ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

**Том 58** **2022** **Вып. 4**

УДК 621.391 : 519.72

© 2022 г. **А. Джанабекова, Г.А. Кабатянский**[1], **И. Камель, Т.Ф. Рабие**

# НЕПЕРЕКРЫВАЮЩИЕСЯ ВЫПУКЛЫЕ МНОГОГРАННИКИ С ВЕРШИНАМИ ИЗ БУЛЕВА КУБА И ДРУГИЕ ЗАДАЧИ ТЕОРИИ КОДИРОВАНИЯ

Устанавливается связь между несколькими задачами, которые, на первый взгляд, довольно далеки друг от друга, и формулируется ряд открытых проблем.

*Ключевые слова*: выпуклые многогранники, булев куб, групповое тестирование, поиск фальшивых монет, сигнатурные коды для каналов с множественным доступом и шумом, мультимедийные коды цифровых отпечатков пальцев, определение носителя вектора по линейным измерениям с шумом.

**DOI:** 10.31857/S0555292322040052, **EDN:** ECJVMW

## § 1. Неперекрывающиеся выпуклые многогранники с вершинами из булева куба, поиск фальшивых монет на точных весах и другие задачи

Начнем с выпуклых многогранников, вершины которых берутся из булева куба $B^n = \{0,1\}^n \subset \mathbb{R}^n$. Мы будем рассматривать множества $A \subset B^n$ мощности не более $t$ и их выпуклые оболочки (выпуклые многогранники)

$$\langle A\rangle = \left\{ \sum_{\boldsymbol{a}\in A} \lambda_{\boldsymbol{a}} \boldsymbol{a} : \sum_{\boldsymbol{a}\in A} \lambda_{\boldsymbol{a}} = 1,\ \lambda_{\boldsymbol{a}} \geqslant 0 \right\}.$$

Точка

$$\boldsymbol{x} = \sum_{\boldsymbol{a}\in A} \lambda_{\boldsymbol{a}} \boldsymbol{a}$$

с неотрицательными коэффициентами $\lambda_{\boldsymbol{a}}$, такими что $\sum\limits_{\boldsymbol{a}\in A} \lambda_{\boldsymbol{a}} = 1$, называется *выпуклой комбинацией* точек множества $A$, а если коэффициенты $\lambda_{\boldsymbol{a}} > 0$ для всех точек из $A$, то про точку $\boldsymbol{x}$ будем говорить, что она представима в виде *строго выпуклой комбинации* точек из $A$ и называть ее *строго внутренней точкой* выпуклой оболочки $\langle A\rangle$.

Определение 1. Множество $\mathcal{C}$ вершин $n$-мерного булева куба $B^n$ называется *$t$-независимым*, если для любых его двух непересекающихся подмножеств $A, B \subset \mathcal{C}$, таких что $|A|, |B| \leqslant t$, их выпуклые оболочки не пересекаются.

Это определение, как мы сейчас покажем, эквивалентно следующему.

[1] Исследование выполнено при финансовой поддержке Российского научного фонда в рамках гранта РНФ 22-41-02028.

Определение 2. Множество $\mathcal{C}$ вершин $n$-мерного булева куба $B^n$ называется *$t$-независимым*, если для любых двух различных подмножеств $A, B \subset \mathcal{C}$, таких что $|A|, |B| \leqslant t$, их выпуклые оболочки не пересекаются по строго внутренней точке.

Предложение 1. *Определения* 1 *и* 2 *эквивалентны.*

Доказательство. $1 \to 2$. Пусть это не так, т.е. условия определения 1 выполнены, но имеются два различных подмножества $A, B \subset \mathcal{C}$, таких что $|A|, |B| \leqslant t$ и их выпуклые оболочки пересекаются по *строго* внутренней точке $\boldsymbol{x}$. Тем самым,

$$\boldsymbol{x} = \sum_{\boldsymbol{a} \in A} \lambda_{\boldsymbol{a}} \boldsymbol{a} = \sum_{\boldsymbol{b} \in B} \lambda'_{\boldsymbol{b}} \boldsymbol{b}, \tag{1}$$

где все коэффициенты $\lambda$ положительны и

$$\sum_{\boldsymbol{a} \in A} \lambda_{\boldsymbol{a}} = \sum_{\boldsymbol{b} \in B} \lambda'_{\boldsymbol{b}} = 1.$$

Рассмотрим множества

$$\Delta := A \cap B, \quad \Delta_A := \{i \in \Delta : \ \lambda_i > \lambda'_i\}, \quad \Delta_B := \{i \in \Delta : \ \lambda_i < \lambda'_i\}.$$

Из уравнения (1) следует, что

$$\sum_{\boldsymbol{a} \in A - \Delta} \lambda_{\boldsymbol{a}} \boldsymbol{a} + \sum_{\boldsymbol{a} \in \Delta_A} (\lambda_{\boldsymbol{a}} - \lambda'_{\boldsymbol{a}}) \boldsymbol{a} = \sum_{\boldsymbol{b} \in B - \Delta} \lambda'_{\boldsymbol{b}} \boldsymbol{b} + \sum_{\boldsymbol{b} \in \Delta_B} (\lambda'_{\boldsymbol{b}} - \lambda_{\boldsymbol{b}}) \boldsymbol{b}. \tag{2}$$

Легко видеть, что сумма коэффициентов в левой части и правой части уравнения (2) одна и та же, поэтому пронормировав на это число (сумму) левую и правую часть (2), получим, что выпуклые оболочки непересекающихся множеств $(A - \Delta) \cup \cup \Delta_A$ и $(B - \Delta) \cup \Delta_B$ пересекаются. Противоречие.

