Multiplying Matrices Faster Than Coppersmith-Winograd

[Extended Abstract]

Virginia Vassilevska Williams
University of California, Berkeley, and
Stanford University
virgi@eecs.berkeley.edu

ABSTRACT

We develop an automated approach for designing matrix multiplication algorithms based on constructions similar to the Coppersmith-Winograd construction. Using this approach we obtain a new improved bound on the matrix multiplication exponent $\omega < 2.3727$.

Categories and Subject Descriptors

F.2.2 [**ANALYSIS OF ALGORITHMS AND PROBLEM COMPLEXITY**]: Nonnumerical Algorithms and Problems

General Terms

Algorithms, Theory

Keywords

matrix multiplication

1. INTRODUCTION

The product of two matrices is one of the most basic operations in mathematics and computer science. Many other essential matrix operations can be efficiently reduced to it, such as Gaussian elimination, LUP decomposition, the determinant or the inverse of a matrix [1]. Matrix multiplication is also used as a subroutine in many computational problems that, on the face of it, have nothing to do with matrices. As a small sample illustrating the variety of applications, there are faster algorithms relying on matrix multiplication for graph transitive closure (see e.g. [1]), context free grammar parsing [21], and even learning juntas [13].

Until the late 1960s it was believed that computing the product C of two $n \times n$ matrices requires essentially a cubic number of operations, as the fastest algorithm known was the naive algorithm which indeed runs in $O(n^3)$ time. In 1969, Strassen [19] excited the research community by giving the first subcubic time algorithm for matrix multiplication, running in $O(n^{2.808})$ time. This amazing discovery spawned a long line of research which gradually reduced the matrix multiplication exponent ω over time. In 1978, Pan [14] showed $\omega < 2.796$. The following year, Bini et al. [4] introduced the notion of *border rank* and obtained $\omega < 2.78$. Schönhage [17] generalized this notion in 1981, proved his τ-theorem (also called the asymptotic sum inequality), and showed that $\omega < 2.548$. In the same paper, combining his work with ideas by Pan, he also showed $\omega < 2.522$. The following year, Romani [15] found that $\omega < 2.517$. The first result to break 2.5 was by Coppersmith and Winograd [9] who obtained $\omega < 2.496$. In 1986, Strassen introduced his *laser* method which allowed for an entirely new attack on the matrix multiplication problem. He also decreased the bound to $\omega < 2.479$. Three years later, Coppersmith and Winograd [10] combined Strassen's technique with a novel form of analysis based on large sets avoiding arithmetic progressions and obtained the famous bound of $\omega < 2.376$ which has remained unchanged for more than twenty years.

In 2003, Cohn and Umans [8] introduced a new, group-theoretic framework for designing and analyzing matrix multiplication algorithms. In 2005, together with Kleinberg and Szegedy [7], they obtained several novel matrix multiplication algorithms using the new framework, however they were not able to beat 2.376.

Many researchers believe that the true value of ω is 2. In fact, both Coppersmith and Winograd [10] and Cohn et al. [7] presented conjectures which if true would imply $\omega = 2$. Recently, Alon, Shpilka and Umans [2] showed that both the Coppersmith-Winograd conjecture and one of the Cohn et al. [7] conjectures contradict a variant of the widely believed *sunflower* conjecture of Erdös and Rado [11]. Nevertheless, it could be that at least the remaining Cohn et al. conjecture could lead to a proof that $\omega = 2$.

The Coppersmith-Winograd Algorithm.

In this paper we revisit the Coppersmith-Winograd (CW) approach [10]. We give a very brief summary of the approach here; we will give a more detailed account in later sections.

One first constructs an algorithm A which given Q-length vectors x and y for constant Q, computes Q values of the form $z_k = \sum_{i,j} t_{ijk} x_i y_j$, say with $t_{ijk} \in \{0, 1\}$, using a smaller number of products than would naively be necessary. The values z_k do not necessarily have to correspond to entries from a matrix product. Then, one considers the algorithm A^n obtained by applying A to vectors x, y of length Q^n, recursively n times as follows. One splits x and y into

STOC'12, May 19–22, 2012, New York, New York, USA.

Q subvectors of length Q^{n-1}. Then one runs A on x and y treating them as vectors of length Q with entries that are vectors of length Q^{n-1}. When the product of two entries is needed, one uses A^{n-1} to compute it. This algorithm A^n is called the *nth tensor power* of A. Its running time is essentially $O(r^n)$ if r is the number of multiplications performed by A.

The goal of the approach is to show that for very large n one can set enough variables x_i, y_j, z_k to 0 so that running A^n on the resulting vectors x and y actually computes a matrix product. That is, as n grows, some subvectors x' of x and y' of y can be thought to represent square matrices and when A^n is run on x and y, a subvector of z is actually the matrix product of x' and y'.

If A^n can be used to multiply $m \times m$ matrices in $O(r^n)$ time, then this implies that $\omega \leq \log_m r^n$, so that the larger m is, the better the bound on ω.

Coppersmith and Winograd [10] introduced techniques which, when combined with previous techniques by Schönhage [17] and Strassen [20], allowed them to effectively choose which variables to set to 0 so that one can compute very large matrix products using A^n. Part of their techniques rely on partitioning the index triples $i, j, k \in [Q]^n$ into groups and analyzing how "similar" each group g computation $\{z_{kg} = \sum_{i,j:\ (i,j,k)\in g} t_{ijk} x_i y_j\}_k$ is to a matrix product. The similarity measure used is called the *value* of the group.

Depending on the underlying algorithm A, the partitioning into groups varies and can affect the final bound on ω. Coppersmith and Winograd analyzed a particular algorithm A which resulted in $\omega < 2.388$. Then they noticed that if one uses A^2 as the basic algorithm (the "base case") instead, one can obtain the better bound $\omega < 2.376$. They left as an open problem what happens if one uses A^3 as the basic algorithm instead.

Our contribution.

We give a new way to more tightly analyze the techniques behind the Coppersmith-Winograd (CW) approach [10]. We demonstrate the effectiveness of our new analysis by showing that the 8th tensor power of the CW algorithm [10] in fact gives $\omega < 2.3727$. (It is likely that higher tensor powers can give tighter estimates, and this could be the subject of future work.)

There are two main theorems behind our approach. The first theorem takes any tensor power A^n of a basic algorithm A, picks a particular group partitioning for A^n and derives an efficient procedure computing formulas for the values of these groups. The second theorem assumes that one knows the values for A^n and derives an efficient procedure which outputs a (nonlinear) constraint program on $O(n^2)$ variables, the solution of which gives a bound on ω.

We then apply the procedures given by the theorems to the second, fourth and eighth tensor powers of the Coppersmith-Winograd algorithm, obtaining improved bounds with each new tensor power.

Similar to [10], our proofs apply to any starting algorithm that satisfies a simple uniformity requirement which we formalize later. The upshot of our approach is that now any such algorithm and its higher tensor powers can be analyzed entirely by computer. (In fact, our analysis of the 8th tensor power of the CW algorithm is done this way.) The burden is now entirely offloaded to constructing base algorithms satisfying the requirement. We hope that some of the new group-theoretic techniques can help in this regard.

Why wasn't an improvement on CW found in the 1990s?.

After all, the CW paper explicitly posed the analysis of the third tensor power as an open problem.

The answer to this question is twofold. Firstly, several people have attempted to analyze the third tensor power (from personal communication with Umans, Kleinberg and Coppersmith). As the author found out from personal experience, analyzing the third tensor power reveals to be very disappointing. In fact no improvement whatsoever can be found. This finding led some to believe that 2.376 may be the final answer, at least for the CW algorithm.

The second issue is that with each new tensor power, the number of new values that need to be analyzed grows quadratically. For the eighth tensor power for instance, 30 separate analyses are required! Prior to our work, each of these analyses would require a separate application of the CW techniques. It would have required an enormous amount of patience to analyze larger tensor powers, and since the third tensor power does not give any improvement, the prospects looked bleak.

Stothers' work.

We were recently made aware of the thesis work of A. Stothers [18] in which he claims an improvement to ω. Stothers argues that $\omega < 2.3737$ by analyzing the 4th tensor power of the Coppersmith-Winograd construction. Before learning of Stothers' result, we were only able to prove a bound of the form $\omega < 2.375$, as our analysis of some of the "values" was not tight. After seeing a shortcut that Stothers employs in his analysis of the values for the 4th tensor power (we will point out this shortcut in the main text), we were able to tighten the analysis of the values, fully automate our approach, and improve our bound to $\omega < 2.3727$.

There are several differences between our approach and Stothers'. The first is relatively minor: the CW approach requires the use of some hash functions; ours are different and simpler than Stothers'. The main difference is that because of the generality of our analysis, we do not need to fully analyze all groups of each tensor power construction. Instead we can just apply our formulas in a mechanical way. Stothers, on the other hand, did a completely separate analysis of each group.

Finally, Stothers' approach only works for tensor powers up to 4. Starting with the 5-th tensor power, the values of some of the groups begin to depend on more variables and a more careful analysis is needed.

(Incidentally, we also obtain a better bound from the 4th tensor power, $\omega < 2.37293$, however this is an artifact of our optimization software, as we end up solving essentially the same constraint program.)

Preliminaries.

We use the following notation: $[n] := \{1, \ldots, n\}$, and $\binom{N}{[a_i]_{i\in[k]}} := \binom{N}{a_1,\ldots,a_k}$.

We define $\omega \geq 2$ to be the infimum over the set of all reals r such that $n \times n$ matrix multiplication over $\mathbb{Q}$ can be computed in n^r additions and multiplications for some

natural number n. (However, the CW approach and our extensions work over any ring.)

A *three-term arithmetic progression* is a sequence of three integers $a \le b \le c$ so that $b - a = c - b$, or equivalently, $a + c = 2b$. An arithmetic progression is nontrivial if $a < b < c$.

The following is a theorem by Behrend [3] improving on Salem and Spencer [16]. The subset A computed by the theorem is called a *Salem-Spencer* set.

