Consiglio Nazionale delle Ricerche

ISTITUTO DI ELABORAZIONE DELLA INFORMAZIONE

PISA

SOME PROPERTIES OF DISJOINT SUMS OF TENSORS RELATED TO MATRIX MULTIPLICATION

F. Romani

Nota Interna

B80-4

Febbraio 1980

SOME PROPERTIES OF DISJOINT SUMS OF TENSORS

RELATED TO MATRIX MULTIPLICATION

Francesco Romani

Istituto di Elaborazione dell'Informazione,
via S.Maria 46, 56100 Pisa, ITALY.

ABSTRACT

Let t be a disjoint sum of tensors associated to a matrix product. The rank of the s-th tensorial power of t can be bounded by an expression involving the elements of t and an exponent for matrix multiplication. This relation leads to a trascendental equation defining a new exponent for matrix multiplication. The use of this approach allowed reducing to 2.5166 the exponent 2.5218 due to V.Pan, S.Winograd [7,8] and A.Schonhage [9].

Key Words.

Computational Complexity, Matrix Multiplication, Tensor Rank, Exponent.

1. INTRODUCTION

Since V.Strassen in 1969 showed that the complexity of matrix multiplication is lower than $O(n^3)$ operations [10], the problem arised to determine the intrinsic complexity of the problem. A lower bound to the number of operations is n^2, but for ten years the best known upper bound was $O(n^{2.81})$. Recently the use of new techniques allowed to considerably reduce this upper bound. The method of Trilinear Uniting, Aggregating and Canceling [4,5,6,7,8], the introduction of Approximate Algorithms (also called Field Extension methcd) [1,2,3], and the powerful theory of Partial Matrix Multiplication [9], led to an exponent of 2.5218 [7,8,9].

In this paper we start from the final argument of A.Schonhage in [9] to derive some theorems on the rank of the s-th tensorial power of disjoint sums of tensors. The application of these theorems results in an exponent of 2.5166 for matrix multiplication.

2. NOTATION AND PRELIMINARIES

The reader is assumed to be familiar with the theory of matrix multiplication algorithms. Fcr a survey see [9] whose notation is follcwed here with some minor changes.

Let A and B be twc matrices of indeterminates cn some scalar field F. For the sake of simplicity we assume F=R. A detailed discussion of the role of F can be found in [9]. The problem is tc compute the matrix product C=AB.

Let us introduce some definitions.

mam(n) is the total number of arithmetical operations +,-,x needed to compute AB when A and B are square matrices of order n.

$$w = \inf \{ x : mam(n) = O(n^x) \}.$$

m(n) is the total number of products between linear combinations of indeterminates needed to compute AB in the bilinear noncommutative model.

It is well known that $m(n') \le r'$ implies

$$mam(n) = O(n^b), \quad w \le b, \quad b = \log r' / \log n'.$$

The theory of matrix multiplication algorithms is strictly related to the tensor algebra. Here we deal only with 3-dimensional tensors. The tensorial product of three vectors x,y,z is a 3-dimensional tensor, it is denoted with $x \otimes y \otimes z$ and is called triad. Any tensor t can be expressed as a sum of a number of triads. The rank of t is defined to be the length of the minimal decomposition of t into triads:

$$rk(t) = \min \{ r : t = \sum_{j=1}^{r} x_j \otimes y_j \otimes z_j \}.$$

The tensorial product between tensors is defined as a 3-dimensional tensor. By using multi-indices $i=(i',i'')$, $j=(j',j'')$, $k=(k',k'')$, $t=t' \otimes t''$ has elements:

$$t_{ijk} = t'_{i'j'k'} \, t''_{i''j''k''} .$$

$\langle m,n,p\rangle$ is the tensor associated to the product of a matrix m x n with a matrix n x p, i.e.

$$\langle m,n,p\rangle = t_{(i,j)(h,k)(r,s)} = \delta_{jh} \, \delta_{kr} \, \delta_{si} .$$

The following properties hold for tensors associated to matrix products

$rk(\langle n,n,n\rangle) = m(n)$,

$\langle m,n,p\rangle \otimes \langle m',n',p'\rangle = \langle mm',nn',pp'\rangle$,

$\langle m,n,p\rangle^s = \langle m^s,n^s,p^s\rangle$,

$rk(\langle m,n,p\rangle)$ is symmetrical in m,n,p,

$rk(\langle m,n,p\rangle) \leq r$ implies $w \leq 3 \log r / \log mnp$.