$2 \to 1$. Пусть это не так, т.е. условия определения 2 выполнены, но имеются два непересекающихся подмножества $A, B \subset \mathcal{C}$, таких что $|A|, |B| \leqslant t$ и их выпуклые оболочки пересекаются в некоторой точке $\boldsymbol{x}$. Тем самым,

$$\boldsymbol{x} = \sum_{\boldsymbol{a} \in A} \lambda_{\boldsymbol{a}} \boldsymbol{a} = \sum_{\boldsymbol{b} \in B} \lambda'_{\boldsymbol{b}} \boldsymbol{b}, \tag{3}$$

где все коэффициенты $\lambda$ неотрицательны и $\sum\limits_{\boldsymbol{a} \in A} \lambda_{\boldsymbol{a}} = \sum\limits_{\boldsymbol{b} \in B} \lambda'_{\boldsymbol{b}} = 1$. Введем множества

$$\widehat{A} := \{a \in A : \ \lambda_{\boldsymbol{a}} > 0\}, \quad \widehat{B} := \{b \in B : \ \lambda_{\boldsymbol{b}} > 0\}.$$

Тогда $\boldsymbol{x}$ является общей строго внутренней точкой для двух различных (и даже непересекающихся) множеств $\widehat{A}$ и $\widehat{B}$. Противоречие. ▲

Определение 1 представляется нам более естественным, а определение 2 – это новая математическая формулировка задачи о мультимедийных кодах, устойчивых к произвольным линейным атакам коалиций из не более чем $t$ легальных пользователей (см. [1] и обзор [2]). Действительно, линейная атака коалиции $A \subset \mathcal{C}$ означает, что эта коалиция может подставить ложный вектор $\widehat{\boldsymbol{a}}$, сформированный как взвешенная сумма векторов, соответствующих членам коалиции, т.е.

$$\widehat{\boldsymbol{a}} = \sum_{\boldsymbol{a} \in A} p_{\boldsymbol{a}} \boldsymbol{a},$$

где все $p_{\boldsymbol{a}} \geqslant 0$ и $\sum\limits_{\boldsymbol{a} \in A} p_{\boldsymbol{a}} = 1$. Тогда использование $t$-независимого кода гарантирует, согласно определению 2, что данный ложный вектор $\widehat{\boldsymbol{a}}$ является строго внутренним

для единственной коалиции, и тем самым, все пользователи, внесшие *ненулевой* вклад в создание ложного вектора, могут быть однозначно определены. Отметим, что для обычных кодов цифровых отпечатков пальцев [3,4] это свойство – однозначное нахождение коалиции активных пользователей – принципиально недостижимо.

Обозначим через $M_1(t\,|\,n)$ максимальную мощность $t$-независимых множеств $n$-мерного булева куба и рассмотрим частный случай $t = 2$. Определение 1 говорит, что множество $\mathcal{C}$ вершин булева куба является 2-независимым, если для любых четырех различных точек $a, b, c, d \in \mathcal{C}$ отрезки $[a, b]$ и $[c, d]$ не пересекаются. Заметим, что если отрезки $[a, b]$ и $[c, d]$ пересекаются, то точка пересечения – это середина каждого из отрезков. Действительно, любая точка отрезка, концы которого являются вершинами булева куба, имеет все координаты из множества $\{0, \gamma, 1-\gamma, 1\}$ для некоторого $\gamma \leqslant 1/2$. Если $\gamma < 1/2$ (т.е. это не середина отрезка), то по координатам этой точки концы отрезка восстанавливаются однозначно.

Следовательно, условие 2-независимости равносильно тому, что для любых четырех различных точек $a, b, c, d \in \mathcal{C}$ справедливо $a+b \neq c+d$ (то свойство, что никакая точка-вершина куба не может быть выпуклой комбинацией других точек-вершин, очевидно). В свою очередь, это аналог хорошо известного определения последовательности Сидона [5, 6] или $B_2$-последовательности [7], примененного к булевому кубу как подмножеству $\mathbb{R}^n$, или, что то же самое, сигнатурный код для двоичного суммирующего канала с двумя активными пользователями [8, 9].

Пример. Легко проверить, что для кода

$$\mathcal{C} = \{(0,0,0), (0,1,1), (1,1,0), (1,0,1), (1,1,1)\}$$

все попарные суммы векторов различны, т.е. код является 2-независимым.

С другой стороны, несложно показать, что не существует 2-независимого кода мощности 6, и следовательно, $M_1(2\,|\,3) = 5$.

Ниже мы увидим, что при любом фиксированном $t$ величина $M_1(t\,|\,n)$ растет экспоненциально от $n$ и поэтому естественно рассматривать "скорость" $R_t^{(1)}$ ее роста, определяемую как

$$R_t^{(1)} := \lim_{n\to\infty} n^{-1} \log_2 M_1(t\,|\,n), \tag{4}$$

где предел не обязан существовать, и строго говоря, нужно рассматривать соответствующие верхний и/или нижний пределы.

Вернемся к случаю $t = 2$. Так как столбцы проверочной матрицы кода, исправляющего две ошибки, являются $B_2$-последовательностью в группе $\mathbb{Z}_2^n$, т.е. их попарные суммы различны по модулю 2, то следовательно, эти суммы различны и как целые числа. Взяв примитивный код БЧХ (или неприводимый код Гоппы), получим, что $R_2^{(1)} \geqslant 1/2$.

С другой стороны, из очевидного аналога границы Хэмминга

$$M_1(t\,|\,n)(M_1(t\,|\,n) - 1) \leqslant 3^n$$

(см. общий случай (7) ниже) следует, что

$$R_2^{(1)} \leqslant 0{,}5 \log_2 3.$$

Лучшая известная верхняя граница

$$R_2^{(1)} \leqslant 0{,}5753$$

была получена в [7].

Теперь перейдем к тематике комбинаторного поиска, а именно к задаче поиска фальшивых монет на точных весах. Рассмотрим два крайних варианта постановки задачи, различающиеся тем, какая информация известна априори. Пусть имеется $m$ монет, из которых не более $t$ фальшивых. В первой, традиционной постановке задачи предполагается, что все правильные монеты имеют вес $\alpha$, все фальшивые – вес $\beta$, причем величины $\alpha, \beta$ известны заранее. Будем называть это задачей поиска (фальшивых монет) с полной информацией.

