ we have $1 \leq \phi(a) \leq \mathrm{R}_P(a)$. Since $[1, \mathrm{R}_P(a)]$ is compact for every a, the set Y is compact by the Tychonoff theorem [Mun00, Chapter 5]. Every closed subset of a compact set is compact, so it remains to show that $\mathcal{X}$ is closed.

To prove that $\mathcal{X}$ is closed, we will write $\mathcal{X}$ as an intersection of closed sets. These closed sets are given by the defining properties of a monotone homomorphism: additivity, multiplicativity, monotonicity and normalization. Namely, we may write $\mathcal{X} = Z_1 \cap Z_2 \cap Z_3 \cap Z_4$ where

$$\begin{aligned}
Z_1 &= \{(\phi(a))_{a \in \mathcal{R}} : \phi : \mathcal{R} \to \mathbb{R}_{\geq 1} \text{ is additive}\} \\
Z_2 &= \{(\phi(a))_{a \in \mathcal{R}} : \phi : \mathcal{R} \to \mathbb{R}_{\geq 1} \text{ is multiplicative}\} \\
Z_3 &= \{(\phi(a))_{a \in \mathcal{R}} : \phi : \mathcal{R} \to \mathbb{R}_{\geq 1} \text{ is } P\text{-monotone}\} \\
Z_4 &= \{(\phi(a))_{a \in \mathcal{R}} : \phi : \mathcal{R} \to \mathbb{R}_{\geq 1} \text{ is normalized}\}.
\end{aligned}$$

It remains to show for each i that the set Z_i is closed. This we do by showing that Z_i is the inverse image of a closed set under a continuous map. By construction of the topology on $\mathbb{R}_{\geq 1}^{\mathcal{R}}$ we know that for every $a \in \mathcal{R}$ the function $\widehat{a} : \mathbb{R}_{\geq 1}^{\mathcal{R}} \to \mathbb{R}_{\geq 1} : \phi \mapsto \phi(a)$ is continuous, and hence so is any polynomial combination of functions $\widehat{a_i}$ for any collection $a_1, \ldots, a_n \in \mathcal{R}$.

In particular, here is the proof that Z_1 is closed. For every $a, b \in \mathcal{R}$ we define the function $f_{a,b} = \widehat{a} + \widehat{b} - \widehat{(a+b)}$, and we note that $f_{a,b}$ is continuous since it is a polynomial combination of the continuous functions $\widehat{a}$, $\widehat{b}$ and $\widehat{(a+b)}$. Then the inverse image $f_{a,b}^{-1}(\{0\})$ is a closed set, since $\{0\}$ is closed. Then also the intersection $\bigcap_{a,b \in \mathcal{R}} f_{a,b}^{-1}(\{0\})$ is a closed set. This intersection is precisely equal to Z_1, so Z_1 is closed. The proofs that the other sets Z_i are closed are along the same lines and are left to the reader. ☐

As a consequence we get:

Lemma 3.33. *For every $a_1, \ldots, a_k \in \mathcal{R}$ the subset $\mathcal{X}(a_1, \ldots, a_k) \coloneqq \{(\phi(a_1), \ldots, \phi(a_k)) : \phi \in \mathcal{X}\}$ of $\mathbb{R}_{\geq 1}^k$ endowed with the Euclidean topology is compact.*

Proof. Let us first prove that for every $a \in \mathcal{R}$ the set $\mathcal{X}(a) = \{\phi(a) : \phi \in \mathcal{X}\}$ is compact. The set $\{\phi(a) : \phi \in \mathcal{X}(a)\}$ is the image of $\mathcal{X}$ under $\widehat{a}$. The set $\mathcal{X}$ is compact as we have shown and the map $\widehat{a}$ is continuous by definition of the topology on $\mathcal{X}$. Thus $\{\phi(a) : \phi \in \mathcal{X}\}$ is compact. To see that $\mathcal{X}(a_1, \ldots, a_k)$ is compact, we use that every $\hat{a}_i$ is continuous and thus $(\hat{a}_1, \ldots, \hat{a}_k) : \mathcal{X} \to \mathbb{R}_{\geq 1}^k$ is continuous, and thus again its image is compact, which proves the claim. ☐

A topological space $\mathcal{X}$ is called *Hausdorff* if for every two distinct $\phi, \psi \in \mathcal{X}$ there are disjoint open neighborhoods $\phi \in U$, $\psi \in V$, $U \cap V = \emptyset$. The asymptotic spectrum has this property (by an almost trivial proof).

Lemma 3.34. *The asymptotic spectrum $\mathcal{X}$ is Hausdorff.*

Proof. Let $\phi, \psi \in \mathcal{X}$ be distinct, then there is an $a \in \mathcal{R}$ with $\phi(a) \neq \psi(a)$, so $\hat{a}(\phi) \neq \hat{a}(\psi)$. Then there are disjoint opens U' and V' containing $\hat{a}(\phi)$ and $\hat{a}(\psi)$ respectively, and then $U = \hat{a}^{-1}(U')$ and $V = \hat{a}^{-1}(V')$ satisfy the claim. □

3.7. Examples

We return to the main examples of semirings and preorders of Section 2.3 and discuss their asymptotic spectra.

Example 3.35 (Continuous functions)**.** Continuing Example 2.19, let $\mathcal{R} = C(\mathcal{X}, \mathbb{R}_{\geq 1})$ be the semiring of continuous functions to $\mathbb{R}_{\geq 1}$ on some compact set $\mathcal{X}$, with pointwise addition, multiplication and preorder P. Then we find that the asymptotic spectrum $\mathcal{X}_P$ is naturally related to the set $\mathcal{X}$. Namely the P-monotone homomorphisms are precisely the evaluations maps $\hat{s} : f \mapsto f(s)$ for $s \in \mathcal{X}$.

Example 3.36 (Matrices)**.** We already remarked in Example 2.18 that the semiring $\mathcal{R}$ of nonzero matrices under direct sum, tensor product and restriction preorder P is essentially the semiring of natural numbers in disguise. In the language of the asymptotic spectrum, another way of saying this is that the asymptotic spectrum $\mathcal{X}_P$ consists of a single element, which is the matrix rank.

Example 3.37 (Tensors)**.** We continue Example 2.21. Let $\mathcal{R}$ be the semiring of nonzero 3-tensors under direct sum, Kronecker product and restriction preorder P. Contrary to Example 3.35 and Example 3.36, here the asymptotic spectrum $\mathcal{X}_P$ is not fully understood. There are three elements in $\mathcal{X}_P$ that are easy to find. They are the flattening ranks $R_{(1)}$, $R_{(2)}$ and $R_{(3)}$ obtained by flattening the 3-tensor to a matrix (in one of the three ways) and computing the matrix rank. (In Definition 12.6 we will give the precise and general definition of flattening ranks.) Importantly it is known that the asymptotic rank and asymptotic subrank are not elements of $\mathcal{X}_P$.

For the subsemiring $\mathcal{R}_o \subseteq \mathcal{R}$ of oblique tensors, which are tensors such that in some basis the support is an antichain, Strassen [Str91] constructed a continuous family of elements in the corresponding asymptotic spectrum (for the same restriction preorder P), called the support functionals. We will not discuss these in depth here but refer to Section 9.3 for more. These support functionals are able to recover the results in cap sets discussed in Example 2.24 and are closely related to the slice rank [CVZ18, CLZ23]. It is not known whether the support functions are all the elements in the asymptotic spectrum of oblique tensors.

Christandl, Vrana and Zuiddam [CVZ18] constructed a continuous family of elements in the asymptotic spectrum of all complex 3-tensors, called the quantum functionals. We will say more about this in Section 9.3. Also here it is not known whether the quantum functionals are all the elements of the asymptotic spectrum of complex 3-tensors.

Example 3.38 (Graphs)**.** The semiring of non-empty graphs with the disjoint union, strong product and cohomomorphism preorder (Example 2.20) is another example where the asymptotic spectrum $\mathcal{X}_P$ is not known. However, some well-known graph parameters are in $\mathcal{X}_P$, including the Lovász theta function ϑ [Lov79], the fractional Haemers bound [BC19], the fractional orthogonal rank and the fractional clique cover number. The fractional clique cover number in fact turns out to be equal to the asymptotic clique cover number, which is the asymptotic rank in this setting. Thus the asymptotic rank is in $\mathcal{X}_P$ in this setting. The asymptotic subrank, however, which is the Shannon capacity Θ, is known not to be in $\mathcal{X}_P$. We refer to [Zui19, Vra21] for more and will return to this example briefly in Section 9.3.

Recall that the duality theorem for the asymptotic subrank $\widetilde{\mathrm{Q}}(a)$ (Corollary 3.31) relied on the element a being gapped (Definition 3.27), meaning that either there is a number $k \in \mathbb{N}$ such that $a^k \geq_P 2$ or there is a map $\phi \in \mathcal{X}$ such that $\phi(a) = 1$. We mentioned that all our main examples have this property, which we prove here.

Theorem 3.39. *In each of the above examples, every element $a \in \mathcal{R}$ is gapped.*

Proof. For our purposes, we can identify the semiring of matrices (Example 3.36) with the natural numbers $\mathbb{N}$. For every $a \in \mathbb{N}$ either $a^k \geq_P 2$ for some $k \in \mathbb{N}$ or $a = 1$. In the latter case $\phi(a) = 1$ for $\phi \in \mathcal{X}$ (there is only one element in $\mathcal{X}$).

For continuous functions (Example 3.35), either $f^k \geq_P 2$ for some k or there is an s such that $f(s) = 1$ and so for $\hat{s} \in \mathcal{X}$ we have $\hat{s}(f) = 1$.

For graphs (Example 3.38), either G is not a clique in which case $G \geq_P 2$, or G is a clique in which case the fractional clique cover number (which is an element of $\mathcal{X}$) equals 1.

Finally, for tensors (Example 3.37), it is not hard to see with a direction construction that either there is a flattening rank of one, or $T^2 \geq_P 2$. □

From Theorem 3.39 it follows that in the above examples we have (roughly speaking) that inequalities of the form $f(n) \leq_P a^n 2^{o(n)}$ are equivalent to $f(n) \leq_P a^n$ (Lemma 3.28). For our examples, we can prove something slightly stronger: In Section 3.1 we have defined the asymptotic preorder $\widetilde{P}$ of a Strassen preorder P by saying for every $a, b \in \mathcal{R}$ that $a \leq_{\widetilde{P}} b$ if and only if $a^n \leq_P b^n 2^{o(n)}$ (Definition 3.1). We mentioned that in our main examples the asymptotic preorder can equivalently be characterized by replacing $b^n 2^{o(n)}$ with $b^{n+o(n)}$, so that it is phrased purely in terms of powers of a and b. We state this equivalence here, and leave the proofs, which are quite similar to the proof of the above theorem, to the reader.

Theorem 3.40. *In each of the above examples, for every $a, b \in \mathcal{R}$ we have $a^n \leq_P b^n 2^{o(n)}$ if and only if $a^n \leq_P b^{n+o(n)}$.*

Note that the implication $a^n \leq_P b^{n+o(n)} \Rightarrow a^n \leq_P b^n 2^{o(n)}$ follows from the fact that every element $b \in \mathcal{R}$ is bounded from above by some natural number (Archimedean property of the Strassen preorder P), so it remains to prove the reverse implication ($\Leftarrow$).

3.8. Summary

We summarize the section, in particular the duality theorem, and discuss an insightful rephrasing that relates any semiring with Strassen preorder to the semiring of continuous functions.

The setup for the duality theorem is the one that we introduced in Section 2 which we now summarize. We let $\mathcal{R}$ be a commutative semiring with a Strassen preorder P. We recall that a preorder P on $\mathcal{R}$ is a Strassen preorder if (1) for all natural numbers $n, m \in \mathbb{N}_{\geq 1}$ we have $n \leq m$ if and only if $n \leq_P m$, (2) for every $a, b, c, d \in \mathcal{R}$, if $a \leq_P b$ and $c \leq_P d$, then also $a + c \leq_P b + d$ and $ac \leq_P bd$, (3) for every $a, b \in \mathcal{R}$ it holds that $a \leq_P a + b$, and (4) for every $a \in \mathcal{R}$ it holds that $1 \leq_P a \leq_P n$ for some $n \in \mathbb{N}_{\geq 1}$. We also defined in Section 2, for any element $a \in \mathcal{R}$, several notions of rank:

- rank: $\mathrm{R}_P(a) = \min\{n \in \mathbb{N} : a \leq_P n\}$
- subrank: $\mathrm{Q}_P(a) = \max\{n \in \mathbb{N} : n \leq_P a\}$
- asymptotic rank: $\widetilde{\mathrm{R}}_P(a) = \inf\{\mathrm{R}_P(a^m)^{1/m} : m \in \mathbb{N}\}$
- asymptotic subrank: $\widetilde{\mathrm{Q}}_P(a) = \sup\{\mathrm{Q}_P(a^m)^{1/m} : m \in \mathbb{N}\}$

We also defined the asymptotic preorder $\widetilde{P}$ by $a \leq_{\widetilde{P}} b$ if and only if $a^n \leq_P b^n 2^{o(n)}$.

With this setup we introduced in this section the notion of P-monotone homomorphisms, our dual objects. A map $\phi : \mathcal{R} \to \mathbb{R}_{\geq 1}$ is called a P-monotone homomorphism if for all $a, b \in \mathcal{R}$ we have $\phi(ab) = \phi(a)\phi(b)$, $\phi(a+b) = \phi(a) + \phi(b)$, $\phi(1) = 1$ and $a \leq_P b \implies \phi(a) \leq \phi(b)$. The asymptotic spectrum $\mathcal{X} = \mathcal{X}_P$ is the set of all P-monotone homomorphisms $\phi : \mathcal{R} \to \mathbb{R}_{\geq 1}$. This is a compact Hausdorff space in the coarsest topology that makes all evaluation maps $\widehat{s} : \mathcal{X} \to \mathbb{R} : \phi \mapsto \phi(s)$ continuous. In terms of these dual objects we obtain the duality theorem. We defined an element $a \in \mathcal{R}$ to be gapped if either $a^k \geq_P 2$ for some $k \in \mathbb{N}$ or there is a map $\phi \in \mathcal{X}$ such that $\phi(a) = 1$.

Theorem 3.41 (Duality theorem, Corollaries 3.30 and 3.31 and Theorem 3.22)**.** *Let $a, b \in \mathcal{R}$. Then*

- $\widetilde{\mathrm{R}}_P(a) = \max_{\phi \in \mathcal{X}_P} \phi(a)$
- $\widetilde{\mathrm{Q}}_P(a) = \min_{\phi \in \mathcal{X}_P} \phi(a)$ *assuming a is gapped*
- $a \leq_{\widetilde{P}} b \iff \forall \phi \in \mathcal{X}_P,\ \phi(a) \leq \phi(b)$.

One of the main examples of a semiring with a Strassen preorder in Section 2.3 is the semiring of continuous functions Example 2.19. We remarked that this example is in fact "complete" and here we explain how. In Section 4.4 we will go into the details and proofs of this. Let $\mathcal{C} = C(\mathcal{X}, \mathbb{R}_{\geq 1})$ be the semiring of continuous functions from $\mathcal{X}$ to $\mathbb{R}_{\geq 1}$, with pointwise addition, multiplication and preorder. Let $\Phi : \mathcal{R} \to \mathcal{C} : s \mapsto \hat{s}$ where $\hat{s} : \mathcal{X} \to \mathbb{R}$ is the evaluation map $\phi \mapsto \phi(s)$. Then

$$\Phi : \mathcal{R} \to \mathcal{C}$$

is a semiring homomorphism. The duality theorem translates directly:

Theorem 3.42 (Duality theorem, rephrased)**.** *Let $a, b \in \mathcal{R}$. Then*

- $\widetilde{\mathrm{R}}(a) = \max \Phi(a)$
- $\widetilde{\mathrm{Q}}(a) = \min \Phi(a)$ *assuming a is gapped*
- $a \leq_{\widetilde{P}} b \iff \Phi(a) \leq \Phi(b)$.

The theorem tells us that we can map $\mathcal{R}$ to the semiring of continuous functions in a way that the parameters of interest can be read off easily. The claim that the semiring of continuous functions is "complete" becomes apparent with the density theorem that we will prove later (Theorem 4.19) but state here already.

Let $\mathbb{Q}\Phi(\mathcal{R})$ denote the $\mathbb{Q}$-span of $\Phi(\mathcal{R})$ in the ring $C(\mathcal{X}, \mathbb{R})$ of continuous functions from $\mathcal{X}$ to the reals.

Theorem 3.43 (Density theorem)**.** *The set $\mathbb{Q}\Phi(\mathcal{R})$ is dense in $C(\mathcal{X}, \mathbb{R})$ under the sup-norm.*

Unrolling the statement of the density theorem we get that for any $f \in C(\mathcal{X}, \mathbb{R})$ and for any $\varepsilon > 0$ there are $a, b \in \mathcal{R}$ and $n \in \mathbb{N}_{\geq 1}$ such that for all $\phi \in \mathcal{X}$ it holds that $\left|\phi(a)/n - \phi(b)/n - f(\phi)\right| < \varepsilon$.

Combining the density theorem and the duality theorem we find with a short argument (see Section 4.4) that any non-negative continuous function on $\mathcal{X}$ can be approximated with rescaled inequalities $a \leq_{\widetilde{P}} b$ for elements $a, b \in \mathcal{R}$:

Corollary 3.44 (Completeness theorem)**.** *Any element of $C(\mathcal{X}, \mathbb{R}_{\geq 1})$ can be approximated arbitrarily well by differences $\frac{1}{n}\hat{b} - \frac{1}{n}\hat{a}$ for some elements $n \in \mathbb{N}_{\geq 1}$ and $a, b \in \mathcal{R}$ that satisfy $a \leq_{\widetilde{P}} b$*

The completeness theorem says that the collection of all inequalities $a \leq_{\widetilde{P}} b$ is very rich. Namely, any non-negative continuous function on $\mathcal{X}$ that we come up with we can simulate (up to scaling) with the asymptotic preorder in $\mathcal{R}$.

4. Applications, variations and extensions of Strassen duality

In this section we will discuss various applications, connections and variations of Strassen duality results that we have covered in Section 3. We give a high-level introduction to each topic, providing main ideas and results, references for further reading, and some directions for future research.

4.1. Additivity if and only if multiplicativity

As we discussed, the asymptotic rank is a sub-additive and sub-multiplicative parameter in every semiring with a Strassen preorder. Clearly it is additive and multiplicative when applied to pairs of integers in the semiring. What other pairs of elements is it additive on? Multiplicative on? These two interesting questions seem completely unrelated. However, (perhaps counterintuitively) it turns out that they are equivalent, which follows from the duality theorem, as we will see now. This application of the duality theorem was brought to our attention by Ron Holzman (personal communication) and may be thought of as a generalization of Schönhage's tau theorem to any semiring with Strassen preorder.

Let $\mathcal{R}$ be a semiring with a Strassen preorder P. Let $\mathcal{X}$ be the corresponding asymptotic spectrum. We will in this section assume $\mathcal{R}$ to be gapped (Definition 3.27) so that the asymptotic subrank is characterized by the asymptotic spectrum. However, the theorems about asymptotic rank that follow hold already without this assumption of being gapped.

Theorem 4.1 (Additivitiy if and only if multiplicativity)**.** *For any $a, b \in \mathcal{R}$ the following are equivalent:*

- $\widetilde{\mathrm{Q}}(a + b) = \widetilde{\mathrm{Q}}(a) + \widetilde{\mathrm{Q}}(b)$
- $\widetilde{\mathrm{Q}}(ab) = \widetilde{\mathrm{Q}}(a)\,\widetilde{\mathrm{Q}}(b)$
- *there is an element $\phi \in \mathcal{X}$ such that $\phi(a) = \widetilde{\mathrm{Q}}(a)$ and $\phi(b) = \widetilde{\mathrm{Q}}(b)$.*

For any $a, b \in \mathcal{R}$ the following are equivalent:

- $\widetilde{\mathrm{R}}(a + b) = \widetilde{\mathrm{R}}(a) + \widetilde{\mathrm{R}}(b)$
- $\widetilde{\mathrm{R}}(ab) = \widetilde{\mathrm{R}}(a)\,\widetilde{\mathrm{R}}(b)$

- *there is an element $\phi \in \mathcal{X}$ such that $\phi(a) = \widetilde{\mathrm{R}}(a)$ and $\phi(b) = \widetilde{\mathrm{R}}(b)$.*

Theorem 4.1 will follow from a more general theorem that we will state and prove in a moment (Theorem 4.8). We first give some consequences which illustrate the power of Theorem 4.1.

Example 4.2 (The asymptotic rank and asymptotic subrank on univariate polynomials)**.** We begin with a general application of Theorem 4.1. This application generalizes Schönhage's tau theorem (which we will discuss in Section 4.2) from matrix multiplication tensors to any semiring $\mathcal{R}$ with a Strassen preorder P. For any element $a \in \mathcal{R}$ we have by definition of the asymptotic rank and asymptotic subrank that $\widetilde{\mathrm{R}}(a^k) = \widetilde{\mathrm{R}}(a)^k$ and $\widetilde{\mathrm{Q}}(a^k) = \widetilde{\mathrm{Q}}(a)^k$ for every $k \in \mathbb{N}$. Thus by Theorem 4.1 the asymptotic rank and asymptotic subrank are additive on powers of a. For any polynomial $p(x)$ with non-negative integer coefficients, we may consider the element $p(a) \in \mathcal{R}$ and for this element it follows that we have:

Corollary 4.3. *For any $p \in \mathbb{N}[x]$, $\widetilde{\mathrm{R}}(p(a)) = p(\widetilde{\mathrm{R}}(a))$ and $\widetilde{\mathrm{Q}}(p(a)) = p(\widetilde{\mathrm{Q}}(a))$.*

In particular, if the semiring $\mathcal{R}$ is generated by a single element $a \in \mathcal{R}$, meaning that every element in $\mathcal{R}$ is of the form $p(a)$ for some polynomial p with non-negative integer coefficients, then the asymptotic rank $\widetilde{\mathrm{R}}$ and the asymptotic subrank $\widetilde{\mathrm{Q}}$ are points in the asymptotic spectrum $\mathcal{X}$ of $\mathcal{R}$.

Example 4.4 (The Shannon capacity is not always attained at a finite power)**.** It is a natural problem to ask for which elements a in a preordered semiring the asymptotic values $\widetilde{\mathrm{Q}}(a) = \lim_{n\to\infty} \mathrm{Q}(a^n)^{1/n}$ and $\widetilde{\mathrm{R}}(a) = \lim_{n\to\infty} \mathrm{R}(a^n)^{1/n}$ are attained already at a finite power of a, and for which they only approach it at infinity? We consider the semiring $\mathcal{R}$ of graphs with P the cohomomorphism preorder (Example 2.20). For any graph $G \in \mathcal{R}$ we say that the Shannon capacity $\Theta(G) = \lim_{n\to\infty} \alpha(G^{\boxtimes n})^{1/n}$ (which is the asymptotic subrank of G in this setting) is *attained at the nth power* if $\Theta(G)$ equals the normalized independence number $\alpha(G^{\boxtimes n})^{1/n}$ for some fixed value of n. It is easy to find graphs where the Shannon capacity is attained at the first power. In particular, Shannon showed that this is true for all graphs on at most five vertices, except for the five-cycle C_5, for which $\alpha(C_5^{\boxtimes 2})^{1/2} = \sqrt{5} > \alpha(C_5) = 2$ [Sha56]. Lovász proved that the Shannon capacity of C_5 is in fact attained at the second power and thus that $\Theta(C_5) = \sqrt{5}$ [Lov79]. Alon and Lubetzky pointed out that there are graphs for which the Shannon capacity is not attained at any finite power [AL06], and they give as an example the disjoint union of the five-cycle and a single vertex, $C_5 \sqcup K_1$. This easily follows from Theorem 4.1. Indeed, the product $C_5 \boxtimes K_1$ is clearly isomorphic to C_5, and so the Shannon capacity is multiplicative on C_5 and K_1: $\Theta(C_5 \boxtimes K_1) = \Theta(C_5) = \Theta(C_5)\Theta(K_1)$. Therefore, by Theorem 4.1, the Shannon capacity is additive on C_5 and K_1, and so $\Theta(C_5 \sqcup K_1) = \Theta(C_5) + \Theta(K_1) = \sqrt{5} + 1$. One verifies directly that the number $\sqrt{5} + 1$ is not equal to $a^{1/n}$ for any non-negative integers a, n. Therefore, the Shannon capacity $\Theta(C_5 \sqcup K_1)$ is not attained at a finite power.

Example 4.5. Related to Example 4.4 we note in passing that, for the semiring $\mathcal{R}$ of tensors with P the restriction preorder (Example 2.21), the asymptotic rank is also not always attained at a finite power. This is in particular true for our protagonist, the matrix multiplication tensor, as was proved by Coppersmith and Winograd [CW82, Corollary 3.4]. Namely, they showed that if $\mathrm{R}(\mathrm{MM}_n) = n^{\omega_0}$, then $\omega < \omega_0$. In other words, no single algorithm for computing the product of $m \times m$ matrices for some fixed m, can give the optimal algorithm for computing the product of $n \times n$ matrices for arbitrary n by recursive application as in Strassen's algorithm [Str69]. The proof of Coppersmith and Winograd that the matrix multiplication exponent is not attained at a finite power is more involved than the proof that the Shannon capacity is not attained at a finite power, and in particular we do not see how to reproduce it as an application of Theorem 4.1.

Example 4.6 (Equivalence of Shannon's additivity and multiplicativity conjectures)**.** We give another application of Theorem 4.1 to the semiring $\mathcal{R}$ of graphs with P the cohomomorphism preorder (Example 2.20). Shannon made in his seminal paper [Sha56] two conjectures: (1) the Shannon capacity is additive under the disjoint union and (2) the Shannon capacity is multiplicative under the strong product. Haemers disproved the second conjecture in [Hae79] and Alon disproved the first conjecture in [Alo98b]. The tight relation between addition and multiplication provided by Theorem 4.1 implies that the two conjectures are in fact equivalent, and moreover the equivalence is true in a "local" manner: For any two graphs G and H we have $\Theta(G \boxtimes H) > \Theta(G)\Theta(H)$ if and only if $\Theta(G + H) > \Theta(G) + \Theta(H)$.

Example 4.7. It is again natural to relate Example 4.6 to the situation for the semiring $\mathcal{R}$ of tensors with P the restriction preorder. In this situation, it follows already from Strassen's matrix multiplication algorithm [Str69] that the asymptotic rank is not multiplicative and hence not additive. We explain how. Recall from Example 2.22 that the matrix multiplication tensors $\mathrm{MM}_{a,b,c}$ have the recursive property that $\mathrm{MM}_{a,b,c} \otimes \mathrm{MM}_{d,e,f} = \mathrm{MM}_{ad,be,cf}$.[45] Strassen's algorithm implies that $\widetilde{\mathrm{R}}(\mathrm{MM}_{2,2,2}) \leq \mathrm{R}(\mathrm{MM}_{2,2,2}) \leq 7$. On the other hand, we may write $\mathrm{MM}_{2,2,2} = \mathrm{MM}_{2,1,1} \otimes \mathrm{MM}_{1,2,1} \otimes \mathrm{MM}_{1,1,2}$ and it is easy to see that $\widetilde{\mathrm{R}}(\mathrm{MM}_{2,1,1})$, $\widetilde{\mathrm{R}}(\mathrm{MM}_{2,1,1})$ and $\widetilde{\mathrm{R}}(\mathrm{MM}_{2,1,1})$ are all equal to 2. Therefore, we have the strict sub-multiplicativity $\widetilde{\mathrm{R}}(\mathrm{MM}_{2,1,1} \otimes \mathrm{MM}_{1,2,1} \otimes \mathrm{MM}_{1,1,2}) < \widetilde{\mathrm{R}}(\mathrm{MM}_{2,1,1})\, \widetilde{\mathrm{R}}(\mathrm{MM}_{2,1,1})\, \widetilde{\mathrm{R}}(\mathrm{MM}_{2,1,1})$. This example, in which we take a product of three elements instead of two, leads us naturally to the following generalization of Theorem 4.1.

The next theorem generalizes Theorem 4.1 by replacing the sum and the product by polynomials with non-negative integer coefficients. We recall that the asymptotic spectrum $\mathcal{X}$ is the set of P-monotone semiring homomorphisms $\phi : \mathcal{R} \to \mathbb{R}_{\geq 1}$. We recall that for any collection $a_1, \ldots, a_n \in \mathcal{R}$ we defined

$$\mathcal{X}(a_1, \ldots, a_n) \coloneqq \{(\phi(a_1), \ldots, \phi(a_n)) : \phi \in \mathcal{X}\} \subseteq [1, \infty)^n$$

and clearly $\mathcal{X}(a_1, \ldots, a_n) \subseteq \mathcal{X}(a_1) \times \cdots \times \mathcal{X}(a_n)$.

Theorem 4.8. *Fix any elements $a_1, \ldots, a_n$. The following are equivalent:*

(i) *For every polynomial $p \in \mathbb{N}[x_1, \ldots, x_n]$ we have*

$$\widetilde{\mathrm{Q}}(p(a_1, \ldots, a_n)) = p(\widetilde{\mathrm{Q}}(a_1), \ldots, \widetilde{\mathrm{Q}}(a_n)).$$

(ii) *There exists a polynomial $p \in \mathbb{N}[x_1, \ldots, x_n]$, depending on all n variables, such that*

$$\widetilde{\mathrm{Q}}(p(a_1, \ldots, a_n)) = p(\widetilde{\mathrm{Q}}(a_1), \ldots, \widetilde{\mathrm{Q}}(a_n)).$$

(iii) *There exist a spectral point $\phi \in \mathcal{X}$ such that for all $i \in [n]$ it holds that $\widetilde{\mathrm{Q}}(a_i) = \phi(a_i)$.*

(iv) *$\mathcal{X}(a_1, \ldots, a_n)$ contains the point $(\min \mathcal{X}(a_1), \ldots, \min \mathcal{X}(a_n))$.*

[45] We note that in our notation of tensor networks the tensor $\mathrm{MM}_{a,b,c}$ equals $T_{\mathsf{C}_3,(a,b,c)}$. Strassen denotes $\mathrm{MM}_{a,b,c}$ by $\langle a, b, c\rangle$ in his work and this notation is often used in the matrix multiplication literature.

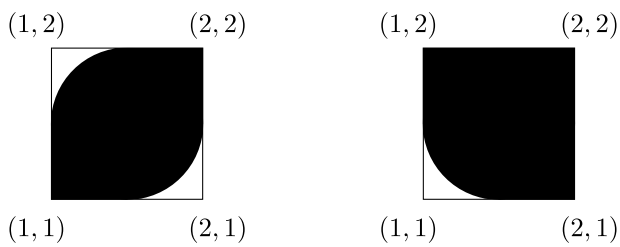
The same holds with asymptotic subrank $\widetilde{\mathrm{Q}}$ *replaced by asymptotic rank* $\widetilde{\mathrm{R}}$ *when in* (iv) *the expression* $(\min \mathcal{X}(a_1), \ldots, \min \mathcal{X}(a_n))$ *is replaced by* $(\max \mathcal{X}(a_1), \ldots, \max \mathcal{X}(a_n))$.

When the conditions of Theorem 4.8 hold we may think of the elements $a_1, \ldots, a_n$ being "free" or "independent" with respect to the asymptotic subrank $\widetilde{\mathrm{Q}}$ (or with respect to the asymptotic rank $\widetilde{\mathrm{R}}$, respectively).

We will give the proof of Theorem 4.8 after some examples.

Example 4.9 (Continuing Example 4.7)**.** Continuing the discussion of Example 4.7 we find that Theorem 4.8 implies that the asymptotic rank of the tensor $p(T_{2,1,1}, T_{1,2,1}, T_{1,1,2})$, for any polynomial $p(x_1, x_2, x_3)$ with non-negative integer coefficients depending on all three variables, is strictly smaller than $p(\widetilde{\mathrm{R}}(T_{2,1,1}), \widetilde{\mathrm{R}}(T_{1,2,1}), \widetilde{\mathrm{R}}(T_{1,1,2}))$. This observation is consistent with Strassen's asymptotic rank conjecture which says that the asymptotic rank of any concise[46] $n_1 \times n_2 \times n_3$ tensor is equal to $\max_i n_i$, and in particular it is consistent with the possibility that the asymptotic rank of $T_{a,b,c}$, which is a concise $ab \times bc \times ca$ tensor, equals $\max\{ab, ac, bc\}$.

Example 4.10. Condition (iv) of Theorem 4.8 has a simple pictorial interpretation. To show this, let $a_1, a_2 \in \mathcal{R}$ (so $n = 2$). Then $\mathcal{X}(a_1)$ and $\mathcal{X}(a_2)$ are compact subsets of $[1, \infty)$, and $\mathcal{X}(a_1, a_2)$ is a compact subset of $\mathcal{X}(a_1) \times \mathcal{X}(a_2)$ (Section 3.6). We will see in Theorem 4.16 that the natural projections $\mathcal{X}(a_1, a_2) \to \mathcal{X}(a_i)$ are surjective. Let us for simplicity assume that $\mathcal{X}(a_1) = [1, 2]$ and $\mathcal{X}(a_2) = [1, 2]$ are closed intervals, and let us then draw illustrations of two possible examples of $\mathcal{X}(a_1, a_2) \subseteq [1, 2] \times [1, 2]$, taking the aforementioned properties into account:



In the illustration on the left, $\mathcal{X}(a_1, a_2)$ contains the point $(\min \mathcal{X}(a_1), \min \mathcal{X}(a_2))$ and thus by Theorem 4.8 in this situation for every polynomial $p \in \mathbb{N}[x_1, x_2]$ we have $\widetilde{\mathrm{Q}}(p(a_1, \ldots, a_n)) = p(\widetilde{\mathrm{Q}}(a_1), \ldots, \widetilde{\mathrm{Q}}(a_n))$. In the illustration on the right, on the other hand, $\mathcal{X}(a_1, a_2)$ does not contain the point $(\min \mathcal{X}(a_1), \min \mathcal{X}(a_2))$ and thus for every polynomial $p \in \mathbb{N}[x_1, x_2]$ that contains both variables x_1 and x_2 we have $\widetilde{\mathrm{Q}}(p(a_1, \ldots, a_n)) \neq p(\widetilde{\mathrm{Q}}(a_1), \ldots, \widetilde{\mathrm{Q}}(a_n))$.

Proof of Theorem 4.8. (i) $\implies$ (ii). This implication is trivial.

(ii) $\implies$ (iii). Let $p(x_1, \ldots, x_n) = \sum_j m_j(x_1, \ldots, x_n)$ be the decomposition of the (given) polynomial p into monomials m_j. Then

$$p(\widetilde{\mathrm{Q}}(a_1), \ldots, \widetilde{\mathrm{Q}}(a_n)) = \sum_j m_j(\widetilde{\mathrm{Q}}(a_1), \ldots, \widetilde{\mathrm{Q}}(a_n)).$$

[46] We say two tensors S and T are equivalent if $S \leq T$ and $T \leq S$. An $n_1 \times n_2 \times n_3$ tensor is *concise* if it is not equivalent to any $m_1 \times m_2 \times m_3$ tensor for which any m_i is strictly smaller than n_i.

By the duality theorem (Corollary 3.31) there exists an element in the asymptotic spectrum $\phi \in \mathcal{X}$ such that $\widetilde{\mathrm{Q}}(p(a_1,\ldots,a_n)) = \phi(p(a_1,\ldots,a_n))$, and so (since ϕ is additive and multiplicative)

$$\widetilde{\mathrm{Q}}(p(a_1,\ldots,a_n)) = \phi(p(a_1,\ldots,a_n)) = \sum_j \phi(m_j(a_1,\ldots,a_n)) = \sum_j m_j(\phi(a_1),\ldots,\phi(a_n)).$$

From the above, we find

$$\sum_j m_j(\widetilde{\mathrm{Q}}(a_1),\ldots,\widetilde{\mathrm{Q}}(a_n)) = \sum_j m_j(\phi(a_1),\ldots,\phi(a_n)).$$

Since for every $i \in [n]$ it holds that $\widetilde{\mathrm{Q}}(a_i) \leq \phi(a_i)$, it follows that for every i we have $\widetilde{\mathrm{Q}}(a_i) = \phi(a_i)$.

(iii) $\implies$ (i). Let $p \in \mathbb{N}[x_1,\ldots,x_n]$ be any polynomial and let $p(x_1,\ldots,x_n) = \sum_j m_j(x_1,\ldots,x_n)$ be the decomposition of the polynomial p into monomials m_j. Using super-additivity and super-multiplicativity of the asymptotic subrank $\widetilde{\mathrm{Q}}$, additivity and multiplicativity of the (given) spectral point ϕ, and the fact that $\phi(b) \geq \widetilde{\mathrm{Q}}(b)$ for any element b, we derive

$$\begin{aligned}
\widetilde{\mathrm{Q}}(p(a_1,\ldots,a_n)) &\geq p(\widetilde{\mathrm{Q}}(a_1),\ldots,\widetilde{\mathrm{Q}}(a_n)) \\
&= p(\phi(a_1),\ldots,\phi(a_n)) \\
&= \phi(p(a_1,\ldots,a_n)) \\
&\geq \widetilde{\mathrm{Q}}(p(a_1,\ldots,a_n))
\end{aligned}$$

and all inequalities must then be equalities. In particular, $p(\widetilde{\mathrm{Q}}(a_1),\ldots,\widetilde{\mathrm{Q}}(a_n)) = p(\phi(a_1),\ldots,\phi(a_n))$.

(iii) $\iff$ (iv). This equivalence follows directly from the fact that $\min \mathcal{X}(a_i) = \widetilde{\mathrm{Q}}(a_i)$ and the definition of $\mathcal{X}(a_1,\ldots,a_n)$ as $\{(\phi(a_1),\ldots,\phi(a_n)) : \phi \in X\}$ and $\mathcal{X}(a_i)$ as $\{\phi(a_i) : \phi \in \mathcal{X}\}$.

The claim for the asymptotic rank $\widetilde{\mathrm{R}}$ is proven in precisely the same way. $\square$

Example 4.11. Condition (ii) of Theorem 4.8 states the existence of a polynomial $p \in \mathbb{N}[x_1,\ldots,x_n]$ depending on all n variables $x_1,\ldots,x_n$ such that $\widetilde{\mathrm{Q}}(p(a_1,\ldots,a_n) = p(\widetilde{\mathrm{Q}}(a_1),\ldots,\widetilde{\mathrm{Q}}(a_n))$. It is natural to ask whether this condition (ii) is equivalent to requiring that $\widetilde{\mathrm{Q}}(a_i a_j) = \widetilde{\mathrm{Q}}(a_i)\,\widetilde{\mathrm{Q}}(a_j)$ for every pair $i,j \in [n]$. The following simple abstract example shows that the answer is no. Namely, the latter condition is strictly weaker. Let $\mathcal{R} \subseteq \mathbb{N}^3$ be the subsemiring generated by the three elements $a_1 = (2,1,1)$, $a_2 = (1,2,1)$ and $a_3 = (1,1,2)$ under pointwise addition and multiplication and endowed with the pointwise preorder (cf. Example 2.19). Then for any element $b \in \mathcal{R}$ the asymptotic subrank $\widetilde{\mathrm{Q}}(b)$ is simply equal to the smallest coefficient $\min(b)$. The asymptotic spectrum $\mathcal{X}$ consists of the three maps $\phi_i : b \mapsto b_i$ for $i \in [3]$. We see directly that $\min(a_i a_j) = 1 = \min(a_i)\min(a_j)$, and so the asymptotic subrank is multiplicative on all products of pairs of generators. However, $\min(a_1 a_2 a_3) = 2 > 1 = \min(a_1)\min(a_2)\min(a_3)$, and so the asymptotic subrank is not multiplicative on the product of all three generators.

In fact, we can also use the concrete matrix multiplication (tensor) setting of Example 4.9 to see an example of the same behavior, but for the asymptotic rank, since in that example the asymptotic rank is multiplicative on any product of pairs of generators, but not multiplicative on the product of all generators by virtue of the fact that the matrix multiplication exponent ω is strictly less than 3.

We have in our treatment above focussed exclusively on the interaction between addition and multiplication, but one can imagine that there are strong interactions with (or among) other interesting operations (on graphs or other objects). For example, in Hedetniemi's conjecture [Hed66] (which has been disproved by Shitov [Shi19] and the counterexamples were strengthened by He and Wigderson [HW21], Zhu [Zhu19], and Wrochna [Wro20]) the relevant operation is the categorical product on graphs. Simonyi [Sim21] studies this operation in the context of the asymptotic spectrum of graphs (cf. Examples 2.20 and 3.38).

4.2. Schönhage's tau theorem and a direct sum theorem for square matrix multiplication

One of the most important theorems for the construction of matrix multiplication algorithms (cf. Example 2.22) is Schönhage's tau theorem [Sch81]. We state and prove this theorem here and explain how it is related to the *additivity if and only if multiplicativity theorem* (Section 4.1) and in particular to the univariate semiring example, Example 4.2.

The tau theorem is situated in the setting of tensors (cf. Example 2.21), where (we recall) R denotes the tensor rank, MM_m denote the matrix multiplication tensors, and $\omega = \log \widetilde{\mathrm{R}}(\mathrm{MM}_2)$ is the matrix multiplication exponent. The earliest and simplest method to upper bound ω uses the simple fact that any rank upper bound $\mathrm{R}(\mathrm{MM}_n) \leq r$ implies the exponent upper bound $n^\omega \leq r$. The tau theorem extends this method to upper bounds on the rank of *direct sums* of matrix multiplication tensors:

Theorem 4.12 (Tau theorem[47][48], Schönhage [Sch81])**.** *If* $\mathrm{R}(\sum_i \mathrm{MM}_{n_i}) \leq r$, *then* $\sum_i n_i^\omega \leq r$.

It is easy to see that Theorem 4.12 (whose original proof was quite complex) follows directly from the following direct sum theorem of Strassen, which relies on Strassen's duality theorem.

Theorem 4.13 (Direct sum theorem, Strassen [Str88])**.** $\widetilde{\mathrm{R}}(\sum_i \mathrm{MM}_{n_i}) = \sum_i \widetilde{\mathrm{R}}(\mathrm{MM}_{n_i})$.

(To prove Theorem 4.12 from Theorem 4.13, we use the fact that asymptotic rank is at most rank, $\widetilde{\mathrm{R}} \leq \mathrm{R}$, and that $\widetilde{\mathrm{R}}(\mathrm{MM}_{n_i}) = {n_i}^\omega$.) Schönhage proved Theorem 4.12 before Strassen's duality, with a proof that is more ad hoc. This partly motivated Strassen to develop his duality theory. Moreover, using duality, the proof not only gets much simpler, but also its tightness follows.

We discuss the proof of Theorem 4.13. It is easy to see that Example 4.2 directly implies the special case of Theorem 4.13 when all n_i are of the form 2^{m_i}. Indeed, in that case, every tensor $\sum_i \mathrm{MM}_{n_i}$ is of the form $p(\mathrm{MM}_2)$ for some polynomial $p \in \mathbb{N}[x]$. We know that in such a univariate setting (Example 4.2) we have that $\widetilde{\mathrm{R}}(p(\mathrm{MM}_2)) = p(\widetilde{\mathrm{R}}(\mathrm{MM}_2))$. It is intuitively clear that this special case for powers of 2 captures the essence of Theorem 4.12. To prove the full Theorem 4.12 one follows a similar argument as the proof of Theorem 4.8 with one additional simple fact:

Lemma 4.14. *Let* $\phi \in \mathcal{X}$ *and* $m \in \mathbb{N}$. *Then* $\phi(\mathrm{MM}_m) = \phi(\mathrm{MM}_2)^{\log m}$.

Proof. This follows from a straightforward approximation argument, using the multiplicativity of ϕ and taking large powers of MM_m to get rid of the "gap" in the approximation. Namely, for any $k \in \mathbb{N}$ we have

$$2^a \leq m^k \leq 2^{a+1}$$

[47]The name "tau theorem" comes from the fact that a variable named τ plays a central role in Schönhage's original proof [Blä13]. The theorem is also referred to as the asymptotic sum inequality [BCS97].

[48]In fact, the full version of Schönhage's tau theorem allows the rank R to be replaced by the border rank $\underline{\mathrm{R}}$, which is the (algebro-geometric) approximative version of rank. Furthermore, it allows the matrix multiplication tensors to be *rectangular*, resulting in: if $\sum_i \underline{\mathrm{R}}(\mathrm{MM}_{a_i,b_i,c_i}) \leq r$, then $\sum_i (a_i b_i c_i)^{\omega/3} \leq r$. We return to this version and further extensions in Section 9.1.

for $a = \lfloor k \log m \rfloor$. Therefore we have the inequalities of matrix multiplication tensors

$$\mathrm{MM}_2^{\otimes a} \leq \mathrm{MM}_m^{\otimes k} \leq \mathrm{MM}_2^{\otimes (a+1)}.$$

We apply ϕ and use multiplicativity of ϕ to get

$$\phi(\mathrm{MM}_2)^{a/k} \leq \phi(\mathrm{MM}_m) \leq \phi(\mathrm{MM}_2)^{(a+1)/k}.$$

Finally, letting k go to infinity, we obtain the claim. $\square$

Now we give the proof of Theorem 4.13 using Strassen duality and Lemma 4.14.

