ILLiad Request Printout

Transaction Number: 507299
Username: 100934397 Name: William Gasarch
ISSN/ISBN: 0747-7171
NotWantedAfter: 11/09/2010
Accept Non English: Yes
Accept Alternate Edition: No
Request Type: Article - Article

Loan Information

LoanAuthor:
LoanTitle:
LoanPublisher:
LoanPlace:
LoanDate:
LoanEdition:
NotWantedAfter: 11/09/2010

Article Information

PhotoJournalTitle: Journal of symbolic computation
PhotoJournalVolume: 9
PhotoJournalIssue: 3
Month:
Year: 1990
Pages: 251- 280
Article Author: Coppersmith, D. + Winograd, S.
Article Title: Matrix Multiplication via arithmetic progressions

Citation Information

Cited In:
Cited Title:
Cited Date:
Cited Volume:
Cited Pages:

OCLC Information

ILL Number:
OCLC Number:
Lending String: Direct Request
Original Loan Author:
Original Loan Title:
Old Journal Title:
Call Number: EPSL Periodical Stacks QA76.95 .J68
Location:

Notes

9/13/2010 8:29:28 AM System 1. No Matching Bib/2. ISSN Search With Too Many Hits.

J. Symbolic Computation (1990) **9**, 251–280

Matrix Multiplication via Arithmetic Progressions

DON COPPERSMITH and SHMUEL WINOGRAD

Department of Mathematical Sciences
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, New York 10598, U.S.A.

(*Received* 17 *May* 1987)

We present a new method for accelerating matrix multiplication asymptotically. This work builds on recent ideas of Volker Strassen, by using a basic trilinear form which is not a matrix product. We make novel use of the Salem-Spencer Theorem, which gives a fairly dense set of integers with no three-term arithmetic progression. Our resulting matrix exponent is 2.376.

1. Introduction.

A matrix multiplication algorithm is usually built as follows. First an algorithm for a small matrix problem is developed. Then a tensor product construction produces from it an algorithm for multiplying large matrices. Several times over the last two decades, the ground rules for constructing the basic algorithm have been relaxed, and with more care in the tensor product construction, it has been shown how to use these more lenient basic constructions to still give efficient algorithms for multiplying large matrices.

Recently Strassen (1986) found a new relaxation of the ground rules. His basic trilinear algorithm computes a trilinear form which is not a matrix product at all. In this trilinear form, the variables are collected into blocks. The block structure (the arrangement of the blocks) is that of a matrix product, and the fine structure (the arrangement of variables within individual blocks) is also that of a matrix product, but the overall structure is not, because the fine structures of different blocks are incompatible. After taking a tensor power of this trilinear form, Strassen operates on the block structure (that of a large matrix product) to reduce it to several block scalar multiplications. Each block scalar multiplication is itself a matrix product (the fine structure), so that he has several disjoint matrix products (sharing no variables). He can then apply Schönhage's τ -theorem to obtain an estimate of the matrix exponent ω:

0747–7171/90/030251+30 $03.00/0

$\omega < 2.479$.

Here we follow Strassen's lead. We use a basic trilinear algorithm closely related to Strassen's. The block structure of our trilinear form is not a matrix product, although the fine structure still is. We use a combinatorial theorem of Salem and Spencer (1942), which gives a fairly dense set of integers containing no three-term arithmetic progression. We hash the indices of the blocks of variables to integers, and set to zero any block of variables not mapping to the Salem-Spencer set. We do this in such a way that if the product $X^{[I]}Y^{[J]}Z^{[K]}$ is contained in our trilinear form, then the hash values $b_X(I)$, $b_Y(J)$, $b_Z(K)$ form an arithmetic progression. So for any product of nonzero blocks $X^{[I]}Y^{[J]}Z^{[K]}$ in our trilinear form, we will get $b_X(I) = b_Y(J) = b_Z(K)$. We choose parameters so that on average each nonzero block of variables is contained in at most one nonzero block product $X^{[I]}Y^{[J]}Z^{[K]}$, and set to zero some blocks of variables to ensure that this condition holds absolutely, not just on average. Then, as Strassen, we have several disjoint matrix products, and can apply Schönhage's τ-theorem to obtain our exponent

$\omega < 2.376$.

The rest of the paper is organized as follows. In Section 2 we review Schönhage's τ-theorem. In Section 3 we present Strassen's construction. Section 4 contains the results of the Salem-Spencer theorem. Section 5 presents an outline of the present construction. In Section 6 we present an example of our construction, which gives an exponent of 2.404. The version presented in Section 7 uses exactly the same ideas, but is complicated by more terms and more indices; this gives an improvement to of 2.388. Section 8 introduces yet more complicated techniques which achieve a slightly better estimate of 2.376. Section 9 contains some related ideas that were not as effective in reducing the exponent. We make miscellaneous remarks in Section 10.

Finally, Section 11 shows how the existence of a certain combinatorial construction would yield $\omega = 2$. We cannot tell whether this construction can be realized.

Earlier versions of this paper appeared as Coppersmith and Winograd (1986) and (1987).

Readers unfamiliar with previous work in matrix multiplication are referred to the excellent survey by Victor Pan (1984).

We are grateful to James
last constituent we needed f
Arnold Schönhage gave a n
James Davenport offered hel

The basic results from "cl
τ-theorem:

Theorem (Schönhage): *(the field of rational functio*

$$\sum_{\ell=1}^{L}\left(\sum_{i,j,h}\alpha_{i,j,h,\ell}x_{i,j}^{(h)}\right)\left($$
$$=\sum_{h}\left(\sum_{i=1}^{m_h}\sum_{j=1}^{n_h}\sum_{k=1}^{p_h}x_{i,j}^{(h)}y$$

is an identity in $x_{i,j}^{(h)}, y_{j,k}^{(h)}, z$ *N × N square matrices in O(*

$$L=\sum_{h}(m_hn_hp_h)^{\tau}.$$

We will also write the e

$$\sum_{\ell=1}^{L}\left(\sum_{i,j,h}\alpha_{i,j,h,\ell}x_{i,j}^{(h)}\right)\left($$

Less formally, the hypot
proximately) compute sim
$m_h \times n_h$ times $n_h \times p_h$ (writ
variable belongs to.

In such a situation, we d

Note: Here we have pr
our construction easier to d

ACKNOWLEDGMENTS

We are grateful to James Shearer for the reference to Behrend's construction, which was the last constituent we needed for the present work. Victor Pan referred us to Salem and Spencer. Arnold Schönhage gave a more symmetric presentation of our starting algorithm in Section 6. James Davenport offered helpful comments on an early draft of the paper.

2. Schönhage's Theorem

The basic results from "classical" matrix multiplication can be summarized by Schönhage's τ-theorem:

Theorem (Schönhage): *Assume given a field F, coefficients $\alpha_{i,j,h,\ell}$, $\beta_{j,k,h,\ell}$, $\gamma_{k,i,h,\ell}$ in $F(\lambda)$ (the field of rational functions in a single indeterminate λ), and polynomials f_g over F, such that*

$$\sum_{\ell=1}^{L}\left(\sum_{i,j,h}\alpha_{i,j,h,\ell}x_{i,j}^{(h)}\right)\left(\sum_{j,k,h}\beta_{j,k,h,\ell}y_{j,k}^{(h)}\right)\left(\sum_{k,i,h}\gamma_{k,i,h,\ell}z_{k,i}^{(h)}\right)$$
$$=\sum_{h}\left(\sum_{i=1}^{m_h}\sum_{j=1}^{n_h}\sum_{k=1}^{p_h}x_{i,j}^{(h)}y_{j,k}^{(h)}z_{k,i}^{(h)}\right)+\sum_{g>0}\lambda^{g}f_g\left(x_{i,j}^{(h)},y_{j,k}^{(h)},z_{k,i}^{(h)}\right)$$

is an identity in $x_{i,j}^{(h)}, y_{j,k}^{(h)}, z_{k,i}^{(h)}, \lambda$. Then given $\varepsilon>0$, one can construct an algorithm to multiply $N\times N$ square matrices in $O(N^{3\tau+\varepsilon})$ operations, where τ satisfies

$$L=\sum_{h}(m_hn_hp_h)^{\tau}. \tag{1}$$

We will also write the error term as $O(\lambda)$, so that the hypothesis becomes

$$\sum_{\ell=1}^{L}\left(\sum_{i,j,h}\alpha_{i,j,h,\ell}x_{i,j}^{(h)}\right)\left(\sum_{j,k,h}\beta_{j,k,h,\ell}y_{j,k}^{(h)}\right)\left(\sum_{k,i,h}\gamma_{k,i,h,\ell}z_{k,i}^{(h)}\right)=\sum_{h}\left(\sum_{i,j,k}x_{i,j}^{(h)}y_{j,k}^{(h)}z_{k,i}^{(h)}\right)+O(\lambda).$$

Less formally, the hypothesis is a trilinear algorithm, using L bilinear multiplications to (approximately) compute simultaneously several independent matrix products, of dimension $m_h\times n_h$ times $n_h\times p_h$ (written $<m_h,n_h,p_h>$). The superscript (h) indicates which matrix the variable belongs to.

In such a situation, we define the **matrix exponent** obtained from the construction as $\omega=3\tau$.

Note: Here we have presented Schönhage's Theorem in its *trilinear* form, which will make our construction easier to describe. The *bilinear* version assumes L rank-1 bilinear forms

$$M_\ell \equiv \left(\sum_{i,j,h} \alpha_{i,j,h,\ell} x_{i,j}^{(h)}\right)\left(\sum_{j,k,h} \beta_{j,k,h,\ell} y_{j,k}^{(h)}\right), \quad \ell = 1, 2, \ldots, L$$

connected by identities

$$v_{i,k}^{(h)} \equiv \sum_{j=1}^{n_h} x_{i,j}^{(h)} y_{j,k}^{(h)} = \sum_{\ell=1}^{L} \gamma_{k,i,h,\ell} M_\ell + O(\lambda), \quad i \le m_h,\ k \le p_h.$$

That is, one forms L bilinear products, each of which is a linear combination of x-variables times a linear combination of y-variables, and expresses the answers $v_{i,k}^{(h)}$ as linear combinations of these products, up to terms of order λ. The *answers* $v_{i,k}^{(h)}$ are viewed as duals to the variables $z_{k,i}^{(h)}$ in the trilinear presentation, and in fact the bilinear presentation is obtained from the trilinear one by identifying coefficients of $z_{k,i}^{(h)}$ in both sides of the equation.