Во второй постановке задачи известно лишь, что все правильные монеты имеют один и тот же вес, а вот веса фальшивых монет могут различаться. Будем называть это задачей поиска с минимальной априорной информацией.

Обозначим через $M_2(t\,|\,n)$ максимальное число монет, среди которых можно найти $t$ фальшивых не более чем за $n$ взвешиваний для задачи поиска с полной информацией, и через $M_3(t\,|\,n)$ – для задачи поиска с минимальной априорной информацией. Также рассмотрим функцию $n(t\,|\,m)$, обратную к $M_2(t\,|\,n)$ и определяемую для задачи поиска с полной информацией как минимальное число взвешиваний, необходимое, чтобы найти среди $m$ монет все фальшивые, если известно, что таковых не более $t$.

Мы рассматриваем только так называемый *неадаптивный поиск*, т.е. когда взвешивания производятся одновременно.

Обозначим вес $j$-й монеты через $x_j$ и рассмотрим вектор $\boldsymbol{x} = (x_1, \ldots, x_m)$, который мы хотим найти с помощью взвешивания групп монет на точных весах. Взвешивание (тест) множества монет $J \subset \{0, 1, \ldots, m\}$ выдает в качестве результата суммарный вес $s(J)$ взвешиваемых монет

$$s(J) = \sum_{j \in J} x_j = (\boldsymbol{x}, \boldsymbol{h}),$$

где $\boldsymbol{h}$ – характеристический вектор множества $J$. Пусть $J_1, \ldots, J_n$ обозначают множества взвешиваемых монет, $\boldsymbol{h}^{(1)}, \ldots, \boldsymbol{h}^{(n)}$ – их характеристические векторы, и пусть $s_k := (\boldsymbol{x}, \boldsymbol{h}^{(k)})$ – результат взвешивания множества $J_k$. Двоичную $(n \times m)$-матрицу $H$, строками которой являются векторы $\boldsymbol{h}^{(1)}, \ldots, \boldsymbol{h}^{(n)}$, будем называть матрицей измерений. Тогда результаты взвешиваний можно записать в виде уравнения

$$H\boldsymbol{x}^T = \sum_{j=1}^{m} x_j \boldsymbol{h}_j = \boldsymbol{s}, \tag{5}$$

где вектор-синдром $\boldsymbol{s} = (s_1, \ldots, s_n)$ состоит из результатов взвешиваний, а $\boldsymbol{h}_j$ – $j$-й столбец матрицы $H$.

В задаче с полной информацией величина $s(J)$ дает нам точное число фальшивых монет среди взвешенных. Действительно,

$$s(J) = \beta|J \cap E| + \alpha(|J| - |J \cap E|) = \alpha|J| + (\beta - \alpha)|J \cap E|,$$

где $E$ – множество номеров фальшивых монет.

Введем модифицированные результаты взвешиваний

$$\widehat{s_k} := \frac{s_k - \alpha \operatorname{wt}(\boldsymbol{h}^{(k)})}{\beta - \alpha}$$

и, соответственно, модифицированный вектор-синдром $\widehat{\boldsymbol{s}} = (\widehat{s}_1, \ldots, \widehat{s}_n)$, где $\operatorname{wt}(\boldsymbol{h}) = |\{i : h_i \neq 0\}|$ – вес Хэмминга вектора $\boldsymbol{h}$. Тем самым, задача свелась к поиску двоичной $(n \times m)$-матрицы $H$ с максимальным $m$ при заданном $n$, для которой

уравнение

$$\widehat{\boldsymbol{s}} = H\boldsymbol{y}^T, \tag{6}$$

называемое далее синдромным, имеет не более одного *двоичного* решения $\boldsymbol{y}$ с весом Хэмминга $\mathrm{wt}(\boldsymbol{y}) \leqslant t$, где $\boldsymbol{y}$ – характеристический вектор множества $E$ номеров фальшивых монет. Это уравнение играет роль, аналогичную роли синдромного уравнения для кодов, исправляющих ошибки. Условие, что все суммы по $t$ или меньше стобцов матрицы $H$, рассматриваемых как вещественные векторы, различны, очевидно равносильно тому, что уравнение (6) имеет не более одного решения веса Хэмминга не более $t$. Напомним, что код $\mathcal{C} \subset B^n$ называется $t$-сигнатурным кодом для двоичного суммирующего канала множественного доступа, если все суммы его слов по $t$ или меньше различны и отличны от нуля (см. [8]). Тем самым, матрица $H$ является искомой тогда и только тогда, когда ее столбцы образуют $t$-сигнатурный код, и величина $M_2(t\,|\,n)$ – это максимально возможная мощность $t$-сигнатурного кода. Взаимосвязи кодов для каналов множественного доступа и комбинаторной теории поиска посвящено множество работ, начиная с [10, 11].

Задача поиска с полной информацией – классическая, она восходит к работе Эрдеша и Реньи [12], где были получены верхние и нижние границы величины $n(t\,|\,m)$ для случая, когда $t = m$. Затем в [13, 14] была найдена лучшая верхняя граница, асимптотически совпавшая с нижней границей из [12], и тем самым, было доказано, что

$$n(m\,|\,m) = \frac{m}{2\log_2 m}(1 + o(1)).$$

Этот результат был обобщен на недвоичный случай в [15].

Так как суммы по $t$ или менее векторов кода $\mathcal{C}$ мощности $m$ принадлежат множеству $\{0, 1, \ldots, t\}^n$ из $(t+1)^n$ элементов и при этом различны, то

$$\sum_{i=0}^{t} \binom{m}{i} \leqslant (t+1)^n. \tag{7}$$

Определим, аналогично (4), "скорость" роста величины $M_2(t\,|\,n)$ как

$$R_t^{(2)} := \lim_{n\to\infty} n^{-1} \log_2 M_2(t\,|\,n).$$

Из (7) следует, что

$$R_t^{(2)} \leqslant \frac{\log_2 t}{t}(1 + o(1)). \tag{8}$$

Отметим, что эту асимптотическую границу можно в два раза улучшить [10].

