

25. Strassen's Fast Multiplication of Matrices Algorithm and Spreadsheet Matrix Multiplications

25.1 Introduction

We will use the notation A_{ij} to indicate the element in the i -th row and j -th column of the matrix A .

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. In the definition of matrix multiplication, each one of these elements is naturally expressed as the sum of n products, each of an element of A with one of B . Matrix multiplication is defined for each element of AB in the following equation

$$(AB)_{ij} = \sum_{k=1}^n A_{ik} B_{kj}$$

and the number of multiplications involved in producing AB in this way is n^3 , since there are n multiplications for each of the n^2 elements of AB .

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. For example, if we produce AB in the usual manner on a spreadsheet, with no additional work we can obtain $A^k B$ as well, for any k .

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, which can be used to solve a set of n -by- n linear equations.

25.2 Fast Matrix Multiplication; Partitioning Matrices

We will describe an algorithm, discovered by V. Strassen and usually called **Strassen's Algorithm**, which allows us to multiply two n -by- n matrices with a number of multiplications that is a small multiple of $n^{(\ln 7)/(\ln 2)}$, when n is of the form 2^k . This means we will be able to multiply matrices using about $n^{2.8}$ multiplications instead of n^3 .

The algorithm is based on three ideas

The first of these is **partitioning matrices**. The basic concept is that we can partition matrices into submatrices, or **blocks**. We do this by drawing lines between certain columns of the matrix, and then drawing lines between certain rows of the matrix. This gives us a number of smaller matrices, which is equal to the product of the number of lines we drew between columns and the number of lines we drew between rows.

Thus, for example, we could divide up a 4-by-4 matrix in the following way:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \rightarrow A = \left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array} \right]$$

Our matrix A is now partitioned into 4 blocks, each of which is a 2-by-2 matrix. We can now write A as a 2-by-2 matrix of 2-by-2 matrices in the following way:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where

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

Now, when we originally defined matrix multiplication, we did so for n -by- n matrices consisting of n^2 elements. However, given n -by- n matrices A and B for which n is factorable into r and s ($n = rs$), we can similarly write A and B as r -by- r matrices whose components are each s -by- s matrices without changing the definition of their matrix multiplication at all.

So, for example, if we have two 4-by-4 matrices A and B, we can write them as we did above:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ and } B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

When we multiply A and B we will use our matrix multiplication equation to get that:

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}$$

Now suppose we have two matrices whose dimensions are 2^k -by- 2^k . We can then partition these into two 2-by-2 matrices whose elements are 2^{k-1} -by- 2^{k-1} matrices. These 2^{k-1} -by- 2^{k-1} matrices themselves can be partitioned into 2-by-2 matrices whose elements are 2^{k-2} -by- 2^{k-2} matrices and so on.

The consequence we draw from this fact is that 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, so long as we can handle 2-by-2 matrices with 7 multiplications in a way that does not involve commutativity.

We saw that it takes n^3 multiplications to multiply two n -by- n matrices. If $n = 2^k$, then this is 2^{3k} multiplications. However, it only takes 7^k multiplications if we can get our new method to work. What is this as a power of 2?

$$2^x = 7^k$$

$$x = \log_2 7^k = k \log_2 7 = k \left(\frac{\ln 7}{\ln 2} \right)$$

So this gives us $2^{(k \ln 7) / (\ln 2)}$ which is approximately $2^{2.8k}$. For an arbitrary n , then, this will give us a small multiple of $n^{2.8}$ multiplications instead of n^3 .

25.3 Representing a Matrix Product as a Single Polynomial

Now we will see how to compute the 4 entries in the product of two 2-by-2 matrices using only 7 multiplications.

Explicitly, given two 2-by-2 matrices A and B, 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 terms into one entity. We can do this by multiplying each one by an **indeterminate** (or **marker 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 marker variables in the polynomial.

Here is where we run into amazing luck. If we call our "variables" (or if you prefer, indeterminates) z_{kj} and multiply the terms above by z_{11} , z_{21} , z_{12} and z_{22} respectively, our polynomial when written out is

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

which is

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

Our luck stems from the fact that this polynomial has quite a bit of symmetry.

In particular, we can interchange the subscripts 1 and 2 in each term of the polynomial, and the polynomial itself does not change. This is because for each term in the polynomial (for example the first term), there is another term that is the same except it has the subscripts 1 and 2 reversed (here the last term). Look at each term of the polynomial above and verify this statement.

Secondly, we can permute the indices j , s , and k , by replacing j by s , s by k , and k by j in the above summation, and we will still get the same polynomial. This is equivalent to the permutation (j,s,k) , in cycle notation (see notes 17). If we perform this permutation, the first and last terms stay fixed, but the term $a_{12}b_{21}z_{11}$ becomes $a_{21}b_{11}z_{12}$ and so on. We end up with the same polynomial as before. The same is true for the permutation (j,k,s) , as you should prove to yourself.