3. Strassen's construction

Strassen has found a new relaxation of the ground rules for the construction of the basic algorithm, that is, he has relaxed the hypotheses of the theorem. A key element in his construction is the observation that, using the ability to multiply a pair of $N \times N$ matrices, one can "approximately" (in the λ sense) multiply $(3/4)N^2$ pairs of independent scalars, that is, compute

$$\sum_{i=1}^{(3/4)N^2} x_i y_i z_i + O(\lambda) \tag{2}$$

where all the x_i, y_i, z_i are independent. Namely, setting

$$g = [(3/2)(N+1)],$$

and multiplying each variable in

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} x_{i,j} y_{j,k} z_{k,i}$$

by an appropriate power of λ, one obtains

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} \left(x_{i,j}\lambda^{i^2+2ij}\right)\left(y_{j,k}\lambda^{j^2+2j(k-g)}\right)\left(z_{k,i}\lambda^{(k-g)^2+2(k-g)i}\right) = \sum_{\substack{i+j+k=g \\ 1\le i,j,k\le N}} x_{i,j} y_{j,k} z_{k,i} + O(\lambda),$$

since the exponent of λ,

$$i^2 + 2ij + j^2 + 2j(k-$$

is zero when $i+j+k=g$ a
the third $k=g-i-j$, each
$\left[(3/4)N^2\right]$ triples (i,j,k), $1 \le$

Strassen uses the foll
aggregation" (1978):

$$\sum_{i=1}^{q}\left(x_0^{[2]} + \lambda x_i^{[1]}\right)\left(y_0^{[1]}\right.$$
$$= \sum_{i=1}^{q}\left(x_i^{[1]} y_0^{[1]} z_i + x_0^{[2]}\right.$$

This gives a basic algorithm

$$\sum_{i=1}^{q}\left(x_i^{[1]} y_0^{[1]} z_i + x_0^{[2]}\right.$$

The superscripts denote in
subscript indices. We can
variables; similarly y_i and
This is the new complicati
a matrix product.

(If we tried to represen
variables z_i are associated
But since they are all assoc
dex.)

The first block, $\sum_i x_i^{[1]}$
(column vector) x is multi
v, which is dual to the vect
size $< 1, 1, q >$. A 1×1 m
a $1 \times q$ matrix (*row* vector
try to add the two blocks

since the exponent of λ,

$$i^2 + 2ij + j^2 + 2j(k-g) + (k-g)^2 + 2(k-g)i = (i+j+k-g)^2,$$

is zero when $i+j+k=g$ and is positive otherwise. Since any two indices i,j uniquely determine the third $k=g-i-j$, each variable $x_{i,j}$ is involved in at most one product. There are about $\left[(3/4)N^2\right]$ triples (i,j,k), $1 \le i,j,k \le N$, $i+j+k=g$. Call this the *matrix-to-scalar* construction.

Strassen uses the following basic trilinear identity, related to Victor Pan's "trilinear aggregation" (1978):

$$\sum_{i=1}^{q}\left(x_0^{[2]} + \lambda x_i^{[1]}\right)\left(y_0^{[1]} + \lambda y_i^{[2]}\right)\left(z_i\lambda^{-1}\right) + \left(x_0^{[2]}\right)\left(y_0^{[1]}\right)\left(-\sum_{i=1}^{q} z_i\lambda^{-1}\right) = \\ = \sum_{i=1}^{q}\left(x_i^{[1]}y_0^{[1]}z_i + x_0^{[2]}y_i^{[2]}z_i\right) + O(\lambda). \tag{3}$$

This gives a basic algorithm, using $q+1$ multiplications to compute a block inner product:

$$\sum_{i=1}^{q}\left(x_i^{[1]}y_0^{[1]}z_i + x_0^{[2]}y_i^{[2]}z_i\right) + O(\lambda).$$

The superscripts denote indices in the block inner product, and are uniquely determined by the subscript indices. We can label x_i and x_0 with different superscripts because they are different variables; similarly y_i and y_0. But the z-variables are involved in both blocks. They are shared. This is the new complication in the basic algorithm. This algorithm does not in itself represent a matrix product.

(If we tried to represent this algorithm as a matrix product, we would find that since all the variables z_i are associated with the same y-variable $y_0^{[1]}$, they must all have the same k-index. But since they are all associated with the same x-variable $x_0^{[2]}$, they must all have the same i-index.)

The first block, $\sum_i x_i^{[1]}y_0^{[1]}z_i$, represents a matrix product of size $< q, 1, 1 >$. A $q \times 1$ matrix (column vector) x is multiplied by a 1×1 matrix (scalar) y_0 to yield a $q \times 1$ matrix (*column* vector) v, which is dual to the vector z. In the second block, $\sum_i x_0^{[2]}y_i^{[2]}z_i$ represents a matrix product of size $< 1, 1, q >$. A 1×1 matrix (scalar) x_0 is multiplied by a $1 \times q$ matrix (row vector) y to yield a $1 \times q$ matrix (*row* vector) v, which is again dual to the vector z. The difficulty comes when we try to add the two blocks. The indices i of v are "schizophrenic": they don't know whether to

behave as row indices or as column indices. Strassen's construction gives a way out of this difficulty.

Take Construction (3) and the two constructions gotten by cyclic permutations of the variables x,y,z, and tensor them together, to get an algorithm requiring $(q+1)^3$ multiplications to compute

$$\sum_{i,j,k=1}^{q} \left(x_{i,j,0}^{[1,1]} y_{0,j,k}^{[1,1]} z_{i,0,k}^{[1,1]} + x_{i,j,k}^{[2,1]} y_{0,j,k}^{[1,1]} z_{i,0,0}^{[1,2]} + x_{i,j,0}^{[1,1]} y_{0,0,k}^{[1,2]} z_{i,j,k}^{[2,1]} + x_{i,j,k}^{[2,1]} y_{0,0,k}^{[1,2]} z_{i,j,0}^{[2,2]} + \right.$$
$$\left. + x_{0,j,0}^{[1,2]} y_{i,j,k}^{[2,1]} z_{i,0,k}^{[1,1]} + x_{0,j,k}^{[2,2]} y_{i,j,k}^{[2,1]} z_{i,0,0}^{[1,2]} + x_{0,j,0}^{[1,2]} y_{i,0,k}^{[2,2]} z_{i,j,k}^{[2,1]} + x_{0,j,k}^{[2,2]} y_{i,0,k}^{[2,2]} z_{i,j,0}^{[2,2]} \right) + O(\lambda).$$

This is a block 2×2 matrix product (indicated by the superscripts). Within each block is a smaller matrix product; for example the third block is the matrix product $\sum_{i,j,k=1}^{q} (x_{i,j,0}^{[1,1]} y_{0,0,k}^{[1,2]} z_{i,j,k}^{[2,1]})$, which can be interpreted as a matrix product of size $< q^2, 1, q >$:

$$\sum_{i,j,k=1}^{q} x_{(i,j),0} y_{0,k} z_{k,(i,j)},$$

with (i,j) acting as the I-index (shared by x and z and taking on q^2 values), 0 acting as the J-index, and k acting as the K-index.

Taking the N^{th} tensor power of this algorithm, one gets an algorithm, requiring $(q+1)^{3N}$ multiplications, and producing a block $2^N \times 2^N$ matrix product, each block of which is a matrix product of some size $< m,n,p >$ where $mnp = q^{3N}$. Applying the matrix-to-scalar construction to the block structure, one then obtains $(3/4)(2^N)^2$ independent matrix products, each of some size $< m,n,p >$ where $mnp = q^{3N}$. Applying the τ-theorem, one gets

$$\omega \leq 3\tau_N, \qquad (q+1)^{3N} = (3/4)2^{2N}(q^{3N})^{\tau_N}.$$

Taking N^{th} roots and letting N grow, the (3/4) becomes insignificant, and we have

$$\omega \leq 3\tau, \quad (q+1)^3 = 2^2 q^{3\tau}.$$

Letting $q = 5$, Strassen obtains

$$\omega \leq \log(6^3/2^2)/\log 5 = \log_5 54 \simeq 2.4785.$$

We will use the following t

Theorem (Salem and S

there is a set B of $M' > M$

$$0 < b_1 < b_2 < \cdots < b_M$$

with no three terms in an a

for $b_i, b_j, b_k \in B$,

We will be considering
ments of the Salem-Spence
mod M:

for $b_i, b_j, b_k \in B$,

Previous authors in this fie
hashing and counting argu
modification of Strassen's
the $3N^{th}$ tensor power, yi
about $(27/4)^N$ out of these
of variables $X^{[I]}$, $Y^{[J]}$, o
independent in the sense t
which is *not* in our chosen
absent from our chosen bl
block products, we will set
done that, we will use Sch

By indirect arguments
products. We will start w
mod M ($\mathbb{Z}_M$), in such a wa
$b_X(I)$, $b_Y(J)$, $b_Z(K)$ form
theorem to control the exi

4. The Salem-Spencer Theorem

We will use the following theorem of Salem and Spencer (1942); see also Behrend (1946).

Theorem (Salem and Spencer): *Given $\varepsilon > 0$, there exists $M_\varepsilon \simeq 2^{c/\varepsilon^2}$ such that for all $M > M_\varepsilon$, there is a set B of $M' > M^{1-\varepsilon}$ distinct integers*

$$0 < b_1 < b_2 < \cdots < b_{M'} < M/2$$

with no three terms in an arithmetic progression:

$$\text{for } b_i, b_j, b_k \in B, \qquad b_i + b_j = 2b_k \quad \text{iff} \quad b_i = b_j = b_k.$$

We will be considering the ring $\mathbb{Z}_M$ of integers modulo M, where M is odd. Because the elements of the Salem-Spencer set satisfy $0 < b_i < M/2$, no three can form an arithmetic progression mod M:

$$\text{for } b_i, b_j, b_k \in B, \qquad b_i + b_j \equiv 2b_k \bmod M \quad \text{iff} \quad b_i = b_j = b_k. \tag{4}$$

5. New Construction: Outline

Previous authors in this field have exhibited their algorithms directly, but we will have to rely on hashing and counting arguments to show the existence of a suitable algorithm. We start with a modification of Strassen's starting algorithm, producing 3 (instead of 2) block products. We take the $3N^{th}$ tensor power, yielding $3^{3N} = 27^N$ block products. We will show that we can choose about $(27/4)^N$ out of these 27^N products, which are *independent* in the sense that a given block of variables $X^{[I]}$, $Y^{[J]}$, or $Z^{[K]}$ will occur in at most one of our chosen products, and *strongly independent* in the sense that for any block product $X^{[I]}Y^{[J]}Z^{[K]}$ in the original tensor power which is *not* in our chosen set, at least one of its blocks of variables ($X^{[I]}$, $Y^{[J]}$ or $Z^{[K]}$) will be absent from our chosen block products. Thus, by setting to zero any variables not in our chosen block products, we will set to zero all other products in the original tensor power. Once we have done that, we will use Schönhage's τ-theorem to provide an estimate of ω.

By indirect arguments we will show the existence of the strongly independent set of block products. We will start with a hash function from the set of block indices to the set of integers mod M ($\mathbb{Z}_M$), in such a way that if $X^{[I]}Y^{[J]}Z^{[K]}$ appears in the tensor power, the hashed indices $b_X(I)$, $b_Y(J)$, $b_Z(K)$ form a three-term arithmetic progression. We will use the Salem-Spencer theorem to control the existence of such arithmetic progressions. Finally, among the large class

of hash functions available, a counting argument will show that at least one of them has the desired performance.