Теперь перейдем к задаче поиска в условиях минимальной информации. Эта задача была поставлена и во многом решена в работе [16]. Начнем с замечания из [16] о том, что неизвестный вес $\alpha$ правильной монеты можно найти взвешиванием поодиночке произвольных $2t+1$ монет, поскольку вес $\alpha$ появится среди $2t+1$ результатов взвешивания как минимум $t+1$ раз.

Зная вес $\alpha$, перепишем синдромное уравнение (5) в виде

$$\boldsymbol{s} = \sum_{j=1}^{m} x_j \boldsymbol{h}_j = \alpha \sum_{j=1}^{m} \boldsymbol{h}_j + \sum_{j\in E} (x_j - \alpha)\boldsymbol{h}_j = \alpha \sum_{j=1}^{m} \boldsymbol{h}_j + H\boldsymbol{y}^T, \tag{9}$$

где вектор $\boldsymbol{y} = (y_1, \ldots, y_m)$ состоит из новых, искусственно введенных весов $y_j := x_j - \alpha$, а $E \subset \{0, 1, \ldots, m\}$, как и выше, – множество позиций фальшивых монет.

Приведем синдромное уравнение к виду

$$\widehat{\boldsymbol{s}} = H\boldsymbol{y}^T, \tag{10}$$

где вектор

$$\widehat{\boldsymbol{s}} = \boldsymbol{s} - \alpha \sum_{j=1}^{m} \boldsymbol{h}_j$$

будем называть модифицированным синдромом. Напомним, что *вектор* $\boldsymbol{y}$ *называется* $t$*-разреженным, если мощность его носителя* $\operatorname{supp}(\boldsymbol{y}) := \{i : y_i \neq 0\}$ *не превышает* $t$. Следовательно, задача поиска фальшивых монет в предположении, что вес $\alpha$ правильной монеты известен, равносильна задаче о максимально возможном числе столбцов $m$ в *двоичной* матрице с заданным числом строк $n$, такой что уравнение (10) при любом $\widehat{\boldsymbol{s}}$ имеет не более одного $t$-разреженного решения $\boldsymbol{y} \in \mathbb{R}^m$. Это условие, в свою очередь, равносильно тому, что любые $2t$ столбцов матрицы линейно независимы.

Легко видеть, что все три задачи могут быть переформулированы как условие, что синдромное уравнение (10) имеет не более одного $t$-разреженного решения $\boldsymbol{y}$, где сами задачи различаются соответствующими дополнительными ограничениями на координаты вектора $\boldsymbol{y}$.

А именно, определения 1 или 2 приводят к ограничению, что рассматриваются только стохастические векторы-решения $\boldsymbol{y}$, т.е. такие, что все $y_i \geqslant 0$ и $\sum\limits_{i=1}^{m} y_i = 1$.

Вторая задача – поиск фальшивых монет с полной информацией – описывается тем ограничением, что вектор $\boldsymbol{y}$ двоичный.

Наконец, третья задача – при дополнительном предположении, что известен вес правильной монеты – не накладывает никаких ограничений на вектор $\boldsymbol{y}$, кроме того, что он $t$-разрежен. Эта задача, как мы только что отмечали, равносильна задаче о максимальном *двоичном* коде, в котором любые $2t$ векторов линейно независимы над полем $\mathbb{R}$.

Тем самым,

$$m^*(t\,|\,n) \leqslant M_1(t\,|\,d), \quad m^*(t\,|\,d) \leqslant M_2(t\,|\,d)$$

и

$$m^*(t\,|\,d) + 2t + 1 \leqslant M_3(t\,|\,d),$$

где $m^*(t\,|\,n)$ обозначает максимальную мощность $n$-мерного двоичного кода, в котором любые $2t$ векторов линейно независимы над $\mathbb{R}$.

Отметим, что в [16] была доказана асимптотическая нижняя граница

$$R^*(t) := \limsup_{n\to\infty} n^{-1} \log_2 m^*(t\,|\,n) = \Omega(t^{-1}\log t),$$

которая тем самым справедлива для скоростей оптимальных кодов во всех рассматриваемых задачах.

С другой стороны, если во второй задаче предположить, что число фальшивых монет заранее известно и равно $t$, то это требование оказывается самым слабым из всех перечисленных выше задач. Поэтому для всех трех задач справедливо неравенство

$$\binom{m}{t} \leqslant (t+1)^n$$

(ср. неравенство (7)), и следовательно, для них справедлива асимптотическая верхняя оценка (8) на скорость соответствующих кодов.

Таким образом, мы установили следующее:

*Для всех трех задач порядок скорости наилучшего кода равен* $t^{-1}\log t$.

## § 2. Ошибки в измерениях и каналах

Перейдем теперь к рассмотрению случая, когда имеются ошибки, например, когда в задаче поиска появляются ошибки во взвешиваниях, или когда ошибки возникают на выходе двоичного суммирующего канала. Как мы увидим ниже, наличие ошибок вносит существенные различия в описанные выше три задачи. Так, например, определения 1 и 2 перестают быть эквивалентными.

Расширим определение 1 на случай ошибок. Очевидно, что для двоичных векторов евклидово расстояние и расстояние Хэмминга связаны соотношением $D^2(\boldsymbol{a},\boldsymbol{b}) = d(\boldsymbol{a},\boldsymbol{b})$ для любых $\boldsymbol{a},\boldsymbol{b}\in B^n$.