THEOREM 1. *There exists an absolute constant c such that for every $N \ge \exp(c^2)$, one can construct in poly(N) time a subset $A \subset [N]$ with no three-term arithmetic progressions and $|A| > N\exp(-c\sqrt{\log N})$.*

The following lemma is needed in our analysis. Incidentally, a similar lemma appears in [7] in a slightly different context.

LEMMA 1. *Let k be a constant and N be sufficiently large. Let B_i be fixed for $i \in [k]$. Let a_i for $i \in [k]$ be variables such that $a_i \ge 0$ and $\sum_i a_i = 1$. Then the quantity*

$$\binom{N}{[a_i N]_{i\in[k]}} \prod_{i=1}^{k} B_i^{a_i N}$$

is maximized for the choices $a_i = B_i / \sum_{j=1}^{k} B_j$ for all $i \in [k]$ and for these choices it is at least

$$\left(\sum_{j=1}^{k} B_j\right)^N \Big/ \ (N+1)^k.$$

PROOF. We will prove the lemma by induction on k. Suppose that $k = 2$ and consider

$$\binom{N}{aN} x^{aN} y^{N(1-a)} = y^N \binom{N}{aN} (x/y)^{aN},$$

where $x \le y$.

When $(x/y) \le 1$, the function $f(a) = \binom{N}{aN}(x/y)^{aN}$ of a is concave for $a \le 1/2$. Hence its maximum is achieved when $\partial f(a)/\partial a = 0$. Consider $f(a)$: it is $N!/((aN)!(N(1-a))!)(x/y)^{aN}$. We can take the logarithm to obtain $\ln f(a) = \ln(N!) + Na\ln(x/y) - \ln(aN!) - \ln((N(1-a))!)$. $f(a)$ grows exactly when $a\ln(x/y) - \ln(aN!)/N - \ln(N(1-a))!/N$ does. Taking Stirling's approximation, we obtain

$$a\ln(x/y) - \ln(aN!)/N - \ln(N(1-a))!/N =$$

$$a\ln(x/y) - a\ln(a) - (1-a)\ln(1-a) - \ln N - O((\log N)/N).$$

Since N is large, the $O((\log N)/N)$ term is negligible. Thus we are interested in when $g(a) = a\ln(x/y) - a\ln(a) - (1-a)\ln(1-a)$ is maximized. Because of concavity, for $a \le 1/2$ and $x \le y$, the function is maximized when $\partial g(a)/\partial a = 0$, i.e. when

$$0 = \ln(x/y) - \ln(a) - 1 + \ln(1-a) + 1 = \ln(x/y) - \ln(a/(1-a)).$$

Hence $a/(1-a) = x/y$ and so $a = x/(x+y)$.

Furthermore, since the maximum is attained for this value of a, we get that for each $t \in \{0, \ldots, N\}$ we have that $\binom{N}{t} x^t y^{N-t} \le \binom{N}{aN} x^{aN} y^{N(1-a)}$, and since $\sum_{t=0}^{N} \binom{N}{t} x^t y^{N-t} = (x+y)^N$, we obtain that for $a = x/(x+y)$,

$$\binom{N}{aN} x^{aN} y^{N(1-a)} \ge (x+y)^N/(N+1).$$

Now let's consider the case $k > 2$. First assume that the B_i are sorted so that $B_i \le B_{i+1}$. Since $\sum_i a_i = 1$, we obtain

$$\binom{N}{[a_i]_{i\in[k]}} \prod_{i=1}^{k} B_i = \left(\sum_i B_i\right)^N \binom{N}{[a_i]_{i\in[k]}} \prod_{i=1}^{k} b_i,$$

where $b_i = B_i / \sum_j B_j$. We will prove the claim for $\binom{N}{[a_i]_{i\in[k]}} \prod_{i=1}^{k} b_i$, and the lemma will follow for the B_i as well. Hence we can assume that $\sum_i b_i = 1$.

Suppose that we have proven the claim for $k-1$. This means that in particular

$$\binom{N - a_1 N}{[a_j N]_{j\ge 2}} \prod_{j=2}^{k} b_j^k \ge \left(\sum_{j=2}^{k} b_j\right)^{N-a_1 N} /(N+1)^{k-1},$$

and the quantity is maximized for $a_j N/(N - a_1 N) = b_j / \sum_{j\ge 2} b_j$ for all $j \ge 2$.

Now consider $\binom{N}{a_1 N} b_1^{a_1 N} \left(\sum_{j=2}^{k} b_j\right)^{N-a_1 N}$. By our base case we get that this is maximized and is at least $(\sum_{j=1}^{k} b_j)^N/N$ for the setting $a_1 = b_1$. Hence, we will get

$$\binom{N}{[a_j N]_{j\in[k]}} \prod_{j=1}^{k} b_j^k \ge \left(\sum_{j=1}^{k} b_j\right)^N /(N+1)^k,$$

for the setting $a_1 = b_1$ and for $j \ge 2$, $a_j N/(N - a_1 N) = b_j / \sum_{j\ge 2} b_j$ implies $a_j/(1-b_1) = b_j/(1-b_1)$ and hence $a_j = b_j$. We have proven the lemma. □

1.1 A brief summary of the techniques used in bilinear matrix multiplication algorithms

A full exposition of the techniques can be found in the book by Bürgisser, Clausen and Shokrollahi [6]. The lecture notes by Bläser [5] are also a nice read.

Bilinear algorithms and trilinear forms.

Matrix multiplication is an example of a trilinear form. $n \times n$ matrix multiplication, for instance, can be written as

$$\sum_{i,j\in[n]} \sum_{k\in n} x_{ik} y_{kj} z_{ij},$$

which corresponds to the equalities $z_{ij} = \sum_{k\in n} x_{ik} y_{kj}$ for all $i, j \in [n]$. In general, a trilinear form has the form $\sum_{i,j,k} t_{ijk} x_i y_j z_k$ where i, j, k are indices in some range and t_{ijk} are the coefficients which define the trilinear form; t_{ijk} is also called a tensor. The trilinear form for the product of a $k \times m$ by an $m \times n$ matrix is denoted by $\langle k, m, n\rangle$.

Strassen's algorithm for matrix multiplication is an example of a bilinear algorithm which computes a trilinear form. A bilinear algorithm is equivalent to a representation of a trilinear form of the following form:

$$\sum_{i,j,k} t_{ijk} x_i y_j z_k = \sum_{\lambda=1}^{r} \left(\sum_i \alpha_{\lambda,i} x_i\right) \left(\sum_j \beta_{\lambda,j} y_j\right) \left(\sum_k \gamma_{\lambda,k} z_k\right).$$

Given the above representation, the algorithm is then to first compute the r products $P_\lambda = (\sum_i \alpha_{\lambda,i} x_i)(\sum_j \beta_{\lambda,j} y_j)$ and then for each k to compute $z_k = \sum_\lambda \gamma_{\lambda,k} P_\lambda$.

The minimum number of products r in a bilinear construction is called the rank of the trilinear form (or its tensor). It is known that the rank of 2×2 matrix multiplication is 7, and hence Strassen's bilinear algorithm is optimal for the

product of 2×2 matrices. A basic property of the rank R of matrix multiplication is that $R(\langle k,m,n\rangle) = R(\langle k,n,m\rangle) = R(\langle m,k,n\rangle) = R(\langle m,n,k\rangle) = R(\langle n,m,k\rangle) = R(\langle n,k,m\rangle)$. This property holds in fact for any tensor and the tensors obtained by permuting the roles of the x, y and z variables.

Any algorithm for $n \times n$ matrix multiplication can be applied recursively k times to obtain a bilinear algorithm for $n^k \times n^k$ matrices, for any integer k. Furthermore, one can obtain a bilinear algorithm for $\langle k_1k_2, m_1m_2, n_1n_2\rangle$ by splitting the $k_1k_2 \times m_1m_2$ matrix into blocks of size $k_1 \times m_1$ and the $m_1m_2 \times n_1n_2$ matrix into blocks of size $m_1 \times n_1$. The one can apply a bilinear algorithm for $\langle k_2, m_2, n_2\rangle$ on the matrix with block entries, and an algorithm for $\langle k_1, m_1, n_1\rangle$ to multiply the blocks. This composition multiplies the ranks and hence $R(\langle k_1k_2, m_1m_2, n_1n_2\rangle) \leq R(\langle k_1, m_1, n_1\rangle) \cdot R(\langle k_2, m_2, n_2\rangle)$. Because of this, $R(\langle 2^k, 2^k, 2^k\rangle) \leq (R(\langle 2,2,2\rangle))^k = 7^k$ and if $N = 2^k$, $R(\langle N,N,N\rangle) \leq 7^{\log_2 N} = N^{\log_2 7}$. Hence, $\omega \leq \log_N R(\langle N,N,N\rangle)$.

In general, if one has a bound $R(\langle k,m,n\rangle) \leq r$, then one can symmetrize and obtain a bound on $R(\langle kmn, kmn, kmn\rangle) \leq r^3$, and hence $\omega \leq 3\log_{kmn} r$.

The above composition of two matrix product trilinear forms to form a new trilinear form is called the *tensor product* $t_1 \otimes t_2$ of the two forms t_1, t_2. For two generic trilinear forms $\sum_{i,j,k} t_{ijk}x_iy_jz_k$ and $\sum_{i',j',k'} t'_{ijk}x_{i'}y_{j'}z_{k'}$, their tensor product is the trilinear form

$$\sum_{(i,i'),(j,j'),(k,k')} (t_{ijk}t'_{i'j'k'})x_{(i,i')}y_{(j,j')}z_{(k,k')},$$

i.e. the new form has variables that are indexed by pairs if indices, and the coordinate tensors are multiplied.