Notice also that we can combine the operations of switching indices 1 and 2 and permuting the indices, because neither of them separately changes the polynomial and thus they also do not change it in tandem.

Thus, there is a group of index changes that leave our polynomial unchanged. This is a group of symmetry operations as we saw in notes 17. This group has the following six elements:

- 1) The identity
- 2) permute (j,s,k)
- 3) permute (j,k,s)
- 4) switch indices 1 and 2
- 5) switch 1 and 2 and permute (j,s,k)
- 6) switch 1 and 2 and permute (j,k,s)

The 8 terms of our polynomial form two **orbits** under the action of this group. An orbit is a set of elements where if you apply a group operation to one of them, you get another element of the set. For any two elements of an orbit, you can get from one to another by a finite number of group operations. A **stabilizer** of an orbit is an element of the group such that if you apply it to an element of the orbit, you get that same element as an answer.

One orbit in our polynomial is the pair consisting of the first and last terms, whose indices are all the same. These are stabilized by the identity and by the permutations (j,s,k) and (j,k,s) .

The other orbit consists of the remaining six middle terms. Notice that each of these terms has one factor with repeated indices and two factors whose indices are each 12 or 21. If, for example, we apply (j,s,k) to one of them, we will get another one, if we apply (j,s,k) to that one, we will get a third, and so on until we get all the way back to the one we started with, going through every element of the orbit exactly once.

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

25.4 The last idea

What we want to do is find 7 products that when added together will give us our polynomial. First, we introduce a product that will give us the first and last entries of our polynomial.

We can find a single product, which has all the same symmetries as our polynomial, and when multiplied out gives us the terms we want (along with some others):

$$(a_{11} + a_{22})(b_{11} + b_{22})(z_{11} + z_{22})$$

If we multiply 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 6 terms in our polynomial that are not present in this product.

In fact, the difference between our polynomial and the product

$$(a_{11} + a_{22})(b_{11} + b_{22})(z_{11} + z_{22})$$

is the 6 middle terms of our polynomial minus the six middle terms of the above product. This difference, then, is the twelve-term polynomial:

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

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

If you look at this difference, it consists of two complete orbits under our symmetry group. In each term of the first orbit, one of the factors has two of the same index and the others have one of each index, but in opposite order. (For example, $a_{21}b_{11}z_{12}$ is in this orbit). In the other orbit, each factor has two indices that are the same. (For example, $a_{22}b_{11}z_{11}$).

We want to find an invariant that will give us this difference polynomial and is the sum of six products.

How can we get an invariant here?

There is always a sure-fire way to find an invariant in situations like this: *Take one asymmetric term and look at the sum of the terms in its orbit.* This orbit sum will always be invariant.

Thus, we want to find a pair of terms in our difference polynomial, one positive and one negative, which can be gotten by one product. We can apply each of the symmetries in our group to it and sum the results.

Each term in our product, when summed, will produce an invariant, consisting of six terms. Since the difference polynomial we seek to create consists of the difference of two orbit sums, if we can find a single product that handles one from each orbit, by symmetry, their orbit sums will have them all. So, if we can find a single product that adds up to two of our terms, one of each sign, it will do what we need as long as it does not add additional terms of its own.

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

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

When we multiply this out, we get the two terms above that we wanted, but also two extra terms, namely

$$a_{12}b_{22}z_{11} \text{ and } -a_{22}b_{21}z_{11}$$

Here is the great thing. These two terms are in the same orbit under our group action (you can get the second one using the permutation (j,s,k)) and have the opposite sign. Under the group action, each term here will have an orbit consisting of all possible terms with one 11 index pair, one 22 pair, and a 12 or 21 pair. In other words, the sum of each of these terms over the orbit will be exactly the same, and so the sum of both will cancel to 0.

Furthermore, 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 that, 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 gives us our polynomial (see exercises for more info).

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

Thus, the first term, $(a_{11} + a_{22})(b_{11} + b_{22})(z_{11} + z_{22})$, says we put the product $(a_{11} + a_{22})(b_{11} + b_{22})$ in both the 11 and 22 entries of the product matrix. Similarly, the second product, $(a_{12} - a_{22})(b_{21} + b_{22})z_{11}$, goes in the 11 entry of the product matrix. **Note: z_{21} means that the accompanying product goes in the 12 position not the 21 position. Similarly, z_{12} corresponds to the 21 position.**

To take the product of two 2^k -by- 2^k matrices, we must form the above combinations of a's and b's k times, each time using 7 multiplications.

On top of the 7 products, it takes 10 additions or subtractions to compute each of the 7 terms.

Then, once the results are obtained for the multiplications, they must be reassembled into the product matrix. This requires an additional 8 additions or subtractions.