Определение 3. Множество $\mathcal{C}$ вершин $n$-мерного булева куба $B^n$ называется $(t,\delta)$-независимым кодом, если для любых его двух непересекающихся подмножеств $A,B\subset\mathcal{C}$, таких что $|A|,|B|\leqslant t$, их выпуклые оболочки находятся на евклидовом расстоянии не меньше чем $\delta$ друг от друга, т.е.

$$D(\langle A\rangle,\langle B\rangle) := \min_{\substack{\boldsymbol{a}\in\langle A\rangle\\ \boldsymbol{b}\in\langle B\rangle}} D(\boldsymbol{a},\boldsymbol{b}) \geqslant \delta. \tag{11}$$

Мы отмечали в § 1, что определение 2 является математической формулировкой задачи о мультимедийных кодах отпечатков пальцев, устойчивых к линейным атакам коалиций, где под линейной атакой коалиции $A\subset\mathcal{C}$ понимается создание ложного вектора $\widehat{\boldsymbol{a}} = \sum\limits_{\boldsymbol{a}\in A} p_{\boldsymbol{a}}\boldsymbol{a}$, где все $p_{\boldsymbol{a}}\geqslant 0$ и $\sum\limits_{\boldsymbol{a}\in A} p_{\boldsymbol{a}} = 1$, и нужно уметь находить $A$ по $\widehat{\boldsymbol{a}}$. Следовательно, использование $(t,\delta)$-независимого кода гарантирует, что если происходящие ошибки имеют евклидову длину не более $\delta/2$, то любые две коалиции, способные породить данный ложный вектор, имеют непустое попарное пересечение. Это свойство аналогично так называемому свойству secure frameproof, введенному ранее для обычных кодов цифровых отпечатков пальцев. Однако оно не только не позволяет найти всю коалицию, но даже не гарантирует нахождение хотя бы одного участника коалиции, так как пересечение всех коалиций, способных породить данный вектор, может быть пусто (см. [3]).

Отметим, что произвольное $t$-независимое множество $\mathcal{C}$ является одновременно и $(t,\delta)$-независимым кодом. Действительно, достаточно положить

$$\delta = d_t(\mathcal{C}) := \min_{\substack{A,B\subset\mathcal{C}\\ A\cap B=\varnothing\\ |A|=|B|=t}} D(\langle A\rangle,\langle B\rangle). \tag{12}$$

Естественно рассмотреть величину $M_1(t,\delta\,|\,n)$, равную максимальной мощности $(t,\delta)$-независимого кода в $n$-мерном булевом кубе. Нам неизвестно, как ведет себя величина $\delta(t)$, такая что $M_1(t,\delta\,|\,n)$ растет экспоненциально от $n$ при $\delta<\delta(t)$ и фиксированном $t$. Более того, нам это неизвестно даже для случая $t=2$.

Заметим, что если взять в качестве кода $\mathcal{C}$ двоичный код длины $n$ с минимальным расстоянием Хэмминга $d=\tau n$, где $0<\tau<1/2$, и мощности $M = 2^{(R(\tau)+o(1))n}$ с $R(\tau)>0$, то все вершины будут на попарном евклидовом расстоянии не менее чем $\sqrt{\tau n}$, однако это не гарантирует даже свойство $t$-независимости (т.е. выпуклые оболочки могут пересекаться).

Согласно определению 2 множество $\mathcal{C} \subset B^n$ называется $t$-независимым, если для любых двух различных подмножеств $A, B \subset \mathcal{C}$ мощности не более $t$ каждое их выпуклые оболочки не пересекаются по строго внутренней точке. Очевидно, что сколь угодно малая ошибка может нарушить это свойство.

Рассмотрим для примера $t = 2$, $A = \{\boldsymbol{a}, \boldsymbol{c}\}$, $B = \{\boldsymbol{b}, \boldsymbol{c}\}$ и две внутренние точки:

$$\boldsymbol{x_a} = \varepsilon\boldsymbol{a} + (1-\varepsilon)\boldsymbol{c} \in A \quad \text{и} \quad \boldsymbol{x_b} = \varepsilon\boldsymbol{b} + (1-\varepsilon)\boldsymbol{c} \in B,$$

находящиеся на расстоянии $\varepsilon\|\boldsymbol{a} - \boldsymbol{b}\|_2$ друг от друга. Иначе говоря, малые ошибки $-\varepsilon\boldsymbol{a}$ и $-\varepsilon\boldsymbol{b}$, соответственно, переведут эти точки в точку $(1-\varepsilon)\boldsymbol{c}$.

Это замечание может быть интерпретировано как то, что мультимедийные коды отпечатков пальцев не способны полностью найти коалицию недобросовестных пользователей в условиях общей линейной атаки и сколь угодно малого целенаправленного шума (см. [17]). Тем не менее, как было показано в [18], такие коды существуют, если ограничиться атакой усреднения, т.е. когда ложный вектор $\widehat{\boldsymbol{a}} = |A|^{-1} \sum\limits_{\boldsymbol{a}\in A} \boldsymbol{a}$. Иначе говоря, мы ослабляем условие на расстояние между выпуклыми оболочками различных $t$-подмножеств, а именно требуем только, чтобы центры масс выпуклых оболочек были друг от друга на расстоянии не меньше заданного порога, равного удвоенной длине ошибки. Если мощность коалиции известна априори, то возникающая задача превращается в задачу 2 поиска фальшивых монет при полной информации и ошибках измерений, или, что равносильно, задачу о сигнатурных кодах для двоичного суммирующего канала, исправляющих ошибки на выходе канала.

Код $\mathcal{C} \subset B^n$ будем называть $(t, \delta)$-сигнатурным кодом для двоичного суммирующего канала, если любые две суммы его слов по $t$ или меньше не просто различны, а находятся на евклидовом расстоянии не менее $\delta$ друг от друга. Таким образом, $(t, \delta)$-сигнатурный код способен правильно находить группу из не более чем $t$ активных пользователей в условиях, когда выход двоичного суммирующего канала может быть искажен ошибкой длины (евклидовой) меньше $\delta/2$.

Напомним конструкцию $(t, \delta)$-сигнатурных кодов из [18]. Рассмотрим двоичный $n$-мерный код $\mathcal{H}$ мощности $m$, который состоит из столбцов проверочной $(n \times m)$-матрицы двоичного примитивного кода БЧХ длины $m = 2^{n/t} - 1$ или неприводимого кода Гоппы длины $m = 2^{n/t}$, исправляющих $t$ ошибок, где $n$ – число проверочных символов данных кодов. Кроме того, пусть $V$ – линейный двоичный код длины $N$ с $n$ информационными символами и минимальным кодовым расстоянием $d(V)$, а $\varphi\colon B^n \to V \subset B^N$ – произвольное систематическое кодирование этого кода.