Proof of Theorem 4.13. Strassen duality (Corollary 3.30) gives

$$\widetilde{\mathrm{R}}\Big(\sum_i \mathrm{MM}_{n_i}\Big) = \max_{\phi \in \mathcal{X}} \phi\Big(\sum_i \mathrm{MM}_{n_i}\Big).$$

Using additivity of ϕ and Lemma 4.14 we get

$$\widetilde{\mathrm{R}}\Big(\sum_i \mathrm{MM}_{n_i}\Big) = \max_{\phi \in \mathcal{X}} \sum_i \phi(\mathrm{MM}_2)^{\log n_i}.$$

Since the function $x \mapsto \sum_i x^{\log n_i}$ is monotonically increasing, we find that

$$\max_{\phi \in \mathcal{X}} \sum_i \phi(\mathrm{MM}_2)^{\log n_i} = \sum_i \big(\max_{\phi \in \mathcal{X}} \phi(\mathrm{MM}_2)\big)^{\log n_i}.$$

Using Strassen duality to get $\max_{\phi \in \mathcal{X}} \phi(\mathrm{MM}_2) = \widetilde{\mathrm{R}}(\mathrm{MM}_2)$, we get

$$\widetilde{\mathrm{R}}\Big(\sum_i \mathrm{MM}_{n_i}\Big) = \sum_i \widetilde{\mathrm{R}}(\mathrm{MM}_2)^{\log n_i}.$$

Finally, using the duality (Corollary 3.30) and Lemma 4.14 once more, we find

$$\widetilde{\mathrm{R}}(\mathrm{MM}_2)^{\log n_i} = (\max_{\phi \in \mathcal{X}} \phi(\mathrm{MM}_2))^{\log n_i} = \max_{\phi \in \mathcal{X}} \phi(\mathrm{MM}_2)^{\log n_i} = \max_{\phi \in \mathcal{X}} \phi(\mathrm{MM}_{n_i}) = \widetilde{\mathrm{R}}(\mathrm{MM}_{n_i}),$$

which finishes the proof. $\square$

4.3. Lifting between semirings

It often occurs that two different semirings with preorders are related to each other. A natural such relation is via semiring homomorphisms that are order embeddings:

Definition 4.15. Given a semiring $\mathcal{R}$ with a Strassen preorder P and another semiring $\mathcal{R}'$ with a Strassen preorder P', a map

$$f : \mathcal{R} \to \mathcal{R}'$$

is called a *semiring homomorphism* if for every $a, b \in \mathcal{R}$ it holds that $f(ab) = f(a)f(b)$, and $f(a+b) = f(a) + f(b)$, and $f(1) = 1$. We call f an *order embedding* if for every $a, b \in \mathcal{R}$, we have that $a \leq_P b$ if and only if $f(a) \leq_{P'} f(b)$.

If two semirings are related by a semiring homomorphism that is an order embedding, then the corresponding asymptotic spectra are also related in the following strong sense:

Theorem 4.16. *Let $\mathcal{R}$ be a semiring with a Strassen preorder P and let $\mathcal{R}'$ be a semiring with a Strassen preorder P'. Let*

$$f : \mathcal{R} \to \mathcal{R}'$$

be a semiring homomorphism that is an order embedding. Then the map from the asymptotic spectrum of $\mathcal{R}'$ to the asymptotic spectrum of $\mathcal{R}$,

$$f^* : \mathcal{X}(\mathcal{R}') \to \mathcal{X}(\mathcal{R}),$$

which maps ϕ to the composition $\phi \circ f$, is surjective.

It is instructive to think of the special case of Theorem 4.16 where $\mathcal{R}$ is a subsemiring of $\mathcal{R}'$.[49] In this case, Theorem 4.16 says that the map $\mathcal{X}(\mathcal{R}') \to \mathcal{X}(\mathcal{R})$ that maps ϕ to the restriction $\phi|_{\mathcal{R}}$, is surjective. In other words, any monotone homomorphism $\psi \in \mathcal{X}(\mathcal{R})$ can be *extended* or *lifted* to a monotone homomorphism $\phi \in \mathcal{X}(\mathcal{R}')$ that is defined on the larger semiring $\mathcal{R}'$. This means that once we have understood the asymptotic spectrum of $\mathcal{R}'$ we learn also precisely the asymptotic spectrum of $\mathcal{R}$.

We will defer the proof of Theorem 4.16 for a moment, as it will follow naturally from the density theorem that we discuss in the next section. We give examples that illustrate the power of Theorem 4.16.

Example 4.17 (Directed and undirected graphs)**.** We have discussed earlier the asymptotic spectrum of (undirected) graphs in Example 2.20 to understand the Shannon capacity. It is natural to consider the Shannon capacity of *directed* graphs (introduced in [KS92] and [GKV92], see also [Alo98a]) and ask how the corresponding asymptotic spectrum relates to the one for undirected graphs. As in Example 2.20, let $\mathcal{R}$ be the semiring of undirected graphs with addition given by disjoint union and multiplication given by the strong product (defined by taking the tensor product of the adjacency matrices where we ensure the diagonal elements to be 1), and with P the cohomomorphism preorder, which is a Strassen preorder. Let $\mathcal{R}'$ be the semiring of *directed* graphs with addition given by disjoint union and multiplication given by the strong product. A cohomomorphism of directed graphs $f : G \to H$ is a map $V(G) \to V(H)$ such that if $u \neq v$ and $(u, v) \notin E(G)$, then $f(u) \neq f(v)$ and $(f(u), f(v)) \notin E(H)$. We define the cohomomorphism preorder P' on $\mathcal{R}'$ by saying that $G \leq_{P'} H$ if and only if there is a cohomomorphism $G \to H$. Let

$$g : \mathcal{R} \to \mathcal{R}'$$

be the injective map that maps any undirected graph G to the associated bidirected graph, which is the directed graph with vertex set $V(G)$ and edge set $\{(u, v) : \{u, v\} \in E(G)\} \cup \{(v, u) : \{u, v\} \in E(G)\}$. Then g is clearly a semiring homomorphism that is an order embedding. Thus the map from the asymptotic spectrum of $\mathcal{R}'$ to the asymptotic spectrum of $\mathcal{R}$,

$$g^* : \mathcal{X}(\mathcal{R}') \to \mathcal{X}(\mathcal{R}),$$

[49] Although this is technically a special case, it is essentially the only case. Indeed, the definition of order embedding implies that f is "essentially" injective: if $f(x) = f(y)$, then we must have $x \leq y$ and $y \leq x$, so that x and y are indistinguishable under the ordering P. So if we work with partial orders rather than preorders, any semiring homomorphism that is an order embedding must actually be an inclusion of one semiring in another.

which maps $\phi \in \mathcal{X}(\mathcal{R}')$ to the composition $\phi \circ g \in \mathcal{X}(\mathcal{R})$, is a surjective map by Theorem 4.16. In other words, any element in the asymptotic spectrum of (undirected) graphs $\mathcal{X}(\mathcal{R})$ can be extended or lifted to an element in the asymptotic spectrum of directed graphs $\mathcal{X}(\mathcal{R}')$. This implies, for instance, that the Lovász theta function ϑ on undirected graphs can be extended to a Lovász theta function $\hat{\vartheta} \in \mathcal{X}(\mathcal{R}')$ on directed graphs.

Example 4.18 (Symmetric tensors and symmetric restriction)**.** We have discussed earlier the asymptotic spectrum of tensors (Example 2.21). An important class of tensors, studied already since at least Sylvester [Syl52], is that of symmetric tensors. Symmetric tensors are very natural and common as they are essentially equivalent to homogeneous polynomials. Of particular interest is a symmetric version of tensor rank known as Waring rank. Symmetric tensors also form a semiring and there is a natural way of defining a Strassen preorder on this semiring called symmetric restriction, which induces the Waring rank as its rank.

It is natural to ask how the asymptotic spectrum of symmetric tensors with symmetric restriction relates to the ordinary asymptotic spectrum of tensors. The answer turns out to be simple, but the proof is too involved to discuss here and we refer to [CFTZ21]. As in Example 2.21, let $\mathcal{R}$ be the semiring of 3-tensors with addition given by the direct sum and multiplication given by the tensor product, and with P the restriction preorder. The symmetric group $\mathfrak{S}_3$ acts on $\mathcal{R}$ by permuting the coordinates. In other words, for any $\pi \in \mathfrak{S}_3$ and any tensor $T = (T_{i,jk})_{i,j,k} \in \mathcal{R}$, π maps T to $(T_{\pi(i),\pi(j),\pi(k)})_{i,j,k}$. The tensor T is called *symmetric* if it is invariant under the action of $\mathfrak{S}_3$. Let $\mathcal{R}_s$ be the semiring of symmetric 3-tensors. This is a subsemiring of $\mathcal{R}$. On symmetric tensors there is a natural preorder P_s called *symmetric restriction*. This preorder P_s on $\mathcal{R}$ is defined by saying for any tensors $T \in \mathbb{F}^{n_1 \times n_2 \times n_3}$ and $T' \in \mathbb{F}^{n_1' \times n_2' \times n_3'}$ that $T \leq_{P_s} T'$ if there is a matrix A of the appropriate size such that $T = (A, A, A) \cdot T'$ where $((A, A, A) \cdot T')_{u,v,w} = \sum_{i,j,k} A_{u,i} A_{v,j} A_{w,k} T'_{i,j,k}$.

Let $f : \mathcal{R}_s \to \mathcal{R}$ be the natural embedding of the semiring of symmetric tensors in the semiring of all tensors. We note that this is not an order embedding from $(\mathcal{R}_s, P_s)$ to $(\mathcal{R}, P)$! However, it is an order embedding from $(\mathcal{R}_s, P)$ to $(\mathcal{R}, P)$. Let $\mathcal{X}(\mathcal{R}, P)$ be the asymptotic spectrum of $\mathcal{R}$ with respect to P and define $\mathcal{X}(\mathcal{R}_s, P)$ and $\mathcal{X}(\mathcal{R}_s, P_s)$ analogously. It follows from Theorem 4.16 that the map $f^* : \mathcal{X}(\mathcal{R}, P) \to \mathcal{X}(\mathcal{R}_s, P)$ which maps $\phi \in \mathcal{X}(\mathcal{R}, P)$ to the composition $\phi \circ f \in \mathcal{X}(\mathcal{R}_s, P)$, is a surjective map. It is proven in [CFTZ21] that the asymptotic restriction preorder $\widetilde{P}$ and the asymptotic symmetric restriction preorder $\widetilde{P_s}$ coincide. It thus follows that $\mathcal{X}(\mathcal{R}_s, P) = \mathcal{X}(\mathcal{R}_s, P_s)$. We conclude that f^* is a surjective map $\mathcal{X}(\mathcal{R}, P) \to \mathcal{X}(\mathcal{R}_s, P_s)$. In other words, the asymptotic spectrum of symmetric tensors under the symmetric restriction preorder is completely determined by the asymptotic spectrum of tensors.

In the above two examples Example 4.17 and Example 4.18, the two semirings $\mathcal{R}$ and $\mathcal{R}'$ were similar in nature and differed only in a mild way (directed vs. undirected graphs, symmetric tensors with symmetric restriction vs. tensors with restriction). One can imagine that there are more elaborate ways to relate different asymptotic spectra to each other. We see this happening intuitively in the semiring of graphs (Example 2.20). Namely, in that setting, the fractional Haemers bound, which is an element in the asymptotic spectrum of graphs, is defined in terms of the matrix rank, which is itself an element (the only element) in the asymptotic spectrum of matrices. One may ask, for example, whether the Lovász theta function similarly arises from an element in the asymptotic spectrum of a natural semiring, and whether there is a theory that explains such connections more broadly.

4.4. Complete semirings and the density theorem

There is a collection of semirings with Strassen preorder that is "complete" for the theory. Recall from Example 2.17 that a natural instance of a semiring with a Strassen preorder is the semiring $C(\mathcal{X}, \mathbb{R}_{\geq 1})$ of continuous functions from a space $\mathcal{X}$ (which here we will require to be compact and Hausdorff for technical reasons that will become clear) to $\mathbb{R}_{\geq 1}$, under pointwise addition, multiplication and preorder. This semiring is "complete" in the sense that *any* semiring with a Strassen preorder can be naturally embedded into a semiring of this form for some $\mathcal{X}$ (namely for $\mathcal{X}$ equal to the asymptotic spectrum), in a way that the $\mathbb{Q}$-span of the image is *dense* in $C(\mathcal{X}, \mathbb{R})$. The following density theorem explains this.

Let $\mathcal{R}$ be a semiring with a Strassen preorder P and let $\mathcal{X}$ denote the asymptotic spectrum. Recall that $\mathcal{X}$ with the weakest topology that makes the evaluation functions continuous is a compact space (Section 3.6). We will also need that $\mathcal{X}$ is Hausdorff, which we proved in Lemma 3.34. Let $C(\mathcal{X}, \mathbb{R})$ be the set of continuous functions from $\mathcal{X}$ to the reals. For every $a \in \mathcal{R}$ let $\hat{a} : \mathcal{X} \to \mathbb{R}_{\geq 1}$ be the evaluation map defined by $\phi \mapsto \phi(a)$. Let $\hat{\mathcal{R}} = \{\hat{a} : a \in \mathcal{R}\}$ denote the set of all evaluation maps. Let $\mathbb{Q}\hat{\mathcal{R}}$ denote the $\mathbb{Q}$-subspace of $C(\mathcal{X}, \mathbb{R})$ spanned by $\hat{\mathcal{R}}$.

Theorem 4.19 (Density theorem)**.** *The set $\mathbb{Q}\hat{\mathcal{R}}$ is dense in $C(\mathcal{X}, \mathbb{R})$ under the sup-norm. That is, for any $f \in C(\mathcal{X}, \mathbb{R})$, for every $\varepsilon > 0$, there are $a, b \in \mathcal{R}$ and $n \in \mathbb{N}_{\geq 1}$ such that for all $\phi \in \mathcal{X}$ it holds that*

$$\Big|\frac{\hat{b}(\phi)}{n} - \frac{\hat{a}(\phi)}{n} - f(\phi)\Big| = \Big|\frac{\phi(b)}{n} - \frac{\phi(a)}{n} - f(\phi)\Big| < \varepsilon.$$

The density theorem follows from the classical Stone–Weierstrass theorem. We need some terminology to state it. We say that a subset $A \subseteq C(\mathcal{X}, \mathbb{R})$ is a $\mathbb{Q}$-subalgebra if A is a $\mathbb{Q}$-subspace which is closed under pointwise multiplication. We call $\mathcal{X}$ Hausdorff if every two distinct points can be separated by open neighborhoods. We say that A separates the elements of $\mathcal{X}$ if for any two elements $s, t \in \mathcal{X}$ there is an element $f \in A$ such that $f(s) \neq f(t)$. The sup-norm on $C(\mathcal{X}, \mathbb{R})$ is defined by $\|f\| = \sup_{s \in \mathcal{X}} |f(s)|$.

Theorem 4.20 (Stone–Weierstrass theorem)**.** *Let $\mathcal{X}$ be a compact Hausdorff space and let A be a $\mathbb{Q}$-subalgebra of $C(\mathcal{X}, \mathbb{R})$ that contains a nonzero constant function. Then A is dense in $C(\mathcal{X}, \mathbb{R})$ under the sup-norm if and only if A separates the elements of $\mathcal{X}$.*

Proof of Theorem 4.19. Let $A = \mathbb{Q}\hat{\mathcal{R}}$. Then A is a $\mathbb{Q}$-subalgebra of $C(\mathcal{X}, \mathbb{R})$. Also, A contains the nonzero constant function $\hat{1}$. If $\phi, \psi \in \mathcal{X}$ are different, then there is an $a \in \mathcal{R}$ such that $\phi(a) \neq \psi(a)$ and hence $\hat{a}(\phi) \neq \hat{a}(\psi)$. Thus A separates the elements of $\mathcal{X}$. We endowed $\mathcal{X}$ with the coarsest topology that makes all evaluation maps $\hat{a}$ for $a \in \mathcal{R}$ continuous. Then $\mathcal{X}$ is compact (Lemma 3.32) and Hausdorff (Lemma 3.34) Therefore, by the Stone–Weierstrass theorem (Theorem 4.20), the set $\mathbb{Q}\hat{\mathcal{R}}$ is dense in $C(\mathcal{X}, \mathbb{R})$ under the sup-norm. $\square$

Remark 4.21. The density theorem Theorem 4.19 says that we can approximate any continuous real-valued function on the asymptotic spectrum $\mathcal{X}$ by spectral differences $\frac{1}{n}\hat{b} - \frac{1}{n}\hat{a}$ for $a, b \in \mathcal{R}, n \in \mathbb{N}$. It is not hard to see that, as a consequence, any *non-negative* continuous real-valued function on $\mathcal{X}$ can be approximated by spectral differences $\frac{1}{n}\hat{b} - \frac{1}{n}\hat{a}$ for $a, b \in \mathcal{R}, n \in \mathbb{N}_{\geq 1}$ that satisfy $a \leq_{\widetilde{P}} b$. The proof for this is as follows. If $f \in C(\mathcal{X}, \mathbb{R}_{\geq 0})$, then for any $\delta > 0$ we have $f + \delta \geq \delta$. Let $\hat{b}_\varepsilon/n_\varepsilon - \hat{a}_\varepsilon/n_\varepsilon$ approximate $f + \delta$. Suppose that for all ε it holds that not $a_\varepsilon \leq_{\widetilde{P}} b_\varepsilon$. Then $\hat{a}_\varepsilon \not\leq \hat{b}_\varepsilon$ so $\hat{a}_\varepsilon/n_\varepsilon \not\leq \hat{b}_\varepsilon/n_\varepsilon$, that is $0 \not\leq \hat{b}_\varepsilon/n_\varepsilon - \hat{a}_\varepsilon/n_\varepsilon$. Now let δ go to zero to prove the claim.

Now that we have the density theorem, we can give the deferred proof of Theorem 4.16.

Proof of Theorem 4.16. Let $\widetilde{P}$ denote the asymptotic preorder on $\mathcal{R}$ corresponding to P. Let $\widetilde{P}'$ denote the asymptotic preorder on $\mathcal{R}'$ corresponding to P'. Let $a, b \in \mathcal{R}$. By the duality theorem applied to $\mathcal{R}$ we have

$$\forall \psi \in \mathcal{X}(\mathcal{R}),\ \psi(a) \leq \psi(b) \quad \iff \quad a \leq_{\widetilde{P}} b.$$

Since f is a homomorphism that is an order embedding we have

$$a \leq_{\widetilde{P}} b \quad \iff \quad f(a) \leq_{\widetilde{P}'} f(b).$$

By the duality theorem applied to $\mathcal{R}'$, and the using the notation f^* for the map $\mathcal{X}(\mathcal{R}') \to \mathcal{X}(\mathcal{R}) : \phi \mapsto \phi \circ f$, we have

$$\begin{aligned} f(a) \leq_{\widetilde{P}'} f(b) \quad &\iff \quad \forall \phi \in \mathcal{X}(\mathcal{R}'),\ \phi(f(a)) \leq \phi(f(b)) \\ &\iff \quad \forall \psi \in f^*(\mathcal{X}(\mathcal{R}')),\ \psi(a) \leq \psi(b). \end{aligned}$$

Write $\mathcal{X} = \mathcal{X}(\mathcal{R})$ and $\mathcal{Y} = f^*(\mathcal{X}(\mathcal{R}'))$. We know that $\mathcal{X} \supseteq \mathcal{Y}$ and we need to prove that $\mathcal{X} = \mathcal{Y}$. Suppose that there is an element $\psi_0 \in \mathcal{X} \setminus \mathcal{Y}$. Since $\mathcal{X}$ is a compact Hausdorff space, there is a continuous function $g : \mathcal{X} \to [-1, 1]$ with $g(\mathcal{Y}) = 1$ and $g(\psi_0) = -1$. The function g can be approximated by elements from $\mathbb{Q}\hat{\mathcal{R}}$ by the density theorem Theorem 4.19, in the following sense. Applying Theorem 4.19 with $\varepsilon = 1/2$, there are $a, b \in \mathcal{R}$ and an $n \in \mathbb{N}_{\geq 1}$ such that for all $\psi \in \mathcal{Y}$ we have

$$\frac{1}{n}(\hat{a}(\psi) - \hat{b}(\psi)) > 1 - \varepsilon = \tfrac{1}{2}$$

and

$$\frac{1}{n}(\hat{a}(\psi_0) - \hat{b}(\psi_0)) < -1 + \varepsilon = -\tfrac{1}{2}.$$

The first implies that $\hat{a} \geq \hat{b}$ on $\mathcal{Y}$, and so $f(a) \geq_{\widetilde{P}} f(b)$ and thus $a \geq_{\widetilde{P}} b$, while the second implies that *not* $\hat{a} \geq \hat{b}$ on $\mathcal{X}$ and so *not* $a \geq_{\widetilde{P}} b$. $\square$

4.5. Barrier theorems

When a mathematical problem resists being solved, it is natural to try to prove that the available proof methods are intrinsically *not powerful enough* to solve the problem. Theorems on the lack of power of a collection of methods are called *barrier theorems*. These are particularly common in complexity theory, including central results like the relativization barrier in uniform complexity [BGS75], the natural proofs barrier in circuit complexity [RR94], and the rank method barrier in arithmetic complexity [EGdOW18] (relevant in particular for tensor rank lower bounds).

In this section we discuss the idea of proving barrier theorems within the framework of Strassen duality. Specifically, we review recent barrier theorems for obtaining faster algorithms for matrix multiplication (cf. Example 2.22). Interestingly, unlike all barrier results above, which "explain" the difficulty of proving lower bounds, these barrier results "explain" the difficulty of improving the best upper bounds. We will further see how Strassen's duality theorem, the existence of the spectrum and its extreme elements, asymptotic rank and asymptotic subrank, play an essential role in the proof and its simplicity.

Recall that the matrix multiplication problem is to determine the matrix multiplication exponent ω, which is the infimum over all numbers $c \in \mathbb{R}$ such that any two $n \times n$ matrices can be multiplied using $\mathcal{O}(n^c)$ arithmetic operations. The matrix multiplication exponent is easily seen to be between 2 and 3. After Strassen's seminal paper [Str69], in which he proved that $\omega \leq \log 7 < 3$, the community developed increasingly sophisticated tools in order to prove increasingly better upper bounds on ω, leading to the currently best upper bound of $\omega < 2.373$ by Coppersmith and Winograd [CW90], Le Gall [LG12] and Alman and Williams [AW21], leaving, however, unanswered the question whether $\omega = 2$ or $\omega > 2$. Drawing a plot of the development of the best upper bound on ω over time reveals that progress has slowed down considerably over the previous thirty years, with the best upper bound going from the $\omega < 2.376$ of Coppersmith and Winograd [CW90] to the $\omega < 2.373$ of Alman and Williams [AW21].

Ambainis, Filmus and Le Gall [AFLG15] explained this slow-down by proving a barrier theorem for a collection of algorithmic techniques that includes the popular techniques of the last thirty years. In particular, they showed that these techniques cannot prove $\omega = 2$, and in fact not even $\omega \leq 2.31$. This barrier theorem was greatly simplified and generalized by Alman and Williams [AW18] to apply to an even larger collection of tools, and subsequently generalized further by Alman [Alm19] and Christandl, Vrana and Zuiddam [CVZ19b].[50] The latter work explicitly proves this barrier within the framework of Strassen duality in a way that we will discuss now.

On a high level, the barrier theorem for upper bounds on matrix multiplication of [CVZ19b] is based on the following ideas. First we assume that the matrix multiplication algorithm is obtained by *reducing* the matrix multiplication problem to an *intermediate problem* (tensor) T for which there is an efficient algorithm (small asymptotic rank). The currently fastest algorithms indeed all have this structure, where T corresponds to a member of a family of problems going back to Coppersmith and Winograd [CW87] (Coppersmith–Winograd tensors), but we may also take T to be other tensors. For some choices of T there will be a barrier and for some there will not be.[51] The first component for the matrix multiplication barriers is that the asymptotic subrank of the matrix multiplication tensor is *maximal* (for an $m^2 \times m^2 \times m^2$ tensor):

Theorem 4.22 (Strassen [Str87])**.** *For every $m \in \mathbb{N}$, $\widetilde{\mathrm{Q}}(\mathrm{MM}_m) = m^2$. In particular, $\widetilde{\mathrm{Q}}(\mathrm{MM}_2) = 4$.*

The second component for the matrix multiplication barriers, for a given intermediate problem T, is to prove that the asymptotic subrank of T is *small*. If this is indeed the case, then there cannot be a large matrix multiplication problem embedded into T, as we will make precise and prove below, making crucial use of Theorem 4.22. Thus, perhaps surprisingly at first, in proving barrier theorems for *asymptotic rank* (the matrix multiplication exponent), we are using the seemingly unrelated *asymptotic subrank*.[52]

Before stating and proving the barrier theorem, we make the above ideas precise using our setup of tensors, preorders and ranks (Example 2.21) and in particular define what it means for T to be an intermediate problem. First let us rephrase the problem of upper bounding ω in terms of

[50] Also Blasiak, Church, Cohn, Grochow, Naslund, Sawin, and Umans [BCC+17a] and Blasiak, Church, Cohn, Grochow, and Umans [BCC+17b] proved barrier theorems for matrix multiplication for a subset of techniques related to the group-theoretic method. These barrier theorems rely on very similar ideas as the later barrier theorems in [AW18] and [CVZ19b].

[51] Indeed, as will become clear, there are choices of T for which no barrier can be obtained, for example, when T is a matrix multiplication tensor or a diagonal tensor. However, those are not interesting choices of T as they do not provide any extra leverage beyond the standard method of proving tensor rank upper bounds.

[52] In this section we only discuss the barrier theorem for *square* matrix multiplication. This has been extended also to *rectangular* matrix multiplication [CGLZ20] where the asymptotic subrank is replaced by elements in the asymptotic spectrum of rectangular matrix multiplication.

asymptotic inequalities in the restriction preorder. Namely, if the inequality

$$(\mathrm{MM}_2)^{\otimes n} \leq (I_2)^{\otimes (cn+o(n))}$$

holds (where I_2 denotes the $2 \times 2 \times 2$ unit tensor, cf. Example 2.21), then $\omega \leq c$, and arbitrary matrix multiplication algorithms correspond precisely to inequalities of this form.[53] In practice, as mentioned before, matrix multiplication algorithms are obtained by reduction to intermediate problems T. Namely, let T be any tensor. Then clearly, if

$$(\mathrm{MM}_2)^{\otimes n} \leq T^{\otimes (an+o(n))} \leq (I_2)^{\otimes (abn+o(n))}, \tag{4.1}$$

then $\omega \leq ab$. Let ω_T denote the smallest (infimum) value of ab for which (4.1) holds. That is, ω_T is the best upper bound on ω that can be obtained by using T as an intermediate tensor in the sense of (4.1). We call ω_T the *T-exponent of matrix multiplication*. In practice, upper bounds on ω are obtained by upper bounding ω_T for some T. The barrier we prove is a lower bound on ω_T depending on T.

Theorem 4.23 (Barrier theorem for matrix multiplication [CVZ19b][54])**.** *For any tensor T, the best upper bound ω_T on the matrix multiplication exponent ω that can be obtained using T as an intermediate tensor, satisfies the inequality*

$$2\,\frac{\log \widetilde{\mathrm{R}}(T)}{\log \widetilde{\mathrm{Q}}(T)} \leq \omega_T.$$

Proof. This proof uses Strassen's duality theorem. Let $\phi \in \mathcal{X}$ be a spectral point. We apply ϕ to both sides of the first inequality in (4.1), take the nth root and let n go to infinity, to get

$$\phi(\mathrm{MM}_2) \leq \phi(T)^a.$$

Then by taking logarithms and by using the lower bound $2 \leq \log \widetilde{\mathrm{Q}}(\mathrm{MM}_2) \leq \log \phi(\mathrm{MM}_2)$ that follows from Theorem 4.22, we get

$$\frac{2}{\log \phi(T)} \leq \frac{\log \phi(\mathrm{MM}_2)}{\log \phi(T)} \leq a. \tag{4.2}$$

Let $\psi \in \mathcal{X}$ be another spectral point. We apply ψ to both sides of the second inequality in (4.1), take the nth root and let n go to infinity, to get

$$\psi(T) \leq 2^b.$$

Then

$$\log \psi(T) \leq b. \tag{4.3}$$

From combining (4.2) and (4.3) we conclude that

$$\frac{2}{\log \phi(T)} \log \psi(T) \leq ab.$$

[53] In our notation we always choose $o(n)$ so that $cn + o(n)$ is an integer.
[54] This barrier is similar to the barrier of Alman [Alm19].

Maximizing the left-hand side over ϕ and ψ, we find

$$2 \max_{\phi,\psi\in\mathcal{X}} \frac{\log \psi(T)}{\log \phi(T)} \leq ab.$$

Applying the duality theorem (Corollaries 3.30 and 3.31) to the left-hand side, we get

$$2\,\frac{\log \widetilde{\mathrm{R}}(T)}{\log \widetilde{\mathrm{Q}}(T)} \leq ab.$$

Since this inequality holds for any value of ab that satisfies (4.1), we may replace the right-hand side by the T-exponent of matrix multiplication ω_T. This proves the claim. □

We now discuss the power and limitations of Theorem 4.23 as an approach to proving barrier results for matrix multiplication. Theorem 4.23 directly implies that for a given intermediate tensor T to be good for proving $\omega = 2$, the *slack* between the asymptotic rank and asymptotic subrank of T,

$$\mathrm{slack}(T) \coloneqq \frac{\log \widetilde{\mathrm{R}}(T)}{\log \widetilde{\mathrm{Q}}(T)},$$

must be equal to 1, or equivalently the asymptotic subrank and asymptotic rank of T must be equal, $\widetilde{\mathrm{Q}}(T) = \widetilde{\mathrm{R}}(T)$. In other words, if $\mathrm{slack}(T)$ is strictly larger than 1, then we cannot prove $\omega = 2$ via T. More precisely, to prove $\omega = 2$ via intermediate tensors, we must at least have a sequence of intermediate tensors with slack *converging* to 1.

How can we prove non-trivial lower bounds on the slack of an explicitly given tensor T? Since we do not have any non-trivial lower bounds on the asymptotic rank of any tensor, the only way to prove non-trivial lower bounds on the slack is by upper bounding the asymptotic subrank. Fortunately, we do know powerful methods to upper bound asymptotic subrank, namely the slice rank [Tao16], support functionals [Str91] and quantum functionals [CVZ18]. Using these tools it can be shown for many explicit tensors that the slack is provably strictly larger than 1. In particular, for the Coppersmith–Winograd tensors CW_q ($q \in \mathbb{N}$) that have been used in all recent matrix multiplication algorithms, the slack satisfies $\mathrm{slack}(\mathrm{CW}_q) \geq 1.08$ and so $\omega_{\mathrm{CW}_q} \geq 2.16$.

General upper bounds are known on the possible values of the slack [CVZ19b]. These follow from the fact that $\widetilde{\mathrm{R}}(T) \leq n^{2\omega/3}$ for any $n\times n\times n$ tensor T,[55] and $\widetilde{\mathrm{Q}}(T) \geq n^{2/3}$ for any *balanced*[56] $n\times n\times n$ tensor T [Str88].[57] In particular, for any balanced tensor T the slack is upper bounded by a constant independent of the dimensions of the tensor, namely (remarkably) the matrix multiplication exponent: $1 \leq \mathrm{slack}(T) \leq \omega$. On the other hand, there are examples of (non-balanced) tensors T with $\widetilde{\mathrm{Q}}(T) = 2$ and $\widetilde{\mathrm{R}}(T) = n$, so that the slack is not upper bounded by any constant: $\mathrm{slack}(T) = \log(n)$ [CVZ19b].

[55] This is proven roughly as follows. Using a construction similar to the proof of the well-known inequality $\mathrm{R}(T) \leq n^2$ we can obtain the inequality $T \leq \mathrm{MM}_{n^2,1,1}$ and its permutations $T \leq \mathrm{MM}_{1,n^2,1}$ and $T \leq \mathrm{MM}_{1,1,n^2}$, where $\mathrm{MM}_{a,b,c}$ denotes the rectangular matrix multiplication tensor. Multiplying these inequalities together gives $T^{\otimes 3} \leq \mathrm{MM}_{n^2,n^2,n^2}$, from which the inequality $\widetilde{\mathrm{R}}(T) \leq n^{2\omega/3}$ follows.

[56] A tensor $T \in V_1 \otimes V_2 \otimes V_3$ is called balanced if every flattening $V_i \to (V_j \otimes V_k)$ of T (for $i, j, k \in [3]$ all different) has full rank and contains a full-rank 2-tensor in its image. In particular, over an algebraically closed field generic tensors are balanced.

[57] Neither the asymptotic rank upper bound $\widetilde{\mathrm{R}}(T) \leq n^{2\omega/3}$ nor the asymptotic subrank lower bound $\widetilde{\mathrm{Q}}(T) \geq n^{2/3}$ for balanced tensors are known to be tight. In particular, it is possible that $\widetilde{\mathrm{R}}(T) = n$ and $\widetilde{\mathrm{Q}}(T) = n$ for balanced T.

There are several extensions and versions of the barrier theorem (Theorem 4.23), namely a more precise version that takes into account *how* the Schönhage tau theorem is applied to obtain a matrix multiplication algorithm [CVZ19b], a version that replaces the restriction preorder by a "smaller" preorder that is relevant for the group-theoretic method (the monomial restriction preorder) [CVZ19b], and a version for rectangular matrix multiplication [CGLZ20]. The latter is an example of the usefulness of the asymptotic spectrum of *rectangular* matrix multiplication, as in the statement of this theorem the slack is replaced by a different optimization over points in the asymptotic spectrum of rectangular matrix multiplication.

We expect that barrier theorems of the above type can be interesting also for other problems within the Strassen duality framework, whenever reductions via "intermediate objects" are involved, for example in the setting of graphs (Example 2.20, cf. [PS19]).

4.6. Related duality theories and Positivstellensätze

Strassen's duality fits in a large web of duality theories and theorems known as Positivstellensätze (positive-locus theorems) in real algebraic geometry, which characterize positivity of polynomials on various sets. Moreover, recently several variations on Strassen's duality theory have been introduced in multiple directions and communities, and with various applications. We give in this section a brief overview of these old and new results.

Strassen's original duality theory [Str86, Str87, Str88, Str91] applies to commutative semirings with a Strassen preorder. Strassen uses a notion of good preorder that is slightly more general than our notion of a Strassen preorder. Our definition of a Strassen preorder is simpler and suffices for our main applications. Strassen developed his theory to understand the complexity of matrix multiplication and therefore focuses on the semiring of tensors with the restriction preorder. A large part of his treatment, however, is general, and thus serves as an invitation to apply his theory to other domains. His proof uses the representation theorem of Becker and Schwartz [BS83] as a black box. Our proof in Section 2 and Section 3 integrates the proof of Strassen with the proof of Becker and Schwartz (as was also done in [Zui18]) and moreover makes the role of fractional rank and fractional subrank explicit.

We may naturally compare Strassen's duality theory to the many celebrated Positivstellensätze. Generally, a Positivstellensatz gives necessary conditions (that are in some versions also sufficient conditions) for a polynomial to be non-negative on a certain set. As mentioned, Strassen's theorem can be obtained from the Becker–Schwartz theorem, which is itself a Positivstellensatz. There are many more variations of Positivstellensätze [Pó28, Kri64, Ste74, Put93, Sch91, Rez95, Jac01] that have many interconnections [Mar08, Sch99]. A simple illustrative example is Pólya's Positivstellensatz, which Fritz [Fri21] shows can be obtained from a generalization of Strassen's duality, and which is as follows.

Theorem 4.24 (Pólya [Pó28]. See also Schweighofer [Sch99, Satz 3.6])**.** *Let $f, g \in \mathbb{R}_{\geq 0}[x_1, \dots, x_n]$ be homogeneous polynomials of the same degree. The following two statements are equivalent:*

(i) *For all $x \in \mathbb{R}^n_{\geq 0} \setminus \{0\}$, $f(x) > g(x)$.*

(ii) *There exists an integer $k \geq 0$ such that $(\sum_{i=1}^n x_i)^k f > (\sum_{i=1}^n x_i)^k g$ coefficient-wise. (This means that every coefficient of $(\sum_{i=1}^n x_i)^k (f - g)$ is strictly positive.)*

The strong similarity between Pólya's Positivstellensatz (Theorem 4.24) and Strassen's duality (Theorem 3.41) is hard to miss. Morally, Pólya's coefficient-wise inequality in (ii) plays the role of Strassen's asymptotic preorder $\widetilde{P}$. Pólya's set $\mathbb{R}^n_{\geq 0} \setminus \{0\}$ in (i) plays the role of Strassen's asymptotic spectrum $\mathcal{X}$. It is not hard to see that (ii) implies (i). In fact, Theorem 4.24 remains true when $\mathbb{R}^n_{\geq 0} \setminus \{0\}$ is replaced by the *compact* subset of sum-one elements, making the analogy to Strassen's compact asymptotic spectrum even clearer.[58] We think that a better understanding of the connections between these Positivstellensätze and their proofs will be very fruitful.

Now we discuss newer work. Strassen's duality theory for semirings has been extended by Fritz [Fri21] and Vrana [Vra22] by relaxing the conditions that are required on the Strassen preorder, specifically, the Archimedean property (boundedness condition) in our Definition 2.4. These extensions have applications for example in quantum information theory [PVW22, BV21] and extensions of Pólya's theorem [Fri21].

In a different direction, Fritz [Fri17][59] introduced and developed a duality theory for commutative semi*groups*[60] with a "good" preorder. This theory is closely related to Strassen's. A commutative semigroup is a very basic structure that appears naturally in many settings. In particular, any commutative semiring carries two commutative semigroups inside, namely one under the multiplication operation and one under the addition operation. In particular, a duality theorem for the Shannon capacity of graphs is proven in terms of multiplicative monotone functions. This precedes the stronger duality for the Shannon capacity of graphs in terms of *additive*, multiplicative monotone functions that was obtained in [Zui19] using Strassen's duality. In [Fri17], connections are noted to the utility theorem of von Neumann and Morgenstern [vNM07] in decision theory and economics, and also to a theorem of Lieb and Yngvason [LY99] in the foundations of thermodynamics.

The applications of the various dualities that have appeared in the aforementioned works have been of either combinatorial (graphs, hypergraphs) or linear algebraic nature (tensors, quantum channels). Robere and Zuiddam [RZ22] introduced the Strassen duality paradigm into the field of boolean function complexity. For this they developed a variation of Strassen's duality that applies to preordered semigroups with extra finiteness conditions. This duality relies on (and is essentially equivalent to) linear programming duality. This leads in particular to a general duality theorem for the amortized circuit complexity of boolean functions for a large collection of boolean circuit models, and sheds new light on the work by Razborov [Raz92] on submodular complexity measures, the work of Potechin [Pot17] on catalytic branching programs, and a general concept of catalysts in circuit complexity which is closely related to the multiplicative and additive cancellation properties of closed Strassen preorders as in our Lemma 3.7.

[58] Note however that, different from Strassen's duality, Pólya's theorem is properly a *Positiv*stellensatz in the sense that both (i) as (ii) have *strict inequalities.* The theorem can in fact be phrased using non-strict inequalities $\geq$ instead of $>$ as is done in [Fri21] (as to obtain a Nichtnegativstellensatz, a non-negative-locus theorem), but care has to be taken in rephrasing (ii).

[59] This work was independent of Strassen's work on semirings.

[60] The difference between a semiring and a semigroup is that the former has an addition and a multiplication operation and the latter has just an addition operation.

Part II
Connectivity and convexity of the asymptotic spectrum

Contents

In Part I we have seen how Strassen's duality associates to any semiring $\mathcal{R}$ with a Strassen preorder P a compact topological space $\mathcal{X}$, which is called the asymptotic spectrum and whose points are called spectral points. The main utility of $\mathcal{X}$ is that it gives a dual characterization of the asymptotic preorder associated to P and (as a consequence) the asymptotic rank and asymptotic subrank (Theorem 3.41).

The goal of Part II is to understand *structural properties* of the asymptotic spectrum $\mathcal{X}$. For example, given a collection of spectral points $\phi \in \mathcal{X}$, can we find more? Concretely, it is natural to ask whether $\mathcal{X}$ has a "convex structure" or in other words whether there is a way of "interpolating" between spectral points. Since an asymptotic spectrum $\mathcal{X}$ can be any (compact, Hausdorff) topological space

(Theorem 4.19) the general answer is "no". However, for specific semirings and preorders of interest the answer is "yes". Such "convex structures" will moreover have interesting implications as we will discuss in Part III.

In order to investigate these convexity properties of asymptotic spectra (which are in principle infinite-dimensional) we will study the finite-dimensional compact sets

$$\mathcal{X}(a_1, \dots, a_k) \coloneqq \{(\phi(a_1), \dots, \phi(a_k)) : \phi \in \mathcal{X}\} \subseteq [1, \infty)^k$$

of simultaneous evaluations of the spectral points at chosen semiring elements $a_1, \dots, a_k \in \mathcal{R}$ (as we defined in Definition 3.21).[61] The objective, broadly speaking, is to derive for the sets $\mathcal{X}(a_1, \dots, a_k)$ sufficient and necessary conditions for several topological and geometric properties (connectedness, log-convexity, log-star-convexity).

At a high level, the approach in Part II follows an intricate sequence of ideas of Strassen, which we try to decouple and motivate in turn. As the starting point, we will discuss in Section 5 how log-convex sets $\mathcal{X} \subseteq [1, \infty)^k$ (that are compact) are characterized by *monomial inequalities* $p(x_1, \dots, x_k) \leq q(x_1, \dots, x_k)$, where $p, q \in \mathbb{N}[x_1, \dots, x_k]$ are monomials. This connection between convexity and monomial inequalities motivates the definition in Section 6 of the *monomial partial order* on polynomials. This is an ingenious and subtle definition Strassen made, which in turn is based on a *type decomposition* of powers of polynomials. This is a natural decomposition that is finer than the usual monomial decomposition and which is suited to the asymptotics that high powers generate. In Section 7 we will then see that connected sets (or rather log-convex ones) are precisely those for which the monomial and pointwise partial order on polynomials agree. Finally, in Section 8 we will develop tools, mainly the *anchor method*, to help establish inequalities in the monomial partial order from the pointwise one. In Section 9 we will discuss implications of convexity properties of asymptotic spectra and other related results about convexity in this context.

The anchor method, an important abstraction of Strassen's original argument, provides a way of proving that a compact set $\mathcal{X} \subseteq [1, \infty)^k$ is log-star-convex by finding a single point in $\mathcal{X}$ with special properties, the *anchor*. While at first we will discuss this without reference to asymptotic spectra, the full version of the anchor method provides a method to prove that a set of the form $\mathcal{X}(a_1, \dots a_k) \subseteq [1, \infty)^k$ for a given semiring $\mathcal{R}$, Strassen preorder P and semiring elements $a_1, \dots, a_k \in \mathcal{R}$, is log-star-convex, and explicitly utilizes the preorder P to do so. This is the only place in Part II where we explicitly use Strassen duality.

Throughout our exposition, we will focus on (and provide full proofs only for) the one-dimensional case $\mathcal{X}(a) \subseteq [1, \infty)$ for $a \in \mathcal{R}$, since in this case convexity and connectedness coincide, it requires simpler notation, and at the same time already captures the proof ideas for the high-dimensional case $\mathcal{X}(a_1, \dots, a_k) \subseteq [1, \infty)^k$. A large part of the discussion will not be specific to the sets $\mathcal{X}(a_1, \dots, a_k) \subseteq [1, \infty)^k$ but will rather apply to *general* compact sets $\mathcal{X} \subseteq [1, \infty)^k$.

Summarizing, Part II reduces proofs of connectedness and convexity properties of the spectrum $\mathcal{X}$ at hand to proving that it contains an anchor. Looking ahead, in Part III we will complete the picture in two ways. First, we will show how to find anchors for spectra of specific semirings within the semiring of tensors, in particular matrix multiplication (following Strassen) and a new setting of tensor networks, thus obtaining a proof of these connectivity properties for them. Second, in Part III

[61] This is almost equivalent to studying the asymptotic spectra of finitely generated semirings (simply taking the subsemiring of $\mathcal{R}$ generated by the a_i), except that because of the way $\mathcal{X}(a_1, \dots, a_k)$ is "encoded", $\mathcal{X}(a_1, \dots, a_k)$ does depend on the choice of those generators. In other words, $\mathcal{X}(a_1, \dots, a_k)$ and $\mathcal{X}(b_1, \dots, b_\ell)$ may be different even if the a_i generate the same semiring as the b_i.

we will discuss the implications of convexity properties of sets $\mathcal{X}(a_1, \ldots, a_k)$ in (the aforementioned and other) applications of the theory of asymptotic spectra.

5. Convexity and monomial inequalities

The connection between convexity and monomial inequalities that we establish here will motivate the definitions in Section 6, and the development of other tools towards proving connectivity in Sections 7 and 8. This section may safely be skipped as we will make sure to recall the definitions when they are used later.

Recall that Strassen's duality associates to any semiring $\mathcal{R}$ with a Strassen preorder P a compact topological space $\mathcal{X}$, the asymptotic spectrum. We want to understand more about the structure of $\mathcal{X}$, which we can naturally do by studying, for any elements $a_1, \ldots, a_k \in \mathcal{R}$, the compact set $\mathcal{X}(a_1, \ldots, a_k) \coloneqq \{(\phi(a_1), \ldots, \phi(a_k)) : \phi \in \mathcal{X}\} \subseteq [1, \infty)^k$, the asymptotic spectrum of $a_1, \ldots, a_k$ (Definition 3.21). In this and following sections we take a step back, forget about semirings, preorders and Strassen duality, and study any compact sets $\mathcal{X} \subseteq [1, \infty)^k$.

