Спецкурс 2020/2021: “Геометрические и комбинаторные свойства матриц и аппроксимация” Блок лекций “Сложность матриц и аппроксимация” Лекция 3: “Жёсткие матрицы”

24 ноября 2020 г.

Жесткость матрицы и линейные схемы

Жесткость матрицы (Rigidity):

$$\mathrm{Rig}(A, r) := \min_{\operatorname{rank} B \leqslant r} \#\{(i, j) \colon A_{i,j} \neq B_{i,j}\}.$$

Разминка!

- вычислите $\mathrm{Rig}(\mathrm{Id}_n, r)$;
- оцените $\mathrm{Rig}(A, r)$ сверху для произвольной матрицы $A \in \mathbb{R}^{n \times n}$;
- оценить $\mathrm{Rig}(\Delta_n, n/100)$ для верхнетреугольной $\{0, 1\}$-матрицы Δ_n.

Понятие жёсткости возникло в теории сложности в контексте *линейных схем*, вычисляющих линейные функций $x \mapsto Ax$. Работа Leslie Valiant-а 1977 года.

Схема состоит из узлов, на входе n узлов-переменных $x_1, \ldots, x_n$, на выходе должны быть узлы y_i, так чтобы $(y_1, \ldots, y_n) = Ax$, промежуточные узлы: элементы сложения (с двумя входами) и умножения на скаляр. Узлы соединены в **ориентированный ацикличный граф**.

- размер схемы = количество рёбер;
- глубина = длина максимального пути;

Можно вычислить любое отображение $\mathbb{F}^n \to \mathbb{F}^n$ схемой размера $O(n^2)$ и глубины $O(\log n)$. Почему?

Theorem (Valiant)

Если линейное отображение $A\colon \mathbb{F}^n \to \mathbb{F}^n$ можно вычислить схемой размера s и глубины d, то для любого $t > 1$

$$\mathrm{Rig}(A, \frac{s \log t}{\log d}) \leqslant n2^{Cd/t}.$$

Следствие: если для некоторых $\varepsilon, \delta > 0$ имеем $\mathrm{Rig}(A, \varepsilon n) \geqslant n^{1+\delta}$), то схема, вычисляющая преобразование $x \mapsto Ax$ и имеющая логарифмическую глубину, имеет размер не менее $c(\varepsilon, \delta) n \log \log n$.

Для случайных матриц над бесконечным полем $\mathrm{Rig}(A, r) = (n - r)^2$.

Проблема: построить явное семейство жёстких матриц $n \times n$: $\mathrm{Rig}(A_n, \varepsilon n) \geqslant n^{1+\delta}$.

Поработаем с ориентированными ацикличными графами.

Назовём *разметкой* графа $G = (V, E)$ отображение $L \colon V \to \mathbb{Z}$, такое что для всякого ребра $(u, v) \in E$ имеем $L(u) < L(v)$.

Если есть разметка $L \colon V \to \{1, 2, \ldots, d\}$, то глубина графа не больше d.

Всякий граф глубины d может быть размечен числами $\{1, 2, \ldots, d\}$. Почему?

Для доказательства теоремы Valiant-а нам потребуется утверждение из теории графов. Его доказал...Valiant.

Lemma

Пусть (V, E) — ориентированный ацикличный граф глубины не выше d и задано число r. Тогда можно удалить из графа не более $|E| \cdot r / \log_2 d$ рёбер так, что глубина оставшегося графа будет не выше $d/2^r$.

В лемме для простоты считаем, что d равно степени двойки.

Доказательство.

Рассмотрим разметку $L\colon V \to \{0, 1, \ldots, d-1\}$ и отождествим $\{0, 1, \ldots, d-1\}$ с двоичными строками длины $\log_2 d$ (двоичная запись числа).

Возьмём ребро $(u, v) \in E$, тогда $L(u) < L(v)$. Пусть старший бит, где отличаются $L(u)$ и $L(v)$ это i-й бит. Через E_i обозначим множество таких рёбер.

