25. Strassen's Fast Multiplication of Matrices Algorithm

1. Introduction

Suppose we want to multiply two n by n matrices, A and B.

Their product, AB, will be an n by n matrix and will therefore have n^2 elements. Each one of these elements is naturally expressed as the sum of n products, each of an element of A with one of B. Thus we have

$$(AB)_{jk} = \Sigma_{s=1}^{s=n} A_{js}B_{sk},$$

and the number of multiplications involved in producing the product in this way is n^3.

In fact, a matrix product of this kind can be obtained using a smaller number of operations, and we will describe how this can be done.

We will also discuss some curious spreadsheet algorithms for manipulating matrices.

Thus, if we produce the product AB in the usual manner on a spreadsheet with no additional work we can obtain A^kB as well, for any k we choose.

Also, with one easy instruction, whose entry is similar to applying the formula for computing the determinant of a 2 by 2 matrix (along with some copying), we can compute the determinant of any square matrix A of any size, and even obtain all the cofactors of the elements A.

2. Fast Matrix Multiplication; Partitioning Matrices

We will describe an algorithm (discovered by V.Strassen) that allows us to multiply two n by n matrices A and B, with a number of multiplications (and additions) which is a small multiple of $n^{(\ln 7)/(\ln 2)}$, when n is of the form 2^k.

The algorithm is based upon three ideas.

The first idea is that of "**partitioning matrices**" which in our context is:

The multiplication of 4 by 4 matrices A and B is equivalent to the multiplication of a pair of 2 by 2 matrices whose elements are each 2 by 2 matrices.

Explicitly, if we describe A and B as

$$\begin{matrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{matrix} \quad \text{and} \quad \begin{matrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \end{matrix}$$

$$a_{41}\ \ a_{42}\ \ a_{43}\ \ a_{44} \qquad\qquad b_{41}\ \ b_{42}\ \ b_{43}\ \ b_{44}$$

Then their product is exactly the same as the product of the matrices

$$\begin{matrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{matrix} \quad \text{and} \quad \begin{matrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{matrix}$$

where we have

$$A_{11} = \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix} \quad A_{21} = \begin{matrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{matrix} \quad A_{12} = \begin{matrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{matrix} \quad A_{22} = \begin{matrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{matrix},$$

and the same for B.

When we perform matrix multiplication here we take the dot products of the rows of A with the columns of B. In doing so we sum over an intermediate variable.

The difference between the normal representation of the product and the "partitioned representation as a 2 by 2 product of 2 by 2 matrices is the difference between representing the indices to be summed over as 1 2 3 and 4, or as 11 12 21 and 22.

This is akin to representing numbers from 1 to 4 either by a pair of binary digits or as one digit mod 4. When we do the multiplication, as long as we sum over all possible middle indices, it makes no difference how we choose to represent them.

And of course the same thing is true for matrices whose dimensions are 2^k by 2^k. We can represent the indices either as integers from 1 to 2^k or as k-tuples, each element of which is 1 or 2.

And with the latter representation we can interpret matrix multiplication as the 2 by 2 product of 2^{k-1} by 2^{k-1} matrices each one of which is a 2 by 2 product of 2^{k-2} by 2^{k-2} matrices, and so on.

The consequence we draw from this fact is: if we can multiply 2 by 2 matrices using only 7 multiplications instead of the usual 8, we can parlay that into multiplying 4 by 4 matrices using 7 multiplications of 2 by 2 matrices each of which requires 7 multiplications of numbers, for a total of 49 multiplications.

Furthermore, by iterating this fact, we can multiply 2^k by 2^k matrices with 7^k multiplications of numbers, so long as we can handle 2 by 2 matrices with 7 multiplications.

You might note here that this kind of matrix partitioning can be done whenever you have n by n matrices and n is the product of a and b. You can then write the product

of n by n matrices as the product of a by a matrices each one of which is a b by b matrix, again by labeling the indices accordingly.

2. Representing a Matrix Product as a Single Polynomial

From now on then we will consider the problem of computing the 4 entries in the product of two 2 by 2 matrices.

Explicitly, given 2 by 2 matrices with elements a_{ij} and b_{jk} we want to compute the four combinations

$$a_{11}b_{11} + a_{12}b_{21},\ a_{11}b_{12} + a_{12}b_{22},\ a_{21}b_{11} + a_{22}b_{21},\ \text{and}\ a_{21}b_{12} + a_{22}b_{22}.$$

The second idea here is that **we will get a better grasp of the problem if we combine these four combinations into one entity. And we can do this by multiplying each one by an indeterminate (or variable) and adding them up.**

The result will be a polynomial, and our task will be both to compute this polynomial from the a's and b's using 7 multiplications, and to find the coefficients of our indeterminates in the polynomial.