The *direct sum* $t_1 \oplus t_2$ of two trilinear forms t_1, t_2 is just their sum, but where the variable sets that they use are disjoint. That is, the direct sum of $\sum_{i,j,k} t_{ijk}x_iy_jz_k$ and $\sum_{i,j,k} t'_{ijk}x_iy_jz_k$ is a new trilinear form with the set of variables $\{x_{i0}, x_{i1}, y_{j0}, y_{j1}, z_{k0}, z_{k1}\}_{i,j,k}$:

$$\sum_{i,j,k} t_{ijk}x_{i0}y_{j0}z_{k0} + t'_{ijk}x_{i1}y_{j1}z_{k1}.$$

A lot of interesting work ensued after Strassen's discovery. Bini et al. [4] showed that one can extend the form of a bilinear construction to allow the coefficients $\alpha_{\lambda,i}, \beta_{\lambda,j}$ and $\gamma_{\lambda,k}$ to be linear functions of the integer powers of an indeterminate, ϵ. Specifically, one considers "approximate" bilinear algorithms of the form:

$$\sum_{\lambda=1}^{r} \left(\sum_i \alpha_{\lambda,i}x_i\right)\left(\sum_j \beta_{\lambda,j}y_j\right)\left(\sum_k \gamma_{\lambda,k}z_k\right) =$$

$$\sum_{i,j,k} t_{ijk}x_iy_jz_k + O(\epsilon),$$

where the $O(\epsilon)$ term hides triples which have coefficients that depend on positive powers of ϵ.

As an example, Bini et al. gave the following construction for three entries of the product of 2×2 matrices in terms of 5 bilinear products:

$$(x_{11}y_{11} + x_{12}y_{21})z_{11} + (x_{11}y_{12} + x_{12}y_{22})z_{12} +$$

$$(x_{21}y_{11} + x_{22}y_{21})z_{21} + O(\epsilon) =$$

$$(x_{12}+\epsilon x_{22})y_{21}(z_{11}+\epsilon^{-1}z_{21}) + x_{11}(y_{11}+\epsilon y_{12})(z_{11}+\epsilon^{-1}z_{12}) +$$

$$x_{12}(y_{11}+y_{21}+\epsilon y_{22})(-\epsilon^{-1}z_{21}) + (x_{11}+x_{12}+\epsilon x_{21})y_{11}(-\epsilon^{-1}z_{12}) +$$

$$(x_{12} + \epsilon x_{21})(y_{11} + \epsilon y_{22})(\epsilon^{-1}z_{12} + \epsilon^{-1}z_{21}).$$

The minimum number of products of such a bilinear algorithm is called the *border rank* $\tilde{R}$ of a trilinear form (or its tensor). Border rank is a stronger notion of rank and it allows for better bounds on ω. Most of the properties of rank also extend to border rank, so that if $\tilde{R}(\langle k,m,n\rangle) \leq r$, then $\omega \leq 3\cdot\log_{kmn} r$. For instance, Bini et al. used their construction above to obtain a border rank of 10 for the product of a 2×2 by a 2×3 matrix and, by symmetry, an upper bound of the border rank of 10^3 for the product of two 12×12 matrices. This gave the new bound of $\omega \leq 3\log_{12} 10 < 2.78$.

Schönhage [17] generalized Bini et al.'s approach and proved his τ-theorem (also known as the asymptotic sum inequality). Up until his paper, all constructions used in designing matrix multiplication alrgorithms explicitly computed a single matrix product trilinear form. Schönhage's theorem allowed a whole new family of contructions. In particular, he showed that constructions that are direct sums of rectangular matrix products suffice to give a bound on ω.

THEOREM 2 (SCHÖNHAGE'S τ-THEOREM). *If for some $r > q$, $\tilde{R}(\bigoplus_{i=1}^{q}\langle k_i, m_i, n_i\rangle) \leq r$, then let τ be defined as $\sum_{i=1}^{q}(k_im_in_i)^{\tau} = r$. Then $\omega \leq 3\tau$.*

2. COPPERSMITH AND WINOGRAD'S ALGORITHM

We recall Coppersmith and Winograd's [10] (CW) construction:

$$\lambda^{-2} \cdot \sum_{i=1}^{q}(x_0 + \lambda x_i)(y_0 + \lambda y_i)(z_0 + \lambda z_i)$$

$$-\lambda^{-3} \cdot (x_0 + \lambda^2 \sum_{i=1}^{q} x_i)(y_0 + \lambda^2 \sum_{i=1}^{q} y_i)(z_0 + \lambda^2 \sum_{i=1}^{q} z_i)$$

$$+(\lambda^{-3} - q\lambda^{-2}) \cdot (x_0 + \lambda^3 x_{q+1})(y_0 + \lambda^3 y_{q+1})(z_0 + \lambda^3 z_{q+1}) =$$

$$\sum_{i=1}^{q}(x_iy_iz_0 + x_iy_0z_i + x_0y_iz_i) + (x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0) + O(\lambda).$$

The construction computes a particular symmetric trilinear form. The indices of the variables are either 0, $q+1$ or some integer in $[q]$. We define

$$p(i) = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } i \in [q] \\ 2 & \text{if } i = q+1 \end{cases}$$

The important property of the CW construction is that for any triple $x_iy_jz_k$ in the trilinear form, $p(i)+p(j)+p(k) = 2$.

In general, the CW approach applies to any construction for which we can define an integer function p on the indices so that there exists a number P so that for every $x_iy_jz_k$ in the trilinear form computed by the construction, $p(i) + p(j) + p(k) = P$. We call such constructions (p, P)-uniform.

DEFINITION 1. *Let p be a function from $[n]$ to $[N]$. Let $P \in [N]$ A trilinear form $\sum_{i,j,k\in[n]} t_{ijk}x_iy_jz_k$ is (p, P)-uniform if whenever $t_{ijk} \neq 0$, $p(i) + p(j) + p(k) = P$. A construction computing a (p, P)-uniform trilinear form is also called (p, P)-uniform.*

Any tensor power of a (p, P)-uniform construction is (p', P') uniform for some p' and P'. There are many ways to define p' and P' in terms of p and P. For the K-th tensor power the variable indices $xindex, yindex, zindex$ are length K sequences of the original indices: $xindex[1], \ldots, xindex[K]$, $yindex[1], \ldots, yindex[K]$ and $zindex[1], \ldots, zindex[K]$. Then, for instance, one can pick p' to be an arbitrary linear combination, so that $p'[xindex] = \sum_i^K a_i \cdot xindex[i]$, and similarly $p'[yindex] = \sum_i^K a_i \cdot yindex[i]$ and $p'[zindex] = \sum_i^K a_i \cdot zindex[i]$. Clearly then $P' = P \sum_i a_i$, and the K-th tensor power construction is (p', P')-uniform.

In this paper we will focus on the case where $a_i = 1$ for all $i \in [K]$, so that $p'[index] = \sum_i^K index[i]$ and $P' = PK$. Similar results can be obtained for other choices of p'.

The CW approach proceeds roughly as follows. Suppose we are given a (p, P)-uniform construction and we wish to derive a bound on ω from it. (The approach only works when the range of p is at least 2.) Let C be the trilinear form computed by the construction and let r be the number of bilinear products performed. If the trilinear form happens to be a direct sum of different matrix products, then one can just apply the Schönhage τ-theorem [17] to obtain a bound on ω in terms of r and the dimensions of the small matrix products. However, typically the triples in the trilinear form C cannot be partitioned into matrix products on disjoint sets of variables.

The first CW idea is to partition the triples of C into groups which look like matrix products but may share variables. Then the idea is to apply procedures to remove the shared variables by carefully setting variables to 0. In the end one obtains a smaller, but not much smaller, number of independent matrix products and can use Schönhage's τ-theorem.

Two procedures are used to remove the shared variables. The first one defines a random hash function h mapping variables to integers so that there is a large set S such that for any triple $x_i y_j z_k$ with $h(x_i), h(y_j), h(z_k) \in S$ one actually has $h(x_i) = h(y_j) = h(z_k)$. Then one can set all variables mapped outside of S to 0 and be guaranteed that the triples are *partitioned* into groups according to what element of S they were mapped to, and moreover, the groups do not share any variables. Since S is large and h maps variables independently, there is a setting of the random bits of h so that a lot of triples (at least the expectation) are mapped into S and are hence preserved by this partitioning step. The construction of S uses the Salem-Spencer theorem and h is a cleverly constructed linear function.

After this first step, the remaining nonzero triples have been partitioned into groups according to what element of S they were mapped to, and the groups do not share any variables. The second step removes shared variables within each group. This is accomplished by a greedy procedure that guarantees that a constant fraction of the triples remain. More details can be found in the next section.

When applied to the CW construction above, the above procedures gave the bound $\omega < 2.388$.

The next idea that Coppersmith and Winograd had was to extend the τ-theorem to Theorem 3 below using the notion of *value* V_τ. The intuition is that V_τ assigns a weight to a trilinear form T according to how "close" an algorithm computing T is to an $O(n^{3\tau})$ matrix product algorithm.

Suppose that for some N, the Nth tensor power of T [1] can be reduced to $\bigoplus_{i=1}^q \langle k_i, m_i, n_i \rangle$ by substitution of variables. Then, as in [10] we introduce the constraint

$$V_\tau(T) \geq \left(\sum_{i=1}^{q} (k_i m_i n_i)^\tau \right)^{1/N}.$$

Furthermore, if π is the cyclic permutation of the x,y and z variables in T, then we also have $V_\tau(T) \geq (V_\tau(T \otimes \pi A \otimes \pi^2 T))^{1/3} \geq (V_\tau(T) V_\tau(\pi T) V_\tau(\pi^2 T))^{1/3}$.

We can give a formal definition of $V_\tau(T)$ as follows. Consider all positive integers N, and all possible ways σ to zero-out variables in the Nth tensor power of T so that one obtains a direct sum of matrix products $\bigoplus_{i=1}^{q(\sigma)} \langle k_i^\sigma, m_i^\sigma, n_i^\sigma \rangle$. Then we can define

$$V_\tau(T) = \limsup_{N \to \infty, \sigma} \left(\sum_{i=1}^{q(\sigma)} (k_i^\sigma m_i^\sigma n_i^\sigma)^\tau \right)^{1/N}.$$

We can argue that for any permutation of the x, y, z variables π, and any N there is a corresponding permutation of the zeroed out variables σ that gives the same (under the permutation π) direct sum of matrix products. Hence $V_\tau(T) \leq V_\tau(\pi T)$ and since T can be replaced with πT and π with π^{-1}, we must have $V_\tau(T) = V_\tau(\pi T)$, thus also satisfying the inequality $V_\tau(T) \geq (V_\tau(T) V_\tau(\pi T) V_\tau(\pi^2 T))^{1/3}$.