We begin even simpler and let $\mathcal{X} \subseteq [1, \infty)$ be any *one-dimensional* compact set. It is not hard to see that $\mathcal{X}$ is naturally characterized by polynomial inequalities, as we will discuss in detail in Section 5.1. Building on this, we will in Section 5.2 discuss how *monomial* inequalities characterize when $\mathcal{X}$ is *connected*. Finally, in Section 5.3 we discuss the straightforward multivariate generalization of this for compact sets $\mathcal{X} \subseteq [1, \infty)^k$ where in particular the monomial inequalities characterize *log-convex* sets.[62]

5.1. Polynomial inequalities characterize closed sets

We introduce some general notation, which we will continue using in later sections. Let f and g be functions $[1, \infty) \to \mathbb{R}_{\geq 0}$. These functions will be given by polynomials in a moment. If for all $s \in \mathcal{X}$ the inequality $f(s) \leq g(s)$ holds, then we write $f \leq g$ on $\mathcal{X}$. We call the set of elements in $[1, \infty)$ where the inequality $f \leq g$ holds the *non-negative locus* of the inequality, and we denote this set by

$$\mathcal{S}(f \leq g) \coloneqq \{s \in [1, \infty) : f(s) \leq g(s)\}.$$

Let $\mathbb{R}_{\geq 0}[x]$ be the set of polynomials in the variable x with non-negative real coefficients, which we naturally view as functions $[1, \infty) \to \mathbb{R}_{\geq 0}$.

We begin with a simple polynomial description of closed subsets of $[1, \infty)$. Note that a subset of $[1, \infty)$ is closed if and only if it is a union of closed intervals with potentially one unbounded interval $[s, \infty)$. In fact, we will characterize the closure of any subset $\mathcal{X} \subseteq [1, \infty)$ using polynomials. For any $\mathcal{X} \subseteq [1, \infty)$ we directly find the inclusion

$$\mathcal{X} \subseteq \bigcap \mathcal{S}(f \leq g), \tag{5.1}$$

where the intersection is over all polynomials $f, g \in \mathbb{R}_{\geq 0}[x]$ such that $f \leq g$ on $\mathcal{X}$. Of course each set $\mathcal{S}(f \leq g)$ is closed, and therefore the right-hand side of (5.1) is closed. Thus, clearly, if $\mathcal{X}$ is not closed, then the inclusion in (5.1) is strict. This naturally leads us to the following characterization of the closure $\overline{\mathcal{X}} \subseteq [1, \infty)$ of $\mathcal{X}$ (which we will use later on).

[62] Connectedness and log-convexity coincide in one dimension.

Lemma 5.1. *Let $\mathcal{X} \subseteq [1, \infty)$. Then the closure $\overline{\mathcal{X}}$ of $\mathcal{X}$ is equal to the intersection of non-negative loci*

$$\overline{\mathcal{X}} = \bigcap \mathcal{S}(f \leq g),$$

ranging over all polynomials $f, g \in \mathbb{R}_{\geq 0}[x]$ such that $f \leq g$ on $\mathcal{X}$. Equivalently, the intersection may be taken over only the polynomials $f, g \in \mathbb{N}[x]$ such that $f \leq g$ on $\mathcal{X}$.

Proof. We first prove $\subseteq$. Let $s \in \overline{\mathcal{X}}$. If $f \leq g$ on $\mathcal{X}$, then also $f(s) \leq g(s)$ by continuity, and so $s \in \mathcal{S}(f \leq g)$. We now prove $\supseteq$. Suppose that $s \notin \overline{\mathcal{X}}$. Then $\varepsilon < (x-s)^2$ on $\overline{\mathcal{X}}$ for small enough $\varepsilon > 0$, since $\overline{\mathcal{X}}$ is closed. Let $f(x) = 2xs + \varepsilon$ and $g(x) = x^2 + s^2$. Then $f \leq g$ on $\overline{\mathcal{X}}$ and $f(s) > g(s)$. This proves that the closure of $\mathcal{X}$ is the intersection of $\mathcal{S}(f \leq g)$ over all $f, g, \in \mathbb{R}_{\geq 0}[x]$ such that $f \leq g$ on $\mathcal{X}$. By ensuring that ε is rational and after an appropriate scaling of both f and g, we may equivalently take the intersection over $f, g \in \mathbb{N}[x]$ such that $f \leq g$ on $\mathcal{X}$. □

It is customary in semi-algebraic geometry to study the non-negative locus $\mathcal{S}(0 \leq h)$ for polynomials $h \in \mathbb{R}[x]$ with potentially negative coefficients, which naturally corresponds to our notion $\mathcal{S}(f \leq g)$ by choosing $f, g \in \mathbb{R}_{\geq 0}[x]$ such that $h = g - f$. Our notation $\mathcal{S}(f \leq g)$ will facilitate the discussion from Section 7 onwards.

To conclude, the subsets $\mathcal{X} \subseteq [1, \infty)$ that are closed are precisely those that are characterized by polynomial inequalities, and they are the intersection of the non-negative loci $\mathcal{S}(f \leq g)$ over all polynomials $f, g \in \mathbb{N}[x]$ such that $f \leq g$ on $\mathcal{X}$.

5.2. Monomial inequalities characterize connected sets

Next we give a polynomial description of connected subsets $\mathcal{X} \subseteq [1, \infty)$. In order to have a polynomial description we must, by the previous subsection, assume that $\mathcal{X}$ is closed. We will also assume that $\mathcal{X}$ is bounded for simplicity of notation and since that will be the case for our applications. In fact, we will give a polynomial description of the *convex hull* of any closed and bounded subset $\mathcal{X} \subseteq [1, \infty)$. So let $\mathcal{X} \subseteq [1, \infty)$ be closed and bounded. The convex hull of $\mathcal{X}$ is given by

$$\operatorname{conv}(\mathcal{X}) = [\min(\mathcal{X}), \max(\mathcal{X})].$$

In terms of sets of the form $\mathcal{S}(f \leq g)$, we see directly that

$$\operatorname{conv}(\mathcal{X}) = \mathcal{S}(\min(\mathcal{X}) \leq x) \cap \mathcal{S}(x \leq \max(\mathcal{X})).$$

With a simple argument we can strengthen this in such a way that the right-hand side ranges over all monomial inequalities on $\mathcal{X}$, as follows.

Lemma 5.2. *Let $\mathcal{X} \subseteq [1, \infty)$ be closed and bounded. Then the convex hull* $\operatorname{conv}(\mathcal{X})$ *of $\mathcal{X}$ is equal to the intersection of non-negative loci*

$$\operatorname{conv}(\mathcal{X}) = \bigcap \mathcal{S}(cx^i \leq dx^j),$$

ranging over the $c, d \in \mathbb{R}_{\geq 1}$ and $i, j \in \mathbb{N}$ such that $cx^i \leq dx^j$ on $\mathcal{X}$. *Equivalently, the intersection may be taken over just the non-negative integers $c, d \in \mathbb{N}$ and $i, j \in \mathbb{N}$ such that $cx^i \leq dx^j$* on $\mathcal{X}$.

Proof. The inclusion $\supseteq$ follows directly from the aforementioned simple observation that

$$\mathrm{conv}(\mathcal{X}) = \mathcal{S}(\min(\mathcal{X}) \leq x) \cap \mathcal{S}(x \leq \max(\mathcal{X})).$$

We prove the other inclusion $\subseteq$. Let $s \in \mathrm{conv}(\mathcal{X})$. Then we have $\min(\mathcal{X}) \leq s \leq \max(\mathcal{X})$. Suppose that $cx^i \leq dx^j$ on $\mathcal{X}$. Either $c/d \leq x^{j-i}$ on $\mathcal{X}$ with $j - i \geq 0$, so $c/d \leq \min(\mathcal{X})^{j-i}$ and thus $c/d \leq s^{j-i}$; or $x^{i-j} \leq d/c$ with $i - j \geq 0$, so $\max(\mathcal{X})^{i-j} \leq d/c$ and thus $s^{i-j} \leq d/c$. In both cases we have $cs^i \leq ds^j$, which proves the claim.

The same statement with the intersection ranging over polynomials with non-negative integer coefficients $c, d \in \mathbb{N}$ holds by a rational approximation argument. ☐

Remark 5.3. In fact, it is clear from the proof that in Lemma 5.2 it suffices to let the intersection range over $cx^i \leq dx^j$ such that $(i, j) = (1, 0)$ or $(i, j) = (0, 1)$. Namely, from $c/d \leq s^{j-i}$ it follows (via taking $|i-j|$th roots) that $(c/d)^{1/(j-i)} \leq s$ and from $s^{i-j} \leq d/c$ it follows that $s \leq (d/c)^{1/(i-j)}$. Thus we have an even stronger characterization of $\mathrm{conv}(\mathcal{X})$ in terms of linear inequalities, namely:

$$\mathrm{conv}(\mathcal{X}) = \bigcap_{\substack{c,d \in \mathbb{R}_{\geq 1}: \\ dx \leq c \text{ on } \mathcal{X}}} \mathcal{S}(dx \leq c) \cap \bigcap_{\substack{c,d \in \mathbb{R}_{\geq 1}: \\ c \leq dx \text{ on } \mathcal{X}}} \mathcal{S}(c \leq dx)$$

and similarly in terms of linear inequalities with non-negative integral coefficients:

$$\mathrm{conv}(\mathcal{X}) = \bigcap_{\substack{c,d \in \mathbb{N}: \\ dx \leq c \text{ on } \mathcal{X}}} \mathcal{S}(dx \leq c) \cap \bigcap_{\substack{c,d \in \mathbb{N}: \\ c \leq dx \text{ on } \mathcal{X}}} \mathcal{S}(c \leq dx).$$

Remark 5.4. In this section we are looking at compact subsets $\mathcal{X} \subseteq [1, \infty)$ for their own sake, but let us in this remark briefly consider the special case of compact subsets $\mathcal{X}(a) \subseteq [1, \infty)$ given by the theory of asymptotic spectra. We know already that generally $\mathcal{X}(a)$ is not connected. Indeed, $\mathcal{X}(a)$ can be any compact subset of $[1, \infty)$. Suppose that $\mathcal{X}(a)$ is connected. Then from the duality theorem we know that $\mathcal{X}(a)$ is the closed interval with endpoints given by the asymptotic subrank and the asymptotic rank of a,

$$\mathcal{X}(a) = [\widetilde{\mathrm{Q}}_P(a), \widetilde{\mathrm{R}}_P(a)]. \tag{5.2}$$

Now Lemma 5.2 says that $\mathcal{X}(a)$ is characterized by monomial inequalities. This means in particular that the asymptotic spectrum can be completely understood by studying inequalities $ca^i \lesssim_{P} da^j$ for $c, d \in \mathbb{N}$. Indeed, it is clear already from (5.2) that it even suffices to study inequalities of the form $a^n \leq_P r$ and $r \leq_P a^n$ for $r, n \in \mathbb{N}$ since these characterize the asymptotic rank $\widetilde{\mathrm{R}}_P(a)$ and the asymptotic subrank $\widetilde{\mathrm{Q}}_P(a)$.

5.3. Multivariate monomial inequalities characterize log-convex sets

In Section 5.2 we have seen that for compact sets $\mathcal{X} \subseteq [1, \infty)$ the convex hull is characterized by the monomial inequalities. Now we will give the natural high-dimensional extension (of which we leave the proofs to the reader). Let $\mathbb{R}_{\geq 0}[x_1, \ldots, x_k]$ be the set of polynomials in the variables $x_1, \ldots, x_k$ with non-negative real coefficients. Extending the previous definition, for polynomials $f, g \in \mathbb{R}_{\geq 0}[x_1, \ldots, x_k]$ we call the set of elements in $[1, \infty)^k$ where the inequality $f \leq g$ holds the *non-negative locus* of the inequality, and we denote this set by

$$\mathcal{S}(f \leq g) \coloneqq \{s \in [1, \infty)^k : f(s_1, \ldots, s_k) \leq g(s_1, \ldots, s_k)\}.$$

Lemma 5.5 (Polynomial inequalities characterize the closure)**.** *Let $\mathcal{X} \subseteq [1,\infty)^k$. Then the closure is given by $\overline{\mathcal{X}} = \bigcap \mathcal{S}(f \leq g)$ where the intersection is ranging over all polynomials $f, g \in \mathbb{R}_{\geq 0}[x_1, \ldots, x_k]$ such that $f \leq g$ on $\mathcal{X}$. Equivalently, the intersection may be taken over only the polynomials $f, g \in \mathbb{N}[x_1, \ldots, x_k]$ such that $f \leq g$ on $\mathcal{X}$.*

For $\mathcal{X} \subseteq [1,\infty)^k$ we define $\log \mathcal{X} \coloneqq \{(\log s_1, \ldots, \log s_k) : s \in \mathcal{X}\}$. For $\mathcal{Y} \subseteq [0,\infty)^k$ we define $2^{\mathcal{Y}} \coloneqq \{(2^{t_1}, \ldots, 2^{t_k}) : t \in \mathcal{Y}\}$. We say $\mathcal{X} \subseteq [1,\infty)^k$ is *log-convex* if $\log \mathcal{X}$ is convex. We define $\mathrm{logconv}(\mathcal{X}) \coloneqq 2^{\mathrm{conv}(\log(\mathcal{X}))}$ to be the *log-convex hull* of $\mathcal{X}$.

Lemma 5.6 (Monomial inequalities characterize the log-convex hull)**.** *Let $\mathcal{X} \subseteq [1,\infty)^k$ be compact. Then*

$$\mathrm{logconv}(\mathcal{X}) = \bigcap \mathcal{S}(f \leq g)$$

where the intersection ranges over all monomials $f, g \in \mathbb{R}_{\geq 0}[x_1, \ldots, x_k]$ such that $f \leq g$ on $\mathcal{X}$. Equivalently, the intersection may be restricted to monomials in $\mathbb{N}[x_1, \ldots, x_k]$, or even stronger, monomials in $\mathbb{N}[x_1, \ldots, x_k]$ such that precisely one is constant and one is linear.

Proof. We give the proof idea. Let $\mathcal{Y} = \log \mathcal{X}$. Then $\mathrm{conv}(\mathcal{Y})$ is characterized the linear inequalities on $\mathcal{Y}$. Exponentiating these inequalities gives the claim. By rational approximation we get the monomials to have non-negative integer coefficients. □

This concludes this section, in which we introduced some basic notions and connections. The take-away message is that polynomial inequalities provide a natural way to understand compact sets $\mathcal{X} \subseteq [1,\infty)^k$ (and in particular, the asymptotic spectrum $\mathcal{X}(a_1, \ldots, a_k)$ of $a_1, \ldots, a_k$) and that there is a close relation between *monomial* inequalities and log-convexity. These ideas are the starting point for the remaining sections in Part II.

6. Type decomposition of polynomials and the monomial partial order

In the previous section we have seen that an asymptotic spectrum $\mathcal{X}$, or any closed subset $\mathcal{X} \subseteq [1,\infty)$, is connected if and only if it is characterized by the monomial inequalities on $\mathcal{X}$. We are interested in the problem of proving, given a semiring and a Strassen preorder, that the asymptotic spectrum is connected. In this section we introduce a basic and general building block for this, which is a *type decomposition* for powers of polynomials (Section 6.1). The main application of this type decomposition is the definition of the *monomial partial order* (in Section 6.2), and its variants (Sections 6.3 and 6.4).

At a high level, for any polynomial f the type decomposition that we will define is a natural way to express the power f^n as a sum $\sum_p [f^n]_p$ of polynomially in n many polynomials $[f^n]_p$ (type components).[63] For any two polynomials f and g, the monomial partial order comes down to saying that f is at most g (on a set $\mathcal{X}$) if for every $n \in \mathbb{N}$ and for every component $[f^n]_p$ appearing in the decomposition of f^n there is a component $[g^n]_q$ appearing in the decomposition of g^n such that $[f^n]_p$ is at most $[g^n]_q$ pointwise on $\mathcal{X}$, up to a subexponential slack factor $2^{o(n)}$. We may call the set $\mathcal{X}$

[63] This is analogous to the standard (but powerful) idea in information theory of the *method of types* [CT12], in which one takes n-samples from a probability distribution and collects samples into type classes according to the frequencies of the symbols. Here an analogous thing happens with monomials of a polynomial raised to the nth power (with appropriate weights added).

on which f is at most g in the monomial partial order, the *domain* of this inequality. The main theorem that we will prove is the important property of the monomial partial order that its domain is connected (and log-convex in the multivariate case). Namely, if f is at most g in the monomial partial order on $\mathcal{X}$, then also f is at most g in the monomial partial order on the *log-convex hull* of $\mathcal{X}$. This property is the main reason for defining the monomial partial order.

To develop the above, it turns out to be mathematically convenient to consider an equivalent definition of the monomial partial order that is phrased in terms of the functions $f_p = \lim_{n\to\infty}([f^n]_p)^{1/n}$ which measure the rate of growth of type components. We introduce these functions together with the type decomposition in Section 6.1. In Section 6.2 we give the definition of the monomial partial order in terms of f_p and prove the main theorem (Theorem 6.10). In Section 6.3 we prove that this definition coincides with the aforementioned "finite" definition of the monomial partial order that is directly in terms of $[f^n]_p$. All of this we will first discuss for univariate polynomials before discussing the straightforward multivariate extension in Section 6.4.

6.1. Asymptotic behavior of polynomials and the type decomposition

In this section, let $f = \sum_{i=0}^{d} f_i x^i \in \mathbb{N}[x]$ be a univariate polynomial of degree d with non-negative integer coefficients $f_i \in \mathbb{N}$.

Definition 6.1. Let $p = (p_0, \ldots, p_d) \in \mathbb{R}^{d+1}$ be a probability vector, which we think of as a probability distribution over the $d+1$ monomials of f, such that the monomial $f_i x^i$ is assigned probability p_i. Let $H(p) \coloneqq -\sum_i p_i \log p_i$ denote the Shannon entropy of p. We denote by $\operatorname{supp}(f) \coloneqq \{i \in \{0,\ldots,d\} : f_i \neq 0\}$ the support of f. We define the function

$$f_p(x) \coloneqq 2^{H(p)} \prod_i f_i^{p_i} x^{i p_i}$$

where the product is over $i \in \operatorname{supp}(f)$.

In the following, all summations and products involving the coefficients of polynomials will be over the support of the polynomial at hand. We see that our function $f_p(x)$ is defined as the p-weighted geometric mean of the monomials $f_i x^i$ multiplied by a scaling factor depending only on p. The reason for the scaling will become clear soon. We will make frequent use of the logarithm of $f_p(x)$,

$$\log f_p(x) = H(p) + \sum_i (p_i \log f_i + i\, p_i \log x).$$

Two things will be important for us. First, the function $\log f_p(x)$ is concave in p, because the entropy is concave, and because all the other terms depend linearly on p. Second, the function $\log f_p$ is linear in $\log x$. That is, the function

$$\log f_p(2^y) = H(p) + \sum_i (p_i \log f_i + i\, p_i y)$$

(obtained by substituting $y = \log x$) is linear in $y \in [0, \infty)$.

The reason for defining $f_p(x)$ is that the family of functions $\{f_p(x)\}_p$, parametrized by all probability vectors $p \in \mathbb{R}^{d+1}$, describes the asymptotic behavior of f, as we will explain next.

We denote by $\mathcal{P}(f) \subseteq \mathbb{R}^{d+1}$ the set of all probability vectors $(p_0, \ldots, p_d)$ of length $d+1$. For any $n \in \mathbb{N}$ we call a probability vector $p = (p_0, \ldots, p_d)$ an *n-type* if every coefficient p_i is an integer multiple of $1/n$. We denote by $\mathcal{P}_n(f) \coloneqq \{(p_0,\ldots,p_d) \in \mathcal{P}(f) : \forall i,\ n\cdot p_i \in \mathbb{N}\}$ the subset of n-types.[64] We may write the nth power of f as the sum

$$f^n = \sum_{p\in\mathcal{P}_n(f)} \binom{n}{pn} \prod_i f_i^{p_i n} x^{i p_i n}, \tag{6.1}$$

where the product is over $i \in \operatorname{supp}(f)$, and where $\binom{n}{pn} = n!/((p_0 n)!\cdots(p_d n)!)$ denotes a multinomial coefficient. In other words, (6.1) is the result of taking the nth power of f and "not fully collecting terms" (see also Remark 6.3 for a formalization of that idea).

Definition 6.2 (The type decomposition of f^n into type components $[f^n]_p$)**.** We call (6.1) the *type decomposition of* f^n. We denote the summands in (6.1) by $[f^n]_p$, that is,

$$[f^n]_p \coloneqq \binom{n}{pn} \prod_i f_i^{p_i n} x^{i p_i n}.$$

We call $[f^n]_p$ the *p-component of* f^n.

Note that $[f^n]_p$ is a single monomial of degree $\sum_i i p_i n$. For a fixed $j \in \mathbb{N}$ there may be multiple probability vectors $p \in \mathcal{P}_n(f)$ for which $\sum_i i p_i n = j$. Therefore, the type decomposition is finer than the monomial decomposition $f^n = \sum_j (f^n)_j\, x^j$. (Indeed, *much* finer, as the monomial decomposition has roughly nd terms, while the type decomposition has roughly n^d terms.)

Remark 6.3. The type decomposition can be seen as a direct sum decomposition if we think of the monomials of f as having different colors. More precisely, we rewrite $f = \sum_{i=0}^d f_i x^i \in \mathbb{N}[x]$ as a linear form in variables $x_0,\ldots,x_d$ in such a way that each degree gets its own variable:

$$f = \sum_{i=0}^d f_i x_i \in \mathbb{N}[x_0, x_1, \ldots, x_d]_1.$$

Then

$$f^n = \sum_{p\in\mathcal{P}_n(f)} \binom{n}{pn} \prod_i f_i^{p_i n} \prod_i x_i^{p_i n}$$

is a direct sum decomposition in the space $\mathbb{N}[\prod_i x_i^{p_i n} : p \in \mathcal{P}_n(f)]$.

The function $f_p(x)$ that we defined in Definition 6.1 measures the asymptotic size of $[f^n]_p$ when n goes to infinity, in the following sense.

Lemma 6.4. *For every $p \in \mathcal{P}_n(f)$ we have*

$$f_p(x)^{n-o(n)} \le [f^n]_p(x) \le f_p(x)^n$$

uniformly in $x \in \mathbb{R}_{\ge 0}$. In particular, the function $f_p(x)$ is the limit

$$f_p(x) = \lim_{k\to\infty} ([f^{kn}]_p(x))^{\frac{1}{kn}}$$

uniformly in $x \in \mathbb{R}_{\ge 0}$. (Note that every n-type is also a kn-type, that is, for every $k \in \mathbb{N}$ the inclusion $\mathcal{P}_n(f) \subseteq \mathcal{P}_{kn}(f)$ holds. Therefore, $[f^{kn}]_p(x)$ is defined for every $p \in \mathcal{P}_n(f)$ and $k \in \mathbb{N}$.)

[64] Note that $\mathcal{P}(f)$ and $\mathcal{P}_n(f)$ depend only on the degree d of f. It will be useful, however, to be able to refer to these sets without worrying about the degree of f, hence our notation.

Proof. We consider the ratio

$$\frac{[f^n]_p}{f_p(x)^n} = \frac{\binom{n}{pn}\prod_i f_i^{p_i n} x^{ip_i n}}{2^{H(p)}\prod_i f_i^{p_i n} x^{ip_i n}} = \frac{\binom{n}{pn}}{2^{H(p)n}}.$$

The claim follows from the standard asymptotic relation [CT12, Theorem 11.1.3] between multinomial coefficients and the Shannon entropy: $2^{H(p)n-o(n)} \leq \binom{n}{pn} \leq 2^{H(p)n}$. ☐

Lemma 6.5. *For every $p \in \mathcal{P}(f)$ and every $s \in \mathbb{R}_{\geq 0}$ we have $f_p(s) \leq f(s)$.*

Proof. By the AM-GM inequality we have for any $g_i, s \in \mathbb{R}_{\geq 0}$ and any probability vector p that

$$\prod_i (g_i s^i)^{p_i} \leq \sum_i p_i g_i s^i.$$

Restrict the product and sum above to those i with $p_i \neq 0$. Let $g_i = p_i^{-1} f_i$. Then

$$f_p(s) = 2^{H(p)} \prod_i (f_i s^i)^{p_i} = \prod_i (p_i^{-1} f_i s^i)^{p_i} \leq \sum_i f_i s^i = f(s).$$

This proves the claim.

Alternatively, the claim follows from Lemma 6.4. Indeed, we may assume that p is an n-type for some $n \in \mathbb{N}$, since $f_p(x)$ is continuous in p. Then for any $k \in \mathbb{N}$ and any $s \in \mathbb{R}_{\geq 0}$ we have that $[f^{kn}]_p(s) \leq f^{kn}(s)$. Taking the kn-th root and letting k go to infinity (Lemma 6.4) we find that $f_p(s) \leq f(s)$ for all $p \in \mathbb{R}_{\geq 0}$. ☐

Crucially, the number of summands appearing in the type decomposition of f^n grows *polynomially* in n, as we make precise in the following lemma. This is where using asymptotics (in n) is powerful, as its allowing of subexponential slack factors swallows polynomial factors easily (and this fact we will be using immediately).

Lemma 6.6 (There are polynomially many n-types.)**.** *For any polynomial f of degree d, the number of n-types $p = (p_0, \ldots, p_d)$ is upper bounded by $|\mathcal{P}_n(f)| \leq (n+1)^{d+1} = \mathrm{poly}(n)$.*

Proposition 6.7. *For every $s \in \mathbb{R}_{\geq 0}$ we have $\max_{p \in \mathcal{P}(f)} f_p(s) = f(s)$.*

Proof. We fix the element s. We already know from Lemma 6.5 that $\max_{p\in\mathcal{P}(f)} f_p(s) \leq f(s)$. It remains to prove that $f(s) \leq \max_{p\in\mathcal{P}(f)} f_p(s)$. There are only polynomially many n-types in $\mathcal{P}(f)$ (Lemma 6.6) and so

$$\begin{aligned} f^n(s) &\leq \mathrm{poly}(n) \max_q [f^n]_q(s) \\ &\leq \mathrm{poly}(n) \max_q (f_q(s))^n, \end{aligned}$$

where the maximum is over $q \in \mathcal{P}_n(f)$. Thus for every $n \in \mathbb{N}$

$$f(s) \leq \mathrm{poly}(n)^{\frac{1}{n}} \max_q f_q(s)$$

where the maximum is still over $q \in \mathcal{P}_n(f)$. By letting n go to infinity we get

$$f(s) \leq \max_p f_p(s)$$

where the maximum is taken over all $p \in \mathcal{P}(f)$. ☐

6.2. The monomial partial order on polynomials

Now that we have types $[f^n]_p$ (which are essentially distributions over monomials) and their asymptotic growth f_p (which we may think of as "generalized monomials"), we want to compare polynomials according to their weightiest monomial (for each possible distribution on the coefficients). We are motivated by Section 5 in which comparison between monomials characterized connectedness. What we aim to do is to use asymptotics to move from polynomial inequalities to such "monomial inequalities", paying in adding quantifiers (which we will have to deal with later). To this end we will define the monomial partial order (Definition 6.8), and we will then prove its main property that the "domain" of such inequalities is connected (Theorem 6.10).

Definition 6.8 (Monomial partial order $\leq_{\mathrm{M}}$)**.** For polynomials $f, g \in \mathbb{N}[x]$ and $\mathcal{X} \subseteq [1, \infty)$ we say that f *is at most* g *monomially on* $\mathcal{X}$ and we write

$$f \leq_{\mathrm{M}} g \text{ on } \mathcal{X},$$

if for every $p \in \mathcal{P}(f)$ there exists a $q \in \mathcal{P}(g)$ such that for all $s \in \mathcal{X}$ it holds that $f_p(s) \leq g_q(s)$.

Note that the monomial partial order is indeed transitive: if $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$ and $g \leq_{\mathrm{M}} h$ on $\mathcal{X}$, then $f \leq_{\mathrm{M}} h$ on $\mathcal{X}$.

We may call the set $\mathcal{X}$ so that $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$ holds, the *domain* of this specific inequality. We stress the subtle behavior of the domain that from $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$ and $f \leq_{\mathrm{M}} g$ on $\mathcal{Y}$ we *cannot* conclude that $f \leq_{\mathrm{M}} g$ on $\mathcal{X} \cup \mathcal{Y}$. This is because of the order of the quantifiers $\exists q$ and $\forall s$ in Definition 6.8.[65]

The monomial partial order is stronger than the pointwise partial order:

Lemma 6.9. *If $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$, then $f \leq g$ on $\mathcal{X}$.*

Proof. For every $s \in \mathcal{X}$ there is an element $q \in \mathcal{P}(g)$ such that

$$f(s) = \max_p f_p(s) \leq g_q(s) \leq g(s)$$

by Lemma 6.5 and Proposition 6.7. □

In contrast to the pointwise partial order on polynomials, from which we cannot infer an inequality on any point from inequalities on other points (except for limit points, cf. Section 5), the monomial partial order allows this—any inequality at the extreme points of an interval interpolate automatically to the whole interval:

Theorem 6.10 (Connectedness[66] of the domain where $f \leq_{\mathrm{M}} g$ holds)**.** *Let $s \leq t \in [1, \infty)$ and let $f, g \in \mathbb{N}[x]$. If $f \leq_{\mathrm{M}} g$ on $\{s, t\}$, then $f \leq_{\mathrm{M}} g$ on $[s, t]$.*

Expanding the statement of Theorem 6.10, it says that if for any $p \in \mathcal{P}(f)$ there is a $q \in \mathcal{P}(g)$ such that $f_p(s) \leq g_q(s)$ and $f_p(t) \leq g_q(t)$, then for any $p \in \mathcal{P}(f)$ there is a (the same) $q \in \mathcal{P}(g)$ such that for all $r \in [s, t]$ we have $f_p(r) \leq g_q(r)$.

[65] For this reason we do not define a locus $\mathcal{S}(f \leq_{\mathrm{M}} g)$ for the monomial partial order analogously to the definition of the non-negative locus $\mathcal{S}(f \leq g)$ of Section 5.1.

[66] It will be clear from the proof that the connectedness follows from log-convexity of the domain. While in the current setting of univariate polynomials connectedness and log-convexity coincide, we will see in Section 6.4 that for multivariate polynomials the analogous lemma gives *log-convexity* of the domain.

Proof. In this proof for the first time the fact that the function $\log f_p(2^y)$ is linear in y enters the picture. By assumption, we have that for every p there is a q such that both $f_p(s) \leq g_q(s)$ and $f_p(t) \leq g_q(t)$. The functions $\log f_p(2^y)$ and $\log g_q(2^y)$ are linear in y. Therefore, we have that the interpolated inequality

$$f_p(2^{\alpha \log s + (1-\alpha)\log t}) \leq g_q(2^{\alpha \log s + (1-\alpha)\log t})$$

holds for every $\alpha \in [0, 1]$. Thus $f \leq_{\mathrm{M}} g$ on $[s, t]$. $\square$

6.3. The finite monomial partial order on polynomials

In the discussion of the monomial partial order $f \leq_{\mathrm{M}} g$ it was convenient to work with general probability vectors $p \in \mathcal{P}(f)$ and $q \in \mathcal{P}(g)$ as in Definition 6.8. In particular, we defined the monomial partial order in terms of the functions f_p. In our upcoming discussion of anchors and application to matrix multiplication and tensor networks, however, it will at some point be more convenient to work only with n-types $p \in \mathcal{P}_n(f)$ and $q \in \mathcal{P}_n(g)$ for n large enough. We will explicitly define a *finite* monomial partial order $\leq_{\mathrm{FM}}$ for this, in terms of the type components $[f^n]_p$. Using Lemma 6.4 and the fact that n-types approximate general probability vectors we give a routine proof that $\leq_{\mathrm{M}}$ and $\leq_{\mathrm{FM}}$ are equivalent.

Definition 6.11 (Finite monomial partial order $\leq_{\mathrm{FM}}$)**.** For polynomials $f, g \in \mathbb{N}[x]$ and $\mathcal{X} \subseteq [1, \infty)$ we say that f *is at most* g *finitely monomially on* $\mathcal{X}$, and we write

$$f \leq_{\mathrm{FM}} g \ \textit{on}\ \mathcal{X},$$

if for every $\varepsilon > 0$ and for large enough n we have that for every $p \in \mathcal{P}_n(f)$ there is a $q \in \mathcal{P}_n(g)$ such that

$$[f^n]_p \leq [g^n]_q \cdot 2^{\varepsilon n} \text{ on } \mathcal{X}.$$

Lemma 6.12 (Equivalence of the monomial and finite monomial partial orders)**.** *For any polynomials $f, g \in \mathbb{N}[x]$ and bounded subset $\mathcal{X} \subseteq [1, \infty)$ we have $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$ if and only if $f \leq_{\mathrm{FM}} g$ on $\mathcal{X}$.*

Proof. Suppose that $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$. Then for every $p \in \mathcal{P}(f)$ there is a $q \in \mathcal{P}(g)$ such that $f_p(x) \leq g_q(x)$ on $\mathcal{X}$. We may approximate probability vectors with n-types to get the following. For every $n \in \mathbb{N}$ for every $p \in \mathcal{P}_n(f)$ there is a $q \in \mathcal{P}_n(g)$ such that

$$f_p(x) \leq g_q(x) \cdot 2^{o(1)} \text{ on } \mathcal{X}.$$

We know (Lemma 6.4) that for every $p \in \mathcal{P}_n(f)$ and $q \in \mathcal{P}_n(g)$ we have

$$[f^n]_p(x) \leq f_p(x)^n$$

and

$$g_q(x)^{n-o(n)} \leq [g^n]_q(x).$$

Thus for every $n \in \mathbb{N}$ for every $p \in \mathcal{P}_n(f)$ there is a $q \in \mathcal{P}_n(g)$ such that

$$[f^n]_p(x) \leq [g^n]_q(x)^{n+o(n)} \cdot 2^{o(n)} \text{ on } \mathcal{X}.$$

From boundedness of $\mathcal{X}$ we have $[g^n]_q(x)^{o(n)} \leq 2^{o(n)}$ on $\mathcal{X}$ and thus we find that $f \leq_{\mathrm{FM}} g$ on $\mathcal{X}$.

For the other direction, suppose that $f \leq_{\mathrm{FM}} g$ on $\mathcal{X}$. This means that for every $n \in \mathbb{N}$ and for every $p \in \mathcal{P}_n(f)$ there is a $q \in \mathcal{P}_n(g)$ such that

$$[f^n]_p(x) \leq [g^n]_q(x) \cdot 2^{o(n)} \text{ on } \mathcal{X}.$$

We know (Lemma 6.4) that for every $p \in \mathcal{P}_n(f)$ and $q \in \mathcal{P}_n(g)$ we have

$$f_p(x)^{n-o(n)} \leq [f^n]_p(x)$$

and

$$[g^n]_q(x) \leq g_q(x)^n.$$

Thus for every $n \in \mathbb{N}$ for every $p \in \mathcal{P}_n(f)$ there is a $q \in \mathcal{P}_n(g)$ such that

$$f_p(x)^n \leq g_q(x)^{n+o(n)} \cdot 2^{o(n)} \text{ on } \mathcal{X}.$$

After taking the nth root and letting n go to infinity gives the claim $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$. $\square$

6.4. Multivariate type decomposition and (finite) monomial partial order

In Sections 6.1, 6.2 and 6.3 we introduced the type decomposition and (finite) monomial partial order for univariate polynomials $f \in \mathbb{N}[x]$ and we discussed and proved their basic properties. The next natural step is to extend these notions to multivariate polynomials $f \in \mathbb{N}[x_1, \ldots, x_k]$.[67]

The multivariate setting is important for applications. For instance, while the univariate setting suffices to study the asymptotic spectrum of square matrix multiplication,[68] the multivariate setting introduced in this section will allow us to study the asymptotic spectrum of *rectangular* matrix multiplication and rectangular versions of Schönhage's tau theorem.[69] These applications we will discuss in detail later (Section 9). We will also remark on a connection between the multivariate type decomposition and tropical polynomials from tropical algebraic geometry.

The extension from univariate to multivariate polynomials is mostly straightforward. We state the extended definitions and lemmas and will not give any proofs in this section as they are essentially the same as the proofs for univariate polynomials. The most notable difference compared to the univariate situation is that the connectedness property in Theorem 6.10 becomes a log-convexity property in the multivariate extension (Theorem 6.19).[70]

In this multivariate setting we let $f = \sum_I f_I x^I \in \mathbb{N}[x_1, \ldots, x_k]$ be a polynomial in k variables x_i with non-negative integer coefficients $f_I \in \mathbb{N}$, where the sum runs over exponent vectors $I \in \mathbb{N}^k$ and we use the monomial notation $x^I \coloneqq x_1^{I_1} \cdots x_k^{I_k}$. Beginning with the basics we define the support, types and n-types: Let $\mathrm{supp}(f) = \{I : f_I \neq 0\}$. Let $\mathcal{P}(f)$ be the set of probability vectors on $\mathrm{supp}(f)$.[71] For $n \in \mathbb{N}$ let $\mathcal{P}_n(f) \coloneqq \{p \in \mathcal{P}(f) : \forall I,\, n \cdot p_I \in \mathbb{N}\} \subseteq \mathcal{P}(f)$ be the subset of n-types.

[67] Strassen discusses the type decomposition and monomial partial order directly in the setting of multivariate polynomials [Str88]. We have chosen to first pay close attention to the univariate setting in Sections 6.1, 6.2 and 6.3, to get the main ideas clear, and only then discuss the multivariate setting in Section 6.4.

[68] Here intuitively we identify the variable x with the matrix multiplication tensor MM_2.

[69] Here we take $k = 3$ and identify the variables x_1, x_2, x_3 with the rectangular matrix multiplication tensors $\mathrm{MM}_{2,1,1}$, $\mathrm{MM}_{1,2,1}$, $\mathrm{MM}_{1,1,2}$ respectively.

[70] Note that connectedness and log-convexity are the same in one dimension.

[71] We note that the definition of $\mathcal{P}(f)$ here is slightly different from the definition we gave in the univariate case, where for simplicity we defined $\mathcal{P}(f)$ as all probability vectors of length $d + 1$ (and had to deal with zero coefficients separately in some proofs). Here the probability vectors are only supported on $\mathrm{supp}(f)$. The notions are essentially the same in practice.

Similarly as before, the number of n-types $|\mathcal{P}_n(f)|$ is upper bounded by a polynomial in n, which will be crucial in many proofs involving the type decomposition. (The dependence on k and $\deg(f)$ is exponential, but we think of these as fixed.)

The type decomposition that we defined for univariate polynomials (Definition 6.2) extends naturally to multivariate polynomials. As for univariate polynomials, the type decomposition of multivariate polynomials is obtained by taking the nth power of the polynomial f and "not collecting terms".

Definition 6.13 (Multivariate type decomposition)**.** For $n \in \mathbb{N}$ we may write $f^n = \sum_p [f_n]_p$ where the sum runs over n-types $p \in \mathcal{P}_n(f)$ and the summands are given by $[f^n]_p \coloneqq \binom{n}{pn} \prod_I (f_I x^I)^{p_I n}$. We call this the *type decomposition* of f^n and we call $[f^n]_p$ the *type components*.

Next, we naturally measure the asymptotic size of the type components with a function that is very similar to the function that we defined in the univariate case, leading to the following definition and (essentially defining) lemma.

Definition 6.14. For $p \in \mathcal{P}(f)$ let $f_p(x_1, \dots, x_k) \coloneqq 2^{H(p)} \prod_I (f_I x^I)^{p_I}$ where the product runs over the k-tuples $I \in \operatorname{supp}(f)$.

Lemma 6.15. *f_p measures the asymptotic size of $[f^n]_p$ when n goes to infinity: for $p \in \mathcal{P}_n(f)$,*

$$f_p(x)^{n-o(n)} \leq [f^n]_p(x) \leq f_p(x)^n$$

uniformly in $x \in \mathbb{R}^k_{\geq 0}$. In particular, uniformly on $\mathbb{R}^k_{\geq 0}$, $f_p = \lim_{k\to\infty}([f^{kn}]_p)^{1/kn}$.

We have the following basic properties of f_p that are proven in a manner almost identical to their univariate versions. These properties are used to prove the lemmas that follow.

Lemma 6.16 (Basic properties of f_p)**.**

(i) $\log f_p(2^{y_1}, \dots, 2^{y_k})$ *is linear in $(y_1, \dots, y_k)$ and concave in p.*

(ii) *For every $(s_1, \dots, s_k) \in \mathbb{R}^k_{\geq 0}$, $\max_{p\in\mathcal{P}(f)} f_p(s_1, \dots, s_k) = f(s_1, \dots, s_k)$.*

(iii) *In particular, for every $p \in \mathcal{P}(f)$, $f_p \leq f$ on $\mathbb{R}^k_{\geq 0}$.*

Now we straightforwardly extend the monomial partial order to multivariate polynomials using the functions f_p:

Definition 6.17 (Monomial partial order $\leq_{\mathrm{M}}$)**.** For $f, g \in \mathbb{N}[x_1, \dots, x_k]$ and $\mathcal{X} \subseteq [1, \infty)^k$ we say that f *is at most g monomially on $\mathcal{X}$* and write

$$f \leq_{\mathrm{M}} g \text{ on } \mathcal{X}$$

if for every $p \in \mathcal{P}(f)$ there exists $q \in \mathcal{P}(g)$ such that $f_p \leq g_q$ on $\mathcal{X}$.

The monomial partial order implies the pointwise partial order on polynomials:

Lemma 6.18. *For $f, g \in \mathbb{N}[x_1, \dots, x_k]$, if $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$, then $f \leq g$ on $\mathcal{X}$.*

Now we state the main theorem of this section, on the domain where the monomial partial order holds. Whereas in the univariate case we proved that this domain is connected, it is clear from inspecting the proof that in the multivariate extension the statement becomes that this domain is *log-convex*. Indeed, for subsets of $[1, \infty)$ (the univariate setting) the notions *connected* and *log-convex* coincide, but for subsets of $[1, \infty)^k$ with $k > 1$ (the multivariate setting) they do not.

For any vector $v = (v_1, \ldots, v_k)$ we let $\log v = (\log v_1, \ldots, \log v_k)$ and $2^v = (2^{v_1}, \ldots, 2^{v_k})$. For any set of vectors S we denote by 2^S the set of vectors $\{2^v : v \in S\}$.

Theorem 6.19 (Log-convexity of the domain where $f \leq_{\mathrm{M}} g$ holds)**.** *Let $s, t \in [1, \infty)^k$ and let $f, g \in \mathbb{N}[x_1, \ldots, x_k]$. If $f \leq_{\mathrm{M}} g$ on $\{s, t\}$, then $f \leq_{\mathrm{M}} g$ on $2^{[\log s, \log t]}$.*

Finally, the notion of finite monomial partial order extends directly to multivariate polynomials, and remains equivalent to the monomial partial order:

Definition 6.20 (Multivariate finite monomial partial order $\leq_{\mathrm{FM}}$)**.** For $f, g \in \mathbb{N}[x_1, \ldots, x_k]$ and $\mathcal{X} \subseteq [1, \infty)^k$ we say that f *is at most g finitely monomially on $\mathcal{X}$*, and write $f \leq_{\mathrm{FM}} g$ on $\mathcal{X}$, if for every $\varepsilon > 0$ and for large enough $n \in \mathbb{N}$ we have that for every $p \in \mathcal{P}_n(f)$ there exists $q \in \mathcal{P}_n(g)$ such that $[f^n]_p \leq [g^n]_q \cdot 2^{\varepsilon n}$ on $\mathcal{X}$.

Lemma 6.21 (Multivariate equivalence of the monomial and finite monomial partial orders)**.** *For any polynomials $f, g \in \mathbb{N}[x_1, \ldots, x_k]$ and bounded subset $\mathcal{X} \subseteq [1, \infty)^k$ we have $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$ if and only if $f \leq_{\mathrm{FM}} g$ on $\mathcal{X}$.*

We recall that $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$ implies $f \leq g$ on $\mathcal{X}$, but not the other way around. In Section 7 we will see precisely when the opposite implication is true for all f, g.

Remark 6.22 (Relation to Tropical Geometry)**.** Some readers may have noticed a similarity between Strassen's type decomposition $[f^n]_p$ and the notion of *tropical polynomials* from tropical geometry. Tropical geometry is an important, growing subarea of real algebraic geometry, aiming to understand (systems of) polynomial equations and inequalities by moving from the usual $(+, \times)$ ring to the $(\max, +)$ semiring (the tropical semiring).[72] Let us say a few words about this connection between the areas (that have seemingly been developed independently). More precisely a *tropical polynomial* is defined [MS15, Sec. 1.1] as a maximization of finitely many linear functions, of the form

$$\max\{a_1 + i_{1,1}x_1 + \cdots + i_{1,n}x_n,\ a_2 + i_{2,1}x_1 + \cdots + i_{2,n}x_n,\ \ldots\},$$

where the a_k are real numbers and the $i_{k,\ell}$ are natural numbers. The type decomposition that we introduced in this section (culminating in Lemma 6.16) naturally suggests a "tropical version" of any polynomial $f = \sum_I f_I x^I \in \mathbb{N}[x_1, \ldots, x_k]$, namely

$$\max_{p \in \mathcal{P}_n(f)} \log\, [f^n]_p. \tag{6.2}$$

The expression in (6.2) is indeed a tropical polynomial in the usual sense, in the logarithmic variables $\log x_1, \ldots, \log x_k$, namely a maximization of the affine forms

$$\log[f^n]_p = \log \binom{n}{pn} + \sum_I p_I n \log f_i + \sum_I p_I n \Big(\sum_i I_i \log x_i\Big)$$

[72] Or equivalently the $(\min, +)$ semiring.

over the finitely many n-types $p \in \mathcal{P}_n(f)$.