And here is where we run into amazing luck. If we call our indeterminates z_{kj} and multiply the combinations above by z_{11}, z_{21}, z_{12} and z_{22} respectively, our polynomial can be written as

$$a_{11}b_{11}z_{11} + a_{12}b_{21}z_{11} + a_{11}b_{12}z_{21} + a_{12}b_{22}z_{21} + a_{21}b_{11}z_{12} + a_{22}b_{21}z_{12} + a_{21}b_{12}z_{22} + a_{22}b_{22}z_{22}$$

which is

$$\mathbf{\Sigma_{j,s,k=1}^{2}\ a_{js}b_{sk}z_{kj}.}$$

Our luck is the fact that **this polynomial has quite a bit of symmetry**.

In particular, we can interchange the subscripts 1 and 2 everywhere, and this combination does not change. Which is to say that for each term (for example the first term) there is another which is its image under changing all the indices (here the last term).

Second, we can permute the indices j s and k, replacing j by s, s by k and k by j, without changing this polynomial, and we can reverse this permutation.

If we perform the first of these two permutations the first and last terms stay fixed, but the term $a_{12}b_{21}z_{11}$ becomes $a_{21}b_{11}z_{12}$ (j=1 s=2 k=1 becomes j=2 s=1 k=1) and so on.

Thus there is a group of index changes that leave our polynomial unchanged, and this group has the following six elements

The identity; (j,s,k); (j,k,s); reverse 1 and 2; reverse 1 and 2 and (j,s,k); reverse 1 and 2 and (j,k,s).

The 8 terms in our polynomial form two orbits under the action of this group.

One is the pair consisting of the first and last term above whose indices are all the same. These are stabilized by the identity (j,s,k) and (j,k,s).

The other orbit consists of the remaining terms. Notice that each of these terms has one factor with repeated indices and two factors whose indices are each 12 or 21.

(By the way, you should read (j,s,k) as: the new value of j is the old value of s, the new value of s is the old value of k, and the new k value is the old j value.)

The next question is: how can we exploit this symmetry to find a way to write our polynomial as the sum of 7 products.

3. The Last Idea

First we must introduce a product that will give us the first and last entries.

We can find **a single product** which has all the same symmetries as our polynomial, that will give us both of these: namely

$$\mathbf{(a_{11} + a_{22})\,(b_{11} + b_{22})\,(z_{11} + z_{22}).}$$

If we write this product out, however, it consists of 8 terms, 2 of which are what we want, namely our first and last terms, but there are 6 terms we do not want and of course 6 terms in our polynomial that are not present in this product.

In fact the difference between our polynomial and this product is given by this twelve term polynomial:

$$\mathbf{a_{12}b_{21}z_{11} + a_{11}b_{12}z_{21} + a_{12}b_{22}z_{21} + a_{21}b_{11}z_{12} + a_{22}b_{21}z_{12} + a_{21}b_{12}z_{22} - a_{11}b_{11}z_{22} - a_{11}b_{22}z_{11} - a_{11}b_{22}z_{22} - a_{22}b_{11}z_{11} - a_{22}b_{11}z_{22} - a_{22}b_{22}z_{11}.}$$

This difference polynomial, being the difference of two polynomials each of which is invariant under the action of our group, is also invariant under this group's action.

If you look at this difference, **it consists of two complete orbits under our symmetry group: both orbits have two indices the same and one different; one has only diagonal matrix elements, the other has off diagonal ones.**

We need to find an invariant that will add the positive terms here and get rid of the negative ones.

How can we get an invariant here?

The obvious way is to find a single asymmetric product that will handle two of the terms we need, apply each of the symmetries in our group to it, and add them up.

This will produce an invariant, consisting of six products, and has a chance of being what we need, since it will certainly fix the 12 terms we want. If it does no other harm it will do what we need here.

Suppose we want a term that will produce $a_{12}b_{21}z_{11} - a_{22}b_{22}z_{11}$. We can try

$$(a_{12} - a_{22})\,(b_{21} + b_{22})\,z_{11}.$$

.