Пусть мы удалили E_i. Тогда из разметки можно удалить i-й бит и свойство разметки будет выполнено!

Т.к. разметка принимает $(\log_2 d - 1)$-битные значения, получится глубина не больше $d/2$.

При удалении r множеств E_i получим глубину не более $d/2^r$.

Выберем $E_{i_1}, \ldots, E_{i_r}$ минимальной мощности. Выбираем r штук из $\log_2 d$ с суммарной мощностью $|E|$, получим $|E_{i_1}| + \ldots + |E_{i_s}| \leqslant |E| \cdot r / \log_2 d$.

$\square$

Перейдём к доказательству теоремы Valiant-a. Пусть для матрицы A есть схема размера s; приблизим A матрицей малого ранга.

Положим $t = 2^r$ и удалим $m \leqslant |E| \cdot r/\log_2 d = sr/\log_2 d$ рёбер так, чтобы остался граф глубины не более d/t.

В каждой вершине графа вычисляется линейная форма от входов $x_1, \ldots, x_n$. Пусть $b_1, \ldots, b_{m'}$ — линейные формы в вершинах на концах удалённых рёбер ($m' \leqslant m$).

Рассмотрим фиксированную выходную вершину. Как она вычисляется? Пройдём от неё вверх и посмотрим: используются либо формы b_i, либо непосредственно входные вершины x_j. В силу того, что глубина графа не более d/t, различных входных вершин не более $2^{d/t}$. Запишем это:

$$y_i = \sum_{k=1}^{m'} \beta_{i,k} b_k(x) + \sum_{j \in \Lambda_i} c_{i,j} x_j, \quad |\Lambda_i| \leqslant 2^{d/t}.$$

$$y_i = \sum_{k=1}^{m'} \beta_{i,k} b_k(x) + \sum_{j \in \Lambda_i} c_{i,j} x_j, \quad |\Lambda_i| \leqslant 2^{d/t}.$$

В матричных терминах:

$$y = Ax, \quad A = \beta B + C,$$

где

- матрица $\beta = (\beta_{i,k})$ размера $n \times m'$,
- матрица B — в которой по строкам стоят коэфф-ты линейных форм $b_k(x)$ — размера $m' \times n$;
- в матрице C не более $2^{d/t}$ ненулевых элементов в каждой строке;

Значит, мы представили A в виде $\beta B + C$, где
$\mathrm{rank}(\beta B) \leqslant m' \leqslant m \leqslant sr / \log_2 d = s \log_2 t / \log_2 d$, $\|C\|_0 \leqslant n2^{d/t}$,

$$\mathrm{Rig}(A, \frac{s \log_2 t}{\log_2 d}) \leqslant n2^{d/t}.$$

Итак, мы доказали, что если A вычисляется простой схемой, то A не слишком жёсткая, т.е. можно приблизить в метрике Хэмминга матрицей малого ранга. Заметим, что это приближение *регулярно*, то есть число отличий в каждой строке небольшое.

Оценки снизу для конкретных матриц

Lemma

Пусть $r \geqslant \log^2 n$. Если в матрице $n \times n$ поменять не более

$$\frac{n(n-r)}{2r+2} \log \frac{n}{r}$$

эл-тов, то некоторый минор $(r+1) \times (r+1)$ будет без изменений.

Предположим, мы сделали изменения в матрице. Рассмотрим двудольный граф с долями $\{v_1, \ldots, v_n\}$ и $\{w_1, \ldots, w_n\}$, где ребро $v_i \mapsto w_j$ проводится для тех (i, j), для которых значение $A_{i,j}$ не изменилось. При этом количество рёбер в графе не меньше

$$n^2 - \frac{n(n-r)}{2r+2} \log \frac{n}{r}.$$

Нам нужно доказать, что полученный граф содержит $K_{r+1,r+1}$. Для этого воспользуемся утверждением из теории графов.