In an asymptotic sense, as we have seen before, the expression $\max_{p\in\mathcal{P}_n(f)} \log\, [f^n]_p$ equals f^n, namely $\max_p \log\, [f^n]_p \leq f^n \leq \mathrm{poly}(n) \max_p \log\, [f^n]_p$. This viewpoint we have formalized earlier by defining the functions f_p that provide the asymptotic growth of the type components $[f^n]_p$ and for which we have the relation

$$\log f = \max_{p\in\mathcal{P}(f)} \log f_p. \tag{6.3}$$

The maximization in (6.3) is not quite a tropical polynomial in the usual sense, but we may think of it as an "infinite, non-integral" tropical polynomial, since the maximization is over the infinitely many elements $p \in \mathcal{P}(f)$ and $\log f_p = H(p) + \sum_I p_I \log f_i + \sum_I p_I \sum_i I_i \log x_i$ has non-integral coefficients $p_I I_i$ appearing.

7. Convexity and the monomial partial order

We introduced in Section 6 the monomial partial order on univariate polynomials $\mathbb{N}[x]$ (Sections 6.1, 6.2 and 6.3). The monomial partial order has the important property that its domain is *connected* (Theorem 6.10). This property we will put to use in this section to characterize *connectedness* of sets $\mathcal{X} \subseteq [1,\infty)$ by the equivalence of the monomial and pointwise partial orders. The two directions of this characterization are proved separately in Sections 7.1 and 7.2. This uses several properties of the functions $f_p(2^y)$ we labored to define above (in particular linearity in y and convexity in p) and connectedness of the domain of the monomial partial order.

After discussing univariate polynomials in Sections 7.1 and 7.2, we will turn to multivariate polynomials $\mathbb{N}[x_1,\ldots,x_k]$ in Section 7.3. The above connectedness characterization extends naturally to this multivariate situation. For this we use the multivariate monomial partial order (Section 6.4) and its main property that its domain is *log-convex* (Theorem 6.19). This partial order straightforwardly leads to a characterization of *log-convexity* of sets $\mathcal{X} \subseteq [1,\infty)^k$ (i.e., convexity of $\log \mathcal{X} \subseteq \mathbb{R}^k_{\geq 0}$) in terms of the monomial partial order.

7.1. Sufficient condition for connectedness

We begin with a sufficient condition for $\mathcal{X} \subseteq [1,\infty)$ to be connected. To show that $\mathcal{X}$ is connected it is sufficient to show that for every $s,t \in \mathcal{X}$ we have $[s,t] \subseteq \mathcal{X}$. Recall that for $f,g \in \mathbb{N}[x]$, $\mathcal{S}(f \leq g) \coloneqq \{s \in [1,\infty) : f(s) \leq g(s)\}$ denotes the locus of the inequality $f \leq g$. The sufficient condition we will prove relies on the following lemma (which is a simple combination of lemmas from Section 6).

Lemma 7.1. *Let $s \leq t \in [1,\infty)$. Let $f,g \in \mathbb{N}[x]$. If $f \leq_{\mathrm{M}} g$ on $\{s,t\}$, then $[s,t] \subseteq \mathcal{S}(f \leq g)$.*

Proof. The domain of the monomial partial order is connected (Theorem 6.10) so from $f \leq_{\mathrm{M}} g$ on $\{s,t\}$ it follows that $f \leq_{\mathrm{M}} g$ on $[s,t]$ by Theorem 6.10. The monomial partial order implies the pointwise partial order (Lemma 6.9), so we get $f \leq g$ on $[s,t]$. This means $[s,t] \subseteq \mathcal{S}(f \leq g)$. ☐

Theorem 7.2 (Sufficient condition for connectedness of $\mathcal{X}$)**.** *Let $\mathcal{X} \subseteq [1,\infty)$ be closed. If the partial orders $\leq$ and $\leq_{\mathrm{M}}$ coincide on all sets $\{s,t\} \subseteq \mathcal{X}$, then $\mathcal{X}$ is connected.*

Proof. Let $s,t \in \mathcal{X}$. We will prove that $[s,t] \in \mathcal{X}$. For any $f,g \in \mathbb{N}[x]$ for which $f \leq g$ on $\mathcal{X}$, we have $f \leq_{\mathrm{M}} g$ on $\{s,t\}$ by assumption, and thus $[s,t] \subseteq \mathcal{S}(f \leq g)$ by Lemma 7.1. Thus $[s,t] \subseteq \bigcap \mathcal{S}(f \leq g)$, where the intersection is over all $f, g \in \mathbb{N}[x]$ for which $f \leq g$ on $\mathcal{X}$. This intersection equals $\overline{\mathcal{X}}$ (Lemma 5.1), which equals X by closedness. This means $[s,t] \subseteq \mathcal{X}$. □

Remark 7.3. The assumption in Theorem 7.2 that $\mathcal{X}$ is *closed* is needed, because without it we could take $\mathcal{X}$ to be $[s,t]$ minus a point and then obtain the contradiction that $[s,t] \subseteq \mathcal{X}$. (This is not surprising as polynomials describe closed sets, cf. Lemma 5.1.)

7.2. Necessary condition for connectedness

Now we prove that the sufficient condition for connectedness given in Theorem 7.2 is a necessary condition. The fact that $\log f_p(2^y)$ is concave in p and linear in y will play a crucial role here, facilitating the use of the von Neumann minimax theorem for changing the order of $\exists q$ and $\forall s$ in the definition of $\leq_{\mathrm{M}}$ (Definition 6.8).

Proposition 7.4 (Von Neumann minimax theorem)**.** *Let $C \subseteq \mathbb{R}^n$ and $D \subseteq \mathbb{R}^m$ be compact and convex sets. Let the function $\phi : C \times D \to \mathbb{R} : (v,w) \mapsto \phi(v,w)$ be continuous and concave in v, and continuous and convex in w. Then*

$$\max_v \min_w \phi(v,w) = \min_w \max_v \phi(v,w),$$

that is, the order of minimizing over w and maximizing over v does not matter.

Now we will prove the converse to Theorem 7.2.

Theorem 7.5 (Necessary condition for connectedness of $\mathcal{X}$)**.** *If $\mathcal{X}$ is connected, then $\leq$ and $\leq_{\mathrm{M}}$ coincide on all subsets $[s,t] \subseteq \mathcal{X}$. That is, for every $[s,t] \subseteq [1,\infty)$ and every $f, g \in \mathbb{N}[x]$, if $f \leq g$ on $[s,t]$, then $f \leq_{\mathrm{M}} g$ on $[s,t]$.*

Proof of Theorem 7.5. For every $p \in \mathcal{P}(f)$ for every $r \in [s,t]$ we have by Lemma 6.5 and Proposition 6.7 that

$$f_p(r) \leq f(r) \leq g(r) = \max_q g_q(r)$$

where the maximization is over $q \in \mathcal{P}(g)$. In other words,

$$\text{for every } p \in \mathcal{P}(f) \text{ for every } r \in [s,t] \text{ there is a } q \in \mathcal{P}(g) \text{ such that } f_p(r) \leq g_q(r). \tag{7.1}$$

We are not done with the proof because the choice of q depends not only on p but also on $r \in [s,t]$. Instead, to satisfy the definition of $\leq_{\mathrm{M}}$ (Definition 6.8), we want to have one q depending on p that works for all $r \in [s,t]$. To get there we will use the minimax theorem (Proposition 7.4). First, fixing p, we rewrite (7.1) as

$$0 \leq \min_r \max_q (\log g_q(r) - \log f_p(r)).$$

The function $\log g_q(2^y) - \log f_p(2^y)$ is concave in q from the convex and compact set $\mathcal{P}(g)$, and convex (in fact, linear) in y from the convex and compact set $[\log s, \log t]$. By the minimax theorem (Proposition 7.4) we have

$$\min_y \max_q (\log g_q(2^y) - \log f_p(2^y)) = \max_q \min_y (\log g_q(2^y) - \log f_p(2^y)).$$

Thus, we have

$$\text{for every } p \in \mathcal{P}(f) \text{ there is a } q \in \mathcal{P}(g) \text{ such that for every } r \in [s,t] \text{ we have } f_p(r) \leq g_q(r).$$

We conclude that $f \leq_{\mathrm{M}} g$ on $[s,t]$. □

7.3. Multivariate sufficient and necessary condition for log-convexity

In Sections 7.1 and 7.2 we proved sufficient and necessary conditions for a set $\mathcal{X} \subseteq [1, \infty)$ to be connected in terms of the monomial partial order on polynomials. We will now discuss how these conditions almost directly extend to a characterization for a subset $\mathcal{X} \subseteq [1, \infty)^k$ to be *log-convex*. We leave the proof of the characterization to the reader as it is essentially the same as for the univariate case, using the multivariate lemmas that we prepared in Section 6.4.

Theorem 7.6. *Let $\mathcal{X} \subseteq [1, \infty)^k$ be closed. The following are equivalent:*

- *$\mathcal{X}$ is log-convex*
- *the partial orders $\leq$ and $\leq_{\mathrm{M}}$ on polynomials in $\mathbb{N}[x_1, \ldots, x_k]$ coincide on $\mathcal{X}$.*

In particular, to show that a closed set $\mathcal{X} \subseteq [1, \infty)^k$ is log-convex it suffices to show for every two points $s, t \in \mathcal{X}$ that for every two polynomials $f, g \in \mathbb{N}[x_1, \ldots, x_k]$ if $f \leq g$ on $\mathcal{X}$, then $f \leq_{\mathrm{M}} g$ on $\mathcal{X}$.

Looking ahead to Section 8, there is a third property besides connectedness and convexity, called *star-convexity*, that is stronger than connectedness and weaker than convexity. Star-convexity will be of interest in applications to matrix multiplication and tensor networks. We will see how to use the methods of this section to prove star-convexity of $\log \mathcal{X}$ for closed sets $\mathcal{X} \subseteq [1, \infty)^k$ and in particular for asymptotic spectra $\mathcal{X}$ of semirings with a Strassen preorder.

8. Star-convexity and anchors

In Section 7 we characterized which closed sets $\mathcal{X} \subseteq [1, \infty)$ are connected and more generally which closed sets $\mathcal{X} \subseteq [1, \infty)^k$ are log-convex. In this section we focus on a property that is weaker than log-convexity, called *log-star-convexity*. We define that a set $A \subseteq \mathbb{R}^k_{\geq 0}$ is *star-convex with respect to the element* $c \in A$ (called a *center*) if for every point $a \in A$ the interval $[a, c]$ if $a \leq c$ or the interval $[c, a]$ if $c \leq a$ is contained in A. We say that $\mathcal{X} \subseteq [1, \infty)^k$ is *log-star-convex* if $\log \mathcal{X}$ is star-convex. (In the one-dimensional case, log-star-convexity coincides with connectedness and log-convexity.) We will introduce a method to prove that a closed set $\mathcal{X} \subseteq [1, \infty)^k$ is log-star-convex. This method we call the *anchor method*. It says that if $\mathcal{X}$ contains a special point, called an *anchor*, then $\mathcal{X}$ is log-star-convex.

The main goal of the anchor method (which we were led to by the previous sections) is to turn an asymptotic inequality in the pointwise partial order into one in the monomial partial order, magically allowing the exchange of quantifiers that are part of it. A central component of the anchor method (that we will make precise) is that a "local" polynomial inequality (satisfied at one point of $\mathcal{X}$, the anchor) implies a related polynomial inequality which is global, over all of $\mathcal{X}$. The hard work in applying the anchor method is to find an anchor in $\mathcal{X}$. (Indeed in Part III we will carry this out for a family of asymptotic spectra.)

As in previous sections (Sections 6 and 7) we will discuss the anchor method for the one-dimensional case $\mathcal{X} \subseteq [1, \infty)$ in detail first (in which we can focus on connectivity) and then for the high-dimensional case $\mathcal{X} \subseteq [1, \infty)^k$, as the one-dimensional proofs are simpler and essentially the same as the high-dimensional proofs (which we leave to the reader). As opposed to previous sections we will in this section for the first time pay special attention to (and use the extra structure of) the closed sets $\mathcal{X}(a_1, \ldots, a_n) \subseteq [1, \infty)^k$ that are given by the theory of asymptotic spectra (Definition 3.21).[73]

In Section 8.1 we introduce a preliminary one-dimensional version of the anchor method that applies to any closed set $\mathcal{X} \subseteq [1, \infty)$. This preliminary version does not make any reference to the theory of asymptotic spectra yet. In Section 8.2 we introduce the stronger version of the anchor method that applies to sets $\mathcal{X}(a) \subseteq [1, \infty)$ for any semiring $\mathcal{R}$, Strassen preorder P and semiring element $a \in \mathcal{R}$. In Section 8.3 we discuss the full multivariate version of the anchor method that applies to any set $\mathcal{X}(a_1, \ldots, a_k) \subseteq [1, \infty)^k$ for semiring elements $a_1, \ldots, a_k \in \mathcal{R}$ (Definition 3.21). This version generalizes the foregoing ones.

Our notion of an anchor abstracts the method that Strassen used to prove the central result in [Str88] that the asymptotic spectrum of rectangular matrix multiplication is log-star-convex (and that the asymptotic spectrum of square matrix multiplication is connected). In Part III we will apply the anchor theorem to prove Strassen's star-convexity and connectedness theorems for matrix multiplication and to generalize these to the broader class of tensor networks.

8.1. Anchor method for connectedness

We begin with a one-dimensional version of the anchor method for closed sets $\mathcal{X} \subseteq [1, \infty)$ (without reference to asymptotic spectra). This method provides a condition for $\mathcal{X}$ to be connected (which in the one-dimensional case coincides with log-star-convexity). We will build on the sufficient condition for connectedness from Section 7.1. The main difference from Section 7.1 is that here we will single out a special element $s \in \mathcal{X}$ (ultimately called *anchor*) and prove connectedness of $\mathcal{X}$ *via* the element s (which in the higher-dimensional setting of Section 8.3 will naturally lead to log-*star*-convexity with respect to s).

Let $\mathcal{X} \subseteq [1, \infty)$ be closed and bounded. Throughout this section $s \in \mathcal{X}$ will be a fixed element. We introduce three conditions for $\mathcal{X}$ being connected:

"$s \in \mathcal{X}$ is an anchor" (Definitions 8.3 and 8.5) $\Longrightarrow$ Condition B $\Longleftrightarrow$ Condition A $\Longleftrightarrow$ $\mathcal{X}$ is connected.

Note that some of the implications are equivalences. Most important for us will be the overall implication that if there is an anchor $s \in \mathcal{X}$, then $\mathcal{X}$ is connected. We begin with the most straightforward condition, Condition A. This condition is a simple consequence of things we have seen before, and in particular crucially relies on the convexity of the monomial partial order $\leq_{\mathrm{M}}$. The only difference from before is that we are singling out one element $s \in \mathcal{X}$ as special.

Condition A. For the fixed element $s \in \mathcal{X}$ the following holds. For every $t \in \mathcal{X}$, for every $f, g \in \mathbb{N}[x]$, if $f \leq g$ on $\mathcal{X}$, then $f \leq_{\mathrm{M}} g$ on $\{s, t\}$.

Lemma 8.1. *Condition A holds if and only if $\mathcal{X}$ is connected.*

Proof. This lemma follows directly from the sufficient and necessary condition for connectedness, Theorems 7.2 and 7.5. Let us also give a direct proof that Condition A implies that $\mathcal{X}$ is connected, using the results of Section 6. Recall that $\leq_{\mathrm{M}}$ is convex in the sense that if $f \leq_{\mathrm{M}} g$ on $\{s, t\}$, then $f \leq_{\mathrm{M}} g$ on $[s, t]$ (Theorem 6.10) and thus $f \leq g$ on $[s, t]$ (Lemma 6.9). Therefore, Condition A implies that $[s, t] \subseteq \mathcal{X}$ for any $t \in \mathcal{X}$. We conclude that $\mathcal{X}$ is connected. $\square$

[73] Recall that $\mathcal{X}(a_1, \ldots, a_k)$ is defined as the set of k-tuples $(\phi(a_1), \ldots, \phi(a_k))$ of evaluations going over all spectral points ϕ in the asymptotic spectrum of the semiring $\mathcal{R}$ and Strassen preorder P, where the a_i are elements of $\mathcal{R}$.

To discuss Condition B, we need some notation. We defined for any polynomial g of degree d the type decomposition of g^n into a sum of polynomials $[g^n]_p$ where p runs over the set $\mathcal{P}_n(g)$ of probability vectors on $\{0, 1, \ldots, d\}$ that are n-types (Definition 6.2). For any subset Q of $\mathcal{P}_n(g)$, we use the notation

$$[g^n]_Q \coloneqq \sum_{q \in Q} [g^n]_q.$$

Condition B. For the fixed element $s \in \mathcal{X}$ the following holds. For every $f, g \in \mathbb{N}[x]$, if $f \leq g$ on $\mathcal{X}$, then for every $n \in \mathbb{N}$ and every $p \in \mathcal{P}_n(f)$ there is a subset $Q \subseteq \mathcal{P}_n(g)$ such that the following two statements are true:

(i) $[f^n]_p \leq 2^{o(n)}[g^n]_Q$ on $\mathcal{X}$

(ii) for every $q \in Q$ we have $[f^n]_p(s) \leq 2^{o(n)}[g^n]_q(s)$.

We stress that part (ii) of Condition B says $\forall q \in Q$ rather than $\exists q \in Q$. This is crucial in the proof of the following lemma. Recall (Lemma 6.12) that the monomial partial order $\leq_{\mathrm{M}}$ is equivalent to the finite monomial partial order $\leq_{\mathrm{FM}}$ of Section 6.3, which is defined by: $f \leq_{\mathrm{FM}} g$ on $\mathcal{X}$ if for large enough n we have that for every $p \in \mathcal{P}_n(f)$ there is $q \in \mathcal{P}_n(g)$ such that $[f^n]_p \leq 2^{o(n)}[g^n]_q$ on $\mathcal{X}$. We will be using this equivalence here.

Lemma 8.2. *Condition B and Condition A are equivalent.*

Proof. The implication from Condition A to Condition B follows by taking $Q = \{q\}$ to be the appropriate singleton. We now prove the important implication from Condition B to Condition A. Suppose that $f \leq g$ on $\mathcal{X}$. For any $t \in \mathcal{X}$, we need to show that $f \leq_{\mathrm{M}} g$ on $\{s, t\}$. Let $p \in \mathcal{P}_n(f)$. From Condition B (i) it follows that, since $\mathcal{P}_n(g)$ has $\mathrm{poly}(n)$ size (Lemma 6.6) and thus $Q \subseteq \mathcal{P}_n(g)$ has $\mathrm{poly}(n)$ size, there exists $q \in Q$ such that

$$[f^n]_p(t) \leq 2^{o(n)}[g^n]_q(t).$$

Condition B (ii) implies that (for the same q)

$$[f^n]_p(s) \leq 2^{o(n)}[g^n]_q(s).$$

This proves $f \leq_{\mathrm{FM}} g$ on $\{s, t\}$ and thus $f \leq_{\mathrm{M}} g$ on $\{s, t\}$, by the equivalence between the finite monomial partial order and the monomial partial order (Section 6.3). ☐

Now we introduce the notion of an anchor in $\mathcal{X}$. We will call this definition "preliminary" as we give a more general definition of anchors in Section 8.2 that references asymptotic spectra (of which this can be seen as a special case for the right choice of semiring and preorder).

Anchors provide a method for realizing Condition B. The high-level idea is that given $f \leq g$ on $\mathcal{X}$ and $p \in \mathcal{P}_n(f)$ to obtain Q, we may simply define Q to consist of all $q \in \mathcal{P}_n(g)$ so that $[f^n]_p(s) \leq 2^{o(n)}[g^n]_q(s)$ (where the $o(n)$ function is well-chosen). Then (ii) is automatically satisfied but rather than (i) we have

$$[f^n]_p \leq f^n \leq g^n = [g^n]_{\overline{Q}} + [g^n]_Q \text{ on } \mathcal{X}$$

where $\overline{Q}$ is the complement of Q. Then to satisfy Condition B it remains to get rid of this extra term $[g^n]_{\overline{Q}}$ on the right-hand side.

This is where the anchor comes in. First, note that for any $f \in \mathbb{N}[x]$ every type component $[f^n]_p$ is of the form mx^k for some $m, k \in \mathbb{N}$ (it is a monomial). In simple terms, the goal of an anchor is to, given an inequality $mx^k \leq h_1(x) + h_2(x)$ on $\mathcal{X}$ for polynomials $h_1, h_2, \in \mathbb{N}[x]$, produce a simplified inequality in which $h_1(x)$ does not appear, in exchange for a small slack: $mx^k \leq 2^{o(k)} h_2(x)$ on $\mathcal{X}$. A simple (but very strong) condition that allows this[74] is

$$h_1(x) \leq mx^{k-c} \text{ on } \mathcal{X} \text{ for some } c \geq 1. \tag{8.1}$$

An anchor $s \in \mathcal{X}$ is a special point that allows us to carry out the above simplification assuming (8.1) *only* at $s \in \mathcal{X}$. The way that the definition of an anchor will then connect to Condition B is that h_2 will be the polynomial $[g^n]_Q$ and h_1 will be the polynomial $[g^n]_{\overline{Q}}$.

Definition 8.3 (Anchor in $\mathcal{X}$, preliminary)**.** We say that $s \in \mathcal{X}$ is an *anchor* if there is a constant $c \in \mathbb{N}$ such that for every $k \in \mathbb{N}$, $k \geq c$, every $m \in \mathbb{N}$ and every polynomials $h_1, h_2 \in \mathbb{N}[x]$, if

(i) $mx^k \leq h_1(x) + h_2(x)$ on $\mathcal{X}$

(ii) $h_1(s) \leq ms^{k-c}$,

then $mx^k \leq 2^{o(k)} h_2(x)$ on $\mathcal{X}$.[75]

An anchor $s \in \mathcal{X}$ is, by definition, a very special element in $\mathcal{X}$. Namely, it allows us, given a rather weak upper bound on the polynomial h_1 at a single (anchor) point s in (ii), to eliminate h_1 from the right-hand side of the inequality in (i) (with a small slack) for *all* points in $\mathcal{X}$. This is the power of an anchor, and also the source of difficulty proving that a point s is an anchor.

The crucial property of anchors is this:

Theorem 8.4 (Anchor theorem, preliminary)**.** *If $\mathcal{X} \subseteq [1, \infty)$ contains an anchor, then $\mathcal{X}$ is connected.*

We will not give the proof of Theorem 8.4 (but the reader may try to prove as an exercise that $s \in \mathcal{X}$ being an anchor implies Condition B). Instead, we prove a slightly more general version of Theorem 8.4 in Section 8.2 in the context of asymptotic spectra.

8.2. Anchors imply connectedness of univariate spectra

Let $\mathcal{R}$ be a semiring with a Strassen preorder P and let $a \in \mathcal{R}$. We will in this section introduce the anchor method for showing that $\mathcal{X} = \mathcal{X}(a) \subseteq [1, \infty)$ is connected. This will generalize the preliminary anchor definition and theorem (Definition 8.3 and Theorem 8.4).[76] In Section 8.3 we will further generalize the anchor method to a multivariate version.

Definition 8.5 (Anchor in $\mathcal{X}$ with respect to $\mathcal{R}$ and P)**.** Let $a \in \mathcal{R}$ be fixed. We say that $s \in \mathcal{X} = \mathcal{X}(a)$ is an anchor with respect to $\mathcal{R}$ and P if for all large enough $k \in \mathbb{N}$ (say $k \geq c$), every $m \in \mathbb{N}$ and every polynomials $h_1, h_2 \in \mathbb{N}[x]$, if

(a) $ma^k \leq_P h_1(a) + h_2(a)$

[74] Assuming $\mathcal{X}$ is also bounded.

[75] Note that h_1, h_2 may depend on k, in particular their degrees may grow with k so that condition (b) cannot be trivially satisfied by letting k grow.

[76] Since we may take $\mathcal{R}$ to be the semiring of polynomials $\mathbb{N}[x]$ with P the pointwise preorder on $\mathcal{X}$.

(b) $h_1(s) \leq m s^{k-c}$

then $m a^k \leq_P 2^{o(k)} h_2(a)$.[77]

Note how the pointwise inequalities on polynomials in the preliminary Definition 8.3 are now replaced by inequalities in the preorder P in Definition 8.5.

Theorem 8.6 (Anchor theorem)**.** *If $s \in \mathcal{X}$ is an anchor with respect to $\mathcal{R}$ and P, then $\mathcal{X}$ is connected.*

The hard work in applying Theorem 8.6 is to *find* an anchor. In Section 12 we will prove that there is an anchor in the asymptotic spectrum of matrix multiplication and tensor networks. This proof will rely on a shifting theorem that we will prove in Section 11.

Proof. Let $s \in \mathcal{X}$ be an anchor with respect to $\mathcal{R}$ and P. We will prove that Condition B holds, which implies that $\mathcal{X}$ is connected (Lemmas 8.1 and 8.2). Let $f, g \in \mathbb{N}[x]$ such that $f \leq g$ on $\mathcal{X}$. Let us fix $p \in \mathcal{P}_n(f)$. The rest of the proof has two parts:

- Construct a subset $Q \subseteq \mathcal{P}_n(g)$ in such a way that part (ii) of Condition B,

$$\text{for every } q \in Q,\ [f^n]_p(s) \leq 2^{o(n)} [g^n]_q(s),$$

  is "trivially" true.

- Use the anchor to prove part (i) of Condition B:

$$[f^n]_p \leq 2^{o(n)} [g^n]_Q \text{ on } \mathcal{X}.$$

As a technical preparation, without loss of generality we have that $\deg(f) \geq c$ and $\deg(g) \geq c$ where c is the parameter dictated by the anchor (Definition 8.5), since otherwise we may multiply both f and g by x^c. We will use this later.

We first deal with part (ii). The choice of Q is practically forced on us, given what we want to achieve. Namely, we let

$$Q \coloneqq \{q \in \mathcal{P}_n(g) : [f^n]_p(s) \leq [g^n]_q(s) \cdot (|\mathcal{P}_n(g)| \cdot s^c \cdot 2^{\delta(n)})\},$$

where $\delta(n)$ is a function in $o(n)$ that we determine in the proof of part (i). Thus $\mathcal{P}_n(g)$ splits into the disjoint union $\mathcal{P}_n(g) = Q \cup \overline{Q}$ of Q and its complement $\overline{Q}$ in $\mathcal{P}_n(g)$. The factor $|\mathcal{P}_n(g)| \cdot s^c \cdot 2^{\delta(n)}$ grows subexponentially in n (Lemma 6.6). It does not play a role in the proof of (ii), but it will make (i) work. By definition of Q, we have for any $q \in Q$ that

$$[f^n]_p(s) \leq 2^{o(n)} [g^n]_q(s).$$

Thus we see that part (ii) of Condition B is satisfied.

Now part (i). Recall that $\mathcal{X} = \mathcal{X}(a)$. From $f \leq g$ on $\mathcal{X}$ we have by Strassen duality (Theorem 3.41) that there is a function $\delta(n) = o(n)$ such that (in the preorder P)

$$f^n(a) \leq_P 2^{\delta(n)} g^n(a).$$

[77] In our application we will take $c = 1$.

(Thus Strassen duality provides us the function $\delta(n) = o(n)$ that we promised to determine when we defined Q above.) Thus letting $h_1 = 2^{\delta(n)}[g^n]_{\overline{Q}}$ and $h_2 = 2^{\delta(n)}[g^n]_Q$ we have

$$[f^n]_p(a) \leq_P f^n(a) \leq_P 2^{\delta(n)} g^n(a) = h_1(a) + h_2(a).$$

In this inequality we need to "get rid of" $h_1(a) = 2^{\delta(n)}[g^n]_{\overline{Q}}(a)$ on the right-hand side to get (i). In order to make use of the anchor s to do this, we observe that by definition of the complement set $\overline{Q}$ we have

$$[g^n]_{\overline{Q}}(s) \leq |\mathcal{P}_n(g)| \cdot \max_{q \in \overline{Q}} [g^n]_q(s) \leq [f^n]_p(s) \cdot s^{-c} \cdot 2^{-\delta(n)}.$$

We thus have $[f^n]_p(a) \leq_P h_1(a) + h_2(a)$ in $\mathcal{R}$ and $h_1(s) \leq [f^n]_p(s) \cdot s^{-c}$. (Note that $\deg(f) \geq c$ implies that $\deg([f^n]_p) \geq c$). We are now precisely in the situation that we can use that s is an anchor with respect to $\mathcal{R}$ and P (Definition 8.5). From this it follows that

$$[f^n]_p(a) \leq_P 2^{o(n)} h_2(a) = 2^{o(n)}[g^n]_Q(a).$$

In particular, using the easy direction of Strassen duality (Theorem 3.41),

$$[f^n]_p(x) \leq 2^{o(n)}[g^n]_Q(x) \text{ on } \mathcal{X}.$$

This gives (i) of Condition B. We conclude that Condition B holds. □

8.3. Anchors imply log-star-convexity of multivariate spectra

In Section 8.2 we proved the one-dimensional anchor theorem (Theorem 8.6). Now we will discuss its *high-dimensional* (or *multivariate*) version. Here for the first time we will be able to see the notion of log-star-convexity properly in action (before, it coincided with connectedness), which is defined as follows:

Definition 8.7 (Star-convexity). Any set $A \subseteq \mathbb{R}^k_{\geq 0}$ is called *star-convex with respect to* $a \in A$ (which is called a *center*) if for any $b \in A$ it holds that the line segment $[a, b]$ is contained in A. We call the set $\mathcal{X} \subseteq [1, \infty)^k$ *log-star-convex with respect to* $a \in \mathcal{X}$ if the set $\log \mathcal{X}$ is star-convex with respect to $\log a$.

A star-convex set need not be convex (Fig. 2). However:

Lemma 8.8. *The set of all centers of A is convex*

Proof. We leave this as an exercise to the reader. □

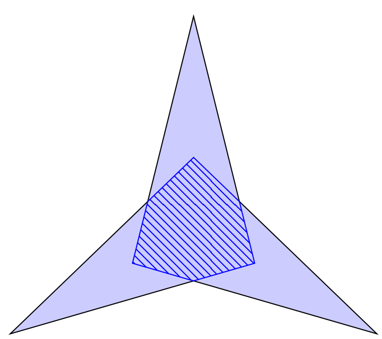


Figure 2: Example of a star-convex set (blue filled region) that is not convex. The subset of centers (striped region) is always convex.

This high-dimensional version of the anchor theorem we will state directly in the language of semirings and preorders. Let $\mathcal{R}$ be a semiring with a Strassen preorder P and let $a_1, \ldots, a_k \in \mathcal{R}$ be finitely many semiring elements. Recall that $\mathcal{X}(a_1, \ldots, a_k) \subseteq [1, \infty)^k$ denotes the asymptotic spectrum evaluated at $a_1, \ldots, a_k$ (Definition 3.21).

Definition 8.9 (Anchor in $\mathcal{X}$ with respect to $\mathcal{R}$ and P)**.** We call $s \in \mathcal{X}(a_1, \ldots, a_k)$ an *anchor* with respect to $\mathcal{R}$ and P if there are $c \in \mathbb{N}^k$ and $\varepsilon(j) = o(j_1 + \cdots + j_k)$ such that the following holds. For all large enough $j_i \in \mathbb{N}$, every $m \in \mathbb{N}$, and every polynomials $h_1, h_2 \in \mathbb{N}[x_1, \ldots, x_k]$, if

(a) $m a_1^{j_1} \cdots a_k^{j_k} \leq_P h_1(a_1, \ldots, a_k) + h_2(a_1, \ldots, a_k)$

(b) $h_1(s_1, \ldots, s_k) \leq m s_1^{j_1 - c_1} \cdots s_k^{j_k - c_k}$

then $m a_1^{j_1} \cdots a_k^{j_k} \leq_P 2^{\varepsilon(j)} h_2(a_1, \ldots, a_k)$.

Theorem 8.10 (Multivariate anchor theorem)**.** *If $s = (s_1, \ldots, s_k) \in \mathcal{X}(a_1, \ldots, a_k)$ is an anchor, then $\mathcal{X}(a_1, \ldots, a_k)$ is log-star-convex with respect to s.*

The proof of Theorem 8.10 is a natural extension of the univariate proof (Theorem 8.6) and we leave it to the reader.

Looking ahead, in Part III we will apply the anchor method in the settings of matrix multiplication and tensor networks to prove log-star-convexity of the relevant asymptotic spectra. Finding anchors in these settings is non-trivial and requires the so-called shifting and compression theorems that we prove in Part III.

9. Applications and extensions of connectedness

In Part II we have discussed several characterizations of connectedness of asymptotic spectra and methods to prove connectedness, log-star-convexity and log-convexity. In this section we will discuss applications these properties of asymptotic spectra, focussing on the task of upper bounding the asymptotic rank (Sections 9.1 and 9.2). Then we will have a high-level discussion of type decompositions in the context of convexity properties of asymptotic spectra (Section 9.3).

9.1. The power of connectedness and convexity in bounding rank

Polynomial inequalities in the semiring provide a central tool for upper bounding the asymptotic rank (or lower bounding the asymptotic subrank). Indeed, Schönhage's tau theorem is a perfect example of this technique, and is behind almost all improvements on the matrix multiplication exponent. In this section we will see that if the asymptotic spectrum is *connected*, this technique is enhanced! Namely, one can obtain bounds from inequalities that in the general case would imply none, and also extract much better bounds from polynomial inequalities than without them.

We use our usual notation: let $\mathcal{R}$ be a semiring, let P be a Strassen preorder on $\mathcal{R}$ and let $a \in \mathcal{R}$ be a semiring element.[78] Let $\mathcal{X}(a) \subseteq [1, \infty)$ be the asymptotic spectrum $\mathcal{X}$ evaluated at the

[78] We may assume that $\mathcal{R}$ is generated by a.

element a. Recall that $\widetilde{P}$ is the asymptotic preorder associated to P, which is defined by $a \leq_{\widetilde{P}} b$ if and only if $a^n \leq_P b^{n+o(n)}$.

For example, $\mathcal{R}$ may be the semiring generated by the matrix multiplication tensor $a = \mathrm{MM}_2$ so that the asymptotic rank $\widetilde{\mathrm{R}}(a) = 2^\omega$ captures the matrix multiplication exponent ω (Example 2.22). Everything we will do also applies (by changing the direction of all inequalities) to lower bounding the asymptotic subrank $\widetilde{\mathrm{Q}}(a)$, for example the Shannon capacity of graphs (Example 2.20), assuming the semiring is gapped. For simplicity we will assume throughout this section that our semiring is gapped, so that the asymptotic subrank is the minimum point in the asymptotic spectrum.

The starting point that we will take is Schönhage's tau theorem for matrix multiplication, or rather the generalized version (namely the version for arbitrary semirings with Strassen preorder) that we discussed in Corollary 4.3 (where we proved it as an application of the "additivity if and only if multiplicativity" theorem Theorem 4.1 that in turn followed from Strassen's duality). This tau theorem, slightly rephrased, provides a way to obtain an upper bound on the asymptotic rank from an inequality in the asymptotic preorder $\widetilde{P}$ in which a polynomial expression of a is upper bounded by a natural number:

Theorem 9.1 (Tau theorem, Corollary 4.3)**.** *For any $n_i, r \in \mathbb{N}$,*

$$\sum_i a^{n_i} \leq_{\widetilde{P}} r \iff \sum_i \widetilde{\mathrm{R}}(a)^{n_i} \leq r.$$

A good way to think about Theorem 9.1, and to remember it, is to draw curves cutting out the asymptotic spectrum $\mathcal{X}(a)$, as follows. Suppose[79] that $\mathcal{X}(a)$ is the disjoint union of two closed intervals. Suppose that $\sum_i a^{n_i} \leq_{\widetilde{P}} r$. Then $\sum_i s^{n_i} \leq r$ for all $s \in \mathcal{X}(a)$. In other words, $r - \sum_i x^{n_i}$ is non-negative on $\mathcal{X}(a)$. Note that $r - \sum_i x^{n_i}$ is non-increasing in x. Drawing the set $\mathcal{X}(a) \in [1, \infty)$ (in black) together with the curve $r - \sum_i x^{n_i}$ (in red), the situation must look like in Fig. 3:

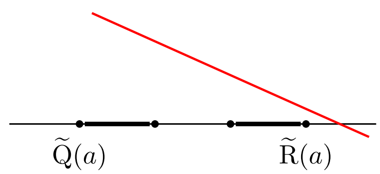


Figure 3: Disconnected spectrum and non-negative monotone curve

The point of the proof of Theorem 9.1 is that, since the red curve is non-increasing, and is non-negative on $\mathcal{X}(a)$, we must have that it crosses the x-axis to the right of the point $\widetilde{\mathrm{R}}(a) \in \mathcal{X}(a)$, and thus $\widetilde{\mathrm{R}}(a)$ is upper bounded by the single root of the red curve.

The power of the tau theorem (Theorem 9.1) is witnessed in practice by the fact that the fastest matrix multiplication algorithms rely on it. The above proof ideas (in particular the usage of Strassen duality) lead in a simple manner to a more general upper bound method. This method does not need any assumptions on the asymptotic spectrum. However, in this generality we will have an implication in one direction only rather than an equivalence. Namely, from *any* inequality $p(a) \leq_{\widetilde{P}} q(a)$ for $p, q \in \mathbb{N}[x]$ we may derive an upper bound on the asymptotic rank $\widetilde{\mathrm{R}}(a)$, in the following way. Recall that $\mathcal{S}(p \leq q) = \{s \in [1, \infty) : p(s) \leq q(s)\}$ denotes the locus of the inequality $p \leq q$.

[79]This assumption is only for the purpose of illustration, as Theorem 9.1 holds regardless of what the asymptotic spectrum looks like, and in particular regardless of whether the asymptotic spectrum is connected or not.

Theorem 9.2. *Let $p, q \in \mathbb{N}[x]$. Then*

$$p(a) \leq_{\widetilde{P}} q(a) \implies \widetilde{\mathrm{R}}(a) \leq \max \mathcal{S}(p \leq q).$$

Theorem 9.2 clearly specializes to Theorem 9.1 by taking $p = \sum_i x^{n_i}$ and $q = r$. Whereas Theorem 9.1 gives an equivalence (i.e., is tight as a method for upper bounding asymptotic rank), in the more general Theorem 9.2 the reverse implication is not always true. Examples can easily be constructed in which $p(\widetilde{\mathrm{R}}(a)) \leq q(\widetilde{\mathrm{R}}(a))$ and $p(\widetilde{\mathrm{Q}}(a)) > q(\widetilde{\mathrm{Q}}(a))$, which prevents the reverse implication. These examples will necessarily come from p and q so that the function $q(x) - p(x)$ is not monotone. (On the other hand, there is a zone of choices for p and q that is more general than Theorem 9.1 but where the converse $\widetilde{\mathrm{R}}(a) \leq \max \mathcal{S}(p \leq q) \implies p(a) \leq_{\widetilde{P}} q(a)$ still holds, for example when $q(x) - p(x)$ has at most one root in $[1, \infty)$.)

Proof. If $p(a) \leq_{\widetilde{P}} q(a)$, then $\mathcal{X}(a) \subseteq \mathcal{S}(p \leq q)$ by definition (the opposite implication is also true by Strassen duality). By Strassen duality, $\widetilde{\mathrm{R}}(a) = \max \mathcal{X}(a)$. Thus $\widetilde{\mathrm{R}}(a) \leq \max \mathcal{S}(p \leq q)$. ☐

Theorem 9.2 does not use any special properties of the semiring and preorder. When we further know that the asymptotic spectrum is *connected* (which we will prove is the case for tensor networks, Theorem 12.23), we can do more. Let us first stick to the previous example and suppose that instead of an inequality $\sum_i a^{n_i} \leq_{\widetilde{P}} r$ we managed to prove an inequality $\sum_i a^{n_i} \leq_{\widetilde{P}} \sum_i a^{m_i}$ for $n_i, m_i \in \mathbb{N}$. What upper bound on $\widetilde{\mathrm{R}}(a)$ does such an inequality give? A priori, none, and it is easy to see this in the illustration below. Again we draw $\mathcal{X}(a)$ (in black) and the curve $\sum_i a^{m_i} - \sum_i a^{n_i}$ (in red). As opposed to the previous illustration, this curve may not be monotone. We know that the curve must be non-negative on $\mathcal{X}(a)$. However, since the curve is not monotone we can be in the unlucky situation (unlucky because upper bounding the value of $\widetilde{\mathrm{R}}(a)$ is our goal) in which the curve cuts out two intervals, as in Fig. 4.

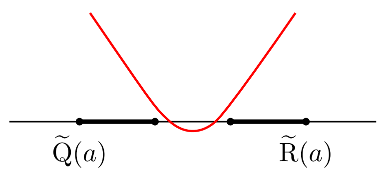


Figure 4: Disconnected spectrum and non-negative non-monotone curve

In this case indeed we do not learn any upper bound on $\widetilde{\mathrm{R}}(a)$, because the locus $\mathcal{S}(p \leq q)$ is unbounded from above.

It is precisely when $\mathcal{X}(a)$ is connected that the above cannot happen and as a result we may obtain an upper bound on $\widetilde{\mathrm{R}}(a)$ from polynomial inequalities that are not monotone, as we will now explain. Suppose that $\mathcal{X}(a)$ is connected and suppose that we know at least one element $t \in \mathcal{X}(a)$. Suppose we prove an inequality $\sum_i a^{n_i} \leq_{\widetilde{P}} \sum_i a^{m_i}$ for $n_i, m_i \in \mathbb{N}$ whose curve $\sum_i x^{m_i} - \sum_i x^{n_i}$ looks as in Fig. 5 (red):

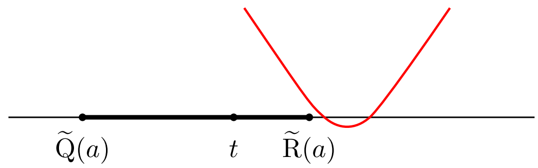


Figure 5: Connected spectrum and non-negative non-monotone curve

We know that the red curve is non-negative on $\mathcal{X}$. The non-negative locus of the red curve is a disjoint union of connected components $\mathcal{Y}_i$. Since we know at least one element $t \in \mathcal{X}(a)$, we know for which i we have $\mathcal{X}(a) \subseteq \mathcal{Y}_i$. And thus for this i we have $\widetilde{\mathrm{R}}(a) \leq \max \mathcal{Y}_i$.

The above discussion leads naturally to the following method for upper bounding asymptotic rank when the asymptotic spectrum is connected. The proof of this theorem will go along the same lines as Theorem 9.2 making essential use of Strassen duality. The method will turn out to be tight in a case corresponding to taking the point t of the above discussion to be $\widetilde{\mathrm{Q}}(a)$.[80] For $t \in \mathcal{S}(p \leq q)$ let $\mathcal{S}(p \leq q)_t$ denote the connected component of $\mathcal{S}(p \leq q)$ that contains t.

Theorem 9.3. *Suppose that $\mathcal{X}(a)$ is connected. Let $p, q \in \mathbb{N}[x]$ and let $t \in \mathcal{X}(a)$. Then*

$$p(a) \leq_{\widetilde{P}} q(a) \implies \widetilde{\mathrm{R}}(a) \leq \max \mathcal{S}(p \leq q)_t.$$

If $\widetilde{\mathrm{Q}}(a) \in \mathcal{S}(p \leq q)$, then

$$p(a) \leq_{\widetilde{P}} q(a) \iff \widetilde{\mathrm{R}}(a) \leq \max \mathcal{S}(p \leq q)_{\widetilde{\mathrm{Q}}(a)}.$$

Note how Theorem 9.3 has the potential to give better upper bounds on asymptotic rank than Theorem 9.2 whenever the locus $\mathcal{S}(p \leq q)$ has more than one connected component, as then $\max \mathcal{S}(p \leq q)_t$ may be strictly smaller than $\max \mathcal{S}(p \leq q)$.

Proof. If $p(a) \leq_{\widetilde{P}} q(a)$, then $\mathcal{X}(a) \subseteq \mathcal{S}(p \leq q)_t$ by definition and using that $\mathcal{X}(a)$ is connected and contains t. (The opposite implication is also true by Strassen duality.) By Strassen duality, $\widetilde{\mathrm{R}}(a) = \max \mathcal{X}(a)$. Thus $\widetilde{\mathrm{R}}(a) \leq \max \mathcal{S}(p \leq q)_t$. This proves the first claim. To prove the reverse implication in the second claim, we use that from the inequality $\widetilde{\mathrm{R}}(a) \leq \max \mathcal{S}(p \leq q)_{\widetilde{\mathrm{Q}}(a)}$ it follows that $[\widetilde{\mathrm{Q}}(a), \widetilde{\mathrm{R}}(a)] \subseteq \mathcal{S}(p \leq q)$ which by Strassen duality gives the inequality $p(a) \leq_{\widetilde{P}} q(a)$. $\square$

The above approach of using polynomial inequalities and the connectedness of the asymptotic spectrum of matrix multiplication may be the way to obtain better bounds on the matrix multiplication exponent. So far, however, this approach has not been used, although the group-theoretic approach for matrix multiplication [CU03, CKSU05, CU13] is in principle able to produce non-monotone inequalities. Also the asymptotic spectrum of graphs is known to be connected [Vra21] and the approach may thus be used to get better bounds on the Shannon capacity of graphs.

