Институт прикладной математики им. М. В. Келдыша
Российской Академии Наук
Московский государственный университет им. М.В. Ломоносова
Механико-математический факультет

ДИСКРЕТНАЯ МАТЕМАТИКА И ЕЕ ПРИЛОЖЕНИЯ

СБОРНИК ЛЕКЦИЙ МОЛОДЕЖНЫХ НАУЧНЫХ ШКОЛ ПО ДИСКРЕТНОЙ МАТЕМАТИКЕ И ЕЕ ПРИЛОЖЕНИЯМ

V

Москва 2009

СОВЕРШЕННОЕ ЛИНЕЙНОЕ ХЕШИРОВАНИЕ В БУЛЕВОМ КУБЕ

А. В. Чашкин

Московский государственный университет им. М. В. Ломоносова, механико-математический факультет, 119992 Москва, Ленинские горы

e-mail: chashkin@inbox.ru

Рассмотрим произвольную область D в n-мерном булевом кубе $\{0,1\}^n$. Совершенным линейным хешированием области D называется линейное инъективное отображение этой области в булев куб меньшей размерности.

1. Инъективные линейные операторы

Большинство утверждений об инъективных на заданной области линейных операторах основаны на следующем простом факте: *линейный оператор f инъективен на области D тогда и только тогда, когда его ядро не пересекается с множеством $D^\star$ попарных сумм элементов области D.* Действительно, если линейный оператор f инъективно действует на области D, т. е. $f(\boldsymbol{x}_i) \neq f(\boldsymbol{x}_j)$ для любых $\boldsymbol{x}_i$ и $\boldsymbol{x}_j$ из D, то

$$f(\boldsymbol{x}_i \oplus \boldsymbol{x}_j) = f(\boldsymbol{x}_i) \oplus f(\boldsymbol{x}_j) \neq \mathbf{0}. \tag{1}$$

Следовательно, $\boldsymbol{x}_i \oplus \boldsymbol{x}_j \notin \ker f$. Поэтому из (1) следует, что множество $D^\star$ и ядро оператора f не пересекаются. Легко видеть, что верно и обратное: если множество $D^\star$ и подпространство $\mathbb{H} \subseteq \{0,1\}^n$ не пересекаются, то $\mathbb{H}$ является ядром линейного оператора, отображающего несовпадающие наборы области D в несовпадающие наборы ее образа. Рассмотрим подпространство $\mathbb{H}$, не имеющее общих наборов с $D^\star$, и линейный оператор f, ядром которого является $\mathbb{H}$. Пусть $\boldsymbol{x}_i$ и $\boldsymbol{x}_j$ — произвольные наборы из D. Так как $\boldsymbol{x}_i \oplus \boldsymbol{x}_j \notin \mathbb{H} = \ker f$, то

$$f(\boldsymbol{x}_i) \oplus f(\boldsymbol{x}_j) = f(\boldsymbol{x}_i \oplus \boldsymbol{x}_j) \neq \mathbf{0},$$

т. е. образы наборов $\boldsymbol{x}_i$ и $\boldsymbol{x}_j$ различны.

Имеет место следующая верхняя оценка на ранг инъективного на произвольном множестве линейного оператора.

Теорема 1. *Для любой области $D \subseteq \{0,1\}^n$, состоящей не более чем из $\sqrt{2^n}$ наборов, найдется инъективный на этой области линейный (m,n)-оператор, для числа компонент которого справедливо неравенство*

$$m \leq \lfloor 2\log_2 |D| \rfloor - 1.$$

Теорема 1 является простым следствием доказываемой далее теоремы 2.

Теорема 2. *Пусть для множества $D^\star$ попарных сумм элементов области $D \subseteq \{0,1\}^n$ справедливо неравенство*

$$2^{m+1} > |D^\star| + 1.$$

Тогда существует инъективный на области D линейный (m,n)-оператор.

Доказательство теоремы 2 основано на последовательном применении ее частного случая — доказываемой ниже леммы.

