Dynamic Programming: Matrix-Chain Multiplication

Yufei Tao's Teaching Team

Department of Computer Science and Engineering
Chinese University of Hong Kong

Matrix-Chain Multiplication

You are given an algorithm $\mathcal{A}$ that, given an $a \times b$ matrix $\boldsymbol{A}$ and a $b \times c$ matrix $\boldsymbol{B}$, can calculate $\boldsymbol{AB}$ in $O(abc)$ time. You need to use $\mathcal{A}$ to calculate the product of $\boldsymbol{A}_1\boldsymbol{A}_2...\boldsymbol{A}_n$ where $\boldsymbol{A}_i$ is an $a_i \times b_i$ matrix for $i \in [1, n]$. This implies that $b_{i-1} = a_i$ for $i \in [2, n]$, and the final result is an $a_1 \times b_n$ matrix.

A trivial strategy is to apply $\mathcal{A}$ to evaluate the product from left to right. However, we may be able to reduce the cost by following a different multiplication order.

Example

Consider $\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3$ where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are $m \times m$ matrices, but $\boldsymbol{A}_3$ is $m \times 1$.

There are two multiplication orders:

- $(\boldsymbol{A}_1\boldsymbol{A}_2)\boldsymbol{A}_3$.
  The cost of computing $\boldsymbol{B} = \boldsymbol{A}_1\boldsymbol{A}_2$ is $O(m \cdot m \cdot m) = O(m^3)$ and $\boldsymbol{B}$ is an $m \times m$ matrix. The cost of $\boldsymbol{B}\boldsymbol{A}_3$ is $O(m \cdot m \cdot 1) = O(m^2)$. The total cost is $O(m^3)$.
- $\boldsymbol{A}_1(\boldsymbol{A}_2\boldsymbol{A}_3)$.
  The cost of computing $\boldsymbol{B} = \boldsymbol{A}_2\boldsymbol{A}_3$ is $O(m \cdot m \cdot 1) = O(m^2)$ and $\boldsymbol{B}$ is an $m \times 1$ matrix. The cost of $\boldsymbol{A}_1\boldsymbol{B}$ is $O(m \cdot m \cdot 1) = O(m^2)$. The total cost is $O(m^2)$.

Parenthesizing $\boldsymbol{A}_1\boldsymbol{A}_2\ldots\boldsymbol{A}_n$ at $\boldsymbol{A}_k$ for some $k \in [1, n-1]$ converts the expression to $(\boldsymbol{A}_1\ldots\boldsymbol{A}_k)(\boldsymbol{A}_{k+1}\ldots\boldsymbol{A}_n)$, after which you can parenthesize each of $\boldsymbol{A}_1\ldots\boldsymbol{A}_i$ and $\boldsymbol{A}_{i+1}\ldots\boldsymbol{A}_n$ recursively.

A **fully parenthesized product** is

- either a single matrix or
- the product of two fully parenthesized products.

For example, if $n = 4$, then $(\boldsymbol{A}_1\boldsymbol{A}_2)(\boldsymbol{A}_3\boldsymbol{A}_4)$ and $((\boldsymbol{A}_1\boldsymbol{A}_2)\boldsymbol{A}_3)\boldsymbol{A}_4$ are fully parenthesized, but $\boldsymbol{A}_1(\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{A}_4)$ is not.

A fully parenthesized product determines a multiplication order that, in turn, determines the computation cost.

Goal: Design an algorithm to find in $O(n^3)$ time a fully parenthesized product with the smallest cost.

Recursive Structure

By parenthesizing at $\boldsymbol{A}_k$, we obtain

$$\underbrace{(\boldsymbol{A}_1 \ldots \boldsymbol{A}_k)}_{\boldsymbol{B}_1} \underbrace{(\boldsymbol{A}_{k+1} \ldots \boldsymbol{A}_n)}_{\boldsymbol{B}_2},$$

where $\boldsymbol{B}_1$ is an $a_1 \times b_k$ matrix and $\boldsymbol{B}_2$ is an $a_{k+1} \times b_n$ matrix.