В [18] было показано, что двоичный код

$$\mathcal{H}' = \varphi(\mathcal{H}) = \{\varphi(h) :\ h \in \mathcal{H}\} \subset V$$

длины $N$ является $(t, \delta)$-сигнатурным кодом с $\delta = \sqrt{d(V)}$. Если взять в качестве $V$ коды мощности $2^{RN}$ с линейно растущим по $N$ расстоянием $d(V) = \tau N$, то получатся $(t, \delta)$-сигнатурные коды с расстоянием $\delta(V)$, растущим как $c\sqrt{N}$. Так как рассматриваемые коды двоичные, то евклидова длина кодовых векторов имеет порядок $\Omega(\sqrt{N})$, и следовательно, минимальное расстояние этих кодов имеет оптимальный порядок $\Omega(\sqrt{N})$. Важно отметить, что данные коды можно строить со сложностью, полиномиальной от $N$, т.е., с *полилогарифмической сложностью по* $m$. Отметим, что случайное кодирование позволяет увеличить скорость в $\log t$ раз и, тем самым, получить коды с оптимальной по порядку скоростью (см. [19]).

Также отметим, что схожая постановка задачи была исследована в [20], где рассматривалась вероятностная модель ошибок (с нормальным распределением).

Перейдем теперь к задаче 3 и рассмотрим наиболее общий вариант постановки задачи, когда, несмотря на ошибки в измерениях, требуется найти не только фальшивые монеты, но и веса всех монет. Заметим, что нахождение веса правильной монеты уже не так очевидно, как в случае, когда ошибок нет, поэтому мы будем рассматривать упрощенный вариант, когда вес правильной монеты известен и равен 0. Для $(n \times m)$-матрицы $H$ измерений и неизвестного $t$-разреженного вектора $\boldsymbol{y} \in \mathbb{R}^m$ соответствующее синдромное уравнение примет вид

$$\widehat{\boldsymbol{s}} = H\boldsymbol{y}^T + \boldsymbol{e}, \tag{13}$$

где $\boldsymbol{e} \in \mathbb{R}^n$ – вектор ошибки длины $\|\boldsymbol{e}\|_2 \leqslant \varepsilon$.

Задача нахождения разреженного решения уравнения (13) стала популярной после основополагающих работ [21, 22], в которых было предложено искать аппроксимацию такого решения уравнения (13) заменой минимизации веса Хэмминга решения на минимизацию его $\ell_1$-нормы. А именно было предложено рассмотреть задачу минимизации $\|\boldsymbol{y}\|_1$ при ограничении $\|\widehat{\boldsymbol{s}} - H\boldsymbol{y}^T\|_2 \leqslant \varepsilon$. Предложенный подход получил название сжатого измерения (compressed sensing). В качестве матриц измерений было предложено использовать матрицы со свойством ограниченной изометрии (restricted isometry property), или, сокращенно, RIP-матрицы.

Матрица $H$ называется $\gamma_t$-RIP-матрицей, если для любого $t$-разреженного вектора $\boldsymbol{y} \in \mathbb{R}^n$ справедливо

$$(1 - \gamma_t)\|\boldsymbol{y}\|_2^2 \leqslant \|H\boldsymbol{y}^T\|_2^2 \leqslant (1 + \gamma_t)\|\boldsymbol{y}\|_2^2. \tag{14}$$

Обозначим через $\boldsymbol{y}^*$ вектор минимальной $\ell_1$-нормы в множестве

$$\big\{\boldsymbol{z} \in \mathbb{R}^n : \ \|\widehat{\boldsymbol{s}} - H\boldsymbol{z}^T\|_2 \leqslant \varepsilon\big\}.$$

Основной результат теории сжатых измерений можно неформально сформулировать следующим образом: если параметр $\gamma_T$, где $T$ в несколько раз больше $t$, достаточно мал, то $\|\boldsymbol{y}^* - \boldsymbol{y}\|_2 \leqslant C\varepsilon$. Приведем в качестве примера точной формулировки теорему 1 из [23]:

"Пусть для матрицы $H$ выполнено $\gamma_{3t} + 3\gamma_{4t} < 2$. Тогда $\|\boldsymbol{y}^* - \boldsymbol{y}\|_2 \leqslant C_t\varepsilon$, где константа $C_t$ зависит только от $\gamma_{4t}$. Например, $C_t \approx 8{,}82$ для $\gamma_{4t} = 1/5$."

В литературе, посвященной теории сжатых измерений, неоднократно указывалось, что для нахождения хорошей аппроксимации решения уравнения (13) достаточно найти носитель неизвестного вектора $\boldsymbol{y}$, а затем применить, например, метод наименьших квадратов. С другой стороны, нахождение хорошего приближения $\boldsymbol{y}^*$ еще не гарантирует, что носители векторов $\boldsymbol{y}^*$ и $\boldsymbol{y}$ совпадают, если не наложить то ограничение, что

$$y_{\min} = \min_{i \in \operatorname{supp}(y)} |y_i| \geqslant \Omega(\varepsilon).$$

В противном случае, как мы уже указывали, точное нахождение носителя невозможно. Тем не менее, если найдено хорошее приближение $\boldsymbol{y}^*$ к искомому вектору $\boldsymbol{y}$, а именно если $\|\boldsymbol{y}^* - \boldsymbol{y}\|_2 \leqslant \widehat{\varepsilon}$ и одновременно $y_{\min} > 2\widehat{\varepsilon}$, то $\operatorname{supp}(\boldsymbol{y}) = \{i : \ |y_i^*| > \widehat{\varepsilon}\}$.

