Алгоритм Копперсмита-Винограда

Мрыхин Михаил

Хотим быстро множить матрицы.

1 Тензорная форма записи

Пишем умножение матриц как $\sum_{i,j,k=1}^{m,n,p} a_{ij}b_{jk}c_{ki}$ для удобства последующих выкладок. Здесь c_{ki} - формальные переменные, двойственные $(AB)_{ik}$ (т.е. коэффициенты при них - элементы матрицы-произведения).

2 τ-теорема Шёнхаге

Если есть такие коэффициенты $\alpha_{i,j,h,l}, \beta_{j,k,h,l}, \gamma_{k,i,h,l}$, что

$$\sum_{l=1}^{L}\left(\sum_{i,j,h}\alpha_{i,j,h,l}x_{i,j}^{(h)}\right)\left(\sum_{j,k,h}\beta_{j,k,h,l}y_{j,k}^{(h)}\right)\left(\sum_{k,i,h}\beta_{k,i,h,l}z_{k,i}^{(h)}\right)=\sum_{h}\left(\sum_{i,j,k=1}^{m_h,n_h,p_h}x_{i,j}^{(h)}y_{j,k}^{(h)}z_{k,i}^{(h)}\right),$$

и при этом $L=\sum_h (m_hn_hp_h)^\tau$, то $\omega \leq 3\tau$. (ω - это нижний предел асимптотической скорости умножения матриц, т.е. мы не обязаны умножать за $O(n^\omega)$, но за $O(n^{\omega+\epsilon})$ для любого $\epsilon > 0$).

Это обобщённая форма того, что происходит в Штрассене - мы пытаемся через небольшое количество произведений линейных комбинаций элементов выразить несколько произведений матриц. Суть доказательства тоже примерно такая же - рекурсивно оцениваем, сколько умножений нам понадобится для асимптотически больших матриц.

Теорема обобщается дальше, если α, β и γ сделать функциями от λ, а произведение матриц считать с точностью до $O(\lambda)$.

3 Теорема Салема-Спенсера

Это было без доказательства:

Для любого $\epsilon > 0$ существует такое M_ϵ, что для $M > M_\epsilon$ найдётся подмножество $B \subset [1, ..., \left\lfloor \frac{M-1}{2} \right\rfloor]$, свободное от арифметических последовательностей длины 3, и при этом $|B| \geq M^{1-\epsilon}$.

4 Суть самой теоремы

Фиксируем q. Будем координатам x, y и z сопоставлять верхние индексы в зависимости от того, нулевая эта координата или нет (0 и 1 соответственно).

Берётся алгоритм, который за $q+2$ умножений приблизительно считает кривоватое произведение векторов $(\sum x_0^{[0]}y_i^{[1]}z_i^{[1]} + x_i^{[1]}y_0^{[0]}z_i^{[1]} + x_i^{[1]}y_i^{[1]}z_0^{[0]})$. Эта конструкция тензорно множится с собой $3N$ раз.

После этого мы хэшируем все тройки в итоговом произведении с помощью случайных весов, помноженных на верхние индексы. Выкидываем (присваиванием одному из множителей 0) все тройки, у которых не ровно треть нулей в каждом индексе, те, у которых хэши не лежат в множестве из теоремы Салема-Спенсера, и те, у которых хэши совпали. После чего замечаем, что оставшиеся тройки не имеют общих переменных, каждая из них считает произведение двух матриц $q^N \times q^N$, а их количество для какого-то выбора весов в хэше достаточно высоко, и применяем теорему Шёнхаге.