6. New Construction: Easy Case

Start with a modification of Strassen's basic algorithm (3); see also Pan (1978). We use $q+2$ multiplications:

$$\begin{aligned}
&\sum_{i=1}^{q} \lambda^{-2}\left(x_0^{[0]} + \lambda x_i^{[1]}\right)\left(y_0^{[0]} + \lambda y_i^{[1]}\right)\left(z_0^{[0]} + \lambda z_i^{[1]}\right) \\
&- \lambda^{-3}\left(x_0^{[0]} + \lambda^2 \sum x_i^{[1]}\right)\left(y_0^{[0]} + \lambda^2 \sum y_i^{[1]}\right)\left(z_0^{[0]} + \lambda^2 \sum z_i^{[1]}\right) \\
&+ \left(\lambda^{-3} - q\lambda^{-2}\right)\left(x_0^{[0]}\right)\left(y_0^{[0]}\right)\left(z_0^{[0]}\right) \\
&= \sum_{i=1}^{q}\left(x_0^{[0]} y_i^{[1]} z_i^{[1]} + x_i^{[1]} y_0^{[0]} z_i^{[1]} + x_i^{[1]} y_i^{[1]} z_0^{[0]}\right) + O(\lambda).
\end{aligned} \tag{5}$$

We have brought the factors λ^{-3}, $\left(\lambda^{-3} - q\lambda^{-2}\right)$ outside in order to reflect the symmetry.

Note. This is equivalent to the *bilinear* algorithm

$$\begin{aligned}
M_i &= \left(x_0^{[0]} + \lambda x_i^{[1]}\right)\left(y_0^{[0]} + \lambda y_i^{[1]}\right), \qquad i = 1, 2, \ldots, q \\
M_{q+1} &= \left(x_0^{[0]} + \lambda^2 \sum x_i^{[1]}\right)\left(y_0^{[0]} + \lambda^2 \sum y_i^{[1]}\right) \\
M_{q+2} &= \left(x_0^{[0]}\right)\left(y_0^{[0]}\right) \\
v_i^{[1]} &\equiv x_0^{[0]} y_i^{[1]} + x_i^{[1]} y_0^{[0]} = \lambda^{-1} M_i - \lambda^{-1} M_{q+1} + O(\lambda), \qquad i = 1, 2, \ldots, q \\
v_0^{[0]} &\equiv \sum_{i=1}^{q} x_i^{[1]} y_i^{[1]} = \sum_{i=1}^{q} \lambda^{-2} M_i - \lambda^{-3} M_{q+1} + \left(\lambda^{-3} - q\lambda^{-2}\right) M_{q+2} + O(\lambda)
\end{aligned}$$

where $v_j^{[J]}$ is the dual to the variable $z_j^{[J]}$, and the equivalence is gotten by identifying coefficients of $z_j^{[J]}$ in both sides of (5).

The x-variables break into two blocks: $X^{[0]} = \{x_0^{[0]}\}$ and $X^{[1]} = \{x_1^{[1]}, \ldots, x_q^{[1]}\}$. Similarly the y-variables break into blocks $Y^{[0]}$ and $Y^{[1]}$, and the z-variables into blocks $Z^{[0]}$ and $Z^{[1]}$. When we *zero* a block $X^{[I]}$ (resp. $Y^{[J]}$, $Z^{[K]}$), we will set to zero all x - (resp. y-, z-) variables with the given index pattern.

Fix $\varepsilon > 0$. Select N large enough so that the M defined below will exceed M_ε from the Salem-Spencer Theorem.

Take the $3N^{th}$ tensor po
be the tensor product of $3N$
Its subscript i will be a vect
scripts j. Its superscript $[I]$
$3N$ superscripts $[J]$. As bef

Set to zero all variables
$2N$ indices of 1; similarly y-
B. Select random integers 0
a *hash* as follows. For each
0 or 1). Define

$$b_X(I) \equiv \sum_{j=1}^{3N} I_j w_j \pmod{}$$

$$b_Y(J) \equiv w_0 + \sum_{j=1}^{3N} J_j w_j$$

$$b_Z(K) \equiv \left(w_0 + \sum_{j=1}^{3N} (2$$

Since M is odd, division by

Notice that for any bl
computed trilinear form, w

$$b_X(I) + b_Y(J) - 2b_Z($$

This follows by considerin

$$I_j + J_j + K_j = 2$$

for each of the three terms

Set to zero all blocks
which $b_Y(J)$ is not in B, a
$X^{[I]} Y^{[J]} Z^{[K]}$ remaining

$$b_X(I) + b_Y(J) \equiv 2b_Z$$

so that

Take the $3N^{th}$ tensor power of Construction (5). Each variable $x_i^{[I]}$ in the tensor power will be the tensor product of $3N$ variables $x_j^{[J]}$, one from each of $3N$ copies of the original algorithm. Its subscript i will be a vector of length $3N$ with entries in $\{0, 1, \ldots, q\}$, made up of the $3N$ subscripts j. Its superscript $[I]$ will be a vector of length $3N$ with entries in $\{0, 1\}$, made up of the $3N$ superscripts $[J]$. As before, $[I]$ is uniquely determined by i.

Set to zero all variables $x_i^{[I]}$ except those for which I has exactly N indices of 0 and exactly $2N$ indices of 1; similarly y- and z-variables. Set $M = 2\binom{2N}{N} + 1$. Construct a Salem-Spencer set B. Select random integers $0 \le w_j < M$, $j = 0,1, \ldots, 3N$. For each superscript $I \in \{0, 1\}^{3N}$, compute a *hash* as follows. For each of the $3N$ index positions j, let I_j denote the j^{th} element of I (either 0 or 1). Define

$$b_X(I) \equiv \sum_{j=1}^{3N} I_j w_j \pmod{M}$$

$$b_Y(J) \equiv w_0 + \sum_{j=1}^{3N} J_j w_j \pmod{M}$$

$$b_Z(K) \equiv \left(w_0 + \sum_{j=1}^{3N} (2 - K_j) w_j \right)/2 \pmod{M}.$$

Since M is odd, division by 2 is well defined.

Notice that for any blocks $X^{[I]}$, $Y^{[J]}$, $Z^{[K]}$ whose product $X^{[I]}Y^{[J]}Z^{[K]}$ appears in the computed trilinear form, we have

$$b_X(I) + b_Y(J) - 2b_Z(K) \equiv 0 \mod M. \tag{6}$$

This follows by considering the contribution of each w_j, noticing that in the basic construction

$$I_j + J_j + K_j = 2$$

for each of the three terms $x_0^{[0]}y_i^{[1]}z_i^{[1]}$, $x_i^{[1]}y_0^{[0]}z_i^{[1]}$, $x_i^{[1]}y_i^{[1]}z_0^{[0]}$.

Set to zero all blocks $X^{[I]}$ for which $b_X(I)$ is not in B. Similarly set to zero all blocks $Y^{[J]}$ for which $b_Y(J)$ is not in B, and blocks $Z^{[K]}$ for which $b_Z(K)$ is not in B. Then for any nonzero term $X^{[I]}Y^{[J]}Z^{[K]}$ remaining in our construction, we have

$$b_X(I) + b_Y(J) \equiv 2b_Z(K) \bmod M, \qquad b_X(I), b_Y(J), b_Z(K) \in B,$$

so that

$$b_X(I) = b_Y(J) = b_Z(K),$$

by the properties of B.

For each element $b \in B$ in the Salem-Spencer set, make a list of triples $(X^{[I]}, Y^{[J]}, Z^{[K]})$ of compatible nonzero blocks, with $b_X(I) = b_Y(J) = b_Z(K) = b$. (A *block* $X^{[I]}$ is the set of q^{2N} variables $x_{I'}^{[I]}$ with nonzero indices in $2N$ specified places, that is, sharing a common superscript I. A *nonzero block* is one which has not yet been set to zero. Blocks $X^{[I]}$, $Y^{[J]}$, $Z^{[K]}$ are *compatible* if the locations of their zero indices are pairwise disjoint.) For each triple $(X^{[I]}, Y^{[J]}, Z^{[K]})$ on the list, if it shares a block (say $Z^{[K]}$) with another triple $(X^{[I']}, Y^{[J']}, Z^{[K]})$ occurring earlier in the list, we set to zero one of the other blocks (say $Y^{[J]}$), and thus eliminate this triple. (If each of $X^{[I]}$, $Y^{[J]}$, $Z^{[K]}$ is shared with previous triples, we will end up eliminating at least two triples by zeroing one block of variables.)

For a fixed element $b \in B$, the expected number of triples in the list, before pruning, is

$$\binom{3N}{N,N,N} M^{-2}.$$

Here $\binom{3N}{N,N,N}$ represents the number of compatible triples $(X^{[I]}, Y^{[J]}, Z^{[K]})$ and the M^{-2} represents the probability of the (independent) events $b_X(I) = b$ and $b_Y(J) = b$. (If both hold, then $b_Z(K) = b$ follows.) That is, for fixed blocks $X^{[I]}$, $Y^{[J]}$, and fixed integer $b \bmod M$, if we randomize the values $w_0, w_1, \ldots, w_{3N}$, then

$$\text{Prob}\{b_X(I) = b_Y(J) = b\} = \text{Prob}\{b_X(I) = b\}\text{Prob}\{b_Y(J) = b\} = M^{-1}M^{-1} = M^{-2},$$

since the sums $b_X(I)$ and $b_Y(J)$ involve different random variables. The expected number of compatible triples $(X^{[I]}, Y^{[J]}, Z^{[K]})$ with $b_X(I) = b_Y(J) = b_Z(K) = b$ is the sum of these probabilities (M^{-2}) over the $\binom{3N}{N,N,N}$ possible triples. We do not need independence among triples, since the expected value of a sum of random variables is the sum of their expected values, regardless of independence.

The expected number of unordered pairs $(X^{[I]}, Y^{[J]}, Z^{[K]})$, $(X^{[I']}, Y^{[J']}, Z^{[K]})$ sharing a Z-block is

$$(1/2)\binom{3N}{N,N,N}\left(\binom{2N}{N,N} - 1\right)M^{-3}.$$

Again $\binom{3N}{N,N,N}$ counts th
$\binom{2N}{N,N} - 1$ counts the block
1/2 eliminates duplica
$((X^{[I']}, Y^{[J']}, Z^{[K]}), (X^{[I]},$
events $b_Z(K) = b$, $b_Y(J) = b$,
because of the presence of th

The expected number of

Suppose we elimina
$((X^{[I]}, Y^{[J]}, Z^{[K]}), (X^{[I']},$
share this Y-block, then sett
$\binom{L}{2} + 1$ pairs, namely all
$\binom{L}{2} + 1 \geq L$, we eliminate a

Lemma. *The expected n*

$$\binom{3N}{N,N,N} M^{-2} - (3/2$$

$$\geq (1/4)\binom{3N}{N,N,N} M^{-}$$

The expected number o

$$H \equiv (1/4)M'\binom{3N}{N,N,N}$$

This expectation H is an a
at least H; fix such a choic

Our algorithm comput
of each block scalar produ

$$< q^N, q^N, q^N >,$$

and all the variables are dis

$$\omega \leq 3\tau_N, \quad (q+2)^{3N}$$

Use Stirling's approximati

Again $\binom{3N}{N,N,N}$ counts the compatible triples $(X^{[I]}, Y^{[J]}, Z^{[K]})$. The binomial coefficient $\binom{2N}{N,N}-1$ counts the blocks $Y^{[J']}$ compatible with $Z^{[K]}$ (other than $Y^{[J]}$ itself). The factor 1/2 eliminates duplicate entries $((X^{[I]}, Y^{[J]}, Z^{[K]}), (X^{[I']}, Y^{[J']}, Z^{[K]}))$ and $((X^{[I']}, Y^{[J']}, Z^{[K]}), (X^{[I]}, Y^{[J]}, Z^{[K]}))$. The factor M^{-3} is the probability of the independent events $b_Z(K) = b$, $b_Y(J) = b$, $b_Y(J') = b$. They are independent even if indices are equal ($J' = I$), because of the presence of the random variable w_0.