It is clear that values are superadditive and supermultiplicative, so that $V_\tau(T_1 \otimes T_2) \geq V_\tau(T_1) V_\tau(T_2)$ and $V_\tau(T_1 \oplus T_2) \geq V_\tau(T_1) + V_\tau(T_2)$.

With this notion of value as a function of τ, we can state an extended τ-theorem, implicit in [10].

THEOREM 3 ([10]). *Let T be a trilinear form such that $T = \bigoplus_{i=1}^q T_i$ and the T_i are independent copies of the same trilinear form T'. If there is an algorithm that computes T by performing at most r multiplications for $r > q$, then $\omega \leq 3\tau$ for τ given by $qV_\tau(T') = r$.*

Theorem 3 has the following effect on the CW approach. Instead of partitioning the trilinear form into matrix product pieces, one could partition it into different types of pieces, provided that their value is easy to analyze. A natural way to partition the trilinear form C is to group all triples $x_i y_j z_k$ for which (i, j, k) are mapped by p to the same integer 3-tuple $(p(i), p(j), p(k))$. This partitioning is particularly good for the CW construction and its tensor powers: in the full version we show for instance that the trilinear form which consists of the triples mapped to $(0, J, K)$ for any J, K is always a matrix product of the form $\langle 1, Q, 1 \rangle$ for some Q.

Using this extra ingredient, Coppersmith and Winograd were able to analyze the second tensor power of their construction and to improve the estimate to the current best bound $\omega < 2.376$.

In the following section we show how with a few extra ingredients one can algorithmically analyze an arbitrary tensor power of any (p, P)-uniform construction. (Amusingly, the algorithms involve the solution of linear systems, indicating that faster matrix multiplication algorithms can help improve the search for faster matrix multiplication algorithms.)

[1] Tensor powers of trilinear forms can be defined analogously to how we defined tensor powers of an algorithm computing them.

1. For each $I, J, K = P\mathcal{K} - I - J$, determine the value V_{IJK} of the trilinear form $\sum_{i,j:\ p(i)=I, p(j)=J} t_{ijk} x_i y_j z_k$, as a nondecreasing function of τ.

2. Define variables a_{IJK} and $\bar{a}_{IJK}$ for $I \le J \le K = P\mathcal{K} - I - J$.

3. Form the linear system: for all I, $A_I = \sum_J \bar{a}_{IJK}$, where $\bar{a}_{IJK} = \bar{a}_{sort(IJK)}$.

4. Determine the rank of the linear system, and if necessary, pick enough variables $\bar{a}_{IJK}$ to place in S and treat as constants, so the system has full rank.

5. Solve for the variables outside of S in terms of the A_I and the variables in S.

6. Compute the derivatives $p_{I'J'K'IJK}$.

7. Form and solve the program below to obtain $\omega \le 3\tau$:

Minimize τ subject to

$$r^{\mathcal{K}} = \prod_{I \le J \le K} \left(\frac{\bar{a}_{IJK}}{a_{IJK}} \right)^{perm(IJK)} \frac{V_{IJK}^{perm(IJK) \cdot a_{IJK}}}{\prod_I A_I^{A_I}},$$

$\bar{a}_{IJK} \ge 0, a_{IJK} \ge 0$ for all I, J, K,

$\sum_{I \le J \le K} perm(IJK) \cdot \bar{a}_{IJK} = 1$,

for all $\bar{a}_{IJK} \in S$:

$$\bar{a}_{IJK} \cdot \prod_{\bar{a}_{I'J'K'} \notin S, p_{I'J'K'IJK} > 0} (\bar{a}_{I'J'K'})^{p_{I'J'K'IJK}} = \prod_{\bar{a}_{I'J'K'} \notin S, p_{I'J'K'IJK} < 0} (\bar{a}_{I'J'K'})^{-p_{I'J'K'IJK}},$$

and unless one is setting $a_{IJK} = \bar{a}_{IJK}$,

$\sum_J a_{IJK} = \sum_J \bar{a}_{IJK}$ for all I.

Figure 1: The procedure to analyze the $\mathcal{K}$ tensor power.

3. ANALYZING ARBITRARY TENSOR POWERS OF UNIFORM CONSTRUCTIONS

Let $\mathcal{K} \ge 2$ be an integer. Let p be an integer function with range size at least 2. We will show how to analyze the $\mathcal{K}$-tensor power of any (p, P)-uniform construction by proving the following theorem:

THEOREM 4. *Given a (p, P)-uniform construction and lower bounds for the values for its $\mathcal{K}$-tensor power, the procedure in Figure 1 outputs a constraint program the solution τ of which implies $\omega \le 3\tau$.*

We begin the proof. Consider the the $\mathcal{K}$-tensor power of a particular (p, P)-uniform construction. Call the trilinear form computed by the construction C. Let r be the bound on the (border) rank of the original construction. Then $r^{\mathcal{K}}$ is a bound on the (border) rank of C.

The variables in C have indices which are $\mathcal{K}$-length sequences of the original indices. Let the index of an x variable be denoted by *xindex*, the index of a y variable by *yindex* and the index of a z variable by *zindex*. Then, for every triple $x_{xindex} y_{yindex} z_{zindex}$ in the trilinear form and any particular position *pos* in the index sequences, $p(xindex[pos]) + p(yindex[pos]) + p(zindex[pos]) = P$. Recall that we defined the extension $\bar{p}$ of p for the $\mathcal{K}$ tensor power as $\bar{p}(index) = \sum_{i=1}^{K} p(index[i])$, and that the $\mathcal{K}$ tensor power is $(\bar{p}, P\mathcal{K})$-uniform.

Now, we can represent C as a sum of trilinear forms $X^I Y^J Z^K$, where $X^I Y^J Z^K$ only contains those triples in C, $x_{xindex} y_{yindex} z_{zindex}$ for which $\bar{p}$ maps *xindex* to I, *yindex* to J and *zindex* to K. That is, if $C = \sum_{ijk} t_{ijk} x_i y_j z_k$, then $X^I Y^J Z^K = \sum_{i,j,k:\ \bar{p}(i)=I, \bar{p}(j)=J} t_{ijk} x_i y_j z_k$. We refer to I,J,K as *blocks*.

Following the CW analysis, we will later compute the value V_{IJK} (as a function of τ) for each trilinear form $X^I Y^J Z^K$ separately. The trilinear forms $X^I Y^J Z^K$ can share variables. For instance, $X^I Y^J Z^K$ and $X^I Y^{J'} Z^{K'}$ share the x variables mapped to block I. We use the CW tools to zero-out some variables until the remaining trilinear forms no longer share variables, and moreover a nontrivial number of the same trilinear form remain so that one can obtain a good estimate on τ and hence ω. We outline the approach in what follows.

Take the N-th tensor power C^N of C for large N; we will eventually let N go to ∞. Now the indices of the variables of C are N-length sequences of $\mathcal{K}$ length sequences. The blocks of C^N are N-length sequences of blocks of C.

We will pick (rational) values $A_I \in [0, 1]$ for every block I of C, so that $\sum_I A^I = 1$. Then we will set to zero all x, y, z variables of C^N the indices of which map to blocks which **do** not have exactly $N \cdot A_I$ positions of block I for every I. (For large enough N, $N \cdot A_I$ is an integer.)

For each triple of blocks of C^N $(\bar{I}, \bar{J}, \bar{K})$ we will consider the trilinear subform of C^N, $X^{\bar{I}} Y^{\bar{J}} Z^{\bar{K}}$, where as before C^N is the sum of these trilinear forms.

Consider values a_{IJK} for all valid block triples I, J, K of C which satisfy

$$A_I = \sum_J a_{IJ(P\cdot\mathcal{K}-I-J)} = \sum_J a_{JI(P\cdot\mathcal{K}-I-J)} = \sum_J a_{(P\cdot\mathcal{K}-I-J)JI}.$$

The values a_{IJK} will correspond to the number of index positions *pos* such that any trilinear form $X^{\bar{I}} Y^{\bar{J}} Z^{\bar{K}}$ of C^N we have that $\bar{I}[pos] = I, \bar{J}[pos] = J, \bar{K}[pos] = K$.

The a_{IJK} need to satisfy the following additional two constraints:

$$1 = \sum_I A_I = \sum_{I,J,K} a_{IJK},$$

and

$$P\mathcal{K} = 3 \sum_I I \cdot A_I.$$

We note that although the second constraint is explicitly stated in [10], it actually automatically holds as a consequence of constraint 1 and the definition of a_{IJK} since

$$3 \sum_I I A_I = \sum_I I A_I + \sum_J J A_J + \sum_K K A_K =$$

$$\sum_{I,J} I \cdot a_{IJ(P\mathcal{K}-I-J)} + \sum_{J,I} J \cdot a_{IJ(P\mathcal{K}-I-J)} + \sum_{K,J} K \cdot a_{(P\mathcal{K}-J-K),J,K} =$$

$$\sum_{I,J} (I + J + (P\mathcal{K} - I - J)) \cdot a_{IJ(P\mathcal{K}-I-J)} =$$

$$P\mathcal{K} \sum_{I,J} a_{IJ(P\mathcal{K}-I-J)} = P\mathcal{K}.$$

Thus the only constraint that needs to be satisfied by the a_{IJK} is $\sum_{I,J,K} a_{IJK} = 1$.

Recall that $\binom{N}{[R_i]_{i\in S}}$ denotes $\binom{N}{R_{i_1},R_{i_2},\ldots,R_{i_{|S|}}}$ where the indices $i_1,\ldots,i_{|S|}$ are the elements of S. When S is implicit, we only write $\binom{N}{[R_i]}$.

By our choice of which variables to set to 0, we get that the number of C^N block triples which still have nonzero trilinear forms is

$$\binom{N}{[N\cdot A_I]}\cdot\left(\sum_{[a_{IJK}]}\prod_I\binom{N\cdot A_I}{[N\cdot a_{IJK}]_J}\right),$$

where the sum ranges over the values a_{IJK} which satisfy the above constraint. This is since the number of nonzero blocks is $\binom{N}{[N\cdot A_I]}$ and the number of block triples which contain a particular X block is exactly $\prod_I\binom{N\cdot A_I}{[N\cdot a_{IJK}]_J}$ for every partition of A_I into $[a_{IJK}]_J$ (for $K = P\mathcal{K} - I - J$).