Лемма 1. *Пусть для множества $D^\star$ попарных сумм элементов области $D \subseteq \{0,1\}^n$ справедливо неравенство*

$$2^n > |D^\star| + 1.$$

Тогда существует инъективный на области D линейный $(n-1,n)$-оператор.

Доказательство. Для построения требуемого линейного оператора достаточно найти в $\{0,1\}^n$ подпространство $\mathbb{H}$, которое не пересекается с множеством $D^\star$ и размерность которого равна единице. Существование такого пространства легко следует из условий леммы. Так как $2^n > |D^\star| + 1$, то среди элементов $\{0,1\}^n$ найдется ненулевой набор $\boldsymbol{y}$, не принадлежащий $D^\star$, который вместе с нулевым набором будет образовывать требуемое одномерное подпространство. Лемма доказана.

Доказательство теоремы 2. Воспользуемся леммой 1. Из этой леммы следует существование такого линейного $(n-1,n)$-оператора f_1, что $f_1(\boldsymbol{x}) \neq f_1(\boldsymbol{y})$ для любых неравных наборов $\boldsymbol{x}$ и $\boldsymbol{y}$ из D. Далее для множества D будем использовать обозначение D_0. Через D_1 обозначим образ области D_0 при действии f_1. Легко видеть, что мощность множества $D_1^\star$,

состоящего из попарных сумм различных элементов множества D_1, не превосходит мощности множества $D_0^\star$. Действительно, если это не так, то в D_0 должны присутствовать такие наборы $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ и $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, что $\boldsymbol{x}_1 \oplus \boldsymbol{x}_2 = \boldsymbol{y}_1 \oplus \boldsymbol{y}_2$ и $f_1(\boldsymbol{x}_1) \oplus f_1(\boldsymbol{x}_2) \neq f_1(\boldsymbol{y}_1) \oplus f_1(\boldsymbol{y}_2)$. Однако очевидно, что только одно из этих соотношений может быть справедливым.

Если $2^{n-1} > |D_0^\star|$, то $2^{n-1} > |D_1^\star|$, и поэтому можно снова воспользоваться леммой 1, применив ее к новому множеству D_1. Из этой леммы следует существование линейного $(n-2, n-1)$-оператора f_2 такого, что $f_2(\boldsymbol{x}) \neq f_2(\boldsymbol{y})$ для любых неравных наборов $\boldsymbol{x}$ и $\boldsymbol{y}$ из D_1. Положим $D_2 = f_1(D_1)$. Как и в предыдущем случае, легко видеть, что $|D_2^\star| \leq |D_1^\star|$. Заметим, что композиция $f_2 \circ f_1$ операторов f_2 и f_1 будет инъективным на D линейным $(n-2, n)$-оператором.

Предположим, что описанную процедуру выполнили в общей сложности $k-1$ раз и для каждого целого i от единицы до $k-1$ получили инъективный на области D_{i-1} линейный $(n-i+1, n-i)$-оператор f_i и лежащее в $\{0,1\}^{n-i}$ множество D_i такие, что $D_i = f_i(D_{i-1})$, $|D_i^\star| \leq |D_{i-1}^\star|$, а композиция $f_{k-1} \circ \cdots \circ f_1$ является инъективным на области D линейным $(n-k+1, n)$-оператором.

Если $2^{n-k+1} > |D_{k-1}^\star| + 1$, то лемму 1 можно применить еще раз. Так как по предположению линейный $(n-k+1, n)$-оператор $f_{k-1} \circ \cdots \circ f_1$ отображает разные наборы области D в разные наборы ее образа D_{k-1}, а линейный $(n-k, n-k+1)$-оператор f_k действует на D_{k-1} инъективно, то легко видеть, что композиция $f_k \circ (f_{k-1} \circ \cdots \circ f_1)$, полученных в результате применения леммы 1 операторов f_i, будет инъективным на области D линейным $(n-k, n)$-оператором.