Lemma (Zarankiewich problem)

Если двудольный граф с долями размера m и n не содержит $K_{s,t}$, то количество рёбер в нём не превосходит

$$(s-1)^{1/t}(n-t+1)m^{1-1/t}+(t-1)m.$$

Матричная формулировка: если в матрице из $\{0,1\}^{m\times n}$ более указанного числа единиц, то найдётся подматрица $s\times t$ из одних единиц.

Доказательство. Пусть $|V_1|=m$, $|V_2|=n$ — доли графа. Рассмотрим t-множества $T\subset V_2$, $|T|=t$. Скажем, что $x\in V_1$ *покрывает* T, если x соединён со всеми элементами T.

Каждый $x\in V_1$ покрывает $\binom{d(x)}{t}$ множеств T. С другой стороны, каждое T покрыто не более чем $(s-1)$ точкой (иначе образуется $K_{s,t}$). Следовательно,

$$\sum_{x\in V_1}\binom{d(x)}{t}\leqslant (s-1)\binom{n}{t}.$$

$$\sum_{x\in V_1}\binom{d(x)}{t}\leqslant (s-1)\binom{n}{t}.$$

Посмотрим на биномиальный коэффициент как на многочлен $f(u)=\binom{u}{t}=u(u-1)\cdots(u-t+1)/t!$, это выпуклая функция при $u\geqslant t$, следовательно,

$$\binom{m^{-1}\sum d(x)}{t}\leqslant \frac{s-1}{m}\binom{n}{t}.$$

Обозначим $y=m^{-1}\sum d(x)=|E|/m$. Тогда

$$\binom{y}{t}\leqslant \frac{s-1}{m}\binom{n}{t}.$$

Ясно, что $y\leqslant n$, поэтому тем более

$$(y-t+1)^t\leqslant \frac{s-1}{m}(n-t+1)^t.$$

Это и есть нужное нам неравенство.

Следствие: если все миноры матрицы A невырождены, то

$$\mathrm{Rig}(A, r) \geqslant \frac{n^2}{4r+4} \log \frac{n}{r}$$

при $\log^2 n \leqslant r \leqslant n/2$.

Примеры:

- матрица Коши $(\frac{1}{x_i+y_j})_{i,j=1}^n$;
- $F = (\omega^{ij})$, ω – примитивный n-корень из единицы.

Матрицы Уолша–Адамара

Вспомним про матрицы Уолша–Адамара H^n:

- размера $2^n \times 2^n$ с элементами ± 1;
- $H^n(x, y) = (-1)^{\langle x, y \rangle}$, $x, y \in \{0, 1\}^n$;
- строки и столбцы ортогональны;
- сложная для коммуникации даже с неограниченной ошибкой: $U(H^n) \geqslant cn$.

Докажем следующую оценку жесткости:

$$\mathrm{Rig}(H^n, r) \geqslant \frac{N^2}{4r}, \quad \text{где } N = 2^n,$$

при условии что r это степень двойки.
Зафиксируем x_1 и y_1, получим разбиение H^n на четыре подматрицы $\pm H^{n-1}$.
Обобщим: возьмём $s \geqslant 1$ и разделим H^n на подматрицы $\pm H^s$.
Получится $N^2/2^{2s}$ штук. Если мы делаем меньше $N^2/4r$ изменений, то в одной из подмариц изменится менее $2^{2s}/4r$ элементов.
Полагаем $2^s = 2r$, тогда на одну из матриц $\pm H^s$, которая имеет ранг $2^s = 2r$, приходится менее $2^{2s}/4r = r$ изменений и у неё останется $\mathrm{rank} \geqslant r$.

Такая оценка впервые была получена в работе Б.С.Кашина и А.А.Разборова (1998) для общих матриц Адамара. Отметим, что она недостаточна для Valiant-жёсткости.