Let $\aleph = \sum_{[a_{IJK}]}\prod_I\binom{N\cdot A_I}{[N\cdot a_{IJK}]_J}$. The current number of nonzero block triples is $\aleph\cdot\binom{N}{[N\cdot A_I]}$.

Our goal will be to process the remaining nonzero triples by zeroing out variables sharing the same block until the remaining trilinear forms corresponding to different block triples do not share variables. Furthermore, we would like for the remaining nonzero trilinear forms to be essentially the same, so that we can use Theorem 3.

This would be the case, if we fix for each I a partition $[a_{IJK}N]_J$ of A_IN: Suppose that each remaining triple $X^{\bar I}Y^{\bar J}Z^{\bar K}$ has exactly $a_{IJK}N$ positions pos such that $\bar I[pos] = I, \bar J[pos] = J, \bar K[pos] = K$. Then all remaining triples would be isomorphic, and each would have value at least $\prod_{I,J}V_{IJK}^{a_{IJK}N}$ by supermultiplicativity.

Suppose that we have fixed a particular choice of the a_{IJK}. We will later show how to pick a choice which maximizes our bound on ω.

The number of small trilinear forms (corresponding to different block triples of C^N) is $\aleph'\cdot\binom{N}{[N\cdot A_I]}$, where

$$\aleph' = \prod_I\binom{N\cdot A_I}{[N\cdot a_{IJK}]_J}.$$

Let us show how to process the triples so that they no longer share variables.

Pick M to be a prime which is $\Theta(\aleph)$. Let S be a Salem-Spencer set of size roughly $M^{1-o(1)}$ as in the Salem-Spencer theorem. The $o(1)$ term will go to 0 when we let N go to infinity. In the following we'll let $|S| = M^{1-\varepsilon}$ and in the end we'll let ε go to 0, similar to [10]; this is possible since our final inequality will depend on $1/M^{\varepsilon/N}$ which goes to 1 as N goes to ∞ and ε goes to 0.

Choose random numbers $w_0, w_1, \ldots, w_N$ in $\{0,\ldots,M-1\}$.

For an index sequence $\bar I$, define the hash functions which map the variable indices to integers, just as in [10]:

$$b_x(\bar I) = \sum_{pos=1}^N w_{pos}\cdot\bar I[pos] \mod M,$$

$$b_y(\bar I) = w_0 + \sum_{pos=1}^N w_{pos}\cdot\bar I[pos] \mod M,$$

$$b_z(\bar I) = 1/2(w_0 + \sum_{pos=1}^N (P\mathcal{K} - w_{pos}\cdot\bar I[pos])) \mod M.$$

Set to 0 all variables with blocks mapping to outside S.

For any triple with blocks $\bar I, \bar J, \bar K$ in the remaining trilinear form we have that $b_x(\bar I) + b_y(\bar J) + 2b_z(\bar K) = 0$. Hence, the hashes of the blocks form an arithmetic progression of length 3. Since S contains no nontrivial arithmetic progressions, we get that for any nonzero triple

$$b_x(\bar I) = b_y(\bar J) = b_z(\bar K).$$

Thus, the Salem-Spencer set S has allowed us to do some partitioning of the triples.

Let us analyze how many triples remain. Since M is prime, and due to our choice of functions, the x,y and z blocks are independent and are hashed uniformly to $\{0,\ldots,M-1\}$. If the I and J blocks of a triple $X^IY^JZ^K$ are mapped to the same value, so is the K block. The expected fraction of triples which remain is hence

$$(M^{1-\varepsilon}/M)\cdot(1/M), \text{ which is } 1/M^{1+\varepsilon}.$$

This holds for the triples that satisfy our choice of partition $[a_{IJK}]$.

The trilinear forms corresponding to block triples mapped to the same value in S can still share variables. We do some pruning in order to remove shared blocks, similar to [10], with a minor change. For each $s\in S$, process the triples hashing to s separately.

We first process the triples that obey our choice of $[a_{IJK}]$, until they do not share any variables. After that we also process the remaining triples, zeroing them out if necessary. (This is slightly different from [10].)

Greedily build a list L of independent triples as follows. Suppose we process a triple with blocks $\bar I, \bar J, \bar K$. If $\bar I$ is among the x blocks in another triple in L, then set to 0 all y variables with block $\bar J$. Similarly, if $\bar I$ is not shared but $\bar J$ or $\bar K$ is, then set all x variables with block $\bar I$ to 0. If no blocks are shared, add the triple to L.

Suppose that when we process a triple $\bar I, \bar J, \bar K$, we find that it shares a block, say $\bar I$, with a triple $\bar I, \bar J', \bar K'$ in L. Suppose that we then eliminate all variables sharing block $\bar J$, and thus eliminate U new triples for some U. Then we eliminate at least $\binom{U}{2}+1$ pairs of triples which share a block: the $\binom{U}{2}$ pairs of the eliminated triples that share block $\bar J$, and the pair $\bar I, \bar J, \bar K$ and $\bar I, \bar J', \bar K'$ which share $\bar I$.

Since $\binom{U}{2}+1 \geq U$, we eliminate at least as many pairs as triples. The expected number of unordered pairs of triples sharing an X (or Y or Z) block and for which at least one triple obeys our choice of $[a_{IJK}]$ is

$$M^{-2-\varepsilon}\left[(1/2)\binom{N}{[N\cdot A_I]}\aleph'(\aleph'-1) + \binom{N}{[N\cdot A_I]}\aleph'(\aleph-\aleph')\right]$$

$$= M^{-2-\varepsilon}\left(\binom{N}{[N\cdot A_I]}\aleph'\right)(\aleph - \aleph'/2 - 1/2).$$

Thus at most this many triples obeying our choice of $[a_{IJK}]$ have been eliminated. Hence the expected number of such triples remaining after the pruning is

$$M^{-1-\varepsilon}\binom{N}{[N\cdot A_I]}\aleph'\cdot\left(1 - \aleph/M + \aleph'/(2M)\right) \geq$$

$$\binom{N}{[N \cdot A_I]} \aleph'/(CM^{1+\varepsilon}),$$

for some constant C (depending on how large we pick M to be in terms of $\aleph$). We can pick values for the variables w_i in the hash functions which we defined so that at least this many triples remain. (Picking these values determines our algorithm.)

We have that

$$\max_{[a_{IJK}]} \prod_I \binom{N \cdot A_I}{[N \cdot a_{IJK}]_J} \leq \aleph \leq$$

$$poly(N) \max_{[a_{IJK}]} \prod_I \binom{N \cdot A_I}{[N \cdot a_{IJK}]_J}.$$

We will approximate $\aleph$ by $\aleph_{\max} = \max_{[a_{IJK}]} \prod_I \binom{N \cdot A_I}{[N \cdot a_{IJK}]_J}$.

We have obtained

$$\Omega\left(\binom{N}{[N \cdot A_I]} \frac{\aleph'}{\aleph_{\max}} \cdot \frac{1}{\text{poly}(N) M^{\varepsilon}}\right)$$

trilinear forms that do not share any variables and each of which has value $\prod_{I,J} V_{IJK}^{a_{IJK} N}$.

If we were to set $\aleph' = \aleph_{\max}$ we would get $\Omega\left(\frac{\binom{N}{[N \cdot A_I]}}{\text{poly}(N) M^{\varepsilon}}\right)$ trilinear forms instead. We use this setting in our analyses, though a better analysis may be possible if you allow $\aleph'$ to vary.

We will see later that the best choice of $[a_{IJK}]$ sets $a_{IJK} = a_{sort(IJK)}$ for each I, J, K, where $sort(IJK)$ is the permutation of IJK sorting them in lexicographic order (so that $I \leq J \leq K$). Since tensor rank is invariant under permutations of the roles of the x, y and z variables, we also have that $V_{IJK} = V_{sort(IJK)}$ for all I, J, K. Let $perm(I, J, K)$ be the number of distinct permutations of I, J, K. That is, if $I = J = K$, $perm(I, J, K) = 1$, if $I = J \neq K$, $perm(I, J, K) = 3$ and finally if $I \neq J \neq K \neq I$, $perm(I, J, K) = 6$.

Recall that r was the bound on the (border) rank of C given by the construction. Then, by Theorem 3, we get the inequality

$$r^{\mathcal{K}N} \geq \binom{N}{[N \cdot A_I]} \frac{\aleph'}{\aleph_{\max}} \frac{1}{\text{poly}(N) M^{\varepsilon}} \prod_{I \leq J \leq K} (V_{IJK}(\tau))^{perm(IJK) \cdot N \cdot a_{IJK}}.$$

Let $\bar{a}_{IJK}$ be the choices which achieve $\aleph_{\max}$ so that $\aleph_{\max} = \prod_I \binom{N \cdot A_I}{[N \cdot \bar{a}_{IJK}]_J}$. Then, by taking Stirling's approximation we get that

$$(\aleph'/\aleph_{\max})^{1/N} = \prod_{IJK} \frac{\bar{a}_{IJK}^{\bar{a}_{IJK}}}{a_{IJK}^{a_{IJK}}}.$$

Taking the N-th root, taking N to go to ∞ and ε to go to 0, and using Stirling's approximation we obtain the following inequality:

$$r^{\mathcal{K}} \geq \prod_{I \leq J \leq K} \left(\frac{\bar{a}_{IJK}^{\bar{a}_{IJK}}}{a_{IJK}^{a_{IJK}}}\right)^{perm(IJK)} \cdot \frac{V_{IJK}^{perm(IJK) \cdot a_{IJK}}}{\prod_I A_I^{A_I}}.$$

If we set $a_{IJK} = \bar{a}_{IJK}$, we get the simpler inequality

$$r^{\mathcal{K}} \geq \prod_{I \leq J \leq K} (V_{IJK})^{perm(IJK) \cdot a_{IJK}} / \prod_I A_I^{A_I},$$

which is what we use in our application of the theorem as it reduces the number of variables and does not seem to change the final bound on ω by much.