Наконец заметим, что при $k \leq n-m$ из условий теоремы и сделанного предположения следуют неравенства

$$2^{n-k+1} \geq 2^{m+1} > |D_0^\star| \geq |D_{k-2}^\star| \geq |D_{k-1}^\star|.$$

Поэтому очевидно, что леммой 1 можно воспользоваться в общей сложности не менее $n-m$ раз, а получившийся в результате линейный (m, n)-оператор $f_{n-m} \circ \cdots \circ f_1$ будет инъективным на области D. Теорема 2 доказана.

Доказательство теоремы 1. Так как

$$2^{(\lfloor 2\log_2 |D| \rfloor - 1)+1} > \frac{|D|^2}{2} \geq \frac{|D|(|D|-1)}{2} + 1 \geq |D^*| + 1,$$

то при некотором $m \leq \lfloor 2\log_2 |D| \rfloor - 1$ в силу теоремы 2 найдется инъективный на D линейный (m, n)-оператор. Теорема 1 доказана.

Теперь покажем, что для любого целого m, не превосходящего $n/2$, в n-мерном булевом кубе найдется область D_m, состоящая из $2^{m+1}-1$ наборов, и такая, что число компонент любого инъективного на этой области линейного оператора не меньше $2m$.

Пусть $m \le n/2$ и $\mathbf{e}_1, \ldots, \mathbf{e}_{2m}$ — первые $2m$ базисных векторов стандартного базиса E_n. Положим

$$D_m = \langle \mathbf{e}_1, \ldots, \mathbf{e}_m \rangle \cup \langle \mathbf{e}_{m+1}, \ldots, \mathbf{e}_{2m} \rangle.$$

Легко видеть, что D_m состоит из $2^{m+1}-1$ различных наборов, а множество $D_m^\star$ попарных сумм наборов из D_m вместе с нулевым набором образуют подпространство размерности $2m$ в $\{0,1\}^n$. Поэтому очевидно, что размерность любого подпространства, не пересекающегося с $D_m^\star$, не меньше чем $n-2m$. Следовательно, ранг любого инъективного на области D_m линейного оператора не превосходит $2m$.

Так как $\lfloor 2\log_2 |D_m| \rfloor - 1 = 2m$, то теорема 1 гарантирует существование инъективного на D_m линейного $(2m, n)$-оператора. Таким образом в общем случае неравенство теоремы 1 является точным и усилить его нельзя.

Далее покажем, что неравенство теоремы 1 является асимптотически точным для почти всех областей из $D_{n,N}$ при условии, что $\frac{\log_2 N}{\log_2 n}$ неограниченно возрастает при $n \to \infty$. Сделаем это следующим образом. Сначала для произвольной области D из $\{0,1\}^n$ введем функцию μ, определив ее равенством

$$\mu(D) = \min \operatorname{rank} f,$$

в котором минимум берется по всем инъективным на области D линейным операторам. Затем величину $\mu(D)$ оценим снизу для почти всех областей из $D_{n,N}$.

Теорема 3. *Пусть $n \le N \le 2\sqrt{n^2 2^n}$. Тогда при $n \to \infty$ для почти всех $D \in D_{n,N}$*

$$\mu(D) \ge 2\log_2 N - 2\log_2 n - 2.$$

Доказательство. Пусть f — произвольный линейный (m,n)-оператор. Через $M(f,N)$ обозначим число областей из $D_{n,N}$, на которых оператор f является инъективным. Легко видеть, что для каждой такой области D никакие два набора из D не принадлежат одному и тому же смежному классу пространства $\{0,1\}^n$ по ядру оператора f. Поэтому для любого оператора ранга $k \le m$

$$M(f,N) = \binom{2^k}{N} 2^{(n-k)N} \le \binom{2^m}{N} 2^{(n-m)N}. \qquad (2)$$

Теперь предположим, что при некоторой постоянной δ не менее чем для $\delta\binom{2^n}{N}$ областей из $D_{n,N}$ среди линейных (m,n)-операторов найдутся инъективные на этих областях операторы. Так как число различных линейных (m,n)-операторов равно 2^{mn}, то в среднем каждый линейный (m,n)-оператор является инъективным не менее чем для $\delta\binom{2^n}{d}2^{-mn}$ областей мощности N. Следовательно, найдется оператор, который будет инъективным по крайней мере для