В ряде работ (см. [24,25] и библиографию в них) был рассмотрен прямой подход к восстановлению носителя вектора с помощью RIP-матриц, что накладывало условия на величину $\gamma_T$ для сравнительно малых $T$. Так, в [24] был предложен алгоритм, однозначно находящий носитель вектора $\boldsymbol{y}$, если выполнены условия

$$\gamma_{t+1}\sqrt{t+1} < 1 \quad \text{и} \quad y_{\min} > 2\varepsilon(1 - \gamma_{t+1}\sqrt{t+1})^{-1}.$$

Хорошо известно, что для RIP-матриц минимальное $n$ имеет вид

$$n = \Omega\Big(t \log \frac{m}{t}\Big),$$

что при фиксированном $t$ и растущем $m$ дает $n = O(t \log m)$. Однако явные конструкции таких матриц неизвестны, тогда как $(t, \delta)$-сигнатурные коды из [18] строятся явно и со сложностью $\mathrm{polylog}(m)$ (см. выше), и при этом имеют ту же асимптотику $n$, что и случайные RIP-матрицы. Недостаток кодов из [18] состоит в том, что они позволяют находить носитель только у таких $t$-разреженных векторов, что все их ненулевые координаты равны 1. Сейчас мы частично устраним этот недостаток.

Вектор $\boldsymbol{y}$ будем называть $\theta$-равномерным в норме $\ell_1$, если

$$\|\boldsymbol{y} - \mathbf{1}_E\|_1 = \sum_{i \in E} |y_i - 1| \leqslant \theta,$$

где $E$ – носитель вектора $\boldsymbol{y}$, а $\mathbf{1}_E$ – характеристический вектор множества $E$.

Предложение 2. *Любой $(t, \delta)$-сигнатурный код $\mathcal{H}$ позволяет находить носитель произвольного $\theta$-равномерного вектора, если $\delta > 2(\theta h + \varepsilon)$, где $h$ – максимальная евклидова длина векторов из $\mathcal{H}$.*

Доказательство. Пусть код $\mathcal{H} \subset B^n$ является $(t, \delta)$-сигнатурным кодом, т.е. для любых двух различных подмножеств $A, B \subset \mathcal{H}$, таких что $|A| \leqslant t$, $|B| \leqslant t$, справедливо

$$\left\| \sum_{\boldsymbol{a} \in A} \boldsymbol{a} - \sum_{\boldsymbol{b} \in B} \boldsymbol{b} \right\|_2 \geqslant \delta.$$

Пусть $\boldsymbol{y}, \boldsymbol{y}'$ – два $\theta$-равномерных $t$-разреженных вектора с различными носителями $A$ и $B$ соответственно, для которых $\|\widehat{\boldsymbol{s}} - H\boldsymbol{y}^T\|_2 \leqslant \varepsilon$ и $\|\widehat{\boldsymbol{s}} - H\boldsymbol{y}'^T\|_2 \leqslant \varepsilon$. Тогда

$$\Delta := \|H\boldsymbol{y}^T - H\boldsymbol{y}'^T\|_2 \leqslant 2\varepsilon.$$

Обозначим $\mu_{\boldsymbol{a}} = 1 - y_{\boldsymbol{a}}$, $\mu_{\boldsymbol{b}} = 1 - y_{\boldsymbol{b}}$ для $\boldsymbol{a} \in A$, $\boldsymbol{b} \in B$ соответственно. Тогда, с другой стороны,

$$\Delta = \left\| \left( \sum_{\boldsymbol{a} \in A} \boldsymbol{a} - \sum_{\boldsymbol{b} \in B} \boldsymbol{b} \right) - \sum_{\boldsymbol{a} \in A} \mu_{\boldsymbol{a}} \boldsymbol{h}_{\boldsymbol{a}} + \sum_{\boldsymbol{b} \in B} \mu_{\boldsymbol{b}} \boldsymbol{h}_{\boldsymbol{b}} \right\|_2 \geqslant \delta - 2\theta h > 2\varepsilon.$$

Полученное противоречие и доказывает утверждение. ▲

Мы *предполагаем*, что это утверждение можно распространить на произвольные $t$-разреженные векторы, или, по крайней мере, на стохастические разреженные векторы. Если эта гипотеза справедлива, то из [19, теорема 1] будет следовать существование матриц измерений с $n = \Omega\Big(t \dfrac{\log m}{\log t}\Big)$, позволяющих находить носитель $t$-разреженных векторов, что и будет асимптотическим ответом для задачи поиска носителя $t$-разреженного вектора для случая, когда $t$ фиксировано, а размерность $m$ вектора стремится к бесконечности.

## СПИСОК ЛИТЕРАТУРЫ


1. *Cheng M., Miao Y.* On Anti-Collusion Codes and Detection Algorithms for Multimedia Fingerprinting // IEEE Trans. Inform. Theory. 2011. V. 57. № 7. P. 4843–4851. `https://doi.org/10.1109/TIT.2011.2146130`