The values V_{IJK} are nondecreasing functions of τ, where $\tau = \omega/3$. The inequality above gives an upper bound on τ and hence on ω.

Computing $\bar{a}_{IJK}$ and a_{IJK}.

Here we show how to compute the values $\bar{a}_{IJK}$ forming $\aleph_{\max}$ and a_{IJK} which maximize our bound on ω.

The only restriction on a_{IJK} is that $A_I = \sum_J a_{IJK} = \sum_J \bar{a}_{IJK}$, and so if we know how to pick $\bar{a}_{IJK}$, we can let a_{IJK} vary subject to the constraints $\sum_J a_{IJK} = \sum_J \bar{a}_{IJK}$. Hence we will focus on computing $\bar{a}_{IJK}$.

Recall that $\bar{a}_{IJK}$ is the setting of the variables a_{IJK} which maximizes $\prod_I \binom{N \cdot A_I}{[N \cdot a_{IJK}]_J}$ for fixed A_I.

Because of our symmetric choice of the A_I, the above is maximized for $\bar{a}_{IJK} = \bar{a}_{sort(IJK)}$, where $sort(IJK)$ is the permutation of I, J, K which sorts them in lexicographic order.

Let $perm(I, J, K)$ be the number of distinct permutations of I, J, K. The constraint satisfied by the a_{IJK} becomes

$$1 = \sum_I A_I = \sum_{I \leq J \leq K} perm(I, J, K) \cdot a_{IJK}.$$

The constraint above together with $\bar{a}_{IJK} = \bar{a}_{sort(IJK)}$ are the only constraints in the original CW paper. However, it turns out that more constraints are necessary for $\mathcal{K} > 2$.

The equalities $A_I = \sum_J \bar{a}_{IJK}$ form a system of linear equations involving the variables $\bar{a}_{IJK}$ and the fixed values A_I. If this system had full rank, then the $\bar{a}_{IJK}$ values (for $\bar{a}_{IJK} = \bar{a}_{sort(IJK)}$) would be determined from the A_I and hence $\aleph$ would be exactly $\prod_I \binom{N \cdot A_I}{[N \cdot \bar{a}_{IJK}]_J}$, and a further maximization step would not be necessary. This is exactly the case for $\mathcal{K} = 2$ in [10]. This is also why in [10], setting $a_{IJK} = \bar{a}_{IJK}$ was necessary.

However, the system of equations may not have full rank. Because of this, let us pick a minimum set S of variables $\bar{a}_{\bar{I}\bar{J}\bar{K}}$ so that viewing these variables as constaints would make the system have full rank.

Then, all variables $\bar{a}_{IJK} \notin S$ would be determined as linear functions depending on the A_I and the variables in S.

Consider the function G of A_I and the variables in S, defined as

$$G = \prod_I \binom{N \cdot A_I}{[N \cdot \bar{a}_{IJK}]_{\bar{a}_{IJK} \notin S}, [N \cdot \bar{a}_{IJK}]_{\bar{a}_{IJK} \in S}}.$$

G is only a function of $\{\bar{a}_{IJK} \in S\}$ for fixed $\{A_i\}_i$. We want to know for what values of the variables of S, G is maximized.

G is maximized when $\prod_{IJ} (\bar{a}_{IJK} N)!$ is minimized, which in turn is minimized exactly when $F = \sum_{IJ} \ln((N \bar{a}_{IJK})!)$ is minimized, where $K = P\mathcal{K} - I - J$.

Using Stirling's approximation $\ln(n!) = n \ln n - n + O(\ln n)$, we get that F is roughly equal to

$$N[\sum_{IJ} \bar{a}_{IJK} \ln(\bar{a}_{IJK}) - \bar{a}_{IJK} + \bar{a}_{IJK} \ln N + O(\log(N \bar{a}_{IJK})/N)] =$$

$$N \ln N + N[\sum_{IJ} \bar{a}_{IJK} \ln(a_{IJK}) - \bar{a}_{IJK} + O(\log(N \bar{a}_{IJK})/N)],$$

since $\sum_{IJ} \bar{a}_{IJK} = \sum_I A_I = 1$. As N goes to ∞, for any fixed setting of the $\bar{a}_{IJK}$ variables, the $O(\log N/N)$ term vanishes, and F is roughly $N \ln N + N(\sum_{IJ} \bar{a}_{IJK} \ln(\bar{a}_{IJK}) -$

$\bar{a}_{IJK}$). Hence to minimize F we need to minimize $f = (\sum_{IJ} \bar{a}_{IJK} \ln(\bar{a}_{IJK}) - \bar{a}_{IJK})$.

We want to know for what values of $\bar{a}_{IJK}$, f is minimized. Since f is convex for positive a_{IJK}, it is actually minimized when $\frac{\partial f}{\partial \bar{a}_{IJK}} = 0$ for every $\bar{a}_{IJK} \in S$. Recall that the variables not in S are linear combinations of those in S.[2]

Taking the derivatives, we obtain for each $\bar{a}_{IJK}$ in S:

$$0 = \frac{\partial f}{\partial \bar{a}_{IJK}} = \sum_{I'J'K'} \ln(\bar{a}_{I'J'K'}) \frac{\partial \bar{a}_{I'J'K'}}{\partial \bar{a}_{IJK}}.$$

We can write this out as

$$1 = \prod_{I'J'K'} (\bar{a}_{I'J'K'})^{\frac{\partial \bar{a}_{I'J'K'}}{\partial \bar{a}_{IJK}}}.$$

Since each variable $\bar{a}_{I'J'K'}$ in the above equality for $\bar{a}_{IJK}$ is a linear combination of variables in S, we have that the exponent $p_{I'J'K'IJK} = \frac{\partial \bar{a}_{I'J'K'}}{\partial \bar{a}_{IJK}}$ is actually a constant, and so we get a system of polynomial equality constraints which define the variables in S in terms of the variables outside of S: for any $\bar{a}_{IJK} \in S$, we get

$$\bar{a}_{IJK} \cdot \prod_{\bar{a}_{I'J'K'} \notin S, p_{I'J'K'IJK} > 0} (\bar{a}_{I'J'K'})^{p_{I'J'K'IJK}} = \prod_{\bar{a}_{I'J'K'} \notin S, p_{I'J'K'IJK} < 0} (\bar{a}_{I'J'K'})^{-p_{I'J'K'IJK}}. \quad (1)$$

Given values for the variables not in S, we can use (1) to get valid values for the variables in S, and hence also for the A_I. For that choice of the A_I, G is maximized for exactly the variable settings we have picked. Now all we have to do is find the correct values for the variables outside of S and for $\bar{a}_{IJK}$, given the constraints $A_I = \sum_J \bar{a}_{IJK}$.

We cannot pick arbitrary values for the variables outside of S. They need to satisfy the following constraints:

- the obtained A_I satisfy $\sum_I A_I = 1$, and
- the variables in S obtained from Equation 1 are nonnegative.

In summary, we obtain the procedure to analyze the $\mathcal{K}$ tensor power shown in Figure 1.

4. ANALYZING THE SMALLER TENSORS

Consider the trilinear form consisting only of the variables from the $\mathcal{K}$ tensor power of C, with blocks I, J, K, where $I+J+K = P \cdot \mathcal{K}$. In this section we will prove the following theorem:

THEOREM 5. *Given a (p, P)-uniform construction C, using the procedure in Figure 2 one can compute lower bounds on the values V_{IJK} for any tensor power of C. The $\mathcal{K}$ tensor power requires $O(\mathcal{K}^2)$ applications of the procedure.*

[2]We could have instead written $f = \sum_{IJ} \bar{a}_{IJK} \ln(\bar{a}_{IJK})$ and minimized f, and the equalities we would have obtained would have been exactly the same since the system of equations includes the equation $\sum_{IJ} \bar{a}_{IJK} = 1$, and although $\partial f / \partial a$ is $\sum_{IJ} \frac{\partial \bar{a}_{IJK} \ln a_{IJK}}{\partial a} = \sum_{IJ} \frac{\partial \bar{a}_{IJK}}{\partial a} (\ln \bar{a} a_{IJK} - 1)$, the -1 in the brackets would be canceled out: if $\bar{a}_{0,0,P\mathcal{K}} = (1 - \sum_{IJ:\ (I,J) \neq (0,0)} \bar{a}_{IJK})$, then $\frac{\partial \bar{a}_{0,0,P\mathcal{K}} \ln \bar{a}_{0,0,P\mathcal{K}}}{\partial a} = \ln(\bar{a}_{0,0,P\mathcal{K}}) \frac{\partial \bar{a}_{0,0,P\mathcal{K}}}{\partial a} + \sum_{I'J':\ (I',J') \neq (0,0)} \frac{\partial \bar{a}_{I'J'K'}}{\partial a}$.

