

矩阵乘法加速

张腾

2022 年 11 月 5 日

设矩阵 $\mathbf{A} = (a_{ij})$ 和 $\mathbf{B} = (b_{ij})$ 是 n 阶方阵，乘积 $\mathbf{C} = (c_{ij})$ 亦是 n 阶方阵，其中

$$c_{ij} = \sum_{k \in [n]} a_{ik} b_{kj}$$

因此按标准的矩阵乘法，计算 \mathbf{C} 的时间开销为 $\Omega(n^3)$ ，事实上这个时间复杂度是可以改进的。

1 分治递归

设计算 \mathbf{C} 的时间开销为 $T(n)$ ，将矩阵分成 $2 \times 2 = 4$ 块，由分块矩阵乘法有

$$\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}$$

其中包含 8 个 $n/2$ 阶方阵相乘、4 个 $n/2$ 阶方阵相加，注意每个 $n/2$ 阶方阵有 $n^2/4$ 个元素，因此共需进行 n^2 次加法，综上有递推关系

$$T(n) = 8 \cdot T(n/2) + c_1 n^2$$

其中 c_1 为单次加法的时间开销。设 $n = 2^k$ ，则

$$\begin{aligned} T(2^k) &= 8^1 \cdot T(2^{k-1}) + 8^0 \cdot c_1 4^k \\ 8^1 \cdot T(2^{k-1}) &= 8^2 \cdot T(2^{k-2}) + 8^1 \cdot c_1 4^{k-1} \\ 8^2 \cdot T(2^{k-2}) &= 8^3 \cdot T(2^{k-3}) + 8^2 \cdot c_1 4^{k-2} \\ &\vdots \\ 8^{k-1} \cdot T(2^1) &= 8^k \cdot T(2^0) + 8^{k-1} \cdot c_1 4^1 \end{aligned}$$

注意 $8^k = n^3$ ， $T(1) = c_2$ 是单次乘法的时间开销，累加可得

$$\begin{aligned} T(n) &= c_2 n^3 + c_1 4^k + 2^1 \cdot c_1 4^k + 2^2 \cdot c_1 4^k + \cdots + 2^{k-1} \cdot c_1 4^k = c_2 n^3 + c_1 4^k \frac{1 - 2^k}{1 - 2} \\ &= c_2 n^3 + c_1 n^2 (n - 1) = (c_2 + c_1) n^3 - c_1 n^2 \end{aligned}$$

即直接分治递归并不能改进时间复杂度。

2 基本想法

要想改进时间复杂度，必须得减少子问题的个数，即乘法的次数。将乘积拉直，易知

$$\begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} \\ \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} \\ \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{A}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{11} & \mathbf{0} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{21} & \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{12} \\ \mathbf{B}_{21} \\ \mathbf{B}_{22} \end{bmatrix} = \tilde{\mathbf{A}} \begin{bmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{12} \\ \mathbf{B}_{21} \\ \mathbf{B}_{22} \end{bmatrix}$$

现假设 $\tilde{\mathbf{A}}$ 可以分解成 m 个“块秩 1 矩阵”的和：

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{A}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{11} & \mathbf{0} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{21} & \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} = \sum_{i \in [m]} \begin{bmatrix} \mathbf{P}_{i1} \\ \mathbf{P}_{i2} \\ \mathbf{P}_{i3} \\ \mathbf{P}_{i4} \end{bmatrix} \mathbf{R}_i \begin{bmatrix} \mathbf{Q}_{i1} \\ \mathbf{Q}_{i2} \\ \mathbf{Q}_{i3} \\ \mathbf{Q}_{i4} \end{bmatrix}^T \quad (1)$$