$$P = \delta\binom{2^n}{N}2^{-mn}$$

различных областей. С другой стороны необходимо, чтобы величина P не превосходила $M(f,N)$. Поэтому из (2) и последнего неравенства

$$\delta\binom{2^n}{N}2^{-mn} \le \binom{2^m}{N}2^{(n-m)N}.$$

Откуда после несложных преобразований получаем

$$\binom{2^n}{N}\Big/\binom{2^m}{N} \le \frac{1}{\delta}2^{mn}2^{(n-m)N}. \tag{3}$$

Легко видеть, что

$$\binom{2^n}{N}\Big/\binom{2^m}{N} = \frac{2^n(2^n-1)\dots(2^n-N+1)}{2^m(2^m-1)\dots(2^m-N+1)} \tag{4}$$

Оценим снизу натуральный логарифм правой части последнего равенства. Так как функция $\ln\frac{a-x}{b-x}$ выпукла вниз при $a > b > 0$ и $x \in [0,b)$, то

$$\ln\frac{2^n(2^n-1)\dots(2^n-N+1)}{2^m(2^m-1)\dots(2^m-N+1)} \ge N\ln\frac{2^n-(N-1)/2}{2^m-(N-1)/2} \ge$$
$$\ge N\ln 2^{n-m} + N\ln\frac{1-(N-1)/2\cdot 2^n}{1-(N-1)/2\cdot 2^m}. \tag{5}$$

Теперь оценим последнее слагаемое в правой части (5). Для этого используем справедливое при $1 > y \ge x \ge 0$ неравенство $\frac{1-x}{1-y} \ge 1-x+y$ и справедливое при $x \in [0,1)$ неравенство $\ln(1+x) \ge \frac{x}{2}$. Так как $m < n$ и $N < 2^m$, то

$$N\ln\frac{1-(N-1)/2\cdot 2^n}{1-(N-1)/2\cdot 2^m} \ge N\ln\Big(1+\frac{N-1}{2}\Big(\frac{1}{2^m}-\frac{1}{2^n}\Big)\Big) \ge \frac{N(N-1)}{4\cdot 2^m}. \tag{6}$$

Из (3)–(6) следует, что

$$\ln\Big(\frac{1}{\delta}2^{mn}2^{(n-m)N}\Big) \geq N\ln 2^{n-m} + \frac{N(N-1)}{4\cdot 2^m},$$

или, после очевидных преобразований,

$$2^m \geq \frac{N(N-1)}{4(nm\ln 2 - \ln\delta)}.$$

Логарифмируя последнее неравенство по основанию 2, видим, что при любой постоянной δ, начиная с некоторого n имеет место неравенство

$$m \geq 2\log_2 N - 2\log_2 n - 2.$$

Теорема доказана.

2. Сложность инъективных линейных операторов

Теорема 4. *Для любой постоянной $\varepsilon > 0$ и любой области $D \subseteq \{0,1\}^n$, состоящей не более чем из $\sqrt{2^n}$ наборов, найдется инъективный на этой области линейный (m,n)-оператор, для числа компонент которого справедливо неравенство*

$$m \leq (2+\varepsilon)\log_2 |D|,$$

и сложность которого есть $\mathcal{O}(n)$.

Будем говорить, что i-я строка двоичной матрицы M покрывает ее j-й столбец, если в M на пересечении i-й строки и j-го столбца стоит единица.

Лемма 2. *Найдется такая постоянная $\delta > 0$, что для любого достаточно большого n существует двоичная матрица $M_{2n,n}$ из $2n$ строк и n столбцов, в каждой строке которой находится ровно 7 единиц и в которой $(\star)$ при любом k, не превосходящем $2n\delta$, любые k строк покрывают более чем $4k$ столбцов.*