9.2. Multivariate bounds from convexity and star-convexity

In one dimension, different notions like connectivity, star-convexity and convexity all coincide. When we move to higher dimension, they become successively stronger, and so do the consequences we

[80] If we happen to know it.

can draw using them from given inequalities. We would like a multivariate analog of Fig. 5 and Theorem 9.3, where we can use non-monotone inequalities to infer bounds on the asymptotic rank. We will be able to do so for rectangular matrix multiplication.

We saw in Section 8 that the anchor method can produce only log-star-convexity (rather than log-convexity), and in particular, connectedness. Strassen [Str88] proved log-star-convexity for the spectrum of rectangular matrix multiplication

$$\mathcal{X}_{\mathrm{rect}} := \mathcal{X}(\mathrm{MM}_{2,1,1}, \mathrm{MM}_{1,2,1}, \mathrm{MM}_{1,1,2}),$$

conjectured that it is actually log-convex, and proved if so how inequalities in that semiring provide bounds on ω. We will show how log-star-convexity is in fact strong enough to provide bounds on ω, albeit somewhat weaker. We thus prove a weaker form of his conjecture.

Throughout this section we use an important correspondence between $\mathcal{X}_{\mathrm{rect}}$ and the spectrum of square matrix multiplication $\mathcal{X}_{\mathrm{sq}} := \mathcal{X}(\mathrm{MM}_{2,2,2})$, which follows from $\mathrm{MM}_2 = \mathrm{MM}_{2,2,2} = \mathrm{MM}_{2,1,1} \otimes \mathrm{MM}_{1,2,1} \otimes \mathrm{MM}_{1,1,2}$ and the following general correspondence. Let $\mathcal{R}$ be any semiring with Strassen preorder. Assume that $\mathcal{R}$ is generated by $a = (a_1, \ldots, a_k)$. Let $p = (p_1, \ldots, p_m)$ be a polynomial map with each p_i having non-negative integer coefficients. Then $b = p(a) = (b_1, \ldots, b_m)$ generates a subsemiring $\mathcal{R}'$, and we have the projection of spectra $\mathcal{X}(b) = p(\mathcal{X}(a))$. So, in particular, we have the projection from $\mathcal{X}_{\mathrm{rect}}$ to $\mathcal{X}_{\mathrm{sq}}$ given simply by the product of the three coordinates: $(u_1, u_2, u_3) \mapsto u_1 u_2 u_3$.[81]

We begin by recalling Schönhage's tau theorem:

Theorem 9.4 (Theorem 4.12)**.** *Let $a_i, b_i, c_i, r \in \mathbb{N}$. Then*

$$\sum_i \mathrm{MM}_{a_i,b_i,c_i} \leq_{\widetilde{P}} I_r \implies \sum_i (a_i b_i c_i)^{\omega/3} \leq r$$

where I_r denotes the $r \times r \times r$ diagonal tensor.[82]

Schönhage used Theorem 9.4 to obtain the upper bound $\omega \leq 2.55$ via the asymptotic inequality $\widetilde{\mathrm{R}}(\mathrm{MM}_{4,1,3} \oplus \mathrm{MM}_{1,8,1}) \leq 13$. Later improved upper bounds have also made use of Theorem 9.4 or a relaxation, called the Coppersmith–Winograd method, in which the left-hand side is replaced by a certain non-direct sum of matrix multiplication tensors.

We now discuss Strassen's result on the generalized tau theorem that can be proven *assuming log-convexity* of the asymptotic spectrum of rectangular matrix multiplication. The main statement that allows this is the following conditional point inclusion in the spectrum:

Lemma 9.5 ([Str88, page 135])**.** *If $\mathcal{X}_{\mathrm{rect}}$ is log-convex, then $(\omega/3, \omega/3, \omega/3) \in \log \mathcal{X}_{\mathrm{rect}}$.*

Proof. We have $\mathrm{MM}_2 = \mathrm{MM}_{2,2,2} = \mathrm{MM}_{2,1,1} \otimes \mathrm{MM}_{1,2,1} \otimes \mathrm{MM}_{1,1,2}$ and so

$$\max_{s \in \mathcal{X}_{\mathrm{rect}}} s_1 s_2 s_3 = \widetilde{\mathrm{R}}(\mathrm{MM}_2) = 2^{\omega}.$$

Let $s = (s_1, s_2, s_3) \in \mathcal{X}_{\mathrm{rect}}$ such that $s_1 s_2 s_3 = 2^{\omega}$. The set $\mathcal{X}_{\mathrm{rect}}$ is S_3-invariant by symmetry of the generators $\mathrm{MM}_{2,1,1}$, $\mathrm{MM}_{1,2,1}$, $\mathrm{MM}_{1,1,2}$. Thus also the cyclic permutations $s' = (s_3, s_1, s_2) \in \mathcal{X}_{\mathrm{rect}}$

[81] As we prove (in Part III) log-star-convexity not only for matrix multiplication but also for tensor networks, one can see that we can make similar conclusions for the natural analog of ω in tensor networks on transitive graphs. We will not pursue this here.

[82] We note for later comparison that a diagonal tensor is a direct sum of (degenerate) matrix multiplication tensors via $I_r = \bigoplus_{i=1}^r \mathrm{MM}_1$.

and $s'' = (s_2, s_3, s_1) \in \mathcal{X}_{\text{rect}}$. Then, by the assumption of log-convexity of $\mathcal{X}_{\text{rect}}$, we get that $(ss's'')^{1/3} \in \mathcal{X}_{\text{rect}}$ and since $(ss's'')^{1/3} = ((s_1s_2s_3)^{1/3}, (s_1s_2s_3)^{1/3}, (s_1s_2s_3)^{1/3}) = (2^{\omega/3}, 2^{\omega/3}, 2^{\omega/3})$, this finishes the proof of the claim. $\square$

Lemma 9.5 in turn implies the following generalized tau theorem for rectangular matrix multiplication. This is a strong form of Schönhage's tau theorem (Theorem 4.12) as the right-hand side is not a constant but also a direct sum of matrix multiplication tensors, so we can get bounds on ω from more general inequalities.

Theorem 9.6 (Strassen [Str88, page 108]). *Let $a_i, b_i, c_i, a'_i, b'_i, c'_i \in \mathbb{N}$. If $\mathcal{X}_{\text{rect}}$ is log-convex, then*

$$\sum_i \mathrm{MM}_{a_i,b_i,c_i} \leq_{\widetilde{P}} \sum_i \mathrm{MM}_{a'_i,b'_i,c'_i} \implies \sum_i (a_ib_ic_i)^{\omega/3} \leq \sum_i (a'_ib'_ic'_i)^{\omega/3}.$$

To prove Theorem 9.6 we need the following lemma.

Lemma 9.7. *Let $\phi \in \mathcal{X}_{\text{rect}}$ and $a, b, c \in \mathbb{N}$. Then*

$$\phi(\mathrm{MM}_{a,b,c}) = \phi(\mathrm{MM}_{2,1,1})^{\log a} \phi(\mathrm{MM}_{1,2,1})^{\log b} \phi(\mathrm{MM}_{1,1,2})^{\log c}.$$

Proof. The proof follows the same simple approximation argument as the proof of the "square version" of this lemma, Lemma 4.14, so we leave this to the reader. $\square$

Proof of Theorem 9.6. Lemma 9.5 says that there is an element $\phi \in \mathcal{X}_{\text{rect}}$ such that

$$\phi(\mathrm{MM}_{2,1,1}) = \phi(\mathrm{MM}_{1,2,1}) = \phi(\mathrm{MM}_{1,1,2}) = 2^{\omega/3}.$$

Applying this ϕ to both sides of the given inequality $\sum_i \mathrm{MM}_{a_i,b_i,c_i} \leq_{\widetilde{P}} \sum_i \mathrm{MM}_{a'_i,b'_i,c'_i}$ and using Lemma 9.7 gives the claim. $\square$

It is not known whether the assumption that $\mathcal{X}_{\text{rect}}$ is log-convex in Lemma 9.5 and Theorem 9.6 is true. Strassen conjectures that it is:

Conjecture 9.8 (Strassen [Str88, page 108]). $\mathcal{X}_{\text{rect}}$ is log-convex.

We will now discuss the role of log-star-convexity in this and a version of the tau theorem that it implies. In Part II we have seen how anchors can be used to prove log-star-convexity of asymptotic spectra. In the upcoming Part III the main goal will be to apply this method to the asymptotic spectra of matrix multiplication and tensor networks to prove that they are log-star-convex:

Theorem 9.9 (Corollary 12.26). *$\log \mathcal{X}_{\text{rect}}$ is star-convex with respect to the triangle*

$$\mathrm{conv}(\{(1,1,0),(1,0,1),(0,1,1)\}).\text{[83]}$$

We can in turn use Theorem 9.9 to prove Strassen's Theorem 9.6 unconditionally, but with an adjusted exponent. The crucial step is the unconditional containment of the following point in the asymptotic spectrum of rectangular matrix multiplication:

[83] Recall that the set of centers of a star-convex set is always convex (Lemma 8.8), so equivalently we could have said that $\log \mathcal{X}_{\text{rect}}$ is star-convex with respect to the three points $\{(1,1,0),(1,0,1),(0,1,1)\}$. However, we will explicitly be using that the full triangle is contained in the set of centers in the proofs below.

Lemma 9.10. $(\frac{4-\omega}{7-2\omega}, \frac{4-\omega}{7-2\omega}, \frac{4-\omega}{7-2\omega}) \in \log \mathcal{X}_{\mathrm{rect}}$.

Proof. This proof will be very similar to the proof of Lemma 9.5, using the weaker log-star-convexity rather than log-convexity. We have $\mathrm{MM}_2 = \mathrm{MM}_{2,2,2} = \mathrm{MM}_{2,1,1} \otimes \mathrm{MM}_{1,2,1} \otimes \mathrm{MM}_{1,1,2}$ and so by Strassen duality we have $\max_{z \in \log \mathcal{X}_{\mathrm{rect}}} z_1 + z_2 + z_3 = \omega$. This means that there exists $z = (z_1, z_2, z_3) \in \log \mathcal{X}_{\mathrm{rect}}$ such that $z_1 + z_2 + z_3 = \omega$. For any such z, by log-star-convexity of $\mathcal{X}_{\mathrm{rect}}$ and using the notation $\Delta_z \coloneqq \mathrm{conv}(\{(1,1,0),(1,0,1),(0,1,1),z\})$, it follows that $\Delta_z \subseteq \log \mathcal{X}_{\mathrm{rect}}$. We only know the *existence* of such a point z, but we may still compute a point of the form $(u,u,u) \in \log \mathcal{X}_{\mathrm{rect}}$ by evaluating

$$\min_{z : z_1 + z_2 + z_3 = \omega} \max\{u : (u,u,u) \in \Delta_z\},$$

where we take a minimum because we do not know what the true value of z is, so we need to assume a worst-case z. We claim that the optimal value is $u = \frac{4-\omega}{7-2\omega}$. In this optimization, by S_3-symmetry of $\mathcal{X}_{\mathrm{rect}}$ and since $\log \mathcal{X}_{\mathrm{rect}}$ is contained in the cube C with vertices $\{0,1\}^3$, we see that we may without loss of generality take z to be the point $(1,1,\omega-2)$ or the point $((\omega-1)/2, (\omega-1)/2, 1)$, which are respectively a vertex of the triangle given by the intersection of the plane $z_1 + z_2 + z_3 = \omega$ and the cube C, and a midpoint on an edge of this triangle.

To compute $\max\{u : (u,u,u) \in \Delta_z\}$ for $z = (1,1,\omega-2)$ we compute the intersection of the line between $(0,0,0)$ and $(1,1,1)$ and the line between $(1/2,1/2,1)$ and $(1,1,\omega-2)$, as drawn in Figure 6.

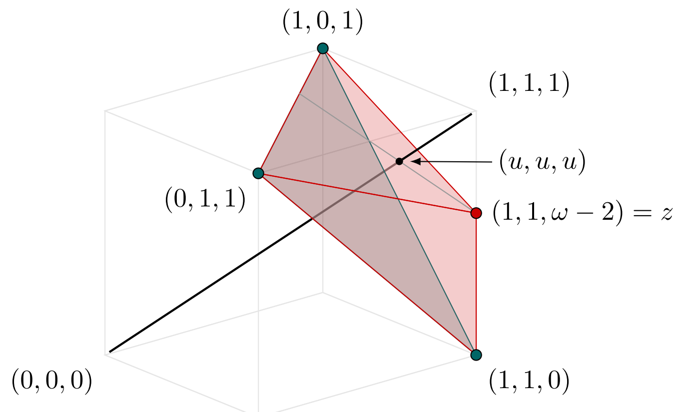


Figure 6: Using the star-convexity of $\log \mathcal{X}_{\mathrm{rect}}$ to find a point of the form (u,u,u) in $\log \mathcal{X}_{\mathrm{rect}}$. The red region is the set Δ_z for $z = (1,1,\omega-2)$.

A straightforward computation gives the optimal value $u = \frac{4-\omega}{7-2\omega}$.

To compute $\max\{u : (u,u,u) \in \Delta_z\}$ for the other candidate $z = ((\omega-1)/2, (\omega-1)/2, 1)$ we similarly compute the intersection of the line between $(0,0,0)$ and $(1,1,1)$ and the line between $((\omega-1)/2, (\omega-1)/2, 1)$ and $(1,1,0)$. This computation gives the optimal value $u = \frac{2}{5-\omega}$.

Comparing $\frac{4-\omega}{7-2\omega}$ and $\frac{2}{5-\omega}$ we see that on the feasible domain $2 \leq \omega \leq 3$ the former is smaller. $\square$

Now from Lemma 9.10 we obtain Strassen's Theorem 9.6, unconditionally, but with an adjusted exponent, namely $\frac{4-\omega}{7-2\omega}$ instead of $\frac{\omega}{3}$:

Corollary 9.11. *Let $a_i, b_i, c_i, a'_i, b'_i, c'_i \in \mathbb{N}$. Then*

$$\sum_i \mathrm{MM}_{a_i,b_i,c_i} \leq_{\widetilde{P}} \sum_i \mathrm{MM}_{a'_i,b'_i,c'_i} \implies \sum_i (a_i b_i c_i)^{\frac{4-\omega}{7-2\omega}} \leq \sum_i (a'_i b'_i c'_i)^{\frac{4-\omega}{7-2\omega}}.$$ [84]

Proof. Lemma 9.10 says that there is an element $\phi \in \mathcal{X}_{\mathrm{rect}}$ such that

$$\phi(\mathrm{MM}_{2,1,1}) = \phi(\mathrm{MM}_{1,2,1}) = \phi(\mathrm{MM}_{1,1,2}) = 2^{\frac{4-\omega}{7-2\omega}}.$$

Applying this ϕ to both sides of the given inequality $\sum_i \mathrm{MM}_{a_i,b_i,c_i} \leq_{\widetilde{P}} \sum_i \mathrm{MM}_{a'_i,b'_i,c'_i}$ and using Lemma 9.7 gives the claim. □

9.3. Type decompositions and convexity

In this subsection we will go on a diversion, and the reader can safely return to it after reading Part III. Recall that a central component in establishing connectivity of certain spectra was a type decomposition of polynomials that we developed in Section 6. There is a remarkable ubiquity of such notions of type decompositions in the context of convexity results in the asymptotic spectra literature of which we will now give a brief survey. These have in common that they "decompose" the nth power of an object as a sum of *polynomially many* (in n) objects (or as an abstract collection of polynomially many objects). One such type decomposition we have already discussed in detail in Part II, and the polynomial growth of the number of components (Lemma 6.6) played a crucial role in the proofs. We will see more kinds of type decompositions here, and how they help us better understand the asymptotic spectrum and its possible convexity structure. We expect such type decompositions to be the key to a general approach to convexity theorems within the theory of asymptotic spectra.

Polynomials. We begin with the type decomposition for powers of polynomials that we have discussed in Section 6, and on which the proof of convexity of the asymptotic spectrum of matrix multiplication and tensor networks relies. We recall that for any polynomial $f = \sum_{i=0}^d f_i x_i \in \mathbb{N}[x]$ the type decomposition of f^n is the sum

$$\sum_p [f^n]_p$$

where p goes over probability distributions on the $d+1$ monomials of f and the components of the decomposition are the monomials $[f^n]_p = \binom{n}{pn} \prod_i f_i^{p_i n} x^{i p_i n}$, which naturally appear when expanding the power f^n without collecting terms of the same degree.

We may think of this type decomposition as a *vanilla* or *domain-agnostic* type decomposition. Namely, it may be applied to any semiring generated by a single element by interpreting its elements as polynomials in $\mathbb{N}[x]$. This can be the subsemiring generated by a single tensor within the semiring of tensors, or the subsemiring generated by a single graph within the semiring of graphs. More generally, the *multivariate* version of the type decomposition for powers of polynomials can be applied to any finitely generated semiring [Str88, Section 5]. An example of this is the semiring generated by the matrix multiplication tensors $\mathrm{MM}_{2,1,1}$, $\mathrm{MM}_{1,2,1}$ and $\mathrm{MM}_{1,1,2}$, which is how the convexity result for matrix multiplication is proven.

[84] Comparing $\frac{4-\omega}{7-2\omega}$ to $\frac{\omega}{3}$, note that for $2 \leq \omega \leq 3$ we have $\frac{4-\omega}{7-2\omega} \leq \frac{\omega}{3}$ with equality when $\omega = 2$ (or when $\omega = 3$).

A distinctive feature of the type decomposition for powers of polynomials is that the components $[f^n]_p$ are again elements of the same univariate semiring $\mathbb{N}[x]$. Interestingly, we will now see that some *specialized* or *domain-specific* type decompositions, two important examples of which we are going to discuss next, do not have this property.

Tensors. We will now discuss two domain-specific type decompositions for powers of tensors (in the setting of Example 2.21). We refer to these as the *classical* decomposition and the *quantum* decomposition.

Classical decomposition of tensors. The classical decomposition is the simplest of the two and is defined for tensors over any ground field. This decomposition is based on a straightforward partitioning of the standard basis of the tensor space. This approach falls under the *method of types* from information theory [CT12, Chapter 11]. This decomposition plays a central role in the construction of Strassen's "support functionals" in [Str91]. We first describe the decomposition and then the related convexity result regarding the support functionals.

The basic ingredient for the classical decomposition is a direct sum decomposition of the vector space $(\mathbb{F}^d)^{\otimes n}$ induced by a set decomposition of the standard basis that respects a natural group action. We define the *type* of an n-tuple $s \in [d]^n$ as the d-tuple $w = (|\{j : s_j = i\}|)_{i\in[d]}$ that counts the number of occurrences of the elements of $[d]$ in s. For every type w we define $[(\mathbb{F}^d)^{\otimes n}]_w$ to be the subspace of $(\mathbb{F}^d)^{\otimes n}$ spanned by the elements $e_{s_1} \otimes \cdots \otimes e_{s_n}$ for all $s \in [d]^n$ with type w. Then

$$(\mathbb{F}^d)^{\otimes n} = \bigoplus_w [(\mathbb{F}^d)^{\otimes n}]_w$$

where the sum is over the types w of all elements in $[d]^n$ (of which there are polynomially many in n). Note that every component $[(\mathbb{F}^d)^{\otimes n}]_w$ is invariant under permuting the n tensor factors of $(\mathbb{F}^d)^{\otimes n}$.

The classical decomposition of powers of tensors $T^{\otimes n} \in (\mathbb{F}^{d_1} \otimes \mathbb{F}^{d_2} \otimes \mathbb{F}^{d_3})^{\otimes n}$ is defined by identifying $(\mathbb{F}^{d_1} \otimes \mathbb{F}^{d_2} \otimes \mathbb{F}^{d_3})^{\otimes n}$ with $(\mathbb{F}^{d_1})^{\otimes n} \otimes (\mathbb{F}^{d_2})^{\otimes n} \otimes (\mathbb{F}^{d_3})^{\otimes n}$ and then decomposing each factor $(\mathbb{F}^{d_i})^{\otimes n}$ to obtain the decomposition

$$(\mathbb{F}^{d_1})^{\otimes n} \otimes (\mathbb{F}^{d_2})^{\otimes n} \otimes (\mathbb{F}^{d_3})^{\otimes n} = \bigoplus_w [(\mathbb{F}^{d_1})^{\otimes n} \otimes (\mathbb{F}^{d_2})^{\otimes n} \otimes (\mathbb{F}^{d_3})^{\otimes n}]_w$$

where the sum is over all triples $w = (w_1, w_2, w_3)$ with w_i a type of an element in $[d_i]^n$ and where $[(\mathbb{F}^{d_1})^{\otimes n} \otimes (\mathbb{F}^{d_2})^{\otimes n} \otimes (\mathbb{F}^{d_3})^{\otimes n}]_w = [(\mathbb{F}^{d_1})^{\otimes n}]_{w_1} \otimes [(\mathbb{F}^{d_2})^{\otimes n}]_{w_2} \otimes [(\mathbb{F}^{d_3})^{\otimes n}]_{w_3}$.

For a tensor $T \in \mathbb{F}^{d_1} \otimes \mathbb{F}^{d_2} \otimes \mathbb{F}^{d_3}$ the *classical decomposition* of the nth power $T^{\otimes n}$, which is an element of $(\mathbb{F}^{d_1})^{\otimes n} \otimes (\mathbb{F}^{d_2})^{\otimes n} \otimes (\mathbb{F}^{d_3})^{\otimes n}$, is the decomposition

$$T^{\otimes n} = \sum_w [T^{\otimes n}]_w$$

where $[T^{\otimes n}]_w$ is the projection of $T^{\otimes n}$ to the subspace $[(\mathbb{F}^{d_1})^{\otimes n} \otimes (\mathbb{F}^{d_2})^{\otimes n} \otimes (\mathbb{F}^{d_3})^{\otimes n}]_w$.

The classical decomposition plays an important role for a construction of Strassen's support functionals [Str91], which are points in the asymptotic spectrum of a certain subsemiring of all tensors. This subsemiring is defined as follows. The support $\mathrm{supp}(T)$ of a tensor $T \in \mathbb{F}^{d_1} \otimes \mathbb{F}^{d_2} \otimes \mathbb{F}^{d_3}$ is called *oblique* if the elements form an antichain with respect to the coordinate-wise order on

$[d_1] \times [d_2] \times [d_3]$. The tensors with oblique support form a semiring. The construction of Strassen's support functionals is based on an optimization of an entropy-like function over all normalized types that appear in the classical decomposition of $T^{\otimes n}$, so that the support functionals measure the size of the components that appear in the decomposition. The support functionals form a family ζ^θ indexed by $\theta = (\theta_1, \theta_2, \theta_3)$ from the probability simplex Θ, a convex set. Thus the support functionals provide a convexly indexed subset of the asymptotic spectrum of oblique tensors. We do not know whether this gives the full spectrum or whether the spectrum itself is convex. The support functional construction covers all known elements in the asymptotic spectrum of oblique tensors. At the vertices $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ of the simplex Θ, the support functional ζ^θ is equal to the respective flattening ranks, and for every tensor T the minimum $\min_\theta \zeta^\theta(T)$ is equal to the asymptotic slice rank of T. We refer to [Str91], [CVZ18], [Zui18], [CLZ23] for more details on the support functionals, and in particular [CGLZ20] for an application to barriers for *rectangular* matrix multiplication (cf. Section 4.5).

Quantum decomposition of tensors. We now discuss the *quantum* decomposition. This decomposition is based on the fundamental *Schur–Weyl duality* of representation theory which is also sometimes referred to as the *quantum method of types* [Har05]. Schur–Weyl duality plays a central role in the construction of the *quantum functionals* in [CVZ18] and geometric complexity theory [BI11]. The quantum decomposition applies only to tensors over the complex numbers. Contrary to the classical decomposition discussed earlier, the quantum decomposition does not correspond to a set decomposition of the standard basis of the tensor space, but rather makes use of subspaces with a different kind of symmetry. We first describe the decomposition and then we discuss two important convexity results related to this decomposition.

The basic ingredient for the quantum decomposition is a direct sum decomposition of the vector space $(\mathbb{C}^d)^{\otimes n}$ that respects a natural group action. Namely, let the symmetric group $\mathfrak{S}_n$ act on $(\mathbb{C}^d)^{\otimes n}$ by permuting the n tensor factors (as before) and let the general linear group GL_d act on V by acting on each of the n tensor factors simultaneously. These actions commute and thus the product group $\mathfrak{S}_n \times \mathrm{GL}_d$ acts accordingly on $(\mathbb{C}^d)^{\otimes n}$. It is a central result in representation theory, called Schur–Weyl duality, that $(\mathbb{C}^d)^{\otimes n}$ decomposes into a direct sum of irreducible representations as

$$(\mathbb{C}^d)^{\otimes n} = \bigoplus_{\lambda \vdash_d n} [(\mathbb{C}^d)^{\otimes n}]_\lambda$$

where the sum is over the partitions λ of n of at most d parts (of which there are polynomially many in n), and such that every component $[(\mathbb{C}^d)^{\otimes n}]_\lambda$ is an irreducible $\mathfrak{S}_n \times \mathrm{GL}_d$ representation of the form $[(\mathbb{C}^d)^{\otimes n}]_\lambda = V_\lambda \otimes W_\lambda$ where V_λ is an irreducible $\mathfrak{S}_n$-representation of type λ and W_λ is an irreducible GL_d-representation of type λ.

We now describe the quantum decomposition of powers of tensors $T^{\otimes n} \in (\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \mathbb{C}^{d_3})^{\otimes n}$ using the decomposition of $(\mathbb{C}^d)^{\otimes n}$. The group $\mathfrak{S}_n \times (\mathrm{GL}_{d_1} \times \mathrm{GL}_{d_2} \times \mathrm{GL}_{d_3})$ acts on the space $(\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \mathbb{C}^{d_3})^{\otimes n}$ similarly as above. We may identify $(\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \mathbb{C}^{d_3})^{\otimes n}$ with the tensor product $(\mathbb{C}^{d_1})^{\otimes n} \otimes (\mathbb{C}^{d_2})^{\otimes n} \otimes (\mathbb{C}^{d_3})^{\otimes n}$ which decomposes as a sum of irreducible representations as

$$(\mathbb{C}^{d_1})^{\otimes n} \otimes (\mathbb{C}^{d_2})^{\otimes n} \otimes (\mathbb{C}^{d_3})^{\otimes n} = \bigoplus_{\lambda \vdash^3_d n} [(\mathbb{C}^{d_1})^{\otimes n} \otimes (\mathbb{C}^{d_2})^{\otimes n} \otimes (\mathbb{C}^{d_3})^{\otimes n}]_\lambda$$

where the sum is over triples of triples $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ with λ_i a partition of n into at most d_i parts (of which there are polynomially many) and each component in turn can be written as

$[(\mathbb{C}^{d_1})^{\otimes n} \otimes (\mathbb{C}^{d_2})^{\otimes n} \otimes (\mathbb{C}^{d_3})^{\otimes n}]_\lambda = [(\mathbb{C}^{d_1})^{\otimes n}]_{\lambda_1} \otimes [(\mathbb{C}^{d_2})^{\otimes n}]_{\lambda_2} \otimes [(\mathbb{C}^{d_3})^{\otimes n}]_{\lambda_3}$, where the $[(\mathbb{C}^{d_i})^{\otimes n}]_{\lambda_i}$ are the irreducible $\mathfrak{S}_n \times \mathrm{GL}_{d_i}$ representations as before.

Now let $T \in \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \mathbb{C}^{d_3}$. The *quantum decomposition* of $T^{\otimes n} \in (\mathbb{C}^{d_1})^{\otimes n} \otimes (\mathbb{C}^{d_2})^{\otimes n} \otimes (\mathbb{C}^{d_3})^{\otimes n}$ is the decomposition

$$T^{\otimes n} = \sum_{\lambda \vdash^3_d n} [T^{\otimes n}]_\lambda$$

where $[T^{\otimes n}]_\lambda$ is the projection of $T^{\otimes n}$ to the subspace $[(\mathbb{C}^{d_1})^{\otimes n} \otimes (\mathbb{C}^{d_2})^{\otimes n} \otimes (\mathbb{C}^{d_3})^{\otimes n}]_\lambda$.

The construction of a family of elements in the asymptotic spectrum of tensors by Christandl, Vrana and Zuiddam [CVZ18] relies on the quantum decomposition. The construction of these elements, called the *quantum functionals*, is based on an optimization of an entropy-like function over the so-called *moment polytope*. The moment polytope is defined as the Euclidean closure of the set of triples $(\lambda_1/n, \lambda_2/n, \lambda_3/n)$ over all $\lambda \vdash^3_d$ such that the component $[T^{\otimes n}]_\lambda$ in the quantum decomposition of $T^{\otimes n}$ is nonzero, and it is a central result that this set is indeed a convex polytope [Nes84, Bri87, Fra02]. Similarly as for the support functionals and the classical decomposition, the quantum functionals should be thought of as measuring the size of the components appearing in the quantum decomposition of powers of a tensor, asymptotically. They form a family F^θ that is indexed by elements $\theta = (\theta_1, \theta_2, \theta_3)$ from the probability simplex Θ, a convex set. Thus the quantum functionals provide a convexly indexed subset of the asymptotic spectrum of tensors. We do not, however, know whether they give the whole spectrum of complex tensors or whether the whole spectrum is a convex set. The quantum functional construction covers all known elements in the asymptotic spectrum. At the vertices $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ of the simplex Θ, the quantum functional F^θ is equal to the respective flattening ranks, and for every tensor T the minimum $\min_\theta F^\theta(T)$ is equal to the asymptotic slice rank of T. We refer to [CVZ18, Zui18, CLZ23] for more details on the quantum functionals, and again [CGLZ20] for an application to barriers for *rectangular* matrix multiplication (cf. Section 4.5)

Graphs. The convexity result of Vrana [Vra21] for graphs that we mentioned in the previous section is also built on a kind of type decomposition for powers of graphs, in the spirit of the method of types and the classical decomposition of tensors. Different from the previous cases, here the type decomposition does not write a power of a graph as a sum (which is disjoint union in this semiring) of graphs. Rather it associates to a power of a graph the collection of induced subgraphs for every choice of a type class in the vertex set. This type decomposition leads to a convex parametrization of the asymptotic spectrum of graphs.

Summarizing, we have touched on several examples of type decompositions — one domain-agnostic type decomposition (for polynomials) and several domain-specific type decompositions (for tensors and graphs) — and related convexity theorems. We believe it will be important in future work to further develop the understanding of type decompositions: What kind of type decompositions are good for proving convexity theorems? Which Strassen preorders allow for such type decompositions?

Part III
Compression of tensors

Contents

In Part I we have introduced the theory of asymptotic spectra, of which the main result is the duality theorem between the asymptotic preorder and the asymptotic spectrum. In Part II we have developed methods to prove structural properties of the asymptotic spectrum (connectedness, log-convexity, log-star-convexity), in particular the anchor method. This theory was developed in full generality, that is, for any semiring with a Strassen preorder.

In Part III we will specialize, and apply the anchor method to the asymptotic spectrum of matrix multiplication and the more general asymptotic spectra of tensor networks. As a consequence we find that all these spectra are log-star-convex, and in particular connected. While Strassen had already proved this result for the asymptotic spectrum of matrix multiplication, we extend his result to all asymptotic spectra of tensor networks.

A good intuitive way to think about what is happening in this part, is as an *error correction* result. This result gives a conversion from an algorithm which solves a large subset of the instances of a problem of a given size n, into an algorithm which solves the problem on *all* instances of a slightly smaller size, say $n/2$. Here the problem at hand is computing the multilinear polynomial associated with a tensor, and a "large subset" above turns out to be (due to our choice of anchors) a high-dimensional linear subspace of tensors. Concretely, for matrix multiplication one can imagine turning a faulty algorithm which multiplies $n \times n$ matrices with an additive error given by a bilinear map of low image dimension, into an algorithm which correctly multiplies *all* $(n/2) \times (n/2)$ matrices. We will discuss this concrete instance in Section 10 first as it is an interesting application in itself and indicates the ingredients that we will need for the general result on tensor networks.

This conversion is broken into two parts, each performing certain linear transformations on linear subspaces of tensors. The first, which is the subject of Section 11, is developing *basis-shifting theorems* for such spaces. These apply linear transformations to shift the location of spanning sets of our linear space; this may be viewed as a special notion of self-reducibility for computing tensors (equivalently, the multilinear polynomials they represent). For appropriately shifted subspaces we then develop, in Section 12, *compression theorems* which shrink the size of the tensor so that the given subspace becomes full-dimensional. This may be viewed as a form of *downward-self-reducibility* or *self-correction*. The combination allows us to prove that certain *flattening ranks* (Definition 12.6) of the tensors are anchors, from which we can deduce connectivity and log-star-convexity of spectra.

Elaborating on the above, we will thus implement the anchor method (Section 8) with the choice of anchor being any flattening rank (Definition 12.6). Central in this is that from the inequality

$$T \leq S + U$$

with S having small flattening rank, we will infer the inequality

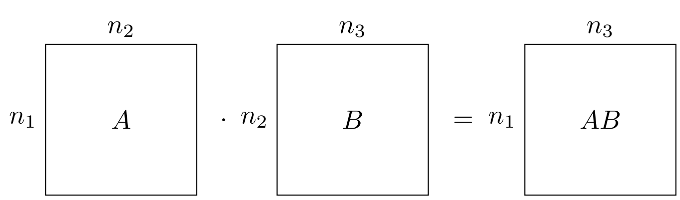
$$T' \leq U$$

with T' slightly smaller than T and of the same structure. In Section 10 we start with doing so for matrix multiplication.

10. Warm-up: compression (and error correction) of matrix multiplication

In this section we give a high-level intuitive description of the compression theorem for the special case of matrix multiplication tensors. We will discuss the important point that this compression theorem implies a surprising *error correction* result for matrix multiplication algorithms, which we think is of independent interest. In our intuitive description we will make sure highlight the important ingredients in the proof of the compression theorem, of which we will see generalizations and proofs in Section 11 and Section 12.

Generally we may think of 3-tensors as bilinear maps. Indeed, matrix multiplication is most naturally understood (and viewed pictorially) as a bilinear map, so that is the language we will use in this section. Let $T(A, B) = AB$ be the bilinear map $\mathbb{F}^{n_1 \times n_2} \times \mathbb{F}^{n_2 \times n_3} \to \mathbb{F}^{n_1 \times n_3}$ that multiplies the $n_1 \times n_2$ matrix A with the $n_2 \times n_3$ matrix B, so that T pictorially performs the following matrix multiplication operation:



The restriction preorder on tensors carries over to bilinear maps as a natural notion of reduction: for any two bilinear maps $S : V_1 \times V_2 \to V_3$ and $T : W_1 \times W_2 \to W_3$ we have $S \leq T$ if S can be obtained from T by applying linear maps to the inputs and output of T, that is, if there are linear maps ϕ_i so that $S(A,B) = \phi_3(T(\phi_1(A), \phi_2(B)))$. Here we will be dealing with matrix multiplication maps and so the spaces V_i and W_i will be matrix spaces.

The error correcting procedure that we will describe in this section achieves the following. Assume that we are handed a "faulty" bilinear map U for computing $n \times n$ matrix multiplication: Rather than computing T (the real matrix multiplication map), it computes $U = T - S$, where subtraction is coordinatewise, and where S is an adversarial bilinear map of low (say $< n^2/4$) image dimension (i.e., flattening rank).[85] Then (using linear operations on the inputs and outputs to U) we can use it to correctly compute the bilinear map T' for (say) $n/2 \times n/2$ matrix multiplication.

Formally, as a starting point we assume that we can write the matrix multiplication map T as a coordinatewise sum

$$T = S + U \tag{10.1}$$

of two bilinear maps S and U that are also of the format $\mathbb{F}^{n_1 \times n_2} \times \mathbb{F}^{n_2 \times n_3} \to \mathbb{F}^{n_1 \times n_3}$.[86] Further, we assume that the image $\mathrm{Im}(S) = \{S(A,B) : A \in \mathbb{F}^{n_1 \times n_2}, B \in \mathbb{F}^{n_2 \times n_3}\}$ of S has low dimension.[87] We will prove that then $T' \leq U$, where T' is a slightly smaller matrix multiplication tensor. For example, if $n_1 = n_2 = n_3 = n$ then we may take T' to be the $n/2 \times n/2$ matrix multiplication tensor.

The proof has two separate parts, which will be (respectively) elaborated in Section 11 and Section 12. The first, applies the same row operation L to the rows of A and AB, and R to the columns of B and AB — these clearly do not affect T (this is called "the invariance property" in Section 12.4). Its purpose is to strategically structure a basis for the (low-dimensional) space $\mathrm{Im}(S)$ (this is called "basis shifting" in Section 11). The second part ("compression", in Section 12) is applying a restriction to T, which combines projections on the A-leg and B-leg of the tensor T, as well as a linear operation Q applied to the "AB-leg" of the tensor T (i.e., the output of T as a bilinear map). These together create a slightly smaller matrix multiplication tensor T', on which the contribution of S is 0. Combined, these two parts prove that $T' \leq U$. We will discuss the two parts further in the rest of this section.

In Section 11 and Section 12 these parts will be explained and proved in the general setting of tensor networks, of which matrix multiplication is the simplest example.

Step 1: Invariance property and matrix subspace basis shifting. As mentioned before, the matrix multiplication map $T(A,B) = AB$ has the obvious invariance property that for any invertible linear operations $L : \mathbb{F}^{n_1} \to \mathbb{F}^{n_1}$ and $R : \mathbb{F}^{n_3} \to \mathbb{F}^{n_3}$ we have

$$L^{-1}T(LA, BR)R^{-1} = L^{-1}LABRR^{-1} = AB.$$

(This invariance of the matrix multiplication map T under this action by L and R, sometimes referred to as *sandwiching action*, we will later generalize to tensor networks as the *invariance property*, Lemma 12.16)

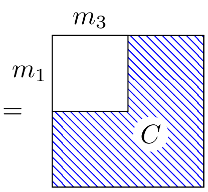
In this first step, we use this invariance property for special L and R applied to both sides of the equality $T = S + U$, to modify S and in particular restructure its image subspace $\mathrm{Im}(S)$. Note that this image subspace $\mathrm{Im}(S) \subseteq \mathbb{F}^{n_1 \times n_3}$ is a subspace of matrices. A good way to view a subspace of matrices (of dimension d) is as a single matrix with entries that are d-vectors. Using the assumption

[85] To appreciate the power of such an adversary, an example of such S is adding, to each of an arbitrary subset of the output matrix of the algorithm (of size $n^2/4$), arbitrary linear forms of all output entries.

[86] Technically the starting point will be that $T \leq S \oplus U$, which implies (10.1) by slightly changing S, U.

[87] This is the same as S (as a tensor) having low flattening rank on the $n_1 \times n_3$ leg.

of low dimension of the matrix subspace $\mathrm{Im}(S)$, say $d = \dim(\mathrm{Im}(S))$, we will be able to find (as will be proved in Section 11) invertible matrices $L \in \mathbb{F}^{n_1 \times n_1}$ and $R \in \mathbb{F}^{n_3 \times n_3}$ such that the matrix subspace $L\,\mathrm{Im}(S)R = \mathrm{Im}(LSR)$ has the following special structure.[88] (This is what we call a basis shifting theorem in Section 11.) The special structure is that the $m_1 \times m_3$ top-left entries of the new subspace $L\,\mathrm{Im}(S)R$, thought of as d-vectors, are in the linear span of the other entries (for parameters m_1, m_3 that depend on the actual dimension of the image of S):

$\mathrm{Im}(LSR) \;=$ 

(10.2)

In the terminology we will formally introduce in Section 11, $\mathrm{Im}(LSR)$ is C-spanning where C is the blue-shaded region in the picture. For example, to get a feeling for the numbers, if in T we have that $n_1 = n_2 = n_3 = n$, and the "low" dimension of $\mathrm{Im}(S)$ was $(1/4)n^2$, then we will get from the shifting theorem (Theorem 11.8) that we may take $m_1 = m_3 = n/2$.

We have thus applied L and R to the output of S to ensure the image subspace has special structure. We now also apply their inverses to the inputs as follows so that we can apply the invariance property of T. To this end we define the bilinear map S' by $S'(A, B) = LS(L^{-1}A, BR^{-1})R$. The subspace $\mathrm{Im}(S')$ equals the subspace $\mathrm{Im}(LSR)$ and is thus still C-spanning (as in (10.2)).

Applying the above maps L and R to both sides of our assumed starting equation $S = T + U$ we find the equation

$$LT(L^{-1}A, BR^{-1})R = LS(L^{-1}A, BR^{-1})R + LU(L^{-1}A, BR^{-1})R.$$

We already know that T has the invariance property that $T(A, B) = LT(L^{-1}A, BR^{-1})R$. So, defining the bilinear map U' by $U'(A, B) = LU(L^{-1}A, BR^{-1})R$ we find

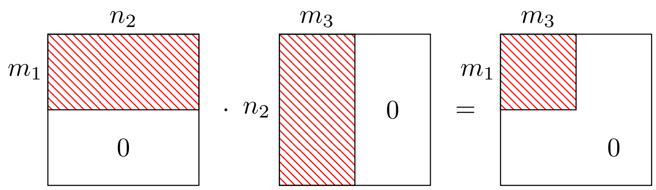
$$T = S' + U'.$$

We have thus transformed the starting equation $T = S + U$ into a new equation $T = S' + U'$ so that the image of S' is C-spanning. In particular, what we will need from this in the next step is that the C-spanning property implies that there is a linear map $Q : \mathbb{F}^{n_1 \times n_3} \to \mathbb{F}^{n_1 \times n_3}$ which adds linear combinations of the entries in the blue region (as in (10.2)) to arbitrary entries in a way that $Q\,\mathrm{Im}(S') = 0$ and the composed map QS' is the zero map.[89]

Step 2: Projection property and compression. In the second step we use the fact (which we will later generalize to tensor networks as the *projection property*, Lemma 12.13) that T becomes a smaller matrix multiplication map T' when we restrict the input matrices appropriately. Namely, we may set certain rows of A and columns of B to 0 so as to enforce the previously defined blue-shaded region C (as in (10.2)) in the product AB to 0. Pictorially this looks as follows:

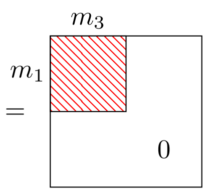
[88] Here LSR denotes the bilinear map obtained by multiplying the output of S, which is a matrix, from the left by L and from the right by R.

[89] Notice that Q acts on one leg of the tensor, so a legitimate restriction operation.



And so the new bilinear map $T'(\ ,\ ) = T(\ ,\ )$ is indeed precisely a slightly smaller matrix multiplication map $\mathbb{F}^{m_1\times n_2} \times \mathbb{F}^{n_2\times m_3} \to \mathbb{F}^{m_1\times m_3}$ (with the m_i slightly smaller than the n_i as will be determined by the shifting theorem).

Now we use the linear map $Q : \mathbb{F}^{n_1\times n_3} \to \mathbb{F}^{n_1\times n_3}$ previously found in Step 1. It is clear that this map Q acts as the identity on the matrix subspace

$\mathrm{Im}(T') =$ 

as it will just be adding zeroes to the red region. Moreover, recall that we constructed Q so that the composition QS' equals zero. We thus find that

$$T'(\ ,\ ) = QT(\ ,\ ) = QS'(\ ,\ ) + QU'(\ ,\ ) = QU'(\ ,\ ),$$

since $QS'(\ ,\ ) = 0$. It follows that

$$T' \leq U' \leq U,$$

which is what we wanted. This finishes the high-level description.

Repeating the "error correction" interpretation above, any erroneous bilinear algorithm U for $n \times n$ matrix multiplication, in which errors are introduced by the addition of a low dimensional bilinear map S, can be converted by linear operations to its inputs and outputs into a correct bilinear $n/2 \times n/2$ matrix multiplication algorithm T'.

11. Basis shifting of tensors of vectors

In this section we develop a method called *basis shifting* for tensors of vectors, extending a result of Strassen [Str88] that powered his proof that the asymptotic spectrum of matrix multiplication is log-convex. We will throughout shorten the term *basis shifting* to just *shifting*. Our proof of shifting is simpler and provides a more precise analysis. Our extension will enable us to prove new convexity results for the asymptotic spectrum of tensor networks.

A *tensor of vectors* is a collection $\mathcal{V} = (v_i : i \in I)$ of vectors $v_i \in \mathbb{F}^k$ that is indexed by a finite set of the form $I = [n_1] \times \cdots \times [n_\ell]$. Thus $\mathcal{V}$ is an ℓ-dimensional array (or tensor) of vectors. The group $\mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_\ell}$ acts naturally on $\mathcal{V}$ by taking linear combinations of the vectors v_i in a way compatible with the structure of I (Definition 11.4). Let $\mathcal{U} = (u_i : i \in I) \in (\mathbb{F}^k)^I$ be a new family obtained from $\mathcal{V}$ by acting with an element of $\mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_\ell}$. *Shifting theorems* say which subsets of $\mathcal{U}$ are a spanning set of $\mathrm{span}(\mathcal{U})$. If $B \subseteq I$ and $\{u_i : i \in B\}$ is a spanning set of $\mathcal{U}$, then we say that $\mathcal{U}$ is B-*spanning*. Our task is to make, via linear transformations taking $\mathcal{V}$ to $\mathcal{U}$, the family $\mathcal{U}$ to be B-spanning for "as many", and "as generally structured" sets B as possible.