The expected number of pairs of triples sharing $Y^{[J]}$, or sharing $X^{[I]}$, is the same.

Suppose we eliminate a block $(Y^{[J]})$ because of a pair of triples $((X^{[I]}, Y^{[J]}, Z^{[K]}), (X^{[I']}, Y^{[J']}, Z^{[K]}))$ sharing a Z-block. If L triples (not yet eliminated) share this Y-block, then setting $Y^{[J]}$ to zero eliminates these L triples, while eliminating at least $\binom{L}{2}+1$ pairs, namely all those sharing $Y^{[J]}$, and at least the pair sharing $Z^{[K]}$. Since $\binom{L}{2}+1 \geq L$, we eliminate at least as many pairs as triples. Thus:

Lemma. *The expected number of triples remaining on each list, after pruning, is at least*

$$\binom{3N}{N,N,N}M^{-2} - (3/2)\binom{3N}{N,N,N}\left[\binom{2N}{N,N}-1\right]M^{-3} \geq (1/4)\binom{3N}{N,N,N}M^{-2}. \tag{7}$$

The expected number of triples remaining on all lists, after pruning, is at least

$$H \equiv (1/4)M'\binom{3N}{N,N,N}M^{-2}. \tag{8}$$

This expectation H is an average over the choices of w_j. There is a choice of w_j which achieves at least H; fix such a choice.

Our algorithm computes at least H block scalar products $X^{[I]}Y^{[J]}Z^{[K]}$. The fine structure of each block scalar product is in fact a matrix product of size

$$< q^N, q^N, q^N >,$$

and all the variables are disjoint (by the Salem-Spencer property). From the τ-theorem we obtain

$$\omega \leq 3\tau_N, \quad (q+2)^{3N} \geq (1/4)M'\binom{3N}{N,N,N}M^{-2}q^{3N\tau_N}.$$

Use Stirling's approximation to obtain

$$(q+2)^{3N} \geq cN^{-1/2+\varepsilon} 3^{3N} 2^{-2N(1+\varepsilon)} q^{3N\tau_N},$$

where c is a constant. Letting ε go to zero and N to infinity, and taking N^{th} roots, we obtain

$$(q+2)^3 \geq (3^3/2^2) q^{3\tau}$$
$$\omega \leq 3\tau \leq \log_q\left(\frac{4(q+2)^3}{27}\right).$$

Setting $q = 8$ we obtain

$$\omega \leq \log_8(4000/27) < 2.40364. \tag{9}$$

7. New Construction: Complicated Version.

In this section and the sequel, we will improve the exponent to 2.388 and then 2.376, by using the same ideas as in the previous section, on more complicated starting algorithms.

Begin with the basic algorithm:

$$\begin{aligned}
&\sum_{i=1}^{q} \lambda^{-2}\left(x_0^{[0]} + \lambda x_i^{[1]}\right)\left(y_0^{[0]} + \lambda y_i^{[1]}\right)\left(z_0^{[0]} + \lambda z_i^{[1]}\right) \\
&- \lambda^{-3}\left(x_0^{[0]} + \lambda^2 \sum x_i^{[1]}\right)\left(y_0^{[0]} + \lambda^2 \sum y_i^{[1]}\right)\left(z_0^{[0]} + \lambda^2 \sum z_i^{[1]}\right) \\
&+ \left[\lambda^{-3} - q\lambda^{-2}\right]\left(x_0^{[0]} + \lambda^3 x_{q+1}^{[2]}\right)\left(y_0^{[0]} + \lambda^3 y_{q+1}^{[2]}\right)\left(z_0^{[0]} + \lambda^3 z_{q+1}^{[2]}\right) \\
&= \sum_{i=1}^{q}\left(x_0^{[0]} y_i^{[1]} z_i^{[1]} + x_i^{[1]} y_0^{[0]} z_i^{[1]} + x_i^{[1]} y_i^{[1]} z_0^{[0]}\right) + \\
&x_0^{[0]} y_0^{[0]} z_{q+1}^{[2]} + x_0^{[0]} y_{q+1}^{[2]} z_0^{[0]} + x_{q+1}^{[2]} y_0^{[0]} z_0^{[0]} + O(\lambda).
\end{aligned} \tag{10}$$

The subscripts now form three classes: $\{0\}$, $\{q+1\}$, and $\{1, 2, \ldots, q\}$, which will again be denoted i. Again the subscripts uniquely determine the superscripts (block indices).

Take the $3N^{th}$ power of this construction. Set $L = [\beta N]$ (greatest integer), where β will be determined later. Set to zero all blocks of variables except those whose superscripts contain exactly $N + L$ indices of 0, $2N - 2L$ indices of 1, and L indices of 2.

Set

$$M = 2\binom{N+L}{L,L,N-L}\binom{2N-2L}{N-L,N-L} + 1.$$

Let I_j pick out the j^{th} index of I as before. Define $b_X(I)$, $b_Y(J)$, $b_Z(K)$ as before, and set to zero any block of variables with $b_X(I)$ (resp. $b_Y(J)$, $b_Z(K)$) not in the Salem-Spencer set. For each b in the Salem-Spencer set, make a list of triples $(X^{[I]}, Y^{[J]}, Z^{[K]})$ of blocks, with $b_X(I) = b_Y(J) = b_Z(K) = b$, and eliminate entries with duplicated blocks.

For a given block $Z^{[K]}$, the number of pairs of blocks $(X^{[I]}, Y^{[J]})$ compatible with $Z^{[K]}$ is

$$\binom{N+L}{L,L,N-L}\binom{2N-2L}{N-L,N-L},$$

since the $N+L$ indices of 0 in K correspond to L instances of (0 in I, 2 in J), L instances of (2 in I, 0 in J), and $N-L$ instances of (1 in I, 1 in J); the $2N-2L$ indices of 1 in K correspond to $N-L$ instances of (1 in I, 0 in J) and $N-L$ instances of (0 in I, 1 in J); and the L instances of 2 in K all have (0 in I, 0 in J). Since M is twice this size, the elimination of duplicates proceeds as before and leaves a constant fraction of the triples intact.

We have M' lists, each with (on average) at least

$$(1/4)\binom{3N}{L,L,L,N-L,N-L,N-L}M^{-2}$$

entries, all having independent variables. (The multinomial coefficient indicates that there are L instances of (2,0,0) as $(x-, y-, z-)$indices, L of (0,2,0), L of (0,0,2), $N-L$ of (1,1,0), etc.). Each entry corresponds to a matrix product of size

$$< q^{N-L}, q^{N-L}, q^{N-L} >.$$

Thus our equation is

$$(q+2)^{3N} \geq (1/4)M'\binom{3N}{L,L,L,N-L,N-L,N-L}M^{-2}q^{3(N-L)\tau_N}$$

$$\simeq cN^{(-1+3\varepsilon/2)}\left[\frac{27}{\beta^{\beta}(1+\beta)^{1+\beta}(2-2\beta)^{2-2\beta}}\right]^N q^{3N(1-\beta)\tau_N c'\varepsilon N}.$$

Letting ε tend to zero and N to infinity, and taking N^{th} roots, we get

$$(q+2)^3 \geq \frac{27}{\beta^{\beta}(1+\beta)^{1+\beta}(2-2\beta)^{2-2\beta}} q^{3(1-\beta)\tau}.$$

For $q = 6$, $\beta = 0.048$, we find

$$\omega \leq 3\tau < 2.38719.$$

8. Coupling the Weights

So far we have assumed that the weights w_j are independent random variables. In this section we will make them dependent: essentially $w_{2j-1} = w_{2j}$. One consequence is that the randomness arguments need to be redone. The advantage is that we are able to gain higher estimates of the "value" of various pieces.

It will be convenient to have a notion of the "value" of a trilinear form (or trilinear algorithm). We define the "value" V_τ of a trilinear form A in terms of the matrix products it can simulate. Suppose the matrix exponent is $\omega = 3\tau$. Suppose that a tensor power of A can be reduced (by substitution of variables) to the approximate computation of several matrix products:

$$A^{\otimes N} \xrightarrow{\lambda} \bigoplus_h < m_h, n_h, p_h > .$$

Then we say

$$V_\tau(A) \geq \left(\sum_h (m_h n_h p_h)^\tau \right)^{1/N}.$$

We also symmetrize: if π is the cyclic permutation of the variables x,y,z in a trilinear form A, then we define

$$V_\tau(A) = (V_\tau(A \otimes \pi A \otimes \pi^2 A))^{1/3}.$$

It is immediate that "value" is *super-multiplicative:*

$$V_\tau(A \otimes B) \geq V_\tau(A) \times V_\tau(B),$$

where "$A \otimes B$" indicates the tensor product. "Valuc" is also *super-additive:*

$$V_\tau(A \oplus B) \geq V_\tau(A) + V_\tau(B),$$

where "$A \oplus B$" indicates that A and B share no variables. If A reduces to B then

$$V_\tau(A) \geq V_\tau(B).$$

For those familiar with Strassen's paper (1986), this notion of value V_τ is intermediate between Strassen's $\underset{\sim}{Q}$ and his $\underset{\sim}{R}$.

We will use this notion to analyze a more complicated version of our present construction, which will yield the exponent of 2.376.

We start with the tensor

variables):

$$(q+2)^2 \to \sum_{i,k=1}^{q} \left(x_{0,0}^{[0]}$$

$$+ x$$

$$+ \sum_{i=1}^{q} \left(x_{0,q}^{[2]}$$

$$+ x$$

$$+ x$$

$$+ \sum_{k=1}^{q} \left(x_{q+}^{[2]}$$

$$+ x$$

$$+ x$$

$$+ \left(x_{q+}^{[4]}$$

$$+ x$$

$$+ x$$

We have divided the $(q+2$

$$X^{[0]} = \{x_{0,0}^{[0]}\}$$

$$X^{[1]} = \{x_{i,0}^{[1]}, x_{0,k}^{[1]}\}$$

$$X^{[2]} = \{x_{q+1,0}^{[2]}, x_{i,k}^{[2]},$$

$$X^{[3]} = \{x_{q+1,k}^{[3]}, x_{i,q+}^{[3]}$$

$$X^{[4]} = \{x_{q+1,q+1}^{[4]}\}.$$

Here i,k denote indices that

have been 2-vectors: $x_{i,0}^{[1,0]}$

Notice that if $X^{[I]} Y^{[J]}$

$$I + J + K = 4.$$

The trilinear form can be w

We start with the tensor square of Construction (10), but we relabel the superscripts (regroup variables):