1. Define variables α_{ijk} and X_i, Y_j, Z_k for all good triples i, j, k.
2. Form the linear system consisting of $X_i = \sum_j \alpha_{ijk}$, $Y_j = \sum_i \alpha_{ijk}$ and $Z_k = \sum_i \alpha_{ijk}$.
3. Determine the rank of the system: it is exactly $\#i + \#j + \#k - 2$ because of the fact that $\sum_i X_i = \sum_j Y_j = \sum_k Z_k$.
4. If the system does not have full rank, then pick enough variables α_{ijk} to treat as constants; place them in a set Δ.
5. Solve the system for the variables outside of Δ in terms of the ones in Δ and X_i, Y_j, Z_k. Now we have $\alpha_{ijk} = \alpha_{ijk}([X_i], [Y_j], [X_k], y \in \Delta)$.
6. Let $W_{i,j,k} = V_{i,j,k} V_{I-i,J-j,K-k}$. Compute for every ℓ,
$$nx_\ell = \prod_{i,j,k} W_{ijk}^{3 \frac{\partial \alpha_{ijk}}{\partial X_\ell}},$$
$$ny_\ell = \prod_{i,j,k} W_{ijk}^{3 \frac{\partial \alpha_{ijk}}{\partial Y_\ell}}, \text{ and,}$$
$$nz_\ell = \prod_{i,j,k} W_{ijk}^{3 \frac{\partial \alpha_{ijk}}{\partial Z_\ell}}.$$
7. Compute for every variable $y \in \Delta$,
$$ny = \prod_{i,j,k} W_{i,j,k}^{\frac{\partial \alpha_{ijk}}{\partial y}}.$$
8. Compute for each α_{ijk} its setting $\alpha_{ijk}(\Delta)$ as a function of the $y \in \Delta$ when $X_\ell = nx_\ell / \sum_i nx_i$, $Y_\ell = ny_\ell / \sum_j ny_j$ and $Z_\ell = nz_\ell / \sum_k nz_k$.
9. Then set
$$V_{IJK} = (\sum_\ell nx_\ell)^{1/3} (\sum_\ell ny_\ell)^{1/3} (\sum_\ell nz_\ell)^{1/3} \prod_{y \in \Delta} ny^y.$$
subject to the constraints on $y \in \Delta$ given by
$$y \geq 0 \text{ for all } y \in \Delta,$$
$$\alpha_{ijk}(\Delta) \geq 0 \text{ for every } \alpha_{ijk} \notin S.$$
10. Find the setting of the $y \in \Delta$ that maximizes the bound on V_{IJK}. For any fixed guess for τ, this is a linear program: Maximize $\sum_{y \in \Delta} y \log ny$ subject to the above linear constraints. Or, alternatively, let V_{IJK} be a function of $y \in \Delta$ and add the above two constraints to the final program in Figure 1 computing ω.

Figure 2: The procedure for computing V_{IJK} for arbitrary tensor powers.

Suppose that we have analyzed the values for some powers $\mathcal{K}'$ and $\mathcal{K} - \mathcal{K}'$ of the trilinear form from the construction with $\mathcal{K}' < \mathcal{K}$. We will show how to inductively analyze the values for the $\mathcal{K}$ power, using the values for these smaller powers. The theorem will follow by noting that the number of values for the $\mathcal{K}$ power is $O(\mathcal{K}^2)$ and that one can use recursion to first compute the values for the $\lfloor \mathcal{K}/2 \rfloor$ and $\lceil \mathcal{K}/2 \rceil$ powers and then combining them to obtain the values for the $\mathcal{K}$ power.

Consider the $\mathcal{K}$ tensor power of the trilinear form C. It

can actually be viewed as the tensor product of the $\mathcal{K}'$ and $\mathcal{K}-\mathcal{K}'$ tensor powers of C.

Recall that the indices of the variables of the $\mathcal{K}$ tensor power of C are $\mathcal{K}$-length sequences of indices of the variables of C. Also recall that if p was the function which maps the indices of C to blocks, then we define $p^{\mathcal{K}}$ to be a function which maps the $\mathcal{K}$ power indices to blocks as $p^{\mathcal{K}}(index) = \sum_{pos} p(index[pos])$.

An index of a variable in the $\mathcal{K}$ tensor power of C can also be viewed as a pair (l,m) such that l is an index of a variable in the $\mathcal{K}'$ tensor power of C and m is an index of a variable in the $\mathcal{K}-\mathcal{K}'$ tensor power of C. Hence we get that $p^{\mathcal{K}}((l,m)) = p^{\mathcal{K}'}(l) + p^{\mathcal{K}-\mathcal{K}'}(m)$.

For any I,J,K which form a valid block triple of the $\mathcal{K}$ tensor power, we consider the trilinear form $T_{I,J,K}$ consisting of all triples $x_i y_j z_k$ of the $\mathcal{K}$ tensor power of the construction for which $p^{\mathcal{K}}(i) = I, p^{\mathcal{K}}(j) = J, p^{\mathcal{K}}(k) = K$.

$T_{I,J,K}$ consists of the trilinear forms $T_{i,j,k} \otimes T_{I-i,J-j,K-k}$ for all i,j,k that form a valid block triple for the $\mathcal{K}'$ power, and such that $I-i, J-j, K-k$ form a valid block triple for the $\mathcal{K}-\mathcal{K}'$ power. Call such blocks i,j,k *good*. Then:

$$T_{IJK} = \sum_{\text{good } ijk} T_{i,j,k} \otimes T_{I-i,J-j,K-k}.$$

(The sum above is a regular sum, not a disjoint sum, so the trilinear forms in it may share indices.)

The above decomposition of T_{IJK} was first observed by Stothers [18] when considering the special cases of the 3rd and 4th tensor powers. This is the shortcut that allowed us to simplify our analysis.

Let $Q_{ijk} = T_{ijk} \otimes T_{I-i,J-j,K-k}$. By supermultiplicativity, the value W_{ijk} of Q_{ijk} satisfies $W_{ijk} \geq V_{ijk} V_{I-i,J-j,K-k}$. If the trilinear forms Q_{ijk} didn't share variables, then we would immediately obtain a lower bound on the value V_{IJK} as $\sum_{ijk} V_{ijk} V_{I-i,J-j,K-k}$. However, the trilinear forms Q_{ijk} may share variables, and we'll apply the techniques from the previous section to remove the dependencies.

To analyze the value V_{IJK} of $T_{I,J,K}$, we first take the N-th tensor power of $T_{I,J,K}$, the N-th tensor power of $T_{K,I,J}$ and the N-th tensor power of $T_{J,K,I}$, and then tensor multiply these altogether. By the definition of value, $V_{I,J,K}$ is at least the $3N$-th root of the value of the new trilinear form.

Here is how we process the N-th tensor power of $T_{I,J,K}$. The powers of $T_{K,I,J}$ and $T_{J,K,I}$ are processed similarly.

We pick values $X_i \in [0,1]$ for each block i of the $\mathcal{K}'$ tensor power of C so that $\sum_i X_i = 1$. Set to 0 all x variables except those that have exactly $X_i \cdot N$ positions of their index which are mapped to $(i, I-i)$ by $(p^{\mathcal{K}'}, p^{\mathcal{K}-\mathcal{K}'})$, for all i.

The number of nonzero x blocks is $\binom{N}{[N \cdot X_i]_i}$.

Similarly pick values Y_j for the y variables, with $\sum_j Y_j = 1$, and retain only those with Y_j index positions mapped to $(j, J-j)$. Similarly pick values Z_k for the z variables, with $\sum_k Z_k = 1$, and retain only those with Z_k index positions mapped to $(k, K-k)$.

The number of nonzero y blocks is $\binom{N}{[N \cdot Y_j]_j}$. The number of nonzero z blocks is $\binom{N}{[N \cdot Z_k]_k}$.

For $i,j,k = P\mathcal{K}' - i - j$ which are valid blocks of the $\mathcal{K}'$ tensor power of C with good i,j,k, let α_{ijk} be variables such that $X_i = \sum_j \alpha_{ijk}$, $Y_j = \sum_i \alpha_{ijk}$ and $Z_k = \sum_i \alpha_{ijk}$.

After taking the tensor product of what is remaining of the Nth tensor powers of $T_{I,J,K}$, $T_{K,I,J}$ and $T_{J,K,I}$, the number of x, y or z blocks is

$$\Gamma = \binom{N}{[N \cdot X_i]} \binom{N}{[N \cdot Y_j]} \binom{N}{[N \cdot Z_k]}.$$

The number of block triples which contain a particular x, y or z block is

$$\aleph = \prod_i \binom{NX_i}{[N\alpha_{ijk}]_j} \prod_j \binom{NY_j}{[N\alpha_{ijk}]_i} \prod_k \binom{NZ_k}{[N\alpha_{ijk}]_i}.$$

Hence the number of triples is $\Gamma \cdot \aleph$.

Set $M = \Theta(\aleph)$ to be a large enough prime greater than $\aleph$. Create a Salem-Spencer set S of size roughly $M^{1-\varepsilon}$. Pick random values $w_0, w_1, w_2, \ldots, w_{3N}$ in $\{0, \ldots, M-1\}$.

The blocks for x, y, or z variables of the new big trilinear form are sequences of length $3N$; the first N positions of a sequence contain pairs $(i, I-i)$, the second N contain pairs $(j, J-j)$ and the last N contain pairs $(k, K-k)$. We can thus represent the block sequences I of the $\mathcal{K}$ tensor power as (I_1, I_2) where I_1 is a sequence of length $3N$ of blocks of the $\mathcal{K}'$ power of C and I_2 is a sequence of length $3N$ of blocks of the $\mathcal{K}-\mathcal{K}'$ power of C (the first N are x blocks, the second N are y blocks and the third N are z blocks).

For a particular block sequence $I = (I_1, I_2)$, we define the hash functions that depend only on I_1:

$$b_x(I) = \sum_{pos=1}^{3N} w_{pos} \cdot I_1[pos] \mod M,$$

$$b_y(I) = w_0 + \sum_{pos=1}^{3N} w_{pos} \cdot I_1[pos] \mod M,$$

$$b_z(I) = 1/2(w_0 + \sum_{pos=1}^{3N} (P\mathcal{K}' - (w_{pos} \cdot I_1[pos]))) \mod M.$$

We then set to 0 all variables that do not have blocks hashing to elements of S. Again, any surviving triple has all variables' blocks mapped to the same element of S. The expected fraction of triples remaining is $M^{1-\varepsilon}/M^2 = 1/M^{1+\varepsilon}$.

As before, we do the pruning of the triples mapped to each element of S separately. The expected number of unordered pairs of triples sharing an x, y or z block is $(3/2)\Gamma\aleph(\aleph - 1)/M^3 \leq \Gamma\aleph/(c \cdot M^2)$ for large constant c, and the number of remaining block triples over all elements of S is $\Omega(\Gamma\aleph/M^{1+\varepsilon}) = \Omega(\Gamma/M^{\varepsilon})$. (Recall that Γ is the number of blocks and $\Gamma\aleph$ was the original number of triples.) Analogously to [10], we will let ε go to 0 and so the expected number of remaining triples is roughly Γ. Hence we can pick a setting of the w_i variables so that roughly Γ triples remain. We have obtained about Γ independent trilinear forms, each of which has value at least

$$\prod_{i,j,k} (V_{i,j,k} \cdot V_{I-i,J-j,K-k})^{3N\alpha_{ijk}}.$$

This follows since values are supermultiplicative.