Given two tensors t' and t'' the disjoint sum $t' \oplus t''$ is formed by packing a copy of t' and a copy of t'' into the opposite corners of a parallelepiped of appropriate size and filling with zeros the other positions. For our purposes $t' \oplus t''$ can be considered equivalent to $t'' \oplus t'$.

The sum $\bigoplus_{i=1}^{k} t$ is denoted with $k*t$.

The following properties hold:

$t \otimes (t' \oplus t'') = (t \otimes t') \oplus (t \otimes t'')$,

$rk(t' \oplus t'') \leq rk(t') + rk(t'')$,

$\langle m,n,p\rangle \oplus \langle m',n',p'\rangle$ is associated to two independent matrix products of size mxn x nxp and m'xn' x n'xp'.

By an approximate decomposition of order h and length r for a tensor t we mean a representation

$$T(l) = \sum_{j=1}^{r} x_j(l) \otimes y_j(l) \otimes z_j(l) = l^h t + O(l^{h+1}),$$

$r_h(t)$ is the minimal length of a decomposition of order h,

$r_0(t) = rk(t)$,

$\min_h r_h(t) = \underline{rk}(t)$ is called the border rank of t.

The following properties hold

$\underline{rk}(t) \leq rk(t)$,

$r_{h'+h''}(t' \otimes t'') \leq r_{h'}(t') + r_{h''}(t'')$,

$r_{sh}(t^s) \leq (r_h(t))^s$,

$rk(t) \leq (1+2h)\, r_h(t)$, [2].

This last property allows to use approximate decompositions to reduce the exponent, i.e.

$\underline{rk}(\langle m,n,p \rangle) \leq r$ implies $w \leq 3 \log r / \log mnp$.

A fundamental theorem by A.Schonhage allows to use approximate decompositions of disjoint sums of tensors to derive new low exponents.

THEOREM 1.

Let $t = \bigoplus_{i=1}^{k} t_i$, $t_i = \langle m_i, n_i, p_i \rangle$, $f_i = m_i n_i p_i$

Then $\underline{rk}(t) \le r$ implies $w \le 3x$ where x is the unique solution of the equation

$$\sum_{i=1}^{k} f_i^x = r.$$

The trilinear Aggregating Uniting and Canceling technique can be used to obtain good approximate decompositions of disjoint sum of tensors. Pan and Winograd proved that $\underline{rk}(t \oplus t' \oplus t'') \le 2(n+2)(k+1)$ where $t = \langle 1,k,2n \rangle$, $t' = \langle n,2,k \rangle$, $t'' = \langle 2k,n,1 \rangle$ [7,8]. Applying theorem 1 with n=11 and k=5 they get $w \le 2.5218127...$.

In his paper [9] Schonhage noted that 2k+2 triads in the Pan's decomposition can be joined into k+1 trilinear forms associated to the tensor $\langle 1,1,2 \rangle$, i.e.

$\underline{rk}(t \oplus t' \oplus t'') \le rk[2(n+1)(k+1) * \langle 1,1,1 \rangle \oplus (k+1) * \langle 1,1,2 \rangle]$.

Symmetrizing $t \oplus t' \oplus t''$ he obtains an expression containing $(k+1)^3 * \langle 2,2,2 \rangle$ which allows him to use Strassen's algorithm to save $(k+1)^3$ scalar products. Thus the exponent is reduced to $w \le 2.5218006...$.