$$
\begin{aligned}
(q+2)^2 \to \sum_{i,k=1}^{q} \big(& x_{0,0}^{[0]} y_{i,k}^{[2]} z_{i,k}^{[2]} + x_{0,k}^{[1]} y_{i,0}^{[1]} z_{i,k}^{[2]} + x_{0,k}^{[1]} y_{i,k}^{[2]} z_{i,0}^{[1]} + x_{i,0}^{[1]} y_{0,k}^{[1]} z_{i,k}^{[2]} + x_{i,k}^{[2]} y_{0,0}^{[0]} z_{i,k}^{[2]} + \\
& + x_{i,k}^{[2]} y_{0,k}^{[1]} z_{i,0}^{[1]} + x_{i,0}^{[1]} y_{i,k}^{[2]} z_{0,k}^{[1]} + x_{i,k}^{[2]} y_{i,0}^{[1]} z_{0,k}^{[1]} + x_{i,k}^{[2]} y_{i,k}^{[2]} z_{0,0}^{[0]} \big) \\
+ \sum_{i=1}^{q} \big(& x_{0,q+1}^{[2]} y_{i,0}^{[1]} z_{i,0}^{[1]} + x_{0,0}^{[0]} y_{i,q+1}^{[3]} z_{i,0}^{[1]} + x_{0,0}^{[0]} y_{i,0}^{[1]} z_{i,q+1}^{[3]} + \\
& + x_{i,q+1}^{[3]} y_{0,0}^{[0]} z_{i,0}^{[1]} + x_{i,0}^{[1]} y_{0,q+1}^{[2]} z_{i,0}^{[1]} + x_{i,0}^{[1]} y_{0,0}^{[0]} z_{i,q+1}^{[3]} + \\
& + x_{i,q+1}^{[3]} y_{i,0}^{[1]} z_{0,0}^{[0]} + x_{i,0}^{[1]} y_{i,q+1}^{[3]} z_{0,0}^{[0]} + x_{i,0}^{[1]} y_{i,0}^{[1]} z_{0,q+1}^{[2]} \big) \\
+ \sum_{k=1}^{q} \big(& x_{q+1,0}^{[2]} y_{0,k}^{[1]} z_{0,k}^{[1]} + x_{q+1,k}^{[3]} y_{0,0}^{[0]} z_{0,k}^{[1]} + x_{q+1,k}^{[3]} y_{0,k}^{[1]} z_{0,0}^{[0]} + \\
& + x_{0,0}^{[0]} y_{q+1,k}^{[3]} z_{0,k}^{[1]} + x_{0,k}^{[1]} y_{q+1,0}^{[2]} z_{0,k}^{[1]} + x_{0,k}^{[1]} y_{q+1,k}^{[3]} z_{0,0}^{[0]} + \\
& + x_{0,0}^{[0]} y_{0,k}^{[1]} z_{q+1,k}^{[3]} + x_{0,k}^{[1]} y_{0,0}^{[0]} z_{q+1,k}^{[3]} + x_{0,k}^{[1]} y_{0,k}^{[1]} z_{q+1,0}^{[2]} \big) \\
+ \big(& x_{q+1,q+1}^{[4]} y_{0,0}^{[0]} z_{0,0}^{[0]} + x_{q+1,0}^{[2]} y_{0,q+1}^{[2]} z_{0,0}^{[0]} + x_{q+1,0}^{[2]} y_{0,0}^{[0]} z_{0,q+1}^{[2]} + \\
& + x_{0,q+1}^{[2]} y_{q+1,0}^{[2]} z_{0,0}^{[0]} + x_{0,0}^{[0]} y_{q+1,q+1}^{[4]} z_{0,0}^{[0]} + x_{0,0}^{[0]} y_{q+1,0}^{[2]} z_{0,q+1}^{[2]} + \\
& + x_{0,q+1}^{[2]} y_{0,0}^{[0]} z_{q+1,0}^{[2]} + x_{0,0}^{[0]} y_{0,q+1}^{[2]} z_{q+1,0}^{[2]} + x_{0,0}^{[0]} y_{0,0}^{[0]} z_{q+1,q+1}^{[4]} \big)
\end{aligned}
\tag{11}
$$

We have divided the $(q+2)^2$ x-variables into five blocks:

$$
\begin{aligned}
X^{[0]} &= \{x_{0,0}^{[0]}\} \\
X^{[1]} &= \{x_{i,0}^{[1]}, x_{0,k}^{[1]}\} \\
X^{[2]} &= \{x_{q+1,0}^{[2]}, x_{i,k}^{[2]}, x_{0,q+1}^{[2]}\} \\
X^{[3]} &= \{x_{q+1,k}^{[3]}, x_{i,q+1}^{[3]}\} \\
X^{[4]} &= \{x_{q+1,q+1}^{[4]}\}.
\end{aligned}
$$

Here i,k denote indices that range from 1 to q. In the original tensor square, the superscripts would have been 2-vectors: $x_{i,0}^{[1,0]}$. We have added the two elements to form a single superscript: $x_{i,0}^{[1]}$.

Notice that if $X^{[I]}Y^{[J]}Z^{[K]}$ appears in the trilinear form, then

$$I + J + K = 4.$$

The trilinear form can be written in block form as:

$$\sum_{I+J+K=4} X^{[I]}Y^{[J]}Z^{[K]}.$$

There are four types of terms in this trilinear form.

$$X^{[0]}Y^{[0]}Z^{[4]} = x^{[0]}_{0,0}y^{[0]}_{0,0}z^{[4]}_{q+1,q+1}. \tag{a}$$

This is a matrix product of size $<1,1,1>$, whose "value" is 1. There are three such terms: $X^{[0]}Y^{[0]}Z^{[4]}$, $X^{[0]}Y^{[4]}Z^{[0]}$, $X^{[4]}Y^{[0]}Z^{[0]}$.

$$X^{[0]}Y^{[1]}Z^{[3]} = \sum_{i=1}^{q} x^{[0]}_{0,0}y^{[1]}_{i,0}z^{[3]}_{i,q+1} + \sum_{k=1}^{q} x^{[0]}_{0,0}y^{[1]}_{0,k}z^{[3]}_{q+1,k}. \tag{b}$$

This is a matrix product of size $<1,1,2q>$ (i.e. a scalar $x^{[0]}_{0,0}$ times the vector $<y^{[1]}_{i,0}, y^{[1]}_{0,k}>$). Its "value" is $(2q)^{\tau}$. There are six such terms: $X^{[0]}Y^{[1]}Z^{[3]}$, $X^{[0]}Y^{[3]}Z^{[1]}$, $X^{[1]}Y^{[0]}Z^{[3]}$, $X^{[1]}Y^{[3]}Z^{[0]}$, $X^{[3]}Y^{[0]}Z^{[1]}$, $X^{[3]}Y^{[1]}Z^{[0]}$.

$$X^{[0]}Y^{[2]}Z^{[2]} = x^{[0]}_{0,0}y^{[2]}_{q+1,0}z^{[2]}_{0,q+1} + x^{[0]}_{0,0}y^{[2]}_{0,q+1}z^{[2]}_{q+1,0} + \sum_{i,k=1}^{q} x^{[0]}_{0,0}y^{[2]}_{i,k}z^{[2]}_{i,k}. \tag{c}$$

This is another matrix product, of size $<1,1,q^2+2>$, with "value" $(q^2+2)^{\tau}$. There are three such terms: $X^{[0]}Y^{[2]}Z^{[2]}$, $X^{[2]}Y^{[0]}Z^{[2]}$, $X^{[2]}Y^{[2]}Z^{[0]}$.

$$\begin{aligned} X^{[1]}Y^{[1]}Z^{[2]} = &\sum_{i=1}^{q} x^{[1]}_{i,0}y^{[1]}_{i,0}z^{[2]}_{0,q+1} + \sum_{k=1}^{q} x^{[1]}_{0,k}y^{[1]}_{0,k}z^{[2]}_{q+1,0} + \\ &+ \sum_{i,k=1}^{q} x^{[1]}_{i,0}y^{[1]}_{0,k}z^{[2]}_{i,k} + \sum_{i,k=1}^{q} x^{[1]}_{0,k}y^{[1]}_{i,0}z^{[2]}_{i,k}. \end{aligned} \tag{d}$$

There are three such terms: $X^{[1]}Y^{[1]}Z^{[2]}$, $X^{[1]}Y^{[2]}Z^{[1]}$, $X^{[2]}Y^{[1]}Z^{[1]}$. In the lemma at the end of this section, we will find that, if $q \geq 3$, the "value" of this term is at least $2^{2/3}q^{\tau}(q^{3\tau}+2)^{1/3}$.

Take the N^{th} tensor

α_ℓ, $0 \leq \ell \leq 4$, be positive r

$$\sum_{\ell=0}^{4} \alpha_\ell = 1, \qquad \sum_{\ell=0}^{4} \ell\alpha$$

Let A_ℓ be integer approxim

$$\sum_{\ell=0}^{4} A_\ell = N, \qquad \sum_{\ell=0}^{4} \ell$$

Retain only those blocks of

$$\#\{j \mid 1 \leq j \leq N,\ l_j = \ell\}$$

setting the others to zero, w

Let M'' be the number

We have

$$M'' = \sum_{\{\eta_{\ell,m,n}\}} \frac{\prod_{0\leq$$

where $\{\eta_{\ell,m,n}\}$ range over

$$\sum_{m,n} \eta_{\ell,m,n} = A_\ell$$
$$\sum_{\ell,n} \eta_{\ell,m,n} = A_m$$
$$\sum_{\ell,m} \eta_{\ell,m,n} = A_n$$

and the only nonzero valu

in (12) is maximized at

Take the N''' tensor power of Construction (11), where N is divisible by 3. Let α_ℓ, $0 \le \ell \le 4$, be positive real numbers (determined below) such that

$$\sum_{\ell=0}^{4} \alpha_\ell = 1, \qquad \sum_{\ell=0}^{4} \ell \alpha_\ell = 4/3.$$

Let A_ℓ be integer approximations to $\alpha_\ell N$ such that

$$\sum_{\ell=0}^{4} A_\ell = N, \qquad \sum_{\ell=0}^{4} \ell A_\ell = 4N/3.$$

Retain only those blocks of variables $(X^{[I]}, Y^{[J]}, Z^{[K]})$ such that

$$\#\{j \mid 1 \le j \le N,\ I_j = \ell\} = A_\ell,$$

setting the others to zero, where as before I_j picks out the j^{th} index position.

Let M'' be the number of nonzero triples $(X^{[I]}, Y^{[J]}, Z^{[K]})$ containing a given block $X^{[I]}$. We have

$$M'' = \sum_{\{\eta_{\ell,m,n}\}} \frac{\prod_{0 \le \ell \le 4} A_\ell!}{\prod_{\ell+m+n=4} \eta_{\ell,m,n}!}, \tag{12}$$

where $\{\eta_{\ell,m,n}\}$ range over partitions of N such that

$$\sum_{m,n} \eta_{\ell,m,n} = A_\ell$$
$$\sum_{\ell,n} \eta_{\ell,m,n} = A_m$$
$$\sum_{\ell,m} \eta_{\ell,m,n} = A_n$$

and the only nonzero values of $\eta_{\ell,m,n}$ occur with $\ell + m + n = 4$, $0 \le \ell, m, n \le 4$. The summand in (12) is maximized at

$$
\begin{aligned}
&\eta_{\ell,m,n} = \gamma_{\ell,m,n} \\
&\gamma_{0,0,4} = \gamma_{0,4,0} = \gamma_{4,0,0} = \hat{a} \\
&\gamma_{0,1,3} = \gamma_{0,3,1} = \gamma_{1,0,3} = \gamma_{1,3,0} = \gamma_{3,0,1} = \gamma_{3,1,0} = \hat{b} \\
&\gamma_{0,2,2} = \gamma_{2,0,2} = \gamma_{2,2,0} = \hat{c} \\
&\gamma_{1,1,2} = \gamma_{1,2,1} = \gamma_{2,1,1} = \hat{d} \\
&A_0 = 2\hat{a} + 2\hat{b} + \hat{c} \\
&A_1 = 2\hat{b} + 2\hat{d} \\
&A_2 = 2\hat{c} + \hat{d} \\
&A_3 = 2\hat{b} \\
&A_4 = \hat{a}.
\end{aligned} \tag{13}
$$

In later calculations we will approximate M'' by its largest term, times a polynomial N^p in N.