The final inequality becomes

$$V_{I,J,K}^{3N} \geq \binom{N}{[N \cdot X_i]} \binom{N}{[N \cdot Y_j]} \binom{N}{[N \cdot Z_k]} \prod_{i,j,k} (V_{i,j,k} \cdot V_{I-i,J-j,K-k})^{3N\alpha_{ijk}}.$$

Recall that we have equalities $X_i = \sum_j \alpha_{ijk}$, $Y_j = \sum_i \alpha_{ijk}$, and $Z_k = \sum_i \alpha_{ijk}$. If we fix X_i, Y_j, Z_k over all i, j, k, this forms a linear system.

The linear system does not necessarily have full rank, and so we pick a minimum set Δ of variables α_{ijk} so that if they are treated as constants, the linear system has full rank, and the variables outside of Δ can be written as linear combinations of variables in Δ and of X_i, Y_j, Z_k.

Now we have that for every α_{ijk},

$$\alpha_{ijk} = \sum_{y \in \Delta \cup \{X_{i'}, Y_{j'}, Z_{k'}\}_{i',j',k'}} y \frac{\partial \alpha_{ijk}}{\partial y},$$

where for all $\alpha_{ijk} \notin \Delta$ we use the linear function obtained from the linear system.

Let $\delta_{ijk} = \sum_{y \in \Delta} y \frac{\partial \alpha_{ijk}}{\partial y}$. Let $W_{i,j,k} = V_{i,j,k} \cdot V_{I-i,J-j,K-k}$. Then,

$$W_{i,j,k}^{\alpha_{ijk}} = W_{ijk}^{\sum_i X_i \frac{\partial \alpha_{ijk}}{\partial X_i}} W_{ijk}^{\sum_i Y_j \frac{\partial \alpha_{ijk}}{\partial Y_j}} W_{ijk}^{\sum_k Z_k \frac{\partial \alpha_{ijk}}{\partial Z_k}} W_{i,j,k}^{\delta_{ijk}}.$$

Define $nx_\ell = \prod_{i,j,k} W_{ijk}^{3\frac{\partial \alpha_{ijk}}{\partial X_\ell}}$ for any ℓ. Set $\bar{nx}_\ell = \frac{nx_\ell}{\sum_{\ell'} nx_{\ell'}}$.

Define similarly $ny_\ell = \prod_{i,j,k} W_{ijk}^{3\frac{\partial \alpha_{ijk}}{\partial Y_\ell}}$, $nz_\ell = \prod_{i,j,k} W_{ijk}^{3\frac{\partial \alpha_{ijk}}{\partial Z_\ell}}$, setting $\bar{ny}_\ell = \frac{ny_\ell}{\sum_{\ell'} ny_{\ell'}}$ and $\bar{nz}_\ell = \frac{nz_\ell}{\sum_{\ell'} nz_{\ell'}}$.

Consider the right hand side of our inequality for V_{IJK}:

$$\binom{N}{[N \cdot X_i]}\binom{N}{[N \cdot Y_j]}\binom{N}{[N \cdot Z_k]} \prod_{i,j,k} W_{i,j,k}^{3N\alpha_{ijk}} =$$

$$\binom{N}{[N \cdot X_i]} \prod_\ell nx_\ell^{NX_\ell} \binom{N}{[N \cdot Y_j]} \prod_\ell ny_\ell^{NY_\ell} \cdot$$

$$\cdot \binom{N}{[N \cdot Z_k]} \prod_\ell nz_\ell^{NZ_\ell} \prod_{i,j,k} W_{i,j,k}^{(\sum_{y \in \Delta} y \frac{\partial \alpha_{ijk}}{\partial y})}.$$

By Lemma 1, the above is maximized for $X_\ell = n\bar{x}_\ell$, $Y_\ell = n\bar{y}_\ell$, and $Z_\ell = n\bar{z}_\ell$ for all ℓ, and for these settings, for instance $\binom{N}{[N \cdot X_i]} \prod_\ell nx_\ell^{NX_\ell}$ is essentially $(\sum_\ell nx_\ell)^N / \operatorname{poly}(N)$, and hence after taking the $3N$th root and letting N go to ∞, we obtain

$$V_{I,J,K} \geq$$

$$(\sum_\ell nx_\ell)^{1/3} (\sum_\ell ny_\ell)^{1/3} (\sum_\ell nz_\ell)^{1/3} \prod_{i,j,k} W_{i,j,k}^{(\sum_{y \in \Delta} y \frac{\partial \alpha_{ijk}}{\partial y})}.$$

If $\Delta = \emptyset$, then the above is a complete formula for $V_{I,J,K}$. Otherwise, to maximize the lower bound on $V_{I,J,K}$ we need to pick values for the variables in Δ, while still preserving the constraints that the values for the variables outside of Δ (which are obtained from our settings of the X_i, Y_j, Z_k and the values for the Δ variables) are nonnegative.

We obtain the procedure for computing lower bounds on the values $V_{I,J,K}$ shown in Figure 2.

4.1 Powers of two

Because the constraint program in the previous section is tricky to solve, we want to be able to reduce the number of variables. It turns out that when the tensor power $\mathcal{K}$ is a power of 2, say $\mathcal{K} = 2^\kappa$, we can use $\mathcal{K}' = \mathcal{K} - \mathcal{K}' = 2^{\kappa-1}$ and we can reduce the number of variables (roughly by half) by exploiting the symmetry. In the full version we show how to modify the procedure in Figure 2.

5. CONCLUSION

We applied our procedures to the second, third, fourth and eighth tensor powers. By applying our procedures to the second tensor power of the CW construction, we obtained the same constraint program that Coppersmith and Winograd obtained, thus obtaining $\omega < 2.376$. Applying them to the fourth tensor power we obtained the same constraint program as in Stothers' thesis, however we were able to obtain an improved solution $\omega < 2.37293$ by using better optimization procedures. Applying the procedures to the eighth tensor power was done entirely by computer, using a combination of Maple and C++ and the nonlinear optimization software NLOPT. We are not certain whether the optimization software we used to solve the resulting constraint program found the optimum solution. However, we found a feasible solution which shows that $\omega < 2.3727$. It is possible that a better bound can be found for the eighth tensor power, and we believe that higher tensor powers of the CW construction should improve the bound on ω further. Nevertheless, it seems that in order to approach $\omega = 2$, we need a new basic construction. It is very possible that a combination of the group theoretic approach of [7] and our techniques can lead to further improvements.

Acknowledgments.

The author is grateful to Satish Rao for encouraging her to explore the matrix multiplication problem more thoroughly and to Ryan Williams for his unconditional support. The author would also like to thank François Le Gall who alerted her to Stothers' work, suggested the use of NLOPT, and pointed out that the feasible solution obtained by Stothers for his 4th tensor power constraint program is not optimal and that one can obtain $\omega < 2.37294$ with a different setting of the parameters[12].

The author was supported by NSF Grants CCF-0830797 and CCF-1118083 at UC Berkeley, and by NSF Grants IIS-0963478 and IIS-0904325, and an AFOSR MURI Grant, at Stanford University.

6. REFERENCES

[1] A. V. Aho, J. E. Hopcroft, and J. Ullman. The design and analysis of computer algorithms. *Addison-Wesley Longman Publishing Co., Boston, MA*, 1974.

[2] N. Alon, A. Shpilka, and C. Umans. On sunflowers and matrix multiplication. *ECCC TR11-067*, 18, 2011.

[3] F. A. Behrend. On the sets of integers which contain no three in arithmetic progression. *Proc. Nat. Acad. Sci.*, pages 331–332, 1946.

[4] D. Bini, M. Capovani, F. Romani, and G. Lotti. $O(n^{2.7799})$ complexity for $n \times n$ approximate matrix multiplication. *Inf. Process. Lett.*, 8(5):234–235, 1979.

[5] M. Bläser. Complexity of bilinear problems (lecture notes scribed by F. Endun). `http://www-cc.cs.uni-saarland.de/teaching/SS09/ComplexityofBilinearProblems/script.pdf`, 2009.

[6] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic complexity theory, Grundlehren der*

mathematischen Wissenschaften. Springer-Verlag, Berlin, 1996.

[7] H. Cohn, R. Kleinberg, B. Szegedy, and C. Umans. Group-theoretic algorithms for matrix multiplication. In *Proc. FOCS*, volume 46, pages 379–388, 2005.

[8] H. Cohn and C. Umans. A group-theoretic approach to fast matrix multiplication. In *Proc. FOCS*, volume 44, pages 438–449, 2003.

[9] D. Coppersmith and S. Winograd. On the asymptotic complexity of matrix multiplication. In *Proc. SFCS*, pages 82–90, 1981.

[10] D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic progressions. *J. Symbolic Computation*, 9(3):251–280, 1990.

[11] P. Erdös and R. Rado. Intersection theorems for systems of sets. *J. London Math. Soc.*, 35:85–90, 1960.

[12] F. Le Gall. Personal communication.

[13] E. Mossel, R. O'Donnell, and R. A. Servedio. Learning juntas. In *Proc. STOC*, pages 206–212, 2003.

[14] V. Y. Pan. Strassen's algorithm is not optimal. In *Proc. FOCS*, volume 19, pages 166–176, 1978.

[15] F. Romani. Some properties of disjoint sums of tensors related to matrix multiplication. *SIAM J. Comput.*, pages 263–267, 1982.

[16] R. Salem and D. Spencer. On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci.*, 28(12):561–563, 1942.

[17] A. Schönhage. Partial and total matrix multiplication. *SIAM J. Comput.*, 10(3):434–455, 1981.

[18] A. Stothers. *Ph.D. Thesis, U. Edinburgh*, 2010.

[19] V. Strassen. Gaussian elimination is not optimal. *Numer. Math.*, 13:354–356, 1969.

[20] V. Strassen. Relative bilinear complexity and matrix multiplication. *J. reine angew. Math. (Crelles Journal)*, 375–376:406–443, 1987.

[21] L. G. Valiant. General context-free recognition in less than cubic time. *Journal of Computer and System Sciences*, 10:308–315, 1975.