This will do the right thing bby giving us the two terms we want but multiplying it out produces four terms:, it produces two weird extra terms, namely

$$\mathbf{a_{12}\,b_{22}z_{11}} \textbf{ and } \mathbf{-a_{22}b_{21}\,z_{11}}.$$

And here is the great thing. **These two terms are in the same orbit under our group action and have the opposite sign**. Thus if we apply all our group operations to them and add up the results, we will get that orbit minus itself or 0.

And the terms we want will give us the difference in orbits that we want.

Which means we have the answer!

Explicitly, the six products that we want which, when added to $(a_{11} + a_{22})\,(b_{11} + b_{22})\,(z_{11} + z_{22})$ give us our polynomial, are

$$(a_{12} - a_{22})\,(b_{21} + b_{22})\,z_{11}$$

$$(a_{21} - a_{11})\,(b_{12} + b_{11})\,z_{22}$$

.

$$a_{11}\,(b_{12} - b_{22})\,(z_{21} + z_{22})$$

$$a_{22}\,(b_{21} - b_{11})\,(z_{12} + z_{11})$$

$$(a_{21} + a_{22})\,b_{11}\,(z_{12} - z_{22})$$

.

$$(a_{12} + a_{11})\,b_{22}\,(z_{21} - z_{11}).$$

.

And indeed, the sum of the seven products indicated here give us our polynomial.

Let us recall what this means. The products indicated are the products of the a's and b's here. The z's tell us where the products go in the product matrix.

Thus in the first term, we must put the first product, $(a_{11} + a_{22})(b_{11} + b_{22})$ in both the 11 and 22 entries of the product matrix. Similarly, the second product, $(a_{12} - a_{22})(b_{21} + b_{22})$ goes in the 11 entry, the last, $(a_{12} + a_{11})\, b_{22}$ appears with a positive sign in the 12 entry and with a negative sign in the 11 entry of the matrix.

To take the products indicated of 2^j by 2^j matrices requires first forming the combinations of a's and b's necessary to take them.

There are 5 additions or subtractions of **a** matrices and the same number of operations on **b** matrices.

Then, once the results are obtained for these multiplications, they must be reassembled into the product matrix. This requires an additional 8 additions or subtractions of such matrices for each product of same, for a total of 18 additions or subtractions of 2^j by 2^j matrices for each multiplication of same.

To handle a 2^k by 2^k multiplication, we have seen that we must perform

$1\ 2^k$ level product, $7\ 2^{k-1}$ level products, $7^2\ 2^{k-2}$ level products, and so on.

This will require, as we have noted **a total of 7^k multiplications of numbers.**

And how many additions of numbers?

At level 2^k there will be 18 additions each of 2^{2k-2} matrix elements (namely of all the elements of the matrices of half the size of the original matrix that form the elements of the top level 2 by 2 matrix)

At the next level there will 7/4 as many sums: the 7 comes from there being 7 times as many matrices to deal with, the 4 from the fact that they have half the size and hence a quarter as many elements.

So the answer is $18*2^{2k-2}*(1+ 7/4 + (7/4)^2 + \ldots(7/4)^{k-1})$ which works out to be $6*(7^k-4^k)$.

Thus even the number of additions grows as 7^k.

This procedure can be implemented on a spreadsheet without too much difficulty for 4 by 4 or 8 by 8 matrices, and with a program you could handle any size.

To do it you must form the 7 combinations of a's and b's to be multiplied. Then multiply together the appropriate combinations, then put them in the right places in the resulting matrix. (You have to remember that the coefficient of z_{12} goes into the 21 element of the product matrix-which is the transpose of what you might think, but that is the only tricky point.)

4. Matrix Magic on a Spreadsheet

The act of matrix multiplication in the ordinary way is easier to implement on a spreadsheet.

In fact it can be accomplished with **one instruction suitably copied**.

Thus if you enter your k by k matrix A in k rows and columns and put B somewhere next to it, you can place the product AB similarly next to B, by entering the dot product of the first row of A with the first column of B in its upper left corner.

If you put dollar signs on all occurrences of the middle index, (the one summed over) when you copy this entry into the k rows and columns starting from it, you get the product AB, since the other indices will vary and give you the dot product of the rows of A with the columns of B.

But here is where magic occurs. If you copy further to the right, beyond where the matrix AB should be, you find another matrix and another and another.

The spreadsheet iterates the matrix multiplication. So what you get after the product AB is the product of A with it, namely A^2B, then A^3B, and so on.

And you get all this at the cost only of copying one entry.