其中 \mathbf{R}_i 只由 $\mathbf{A}_{11}, \mathbf{A}_{12}, \mathbf{A}_{21}, \mathbf{A}_{22}$ 进行加减运算得到且 $\mathbf{P}_{i1}, \dots, \mathbf{P}_{i4}, \mathbf{Q}_{i1}, \dots, \mathbf{Q}_{i4} \in \{\pm \mathbf{I}, \mathbf{0}\}$ 。则

$$\begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} \\ \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} \\ \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix} = \sum_{i \in [m]} \begin{bmatrix} \mathbf{P}_{i1} \\ \mathbf{P}_{i2} \\ \mathbf{P}_{i3} \\ \mathbf{P}_{i4} \end{bmatrix} \mathbf{R}_i \begin{bmatrix} \mathbf{Q}_{i1} \\ \mathbf{Q}_{i2} \\ \mathbf{Q}_{i3} \\ \mathbf{Q}_{i4} \end{bmatrix}^T \begin{bmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{12} \\ \mathbf{B}_{21} \\ \mathbf{B}_{22} \end{bmatrix} = \sum_{i \in [m]} \begin{bmatrix} \mathbf{P}_{i1} \\ \mathbf{P}_{i2} \\ \mathbf{P}_{i3} \\ \mathbf{P}_{i4} \end{bmatrix} \mathbf{R}_i \mathbf{S}_i = \sum_{i \in [m]} \begin{bmatrix} \mathbf{P}_{i1} \\ \mathbf{P}_{i2} \\ \mathbf{P}_{i3} \\ \mathbf{P}_{i4} \end{bmatrix} \mathbf{T}_i$$

其中 $\mathbf{S}_i = \mathbf{Q}_{i1}\mathbf{B}_{11} + \mathbf{Q}_{i2}\mathbf{B}_{12} + \mathbf{Q}_{i3}\mathbf{B}_{21} + \mathbf{Q}_{i4}\mathbf{B}_{22}$ 只由 $\mathbf{B}_{11}, \mathbf{B}_{12}, \mathbf{B}_{21}, \mathbf{B}_{22}$ 进行加减运算得到。计算全部 m 个 $\mathbf{T}_i = \mathbf{R}_i \mathbf{S}_i$ 会产生 m 个子问题。又 $\mathbf{P}_{i1}, \dots, \mathbf{P}_{i4} \in \{\pm \mathbf{I}, \mathbf{0}\}$ ，因此

$$\begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} \\ \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} \\ \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix} = \sum_{i \in [m]} \begin{bmatrix} \mathbf{P}_{i1} \\ \mathbf{P}_{i2} \\ \mathbf{P}_{i3} \\ \mathbf{P}_{i4} \end{bmatrix} \mathbf{T}_i = \begin{bmatrix} \mathbf{P}_{11}\mathbf{T}_1 + \dots + \mathbf{P}_{m1}\mathbf{T}_m \\ \mathbf{P}_{12}\mathbf{T}_1 + \dots + \mathbf{P}_{m2}\mathbf{T}_m \\ \mathbf{P}_{13}\mathbf{T}_1 + \dots + \mathbf{P}_{m3}\mathbf{T}_m \\ \mathbf{P}_{14}\mathbf{T}_1 + \dots + \mathbf{P}_{m4}\mathbf{T}_m \end{bmatrix}$$

只由 $\mathbf{T}_1, \dots, \mathbf{T}_m$ 进行加减运算得到。综上，关键就是如何使式 (1) 中的 $m < 8$ 。

下面给出一个 $m = 7$ 的分解方法，首先去掉左上的 \mathbf{A}_{11} 和右下的 \mathbf{A}_{22}

$$\begin{aligned} \tilde{\mathbf{A}} - \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{A}_{11} & \mathbf{0} & \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} & \mathbf{A}_{22} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{A}_{12} - \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{11} - \mathbf{A}_{22} & \mathbf{0} & \mathbf{A}_{12} - \mathbf{A}_{22} \\ \mathbf{A}_{21} - \mathbf{A}_{11} & \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{21} - \mathbf{A}_{22} & \mathbf{0} & \mathbf{0} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{11} - \mathbf{A}_{22} & \mathbf{A}_{11} - \mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{11} & \mathbf{A}_{22} - \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{A}_{12} - \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{22} - \mathbf{A}_{11} & \mathbf{A}_{12} - \mathbf{A}_{22} \\ \mathbf{A}_{21} - \mathbf{A}_{11} & \mathbf{A}_{11} - \mathbf{A}_{22} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{21} - \mathbf{A}_{22} & \mathbf{0} & \mathbf{0} \end{bmatrix} \end{aligned}$$