In the following we develop the idea of Schonhage to use higher powers of $(t \oplus t' \oplus t'')$.

3. THE RANK OF POWERS OF DISJOINT SUMS OF TENSORS

We premit two simple lemmas.

LEMMA 1.

Let $b>w$, then $m(n) \leq c\, n^{b}$ for any n.

Proof: by definition of w there exists a n' such that

$$\text{for } n>n' \quad m(n) \leq c'\, n^{b}$$

$$\text{and for } n\leq n' \quad m(n) \leq n'^{3}$$

$$\text{then } m(n) \leq n'^{3}\, c'\, n^{b} \leq c\, n^{b} \quad \text{for any } n.$$

LEMMA 2.

Let $b>w$, $m\leq n\leq p$ then

$$rk(<m,n,p>) \leq c'' m^{b-2}\, np.$$

Proof:

$$rk(<m,n,p>) \leq rk(<m,m,m> \otimes <1,\lceil n/m\rceil,\lceil p/m\rceil>) \leq c\, m^{b} \lceil n/m\rceil \lceil p/m\rceil$$

and $\lceil n/m\rceil \geq 1$ and $\lceil p/m\rceil \geq 1$ then

$$rk(<m,n,p>) \leq 4c\, m^{b}\, (n/m)(p/m) = c'' m^{b-2}\, np.$$

Now it can be proved the central theorem of the paper.

THEOREM 2.

$$\text{Let } t = \bigoplus_{i=1}^{k} <m_i, n_i, p_i>, \quad b>w,$$

then

$$rk(t^S) \leq cS^k \left(\max \left[\sum_{i=1}^{k} m_i^{b-2} n_i p_i \right] \right)^S$$

where the max is taken over the permutations of m,n,p.

Proof:

$$rk(t^S) \leq rk \left[\bigoplus_{s_1+s_2+\dots s_k=S} \frac{S!}{s_1!s_2!\dots s_k!} \langle \prod m_i^{s_i}, \prod n_i^{s_i}, \prod p_i^{s_i} \rangle \right] \leq$$

$$\leq \sum_{s_1+s_2+\dots s_k=S} \frac{S!}{s_1!s_2!\dots s_k!} rk(\langle \prod m_i^{s_i}, \prod n_i^{s_i}, \prod p_i^{s_i} \rangle).$$

Let $s_1, s_2, \dots s_k$ be the k-ple for which the corresponding term in the above expansion is maximal, and assume

$$\prod m_i^{s_i} \leq \prod n_i^{s_i} \leq \prod p_i^{s_i} \qquad (1)$$

then

$$rk(t^S) \leq S^k \frac{S!}{s_1!s_2!\dots s_k!} rk(\langle \prod m_i^{s_i}, \prod n_i^{s_i}, \prod p_i^{s_i} \rangle) \leq$$

$$\leq cS^k \frac{S!}{s_1!s_2!\dots s_k!} \prod_{i=1}^{k} m_i^{(b-2)s_i} n_i^{s_i} p_i^{s_i}.$$

Obviously one term of a multinomial expansion is less than the whole expression, hence:

$$rk(t^S) \le cS^k \left(\sum_{i=1}^{k} m_i^{b-2} n_i p_i \right)^S .$$

This formula holds under the assumption (1). But it is not known for which index the minimum is attained, hence:

$$rk(t^S) \le cS^k \left(\max \left[\sum_{i=1}^{k} m_i^{b-2} n_i p_i , \sum_{i=1}^{k} m_i n_i^{b-2} p_i , \sum_{i=1}^{k} m_i n_i p_i^{b-2} \right] \right)^S .$$

COROLLARY 1.

If the set of disjcint tensors in t is symmetrical in m,n,p the three expressions are equal and

$$rk(t^S) \le cS^k \left(\sum_{i=1}^{k} m_i^{b-2} n_i p_i \right)^S .$$

REMARK.

Theorem 2 can be considered as a weak converse of theorem 1.