The simplest shifting theorem, for when the dimension $\ell = 1$, and the index set is $I = [n]$, is as follows. Let $d \coloneqq \dim(\mathrm{span}(\mathcal{V}))$. Suppose $d \leq b = (b+1) - 1 \leq n$.[90] Clearly there is a subset $B \subseteq I$ of cardinality b such that $\{v_i : i \in B\}$ is a spanning set of $\mathcal{V}$ and so $\mathcal{V}$ is B-spanning. Let $\mathcal{U} = (u_i : i \in I)$ be obtained from $\mathcal{V}$ by taking random linear combinations of the v_i. Then with high probability (which goes to 1 when the size of the field $\mathbb{F}$ grows) $\mathcal{U}$ is B-spanning for *any* subset $B \subseteq [n]$ of cardinality b (Lemma 11.6). The proof of this statement is a simple application of the Schwartz–Zippel lemma [DL78, Sch80, Zip79] to matrix minors.

The next case, when the dimension $\ell = 2$, and $I = [n_1] \times [n_2]$, is much more interesting. Strassen proved, implicitly, the following shifting theorem for this case. Let $d \coloneqq \dim(\mathrm{span}(\mathcal{V}))$. Let $b_i \leq n_i$ be integers such that $d \leq (b_1+1)(b_2+1) - 1$. Let the family $\mathcal{U} = (u_i : i \in I)$ be obtained from $\mathcal{V}$ by taking random linear combination of the v_i by acting with $\mathrm{GL}_{n_1} \times \mathrm{GL}_{n_2}$. Then, with high probability (which goes to 1 when the size of the field $\mathbb{F}$ grows), for any set $B = (B_1 \times [n_2]) \cup ([n_1] \times B_2) \subseteq I$ where B_i has cardinality b_i, the family $\mathcal{U}$ is B-spanning (Theorem 11.8). Our proof of this theorem relies on a conditional version of Lemma 11.6 which effectively reduces the problem to dimension 1 via a combinatorial argument. For sets B that have this structure, this theorem is optimal in the sense that the b_i cannot be chosen smaller.

As a consequence of the generality and simplicity of our proof, we can extend the shifting theorem of Strassen to the general case $I = [n_1] \times \cdots \times [n_\ell]$. Let $d \coloneqq \dim(\mathrm{span}(\mathcal{V}))$. Let $b_i \leq n_i$ be integers such that $d \leq (b_1+1)\cdots(b_\ell+1) - 1$. Let $\mathcal{U} = (u_i : i \in I)$ be obtained from $\mathcal{V}$ by taking random linear combination of the v_i by acting with $\mathrm{GL}_{n_1} \times \cdots \times \mathrm{GL}_{n_\ell}$. Then, with high probability (which goes to 1 when the size of the field $\mathbb{F}$ grows), for any set $B = (B_1 \times [n_2] \times \cdots \times [n_\ell]) \cup ([n_1] \times B_2 \times \cdots \times [n_\ell]) \cup ([n_1] \times [n_2] \times \cdots \times B_\ell) \subseteq I$ where B_i has cardinality b_i, the family $\mathcal{U}$ is B-spanning (Theorem 11.10). We prove Theorem 11.10 by proving a conditional version, by induction over ℓ, with a conditional version of Lemma 11.6 as the base case. For sets B that have this structure, this theorem is optimal in the sense that the b_i cannot be chosen smaller.

The organization of this section is as follows. We begin with setting up formal definitions around tensors of vectors and conditional dimension in Section 11.1. After that we gradually build towards the most general result in Section 11.4. We warn the reader that while the proofs get more elaborate notationally in each subsection, the main idea stays the same. We will end with a discussion of the shifting result in the language of subspaces of tensors and how it relates to other lines of work regarding rank and dimension for linear subspaces of matrices and tensors in Section 11.5.

11.1. Conditional dimension and group action on tensors of vectors

We start by setting up some notation for this section. Let $\mathbb{F}$ be a field.

Definition 11.1 (Conditional dimension)**.** Let V and W be subspaces of $\mathbb{F}^k$. As usual, we denote by $\dim(V)$ the dimension of V. We define the *conditional dimension of V given W* as

$$\dim(V|W) \coloneqq \dim(V+W) - \dim(W).\text{[91]}$$

[90] It will soon become clear why we write b as $(b+1)-1$.

[91] Our notation for conditional entropy resembles the standard notation used in information theory for conditional entropy $H(X|Y)$, which has an analogous definition $H(X|Y) = H(X,Y) - H(Y)$.

The conditional dimension can be thought of as the dimension of a quotient vector space,

$$\dim(V|W) = \dim((V + W)/W).$$

Equivalently, via the general equality $\dim(V + W) = \dim(V) + \dim(W) - \dim(V \cap W)$, we may write the conditional dimension as

$$\dim(V|W) = \dim(V) - \dim(V \cap W).$$

Let $\mathcal{V}$ and $\mathcal{W}$ be subsets of $\mathbb{F}^k$ that are not necessarily subspaces. We naturally define the *dimension of* $\mathcal{V}$ as

$$\dim(\mathcal{V}) \coloneqq \dim(\operatorname{span}(\mathcal{V})).$$

We define the *conditional dimension of* $\mathcal{V}$ *given* $\mathcal{W}$ as

$$\dim(\mathcal{V}|\mathcal{W}) \coloneqq \dim(\operatorname{span}(\mathcal{V})|\operatorname{span}(\mathcal{W})).$$

Remark 11.2. The conditional dimension satisfies

$$0 \leq \dim(V|W) \leq \dim(V).$$

These inequalities are tight. Namely, we have $\dim(V|W) = 0$ if and only if $V \subseteq W$. On the other hand, we have $\dim(V|W) = \dim(V)$ if and only if $V \cap W = \{0\}$.

Let I be a finite index set. Suppose that for each $i \in I$ we are given a vector $v_i \in \mathbb{F}^k$. Let us denote this tensor of vectors by $\mathcal{V}$, that is,

$$\mathcal{V} = (v_i : i \in I).$$

The structure of the index set I will not play a role until we define a group action on tensors of vectors.

Definition 11.3 (B-spanning tensor of vectors)**.** Let $\mathcal{V} = (v_i : i \in I) \subseteq (\mathbb{F}^k)^I$. Let $B \subseteq I$ be a subset of the index set. We will use the notation $\overline{B} \coloneqq I \setminus B$ to denote the *complement* of B in I. We define the family $\mathcal{V}_B$ by

$$\mathcal{V}_B \coloneqq (v_i : i \in B).$$

We call $\mathcal{V}_B$ the *restriction of* $\mathcal{V}$ *to* B. We thus have $\mathcal{V} = \mathcal{V}_B \cup \mathcal{V}_{\overline{B}}$. We say that the family $\mathcal{V}$ is B*-spanning* if

$$\dim(\mathcal{V}|\mathcal{V}_B) = 0.$$

In other words, $\mathcal{V}$ is B-spanning if $\dim(\mathcal{V}_B) = \dim(\mathcal{V})$, or equivalently, if $\operatorname{span}(\mathcal{V}_B) = \operatorname{span}(\mathcal{V})$.

Definition 11.4 (Group action on tensors of vectors)**.** Let $\mathcal{V} = (v_i : i \in I)$ be a tensor of vectors that is indexed by $I = [n_1] \times \cdots \times [n_\ell]$, so that $\mathcal{V}$ has the structure of an ℓ-dimensional array of vectors $v_{(i_1,\ldots,i_\ell)} \in \mathbb{F}^k$. Let $\Gamma = \mathrm{GL}_{n_1}(\mathbb{F}) \times \cdots \times \mathrm{GL}_{n_\ell}(\mathbb{F})$ be the group of ℓ-tuples of invertible matrices of dimension $n_i \times n_i$, respectively. We define the action of Γ on $\mathcal{V}$ in the same way that Γ would act on an ℓ-tensor of format $n_1 \times \cdots \times n_\ell$. Namely, for $M = (M_1, \ldots, M_\ell) \in \Gamma$ we define the family

$$M \cdot \mathcal{V} \subseteq \mathbb{F}^k$$

by $M \cdot \mathcal{V} = (u_{(i_1,\ldots,i_\ell)} : i \in I)$ where

$$u_{(i_1,\ldots,i_\ell)} = \sum_{(j_1,\ldots,j_\ell)\in I} v_{(j_1,\ldots,j_\ell)} \, (M_1)_{i_1,j_1} \cdots (M_\ell)_{i_\ell,j_\ell}.$$

For example, when $I = [n]$ the action of $\Gamma = \mathrm{GL}_n(\mathbb{F})$ on $\mathcal{V} = (v_1, \ldots, v_n)$ simply takes linear combinations of the vectors v_i. Or, when $I = [n_1] \times [n_2]$ the action of $\Gamma = \mathrm{GL}_{n_1}(\mathbb{F}) \times \mathrm{GL}_{n_2}(\mathbb{F})$ on $\mathcal{V} = (v_{(i,j)})_{(i,j)\in I}$ takes linear combinations of the vectors $v_{(i,j)}$.

11.2. Shifting of one-dimensional tensors of vectors

The shifting theorem that we will discuss now is the simplest one and demonstrates the idea of shifting. We will use it later to prove the other shifting theorems.

Let $I = [n]$ and let $\mathcal{V} = (v_i : i \in I) \in (\mathbb{F}^k)^I$. As we discussed in Section 11.1, we think of $\mathcal{V}$ as a one-dimensional array of vectors v_i and the group $\Gamma = \mathrm{GL}_n(\mathbb{F})$ acts on $\mathcal{V}$ by taking linear combinations of the vectors v_i. Namely, for every $M \in \Gamma$ we defined $M \cdot \mathcal{V} = (u_i : i \in I)$ where $u_i = \sum_{j\in I} v_j \, M_{i,j}$.

We begin with a trivial fact and then give a more robust version of that fact which follows from the Schwartz–Zippel lemma.

Fact 11.5. *Let $d = \dim(\mathcal{V})$. Let b be an integer for which $d \leq b \leq n$. Then there exists a subset $B \subseteq [n]$ of cardinality $|B| = b$ such that $\mathcal{V}$ is B-spanning.*

For any subset $S \subseteq \mathbb{F}$ we let $\mathrm{GL}_n(S)$ denote the subset of matrices in $\mathrm{GL}_n(\mathbb{F})$ with coefficients in S. We use this notation to give concrete bounds (using the Schwartz-Zippel lemma) on the measures (ratio) of "good" matrices in the results that follow. In the following, S will always be a finite subset. The field $\mathbb{F}$ is arbitrary.

Lemma 11.6. *Let $S \subseteq \mathbb{F}$ be finite. Let $d = \dim(\mathcal{V})$. Let b be an integer for which $d \leq b \leq n$.*

(i) *Except for a measure at most $d/|S|$ of all $M \in \mathrm{GL}_n(S)$, the family $M \cdot \mathcal{V}$ is $[b]$-spanning.*

(ii) *Except for a measure at most $\binom{n}{b} d/|S|$ of all $M \in \mathrm{GL}_n(S)$, for every subset $B \subseteq [n]$ of cardinality b, the family $M \cdot \mathcal{V}$ is B-spanning.*

Proof. (i) Let $\mathcal{U} = \{u_1, \ldots, u_n\} = M \cdot \mathcal{V}$ for any $M \in \mathrm{GL}_n(\mathbb{F})$. The family $\mathcal{U}$ is *not* $[b]$-spanning if and only if the span of $u_1, \ldots, u_b$ has dimension strictly less than d. The span of $u_1, \ldots, u_b$ has dimension strictly less than d if and only if every $d \times d$ minor of the matrix with columns $u_1, \ldots, u_b$ is zero. Every u_i is a linear combination of the v_j, namely $u_i = \sum_j v_j M_{i,j}$. Therefore, there is a finite collection of polynomials $\{f_k\}$ in the coefficients of M and of degree d (namely, the aforementioned $d \times d$ minors) so that: for all k we have $f_k(M) = 0$ if and only if $\mathcal{U}$ is not $[b]$-spanning. Since $d \leq b$, we know there exists an $M \in \mathrm{GL}_n(\mathbb{F})$ such that $\mathcal{U}$ is $[b]$-spanning, and therefore at least one of the f_k is not the zero polynomial. For this nonzero polynomial g we have by Schwartz–Zippel that for the coefficients $M_{i,j}$ chosen randomly, independently and uniformly from the finite set $S \subseteq \mathbb{F}$ we have

$$\Pr_M[\mathcal{U} \text{ is not } [b]\text{-spanning}] = \Pr_M[\forall k \, f_k(M) = 0] \leq \Pr_M[g(M) = 0] \leq \frac{d}{|S|}.$$

This proves the claim.

(ii) This claim follows from claim (i) by taking a union bound over all possible choices of B. $\square$

We will need the following more refined (and more versatile) version of the above Lemma 11.6 that applies to conditional dimensions (Definition 11.1). To see how Lemma 11.6 relates to its refinement, recall that Lemma 11.6 (i) says that, except for a measure at most $\dim(\mathcal{V})/|S|$ of all elements $M \in \mathrm{GL}_n(S)$, the family $M \cdot \mathcal{V}$ is $[b]$-spanning. For $\mathcal{U} = M \cdot \mathcal{V}$, saying $\mathcal{U}$ is $[b]$-spanning is equivalent to saying that $\dim(\mathcal{U} \mid \mathcal{U}_{[b]}) = 0$. Rather than obtaining $\dim(\mathcal{U} \mid \mathcal{U}_{[b]}) = 0$ we will now be aiming to get $\dim(\mathcal{U} \mid \mathcal{W} \cup \mathcal{U}_{[b]}) = 0$ for some arbitrary given tensor of vectors $\mathcal{W}$.

Lemma 11.7. *Let $S \subseteq \mathbb{F}$ be finite. Let $\mathcal{W}$ be an arbitrary tensor of vectors in $\mathbb{F}^k$ and let $d = \dim(\mathcal{V}|\mathcal{W})$ be the conditional dimension of $\mathcal{V}$ given $\mathcal{W}$. Let b be any integer such that $d \leq b \leq n$.*

(i) *Except for a measure at most $d/|S|$ of all $M \in \mathrm{GL}_n(S)$, for the family $\mathcal{U} = M \cdot \mathcal{V}$, we have that $\dim(\mathcal{U}|\mathcal{W} \cup \mathcal{U}_{[b]}) = 0$.*

(ii) *Except for a measure at most $\binom{n}{b} d/|S|$ of all $M \in \mathrm{GL}_n(S)$, for the family $\mathcal{U} = M \cdot \mathcal{V}$, for every set $B \subseteq [n]$ of cardinality b, we have that $\dim(\mathcal{U}|\mathcal{W} \cup \mathcal{U}_B) = 0$.*

(iii) *Let $\mathcal{V}^1, \ldots, \mathcal{V}^m$ be a collection of tensors $\mathcal{V}^j = (v_i^j : i \in [n])$ and define the conditional dimensions $d^{(j)} = \dim(\mathcal{V}^j|\mathcal{W} \cup \mathcal{V}^1 \cup \cdots \cup \mathcal{V}^{j-1})$. Let $b^{(j)}$ be integers for which $d^{(j)} \leq b^{(j)} \leq n$. Except for a measure at most $\sum_{j=1}^m \binom{n}{b^{(j)}} d^{(j)}/|S|$ of all $M \in \mathrm{GL}_n(S)$, for $\mathcal{U}^j = M \cdot \mathcal{V}^j$, for every collection of sets $B^j \subseteq [n]$ each of size $b^{(j)}$, we have that*

$$\dim(\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^m|\mathcal{W} \cup \mathcal{U}^1_{B^1} \cup \cdots \cup \mathcal{U}^m_{B^m}) = 0.$$

Proof. (i) This claim can be obtained from Lemma 11.6 (i) by applying it to the quotient space $(V + W)/W$. This claim can also be proven directly, similarly to the proof of Lemma 11.6 (i). Namely, in the proof of Lemma 11.6 (i) we took $\mathcal{U} = \{u_1, \ldots, u_n\} = M \cdot \mathcal{V}$ for any $M \in \mathrm{GL}_n(S)$. We then considered the matrix with columns $u_1, \ldots, u_b$ and reasoned about the fraction of choices of M for which this matrix has rank strictly less than d. For the proof of claim (i) rather than taking this matrix we take the matrix with columns $u_1, \ldots, u_b, w_1, \ldots, w_r$ where the w_i are elements of $\mathcal{W}$ so that $v_1, \ldots, v_n, w_1, \ldots, w_r$ span $\mathrm{span}(\mathcal{V}) + \mathrm{span}(\mathcal{W})$. Then we consider the $(d+r) \times (d+r)$ minors of this matrix corresponding to all the submatrices that contain d of the columns $u_1, \ldots, u_b$ and all the columns $w_1, \ldots, w_r$, and then proceed as in the proof of Lemma 11.6 (i) to finish the proof, noting that these minors are of degree d in the coefficients of M.

(ii) This claim follows from claim (i) by a union bound over all possible choices of B.

(iii) The proof is by induction on m. The base case $\dim(\mathcal{U}^1|\mathcal{W} \cup \mathcal{U}^1_{B^1}) = 0$ follows from (ii). For the induction step, we assume that

$$\dim(\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^{m-1} \mid \mathcal{W} \cup \mathcal{U}^1_{B^1} \cup \cdots \cup \mathcal{U}^{m-1}_{B^{m-1}}) = 0$$

holds for every collection of sets B^j $(j = 1, \ldots, m-1)$, for all $M \in \mathrm{GL}_n(S)$ except for a measure at most $\sum_{j=1}^{m-1} \binom{n}{b^{(j)}} d^{(j)}/|S|$. This implies $\mathrm{span}(\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^{m-1}) = \mathrm{span}(\mathcal{W} \cup \mathcal{U}^1_{B^1} \cup \cdots \cup \mathcal{U}^{m-1}_{B^{m-1}})$ and therefore

$$\begin{aligned}
&\dim(\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^{m-1} \cup \mathcal{U}^m \mid \mathcal{W} \cup \mathcal{U}^1_{B^1} \cup \cdots \cup \mathcal{U}^{m-1}_{B^{m-1}} \cup \mathcal{U}^m_{B^m}) \\
&\quad = \dim(\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^{m-1} \cup \mathcal{U}^m \mid \mathcal{W} \cup \mathcal{U}^1 \cup \cdots \cup \mathcal{U}^{m-1} \cup \mathcal{U}^m_{B^m}) \\
&\quad = \dim(\mathcal{W}' \cup \mathcal{U}^m|\mathcal{W} \cup \mathcal{W}' \cup \mathcal{U}^m_{B^m}) \\
&\quad = \dim(\mathcal{U}^m|\mathcal{W} \cup \mathcal{W}' \cup \mathcal{U}^m_{B^m})
\end{aligned}$$

where $\mathcal{W}' = \mathcal{U}^1 \cup \cdots \cup \mathcal{U}^{m-1}$. By (ii) we have that $\dim(\mathcal{U}^m|\mathcal{W} \cup \mathcal{W}' \cup \mathcal{U}^m_{B^m}) = 0$ for all B^m except for a measure at most $\binom{n}{b^{(m)}} d^{(m)}/|S|$ of all $M \in \mathrm{GL}_n(S)$. Using the union bound we get the claim. $\square$

11.3. Shifting of two-dimensional tensors of vectors

In the previous subsection we discussed shifting for one-dimensional tensors of vectors. We now move up one dimension and prove a generalization of the shifting theorem for two-dimensional tensors of vectors of Strassen. Let $I = [n_1] \times [n_2]$ and let $\mathcal{V} = (v_{i,j} : (i,j) \in I) \in (\mathbb{F}^k)^I$. We think of $\mathcal{V}$ as a matrix of vectors $v_{i,j}$. The group $\Gamma = \mathrm{GL}_{n_1}(\mathbb{F}) \times \mathrm{GL}_{n_2}(\mathbb{F})$ acts on $\mathcal{V}$ by taking linear combinations of the rows and columns of $\mathcal{V}$ viewed as a matrix, as defined in Section 11.1. Our goal is to shift $\mathcal{V}$ to a set B which is the union of some rows and some columns by acting with Γ.

Theorem 11.8 (Strassen)**.** *Let $S \subseteq \mathbb{F}$ be finite. Let $d = \dim(\mathcal{V})$. Suppose that $b_1 \leq n_1$ and $b_2 \leq n_2$ are integers such that $d \leq (b_1+1)(b_2+1) - 1$.*

(i) *For all but a measure at most $d/|S|$ of all $M_1 \in \mathrm{GL}_{n_1}(S)$ and a measure at most $d/|S|$ of all $M_2 \in \mathrm{GL}_{n_2}(S)$ we have that $\mathcal{U} = (M_1, M_2) \cdot \mathcal{V}$ is B-spanning for $B = ([b_1] \times [n_2]) \cup ([n_1] \times [b_2])$.*

(ii) *For all but a measure at most $\binom{n_1}{b_1} d/|S|$ of all $M_1 \in \mathrm{GL}_{n_1}(S)$ and a measure at most $\binom{n_2}{b_2} d/|S|$ of all $M_2 \in \mathrm{GL}_{n_2}(S)$ we have that $\mathcal{U} = (M_1, M_2) \cdot \mathcal{V}$ is B-spanning for all B of the form $B = (B_1 \times [n_2]) \cup ([n_1] \times B_2)$ with $|B_1| = b_1$ and $|B_2| = b_2$.*

Proof. (i) Let $\mathcal{U} = (M_1, M_2) \cdot \mathcal{V}$ for any $M_1 \in \mathrm{GL}_{n_1}(S)$ and $M_2 \in \mathrm{GL}_{n_2}(S)$. Then $\mathcal{U} \in (\mathbb{F}^k)^{[n_1] \times [n_2]}$ is a matrix of vectors. We let $\mathcal{U}^1, \ldots, \mathcal{U}^{n_2} \in (\mathbb{F}^k)^{[n_1]}$ be the "columns" of $\mathcal{U}$, so that each $\mathcal{U}^i$ is a vector of vectors. We define the conditional dimensions $d_i = \dim(\mathcal{U}^i \mid \mathcal{U}^1 \cup \cdots \cup \mathcal{U}^{i-1})$ for $i = 1, \ldots, n_2$. The numbers d_i depend on the choice of M_2 (and they do not depend on the choice of M_1). We are interested in the M_2 that have the following property among all choices of M_2: d_1 is maximized, and conditioned on d_1 being maximized, d_2 is maximized, etc. for the other d_i. For such M_2 we have that the d_i satisfy $d_1 \geq d_2 \geq \cdots \geq d_{n_2}$. Considering $d_i \times d_i$ minors and using a union bound as in the proof of Lemma 11.6 we find that all $M_2 \in \mathrm{GL}_{n_2}(S)$ satisfy the above property except for a measure at most $\sum_{j=1}^{n_2} d_j / |S| = d/|S|$.

Given any maximizing choice of M_2, we have, except for a measure $\sum_{j=1}^{n_2} d_j/|S| = d/|S|$ of all matrices $M_1 \in \mathrm{GL}_{n_1}(S)$, for $B^j = [d_j] \times \{j\}$ that $\dim(\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^j | \mathcal{U}^1_{B^1} \cup \cdots \cup \mathcal{U}^j_{B^j}) = 0$ by an argument similar to the proof of Lemma 11.7 (iii). Suppose that $\cup_j B^j$ is not contained in B. Then $d_{b_2+1} \geq b_1 + 1$. It follows that $d = \sum_{i=1}^{n_2} d_i \geq \sum_{i=1}^{b_2+1} d_i \geq (b_2+1)(b_1+1)$. This contradicts our assumption about the dimension d. We conclude that $\cup_j B^j$ is contained in B. This implies that $\mathcal{U}^1_{B^1} \cup \cdots \cup \mathcal{U}^j_{B^j}$ is a subset of $\mathcal{U}_B$, so $\mathcal{U}$ is B-spanning.

(ii) This claim follows from claim (i) by a union bound over all possible choices for B. $\square$

We will need the following more refined version of Theorem 11.8 that applies to conditional dimensions.

Theorem 11.9 (Strassen)**.** *Let $S \subseteq \mathbb{F}$ be finite. Let $\mathcal{W}$ be an arbitrary tensor of vectors in $\mathbb{F}^k$ and let $d = \dim(\mathcal{V}|\mathcal{W})$ be the conditional dimension of $\mathcal{V}$ given $\mathcal{W}$. Suppose that $b_1 \leq n_1$ and $b_2 \leq n_2$ are integers such that $d \leq (b_1+1)(b_2+1) - 1$.*

(i) *For all but a measure at most $d/|S|$ of all $M_1 \in \mathrm{GL}_{n_1}(S)$ and a measure at most $d/|S|$ of all $M_2 \in \mathrm{GL}_{n_2}(S)$ we have that*

$$\dim(\mathcal{U} \mid \mathcal{W} \cup \mathcal{U}_B) = 0$$

for $\mathcal{U} = (M_1, M_2) \cdot \mathcal{V}$ and for $B = ([b_1] \times [n_2]) \cup ([n_1] \times [b_2])$.

(ii) *For all but a measure at most* $\binom{n_1}{b_1}d/|S|$ *of all* $M_1 \in \mathrm{GL}_{n_1}(S)$ *and a measure at most* $\binom{n_2}{b_2}d/|S|$ *of all* $M_2 \in \mathrm{GL}_{n_2}(S)$ *we have that*

$$\dim(\mathcal{U} \mid \mathcal{W} \cup \mathcal{U}_B) = 0$$

for $\mathcal{U} = (M_1, M_2) \cdot \mathcal{V}$ *and all* B *of the form* $B = (B_1 \times [n_2]) \cup ([n_1] \times B_2)$ *with* $|B_1| = b_1$ *and* $|B_2| = b_2$*.*

(iii) *Let* $\mathcal{W}$ *be an arbitrary family of elements in* $\mathbb{F}^k$*. Let* $\mathcal{V}^1, \dots, \mathcal{V}^m$ *be a collection of tensors of elements in* $\mathbb{F}^k$ *each indexed by* I*. Define the conditional dimensions* $d^{(j)} = \dim(\mathcal{V}^j \mid \mathcal{W} \cup \mathcal{V}^1 \cup \dots \cup \mathcal{V}^{j-1})$*. Let* $b_1^{(j)} \le n_1$ *and* $b_2^{(j)} \le n_2$ *be integers such that* $d^{(j)} \le (b_1^{(j)}+1)(b_2^{(j)}+1)-1$*. Except for a measure at most* $\sum_j \binom{n_1}{b_1^{(j)}} d^{(j)}/|S|$ *of all* $M_1 \in \mathrm{GL}_{n_1}(S)$ *and a measure at most* $\sum_j \binom{n_2}{b_2^{(j)}} d^{(j)}/|S|$ *of all* $M_2 \in \mathrm{GL}_{n_1}(S)$*, for every collection of sets* B^j *as above, we have that*

$$\dim(\mathcal{U}^1 \cup \dots \cup \mathcal{U}^m \mid \mathcal{W} \cup \mathcal{U}^1_{B^1} \cup \dots \cup \mathcal{U}^m_{B^m}) = 0,$$

where $\mathcal{U}^j = (M_1, M_2) \cdot \mathcal{V}^j$*.*

Proof. (i) The proof of this claim is the same as the proof of claim (i) of Theorem 11.8 except that all dimensions are taken conditional given $\mathcal{W}$.

(ii) This claim follows from claim (i) by a union bound over all possible choices of B.

(iii) The argument is the same as in the proof of Lemma 11.7 (iii). $\square$

11.4. Shifting of high-dimensional tensors of vectors

We now extend the shifting results that we proved in the previous subsections to arrays of any dimension $\ell \in \mathbb{N}$. Let $I = [n_1] \times \dots \times [n_\ell]$ and let $\mathcal{V} = (v_{i_1,\dots,i_\ell} : (i_1, \dots, i_\ell) \in I) \in (\mathbb{F}^k)^I$. We think of $\mathcal{V}$ as an ℓ-dimensional array of vectors. In this setting we are allowed to apply to $\mathcal{V}$ linear transformations from $\Gamma = \mathrm{GL}_{n_1}(\mathbb{F}) \times \dots \times \mathrm{GL}_{n_\ell}(\mathbb{F})$ as defined in Section 11.1. The goal is to shift $\mathcal{V}$ to a set B where B is a union of slices of I by acting with Γ.

Theorem 11.10. *Let* $S \subseteq \mathbb{F}$ *be finite. Let* $d = \dim(\mathcal{V})$*. Suppose that* $b_1 \le n_1, \dots, b_\ell \le n_\ell$ *are integers such that* $d \le (b_1 + 1) \cdots (b_\ell + 1) - 1$*.*

(i) *For all but a measure at most* $d/|S|$ *of all* $M_i \in \mathrm{GL}_{n_i}(S)$ *for each* $i = 1, \dots, \ell$ *we have that* $\mathcal{U} = (M_1, \dots, M_\ell) \cdot \mathcal{V}$ *is* B*-spanning for*

$$B = ([b_1] \times [n_2] \times \dots \times [n_\ell]) \cup ([n_1] \times [b_2] \times \dots \times [n_\ell]) \cup \dots \cup ([n_1] \times [n_2] \times \dots \times [b_\ell]).$$

(ii) *For all but a measure at most* $\binom{n_i}{b_i}d/|S|$ *of all* $M_i \in \mathrm{GL}_{n_1}(S)$ *for each* $i = 1, \dots, \ell$ *we have that* $\mathcal{U} = (M_1, \dots, M_\ell) \cdot \mathcal{V}$ *is* B*-spanning for every* B *of the form*

$$B = (B_1 \times [n_2] \times \dots \times [n_\ell]) \cup ([n_1] \times B_2 \times \dots \times [n_\ell]) \cup \dots \cup ([n_1] \times [n_2] \times \dots \times B_\ell)$$

with $|B_i| = b_i$*.*

Theorem 11.10 follows directly from the following more general theorem that applies to conditional dimensions.

Theorem 11.11. *Let $S \subseteq \mathbb{F}$ be finite.*

(i) *Let $\mathcal{W}$ be an arbitrary tensor of vectors in $\mathbb{F}^k$ and let $d = \dim(\mathcal{V}|\mathcal{W})$. Suppose that $b_1 \leq n_1, \dots, b_\ell \leq n_\ell$ are integers such that $d \leq (b_1+1)\cdots(b_\ell+1)-1$. For all but a measure at most $d/|S|$ of all $M_i \in \mathrm{GL}_{n_i}(S)$ for each $i = 1, \dots, \ell$ we have that $\dim(\mathcal{U} \mid \mathcal{W} \cup \mathcal{U}_B) = 0$ for $\mathcal{U} = (M_1, \dots, M_\ell) \cdot \mathcal{V}$ and for*

$$B = ([b_1] \times [n_2] \times \cdots \times [n_\ell]) \cup ([n_1] \times [b_2] \times \cdots \times [n_\ell]) \cup \cdots \cup ([n_1] \times [n_2] \times \cdots \times [b_\ell]).$$

(ii) *In the same setting as in (i) the following is true. For all but a measure at most $\binom{n_i}{b_i} d/|S|$ of all $M_i \in \mathrm{GL}_{n_1}(S)$ for each $i = 1, \dots, \ell$ we have that $\dim(\mathcal{U} \mid \mathcal{W} \cup \mathcal{U}_B) = 0$ for $\mathcal{U} = (M_1, \dots, M_\ell) \cdot \mathcal{V}$ and all B of the form*

$$B = (B_1 \times [n_2] \times \cdots \times [n_\ell]) \cup ([n_1] \times B_2 \times \cdots \times [n_\ell]) \cup \cdots \cup ([n_1] \times [n_2] \times \cdots \times B_\ell)$$

with $|B_i| = b_i$.

(iii) *Let $\mathcal{V}^1, \dots, \mathcal{V}^m$ be a collection of tensors of vectors in $\mathbb{F}^k$ each indexed by I. Define the conditional dimensions $d^{(j)} = \dim(\mathcal{V}^j \mid \mathcal{W} \cup \mathcal{V}^1 \cup \cdots \cup \mathcal{V}^{j-1})$ for $j = 1, \dots, m$. Let $b_i^{(j)} \leq n_i$ for $i = 1, \dots, \ell$ and $j = 1, \dots, m$ be integers such that $d^{(j)} \leq (b_1^{(j)}+1)\cdots(b_\ell^{(j)}+1)-1$. Except for a measure at most $\binom{n}{b_i^{(j)}} d^{(j)}/|S|$ of all $M_i \in \mathrm{GL}_{n_i}(S)$ for each $i = 1, \dots, \ell$ for*

$$B^j = ([b_1^{(j)}] \times [n_2] \times \cdots \times [n_\ell]) \cup ([n_1] \times [b_2^{(j)}] \times \cdots \times [n_\ell]) \cup \cdots \cup ([n_1] \times [n_2] \times \cdots \times [b_\ell^{(j)}])$$

we have that $\dim(\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^m \mid \mathcal{W} \cup \mathcal{U}^1_{B^1} \cup \cdots \cup \mathcal{U}^m_{B^m}) = 0$, where $\mathcal{U}^j = (M_1, \dots, M_\ell) \cdot \mathcal{V}^j$.

(iv) *Let $\mathcal{V}^1, \dots, \mathcal{V}^m$ be a collection of tensors of vectors in $\mathbb{F}^k$ each indexed by I. Define the conditional dimensions $d^{(j)} = \dim(\mathcal{V}^j \mid \mathcal{W} \cup \mathcal{V}^1 \cup \cdots \cup \mathcal{V}^{j-1})$ for $j = 1, \dots, m$. Let $b_i^{(j)} \leq n_i$ for $i = 1, \dots, \ell$ and $j = 1, \dots, m$ be integers such that $d^{(j)} \leq (b_1^{(j)}+1)\cdots(b_\ell^{(j)}+1)-1$. Except for a measure at most $\sum_j \binom{n}{b_i^{(j)}} d^{(j)}/|S|$ of all $M_i \in \mathrm{GL}_{n_i}(S)$ for each $i = 1, \dots, \ell$ for every collection of sets B^j as above, we have that $\dim(\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^m \mid \mathcal{W} \cup \mathcal{U}^1_{B^1} \cup \cdots \cup \mathcal{U}^m_{B^m}) = 0$, where $\mathcal{U}^j = (M_1, \dots, M_\ell) \cdot \mathcal{V}^j$.*

Proof. We prove the three claims by induction on ℓ.

(i) Let $\mathcal{U} = (M_1, \dots, M_\ell) \cdot \mathcal{V}$ for any $M_i \in \mathrm{GL}_{n_i}(S)$. Let $\mathcal{U}^1, \dots, \mathcal{U}^{n_\ell}$ be the slices of $\mathcal{U}$ along the last index. Let $d_i = \dim(\mathcal{U}^i \mid \mathcal{W} \cup \mathcal{U}^1 \cup \cdots \cup \mathcal{U}^{i-1})$ for $i = 1, \dots, n_\ell$. There is a choice of $M_\ell \in \mathrm{GL}_{n_\ell}(S)$ so that, among all choices of $M_\ell \in \mathrm{GL}_{n_\ell}(S)$, the value of d_1 is maximized, and conditionally the value of d_2 is maximized, etc. These values of $d_1 \geq d_2 \geq \cdots \geq d_{n_\ell}$ are attained for all $M_\ell \in \mathrm{GL}_{n_\ell}(S)$ except for a measure at most $\sum_{j=1}^{n_\ell} d_j / |S| = d/|S|$.

Fix any maximizing choice of M_ℓ. Suppose that $d_{b_\ell+1} \geq (b_1+1)\cdots(b_{\ell-1}+1)$. Then $d = \sum_{i=1}^{n_\ell} d_i \geq \sum_{i=1}^{b_\ell+1} d_i \geq (b_\ell+1)(b_{\ell-1}+1)\cdots(b_1+1)$. This contradicts our assumption about the dimension d. We conclude that $d_{n_\ell} \leq \cdots \leq d_{b_\ell+1} \leq (b_1+1)\cdots(b_{\ell-1}+1)-1$. By induction on ℓ we have that, except for a measure $\sum_{j=1}^{n_2} d_j/|S| = d/|S|$ of all matrices $M_i \in \mathrm{GL}_{n_1}(S)$, for $B^j = [b_1] \times \cdots \times [b_{\ell-1}] \times \{j\}$ for $j \geq b_\ell + 1$ that $\dim(\mathcal{U}^1 \cup \cdots \cup \mathcal{U}^j | \mathcal{W} \cup \mathcal{U}^1_{B^1} \cup \cdots \cup \mathcal{U}^j_{B^j}) = 0$ by Claim (iii). Note that $\cup_j B^j$ is contained in B.

(ii) This claim follows from claim (i) by a union bound over all possible choices for B.

(iii) The proof is the same as the proof of Lemma 11.7 (iii) and Theorem 11.9 (iii), using Claim (ii).

(iv) Follows from (iii) by taking a union bound over all possible choices of B^j. ☐

11.5. Discussion

We have reproved the shifting theorem of Strassen, we have extended it to the high-dimensional case ($\ell > 2$), and we have for finite fields provided precise bounds on the measure of good linear transformations. We end this section by restating the shifting theorem in terms of subspaces of tensors, and comparing the shifting theorem to other work on subspaces of matrices and tensors and in particular the notion of completely entangled subspaces in quantum information theory.

Shifting of subspaces of tensors. We will reformulate the shifting theorem in the language of subspaces of tensors. We must first point out the (straightforward) correspondence between tensors of vectors and subspaces of tensors. Let $\mathcal{V} = (v_i : i \in I) \in (\mathbb{F}^k)^I$ be a tensor of vectors indexed by $I = [n_1] \times \cdots \times [n_\ell]$. Thinking of the vectors v_i as the rows in an $I \times k$ matrix, we let $u_1, \ldots, u_k \in \mathbb{F}^I$ be the columns of this matrix. We think of the u_i as elements of the tensor space $\mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_\ell}$, and we let $V = \mathrm{span}\{u_1, \ldots, u_k\} \subseteq \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_\ell}$. Under this translation from tensors of vectors to tensor subspaces, the prototypical shifting theorem Theorem 11.10 (ii) says:

Theorem 11.12 (Theorem 11.10 (ii))**.** *Let $V \subseteq \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_\ell}$ be a d-dimensional subspace. Suppose that $0 \leq b_i \leq n_i$ for $i = 1, \ldots, \ell$ are integers such that $d \leq (b_1+1)\cdots(b_\ell+1) - 1$. Then there exist subspaces $W_i \subseteq \mathbb{F}^{n_i}$ of dimension $n_i - b_i$, such that $V \cap (W_1 \otimes \cdots \otimes W_\ell) = 0$.*

In this formulation it is particularly easy to see that Theorem 11.12 (and thus all foregoing formulations) is optimal, in the following sense. Let $0 \leq b_i < n_i$ for $i = 1, \ldots, \ell$ be integers. Let $V = V_1 \otimes \cdots \otimes V_\ell \subseteq \mathbb{F}^{n_1 \times \cdots \times n_\ell}$ be the subspace defined by letting $V_i \subseteq \mathbb{F}^{n_i}$ be any subspace of dimension $b_i + 1$ for $i = 1, \ldots, \ell$. Then $\dim(V) = (b_1+1)\cdots(b_\ell+1)$, and so the assumption of Theorem 11.12 is violated by one. Let $W_i \subseteq \mathbb{F}^{n_i}$ for $i = 1, \ldots, \ell$ be any subspace of dimension $n_i - b_i$. Then $V_i \cap W_i \neq 0$ since $\dim(V_i) + \dim(W_i) > n_i$. Therefore, $V \cap (W_1 \otimes \cdots \otimes W_\ell) \neq 0$. This means that the assumption $d \leq (b_1+1)\cdots(b_\ell+1) - 1$ in Theorem 11.12 cannot be relaxed.

Dimension bounds guaranteeing low rank. Our shifting theorems complement a long line of work on matrix and tensor subspaces. These results can be phrased as bounds on the dimension of a subspace that guarantee the existence of a nonzero element of low rank. In order to relate this to our results, we restate our Theorem 11.12 for the special case $b_i = n_i - 1$ for all $i \in [\ell]$:

Theorem 11.13 (Theorem 11.12, special case)**.** *Let $V \subseteq \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_\ell}$ be a d-dimensional subspace. If $d < n_1 \cdots n_\ell$, then there are subspaces $W_i \subseteq \mathbb{F}^{n_i}$ with $\dim W_i = 1$ such that $V \cap (W_1 \otimes \cdots \otimes W_\ell) = 0$. Moreover, this is clearly optimal: if $\dim V = n_1 \cdots n_\ell$, then $V = \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_\ell}$ and thus for all subspaces $W_i \subseteq \mathbb{F}^{n_i}$ with $\dim W_i = 1$ we have $V \cap (W_1 \otimes \cdots \otimes W_\ell) \neq 0$.*

We compare Theorem 11.13 to a result of Parthasarathy [Par04] and Wallach [Wal02, Section 4]. We define a *completely entangled subspace* $V \subseteq \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_\ell}$ to be any subspace V such that for every nonzero $v \in V$ there are no elements $w_i \in \mathbb{F}^{n_i}$ such that $v = w_1 \otimes \cdots \otimes w_\ell$.[92] In other words, a completely entangled subspace contains no elements of rank one. Equivalently, phrased contrapositively, V is *not* a completely entangled subspace if and only if there are subspaces $W_i \subseteq \mathbb{F}^{n_i}$ with $\dim W_i = 1$ such that $V \cap (W_1 \otimes \cdots \otimes W_\ell) \neq 0$. Parthasarathy and Wallach independently proved the following tight upper bound on the dimension of completely entangled subspaces over the complex numbers, which we phrase contrapositively:

[92] Parthasarathy and Wallach only work over the field $\mathbb{F} = \mathbb{C}$ of complex numbers, which is the relevant field for quantum information theory. However, the definition of a completely entangled subspace makes sense over any field.

Theorem 11.14 (Parthasarathy [Par04] and Wallach [Wal02]). *Let $V \subseteq \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_\ell}$ be a d-dimensional subspace. If $d > n_1 \cdots n_\ell - \sum_i (n_i - 1) - 1$, then there are subspaces $W_i \subseteq \mathbb{C}^{n_i}$ with $\dim W_i = 1$ such that $V \cap (W_1 \otimes \cdots \otimes W_\ell) \neq 0$. Moreover, this is optimal: there is a subspace V with $\dim V = n_1 \cdots n_\ell - \sum_i (n_i - 1) - 1$ such that for all subspaces $W_i \subseteq \mathbb{C}^{n_i}$ with $\dim W_i = 1$ it holds that $V \cap (W_1 \otimes \cdots \otimes W_\ell) = 0$.*

Comparing Theorem 11.13 to Theorem 11.14, note how the first characterizes when we can find one-dimensional subspaces W_i such that $V \cap (W_1 \otimes \cdots \otimes W_\ell) = 0$, while the second characterizes when we can find one-dimensional subspaces W_i such that $V \cap (W_1 \otimes \cdots \otimes W_\ell) \neq 0$. The assumption $d < n_1 \cdots n_\ell$ in Theorem 11.13 and the assumption $d > n_1 \cdots n_\ell - \sum_i (n_i - 1) - 1$ in Theorem 11.14 may overlap. In that case there are subspaces W_i with $\dim W_i = 1$ such that $V \cap (W_1 \otimes \cdots \otimes W_\ell) = 0$, while, there are (other) subspaces W_i' with $\dim W_i' = 1$ such that $V \cap (W_1' \otimes \cdots \otimes W_\ell') \neq 0$.

For $\ell = 2$ (subspaces of matrices), an extension of Theorem 11.14 was obtained by Cubitt, Montanaro and Winter [CMW08], who proved a tight upper bound on the dimension of any complex matrix subspace in which all nonzero elements have rank at least a given number r. (Theorem 11.14 covers the special case $r = 1$.) They show:

Theorem 11.15 (Cubitt, Montanaro and Winter [CMW08]). *Let $V \subseteq \mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2}$ be a d-dimensional subspace. If $d > (n_1 - r + 1)(n_2 - r + 1)$, then V contains a nonzero element of rank strictly less than r; in other words, there are subspaces $W_i \subseteq \mathbb{C}^{n_i}$ with $\dim W_i = r - 1$ such that $V \cap (W_1 \otimes W_2) \neq 0$. Moreover this is optimal: there is a subspace V with $\dim(V) = (n_1 - r + 1)(n_2 - r + 1)$ such that all nonzero elements in V have rank at least r; that is, for all subspaces $W_i \subseteq \mathbb{C}^{n_i}$ with $\dim W_i = r - 1$ it holds that $V \cap (W_1 \otimes W_2) = 0$.*

Dimension bounds guaranteeing high rank. Rather than guaranteeing the existence of a nonzero element of *low* rank, Dieudonné [Die49] and Flanders [Fla62] gave a dimension bound guaranteeing the existence of an element of *high* rank. Meshulam [Mes85] extended this to arbitrary fields:

Theorem 11.16 (Meshulam [Mes85], extending Flanders [Fla62]). *Let $V \subseteq \mathbb{F}^{n_1} \otimes \mathbb{F}^{n_2}$ be a d-dimensional subspace. If $d > r \max_i n_i$, then V contains an element of rank strictly larger than r. Moreover, this is optimal: there is a subspace V with $\dim(V) = r \max_i n_i$ such that all elements in V have rank at most r.*

The optimality of the dimension bound in Theorem 11.16 is directly verified by considering subspaces V of the form $\mathbb{F}^{n_1} \otimes \mathbb{F}^r$ and $\mathbb{F}^r \otimes \mathbb{F}^{n_2}$.