$$\begin{aligned}
&= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathbf{A}_{11} - \mathbf{A}_{22} & \mathbf{A}_{11} - \mathbf{A}_{22} & 0 \\ 0 & \mathbf{A}_{22} - \mathbf{A}_{11} & \mathbf{A}_{22} - \mathbf{A}_{11} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{A}_{22} - \mathbf{A}_{12} & \mathbf{A}_{12} - \mathbf{A}_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \mathbf{A}_{12} - \mathbf{A}_{11} & 0 \\ 0 & 0 & \mathbf{A}_{12} - \mathbf{A}_{11} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\
&\quad + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \mathbf{A}_{21} - \mathbf{A}_{11} & \mathbf{A}_{11} - \mathbf{A}_{21} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \mathbf{A}_{21} - \mathbf{A}_{22} & 0 & 0 \\ 0 & \mathbf{A}_{21} - \mathbf{A}_{22} & 0 & 0 \end{bmatrix} \\
\Rightarrow \tilde{\mathbf{A}} &= \begin{bmatrix} \mathbf{I} \\ 0 \\ \mathbf{I} \\ 0 \end{bmatrix} \mathbf{A}_{11} \begin{bmatrix} \mathbf{I} \\ 0 \\ \mathbf{I} \\ 0 \end{bmatrix}^\top + \begin{bmatrix} 0 \\ \mathbf{I} \\ 0 \\ \mathbf{I} \end{bmatrix} \mathbf{A}_{22} \begin{bmatrix} 0 \\ \mathbf{I} \\ 0 \\ \mathbf{I} \end{bmatrix}^\top + \begin{bmatrix} 0 \\ \mathbf{I} \\ -\mathbf{I} \\ 0 \end{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{22}) \begin{bmatrix} 0 \\ \mathbf{I} \\ \mathbf{I} \\ 0 \end{bmatrix}^\top + \begin{bmatrix} 0 \\ \mathbf{I} \\ 0 \\ 0 \end{bmatrix} (\mathbf{A}_{12} - \mathbf{A}_{22}) \begin{bmatrix} 0 \\ 0 \\ -\mathbf{I} \\ \mathbf{I} \end{bmatrix}^\top \\
&\quad + \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \\ 0 \\ 0 \end{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}) \begin{bmatrix} 0 \\ 0 \\ -\mathbf{I} \\ 0 \end{bmatrix}^\top + \begin{bmatrix} 0 \\ 0 \\ \mathbf{I} \\ 0 \end{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{21}) \begin{bmatrix} -\mathbf{I} \\ \mathbf{I} \\ 0 \\ 0 \end{bmatrix}^\top + \begin{bmatrix} 0 \\ 0 \\ \mathbf{I} \\ \mathbf{I} \end{bmatrix} (\mathbf{A}_{21} - \mathbf{A}_{22}) \begin{bmatrix} 0 \\ \mathbf{I} \\ 0 \\ 0 \end{bmatrix}^\top
\end{aligned}$$