В работе 2016 года J.Alman, R. Williams получили неожиданный результат – матрицы Уолша–Адамара не являются жёсткими!

Theorem (Alman, Williams)

Для любого поля $\mathbb{F}$*, достаточно малого* $\varepsilon > 0$*,* $N = 2^n$*, имеем*

$$\mathrm{Rig}^{\mathbb{F}}(H^n, N^{1-c\varepsilon^2}) \leqslant N^{1+c\varepsilon \log(1/\varepsilon)}.$$

Доказательство (случай $\mathbb{F} \subset \mathbb{Q}$)

Пусть $p(x, y)$ — полином от $2n$ переменных ($x, y \in \{0,1\}^n$), состоящий из m мономов. Тогда матрица

$$M(x, y) = p(x, y), \quad x, y \in \{0,1\}^n,$$

имеет $\operatorname{rank} M \leqslant m$.
Действительно, каждый моном имеет вид $u(x)v(y)$ и представляет одноранговую матрицу.

В нашем случае можно взять $p(x, y) = R(x_1 y_1 + \ldots + x_n y_n)$, где полином R альтернирует: $R(j) = (-1)^j$ для $2n\varepsilon \leqslant j \leqslant (1+\varepsilon)n$.
Чтобы уменьшить количество мономов, мы заменим $x_i^m y_i^m \mapsto x_i y_i$; это не изменит значения полинома для булевых векторов.
Количество мономов оценивается

$$m(p) \leqslant \sum_{s=0}^{\deg Q} \binom{n}{s} \leqslant 2^{h_2(r/n)} \leqslant 2^{n(1-c\varepsilon^2)}.$$

Полагаем $M(x, y) = p(x, y)$, $\operatorname{rank} M \leqslant m(p)$.

По построению $M(x, y) = H^n(x, y)$ при $\langle x, y \rangle \in [2n\varepsilon, (1/2 + \varepsilon)n]$:

$$M(x, y) = p(x, y) = Q(\langle x, y \rangle) = (-1)^{\langle x, y \rangle}.$$

Исправление M: полагаем $M'(x, y) = M(x, y)$ для "ядра" $\|x\|_1, \|y\|_1 \in [(1/2 - \varepsilon)n, (1/2 + \varepsilon)n]$, и $M'(x, y) = H^n(x, y)$ иначе. Изменения касаются малого кол-ва строк/столбцов и не сильно увеличат ранг.
Отличие $M'(x, y) \neq H^n(x, y)$ может быть только для пар (x, y) вне ядра и только для $\langle x, y \rangle \notin [2n\varepsilon, (1/2 + \varepsilon)n]$. Следовательно, все расхождения содержат среди пар (x, y):

$$\begin{cases} \langle x, y \rangle < 2n\varepsilon, \\ \|x\|_1 \in [(1/2 - \varepsilon)n, (1/2 + \varepsilon)n], \\ \|x\|_1 \in [(1/2 - \varepsilon)n, (1/2 + \varepsilon)n], \end{cases} \qquad (*)$$

Упражнение: для фиксированного x существует не более $2^{c\varepsilon \log(1/\varepsilon)n}$ таких y, что выполнено (*). Матрица M' даёт нужное приближение для H^n.

Пример жёсткой матрицы. Пусть $p_{i,j}$, $1 \leqslant i, j \leqslant n$ — различные простые числа (например, первые n^2 простых). Рассмотрим матрицу $P = (\sqrt{p_{i,j}}) \in \mathbb{R}^{n \times n}$.

Statement

$\mathrm{Rig}(P, n/17) \geqslant n^2/17.$

Это утверждение следует из теоремы

Theorem

Пусть $A \in \mathbb{R}^{n \times n}$ и $1 \leqslant r \leqslant n$. Если любые nr произведений различных элементов A линейно независимы над $\mathbb{Q}$, то

$$\mathrm{Rig}(A, r) \geqslant n(n - 16r).$$

Известно, что все числа вида $\sqrt{k}$, где k свободно от квадратов, линейно независимы над $\mathbb{Q}$. Следовательно, к матрице P применима данная теорема.