The total cost is

$$\text{cost of computing } \boldsymbol{B}_1 + \text{cost of computing } \boldsymbol{B}_2 + O(a_1 b_k b_n).$$

We define $cost(i, j)$, where $1 \le i \le j \le n$, to be the smallest achievable cost for calculating $\boldsymbol{A}_i ... \boldsymbol{A}_j$. Our objective is to calculate $cost(1, n)$.

If we parenthesize $\boldsymbol{A}_i ... \boldsymbol{A}_j$ at $\boldsymbol{A}_k$, we obtain

$$\underbrace{(\boldsymbol{A}_i ... \boldsymbol{A}_k)}_{cost(i,k)} \underbrace{(\boldsymbol{A}_{k+1} ... \boldsymbol{A}_j)}_{cost(k+1,j)}.$$

The total cost is

$$cost(i, k) + cost(k + 1, j) + O(a_i b_k b_j).$$

To attain $cost(i,j)$, we should try all possible parenthesizations of $\boldsymbol{A}_i ... \boldsymbol{A_j}$. This implies:

$$cost(i,j) = \begin{cases} O(1) & \text{if } i = j \\ \min_{k=i}^{j-1}(cost(i,k) + cost(k+1,j) + O(a_i b_k b_j)) & \text{if } i < j \end{cases}$$

By dyn. programming, we can compute $cost(1,n)$ in $O(n^3)$ time.

Consider $\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{A}_4$ where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are $m \times m$ matrices, $\boldsymbol{A}_3$ is $m \times 1$, and $\boldsymbol{A}_4$ is $1 \times m$.

$i \backslash j$	1	2	3	4
1				$cost(1,4)$
2	0			
3	0	0		
4	0	0	0	

After solving all subproblems, we obtain:

$i \backslash j$	1	2	3	4
1	$O(1)$	$O(m^3)$	$O(m^2)$	$O(m^2)$
2	0	$O(1)$	$O(m^2)$	$O(m^2)$
3	0	0	$O(1)$	$O(m^2)$
4	0	0	0	$O(1)$

Next, we apply the "piggyback technique" to generate an optimal parenthesization.

Define $bestSub(i,j) =$

- nil, if $i = j$;
- k, if the best parenthesization for $\boldsymbol{A}_i \boldsymbol{A}_{i+1} \ldots \boldsymbol{A}_j$ is $(\boldsymbol{A}_i \ldots \boldsymbol{A}_k)(\boldsymbol{A}_{k+1} \ldots \boldsymbol{A}_j)$.

$i \backslash j$	1	2	3	4
1	$O(1)$	$O(m^3)$	$O(m^2)$	$O(m^2)$
2	0	$O(1)$	$O(m^2)$	$O(m^2)$
3	0	0	$O(1)$	$O(m^2)$
4	0	0	0	$O(1)$

After $cost(i,j)$ is ready for all i,j, we can compute all $bestSub(i,j)$ in $O(n^3)$ time.

$i \backslash j$	1	2	3	4
1	$O(1)$	$O(m^3)$	$O(m^2)$	$O(m^2)$
2	0	$O(1)$	$O(m^2)$	$O(m^2)$
3	0	0	$O(1)$	$O(m^2)$
4	0	0	0	$O(1)$

$\boldsymbol{A}_1$: $m \times m$
$\boldsymbol{A}_2$: $m \times m$
$\boldsymbol{A}_3$: $m \times 1$
$\boldsymbol{A}_4$: $1 \times m$

Example:
$bestSub(1,4) = 3$, i.e., the best way to calculate $\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{A}_4$ is $(\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3)\boldsymbol{A}_4$.

Similarly, $bestSub(1,3) = 1$, i.e., the best way to calculate $\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3$ is $\boldsymbol{A}_1(\boldsymbol{A}_2\boldsymbol{A}_3)$.

Therefore, an optimal fully parenthesized product of $\boldsymbol{A}_1\boldsymbol{A}_2\boldsymbol{A}_3\boldsymbol{A}_4$ is $(\boldsymbol{A}_1(\boldsymbol{A}_2\boldsymbol{A}_3))\boldsymbol{A}_4$.