Briët [Bri21] extended Meshulam's theorem to subspaces of tensors $V \subseteq \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_\ell}$ (for prime fields $\mathbb{F} = \mathbb{F}_p$) by proving that if the dimension of V is large enough, then V contains a large subspace all of whose nonzero elements have large analytic rank.[93] Analytic rank, an extension of matrix rank to tensors (with coefficients in $\mathbb{F}_p$), was introduced by Gowers and Wolf [GW11], further developed by Lovett [Lov19] and extended to fields of all characteristics by Kopparty, Moshkovitz and Zuiddam [KMZ20], where it was used to prove exact optimality of Strassen's border subrank construction for matrix multiplication tensors [Str87].

[93] Therefore, by the result of Cohen and Moshkovitz [CM21], these tensors also have high partition rank, a notion introduced by Naslund [Nas20] that is closely related to the slice rank of Tao [Tao16]

Further directions. We have provided only a limited survey of all the results in this area. For a much broader overview, especially for matrix subspaces, we refer to [Mes18]. The last word on the types of dimension bounds that we have discussed here, and their applications, has clearly not been said. In particular, we are confident that the shifting theorem of Strassen and our extension will find applications. We will develop one such application in the next section.

12. Compression (and error correction) of tensor networks

The study of tensor networks is a vast and active area of research in both mathematics and physics. Thus, there is no way we can do justice to it, and we will content ourselves here with providing a self-contained description of the definitions we need to describe our results, referring to the literature (e.g., the survey [Orú14]) for a thorough introduction. On a high level, tensor networks are a way to compactly represent large tensors as being composed of smaller, simpler ones, which sit on the vertices of a graph. These efficient representations are desirable for algorithmic as well as analytical study of quantum many-body systems.

The matrix multiplication tensors are a special case of tensor networks. Namely, they have a natural representation as a tensor network with three vertices connected by three edges in a triangle, and we will be interested in a class of tensors that are a natural generalization of the matrix multiplication tensors from the perspective of tensor networks. We will apply the compression theorem for tensors of vectors of Section 11 to this class of tensors to obtain a *compression theorem of tensor networks*, which roughly says the following:

> For every tensor network T, if there is a restriction $T \leq S \oplus U$ for two tensors S and U[94] such that S has "low rank", then there is a tensor network T' with the same graph structure as T, and only slightly smaller "dimension", such that $T' \leq U$. (12.1)

Statement (12.1), or rather its natural extension to direct sums of tensor networks, leads to a proof that the flattening ranks (defined as the matrix ranks of flattenings of the tensor, Definition 12.6) are anchors (as defined in Section 8) in the asymptotic spectra of tensor networks. This will imply the main theorem of this section, which says that the asymptotic spectra of tensor networks are log-star-convex with respect to the flattening ranks.

This section is organized as follows. In Sections 12.1 and 12.2 we introduce our basic notation around tensor networks. In Sections 12.3 and 12.4 we prove the projection property and invariance property of tensor networks, which are crucial in the later proofs. In Section 12.5 we state precisely and prove the compression theorem of tensor networks described in (12.1). In Section 12.6 we consider the direct sum version of this and see how this gives that the flattening ranks are anchors in the sense of Section 8. Finally, in Section 12.7 we deduce using the anchors the main theorem for the asymptotic spectra of tensor networks, namely that they are log-star-convex with respect to the flattening ranks.

[94] Or equivalently, a restriction $T \leq S + U$ for tensors S, U that live in the same space as T.

12.1. Tensor networks

The kind of tensor network that we consider is specified by a graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ and a vector of integer weights $n \in \mathbb{N}^{\mathsf{E}}$ on the edges E.[95] In this way every graph will define an infinite family of tensors $T_{\mathsf{G},n}$ parametrized by the edge weights n.

Definition 12.1 (Tensor network)**.** Let $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ be a graph and let $n = (n_{\mathsf{e}})_{\mathsf{e}} \in \mathbb{N}^{\mathsf{E}}$ be a vector of integer weights on the edges E. For every vertex $\mathsf{v} \in \mathsf{V}$ we let $\mathsf{E}(\mathsf{v}) \subseteq \mathsf{E}$ denote the set of edges touching vertex v, that is, $\mathsf{E}(\mathsf{v}) \coloneqq \{\mathsf{e} \in \mathsf{E} : \mathsf{e} = \{\mathsf{v}, \mathsf{w}\} \text{ for some } \mathsf{w} \in \mathsf{V}\}$. We denote the standard basis elements of the vector space $\mathbb{F}^{n_{\mathsf{e}}}$ by f_j for $j \in [n_{\mathsf{e}}]$. For every vertex $\mathsf{v} \in \mathsf{V}$ we define the vector space $W_{\mathsf{v}} = \otimes_{\mathsf{e} \in \mathsf{E}(\mathsf{v})} \mathbb{F}^{n_{\mathsf{e}}}$. We define the index set $I = \prod_{\mathsf{e} \in \mathsf{E}} [n_{\mathsf{e}}]$. We define the *tensor network* on the graph G with weight vector n as

$$T_{\mathsf{G},n} \coloneqq \sum_{i \in I} \bigotimes_{\mathsf{v} \in \mathsf{V}} \big(\otimes_{\mathsf{e} \in \mathsf{E}(\mathsf{v})} f_{i_{\mathsf{e}}} \big) \;\in\; \bigotimes_{\mathsf{v} \in \mathsf{V}} W_{\mathsf{v}}.$$

The tensor $T_{\mathsf{G},n}$ defined in Definition 12.1 has both a "fine structure" as an element of the space $\bigotimes_{\mathsf{v} \in \mathsf{V}} \bigotimes_{\mathsf{e} \in \mathsf{E}(\mathsf{v})} \mathbb{F}^{n_{\mathsf{e}}}$ (i.e., as a $(\sum_{\mathsf{v} \in \mathsf{V}} |\mathsf{E}(\mathsf{v})|)$-tensor) and a "coarse structure" as an element of the space $\bigotimes W_{\mathsf{v}}$ (i.e., as a $|\mathsf{V}|$-tensor), depending on how we group the tensor legs. We will use both points of view.

As mentioned before, we think of every graph G as defining a family of tensor networks $T_{\mathsf{G},n}$ parametrized by edge weights n. And so when we prove an inequality of the type mentioned in (12.1), the tensor network T' is defined on the same graph as T, only with slightly smaller weights. We now give a few examples to illustrate this point, and the utility and simplicity of this description of tensors.

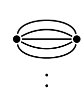
Example 12.2 (One edge)**.** The simplest tensor network, the tensor network on the graph with a single edge K_2

with weight n, is:

$$T_{\mathsf{K}_2,n} = \sum_{i \in [n]} f_i \otimes f_i \in \mathbb{F}^n \otimes \mathbb{F}^n.$$

As an $n \times n$ matrix, $T_{\mathsf{K}_2,n}$ is simply the identity matrix. Thought of as a linear map $\mathbb{F}^n \to \mathbb{F}^n$, $T_{\mathsf{K}_2,n}$ is the identity map. Thought of as a bilinear map $\mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$, $T_{\mathsf{K}_2,n}$ is the standard inner product.

Example 12.3 (Multiple edges between two nodes)**.** The tensor network on the graph with multiple edges between two nodes



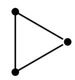
with weight $(n_{\mathsf{e}})_{\mathsf{e}}$, is:

$$T_{\mathsf{G},(n_{\mathsf{e}})_{\mathsf{e}}} = \sum_{i \in \prod_{\mathsf{e} \in \mathsf{E}} [n_{\mathsf{e}}]} \big(\otimes_{\mathsf{e} \in \mathsf{E}} f_{i_{\mathsf{e}}}\big) \otimes \big(\otimes_{\mathsf{e} \in \mathsf{E}} f_{i_{\mathsf{e}}}\big) \;\in\; \big(\otimes_{\mathsf{e} \in \mathsf{E}} \mathbb{F}^{n_{\mathsf{e}}}\big) \otimes \big(\otimes_{\mathsf{e} \in \mathsf{E}} \mathbb{F}^{n_{\mathsf{e}}}\big).$$

[95] To the reader who is familiar with tensor networks, we note that we will only consider tensor networks in which the tensors to be contracted are symbolic tensors, that is, tensors of variables. In that way, for us, every tensor network on k nodes represents a k-tensor. In the language of quantum information theory, our kind of tensor networks represent quantum systems in which k parties are sharing maximally entangled bipartite quantum states (EPR pairs) according to a certain graph structure. They have sometimes been called *graph tensors* [CVZ19a].

We see that this tensor network is essentially equivalent to the tensor network on one edge with weight $\prod_{\mathsf{e}\in\mathsf{E}} n_{\mathsf{e}}$.

Example 12.4 (Matrix multiplication)**.** We have on the cycle graph C_3



with weight (n_1, n_2, n_3) the tensor network

$$T_{\mathsf{C}_3,(n_1,n_2,n_3)} = \sum_{i=(i_1,i_2,i_3)\in\prod_{\mathsf{e}}[n_{\mathsf{e}}]} (f_{i_1}\otimes f_{i_2})\otimes(f_{i_2}\otimes f_{i_3})\otimes(f_{i_3}\otimes f_{i_1}) \in (\mathbb{F}^{n_1}\otimes\mathbb{F}^{n_2})\otimes(\mathbb{F}^{n_2}\otimes\mathbb{F}^{n_3})\otimes(\mathbb{F}^{n_3}\otimes\mathbb{F}^{n_1}).$$

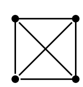
This tensor network equals the matrix multiplication tensor. Thought of as a bilinear map $(\mathbb{F}^{n_1}\otimes\mathbb{F}^{n_3})\times(\mathbb{F}^{n_2}\otimes\mathbb{F}^{n_3})\to\mathbb{F}^{n_1}\otimes\mathbb{F}^{n_3}$, it is the map that multiplies an $n_1\times n_2$ matrix A with an $n_2\times n_3$ matrix B resulting in an $n_1\times n_3$ matrix C.[96] Thought of as a trilinear map $(\mathbb{F}^{n_1}\otimes\mathbb{F}^{n_3})\times(\mathbb{F}^{n_2}\otimes\mathbb{F}^{n_3})\otimes\mathbb{F}^{n_3}\otimes\mathbb{F}^{n_1}\to\mathbb{F}$, the tensor network is the trace of the product of three matrices: $\mathrm{tr}(ABC)=\sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\sum_{i_3=1}^{n_3} A_{i_1,i_2}B_{i_2,i_3}C_{i_3,i_1}$.

Note — we will use this observation later — how any tensor network on $\mathsf{G}=(\mathsf{V},\mathsf{E})$ can be built as the tensor product of copies of the simplest tensor network $T_{\mathsf{K}_2,n_{\mathsf{e}}}$ with one copy for every edge $\mathsf{e}\in\mathsf{E}$. For example, for the tensor network on C_3 we have, under the natural reordering, the equality

$$\sum_{i\in\prod_{\mathsf{e}}[n_{\mathsf{e}}]}(f_{i_1}\otimes f_{i_2})\otimes(f_{i_2}\otimes f_{i_3})\otimes(f_{i_3}\otimes f_{i_1}) = \Big(\sum_{i\in[n_1]} f_{i_1}\otimes 1\otimes f_{i_1}\Big)\otimes\Big(\sum_{i\in[n_2]} f_{i_2}\otimes f_{i_2}\otimes 1\Big)\otimes\Big(\sum_{i\in[n_3]} 1\otimes f_{i_3}\otimes f_{i_3}\Big).$$

This simple observation will allow us to derive several properties of $T_{\mathsf{G},n}$ from $T_{\mathsf{K}_2,n}$ later.

Example 12.5 (Higher-order networks)**.** We have on the complete graph K_4



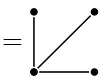
with weight (n,n,n,n) the tensor network:

$$T_{\mathsf{K}_4,(n,\ldots,n)} = \sum_{i\in[n]^6}(f_{i_1}\otimes f_{i_2}\otimes f_{i_3})\otimes(f_{i_1}\otimes f_{i_2}\otimes f_{i_4})\otimes(f_{i_2}\otimes f_{i_4}\otimes f_{i_5})\otimes(f_{i_3}\otimes f_{i_5}\otimes f_{i_6}) \in (\mathbb{F}^n\otimes\mathbb{F}^n\otimes\mathbb{F}^n)^{\otimes 4}.$$

Again we can think of this tensor in several ways. One is as a trilinear map $(\mathbb{F}^n\otimes\mathbb{F}^n\otimes\mathbb{F}^n)^{\times 3}\to\mathbb{F}^n\otimes\mathbb{F}^n\otimes\mathbb{F}^n$ that generalizes matrix multiplication to a product of three 3-tensors. Another is as a multilinear map $(\mathbb{F}^n\otimes\mathbb{F}^n\otimes\mathbb{F}^n)^{\times 4}\to\mathbb{F}$ given by $\sum A_{i_1,i_2,i_3}B_{i_1,i_2,i_4}C_{i_2,i_4,i_5}D_{i_3,i_5,i_6}$ where we sum over $i\in[n]^6$, which generalizes the trace of the product of three matrices.

[96] up to a computationally irrelevant matrix transpose

Finally, one may of course take other graphs, and in particular graphs for which the vertices have different degrees. For example, we have on the graph

$\mathsf{G} =$ 

with weight (n, n, n) the tensor network:

$$T_{\mathsf{G},(n,n,n)} = \sum_{i \in [n]^3} (f_{i_1} \otimes f_{i_2} \otimes f_{i_3}) \otimes f_{i_1} \otimes f_{i_2} \otimes f_{i_3} \in (\mathbb{F}^n \otimes \mathbb{F}^n \otimes \mathbb{F}^n) \otimes \mathbb{F}^n \otimes \mathbb{F}^n \otimes \mathbb{F}^n.$$

12.2. Tensors of vectors associated to a tensor network

Recall that for every graph $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ and edge weights $n \in \mathbb{N}^{\mathsf{E}}$, the tensor network $T_{\mathsf{G},n}$ lives in the vector space $\bigotimes_{\mathsf{v}\in\mathsf{V}} W_{\mathsf{v}}$ where $W_{\mathsf{v}} = \otimes_{\mathsf{e}\in\mathsf{E}(\mathsf{v})} \mathbb{F}^{n_{\mathsf{e}}}$. In order to carry over the compression theorem of the previous section to tensors $T \in \bigotimes_{\mathsf{v}\in\mathsf{V}} W_{\mathsf{v}}$, we will associate to T a tensor of vectors and apply the compression theorem to that tensor of vectors.

Our definition of the tensor of vectors associated to a tensor is very straightforward and is based on the basic notion of a *flattening*. First we give the definition of this general notion of a flattening of a tensor $T \in V_1 \otimes \cdots \otimes V_k$ and the closely related notion of flattening ranks.

Definition 12.6 (Flattening and flattening ranks)**.** Let $T \in V_1 \otimes \cdots \otimes V_k$ be a tensor. For any subset $S \subseteq [n]$ that is not empty and not $[n]$, we may group together the spaces V_i to obtain the space $(\bigotimes_{j\in S} V_j) \otimes (\bigotimes_{j\notin S} V_j)$ of tensors of order two. Under this grouping operation

$$V_1 \otimes \cdots \otimes V_k \to \big(\bigotimes_{j\in S} V_j\big) \otimes \big(\bigotimes_{j\notin S} V_j\big),$$

the tensor T becomes a tensor of order two, which we call the $\bigotimes_{j\in S} V_j$ *versus* $\bigotimes_{j\notin S} V_j$ *flattening* or *flattening with respect to* S or S*-flattening*. An important and useful parameter of a tensor in this context is defined by identifying the flattened tensor with a matrix and taking its matrix rank. This parameter is called the *flattening rank of* T *with respect to the* $\bigotimes_{j\in S} V_j$ *versus* $\bigotimes_{j\notin S} V_j$ *flattening* or S*-flattening rank*.

Lemma 12.7. *The flattening ranks are in the asymptotic spectrum of tensors.*

Proof. This follows immediately from the fact that matrix rank is in the asymptotic spectrum of matrices and the fact that the grouping operation commutes with direct sum and tensor product. ☐

Definition 12.8 (Tensor of vectors relative to w)**.** Let $\mathsf{w} \in \mathsf{V}$ be a distinguished vertex. For any tensor $T \in \bigotimes_{\mathsf{v}\in\mathsf{V}} W_{\mathsf{v}}$, not necessarily a tensor network, we group the spaces W_{v} together to obtain the space $W_{\mathsf{w}} \otimes (\bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} W_{\mathsf{v}})$ of tensors of order two, and accordingly write T in its flattened form

$$T = \sum_{i\in I} \big(\otimes_{\mathsf{e}\in\mathsf{E}(\mathsf{w})} f_{i_{\mathsf{e}}}\big) \otimes T_i \;\in\; W_{\mathsf{w}} \otimes \Big(\bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} W_{\mathsf{v}}\Big)$$

for tensors $T_i \in \bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} W_{\mathsf{v}}$ where the sum goes over $i \in I = \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})} [n_{\mathsf{e}}]$. We define the tensor of vectors $\mathcal{V}_{\mathsf{w}}(T)$ indexed by I as

$$\mathcal{V}_{\mathsf{w}}(T) \coloneqq (T_i : i \in I).$$

We call $\mathcal{V}_{\mathsf{w}}(T)$ the *tensor of vectors associated to* T *relative to* w.

Example 12.9. The simplest tensor network, on the graph K_2, has the associated tensor of vectors

$$\mathcal{V}_1(T_{\mathsf{K}_2,n}) = (f_i : i \in [n]).$$

The tensor network on the cycle graph C_3 has the associated tensor of vectors

$$\mathcal{V}_1(T_{\mathsf{C}_3,(n_1,n_2,n_3)}) = \Big(\sum_{i_3\in[n_3]}(f_{i_2}\otimes f_{i_3})\otimes(f_{i_3}\otimes f_{i_1}) : (i_1,i_2)\in[n_1]\times[n_2]\Big).$$

Every tensor T is uniquely determined by the associated tensor of vectors $\mathcal{V}_{\mathsf{w}}(T)$. We finish by pointing out how the basic notions of flattening rank and restriction of tensors carry over to the associated tensor of vectors in the following important remark.

Remark 12.10 (Flattening rank and restriction for tensors of vectors)**.** Recall that we have defined the dimension of a tensor of vectors as the dimension of its span, and so for every tensor $T \in \bigotimes_{\mathsf{v}\in\mathsf{V}} W_{\mathsf{v}}$ we have $\dim\mathcal{V}_{\mathsf{w}}(T) = \dim\operatorname{span}\mathcal{V}_{\mathsf{w}}(T)$. Since the rank of a matrix is equal to the dimension of its image, this dimension $\dim\mathcal{V}_{\mathsf{w}}(T)$ is equal to the flattening rank of T with respect to the W_{w} versus $\bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} W_{\mathsf{v}}$ flattening, that is, the w-flattening rank (Definition 12.6). Next we discuss restriction. We first recall the general definition. Let V_j, W_j be vector spaces and let $T \in V_1\otimes\cdots\otimes V_k$ and $S\in W_1\otimes\cdots\otimes W_k$ be tensors. We say that T *restricts to* S and write $T \geq S$ if there are linear maps $L_j : V_j \to W_j$ such that $(L_1\otimes\cdots\otimes L_k)T = S$. Then, for $T, S \in \bigotimes_{\mathsf{v}\in\mathsf{V}} W_{\mathsf{v}}$ we have $T\geq S$ if and only if there are linear maps $L_{\mathsf{v}} : W_{\mathsf{v}} \to W_{\mathsf{v}}$ such that $\mathcal{V}_{\mathsf{w}}(S)$ is in the linear span of $((\bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} L_{\mathsf{v}})T_i : i\in I) = \mathcal{V}_{\mathsf{w}}((\mathrm{id}_{\mathsf{w}}\otimes\bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} L_{\mathsf{v}})T)$.

12.3. Projection property of tensor networks

Tensor networks have two important properties (for our purposes), namely the *projection property* (Lemma 12.13) and the *invariance property* (Lemma 12.16). In this section we discuss the projection property.

The idea of the projection property is that for any tensor network $T_{\mathsf{G},n}$ we can reduce the weights n_{e} by applying a projection P to the tensor network that has a very simple form. Namely, this projection P can be chosen in such a way that it has a product form $P = \bigotimes_{\mathsf{v}\in\mathsf{V}} P_{\mathsf{v}}$ and given any distinguished vertex w we can choose P so that P_{w} is the identity map.

We will see that the projection property is trivial for a tensor network on one edge, and from this the projection property for an arbitrary tensor network will follow almost directly.

Lemma 12.11 (Projection property for one edge)**.** *Let $m, n \in \mathbb{N}$ such that $m \leq n$. Let $P : \mathbb{F}^n \to \mathbb{F}^m$ be the restriction to the first m coordinates, that is, P is the linear map defined by $Pf_i = f_i$ for $i \leq m$ and $Pf_i = 0$ for $i > m$, where $f_1, \ldots, f_n$ is the standard basis of $\mathbb{F}^n$. Then*

$$(\mathrm{id}\otimes P)\Big(\sum_{i\in[n]} f_i\otimes f_i\Big) = (P\otimes\mathrm{id})\Big(\sum_{i\in[n]} f_i\otimes f_i\Big) = (P\otimes P)\Big(\sum_{i\in[n]} f_i\otimes f_i\Big) = \sum_{i\in[m]} f_i\otimes f_i.$$

That is,

$$(\mathrm{id}\otimes P)\,T_{\mathsf{K}_2,n} = (P\otimes\mathrm{id})\,T_{\mathsf{K}_2,n} = (P\otimes P)\,T_{\mathsf{K}_2,n} = T_{\mathsf{K}_2,m}.$$

Proof. This is immediate. □

Note how in Lemma 12.11 the projection P may be applied either to the first tensor leg, the second tensor leg, or both tensor legs in order to obtain $T_{\mathsf{K}_2,m}$ from $T_{\mathsf{K}_2,n}$. We will use this freedom when we apply Lemma 12.11 to multiple edges in a tensor network to make sure that to one specific tensor leg (corresponding to a distinguished vertex w in the underlying graph) only identity maps are applied.

We will now see how from the projection property for one edge the projection property for an arbitrary tensor network follows, by applying the former to every edge of the graph. Recall that any tensor network on $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ can be built as the tensor product of copies of $T_{\mathsf{K}_2,n_\mathsf{e}}$ with one copy for every edge $\mathsf{e} \in \mathsf{E}$.

Example 12.12. For example, the tensor network $T_{\mathsf{C}_3,(n_1,n_2,n_3)}$ on the cycle graph C_3 is the tensor product of three "copies" of $T_{\mathsf{K}_2,n_\mathsf{e}}$ where each copy is extended to a 3-tensor by tensoring with 1, in the following way

$$\begin{aligned}&\sum_{i\in\prod_\mathsf{e}[n_\mathsf{e}]} (f_{i_1}\otimes f_{i_2})\otimes(f_{i_2}\otimes f_{i_3})\otimes(f_{i_3}\otimes f_{i_1})\\ &\qquad = \Big(\sum_{i\in[n_1]} f_{i_1}\otimes 1\otimes f_{i_1}\Big)\otimes\Big(\sum_{i\in[n_2]} f_{i_2}\otimes f_{i_2}\otimes 1\Big)\otimes\Big(\sum_{i\in[n_3]} 1\otimes f_{i_3}\otimes f_{i_3}\Big).\end{aligned}$$

We now want to apply projections to reduce the dimensions n_e. For any $m_\mathsf{e} \le n_\mathsf{e}$ let $P_\mathsf{e} : \mathbb{F}^{n_\mathsf{e}} \to \mathbb{F}^{m_\mathsf{e}}$ be the linear map that restricts every vector to the first m_e coordinates. From Lemma 12.11 we know how to use the projections P_e to reduce dimensions:

$$\begin{aligned}(\mathrm{id}\otimes\mathrm{id}\otimes P_1)\Big(\sum_{i\in[n_1]} f_{i_1}\otimes 1\otimes f_{i_1}\Big) &= \sum_{i\in[m_1]} f_{i_1}\otimes 1\otimes f_{i_1}\\ (\mathrm{id}\otimes P_2\otimes\mathrm{id})\Big(\sum_{i\in[n_2]} f_{i_2}\otimes f_{i_2}\otimes 1\Big) &= \sum_{i\in[m_2]} f_{i_2}\otimes f_{i_2}\otimes 1\\ (\mathrm{id}\otimes P_3\otimes P_3)\Big(\sum_{i\in[n_3]} 1\otimes f_{i_3}\otimes f_{i_3}\Big) &= \sum_{i\in[m_3]} 1\otimes f_{i_3}\otimes f_{i_3}.\end{aligned}$$

Recall that Lemma 12.11 left us some choice of where to apply the projections P_i and where the identities, and note how we used this freedom so that to the first tensor leg we only apply identities. Tensoring these projections together we find that

$$((\mathrm{id}\otimes\mathrm{id})\otimes(P_2\otimes P_3)\otimes(P_3\otimes P_1))\,T_{\mathsf{C}_3,n} = T_{\mathsf{C}_3,m}.$$

Indeed to the first tensor leg we only apply the identity map.

The construction of the projection in Example 12.12 immediately generalizes to any tensor network:

Lemma 12.13 (Projection property)**.** *Let $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ be a graph and let $\mathsf{w} \in \mathsf{V}$ be a distinguished vertex. Let $n = (n_\mathsf{e})_\mathsf{e}, m = (m_\mathsf{e})_\mathsf{e} \in \mathbb{N}^\mathsf{E}$ be two weightings of the edges such that $m_\mathsf{e} \le n_\mathsf{e}$ for every $\mathsf{e} \in \mathsf{E}$. For every $\mathsf{e} \in \mathsf{E}$ let $P_\mathsf{e} : \mathbb{F}^{n_\mathsf{e}} \to \mathbb{F}^{m_\mathsf{e}}$ be the restriction to the first m_e coordinates. For every $\mathsf{v} \in \mathsf{V}$ let $P_\mathsf{v} : W_\mathsf{v} \to W_\mathsf{v}$ be defined as the product $P_\mathsf{v} = \otimes_{\mathsf{e}\in\mathsf{E}(\mathsf{v})} P_\mathsf{e}$. Let id_w denote the identity map on W_w. Then*

$$(\mathrm{id}_\mathsf{w}\otimes\bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} P_\mathsf{v})\,T_{\mathsf{G},n} = T_{\mathsf{G},m}.$$

Proof. The claim follows from observing that the tensor network $T_{\mathsf{G},n}$ is the tensor product of tensor networks of the form

$$\sum_{i\in[n_{\mathsf{e}}]} f_i \otimes f_i \otimes 1 \otimes \cdots \otimes 1 \in \mathbb{F}^{n_{\mathsf{e}}} \otimes \mathbb{F}^{n_{\mathsf{e}}} \otimes \mathbb{F} \otimes \cdots \otimes \mathbb{F},$$

one for every edge $\mathsf{e} \in \mathsf{E}$, and applying Lemma 12.11 to each edge in a way that to the tensor leg corresponding to w we only apply the identity map. ☐

12.4. Invariance property of tensor networks

The second important property is that tensor networks are invariant under a natural group action. This action is changing basis on the two sides of the same edge. Again, this property is trivial for a tensor network on a single edges, and from this the property for an arbitrary tensor network follows almost immediately. For $h \in \mathrm{GL}_n$ let h^T denote the transpose.

Lemma 12.14 (Invariance property for one edge)**.** *For any $g \in \mathrm{GL}_n$ we have in $\mathbb{F}^n \otimes \mathbb{F}^n$ the equality*

$$(g \otimes (g^{-1})^\mathsf{T}) \sum_{i\in[n]} f_i \otimes f_i = \sum_{i\in[n]} f_i \otimes f_i.$$

Proof. More generally, let $g, h \in \mathrm{GL}_n$ and let g_i, h_i denote their column vectors. Then

$$\sum_{i\in[n]} g f_i \otimes h f_i = \sum_{i\in[n]} g_i \otimes h_i = \sum_{j,k\in[n]} (gh^\mathsf{T})_{j,k}\, f_j \otimes f_k.$$

Setting $h = (g^{-1})^\mathsf{T}$ we have $(gh^\mathsf{T})_{j,k} = \delta_{i=j}$, which proves the claim. ☐

The invariance property for one edge directly extends to tensor networks $T_{\mathsf{G},n}$ for arbitrary graphs G, by applying the invariance property for one edge to all (or some) of the edges of G.

Example 12.15. For example, for the tensor network on the cycle graph C_3 we have the invariance

$$(g_1 \otimes g_2) \otimes ((g_2^{-1})^\mathsf{T} \otimes \mathrm{id}) \otimes (\mathrm{id} \otimes (g_1^{-1})^\mathsf{T})\, T_{\mathsf{C}_3,n} = T_{\mathsf{C}_3,n}$$

for $g_1 \in \mathrm{GL}_{n_1}$ and $g_2 \in \mathrm{GL}_{n_2}$, since

$$\begin{aligned}
&\sum_{i\in\prod_{\mathsf{e}}[n_{\mathsf{e}}]} (g_1 f_{i_1} \otimes g_2 f_{i_2}) \otimes ((g_2^{-1})^\mathsf{T} f_{i_2} \otimes f_{i_3}) \otimes (f_{i_3} \otimes (g_1^{-1})^\mathsf{T} f_{i_1}) \\
&= \Big(\sum_{i\in[n_1]} g_1 f_{i_1} \otimes 1 \otimes (g_1^{-1})^\mathsf{T} f_{i_1}\Big) \otimes \Big(\sum_{i\in[n_2]} g_2 f_{i_2} \otimes (g_2^{-1})^\mathsf{T} f_{i_2} \otimes 1\Big) \otimes \Big(\sum_{i\in[n_3]} 1 \otimes f_{i_3} \otimes f_{i_3}\Big)
\end{aligned}$$

and we apply Lemma 12.14 to every copy of $T_{\mathsf{K}_2,n_{\mathsf{e}}}$. More generally, following the same principle, we have the invariance property that for every $g_1 \in \mathrm{GL}_{n_1}$, $g_2 \in \mathrm{GL}_{n_2}$ and $g_3 \in \mathrm{GL}_{n_3}$ it holds that $(g_1 \otimes g_2) \otimes ((g_2^{-1})^\mathsf{T} \otimes g_3) \otimes ((g_3^{-1})^\mathsf{T} \otimes (g_1^{-1})^\mathsf{T}) T_{\mathsf{C}_3,n} = T_{\mathsf{C}_3,n}$. This action of $\mathrm{GL}_{n_1} \times \mathrm{GL}_{n_2} \times \mathrm{GL}_{n_3}$ on $T_{\mathsf{C}_3,n}$ is sometimes called the *sandwiching action*, and the invariance precisely corresponds to the fact that the trace of the product of three matrices satisfies $\mathrm{tr}(ABC) = \mathrm{tr}((g_1 A g_2)(g_2^{-1} B g_3)(g_3^{-1} C g_1^{-1}))$ (under a standard translation between tensors and multilinear maps).

Following the above idea, we set up the action on arbitrary tensors in $\bigotimes_{\mathsf{v}\in\mathsf{V}} W_\mathsf{v}$. Define the group $\Gamma = \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})} \mathrm{GL}_{n_\mathsf{e}}$. Let $T \in \bigotimes_{\mathsf{v}\in\mathsf{V}} W_\mathsf{v}$ and $g = (g_\mathsf{e})_{\mathsf{e}\in\mathsf{E}(\mathsf{w})} \in \Gamma$. We let g act on T by letting, for every edge $\mathsf{e} = \{\mathsf{w},\mathsf{v}\}$, g_e act on W_w as g_e and on W_v as $(g_\mathsf{e}^{-1})^\mathsf{T}$. That is, $g \cdot T = \big(\bigotimes_{\mathsf{v}\in\mathsf{V}} \tau_\mathsf{v}(g)\big)T$ where, for every vertex $\mathsf{v}\in\mathsf{V}$, we set

$$\tau_\mathsf{v}(g) \coloneqq \begin{cases} g & \text{if } \mathsf{v} = \mathsf{w} \\ (g_\mathsf{e}^{-1})^\mathsf{T} & \text{if } \mathsf{v} \neq \mathsf{w} \text{ and } \mathsf{e} = \{\mathsf{w},\mathsf{v}\} \in \mathsf{E} \\ \mathrm{id} & \text{if } \mathsf{v} \neq \mathsf{w} \text{ and } \{\mathsf{w},\mathsf{v}\} \notin \mathsf{E}. \end{cases}$$

Lemma 12.16 (Invariance property)**.** *Under the above action, for every $g \in \Gamma$, we have*

$$g \cdot T_{\mathsf{G},n} = T_{\mathsf{G},n}.$$

Proof. The claim follows from the fact that the tensor network $T_{\mathsf{G},n}$ is the tensor product of tensor networks of the form

$$\sum_{i\in[n_\mathsf{e}]} f_i \otimes f_i \otimes 1 \otimes \cdots \otimes 1 \in \mathbb{F}^{n_\mathsf{e}} \otimes \mathbb{F}^{n_\mathsf{e}} \otimes \mathbb{F} \otimes \cdots \otimes \mathbb{F},$$

one for every edge $\mathsf{e}\in\mathsf{E}$, and applying Lemma 12.14 to each. □

12.5. Compression theorem for tensor networks

We will now state and prove the general compression theorem for tensor networks (and returning to the tensor notation). Recall that for a tensor $T \in \bigotimes_{\mathsf{v}\in\mathsf{V}} W_\mathsf{v}$ the dimension $\dim(\mathcal{V}_\mathsf{w}(T))$ is precisely equal to the w-flattening rank of the tensor T (Remark 12.10). Via this connection we will prove that the flattening ranks are anchors (Theorems 12.21 and 12.22).

Theorem 12.17 (Compression theorem for tensor networks)**.** *Let $\mathbb{F}$ be sufficiently large. Given a graph $\mathsf{G} = (\mathsf{V},\mathsf{E})$, if $T_{\mathsf{G},n} \leq S \oplus U$ for two tensors S and U, and $\dim(\mathcal{V}_\mathsf{w}(S)) < \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})}(n_\mathsf{e} - m_\mathsf{e} + 1)$ for a weighting $m \in \mathbb{N}^\mathsf{E}$, then $T_{\mathsf{G},m} \leq U$.*

We prove Theorem 12.17 using one of the compression theorems for tensors of vectors of Section 11, which we now recall:

Theorem 12.18 (Theorem 11.10 (ii))**.** *Let $d = \dim(\mathcal{V})$. Suppose that $b_1 \leq n_1, \ldots, b_\ell \leq n_\ell$ are integers such that $d \leq (b_1+1)\cdots(b_\ell+1) - 1$. For all but a measure at most $\binom{n_i}{b_i}d/|\mathbb{F}|$ of all $M_i \in \mathrm{GL}_{n_1}(\mathbb{F})$ for each $i = 1,\ldots,\ell$ we have that $\mathcal{U} = (M_1,\ldots,M_\ell)\cdot\mathcal{V}$ is B-spanning for every B of the form*

$$B = (B_1 \times [n_2] \times \cdots \times [n_\ell]) \cup ([n_1] \times B_2 \times \cdots \times [n_\ell]) \cup \cdots \cup ([n_1] \times [n_2] \times \cdots \times B_\ell)$$

with $|B_i| = b_i$.

As an intermediate step before proving Theorem 12.17 we apply Theorem 11.10 (i) to the tensor of vectors $\mathcal{V} = \mathcal{V}_\mathsf{w}(S)$ associated to a tensor $S \in \bigotimes_{\mathsf{v}\in\mathsf{V}} W_\mathsf{v}$. As before we let $\mathsf{G} = (\mathsf{V},\mathsf{E})$ be a graph, we let $n \in \mathbb{N}^\mathsf{E}$ be weights, we define the space $W_\mathsf{v} = \otimes_{\mathsf{e}\in\mathsf{E}(\mathsf{v})}\mathbb{F}^{n_\mathsf{e}}$, we let $\mathsf{w}\in\mathsf{V}$ be a fixed distinguished vertex, and we define the group $\Gamma = \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})} \mathrm{GL}_{n_\mathsf{e}}$ and the index set $I = \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})}[n_\mathsf{e}]$.

Corollary 12.19. *Let $\mathbb{F}$ be sufficiently large. Let $S \in \bigotimes_{\mathsf{v}\in\mathsf{V}} W_{\mathsf{v}}$. If $m_{\mathsf{e}} \leq n_{\mathsf{e}}$ for all $\mathsf{e} \in \mathsf{E}$ such that*

$$\dim(\mathcal{V}_{\mathsf{w}}(S)) < \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})} (n_{\mathsf{e}} - m_{\mathsf{e}} + 1),$$

then there is an element $g \in \Gamma$ such that $\mathcal{V}_{\mathsf{w}}(g \cdot S)$ is B-spanning for $B = I \setminus \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})}[m_{\mathsf{e}}]$.

Proof. We apply Theorem 12.18 to the tensor of vectors $\mathcal{V}_{\mathsf{w}}(S)$, with $b_{\mathsf{e}} = n_{\mathsf{e}} - m_{\mathsf{e}}$ and $B_{\mathsf{e}} = [n_{\mathsf{e}}] \setminus [m_{\mathsf{e}}]$. Note that the action of Γ on W_{w} is precisely the action as used in the compression theorem, while the action of Γ on $\bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} W_{\mathsf{v}}$ does not affect the property of being B-spanning. □

Proof of Theorem 12.17. Without loss of generality we may assume that $T_{\mathsf{G},n} = S + U$ where S and U are tensors that live in the same space as $T_{\mathsf{G},n}$ and where $+$ denotes coordinate-wise addition. (Namely, suppose $T_{\mathsf{G},n} \leq S \oplus U$, then there are linear maps L_{v} such that $T_{\mathsf{G},n} = \big(\bigotimes_{\mathsf{v}} L_{\mathsf{v}}\big)(S \oplus U)$. Then we simply replace S by $\big(\bigotimes_{\mathsf{v}} L_{\mathsf{v}}\big)S$ and U by $\big(\bigotimes_{\mathsf{v}} L_{\mathsf{v}}\big)U$.) We will show that after suitable linear transformations applied to both sides, the contribution of S to the (smaller) $T_{\mathsf{G},m}$ is zero, proving the theorem.

Recall that Γ acts invariantly on $T_{\mathsf{G},n}$ (invariance lemma, Lemma 12.16). Thus, after acting with Γ on both sides of the equation $T_{\mathsf{G},n} = S + U$, we may assume that $\mathcal{V}_{\mathsf{w}}(S)$ is B-spanning for $B = I \setminus \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})}[m_{\mathsf{e}}]$ (compression of tensor of vectors, Corollary 12.19). Recall (restriction lemma, Lemma 12.13) that for the linear projection $P = (\mathrm{id}_{\mathsf{w}} \otimes \bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} P_{\mathsf{v}})$ we have

$$PT_{\mathsf{G},n} = T_{\mathsf{G},m}.$$

Thus

$$T_{\mathsf{G},m} = PT_{\mathsf{G},n} = PS + PU.$$

We then have

$$\begin{aligned}[\mathcal{V}_{\mathsf{w}}(PS) + \mathcal{V}_{\mathsf{w}}(PU)]_{\overline{B}} &= \mathcal{V}_{\mathsf{w}}(T_{\mathsf{G},m})\\ [\mathcal{V}_{\mathsf{w}}(PS) + \mathcal{V}_{\mathsf{w}}(PU)]_{B} &= 0.\end{aligned}$$

Since $\mathcal{V}_{\mathsf{w}}(S)$ is B-spanning, and P acts only on $\bigotimes_{\mathsf{v}\in\mathsf{V}\setminus\mathsf{w}} W_{\mathsf{v}}$, the family $\mathcal{V}_{\mathsf{w}}(PS)$ is B-spanning. This means that $\mathcal{V}_{\mathsf{w}}(PS)_{\overline{B}}$ is in the span of $\mathcal{V}_{\mathsf{w}}(PS)_{B}$. This implies that there is a way of adding linear combinations of the vectors in $\mathcal{V}_{\mathsf{w}}(PS)_{B}$ to the vectors in $\mathcal{V}_{\mathsf{w}}(PS)_{\overline{B}}$ to make the latter all zero. That is, there exists a matrix of the form

$$Q = \begin{array}{c} \\ \overline{B} \end{array}\!\!\begin{array}{c} \begin{array}{cc} \overline{B} & B \end{array} \\ \begin{pmatrix} I & * \end{pmatrix} \end{array}$$

such that $Q\mathcal{V}_{\mathsf{w}}(PS)$ is all zero. Since $[\mathcal{V}_{\mathsf{w}}(PS) + \mathcal{V}_{\mathsf{w}}(PU)]_{B}$ is all zero, we have that Q acts as the identity on $\mathcal{V}_{\mathsf{w}}(PS) + \mathcal{V}_{\mathsf{w}}(PU)$, that is, $Q(\mathcal{V}_{\mathsf{w}}(PS) + \mathcal{V}_{\mathsf{w}}(PU)) = \mathcal{V}_{\mathsf{w}}(PS) + \mathcal{V}_{\mathsf{w}}(PU)$. We thus find that

$$\mathcal{V}_{\mathsf{w}}(T_{\mathsf{G},m}) = [\mathcal{V}_{\mathsf{w}}(PS) + \mathcal{V}_{\mathsf{w}}(PU)]_{\overline{B}} = [Q\mathcal{V}_{\mathsf{w}}(PS) + Q\mathcal{V}_{\mathsf{w}}(PU)]_{\overline{B}} = [Q\mathcal{V}_{\mathsf{w}}(PU)]_{\overline{B}}.$$

We conclude that $T_{\mathsf{G},m} \leq U$. □

12.6. Compression theorem for direct sums of tensor networks

There is a straightforward extension of Theorem 12.17 to a compression theorem for *direct sums* of tensor networks. Via this compression theorem we will see (as for single tensor networks) that the flattening ranks are anchors in the asymptotic spectra of graph tensors. In Section 12.7 we will discuss the convexity theorems that follow from this and our anchor theorem of Section 8.

The compression theorem for direct sums of tensor networks has a similar form as Theorem 12.17. Namely, for any collection of p tensor networks $T^{(1)}, \dots, T^{(p)}$ on some fixed graph, if there is a restriction to the direct sum $\bigoplus_{i=1}^p T^{(i)} \leq S \oplus U$ for two tensors S and U such that S is "small enough", then there are tensor networks $T'^{(i)}$ "close to" $T^{(i)}$ such that $\bigoplus_{i=1}^p T'^{(i)} \leq U$. The precise theorem is:

Theorem 12.20 (Compression theorem for direct sums of tensor networks)**.** *Let $\mathbb{F}$ be sufficiently large. Let $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ be a graph, let $n^{(i)} \in \mathbb{N}^{\mathsf{E}}$ for $i \in [p]$, and let $\mathsf{w} \in \mathsf{V}$. Suppose that*

$$\bigoplus_{i=1}^{p} T_{\mathsf{G},n^{(i)}} \leq S \oplus U$$

and

$$\dim(\mathcal{V}_{\mathsf{w}}(S)) \leq \sum_{i=1}^{p} \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})} n_{\mathsf{e}}^{(i)} \quad \big(= \dim \mathcal{V}_{\mathsf{w}}\big(\textstyle\bigoplus_{i=1}^p T_{\mathsf{G},n^{(i)}}\big)\big).$$

Then there are integers $0 \leq q^{(i)} \leq \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})} n_{\mathsf{e}}^{(i)}$ for $i \in [p]$ that sum to $\dim(\mathcal{V}_{\mathsf{w}}(S))$, such that the following holds:

For any $m^{(i)} \in \mathbb{N}^{\mathsf{E}}$ such that for every $i \in [p]$ and $\mathsf{e} \in \mathsf{E}$ it holds that $m_{\mathsf{e}}^{(i)} \leq n_{\mathsf{e}}^{(i)}$, and such that for every $i \in [p]$ it holds that $q^{(i)} < \prod_{\mathsf{e}\in\mathsf{E}(\mathsf{w})}(n_{\mathsf{e}}^{(i)} - m_{\mathsf{e}}^{(i)} + 1)$, we have

$$\bigoplus_{i=1}^{p} T_{\mathsf{G},m^{(i)}} \leq U.$$

Proof. The proof of Theorem 12.20 follows along the exact same lines as the proof of Theorem 12.17 above, with the projection property and the invariance property extending directly, except that we replace the compression theorem Theorem 11.10 (i) with its more general version Theorem 11.11 (iv) where we let $\mathcal{V}^i = \mathcal{V}_{\mathsf{w}}(T_{\mathsf{G},n^{(i)}})$. ☐

We now point out the important consequence of Theorem 12.20 that the flattening ranks are anchors in the asymptotic spectra of tensor networks. We will first state and prove this in the univariate sense (Definition 8.5) and then state this (without proof) in the multivariate sense (Definition 8.9). Let $\mathsf{G} = (\mathsf{V}, \mathsf{E})$.

Theorem 12.21. *Let $\mathbb{F}$ be sufficiently large. If $((T_{\mathsf{G},2})^{\otimes k})^{\oplus p} \leq S \oplus U$ for some tensors S and U, and*

$$\dim(\mathcal{V}_{\mathsf{w}}(S)) \leq \dim \mathcal{V}_{\mathsf{w}}((T_{\mathsf{G},2})^{\otimes(k-1)})^{\oplus p} = p\,(2^{|\mathsf{E}(w)|})^{k-1},$$

then $\big((T_{\mathsf{G},2})^{\otimes(k-1)}\big)^{\oplus\lceil p/2\rceil} \leq U$.