2. *Егорова Е.Е., Кабатянский Г.А.* Разделимые коды для защиты мультимедиа от нелегального копирования коалициями // Пробл. передачи информ. 2021. Т. 57. № 2. С. 90–111. https://doi.org/10.31857/S0555292321020066
3. *Boneh D., Shaw J.* Collusion-Secure Fingerprinting for Digital Data // IEEE Trans. Inform. Theory. 1998. V. 44. № 5. P. 1897–1905. https://doi.org/10.1109/18.705568
4. *Barg A., Blakley G.R., Kabatiansky G.A.* Digital Fingerprinting Codes: Problem Statements, Constructions, Identification of Traitors // IEEE Trans. Inform. Theory. 2003. V. 49. № 4. P. 852–865. https://doi.org/10.1109/TIT.2003.809570
5. *Erdős P., Turán P.* On a Problem of Sidon in Additive Number Theory, and on Some Related Problems // J. London Math. Soc. 1941. V. 16. № 4. P. 212–215. https://doi.org/10.1112/jlms/s1-16.4.212
6. *Babai L., Sós V.T.* Sidon Sets in Groups and Induced Subgraphs of Cayley Graphs // European J. Combin. 1985. V. 6. № 2. P. 101–114. https://doi.org/10.1016/S0195-6698(85)80001-9
7. *Cohen G., Litsyn S., Zémor G.* Binary $B_2$-Sequences: A New Upper Bound // J. Combin. Theory Ser. A. 2001. V. 94. № 1. P. 152–155. https://doi.org/10.1006/jcta.2000.3127
8. *Györfi L., Győri S., Laczay B., Ruszinkó M.* Lectures on Multiple Access Channels. Book draft, 2005. Available at http://www.szit.bme.hu/~gyori/AFOSR_05/book.pdf.
9. *Кабатянский Г.А. Лебедев В.С.* О метрической размерности недвоичных пространств Хэмминга // Пробл. передачи информ. 2018. Т. 54. № 1. С. 54–62. http://mi.mathnet.ru/ppi2259
10. *Дьячков А.Г., Рыков В.В.* Об одной модели кодирования для суммирующего канала с множественным доступом // Пробл. передачи информ. 1981. Т. 17. № 2. С. 26–38. http://mi.mathnet.ru/ppi1390
11. *Wolf J.K.* Born Again Group Testing: Multiaccess Communications // IEEE Trans. Inform. Theory. 1985. V. 31. № 2. P. 185–191. https://doi.org/10.1109/TIT.1985.1057026
12. *Erdős P., Rényi A.* On Two Problems of Information Theory // Magyar Tud. Akad. Mat. Kutató Int. Közl. 1963. V. 8. № 1–2. P. 229–243. Available at http://static.renyi.hu/renyi_cikkek/1963_on_two_problems_of_information_theory.pdf
13. *Lindström B.* On a Combinatory Detection Problem. I // Magyar Tud. Akad. Mat. Kutató Int. Közl. 1964. V. 9. № 1–2. P. 195–207.
14. *Cantor D.G., Mills W.H.* Determination of a Subset from Certain Combinatorial Properties // Canad. J. Math. 1966. V. 18. P. 42–48. https://doi.org/10.4153/CJM-1966-007-2
15. *Jiang Z., Polyanskii N.* On the Metric Dimension of Cartesian Powers of a Graph // J. Combin. Theory Ser. A. 2019. V. 165. P. 1–14. https://doi.org/10.1016/j.jcta.2019.01.002
16. *Bshouty N.H., Mazzawi H.* On Parity Check $(0,1)$-Matrix over $\mathbb{Z}_p$ // Proc. 22nd Annu. ACM–SIAM Symp. on Discrete Algorithms (SODA'11). San Francisco, CA. Jan. 23–25, 2011. P. 1383–1394. https://dl.acm.org/doi/10.5555/2133036.2133142
17. *Fan J., Gu Y., Hachimori M., Miao Y.* Signature Codes for Weighted Binary Adder Channel and Multimedia Fingerprinting // IEEE Trans. Inform. Theory. 2021. V. 67. № 1. P. 200–216. https://doi.org/10.1109/TIT.2020.3033445
18. *Егорова Е.Е., Фернандес М., Кабатянский Г.А., Мяо И.* Существование и конструкции мультимедийных кодов, способных находить полную коалицию при атаке усреднения и шуме // Пробл. передачи информ. 2020. Т. 56. № 4. С. 97–108. https://doi.org/10.31857/S0555292320040087
19. *Vorobyev I.* Complete Traceability Multimedia Fingerprinting Codes Resistant to Averaging Attack and Adversarial Noise with Optimal Rate // Des. Codes Cryptogr. 2022. Open Access Article. https://doi.org/10.1007/s10623-022-01144-x
20. *Gkagkos M., Pradhan A.K., Amalladinne V., Narayanan K., Chamberland J-F., Georghiades C.N.* Approximate Support Recovery Using Codes for Unsourced Multiple Access // Proc. 2021 IEEE Int. Symp. on Information Theory (ISIT'2021). Melbourne, Australia. July 12–20, 2021. P. 2948–2953. https://doi.org/10.1109/ISIT45174.2021.9517995
21. *Donoho D.L.* Compressed Sensing // IEEE Trans. Inform. Theory. 2006. V. 52. № 4. P. 1289–1306. https://doi.org/10.1109/TIT.2006.871582

22. *Candès E.J., Tao T.* Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? // IEEE Trans. Inform. Theory. 2006. V. 52. № 12. P. 5406–5425. https://doi.org/10.1109/TIT.2006.885507
23. *Candès E.J., Romberg J.K., Tao T.* Stable Signal Recovery from Incomplete and Inaccurate Measurements // Comm. Pure Appl. Math. 2006. V. 59. № 8. P. 1207–1223. https://doi.org/10.1002/cpa.20124
24. *Wen J., Zhou Z., Wang J., Tang X., Mo Q.* A Sharp Condition for Exact Support Recovery with Orthogonal Matching Pursuit // IEEE Trans. Signal Process. 2017. V. 65. № 6. P. 1370–1382. https://doi.org/10.1109/TSP.2016.2634550
25. *Mehrabi M., Tchamkerten A.* Error-Correction for Sparse Support Recovery Algorithms // Proc. 2021 IEEE Int. Symp. on Information Theory (ISIT'2021). Melbourne, Australia. July 12–20, 2021. P. 1754–1759. https://doi.org/10.1109/ISIT45174.2021.9518027

*Джанабекова Алия*
Московский физико-технический институт
(государственный университет),
факультет управления и прикладной математики,
кафедра проблем передачи информации и анализа данных
dzhanabekova@phystech.edu
*Кабатянский Григорий Анатольевич*
Сколковский институт науки и технологий (Сколтех)
g.kabatyansky@skoltech.ru
*Камель Ибрахим* (Kamel, Ibrahim)
*Рабие Тамер Фарук* (Rabie, Tamer Farouk)
Университет Шарджи, Шарджа, ОАЭ
Kamel@sharjah.ac.ae
trabie@sharjah.ac.ae

Поступила в редакцию
01.11.2022
После доработки
22.11.2022
Принята к публикации
23.11.2022