Для доказательства нам потребуется несколько определений. Пусть $X = (a_1, \dots, a_p)$ — последовательность чисел и $t \in \mathbb{N}$. Определим размерности Shoup–Smolensky:

$$D_t(X) := \dim_{\mathbb{Q}} \langle a_{i_1} \cdots a_{i_t} : 1 \leqslant i_1 < i_2 < \ldots < i_t \leqslant p \rangle,$$

$$D_t^*(X) := \dim_{\mathbb{Q}} \langle a_{i_1} \cdots a_{i_t} : 1 \leqslant i_1 \leqslant i_2 \leqslant \ldots \leqslant i_t \leqslant p \rangle.$$

Выполнены простые свойства:

- $D_t(X) \leqslant D_t^*(X)$;
- $D_t(X) \leqslant \binom{p}{t}$;
- $D_t^*(X) \leqslant \binom{p+t-1}{t}$.
- Пусть $X = (a_1, \dots, a_p)$, $Y = (b_1, \dots, b_q)$, $XY := (a_i b_j)_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant q}}$. Тогда

$$D_t^*(XY) \leqslant D_t^*(X) D_t^*(Y).$$

Верно ли это для D_t?

Для матрицы $A \in \mathbb{R}^{m\times n}$ величины $D_t(A)$ и $D_t^*(A)$ определяются как соответствующие размерности для списка элементов матрицы (в произвольном порядке). Если произведение матриц AB определено, то

$$D_t^*(AB) \leqslant D_t^*(A) D_t^*(B).$$

Statement

Если $A \in \mathbb{R}^{m\times n}$ и $\operatorname{rank} A = r$, то

$$D_t^*(A) \leqslant \binom{mr+t-1}{t}\binom{nr+t-1}{t}.$$

Действительно, матрица ранга r представляется в виде $A = BC$ размеров $m \times r$ и $r \times n$. Далее применяем оценку $D_t^*(X) \leqslant \binom{|X|+t-1}{t}$.

Вернёмся к доказательству теоремы. Нам дано, что все произведения nr элементов линейно независимы. Нужно оценить жёсткость. Предположим, $A = B + C$, где $\operatorname{rank} B \leqslant r$ и $\|C\|_0 \leqslant R$; оценим R снизу.

$$D_t(B) \leqslant D_t^*(B) \leqslant \binom{nr+t}{t}^2.$$

С другой стороны, если рассматривать произведения элементов B, где $C_{i,j} = 0$ (и, следовательно, $B_{i,j} = A_{i,j}$, то они также линейно независимы над $\mathbb{Q}$. Следовательно,

$$D_t(B) \geqslant \binom{n^2-R}{t}.$$

Сравним неравенства:

$$\binom{n^2-R}{t} \leqslant \binom{nr+t}{t}^2.$$

Положим $t = nr$:

$$\binom{n^2 - R}{nr} \leqslant \binom{2nr}{nr}^2 \leqslant 2^{4nr}.$$

Воспользуемся полезным неравенством

$$(n/k)^k \leqslant \binom{n}{k} \leqslant (en/k)^k, \quad 1 \leqslant k \leqslant n.$$

Получим

$$((n^2 - R)/nr)^{nr} \leqslant 2^{4nr},$$

$$(n^2 - R)/nr \leqslant 16, \quad R \geqslant n^2 - 16nr,$$

Ч.т.д.

- S.V. Lokam, “Complexity Lower Bounds using Linear Algebra”, 2009.
- L. Valiant, “Graph-theoretic arguments in low-level Complexity”, 1977.
- J. Alman, R. Williams, “Probabilistic Rank and Matrix Rigidity”, 2016, arXiv:1611.05558.