Theorem 12.21 implies that the flattening rank $T \mapsto \dim \mathcal{V}_{\mathsf{w}}(T)$ is an anchor in the asymptotic spectrum $\mathcal{X}(T_{\mathsf{G},2})$ (Definition 8.5).

Proof. The proof is an application of Theorem 12.20. We will use a simple averaging argument to show that for half of the p copies of $(T_{\mathsf{G},2})^{\otimes k}$ we have a good bound on the (relevant) contribution of S. To these we apply the compression of Theorem 12.20 (hence the exponent $\lceil p/2 \rceil$ in the conclusion of the theorem).

Let $0 \leq q^{(i)} \leq (2^{|\mathsf{E}(\mathsf{w})|})^k$ for $i \in [p]$ be the integers which sum to $\dim \mathcal{V}_{\mathsf{w}}(S)$ obtained from Theorem 12.20 applied to $T_{\mathsf{G},n^{(i)}} = T_{\mathsf{G},2^k}$ for $i \in [p]$. The assumption

$$\dim \mathcal{V}_{\mathsf{w}}(S) \leq p\,(2^{|\mathsf{E}(w)|})^{k-1}$$

implies that for the $\lceil p/2 \rceil$ (or more) smallest values of $q^{(i)}$ over all $i \in [p]$ we have $q^{(i)} \leq (2^{|\mathsf{E}(w)|})^{k-1}$. Let $J \subseteq [p]$ be the set of those $i \in [p]$. Then $|J| = \lceil p/2 \rceil$. For every $i \in J$ we have

$$q^{(i)} \leq (2^{|\mathsf{E}(\mathsf{w})|})^{k-1} < (2^{k-1}+1)^{|\mathsf{E}(\mathsf{w})|} = \prod_{\mathsf{e} \in \mathsf{E}(\mathsf{w})} (2^k - 2^{k-1} + 1).$$

We set $m_{\mathsf{e}}^{(i)} = 2^{k-1}$ for all $i \in J$ and $m_{\mathsf{e}}^{(i)} = 0$ for all $i \in [p] \setminus J$. Then for all $i \in [p]$ we have that $q^{(i)} < \prod_{\mathsf{e} \in \mathsf{E}(\mathsf{w})} (n_{\mathsf{e}}^{(i)} - m_{\mathsf{e}}^{(i)} + 1)$. From Theorem 12.20 we obtain the claim. □

Via a similar proof that also applies Theorem 12.20 (which we leave to the reader), we obtain that the flattening rank is an anchor in the multivariate sense (Definition 8.9). To state this we use the notation $\delta_{\mathsf{e}} \in \mathbb{N}^E$ for the unit vector that satisfies $(\delta_{\mathsf{e}})_{\mathsf{e}} = 1$ and $(\delta_{\mathsf{e}})_{\mathsf{f}} = 0$ for $\mathsf{f} \neq \mathsf{e}$. Then $T_{\mathsf{G},2} = \bigotimes_{\mathsf{e} \in \mathsf{E}} T_{\mathsf{G},2\delta_{\mathsf{e}}}$.

Theorem 12.22. *Let $\mathbb{F}$ be sufficiently large. If $(\bigotimes_{\mathsf{e}} T_{\mathsf{G},2\delta_{\mathsf{e}}}^{\otimes k_{\mathsf{e}}})^{\oplus p} \leq S \oplus U$ and*

$$\dim \mathcal{V}_{\mathsf{w}}(S) \leq \dim \mathcal{V}_{\mathsf{w}}\Big(\big(\bigotimes_{\mathsf{e}} T_{\mathsf{G},2\delta_{\mathsf{e}}}^{\otimes k_{\mathsf{e}}}\big)^{\oplus p} \Big),$$

then $(\bigotimes_{\mathsf{e}} T_{\mathsf{G},2\delta_{\mathsf{e}}}^{\otimes (k_{\mathsf{e}}-1)})^{\oplus \lceil p/2 \rceil} \leq U$.

Theorem 12.22 means precisely that $T \mapsto \dim_{\mathsf{w}}(T)$, which equals the w-flattening rank of T, is an anchor in the asymptotic spectrum $\mathcal{X}((T_{\mathsf{G},2\delta_{\mathsf{e}}})_{\mathsf{e} \in \mathsf{E}}) \subseteq [1,\infty)^{\mathsf{E}}$ (Definition 8.9).

12.7. Log-star-convexity of asymptotic spectra of tensor networks

We finish this section by applying our anchor theorem of Section 8 to the anchors that we have found in Section 12.6. As before we first consider the one-dimensional version of the statement and then the more precise high-dimensional version.

The one-dimensional statement is the following connectedness theorem for the asymptotic spectrum $\mathcal{X}(T_{\mathsf{G},2})$ of the tensor network $T_{\mathsf{G},2}$.

Theorem 12.23 (Connectedness)**.** *Let G be a graph. Then $\mathcal{X}(T_{\mathsf{G},2}) \subseteq [1,\infty)$ is connected.*

In particular, for every graph G the asymptotic spectrum $\mathcal{X}(T_{\mathsf{G},2})$ of $T_{\mathsf{G},2}$ is the closed interval from asymptotic subrank to asymptotic rank, $[\widetilde{\mathrm{Q}}(T_{\mathsf{G},2}), \widetilde{\mathrm{R}}(T_{\mathsf{G},2})]$.

Proof. We know from Theorem 12.21 and Lemma 12.7 that, for any vertex $\mathsf{w} \in \mathsf{V}$ of the graph G the flattening rank $T \mapsto \dim(\mathcal{V}_{\mathsf{w}}(T))$ is an anchor in the asymptotic spectrum $\mathcal{X}(T_{\mathsf{G},2})$ (Definition 8.5). Applying the anchor theorem (Theorem 8.6) proves the claim. □

Similarly, the multivariate anchor of Theorem 12.22 gives the following log-star-convexity theorem for the asymptotic spectrum $\mathcal{X}((T_{\mathsf{G},2\delta_{\mathsf{e}}})_{\mathsf{e}\in\mathsf{E}})$ of the tensors $T_{\mathsf{G},2\delta_{\mathsf{e}}}$ for $\mathsf{e} \in \mathsf{E}$ (which in fact implies Theorem 12.23). Recall that δ_{e} is the vector in $\mathbb{N}^E$ that has value 1 at index e and value 0 elsewhere.

Theorem 12.24 (Log-star-convexity)**.** *Let $\mathsf{G} = (\mathsf{V}, \mathsf{E})$ be a graph. For $\mathsf{e} \in \mathsf{E}$ let $T_{\mathsf{e}} = T_{\mathsf{G},2\delta_{\mathsf{e}}}$. Then*

$$\mathcal{X}((T_{\mathsf{e}})_{\mathsf{e}\in\mathsf{E}}) \subseteq [1, \infty)^{\mathsf{E}}$$

is log-star-convex with respect to the w*-flattening rank for every* $\mathsf{w} \in \mathsf{V}$.

The matrix multiplication tensors MM_n are instances of tensor networks, namely the square matrix multiplication tensors satisfy $\mathrm{MM}_n = T_{\mathsf{C}_3,n}$ and the rectangular matrix multiplication tensors satisfy $\mathrm{MM}_{n_1,n_2,n_3} = T_{\mathsf{C}_3,(n_1,n_2,n_3)}$. In this way we obtain the following special cases of the above theorems, which are among the main results of Strassen [Str88].

Corollary 12.25 (Connectedness for matrix multiplication, Strassen [Str88])**.** *The asymptotic spectrum of matrix multiplication $\mathcal{X}(\mathrm{MM}_2)$ is connected.*

Corollary 12.26 (Log-star-convexity for matrix multiplication, Strassen [Str88])**.** *The asymptotic spectrum of rectangular matrix multiplication $\mathcal{X}_{\mathrm{rect}} = \mathcal{X}(\mathrm{MM}_{2,1,1}, \mathrm{MM}_{1,2,1}, \mathrm{MM}_{1,1,2})$ is log-star-convex with respect to the points $(2, 2, 0), (2, 0, 2), (0, 2, 2)$.*

(Again, Corollary 12.26 implies Corollary 12.25.)

Recall (Lemma 8.8) that by star-convexity the set of centers in $\log \mathcal{X}_{\mathrm{rect}}$ is convex, so it follows from Corollary 12.26 that $\log \mathcal{X}_{\mathrm{rect}}$ is star-convex with respect to the (filled) triangle with vertices $(1, 1, 0)$, $(1, 0, 1)$, $(0, 1, 1)$.

Since the smallest point in $\mathcal{X}(\mathrm{MM}_2)$ is the asymptotic subrank $\widetilde{\mathrm{Q}}(\mathrm{MM}_2)$ and the value of $\widetilde{\mathrm{Q}}(\mathrm{MM}_2)$ is known to be equal to the flattening rank in this case (which is 4) (Theorem 4.22), we have by Corollary 12.25 that $\mathcal{X}(\mathrm{MM}_2) = [4, 2^{\omega}]$, where ω is the matrix multiplication exponent. In Section 9 we have already discussed the consequences of this connectedness theorem (in the form of a generalized Schönhage tau theorem).

13. Open problems and directions

In this section we list a number of (some interrelated) research directions and open problems which naturally arise given the material in this paper. Many of them were discussed in earlier sections in some detail, and so we will be rather brief here.

Tensors and matrix multiplication. Determining the exponent ω of matrix multiplication was the main driving force for developing the theory exposited here, and remains a great mystery today as well. Inventing new ways of finding better upper bounds, lower bounds, and barrier results is the most obvious direction to pursue.

It is a tantalizing possibility that $\omega = 2$, giving (almost) linear time complexity for solving essentially all linear algebra problems. In fact, a far more striking possibility is that the asymptotic rank of *every* tensor is as small as it can be. More precisely, that for every tensor in $(\mathbb{F}^n)^{\otimes d}$, the asymptotic rank is at most n (observe that for matrix multiplication tensors this very general bound implies $\omega = 2$). This would sharply contrast the fact that the (non-asymptotic) rank is typically around n^{d-1} [Str83], but there are no known methods to rule out this possibility. We consider this a major challenge. (See also [BCS97, Problem 15.5].) Related to this, little is known about the precise behavior of rank under powering. (See e.g., [CJZ18, CLGV20, CHL20].)

In fact, another surprising connection between asymptotic rank and ω is known, which is in the opposite direction: a general upper bound on the asymptotic rank of *any* tensor in terms of ω (making in a sense matrix multiplication "universal" for this problem, at least for $d = 3$). Namely, restricting ourselves to $n \times n \times n$ tensors, we have for every T the upper bound $\widetilde{\mathrm{R}}(T) \leq n^{2\omega/3} \leq n^{1.6}$ [Str88, Prop. 3.6] (see also [CVZ21, Prop. 2.12]). (This is in contrast to the situation for tensor rank, where we know by a dimension counting argument that for almost every T we have $\mathrm{R}(T) \geq \Omega(n^2)$.) We state the following bold conjecture: for every T it holds that $\widetilde{\mathrm{R}}(T) = \mathcal{O}(n)$.

Strassen's duality and Positivstellensätze. As discussed in Section 4.6, Strassen's duality theorem may be viewed as a generalization of classical duality theorems in real algebraic geometry, in particular variants of Positivstellensätze. Both Strassen's and the Positivstellensatz characterize unsatisfiable systems of polynomial inequalities, but the nature of the characterization seems very different. While several works explore this connection as we discussed, we feel there is more to understand, especially whether Strassen's theorem has new applications in real algebraic geometry and optimization.

Types. Section 6 and Section 9.3 discuss in some detail notions of "type decompositions" of the summands in a multinomial sum arising from taking a large power of elements in preordered semirings. In Section 6 such a decomposition is developed for polynomials, as part of establishing connectedness of asymptotic spectra, and in Section 9.3 we mention different decompositions in other semirings, for this purpose but also other applications and constructions. We feel that there is room for a general theory of types, which will rely on representation theory. Simply, many preordered semirings have natural symmetries, namely a group action which renders elements of the semiring isomorphic, and several of the type decompositions mentioned are dictated by the irreducible representation of $G \times \mathfrak{S}_n$ where G is the acting group, $\mathfrak{S}_n$ is the symmetric group, and n is the large power taken. It seems like a systematic study of these may yield a nice theory, generalizing type decomposition in information theory, which may lead to better understanding and unification of existing results and perhaps lead to others.

Anchors. In Section 8 we introduced the notion of an *anchor* (a special point in the asymptotic spectrum), and the anchor method for proving connectivity (and even log-star-convexity) of the spectrum. Proving the existence of anchors, which is the hardest part of implementing this method, seem to require certain basis shifting and compression arguments which we develop (respectively) in Sections 11 and 12. We don't know if there are different ways to prove the existence of anchors. While we were able to generalize Strassen's work, and use anchors to prove connectivity of spectra of many more tensors beyond matrix multiplication, we have no idea how general this method is, and to which other semirings it might apply. The semiring of graphs is an interesting case in point, for which connectivity and even log-convexity of the spectrum was proved (by Vrana [Vra21]) using a different method, but we feel it can yield to the anchor method as well. It would be interesting to prove or disprove that if the spectrum is log-convex, all its elements are anchors.

Basis shifting. In Section 11 we prove, via elementary though somewhat subtle linear algebra arguments, several theorems of the following nature. Given a set (actually, a group) of allowed linear transformations on a tensor of vectors, we can make certain subsets linearly independent. In these theorems the allowed transformations are given by the symmetries of tensors. This raises the following general question. Let $\mathcal{V} = \{v_i : i \in I\}$ be a tensor of vectors spanning a d-dimensional subspace over a field $\mathbb{F}$. Let G be a subgroup of $\mathrm{GL}(I, \mathbb{F})$. Characterize the d-subsets of I which can be made linearly independent by applying linear transformations from G to the vectors in $\mathcal{V}$.

Compression. In Section 12 we establish "compression" theorems for multilinear functions associated with certain tensors. At a high level, these theorems have the following general structure, providing a certain kind of "downward self-reducibility", or "average-case to worst-case reduction". Many of the terms below are left vague on purpose.

Consider a (uniform) family of functions $f = \{f_n : \Sigma^n \to \Sigma\}$, on some domain Σ (these can be Boolean functions, or polynomials over some field, or anything else). Assume that for every n we have an algorithm (or circuit) computing f_n on some large, structured subset of Σ^n. Compression above means that we can convert such an algorithm into another, with a similar complexity, which computes $f_{n'}$ on the full domain $\Sigma^{n'}$, for n' not much smaller than n.

For low degree polynomials over large finite fields, average-case to worst-case reductions [Lip89, BF90] provide such compression for sufficiently large subsets, without any loss (namely $n' = n$). Strassen's theorem, generalized in Section 12, provides such a compression for matrix multiplication $f(A, B) = AB$, when A (say) is taken from any *subspace* of matrices with high enough dimension, and (say) $n' = n/2$.

It is natural to wonder for what other general situations one can prove such efficient compression results. The model of computation and resource bound can vary. An interesting case in point (dear to the first author) is the following. Before Reingold's celebrated $\mathsf{SL} = \mathsf{L}$ result, the following slightly weaker result was proved by [GW02]: there is a logspace algorithm which correctly solves undirected graph connectivity, for all but at most $\exp(n^{\varepsilon})$ of the n-vertex graphs. A compression result of the type above would establish $\mathsf{SL} = \mathsf{L}$ immediately.

Asymptotic spectra. Perhaps the most general research direction, which encompasses many of the ones above, is the understanding of the structure of asymptotic spectra of preordered semirings, and applying that understanding to obtain bounds on asymptotic parameters of interest. Numerous examples of such settings are given in Section 2, with many concrete open problems in each. One favorite concrete one is establishing the Shannon capacity of the 7-cycle C_7, which has been open for almost half a century, since Lovász' famous resolution of the Shannon capacity of the 5-cycle C_5.

Real algebraic geometry. The type decomposition for polynomials of Section 6 gives rise to a "tropical version" of polynomials as we discussed in Remark 6.22. We wonder (speculatively) whether there are further connections and relevance of these ideas to real algebraic geometry, for instance to understanding semi-algebraic sets and inequalities in general. In this context we mention the work of Grigoriev and Podolskii on the theory of systems of tropical inequalities [GP18]. Razborov's flag algebra [Raz07] and the theory of graph densities is another asymptotic theory which has recently started using Tropical Geometric methods (Blekherman, Raymond, Singh and Thomas [BRST20]), and relationships with Strassen's theory can be sought.

Acknowledgements. We thank Yuval Wigderson for thoroughly reading the manuscript and pointing out many improvements. We also thank Markus Bläser, Itai Leigh and Maxim van den Berg for much helpful feedback on the manuscript. AW was partially supported by NSF grant CCF-1900460. JZ was supported by the Institute for Advanced Study, Princeton, a Simons Junior Fellowship, and an NWO Veni grant (VI.Veni.212.284).

References

[AD93] Noga Alon and Moshe Dubiner. Zero-sum sets of prescribed size. *Combinatorics, Paul Erdos is Eighty*, 1:33–50, 1993. URL: http://www.math.tau.ac.il/~nogaa/PDFS/egz1.pdf. 30

[AFLG15] Andris Ambainis, Yuval Filmus, and François Le Gall. Fast matrix multiplication: limitations of the Coppersmith-Winograd method (extended abstract). In *Proceedings of the 2015 ACM Symposium on Theory of Computing (STOC 2015)*, pages 585–593. ACM, 2015. doi:10.1145/2746539.2746554. 58

[AL06] Noga Alon and Eyal Lubetzky. The Shannon capacity of a graph and the independence numbers of its powers. *IEEE Transactions on Information Theory*, 52(5):2172–2176, 2006. arXiv:0608021, doi:10.1109/TIT.2006.872856. 27, 48

[Alm19] Josh Alman. Limits on the universal method for matrix multiplication. In *Proceedings of the 34th Computational Complexity Conference (CCC 2019)*, pages 12:1–12:24, 2019. arXiv:1812.08731, doi:10.4230/LIPIcs.CCC.2019.12. 58, 59

[Alo98a] Noga Alon. On the capacity of digraphs. *European Journal of Combinatorics*, 19(1):1–5, 1998. doi:10.1006/eujc.1997.0148. 54

[Alo98b] Noga Alon. The Shannon capacity of a union. *Combinatorica*, 18(3):301–310, 1998. doi:10.1007/PL00009824. 5, 27, 49

[AW18] Josh Alman and Virginia Vassilevska Williams. Limits on all known (and some unknown) approaches to matrix multiplication. In *Proceedings of the 59th Annual Symposium on Foundations of Computer Science (FOCS 2018)*, pages 580–591, Oct 2018. arXiv:1810.08671, doi:10.1109/FOCS.2018.00061. 58

[AW21] Josh Alman and Virginia Vassilevska Williams. A refined laser method and faster matrix multiplication. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA 2021)*, pages 522–539. SIAM, 2021. doi:10.1137/1.9781611976465.32. 28, 58

[BC19] Boris Bukh and Christopher Cox. On a fractional version of Haemers' bound. *IEEE Trans. Inf. Theory*, 65(6):3340–3348, 2019. arXiv:1802.00476, doi:10.1109/TIT.2018.2889108. 27, 44

[BCC+17a] Jonah Blasiak, Thomas Church, Henry Cohn, Joshua A. Grochow, Eric Naslund, William F. Sawin, and Chris Umans. On cap sets and the group-theoretic approach to matrix multiplication. *Discrete Anal.*, 2017. arXiv:1605.06702, doi:10.19086/da.1245. 58

[BCC+17b] Jonah Blasiak, Thomas Church, Henry Cohn, Joshua A Grochow, and Chris Umans. Which groups are amenable to proving exponent two for matrix multiplication? *arXiv*, 2017. arXiv:1712.02302. 58

[BCS97] Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi. *Algebraic complexity theory*, volume 315 of *Grundlehren Math. Wiss.* Springer-Verlag, Berlin, 1997. doi:10.1007/978-3-662-03338-8. 52, 122

[BF90] Donald Beaver and Joan Feigenbaum. Hiding instances in multioracle queries. In Christian Choffrut and Thomas Lengauer, editors, *STACS 90, 7th Annual Symposium on Theoretical Aspects of Computer Science, Rouen, France, February 22-24, 1990, Proceedings*, volume 415 of *Lecture Notes in Computer Science*, pages 37–48. Springer, 1990. doi:10.1007/3-540-52282-4_30. 123

[BGS75] Theodore P. Baker, John Gill, and Robert Solovay. Relativizations of the P =? NP question. *SIAM J. Comput.*, 4(4):431–442, 1975. doi:10.1137/0204037. 57

[BI11] Peter Bürgisser and Christian Ikenmeyer. Geometric complexity theory and tensor rank. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing (STOC 2011)*, pages 509–518, 2011. doi:10.1145/1993636.1993704. 94

[Blä13] Markus Bläser. *Fast Matrix Multiplication*. Number 5 in Graduate Surveys. Theory of Computing Library, 2013. doi:10.4086/toc.gs.2013.005. 9, 52

[BR14] Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Trans. Inf. Theory*, 60(10):6058–6069, 2014. arXiv:1106.3595, doi:10.1109/TIT.2014.2347282. 5

[Bri87] Michel Brion. Sur l'image de l'application moment. In *Séminaire d'algèbre Paul Dubreil et Marie-Paule Malliavin (Paris, 1986)*, volume 1296 of *Lecture Notes in Math.*, pages 177–192. Springer, Berlin, 1987. doi:10.1007/BFb0078526. 95

[Bri21] Jop Briët. Subspaces of tensors with high analytic rank. *Online J. Anal. Comb.*, 16(6), 2021. URL: https://hosted.math.rochester.edu/ojac/vol16/216.pdf, arXiv:1908.04169. 109

[BRST20] Grigoriy Blekherman, Annie Raymond, Mohit Singh, and Rekha R. Thomas. Tropicalization of graph profiles. *arXiv*, 2020. arXiv:2004.05207. 123

[BS83] Eberhard Becker and Niels Schwartz. Zum Darstellungssatz von Kadison-Dubois. *Arch. Math. (Basel)*, 40(5):421–428, 1983. doi:10.1007/BF01192806. 11, 61

[Bür90] Peter Bürgisser. *Degenerationsordnung und Trägerfunktional bilinearer Abbildungen*. PhD thesis, Universität Konstanz, 1990. http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-20311. 4, 10

[BV21] Gergely Bunth and Péter Vrana. Asymptotic relative submajorization of multiple-state boxes. *Lett. Math. Phys.*, 111(4), 2021. arXiv:2007.11258, doi:10.1007/s11005-021-01430-0. 62

[CFTZ21] Matthias Christandl, Omar Fawzi, Hoang Ta, and Jeroen Zuiddam. Communication complexity, corner-free sets and the symmetric subrank of tensors, 2021. arXiv:2104.01130. 55

[CGLZ20] Matthias Christandl, François Le Gall, Vladimir Lysikov, and Jeroen Zuiddam. Barriers for rectangular matrix multiplication, 2020. arXiv:2003.03019. 58, 61, 94, 95

[CHL20] Austin Conner, Hang Huang, and J. M. Landsberg. Bad and good news for Strassen's laser method: Border rank of the 3×3 permanent and strict submultiplicativity, 2020. arXiv:2009.11391. 122

[CJZ18] Matthias Christandl, Asger Kjærulff Jensen, and Jeroen Zuiddam. Tensor rank is not multiplicative under the tensor product. *Linear Algebra Appl.*, 543:125–139, 2018. doi:10.1016/j.laa.2017.12.020. 122

[CKSU05] Henry Cohn, Robert Kleinberg, Balazs Szegedy, and Christopher Umans. Group-theoretic algorithms for matrix multiplication. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005)*, pages 379–388. IEEE, 2005. doi:10.1109/SFCS.2005.39. 13, 88

[CLGV20] Austin Conner, Joseph M. Landsberg, Fulvio Gesmundo, and Emanuele Ventura. Kronecker powers of tensors and Strassen's laser method. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151, pages 10:1–10:28, 2020. doi:10.4230/LIPIcs.ITCS.2020.10. 122

[CLP17] Ernie Croot, Vsevolod F. Lev, and Péter Pál Pach. Progression-free sets in $\mathbf{Z}_4^n$ are exponentially small. *Ann. of Math. (2)*, 185(1):331–337, 2017. doi:10.4007/annals.2017.185.1.7. 30

[CLZ23] Matthias Christandl, Vladimir Lysikov, and Jeroen Zuiddam. Weighted slice rank and a minimax correspondence to Strassen's spectra. *J. Math. Pures Appl. (9)*, 2023. arXiv:2012.14412, doi:10.1016/j.matpur.2023.02.006. 44, 94, 95

[CM21] Alex Cohen and Guy Moshkovitz. Partition and analytic rank are equivalent over large fields, 2021. arXiv:2102.10509, doi:10.48550/ARXIV.2102.10509. 109

[CMW08] Toby Cubitt, Ashley Montanaro, and Andreas Winter. On the dimension of subspaces with bounded Schmidt rank. *J. Math. Phys.*, 49(2):022107, 6, 2008. doi:10.1063/1.2862998. 109

[CT12] Thomas M. Cover and Joy A. Thomas. *Elements of information theory.* John Wiley & Sons, 2012. doi:10.1002/047174882X. 68, 71, 93

[CU03] Henry Cohn and Christopher Umans. A group-theoretic approach to fast matrix multiplication. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2003)*, pages 438–449. IEEE, 2003. doi:10.1109/SFCS.2003.1238217. 13, 88

[CU13] Henry Cohn and Christopher Umans. Fast matrix multiplication using coherent configurations. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2013)*, pages 1074–1086. SIAM, 2013. doi:10.1137/1.9781611973105.77. 13, 88

[CVZ18] Matthias Christandl, Péter Vrana, and Jeroen Zuiddam. Universal points in the asymptotic spectrum of tensors. In *Proceedings of 50th Annual ACM SIGACT Symposium on the Theory of Computing (STOC 2018)*. ACM, 2018. arXiv:1709.07851, doi:10.1145/3188745.3188766. 44, 60, 94, 95

[CVZ19a] Matthias Christandl, Péter Vrana, and Jeroen Zuiddam. Asymptotic tensor rank of graph tensors: beyond matrix multiplication. *Comput. Complex.*, 28(1):57–111, 2019. arXiv:1609.07476, doi:10.1007/s00037-018-0172-8. 111

[CVZ19b] Matthias Christandl, Péter Vrana, and Jeroen Zuiddam. Barriers for fast matrix multiplication from irreversibility. In *34th Computational Complexity Conference (CCC 2019)*, pages 26:1–26:17, 2019. doi:10.4230/LIPIcs.CCC.2019.26. 58, 59, 60, 61

[CVZ21] Matthias Christandl, Péter Vrana, and Jeroen Zuiddam. Barriers for fast matrix multiplication from irreversibility. *Theory of Computing*, 17(2):1–32, 2021. URL: http://www.theoryofcomputing.org/articles/v017a002, doi:10.4086/toc.2021.v017a002. 122

[CW82] Don Coppersmith and Shmuel Winograd. On the asymptotic complexity of matrix multiplication. *SIAM J. Comput.*, 11(3):472–492, 1982. doi:10.1137/0211038. 48

[CW87] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing (STOC 1987)*, pages 1–6. ACM, 1987. doi:10.1145/28395.28396. 58

[CW90] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. *J. Symbolic Comput.*, 9(3):251–280, 1990. doi:10.1016/S0747-7171(08)80013-2. 58

[Die49] Jean Dieudonné. Sur une généralisation du groupe orthogonal à quatre variables. *Arch. Math.*, 1:282–287, 1949. doi:10.1007/BF02038756. 109

[DL78] Richard A. Demillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Information Processing Letters*, 7(4):193–195, 1978. doi:10.1016/0020-0190(78)90067-4. 101

[DVC00] Wolfgang Dür, Guivre Vidal, and Juan Ignacio Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A (3)*, 62(6):062314, 12, 2000. doi:10.1103/PhysRevA.62.062314. 27

[Ede04] Yves Edel. Extensions of generalized product caps. *Des. Codes Cryptogr.*, 31(1):5–14, 2004. doi:10.1023/A:1027365901231. 30

[EG17] Jordan S. Ellenberg and Dion Gijswijt. On large subsets of $\mathbf{F}_q^n$ with no three-term arithmetic progression. *Ann. of Math. (2)*, 185(1):339–343, 2017. doi:10.4007/annals.2017.185.1.8. 30

[EGdOW18] Klim Efremenko, Ankit Garg, Rafael Mendes de Oliveira, and Avi Wigderson. Barriers for rank methods in arithmetic complexity. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, 2018. doi:10.4230/LIPIcs.ITCS.2018.1. 57

[Fla62] Harley Flanders. On spaces of linear transformations with bounded rank. *Journal of the London Mathematical Society*, 1(1):10–16, 1962. doi:10.1112/jlms/s1-37.1.10. 109

[Fra02] Matthias Franz. Moment polytopes of projective G-varieties and tensor products of symmetric group representations. *J. Lie Theory*, 12(2):539–549, 2002. URL: http://emis.ams.org/journals/JLT/vol.12_no.2/16.html. 95

[Fri17] Tobias Fritz. Resource convertibility and ordered commutative monoids. *Math. Structures Comput. Sci.*, 27(6):850–938, 2017. doi:10.1017/S0960129515000444. 10, 62

[Fri21] Tobias Fritz. A generalization of strassen's positivstellensatz. *Communications in Algebra*, 49(2):482–499, 2021. arXiv:1810.08667, doi:10.1080/00927872.2020.1803344. 10, 11, 61, 62

[GKV92] L. Gargano, J. Körner, and U. Vaccaro. Qualitative independence and Sperner problems for directed graphs. *J. Combin. Theory Ser. A*, 61(2):173–192, 1992. doi:10.1016/0097-3165(92)90016-N. 54

[GNW11] Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's XOR-lemma. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, volume 6650 of *Lecture Notes in Computer Science*, pages 273–301. Springer, 2011. doi:10.1007/978-3-642-22670-0_23. 6

[GP18] Dima Grigoriev and Vladimir V. Podolskii. Tropical effective primary and dual Nullstellensätze. *Discrete & Computational Geometry*, 59(3):507–552, 2018. doi:10.1007/s00454-018-9966-3. 123

[GW02] Oded Goldreich and Avi Wigderson. Derandomization that is rarely wrong from short advice that is typically good. In *Randomization and Approximation Techniques, 6th International Workshop (RANDOM 2002)*, pages 209–223, 2002. doi:10.1007/3-540-45726-7_17. 123

[GW11] William Timothy Gowers and Julia Wolf. Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$. *Geom. Funct. Anal.*, 21(1):36–69, 2011. doi:10.1007/s00039-010-0106-3. 109

[Hae79] Willem Haemers. On some problems of Lovász concerning the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(2):231–232, 1979. doi:10.1109/TIT.1979.1056027. 27, 49

[Har05] Aram W. Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory.* PhD thesis, MIT, 2005. arXiv:quant-ph/0512255. 94

[Hås90] Johan Håstad. Tensor rank is NP-complete. *J. Algorithms*, 11(4):644–654, 1990. doi:10.1016/0196-6774(90)90014-6. 28

[Hed66] Stephen Hedetniemi. Homomorphisms of graphs and automata, technical report 03105-44-t, 1966. 52

[HN04] Pavol Hell and Jaroslav Nesetril. *Graphs and homomorphisms*, volume 28 of *Oxford lecture series in mathematics and its applications.* Oxford University Press, 2004. doi:10.1093/acprof:oso/9780198528173.001.0001. 26

[Hol07] Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC 2007)*, pages 411–419, 2007. doi:10.1145/1250790.1250852. 6

[HW21] Xiaoyu He and Yuval Wigderson. Hedetniemi's conjecture is asymptotically false. *Journal of Combinatorial Theory, Series B*, 146:485–494, 2021. doi:10.1016/j.jctb.2020.03.003. 52

[Jac01] Thomas Jacobi. A representation theorem for certain partially ordered commutative rings. *Math. Z.*, 237(2):259–273, 2001. doi:10.1007/PL00004868. 61

[JM64] Arno Jaeger and Bertram Mond. On direct sums and tensor products of linear programs. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 3:19–31, 1964. doi:10.1007/BF00531681. 6

[Juh82] Ferenc Juhász. The asymptotic behaviour of Lovász' θ function for random graphs. *Combinatorica*, 2(2):153–155, 1982. doi:10.1007/BF02579314. 27

[KMZ20] Swastik Kopparty, Guy Moshkovitz, and Jeroen Zuiddam. Geometric Rank of Tensors and Subrank of Matrix Multiplication. In *Proceedings of the 35th Computational Complexity Conference (CCC 2020)*, pages 35:1–35:21, 2020. arXiv:2002.09472, doi:10.4230/LIPIcs.CCC.2020.35. 109

[Kri64] Jean-Louis Krivine. Anneaux préordonnés. *J. Anal. Math.*, 12:307–326, 1964. doi:10.1007/BF02807438. 11, 61

[KS92] János Körner and Gábor Simonyi. A Sperner-type theorem and qualitative independence. *J. Combin. Theory Ser. A*, 59(1):90–103, 1992. doi:10.1016/0097-3165(92)90100-9. 54

[LG12] François Le Gall. Faster algorithms for rectangular matrix multiplication. In *IEEE 53rd Annual Symposium on Foundations of Computer Science (FOCS 2012)*, pages 514–523. IEEE, 2012. doi:10.1109/FOCS.2012.80. 58

[Lip89] Richard J. Lipton. New directions in testing. In Joan Feigenbaum and Michael Merritt, editors, *Distributed Computing And Cryptography, Proceedings of a DIMACS Workshop*, volume 2 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 191–202, 1989. doi:10.1090/dimacs/002/13. 123

[Lov79] László Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979. doi:10.1109/TIT.1979.1055985. 6, 26, 27, 44, 48

[Lov19] Shachar Lovett. The analytic rank of tensors and its applications. *Discrete Anal.*, 2019. doi:10.19086/da.8654. 109

[LY99] Elliott H. Lieb and Jakob Yngvason. The physics and mathematics of the second law of thermodynamics. *Phys. Rep.*, 310(1):96, 1999. doi:10.1016/S0370-1573(98)00082-9. 62

[Mar08] Murray Marshall. *Positive polynomials and sums of squares*, volume 146 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2008. doi:10.1090/surv/146. 9, 38, 61

[Mau98] Franz Mauch. *Ein Randverteilungsproblem und seine Anwendung auf das asymptotische Spektrum bilinearer Abbildungen*. PhD thesis, Universität Konstanz, 1998. http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-20543. 4

[Mes85] Roy Meshulam. On the maximal rank in a subspace of matrices. *Quart. J. Math. Oxford Ser. (2)*, 36(142):225–229, 1985. doi:10.1093/qmath/36.2.225. 109

[Mes95] Roy Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *Journal of Combinatorial Theory, Series A*, 71(1):168 – 172, 1995. doi:10.1016/0097-3165(95)90024-1. 30

[Mes18] Roy Meshulam. Extremal problems for spaces of matrices. Institute for Advanced Study, Princeton, June 2018. URL: https://www.ias.edu/video/OCIT/2018/0608-RoyMeshulam. 110

[Mon65] Bertram Mond. On direct sums and tensor products of non-linear programs. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 4:27–39, 1965. doi:10.1007/BF00535480. 6

[MS15] Diane Maclagan and Bernd Sturmfels. *Introduction to tropical geometry*, volume 161. American Mathematical Society, 2015. doi:10.1090/gsm/161. 76

[Mun00] James R. Munkres. *Topology*. Prentice Hall, second edition, 2000. 43

[Nas20] Eric Naslund. The partition rank of a tensor and k-right corners in $\mathbb{F}_q^n$. *J. Combin. Theory Ser. A*, 174:105190, 25, 2020. doi:10.1016/j.jcta.2019.105190. 109

[Nes84] Linda Ness. A stratification of the null cone via the moment map. *Amer. J. Math.*, 106(6):1281–1329, 1984. With an appendix by David Mumford. doi:10.2307/2374395. 95

[NRS98] Noam Nisan, Steven Rudich, and Michael Saks. Products and help bits in decision trees. *SIAM Journal on Computing*, 28(3):1035–1050, 1998. doi:10.1137/S0097539795282444. 6

[Orú14] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, 2014. doi:10.1016/j.aop.2014.06.013. 110

[Pan72] Victor Ya. Pan. On schemes for the evaluation of products and inverses of matrices. *Uspekhi Matematicheskikh Nauk*, 27(5):167, 1972. (In Russian). 28

[Pan78] Victor Ya. Pan. Strassen's algorithm is not optimal. Trilinear technique of aggregating, uniting and canceling for constructing fast algorithms for matrix operations. In *19th Annual Symposium on Foundations of Computer Science (FOCS 1978)*, pages 166–176. IEEE, 1978. doi:10.1109/SFCS.1978.34. 28

[Par04] K. R. Parthasarathy. On the maximal dimension of a completely entangled subspace for finite level quantum systems. *Proceedings Mathematical Sciences*, 114(4):365–374, November 2004. doi:10.1007/BF02829441. 108, 109

[Pot17] Aaron Potechin. A note on amortized branching program complexity. In *32nd Computational Complexity Conference (CCC 2017)*, volume 79, pages 4:1–4:12, 2017. doi:10.4230/LIPIcs.CCC.2017.4. 62

[PS98] George Pólya and Gabor Szegő. *Problems and theorems in analysis. I.* Classics in Mathematics. Springer-Verlag, Berlin, 1998. Series, integral calculus, theory of functions, Translated from the German by Dorothee Aeppli, Reprint of the 1978 English translation. doi:10.1007/978-3-642-61905-2. 21

[PS19] Sven C. Polak and Alexander Schrijver. New lower bound on the Shannon capacity of C7 from circular graphs. *Information Processing Letters*, 143:37 – 40, 2019. doi:https://doi.org/10.1016/j.ipl.2018.11.006. 26, 61

[Put93] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.*, 42:969–984, 1993. doi:10.1512/iumj.1993.42.42045. 11, 61

[PVW22] Christopher Perry, Péter Vrana, and Albert H. Werner. The semiring of dichotomies and asymptotic relative submajorization. *IEEE Trans. Inform. Theory*, 68(1):311–321, 2022. doi:10.1109/TIT.2021.3117440. 62

[Pó28] G. Pólya. Über positive Darstellung von Polynomen. *Vierteljschr, Naturforsch. Ges. Zürich*, 73:141–145, 1928. 61

[Raz92] A. A. Razborov. On submodular complexity measures. In *Boolean function complexity*, volume 169 of *London Math. Soc. Lecture Note Ser.*, pages 76–83. Cambridge Univ. Press, Cambridge, 1992. doi:10.1017/CBO9780511526633.007. 62

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998. doi:10.1137/S0097539795280895. 6

[Raz07] Alexander A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007. doi:10.2178/jsl/1203350785. 123

[Raz09] Ran Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. *J. ACM*, 56(2):Art. 8, 17, 2009. doi:10.1145/1502793.1502797. 27

[Rez95] Bruce Reznick. Uniform denominators in Hilbert's seventeenth problem. *Math. Z.*, 220(1):75–97, 1995. doi:10.1007/BF02572604. 61

[RR94] Alexander A. Razborov and Steven Rudich. Natural proofs. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing (STOC 1994)*, pages 204–213. ACM, 1994. doi:10.1145/195058.195134. 57

[RZ22] Robert Robere and Jeroen Zuiddam. Amortized circuit complexity, formal complexity measures, and catalytic algorithms. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS 2021)*, pages 759–769, 2022. URL: https://eccc.weizmann.ac.il/report/2021/035/, doi:10.1109/FOCS52979.2021.00079. 10, 11, 62

[Sch80] Jacob T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27(4):701–717, 1980. doi:10.1145/322217.322225. 101

[Sch81] Arnold Schönhage. Partial and total matrix multiplication. *SIAM J. Comput.*, 10(3):434–455, 1981. doi:10.1137/0210032. 9, 52

[Sch91] Konrad Schmüdgen. The K-moment problem for compact semi-algebraic sets. *Math. Ann.*, 289:203–206, 1991. doi:10.1007/BF01446568. 11, 61

[Sch99] Markus Schweighofer. Algorithmische Beweise für Nichtnegativ- und Positivstellensätze. *Master's thesis, Universität Passau*, 136, 1999. 61

[Sch03] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003. 21

[Sha56] Claude E. Shannon. The zero error capacity of a noisy channel. *Institute of Radio Engineers, Transactions on Information Theory*, IT-2(September):8–19, 1956. doi:10.1109/TIT.1956.1056798. 25, 27, 48, 49

[Shi19] Yaroslav Shitov. Counterexamples to Hedetniemi's conjecture. *Annals of Mathematics*, 190(2):663–667, 2019. doi:10.4007/annals.2019.190.2.6. 10, 52

[Sim21] Gábor Simonyi. Shannon capacity and the categorical product. *Electron. J. Comb.*, 28(1):P1.51, 2021. arXiv:1911.00944, doi:10.37236/9113. 52

[Ste74] Gilbert Stengle. A Nullstellensatz and a Positivstellensatz in semialgebraic geometry. *Math. Ann.*, 207:87–97, 1974. doi:10.1007/BF01362149. 11, 61

[Str69] Volker Strassen. Gaussian elimination is not optimal. *Numer. Math.*, 13(4):354–356, 1969. doi:10.1007/BF02165411. 7, 28, 48, 49, 58

[Str72] Volker Strassen. *Evaluation of Rational Functions*, pages 1–10. Springer US, Boston, MA, 1972. doi:10.1007/978-1-4684-2001-2_1. 28

[Str83] Volker Strassen. Rank and optimal computation of generic tensors. *Linear algebra and its applications*, 52:645–685, 1983. doi:10.1016/0024-3795(83)80041-X. 122

[Str86] Volker Strassen. The asymptotic spectrum of tensors and the exponent of matrix multiplication. In *Proceedings of the 27th Annual Symposium on Foundations of Computer Science (FOCS 1986)*, pages 49–54, 1986. doi:10.1109/SFCS.1986.52. 4, 61

[Str87] Volker Strassen. Relative bilinear complexity and matrix multiplication. *J. Reine Angew. Math.*, 375/376:406–443, 1987. doi:10.1515/crll.1987.375-376.406. 4, 58, 61, 109

[Str88] Volker Strassen. The asymptotic spectrum of tensors. *J. Reine Angew. Math.*, 384:102–152, 1988. doi:10.1515/crll.1988.384.102. 4, 27, 29, 38, 52, 60, 61, 74, 80, 89, 90, 92, 100, 121, 122

[Str91] Volker Strassen. Degeneration and complexity of bilinear maps: some asymptotic spectra. *J. Reine Angew. Math.*, 413:127–180, 1991. doi:10.1515/crll.1991.413.127. 4, 44, 60, 61, 93, 94

[Str05] Volker Strassen. Komplexität und Geometrie bilinearer Abbildungen. *Jahresber. Deutsch. Math.-Verein.*, 107(1):3–31, 2005. 4

[Syl52] J. J. Sylvester. On the principles of the calculus of forms. *Cambridge and Dublin Mathematical Journal*, page 52–97, 1852. 55

[Tao08] Terence Tao. *Structure and randomness: pages from year one of a mathematical blog.* American Mathematical Soc., 2008. 30

[Tao16] Terence Tao. A symmetric formulation of the Croot–Lev–Pach–Ellenberg–Gijswijt capset bound. https://terrytao.wordpress.com, 2016. 27, 60, 109

[Tob91] Verena Tobler. *Spezialisierung und Degeneration von Tensoren.* PhD thesis, Universität Konstanz, 1991. http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-20324. 4

[Val79] Leslie G. Valiant. Completeness classes in algebra. In *Proceedings of the 11h Annual ACM Symposium on Theory of Computing (STOC 1979)*, pages 249–261, 1979. doi:10.1145/800135.804419. 28

[vNM07] John von Neumann and Oskar Morgenstern. *Theory of games and economic behavior.* Princeton University Press, anniversary edition, 2007. 62

[Vra21] Péter Vrana. Probabilistic refinement of the asymptotic spectrum of graphs. *Combinatorica*, 41(6):873–904, 2021. arXiv:1903.01857, doi:10.1007/s00493-020-4324-5. 45, 88, 95, 122

[Vra22] Péter Vrana. A generalization of Strassen's theorem on preordered semirings. *Order*, 39(2):209–228, 2022. doi:10.1007/s11083-021-09570-7. 62

[VW07] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC 2007)*, pages 141–154. IEEE, 2007. doi:10.1109/CCC.2007.15. 6

[Wal02] Nolan R. Wallach. An unentangled Gleason's theorem. In *Quantum computation and information*, volume 305 of *Contemp. Math.*, pages 291–298. American Mathematical Society, 2002. doi:10.1090/conm/305/05226. 108, 109

[Wro20] Marcin Wrochna. Smaller counterexamples to hedetniemi's conjecture, 2020. arXiv:2012.13558. 52

[Zhu19] Xuding Zhu. On hedetniemi's conjecture and the Poljak-Rodl function, 2019. arXiv:1911.12015. 52

[Zip79] Richard Zippel. *Probabilistic algorithms for sparse polynomials.* Springer, 1979. doi:10.1007/3-540-09519-5_73. 101

[Zui18] Jeroen Zuiddam. *Algebraic complexity, asymptotic spectra and entanglement polytopes.* PhD thesis, University of Amsterdam, 2018. 11, 15, 61, 94, 95

[Zui19] Jeroen Zuiddam. The asymptotic spectrum of graphs and the Shannon capacity. *Combinatorica*, 39(5):1173–1184, 2019. arXiv:1807.00169, doi:10.1007/s00493-019-3992-5. 45, 62