E.g. let $t = \bigoplus_{i=1}^{k} \langle m_i, m_i, m_i \rangle$, then from theorem 1 it follows

$$\underline{rk}(t) \le r = \sum_{i=1}^{k} m_i^{3x} \quad \text{implies } w \le 3x,$$

and from theorem 2

$$w \leq 3x \text{ implies } rk(t^s) \leq cs^k \left(\sum_{i=1}^{k} m_i^{3x} \right)^s$$

COROLLARY 2. Let

$t_1 = d*<1,1,1>$, $t_2 = e*<1,1,k>$, $t'_2 = e*<1,k,1>$, $t''_2 = e*<k,1,1>$,

$t = (t_1 \oplus t_2) \otimes (t_1 \oplus t'_2) \otimes (t_1 \oplus t''_2)$, $b > w$, then

$$rk(t^s) \leq cs^8 [(d+ek)^2 (d+ek^{b-2})]^s$$

Proof: the set of elements of t is symmetrical, in fact

$$t = d^3*<1,1,1> \oplus d^2e*(<1,1,k> \oplus <1,k,1> \oplus <k,1,1>) \oplus$$
$$\oplus de^2*(<1,k,k> \oplus <k,1,k> \oplus <k,k,1>) \oplus e^3*<k,k,k>.$$

Applying corollary 1 we get

$$rk(t^s) \leq cs^8 [d^3 + d^2e(2k+k^{b-2}) + de^2(2kk^{b-2}+k^2) + e^3(k^2k^{b-2})]^s =$$

$$= cs^8 [(d+ek)^2 (d+ek^{b-2})]^s .$$

4. APPLICATION TO MATRIX MULTIPLICATION

Let t be a sum of r triads; t can be viewed as the homomorphic image of the tensor $T=r*<1,1,1>$. In such a case we write $t \to T$. It easy to see that

$$\sum_{j=1}^{k} x \otimes y_j \otimes z_j \quad \to \quad \langle 1,1,k \rangle.$$

Obviously $t \to T$ implies $rk(t) \le rk(T)$, moreover $t' \to T'$, $t'' \to T''$ implies $t' \otimes t'' \to T' \otimes T''$.

The same considerations can be made for sums of triads depending on a variable l. Thus the following theorems can be stated.

THEOREM 3.

$$l^{h'} t' + O(l^{h'+1}) = T'(l) \to t'_1 \quad \text{and}$$

$$l^{h''} t'' + O(l^{h''+1}) = T''(l) \to t''_1 \quad \text{imply}$$

$$\underline{rk}(t' \otimes t'') \le rk(t'_1 \otimes t''_1).$$

Proof.

From the definition of $\to$ and of border rank it follows:

$$\underline{rk}(t' \otimes t'') \le r_{h'+h''}(t' \otimes t'') \le rk(T'(l) \otimes T''(l)) \le rk(t'_1 \otimes t''_1).$$

THEOREM 4.

Let $t = \sum_{i=1}^{z} \bar{t}_i$, $\bar{t}_i = \langle m_i, n_i, p_i \rangle$, $m_i n_i p_i = f_i$,

and let

$$l^{h} t + O(l^{h+1}) = T(l) \to d*\langle 1,1,1 \rangle \oplus e*\langle 1,1,k \rangle.$$

Then the unique solution of the equation

$$\left(\sum_{i=1}^{z} f_i^{x}\right)^3 = (d+ek)^2 (d+ek^{3x-2})$$

satisfies $w \le 3x$.