Доказательство. Пусть R — множество всех матриц, состоящих из $2n$ строк и n столбцов, во всех строках которых находится ровно 7 единиц. Оценим величину N, равную отношению числа тех матриц из R, которые не обладают свойством $(\star)$, к числу всех матриц из R. Нетрудно видеть, что k строк, в которых единицы сосредоточены на пересечении не более чем с $4k$ столбцами, можно выбрать $\binom{2n}{k}$ способами, а соответствующие им столбцы — $\binom{n}{4k}$ способами, единицы в выбранных строках можно расставить

не более чем $\binom{4k}{7}^k$ способами, в оставшихся строках это можно сделать $\binom{n}{7}^{2n-k}$ способами. Поэтому

$$
\begin{aligned}
N &\le \sum_{k=1}^{2n\delta} \binom{2n}{k}\binom{n}{4k}\binom{4k}{7}^k\binom{n}{7}^{2n-k}\binom{n}{7}^{-2n} = \\
&= \sum_{k=1}^{2n\delta} \binom{2n}{k}\binom{n}{4k}\binom{4k}{7}^k\binom{n}{7}^{-k} \le \\
&\le \sum_{k=1}^{2n\delta} \Big(\frac{3\cdot 2n}{k}\Big)^k \Big(\frac{3\cdot n}{4k}\Big)^{4k} \Big(\frac{4k(4k-1)\dots(4k-6)}{n(n-1)\dots(n-6)}\Big)^k \le \\
&\le \sum_{k=1}^{2n\delta} \Big(\frac{3\cdot 2n}{k}\Big)^k \Big(\frac{3n}{4k}\Big)^{4k} \Big(\frac{4k}{n}\Big)^{7k} = \sum_{k=1}^{2n\delta} 3^{5k}2^{7k}\Big(\frac{k}{n}\Big)^{2k} < \sum_{k=1}^{2n\delta} \big(3^5 2^7 \delta^2\big)^k .
\end{aligned}
$$

Нетрудно видеть, что при при выполнении неравенства $3^5 2^7 \delta^2 \le 2^{-1}$ (которое, очевидно, справедливо при $\delta < 2^{-8}$) отношение N будет меньше единицы, и, следовательно, найдется матрица, удовлетворяющая условиям леммы. Лемма доказана.

Лемма 3. *Найдется такая постоянная $\delta > 0$, что для любого достаточно большого n существует двоичная матрица $M_{2n,n}$ из $2n$ строк и n столбцов, в каждой строке которой находится ровно 7 единиц и в которой при любом k, не превосходящем $2n\delta$, в каждой подматрице, образованной k строками, найдется более k столбцов содержащих ровно один единичный элемент.*

Доказательство. Пусть матрица M удовлетворяет условию $(\star)$ из леммы 2. В этой матрице произвольным образом выберем k строк и составим из них подматрицу M' матрицы M. Пусть R_1 обозначает число столбцов, покрываемых ровно одной из выбранных строк, а $R_{\ge 2}$ — число столбцов, покрываемых более чем одной такой строкой. Другими словами, R_1 равно числу столбцов подматрицы M' содержащих ровно по одному единичному элементу, а $R_{\ge 2}$ равно числу столбцов с более чем одним единичным элементом. В силу леммы 2 величины R_1 и $R_{\ge 2}$ удовлетворяют следующим неравенствам

$$
\begin{aligned}
R_1 + R_{\ge 2} &> 4k, \\
R_1 + 2R_{\ge 2} &\le 7k.
\end{aligned}
$$

Исключая из этих неравенств $R_{\geq 2}$, имеем

$$R_1 > 4k - R_{\geq 2} \geq 4k - \frac{1}{2}(7k - R_1) = \frac{1}{2}k + \frac{1}{2}R_1,$$

т. е. $R_1 > k$. Лемма доказана.