Set $M = 2M'' + 1$. Construct the Salem-Spencer set. Choose random weights w_j, $0 \le j \le N$. Compute the hash as before, with a minor change (4 replacing 2) in the definition on the z-indices:

$$
\begin{aligned}
b_X(I) &\equiv \sum_{j=1}^{N} I_j w_j \bmod M \\
b_Y(J) &\equiv w_0 + \sum_{j=1}^{N} J_j w_j \bmod M \\
b_Z(K) &\equiv \left[w_0 + \sum_{j=1}^{N} (4 - K_j) w_j \right] /2 \bmod M.
\end{aligned}
$$

Retain only those variables mapping into B, setting others to zero. As before, a nonzero triple $X^{[I]}Y^{[J]}Z^{[K]}$ remaining in the trilinear form will have $b_X(I) = b_Y(J) = b_Z(K) \in B$.

After the usual pruning, we have approximately

$$\binom{N}{A_0, A_1, A_2, A_3, A_4}$$

triples of blocks $(X^{[I]}, Y^{[J]}, Z^{[K]})$ remaining. For a good portion of these triples (at least a fraction N^{-p} of the total) the N indices $j = 1,2,\ldots,N$ will contain about $\gamma_{\ell,m,n}$ instances of $X^{[\ell]}Y^{[m]}Z^{[n]}$. So the value of each triple of blocks is about

$$[1]^{3\hat{a}}[(2q^{\tau})]^{6\hat{b}}[(q^2+2)^{\tau}]^{3\hat{c}}[2^{2/3}q^{\tau}(q^{3\tau}+2)^{1/3}]^{3\hat{d}}$$
$$= (2q)^{6\tau\hat{b}}(q^2+2)^{3\tau\hat{c}}[4q^{3\tau}(q^{3\tau}+2)]^{\hat{d}}.$$

Thus our auxiliary equation is

$$(q+2)^{2N} \geq N^{-p}\binom{N}{A_0, A_1, A_2, A_3, A_4}(2q)^{6\tau\hat{b}}(q^2+2)^{3\tau\hat{c}}[4q^{3\tau}(q^{3\tau}+2)]^{\hat{d}}.$$

We want to choose $\hat{a}$, $\hat{b}$, $\hat{c}$, $\hat{d}$ to maximize the right-hand side, subject to

$$3\hat{a} + 6\hat{b} + 3\hat{c} + 3\hat{d} = N,$$

and with A_ρ defined by (13).

In fact, letting $\hat{a} = \bar{a}N$, $\hat{b} = \bar{b}N$, $\hat{c} = \bar{c}N$, $\hat{d} = \bar{d}N$, letting N grow and taking N^{th} roots, we get

$$(q+2)^2 = \frac{(2q)^{6\tau\bar{b}}(q^2+2)^{3\tau\bar{c}}[4q^{3\tau}(q^{3\tau}+2)]^{\bar{d}}}{(2\bar{a}+2\bar{b}+\bar{c})^{2\bar{a}+2\bar{b}+\bar{c}}(2\bar{b}+2\bar{d})^{2\bar{b}+2\bar{d}}(2\bar{c}+\bar{d})^{2\bar{c}+\bar{d}}(2\bar{b})^{2\bar{b}}(\bar{a})^{\bar{a}}}.$$

We wish to minimize τ with respect to $\bar{a}$, $\bar{b}$, $\bar{c}$, $\bar{d}$, q, subject to

$$3\bar{a} + 6\bar{b} + 3\bar{c} + 3\bar{d} = 1$$
$$\bar{a}, \bar{b}, \bar{c}, \bar{d} > 0$$
$$q \geq 3, \quad q \in \mathbb{Z}.$$

We find that

$$\bar{a} = 0.000233$$
$$\bar{b} = 0.012506$$
$$\bar{c} = 0.102546$$
$$\bar{d} = 0.205542$$
$$q = 6$$

gives an exponent of

$$\omega \leq 3\tau < 2.375477.$$

DISCUSSION

Here we essentially set $w_{2j-1} = 2_{2j}$, when we grouped together into one block the variables $x_{i,0}^{[1]}, x_{0,k}^{[1]}$, and similar groupings for the other blocks. Other linear dependences among the w_j remain to be investigated, including:

$$w_{3j-2} = w_{3j-1} = w_{3j}$$

or

$$w_{2j-1} = -2w_{2j}.$$

PROOF OF THE LEMMA

Lemma. *Assume $q \geq 3$. The "value" V_τ of the trilinear form*

$$\sum_{i=1}^{q} x_{i,0}^{[1,0]} y_{i,0}^{[1,0]} z_{0,q+1}^{[0,2]} + \sum_{k=1}^{q} x_{0,k}^{[0,1]} y_{0,k}^{[0,1]} z_{q+1,0}^{[2,0]} + \\ + \sum_{i,k=1}^{q} x_{i,0}^{[1,0]} y_{0,k}^{[0,1]} z_{i,k}^{[1,1]} + \sum_{i,k=1}^{q} x_{0,k}^{[0,1]} y_{i,0}^{[1,0]} z_{i,k}^{[1,1]}$$

is at least $2^{2/3} q^{\tau} (q^{3\tau} + 2)^{1/3}$.

Note: The superscripts are different from those in the application of this lemma, but they are still uniquely determined by the subscripts.

Proof. Take the $2N^{th}$ tensor power. Retain only those X-blocks with exactly N indices of [1,0] and N of [0,1]. Retain only those Y-blocks with exactly N indices of [1,0] and N of [0,1]. Retain those Z-blocks with exactly L indices of [2,0], exactly L of [0,2], and $2G = 2N - 2L$ of [1,1], where

$$L = \left\lfloor \frac{2}{q^{3\tau} + 2} N \right\rfloor \quad \text{and} \quad G = N - L.$$

The number of X-blocks is $\binom{2N}{N}$. The number of nonzero triples $(X^{[I]} Y^{[J]} Z^{[K]})$ containing a given X-block is

$$\binom{L+G}{G}^2.$$

The same numbers hold for

Set

$$M = 4\binom{L+G}{G}^2 + 1.$$

Construct a Salem-Spencer

B. Eliminate instances wher

ples.

There are $\binom{2N}{L,L,2G}$ Z-

Counting only those with b

$$\binom{2G}{G} M^{-1} \simeq \frac{1}{4} \binom{2G}{G}$$

A calculation shows that H

This is, in Strassen's ter

consisting of matrix produc

In short, we have $\binom{2L}{L,L}$

The auxiliary equation is:

$$V_\tau^{2N} \simeq \binom{2N}{L,L,2G} \left[\right.$$

$$\simeq \frac{(2N)^{2N}}{L^L L^L (2G)^{2G}}$$

$$V_\tau^{3N} \simeq \frac{(2N)^{3N}}{L^L (2G)^G N^{2N}}$$

$$\simeq 2^{2N} q^{3N\tau} \binom{N}{L}$$

Selecting

$$\frac{L}{L+G} = \frac{2}{q^{3\tau} + 2},$$

maximizes this estimate at

The same numbers hold for Y-blocks, but not Z-blocks; see below.

Set

$$M = 4\binom{L+G}{G}^2 + 1.$$

Construct a Salem-Spencer set B, define a hash as before, and set to zero blocks not hashing into B. Eliminate instances where an X-block or a Y-block is shared among two or more nonzero triples.

There are $\binom{2N}{L,L,2G}$ Z-blocks, and to each there are initially $\binom{2G}{G}$ triples $(X^{[I]}Y^{[J]}Z^{[K]})$. Counting only those with $b_X(I) = b_Y(J) = b_Z(K)$ we have about

$$\binom{2G}{G}M^{-1} \simeq \frac{1}{4}\binom{2G}{G}\binom{L+G}{G}^{-2} \equiv H$$

A calculation shows that $H >> 1$ if $G/L > 3.41$.

This is, in Strassen's terminology, a C-tensor over $< 1,H,1 >$, where C is the class of tensors consisting of matrix products $< m,n,p >$ with $mnp = (q^2)^{2G}(q)^{2L} = q^{4G+2L}$.

In short, we have $\binom{2L+2G}{L,L,2G}$ disjoint objects, each of which is a C-tensor over $< 1,H,1 >$. The auxiliary equation is:

$$V_\tau^{2N} \simeq \binom{2N}{L,L,2G}\left[\frac{\binom{2G}{G}}{4\binom{N}{G}^2}\right]^{2/3}(q^{4G+2L})^\tau$$

$$\simeq \frac{(2N)^{2N}}{L^L L^L (2G)^{2G}}\left[\frac{(2G)^{2G}}{G^G G^G}\,\frac{G^G L^L G^G L^L}{4N^N N^N}\right]^{2/3} q^{(4G+2L)\tau}$$

$$V_\tau^{3N} \simeq \frac{(2N)^{3N}}{L^L (2G)^G N^{2N}}\, q^{(6G+3L)\tau} \simeq \frac{2^{2N+L}N^N}{L^L G^G}\, q^{3N\tau} q^{3G\tau}$$

$$\simeq 2^{2N} q^{3N\tau}\binom{N}{L} 2^L (q^{3\tau})^G.$$

Selecting

$$\frac{L}{L+G} = \frac{2}{q^{3\tau}+2}, \qquad \frac{G}{L+G} = \frac{q^{3\tau}}{q^{3\tau}+2}$$

maximizes this estimate at

$$2^{2N} q^{3N\tau} (q^{3\tau} + 2)^N$$

and, taking $3N^{th}$ roots with N large,

$$V_\tau \geq 2^{2/3} q^\tau (q^{3\tau} + 2)^{1/3}.$$

Finally, the estimate

$$\frac{G}{L} = \frac{q^{3\tau}}{2} > 3.41$$

is ensured if $q \geq 3$, since $3\tau \geq 2$.

9. Related schemes

We developed a few other techniques on the way to the present paper. While these did not lead to a lower matrix exponent, they may be of interest when applied some other starting algorithm.

ASYMMETRIC VERSION

The first technique we present is an asymmetric version of the technique presented in the previous two sections. We start with an asymmetric basic algorithm, with a block structure and a fine structure. After a suitable tensor power, instead of pulling out several block scalar products (whose fine structure is that of matrix products), we will instead pull out several block vector products. In Strassen's terminology, we will be developing several C-tensors over $< 1,H,1 >$ for some large value of H. This is made possible by the Salem-Spencer Theorem, as before. We then use Strassen's machinery to estimate the value of these C-tensors over $< 1,H,1 >$.