Proof: Let

$t_1 = d*<1,1,1>$, $t_2 = e*<1,1,k>$, $t'_2 = e*<1,k,1>$, $t''_2 = e*<k,1,1>$,

and t', t'' the tensors obtained from t with the corresponding permutations. We have

$$l^{3h}(t\otimes t'\otimes t'') + O(l^{3h+1}) = T^*(l) \to (t_1 \oplus t_2) \otimes (t_1 \oplus t'_2) \otimes (t_1 \oplus t''_2) = t_3$$

and

$$l^{3sh}(t\otimes t'\otimes t'')^s + O(l^{3sh+1}) = T_s^*(l) \to t_3^s .$$

Applying corollary 2 we obtain

$$rk[(t\otimes t'\otimes t'')^s] \le rk(t_3^s) \le cs^8 [(d+ek)^2 (d+ek^{b(o)-2})]^s ,$$

for any s and $b(o) > w$.

Now $(t\otimes t'\otimes t'')^s$ has z^{3s} independent components. Their volumes are given by the terms of the expansion of

$(\sum f_i)^{3s}$. Then by theorem 1 the solution of

$$\left(\sum_{i=1}^{z} f_i^{x}\right)^{3s} = cs^8 [(d+ek)^2 (d+ek^{b(o)-2})]^s$$

satisfies $w \le 3x$. The solution of this equation depends on c and s. But $x' = \inf\{x(s,c), s \in N\}$ is the solution of the equation

$$\left(\sum_{i=1}^{z} f_i^{x'}\right)^3 = (d+ek)^2 (d+ek^{b(0)-2})$$

and $w \le 3x'$.

It is easy to see that substituting to $b(0)$ the new value $b(1)=3x'$ and iterating the process, the resulting values converge to the unique solution of the equation:

$$\left(\sum_{i=1}^{z} f_i^{b/3}\right)^3 = (d+ek)^2 (d+ek^{b/3-2})$$

and any value of the sequence $\{b(0), b(1), \ldots\}$ is an upper bound for w.

COROLLARY 3. $w \le 2.516648\ldots$.

Proof: Pan presented a decomposition $T(1)$ for $t=\langle 1,k,2n\rangle \oplus \langle n,2,k\rangle \oplus \langle 2k,n,1\rangle$ [7,8] and Schonhage proved $T(1) \rightarrow [2(n+1)(k+1) * \langle 1,1,1\rangle \oplus (k+1) * \langle 1,1,2\rangle]$ [9]. Then from theorem 4 $w \le b$, where b is the solution of

$$27(2kn)^b = [2(n+2)(k+1))^2 (2(n+1)(k+1)+(k+1)2^{b-2}].$$

In fact the symmetrization of t yields 27 independent components of the same volume $(2kn)^3$.

The minimal value of b is attained for $n=10, k=5$, i.e.

$$100^b = 144^2 (132+6\ 2^{b-2})/27 \text{ gives } b=2.516648\ldots .$$

REFERENCES

[1] D.Bini, M.Capovani, G.Lotti, F.Romani, $o(n^{2.7799})$ Complexity for nxn Approximate Matrix Multiplication. Information Processing Lett., 8 (1979), pp.234-235.

[2] D.Bini, Relations between Exact and Approximate Bilinear Algorithms. Applications. (to appear in Calcolo).

[3] D.Bini, G.Lotti, F.Romani, Approximate Solutions for the Bilinear Form Computational Problem. (to appear in SIAM J. on Computing).

[4] V.Ya.Pan, Strassen's Algorithm is not Optimal. Proc 19-th Ann. Symp. on Foundations of Computer Science, (1978), pp.166-176.

[5] V.Ya.Pan, New Fast Algorithms for Matrix Operations. (to appear in SIAM J. on Computing).

[6] V.Ya.Pan, Field Extension and Trilinear Aggregating, Uniting and Canceling for the Acceleration of Matrix Multiplications. Proc 20-th Ann. Symp. on Foundations of Computer Science, (1979), pp.28-38.

[7] V.Ya.Pan, S.Winograd, (to appear).

[8] V.Ya.Pan, New Combinations of Methods for the Acceleration of Matrix Multiplication (to appear).

[9] A.Schonhage, Partial and Total Matrix Multiplication. (to appear).

[10] V.Strassen, Gaussian Elimination is not Optimal. Numer. Math. 13(1969), pp.354-356.