Лемма 4. *Пусть $M_{2n,n}$ — матрица из леммы 3. Тогда для любого двоичного набора v веса k, где $k \leq 2n\delta$, произведение $\boldsymbol{v} \cdot M_{2n,n}$ содержит более k единичных элементов.*

Доказательство. Пусть в наборе $\boldsymbol{v}$ компоненты $\boldsymbol{v}_{i_1}, \ldots, v_{i_k}$ ненулевые. В матрице $M_{2n,n}$ рассмотрим подматрицу M, образованную строками с номерами $i_1, \ldots, i_k$. В силу леммы 3 в этой подматрице найдутся столбцы с номерами $j_1, \ldots, j_s$ где $s > k$, каждый из которых содержит ровно один единичный элемент. Легко видеть, что для любого j_i из $\{j_1, \ldots, j_s\}$ скалярное произведение $\boldsymbol{v}$ и j_i-го столба матрицы $M_{2n,n}$ равно единице. Следовательно, произведение $\boldsymbol{v} \cdot M_{2n,n}$ содержит $s > k$ единичных элементов. Лемма доказана.

Лемма 5. *Существует такая постоянная $0 < \gamma < 1$, что для любого достаточно большого n найдется такое линейное отображение $\mathcal{G}_{n,4n}$ из $\{0,1\}^n$ в $\{0,1\}^{4n}$, что $\|\mathcal{G}_{n,4n}(\boldsymbol{v})\| \geq \gamma n$ для любого ненулевого вектора $\boldsymbol{v}$ и $L(\mathcal{G}_{n,4n}) = \mathcal{O}(n)$.*

Доказательство. Лемму докажем индукцией по двоичному логарифму n. В основание индукции положим отображение $\mathcal{G}_{m,4m}(\boldsymbol{v}) = (\boldsymbol{v}, \boldsymbol{v}, \boldsymbol{v}, \boldsymbol{v})$, где m — минимально возможное, при котором существует матрица $M_{4m,2m}$. Очевидно, что в этом случае $\gamma = 4/m$ и $L(\mathcal{G}_{m,4m}) = 0$.

Теперь допустим, что удовлетворяющее условиям леммы линейное отображение $\mathcal{G}_{n,4n}$ с матрицей $G_{n,4n}$ существует при некотором $n \geq m$. Используя это отображение построим отображение $\mathcal{G}_{2n,8n}$. Пусть $\boldsymbol{v}$ — вектор длины $2n$, $\boldsymbol{v}' = \boldsymbol{v} \cdot M_{2n,n}$ — вектор длины n, $\boldsymbol{w} = \boldsymbol{v}' \cdot G_{n,4n}$ — вектор длины $4n$, $\boldsymbol{u} = \boldsymbol{w} \cdot M_{4n,2n}$ — вектор длины $2n$. Тогда $\mathcal{G}_{2n,8n}(\boldsymbol{v}) = (\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{u})$. Покажем, что неравенство

$$\|\mathcal{G}_{2n,8n}(\boldsymbol{v})\| \geq 2n\gamma \tag{7}$$

справедливо для любого ненулевого вектора $\boldsymbol{v}$ длины $2n$ и $\gamma = \min(4/m, \delta)$, где δ — постоянная из леммы 4.

Если $\|\boldsymbol{v}\| \geq 2n\gamma$, то, очевидно, имеет место и неравенство (7). Если вес ненулевого вектора $\boldsymbol{v}$ меньше чем $2n\gamma$, то в силу леммы 4 вес вектора $\boldsymbol{v}'$ больше нуля, и в силу предположения индукции $\|\boldsymbol{w}\| \geq n\delta$. Если при этом

справедливо более сильное неравенство $\|\boldsymbol{w}\| \geq 2n\gamma$, то (7) также справедливо. Если же $n\gamma < \|\boldsymbol{w}\| < 2n\gamma$, то в силу леммы 4 вес вектора $\boldsymbol{u}$ больше веса вектора $\boldsymbol{w}$ и поэтому $\|\mathcal{G}_{2n,8n}(\boldsymbol{v})\| > \|\boldsymbol{w}\| + \|\boldsymbol{u}\| > 2n\gamma$.

Покажем, что $L(\mathcal{G}_{n,4n}) = \mathcal{O}(n)$. Допустим, что $L(\mathcal{G}_{n,4n}) \leq 42n$ при $n > m$. Так как сложность умножения матрицы на вектор не превосходит числа единичных элементов матрицы, то

$$L(\mathcal{G}_{2n,8n}) \leq L(M_{2n,n}) + L(\mathcal{G}_{n,4n}) + L(M_{4n,2n}) \leq \\ \leq 14n + 42n + 28n = 42 \cdot 2n.$$

Лемма доказана.