We illustrate with a basic algorithm closely related to Strassen's. It uses $q + 1$ multiplications.

$$\sum_{i=1}^{q} (x_0^{[0]} + \lambda x_i^{[1]})(y_0^{[0]} + \lambda y_i^{[1]})(z_i^{[1]} \lambda^{-1}) + (x_0^{[0]})(y_0^{[0]})(z_{q+1}^{[2]} - \sum z_i^{[1]} \lambda^{-1})$$
$$= \sum_{i=1}^{q} (x_i^{[1]} y_0^{[0]} z_i^{[1]} + x_0^{[0]} y_i^{[1]} z_i^{[1]}) + x_0^{[0]} y_0^{[0]} z_{q+1}^{[2]} + O(\lambda). \tag{14}$$

Later we will determine real numbers $\alpha, \beta > 0$ with $2\alpha + \beta = 1$. Choose ε small and N "large enough". Set

$L = [\alpha N]$, $G = N - 2$

Construct a Salem-Spencer

Take the N^{th} tensor pow

having exactly $G + L$ indice

and $2L$ indices of 1. Select

and set to zero any variabl

$(X^{[I]}, Y^{[J]}, Z^{[K]})$ whose p

$$b_X(I) = b_Y(J) = b_Z(K)$$

Any X-block X_I is comp

correspond to indices of 1 i

dices of 2 in K, and L t

$b_Z(K) = b_X(I)$ is about $1/2$

Y-block. Similarly,

$(X^{[I]}, Y^{[J]}, Z^{[K]}), (X^{[I']},$

is involved in at most one t

We do not make such a

in a large number of triples

$$(X^{[I_h]}, Y^{[J_h]}, Z^{[K]}),$$

For a given Z-block, the e

$$\mu = \binom{2L}{L} M^{-1}.$$

As before, given two blo

$b_X(I) = b_Z(K)$ and $b_X(I') =$

in the number of triples (b

$$\sigma^2 = \binom{2L}{L}(M^{-1} - M$$

Then Chebyshev's inequali

$$\text{Prob}(\#blocks < \mu -$$

$$L = [\alpha N],\ G = N - 2L,\ M = 2\binom{L+G}{L} + 1.$$

Construct a Salem-Spencer set B using this value of M.

Take the N^{th} tensor power of Construction (10). Set to zero all blocks except: X - or Y-blocks having exactly $G + L$ indices of 0 and L indices of 1, and Z-blocks having exactly G indices of 2 and $2L$ indices of 1. Select random integers $w_0, w_1, \ldots, w_N$ mod M, compute the hash as before, and set to zero any variable not hashing to an element of B. Then any nonzero triple of blocks $(X^{[I]}, Y^{[J]}, Z^{[K]})$ whose product appears in the construction has, as before,

$$b_X(I) = b_Y(J) = b_Z(K) \in B.$$

Any X-block X_I is compatible with exactly $\binom{L+G}{L}$ Z-blocks Z_K: the L indices of 1 in I all correspond to indices of 1 in K, while of the $G + L$ indices of 0 in I, exactly G correspond to indices of 2 in K, and L to indices of 1. Thus the expected number of Z-blocks such that $b_Z(K) = b_X(I)$ is about 1/2. If there are more than two such Z-blocks, zero a corresponding Y-block. Similarly, if a Y-block is involved in at least two triples $(X^{[I]}, Y^{[J]}, Z^{[K]}), (X^{[I']}, Y^{[J]}, Z^{[K']})$, we zero one of the X-blocks. Thus each X- or Y-block is involved in at most one triple $(X^{[I]}, Y^{[J]}, Z^{[K]})$ after this pruning.

We do not make such a requirement on the Z-blocks; we allow each Z-block to be involved in a large number of triples:

$$(X^{[I_h]}, Y^{[J_h]}, Z^{[K]}), \quad 1 \le h \le H.$$

For a given Z-block, the expected number of such triples (before pruning) is

$$\mu = \binom{2L}{L} M^{-1}.$$

As before, given two blocks $X^{[I]}, X^{[I']}$ compatible with a given block Z_K, the events that $b_X(I) = b_Z(K)$ and $b_X(I') = b_Z(K)$ are (pairwise) independent, so that we can estimate the variance in the number of triples (before pruning) as

$$\sigma^2 = \binom{2L}{L}(M^{-1} - M^{-2}).$$

Then Chebyshev's inequality can be applied to say that

$$\text{Prob}(\#blocks < \mu - 3\sigma) \le \frac{1}{10}.$$

For a given block $Z^{[K]}$ with $b_Z(K) \in B$, the probability of at least $\mu - 3\sigma$ triples $(X^{[I_h]}, Y^{[J_h]}, Z^{[K]})$ before pruning (i.e. $b_X(I_h) = b_Y(J_h) = b_Z(K)$) is at least 9/10. Pruning eliminates only a constant fraction of the triples. Together with the observation that $\sigma << \mu$, we obtain:

Lemma: *There are constants c_1, c_2 such that, averaging over the choice of w_j, the expected number of blocks $Z^{[K]}$ with at least $c_1\binom{2L}{L}M^{-1}$ associated triples $(X^{[I_h]}, Y^{[J_h]}, Z^{[K]})$ remaining after pruning, is at least $c_2 M' M^{-1}\binom{2L+G}{G}$.*

That is, we will have at least $c_2 M' M^{-1}\binom{2L+G}{G}$ independent objects, each of which is a C-tensor over $< 1, c_1\binom{2L}{L}M^{-1}, 1 >$, where C is the class of matrix multiplication problems $< m, n, p >$ with $mnp = q^{2L}$.

Applying Strassen's formula, we get

$$\omega \le 3\tau_N, \quad (q+1)^N \ge c_2 M' M^{-1}\binom{2L+G}{G}\left[c_1\binom{2L}{L}M^{-1}\right]^{2/3} q^{(2L\tau_N)}.$$

Letting ε shrink and N grow, and taking N^{th} roots,

$$\omega \le 3\tau, \quad (q+1) \ge (2\alpha)^{-2\alpha}\beta^{-\beta}\left[2^{2\alpha}\frac{\alpha^{\alpha}\beta^{\beta}}{(\alpha+\beta)^{\alpha+\beta}}\right]^{2/3} q^{2\alpha\tau}.$$

(As a consistency check, note that the limiting case, $\alpha = 0.5$, $\beta = 0$, recovers Strassen's formula $q + 1 = 2^{2/3} q^{\tau}$, by disallowing terms containing z_{q+1}.) Selecting $q = 4$ and $\alpha = 0.485$, we optimize ω at 2.4602. This exponent is not as good as that obtained in the previous section, but the techniques exposed here may be more widely applicable.

"STRASSEN SQUARED"

Another possibility is to iterate Strassen's construction. Where Strassen develops a C-tensor over $< 1,k,1 >$, where C is a class of matrix products $< m,n,p >$ for a fixed value of mnp, we instead develop a D-tensor over $< 1,1,k >$, where D is a class of C-tensors over $< 1, k', 1 >$, with C again a class of matrix products.

To illustrate, start with Construction (5) from Section 5, using $q + 2$ multiplications to obtain

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

Replace λ by λ^2 throughout. Take the tensor product of this construction with itself; $(q+2)^2$ multiplications yield

$$\sum_{i=1}^{q}\sum_{j=1}^{q}(x_{0,0}y_{i,j}z_{i,j} + x_{0,j}y_{i,0}z_{i,j} + x_{0,j}y_{i,j}z_{i,0} + x_{i,0}y_{0,j}z_{i,j} + x_{i,j}y_{0,0}z_{i,j}$$
$$+ x_{i,j}y_{0,j}z_{i,0} + x_{i,0}y_{i,j}z_{0,j} + x_{i,j}y_{i,0}z_{0,j} + x_{i,j}y_{i,j}z_{0,0}) + O(\lambda^2).$$

Set $x_{0,0}$ to 0. Multiply $x_{i,j}$ by λ, and multiply $y_{0,0}$ and $z_{0,0}$ by λ^{-1}. This kills three terms:

$$x_{0,0}y_{i,j}z_{i,j}, \quad x_{i,j}y_{0,j}z_{i,0}, \quad x_{i,j}y_{i,0}z_{0,j}.$$

We are left with

$$\sum_{i=1}^{q}\sum_{j=1}^{q}[(x_{0,j}y_{i,0}z_{i,j} + x_{i,0}y_{0,j}z_{i,j} + x_{i,j}y_{0,0}z_{i,j}) + (x_{0,j}y_{i,j}z_{i,0} + x_{i,0}y_{i,j}z_{0,j} + x_{i,j}y_{i,j}z_{0,0})] + O(\lambda).$$

The first parenthesized expression has the form of a C-tensor over $<1,3,1>$, where C is the class of matrix products $<m,n,p>$ with $mnp = q^2$; the second is a C-tensor over $<3,1,1>$. Let D denote the class of C-tensors over $<m',n',p'>$ with $m'n'p' = 3$. The two parenthesized expressions share x-variables, and they fit together as a D-tensor over $<1,1,2>$. By Strassen's machinery, we can write:

$$\omega \le 3\tau, \quad (q+2)^2 = 2^{2/3}3^{2/3}q^{2\tau}$$
$$\omega \le \log_q\left(\frac{(q+2)^3}{6}\right).$$

Setting $q = 9$, we get $\omega \le \log_9(11^3/6) < 2.459$. Again, this is not as profitable as the development in Section 6, where the corresponding equation was

$$\omega \le \log_q\left(\frac{(q+2)^3}{6.75}\right).$$

10. Remarks

We are applying the τ-theorem in a special case, namely when all the matrix products are of the same shape. Examining the proof of the τ-theorem, one finds that this eliminates some of the sources of inefficiency for matrices of moderate size. Also, by altering the construction slightly, we can eliminate the use of λ, another contributor to the inefficiency inherent in the τ-theorem.

Nonetheless, the large values of M dictated by the Salem-Spencer theorem still make the present algorithm wildly impractical for any conceivable applications.

A remarkable feature of the present approach is that we can use basic constructions in which the number of x-variables is equal to the number of multiplications, as is the number of y- or z-variables. (In Section 7, this number is $q+2$.) Before Strassen's 1986 paper, this was not possible. Coppersmith and Winograd (1982) had shown that for any basic algorithm (approximately) computing (several) matrix products, the number of multiplications had to strictly exceed the number of x -variables (or else y- or z-variables), except in trivial cases. This had led to the acceleration techniques of Coppersmith and Winograd (1982), but had also implied that one could never achieve $\omega = 2$ by starting with a *fixed, finite-size basic algorithm.* But now Strassen's techniques have removed the necessity of starting with a matrix product, and, coupled with the techniques of the present paper, allow the hope of someday achieving $\omega = 2$. (See the following section.)