You may find this quite mundane, but it seems remarkable to me. But here is something even you will find remarkable.

5. Determinants and Cofactors with a Spreadsheet

Lewis Carroll, the author of Alice in Wonderland, was a mathematician and the discoverer of a useful theorem about determinants.

It can be stated as follows.

Suppose A is an n by n matrix, and suppose we define A_{jk} to be the matrix obtained from A by removing its j-th row and k-th column.

Similarly let us define $A_{jk,lm}$ to be the matrix obtained from A by omitting its j-th and k-th rows, and l-th and m-th columns.

We denote the determinant of a matrix by |A|

Carroll's (or Dodson's) theorem then takes the form

$$|\mathbf{A}||\mathbf{A}_{\mathbf{jk},\mathbf{jk}}| = |\mathbf{A}_{\mathbf{jj}}||\mathbf{A}_{\mathbf{kk}}| - |\mathbf{A}_{\mathbf{jk}}||\mathbf{A}_{\mathbf{kj}}|.$$

Here a 0 by 0 determinant is defined to be 1.

The theorem is of particular interest when we choose j=1 and k=n.

It then states that the determinant of a matrix multiplied by the determinant of the matrix obtained by throwing away its outside rows and columns, is the product of the determinants of submatrices obtained by throwing away the top row and left column with the one obtained by throwing away the last row and right column, less the product of the determinant obtained after throwing away the top row and right column and the determinant obtained after throwing away the last row and left column.

This definitely gives the 2 by 2 determinant, and Dodson noticed that it generalizes the 2 by 2 formula to larger matrices.

In fact this theorem can be taken as a definition of larger determinants, so long as the determinant of the matrix obtained by throwing away all borders is non-zero.

The wonderful thing about this theorem is that you can implement it on a spreadsheet with one instruction, and it will, if you start with a matrix of 1's (representing 0 by 0 determinants) then enter your matrix, then enter one entry and copy it down and across, it will first compute the two by two determinants of submatrices consisting of adjacent rows and columns, then similar 3 by 3 determinants then 4 by 4 etc., until it produces the determinant of your original matrix.

It can fail if you try to divide by a determinant that is 0; but this can be avoided by adding a very small increment appropriately to elements of the original matrix to make all determinants obtained slightly different from 0. By varying these increments if necessary you can eliminate any errors they might introduce.

What is the magic entry?

Choose a blank space q rows below the first column of your matrix, and enter the 2 by 2 determinant of the first two rows and columns of your matrix divided by the entry q-1 rows above the top of your matrix, and one row to its right, (which should be a 1).

You must prepare by filling the squares above your matrix up to the q-1'st row by 1's.

This algorithm computes the determinant with a cubic number of operations, since it has to compute the sum of $1 + 2^2 + 3^2 + \ldots (n-1)^2$ determinants each of which involves two multiplications a subtraction and a division.

In the previous iteration, just before computing the determinant, this procedure produces the determinants of matrices obtained by omitting the last row and column, the

last row and first column and under these the determinants of matrices obtained by omitting the first row and last column and first row and first column.

These are all plus or minus cofactors of elements in the omitted places.

If you want all the cofactors, you can obtain them by copying columns 1 through n-1 immediately to the right of the n-th column, (you can do that with one instruction (=top left element) copied into the n-1 next columns) and copying the firsr n-1 rows of the resulting matrix (similarly) into the n-1 rows immediately beneath it, before you enter the instruction for your algorithm. You will then have a 2n-1 by 2n-1 square matrix.

This will give you at the end a whole matrix of evaluations of the determinant, and on the previous iteration will give you the cofactors, up to sign, starting in the second row and second column.

The reason for this is that when you do this at the next to the last iteration each adjacent submatrix will omit exactly one row and column and will be a cofactor of the entry in that row and column, up to a sign. 0.

When n is odd, these values will actually be the cofactors. When n is even, however, the signs of every second one must be reversed to get the cofactors.

Recall that the cofactors of the matrix divided by the determinant give the transpose of the inverse of the matrix. Thus this simple algorithm gives all the information you need to deduce the inverse of the matrix.

Exercises:1. Form a spreadsheet that sets up the matrix multiplication and determinant and inverse finding algorithms described in the last two sections. Use the latter to find the inverse of a random 5by 5 matrix and test it by matrix multiplying it by the original matrix using the former. Arrange to be able to do this for any 5 by 5 matrix which doesn't cause you to divide by 0.