Therefore, to handle a 2^k -by- 2^k multiplication, we have seen that we must perform one 2^k level product, $7 \cdot 2^{k-1}$ level products, $7^2 \cdot 2^{k-2}$ level products and so on. This will require, as we have noted, a total of 7^k multiplications or numbers.

How many additions of numbers?

At level 2^k , there will be 18 additions, each of 2^{2k-2} 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 be $7/4$ as many sums: the 7 comes from there being 7 times as many matrix products that we have to perform, and the 4 comes from the fact that they each have half the size and hence a quarter as many elements.

So the number of additions is $18 \cdot 2^{2k-2} \cdot (1 + 7/4 + (7/4)^2 + \dots + (7/4)^{k-1})$, which works out to be

$$6 \cdot (7^k - 4^k)$$

Thus, the number of additions grows as $6 \cdot 7^k$, which is $6 \cdot n^{(\ln 7)/(\ln 2)}$.

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

To do so, you must form the 7 combinations of a's and b's to be multiplied. Then, multiply together the appropriate combinations, and 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 spot of the product matrix).

25.5 Matrix Magic on a Spreadsheet

The act of matrix multiplication in the ordinary way is easy 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 the upper left corner of AB.

If you put dollar signs on all occurrences of the middle index (the one summed over) when you copy this entry k times across and down (forming a k-by-k matrix), 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.

Here is where the magic occurs. If you copy further to the right, beyond where the matrix AB should be, you find another k-by-k matrix, and another, and so on.

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

The best part is that you get all this at the cost only of copying one entry.

25.6 Determinants and Cofactors with a Spreadsheet

Lewis Carroll, the author of *Alice in Wonderland*, was also a mathematician and discovered a useful theorem about determinants. It can be stated as follows.

Suppose A in an n-by-n matrix, and suppose that 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 its l-th and m-th columns.

We denote the determinant of the matrix A by |A|.

Thus, **Carroll's (or Dodson's) Theorem** takes the form:

$$|A||A^{jk,jk}| = |A^{jj}||A^{kk}| - |A^{jk}||A^{kj}|$$

(Here, the determinant of a 0-by-0 matrix is defined to be 1)

This 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 matrix obtained by throwing away its outermost rows and columns, is the product of the determinant of the submatrix obtained by throwing away the top row and the left column with the one obtained by throwing away the last row and right column, minus 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.

It is quite apparent that the above theorem will give us the formula for calculating determinants in the 2-by-2 case. This is because when you remove a row and a column from a 2-by-2 matrix you are left with a single element of the matrix. If you look at the left side of the above formula, in the case of a 2-by-2 matrix it simplifies to the familiar formula as we see below:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} b & b \\ c & d \end{vmatrix} = \begin{vmatrix} b & b \\ c & d \end{vmatrix} \begin{vmatrix} a & b \\ c & d \end{vmatrix} - \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

$$\Rightarrow \det = da - bc$$

It was Dodson who noticed that the 2-by-2 formula generalized to larger matrices.

In fact, this theorem can be taken as an inductive definition of the concept of a determinant, so long as the determinant of the matrix obtained by throwing away all outside rows and columns ($A^{jk,jk}$) is non-zero.

The most wonderful thing about this theorem is that you can implement it on a spreadsheet with one instruction. Then, if you start with an n -by- n matrix of 1's (representing the 0-by-0 determinants), then enter your matrix below this matrix, and then enter the key instruction and copy it down and across, it will first compute the 2-by-2 determinants of the 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 number appropriately to certain elements of the original matrix in

order to make all the necessary determinants slightly different than 0. By varying these increments, if necessary, you can eliminate any errors they might introduce.

So, what is this magic instruction?

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 $(q - 1)$ -th row with 1's.

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

In the iteration just before computing the determinant itself, 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 first $n - 1$ 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 same determinant, and on the previous iteration it 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 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.

When n is odd, these values will actually be the cofactors. When n is even, however, the signs of every second entry 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.

Another thing you can do is solve n linear equations in n variables using Cramer's rule and the above procedure. To do this, you enter your matrix followed by the right hand side followed by copies of the first $n - 1$ columns of your matrix.

The row in which you compute the n -by- n determinants of your matrix will contain after it all the numerators you need for Cramer's rule, though some may have the wrong sign, depending on oddness and evenness. From these, you can read off the solutions to your linear equations.

To make all this work with large matrices, you must take steps to avoid dividing by 0 in your process (you can usually get by without it for 3-by-3 matrices).

A way to do this that seems to work is to set up a supplementary matrix of the same shape as the original one, and adding it to the original. This new matrix should be very small (say all entries on the order of 10^{-8}) in a way that does not produce a 0 determinant easily. $10^{-8} \ln(j + k)$ seems to work pretty well, where j and k are the row and column indices.

Exercises

Exercise 1 Multiply out the 7 products above to show that they actually give us the correct product of two 2-by-2 matrices

Exercise 2 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 5-by-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 that does not cause you to divide by 0.

~Edited by Jacob Green