An open question is a possible analogue of the Salem-Spencer theorem in characteristic two: For arbitrary $\varepsilon > 0$ and large enough N, do there exist subsets A, B, C of $(\mathbb{Z}_2)^N$ of size $|A| = |B| = |C| = (2^N)^{1-\varepsilon}$, such that each element $a \in A$, $b \in B$, or $c \in C$ is involved in exactly one relation of the form

$$a + b + c = 0?$$

That is,

$$\begin{aligned}
&\forall a \in A \ \exists! b \in B,\ c \in C: \quad a + b + c = 0 \\
&\forall b \in B \ \exists! a \in A,\ c \in C: \quad a + b + c = 0 \\
&\forall c \in C \ \exists! a \in A,\ b \in B: \quad a + b + c = 0
\end{aligned}$$

With such a construction, the present techniques would become applicable to a wider class of starting algorithms. For example, our first algorithm involved three block products, $X^{[0]}Y^{[1]}Z^{[1]}$, $X^{[1]}Y^{[0]}Z^{[1]}$, $X^{[1]}Y^{[1]}Z^{[0]}$, so that if $X^{[I]}Y^{[J]}Z^{[K]}$ appeared we knew $I + J + K = 2$. If in addition the block product $X^{[0]}Y^{[0]}Z^{[0]}$ appeared, our condition on the indices would become $I + J + K = 0 \in \mathbb{Z}_2$. A characteristic-two version of the Salem-Spencer theorem would become useful in that situation.

11. Can We Achieve $\omega = 2$?

In closing, we sketch the relation between a hypothetical combinatorial construction and the elusive

$$\omega \stackrel{?}{=} 2.$$

Definition. An Abelian group G (with at least two elements) and a subset S of G satisfy the *no three disjoint equivoluminous subsets* property if: whenever T_1, T_2, T_3 are three disjoint subsets of S, not all empty, they cannot all have the same sum in G:

$$\sum_{s \in T_1} s \neq \sum_{s \in T_2} s \quad \text{or} \quad \sum_{s \in T_2} s \neq \sum_{s \in T_3} s.$$

Assume for now that we can find a sequence of pairs G,S with the no three disjoint equivoluminous subsets property, such that $(\log_2|G|)/|S|$ approaches 0.

If the trilinear form

$$x_0y_1z_2 + x_0y_2z_1 + x_1y_0z_2 + x_2y_0z_1 + x_1y_2z_0 + x_2y_1z_0$$

had border rank 3, then by the techniques of Chapter 6 (easy case) with $q = 2$, we could prove $\omega = 2$. Unfortunately, this form has border rank 4. But if the n^{th} tensor power of this form had border rank $3^{n+o(n)}$, the same techniques would still yield $\omega = 2$. This is our goal.

We will use a Fourier transform to get a related form to have rank 3. This related form has extra terms, which we must cancel somehow. We will multiply variables by roots of unity, take a large tensor power, and take a Fourier transform again, and use the no three disjoint equivoluminous subsets property to arrange the cancellation of the unwanted terms. This will show that the n^{th} tensor power has rank $3^{n+o(n)}$. Then we will apply the techniques from this paper to derive $\omega = 2$, given the existence of (G,S) with $(\log_2|G|)/|S|$ approaching 0.

Let $S = \{s_1, s_2, ..., s_n\}$ and G satisfy the no three disjoint equivoluminous subsets property, with $\varepsilon = (\log_2|G|)/|S|$.

Let χ and ψ be two independent characters on G. Select i from $\{1, 2, ..., n\}$. Evaluate the characters $\chi(s_i)$, $\psi(s_i)$, which are complex numbers of magnitude 1. Select $r = r(\chi, \psi, s_i)$ to satisfy $r^3 = \chi(s_i)/\psi(s_i)$.

Evaluate the following sum, where ω_i ranges over the three complex cube roots of 1:

$$U_i \equiv \sum_{\omega_i} \frac{1}{3} \left(x_0 + \chi(s_i)^{-1} r\omega_i x_1 + r^{-1}\omega_i^2 x_2\right) \times$$
$$\times \left(\chi(s_i) r^{-1} y_0 + \omega_i y_1 + \psi(s_i) r\omega_i^2 y_2\right) \times$$
$$\times \left(r z_0 + \psi(s_i)^{-1} r^{-1} \omega_i z_1 + \omega_i^2 z_2\right)$$

We see that this has trilinear rank at most 3. By multiplying by powers of ω_i and summing over ω_i, we are taking a three-point Fourier transform, and the terms which survive in the product are those for which the exponent of ω_i is divisible by 3:

$$U_i = (x_0 y_1 z_2 + x_0 y_2 z_1 + x_1 y_0 z_2 + x_2 y_0 z_1 + x_1 y_2 z_0 + x_2 y_1 z_0) +$$
$$+ (\chi(s_i) x_0 y_0 z_0) + (\chi(s_i)^{-1} \psi(s_i)^{-1} x_1 y_1 z_1) + (\psi(s_i) x_2 y_2 z_2).$$

We may think of this as

$$U_i = (\text{Good}_i) + \chi(s_i)\text{Bad}1_i + \chi(s_i)^{-1}\psi(s_i)^{-1}\text{Bad}2_i + \psi(s_i)\text{Bad}3_i,$$

where Good_i is the sum ot the first six terms (which we want), and $\text{Bad}\ell_i$ are the unwanted terms.

Now take the *tensor product* of the various U_i:

$$\bigotimes_{1 \le i \le n} U_i = \bigotimes_i \Big((\text{Good}_i) + \chi(s_i)\text{Bad}1_i + \chi(s_i)^{-1}\psi(s_i)^{-1}\text{Bad}2_i + \psi(s_i)\text{Bad}3_i \Big).$$

This is the sum of 4^n terms, each a tensor product of trilinear forms of the sort Good_i or $\text{Bad}\ell_i$. The first term,

$$\text{GOOD} = \bigotimes_i \text{Good}_i$$

is the one we want. For each of the other $4^n - 1$ terms, for $\ell = 1, 2, 3$, let T_ℓ be the set of s_i corresponding to those indices i for which $\text{Bad}\ell_i$ is included in the product. The coefficient of this term is

$$\left(\prod_{i \in T_1} \chi(s_i)\right)\left(\prod_{i \in T_2} \chi(s_i)^{-1}\psi(s_i)^{-1}\right)\left(\prod_{i \in T_3} \psi(s_i)\right)$$
$$= \chi\left(\sum_{i \in T_1} s_i - \sum_{i \in T_2} s_i\right) \times \psi\left(-\sum_{i \in T_2} s_i + \sum_{i \in T_3} s_i\right).$$

Since the T_ℓ are disjoint and not all empty, the no three disjoint equivoluminous subsets property implies that either the argument of χ or the argument of ψ is not the identity element of G.

Now sum over the variou
multiplied by $|G|^2$, which
summing over all characte

This shows that the tri

$$\bigotimes_{1 \le i \le n} \text{Good}_i = \bigotimes_{1 \le i \le n} (x$$

is no more than $|G|^2 3^n$.

Note here that, if neces
and replace S by N copies
$(0,...,0, s_i, 0,...,0)$. Then
equivoluminous subsets p

Following the argume

$$2^{6\varepsilon} 3^3 \ge (27/4) 2^{3\tau}$$
$$\omega = 3\tau \le \log_2(2^{6\varepsilon} \times$$

With the assumption that

We have not been abl
approaching 0.

Behrend, F.A. (1946).
progression. *Proc. Nat. A*

Coppersmith, D., Winogr
SIAM Journal on Compu

Coppersmith, D., Winogr
Report RC 12104, IBM T
29, 1986.

Coppersmith, D., Winogr
19th Ann. ACM Symp. o

Pan, V.Ya. (1978). Stra
Uniting and Canceling f
Ann. IEEE Symp. on Fou

Now sum over the various choices of χ and ψ. The "GOOD" term will accumulate (and get multiplied by $|G|^2$, which we can deal with), while all the other terms will vanish, because we are summing over all characters χ (or ψ) at an element of G other than the identity.

This shows that the trilinear rank of

$$\bigotimes_{1\le i\le n} \text{Good}_i = \bigotimes_{1\le i\le n} (x_0y_1z_2 + x_0y_2z_1 + x_1y_0z_2 + x_2y_0z_1 + x_1y_2z_0 + x_2y_1z_0)$$

is no more than $|G|^2 3^n$.

Note here that, if necessary, we can replace G by a Cartesian product of G with itself N times, and replace S by N copies of S, one in each copy of G. A typical element of the large S will be $(0,...,0, s_i, 0,...,0)$. Then $\varepsilon = (\log_2|G|)/|S|$ remains unchanged, and the no three disjoint equivoluminous subsets property is inherited.

Following the arguments from Chapter 6, we would then obtain, by analogy to Equation (9),

$$\begin{aligned} &2^{6\varepsilon}3^3 \ge (27/4)2^{3\tau} \\ &\omega = 3\tau \le \log_2(2^{6\varepsilon} \times 4) = 2 + 6\varepsilon. \end{aligned} \tag{15}$$

With the assumption that $\varepsilon = (\log_2|G|)/|S|$ approaches 0, this would yield $\omega = 2$.

We have not been able to determine whether there exist such pairs (G,S) with $(\log_2|G|)/|S|$ approaching 0.

References.

Behrend, F.A. (1946). On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. USA* **32**, 331-332.

Coppersmith, D., Winograd, S. (1982). On the Asymptotic Complexity of Matrix Multiplication. *SIAM Journal on Computing,* Vol. 11, No. 3, 472-492.

Coppersmith, D., Winograd, S. (1986). Matrix Multiplication via Behrend's Theorem. *Research Report RC 12104,* IBM T.J. Watson Research Center, Yorktown Heights, N.Y., 10598, August 29, 1986.

Coppersmith, D., Winograd, S. (1987). Matrix Multiplication via Arithmetic Progressions. *Proc. 19th Ann. ACM Symp. on Theory of Computing,* 1-6.

Pan, V.Ya. (1978). Strassen Algorithm Is Not Optimal. Trilinear Technique of Aggregating, Uniting and Canceling for Constructing Fast Algorithms for Matrix Multiplication. *Proc. 19th Ann. IEEE Symp. on Foundations of Computer Science,* 166-176.

Pan, V.Ya. (1984). How to Multiply Matrices Faster. *Springer Lecture Notes in Computer Science,* vol. 179.

Schönhage, A. (1981). Partial and Total Matrix Multiplication. *SIAM J. on Computing,* 10, 3, 434-456.

Salem, R., Spencer, D.C. (1942). On sets of integers which contain no three terms in arithmetical progression. *Proc. Nat. Acad. Sci. USA* **28**, 561-563.

Strassen, V. (1986). The Asymptotic Spectrum of Tensors and the Exponent of Matrix Multiplication. *Proc. 27th Ann. IEEE Symp. on Foundations of Computer Science,* 49-54.

J. Symbolic Computation (1990)

Fun
Poly

J(

Departm

If g and h are polyr
sition $f = g(h)$ has
f of degree $n = rs$,
case, compute them
tial algorithms that
algorithm with opti
are multivariate, an
where the character

If F is a field and g, h
composition, and (g, h) i
an essentially unique *cor*
are indecomposable poly
the characteristic p of F

We start with the fo
$r, s \in \mathbf{N}$ with $n = rs$, de
such that $f = g \circ h$. Fo
hard: the security of a
broken by Berkovits and
acteristic zero) were give
version of which appeare
who presented the first
quential time $O(n^2)$, an
algorithm.

The present paper c
ing their basic method:

This work was supported
DFG grant SFB 124, and
visits to Universität des Sa
Chile, Santiago, Chile, and a
University, Canberra, Austr

0747–7171/90/030281+19 $0