Лемма 6. *Пусть целые t и R удовлетворяют неравенствам $R \geq 256$ и $2 \leq t \leq \log_2(\frac{1}{2}\log_2 R)$. Положим $m = \lceil \log_2 R \cdot (1 + 4/2^t) \rceil$. Тогда для любых наборов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R$ из $\{0,1\}^n$, вес каждого из которых не меньше $d = \delta n$, где δ — константа, найдется такой линейный оператор $\mathcal{L}$ из $\{0,1\}^n$ в $\{0,1\}^m$, что $L(\mathcal{L}) = \mathcal{O}(n)$ и ни один из наборов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R$ не отображается этим оператором в нулевой набор.*

Доказательство. Пусть $M(n,m)$ — множество булевых матриц с m строками и n столбцами, p — постоянная из $(0, \frac{1}{2})$. В $M(n,m)$ случайным образом выберем матрицу M, полагая, что элементы M_{ij} этой матрицы выбираются независимо с вероятностями $\mathbf{Pr}(M_{ij} = 1) = p$ и $\mathbf{Pr}(M_{ij} = 0) = 1-p$. Пусть $\boldsymbol{a}$ — двоичный набор длины n и веса d. Найдем вероятность того, что линейный оператор $\mathcal{M}$ с выбранной матрицей M отображает набор $\boldsymbol{a}$ в нулевой набор. Нетрудно видеть, что для i-й компоненты $\mathcal{M}_i$ оператора $\mathcal{M}$ справедливы равенства

$$\mathbf{Pr}(\mathcal{M}_i(\boldsymbol{a}) = 0) = \sum_{k=0}^{\lfloor d/2 \rfloor} \binom{d}{2k} p^{2k}(1-p)^{d-2k},$$

$$\mathbf{Pr}(\mathcal{M}_i(\boldsymbol{a}) = 1) = \sum_{k=0}^{\lfloor d/2 \rfloor} \binom{d}{2k} p^{2k+1}(1-p)^{d-2k-1}.$$

Поэтому, учитывая равенство

$$\sum_{k=0}^{d} \binom{d}{k} (-p)^k (1-p)^{d-k} = ((1-p) - p)^d,$$

находим, что

$$\mathbf{Pr}(\mathcal{M}_i(\boldsymbol{a}) = 0) + \mathbf{Pr}(\mathcal{M}_i(\boldsymbol{a}) = 1) = 1,$$

$$\mathbf{Pr}(\mathcal{M}_i(\boldsymbol{a}) = 0) - \mathbf{Pr}(\mathcal{M}_i(\boldsymbol{a}) = 1) = (1-2p)^d.$$

Таким образом,

$$\mathbf{Pr}(\mathcal{M}_i(\boldsymbol{a}) = 0) = \frac{1}{2}\big(1 + (1-2p)^d\big),$$

и, следовательно, для искомой вероятности справедливо равенство

$$\mathbf{Pr}(\mathcal{M}(\boldsymbol{a}) = 0) = \Big(\frac{1}{2}\Big(1 + (1-2p)^d\Big)\Big)^m.$$

Тогда,

$$\begin{aligned}\mathbf{Pr}(\mathcal{M}(\boldsymbol{a}) = 0) &= \frac{1}{2^m}\Big(1 + (1-2p)^{\frac{1}{2p}\cdot 2pd}\Big)^m \le \frac{1}{2^m}\Big(1 + \frac{1}{2^{2pd}}\Big)^m \le \\ &\le \frac{1}{2^m}\Big(1 + \frac{1}{2^{2pd}}\Big)^{2^{2pd}\cdot m/2^{2pd}} \le \frac{1}{2^m}\cdot 4^{m/2^{2pd}} = 2^{-m(1-2/2^{2pd})}.\end{aligned}$$

Нетрудно видеть, что полученная оценка вероятности $\mathbf{Pr}(\mathcal{M}(\boldsymbol{a}) = 0)$ убывает с ростом d.