3 算法实现

根据上面的分解易知计算

$$\begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \mathbf{S}_3 \\ \mathbf{S}_4 \\ \mathbf{S}_5 \\ \mathbf{S}_6 \\ \mathbf{S}_7 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 & \mathbf{I} & 0 \\ 0 & \mathbf{I} & 0 & \mathbf{I} \\ 0 & \mathbf{I} & \mathbf{I} & 0 \\ 0 & 0 & -\mathbf{I} & \mathbf{I} \\ 0 & 0 & -\mathbf{I} & 0 \\ -\mathbf{I} & \mathbf{I} & 0 & 0 \\ 0 & \mathbf{I} & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{12} \\ \mathbf{B}_{21} \\ \mathbf{B}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{11} + \mathbf{B}_{21} \\ \mathbf{B}_{12} + \mathbf{B}_{22} \\ \mathbf{B}_{12} + \mathbf{B}_{21} \\ \mathbf{B}_{22} - \mathbf{B}_{21} \\ -\mathbf{B}_{21} \\ \mathbf{B}_{12} - \mathbf{B}_{11} \\ \mathbf{B}_{12} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \mathbf{R}_3 \\ \mathbf{R}_4 \\ \mathbf{R}_5 \\ \mathbf{R}_6 \\ \mathbf{R}_7 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} \\ \mathbf{A}_{22} \\ \mathbf{A}_{11} - \mathbf{A}_{22} \\ \mathbf{A}_{12} - \mathbf{A}_{22} \\ \mathbf{A}_{11} - \mathbf{A}_{12} \\ \mathbf{A}_{11} - \mathbf{A}_{21} \\ \mathbf{A}_{21} - \mathbf{A}_{22} \end{bmatrix}$$

共会产生 10 次加减运算, 计算 $\mathbf{T}_1 = \mathbf{R}_1\mathbf{S}_1, \dots, \mathbf{T}_7 = \mathbf{R}_7\mathbf{S}_7$ 共会产生 7 个子问题, 最后计算

$$\begin{aligned}
\begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} \\ \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} \\ \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix} &= \begin{bmatrix} \mathbf{I} \\ 0 \\ \mathbf{I} \\ 0 \end{bmatrix} \mathbf{T}_1 + \begin{bmatrix} 0 \\ \mathbf{I} \\ 0 \\ \mathbf{I} \end{bmatrix} \mathbf{T}_2 + \begin{bmatrix} 0 \\ \mathbf{I} \\ -\mathbf{I} \\ 0 \end{bmatrix} \mathbf{T}_3 + \begin{bmatrix} 0 \\ \mathbf{I} \\ 0 \\ 0 \end{bmatrix} \mathbf{T}_4 + \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \\ 0 \\ 0 \end{bmatrix} \mathbf{T}_5 + \begin{bmatrix} 0 \\ 0 \\ \mathbf{I} \\ 0 \end{bmatrix} \mathbf{T}_6 + \begin{bmatrix} 0 \\ 0 \\ \mathbf{I} \\ \mathbf{I} \end{bmatrix} \mathbf{T}_7 \\
&= \begin{bmatrix} \mathbf{T}_1 + \mathbf{T}_5 \\ \mathbf{T}_2 + \mathbf{T}_3 + \mathbf{T}_4 + \mathbf{T}_5 \\ \mathbf{T}_1 - \mathbf{T}_3 + \mathbf{T}_6 + \mathbf{T}_7 \\ \mathbf{T}_2 + \mathbf{T}_7 \end{bmatrix}
\end{aligned}$$

共会产生 8 次加减运算。

综上，一共会产生 7 个子问题和 18 次加减运算，此时递推关系变成

$$T(n) = 7 \cdot T(n/2) + \frac{18}{4}c_1n^2$$

设 $n = 2^k$ ，则

$$\begin{aligned}T(2^k) &= 7^1 \cdot T(2^{k-1}) + \frac{18}{4}c_14^k \\7^1 \cdot T(2^{k-1}) &= 7^2 \cdot T(2^{k-2}) + 7^1 \frac{18}{4}c_14^{k-1} \\7^2 \cdot T(2^{k-2}) &= 7^3 \cdot T(2^{k-3}) + 7^2 \frac{18}{4}c_14^{k-2} \\&\vdots \\7^{k-1} \cdot T(2^1) &= 7^k \cdot T(2^0) + 7^{k-1} \frac{18}{4}c_14^1\end{aligned}$$

注意 $7^k = (2^{\lg 7})^k = (2^k)^{\lg 7} = n^{\lg 7} \approx n^{2.81}$ ，累加可得

$$T(n) = c_2n^{\lg 7} + \frac{18}{4}c_14^k \frac{1 - (7/4)^k}{1 - (7/4)} = c_2n^{\lg 7} + 6c_1(n^{\lg 7} - n^2) = (c_2 + 6c_1)n^{\lg 7} - 6c_1n^2$$