Положим $p = t/2d$. В этом случае

$$\mathbf{Pr}(\mathcal{M}(\boldsymbol{a}) = 0) \le \frac{1}{2^m}\cdot 4^{m/2^{2pd}} = 2^{-m(1-2/2^t)}.$$

Поэтому для любых $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R$ из A справедливо неравенство

$$\mathbf{Pr}(\mathcal{M}(\boldsymbol{a}_1) = 0 \vee \cdots \vee \mathcal{M}(\boldsymbol{a}_R) = 0) \le R \cdot 2^{-m(1-2/2^t)}.$$

Положим $m = \lceil \log_2 R \cdot (1 + 4/2^t) \rceil$. Тогда при $t \ge 3$

$$m(1-2/2^t) \ge \log_2 R \cdot (1+4/2^t)(1-2/2^t) \ge \log_2 R \cdot (1 + 1/2^t).$$

Подставляя полученную оценку в предыдущее неравенство, имеем

$$\mathbf{Pr}(\mathcal{M}(\boldsymbol{a}_1) = 0 \vee \cdots \vee \mathcal{M}(\boldsymbol{a}_R) = 0) \le R \cdot R^{-1-2^{-t}} = R^{-2^{-t}},$$

откуда после несложных преобразований находим, что

$$\mathbf{Pr}(\mathcal{M}(\boldsymbol{a}_1) = 0 \vee \cdots \vee \mathcal{M}(\boldsymbol{a}_R) = 0) \le \frac{1}{4} \tag{8}$$

при $3 \le t \le \log_2(\frac{1}{2}\log_2 R)$.

Теперь оценим $||M||$ — число единичных элементов в матрице M. Это число является случайной величиной ξ, математическое ожидание $\mathbf{M}\xi$ и

дисперсия $\mathbf{D}\xi$ которой равны, соответственно, $mnp = tmn/2d$ и $nmp(1-p) \leq mnt/2d$. Положим $s = 2\sqrt{\mathbf{D}\xi}$. Тогда из неравенства Чебышева следует, что

$$\mathbf{Pr}(|\xi - \mathbf{M}\xi| \geq s) \leq \frac{\mathbf{D}\xi}{s^2} = \frac{1}{4}. \tag{9}$$

Таким образом, вероятность того, что число ненулевых элементов матрицы M больше $tmn/2d + \sqrt{2tmn/d} < 2tmn/d$ не превосходит $1/4$.

Объединяя неравенства (8) и (9), видим, что

$$\begin{aligned}\mathbf{Pr}(\mathcal{M}(\boldsymbol{a}_1) \neq 0 \,\&\, \ldots \,\&\mathcal{M}(\boldsymbol{a}_R) \neq 0 \,\&\, ||M|| < tm/\delta) = \\ = 1 - \mathbf{Pr}(\mathcal{M}(\boldsymbol{a}_1) = 0 \vee \cdots \vee \mathcal{M}(\boldsymbol{a}_R) = 0 \vee ||M|| \geq tm/\delta) \geq \\ \geq 1 - \mathbf{Pr}(\mathcal{M}(\boldsymbol{a}_1) = 0 \vee \cdots \vee \mathcal{M}(\boldsymbol{a}_R) = 0) - \mathbf{Pr}(||M|| \geq tm/\delta) \geq \frac{1}{2},\end{aligned}$$

где $m = \lceil \log_2 R \cdot (1 + 4/2^t) \rceil$ и $3 \leq t \leq \log_2(\frac{1}{2}\log_2 R)$.

Следовательно, для любых $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R$ найдется линейный оператор $\mathcal{M} : \{0,1\}^n \to \{0,1\}^m$, который не отображает ни один из этих наборов в нулевой набор и в матрице которого находится не более tm/δ единиц. Так как $m \leq n$, а δ и t — константы, то очевидно, что сложность оператора $\mathcal{M}$ есть $\mathcal{O}(n)$. Лемма доказана.