

УДК 621.391 : 519.72

© 2023 г. М. Фернандес¹, Г.А. Кабатянский², С.А. Круглик³, И. Мяо⁴**КОДЫ ДЛЯ ТОЧНОГО НАХОЖДЕНИЯ НОСИТЕЛЯ РАЗРЕЖЕННОГО
ВЕКТОРА ПО ОШИБОЧНЫМ ЛИНЕЙНЫМ ИЗМЕРЕНИЯМ
И ИХ ДЕКОДИРОВАНИЕ**

Построены коды, позволяющие точно находить носитель неизвестного разреженного вектора, у которого модули всех ненулевых координат примерно равны, по результатам линейных измерений в присутствии шума с ограниченной сверху ℓ_p -нормой. Предложен алгоритм декодирования, имеющий асимптотически минимальную сложность.

Ключевые слова: сжатие измерений, носитель разреженного вектора, групповое тестирование, поиск фальшивых монет, сигнатурные коды для суммирующего канала с множественным доступом и шумом, мультимедийные коды цифровых отпечатков пальцев.

DOI: 10.31857/S0555292323010023, **EDN:** JDHVDR**§ 1. Введение**

В данной статье мы продолжаем (см. [1–6]) исследование задачи нахождения носителя неизвестного разреженного вектора $\mathbf{x} \in \mathbb{R}^n$ с помощью m линейных измерений, подверженных влиянию шума. В литературе, посвященной теории сжатых измерений, неоднократно указывалось, что для нахождения хорошей аппроксимации неизвестного K -разреженного вектора $\mathbf{x}$ достаточно найти его носитель $\text{supp}(\mathbf{x}) := \{i : x_i \neq 0\}$, а затем применить, например, метод наименьших квадратов. Напомним, что вектор $\mathbf{x}$ называется K -разреженным, если $|\text{supp}(\mathbf{x})| \leq K$.

С другой стороны, нахождение хорошего приближения $\hat{\mathbf{x}}$ к вектору $\mathbf{x}$ не гарантирует, что носители векторов $\mathbf{x}$ и $\hat{\mathbf{x}}$ совпадают, поэтому задача нахождения носителя разреженного вектора по линейным измерениям представляет и самостоятельный интерес. Кроме того, как неоднократно отмечалось, эта задача равносильна нескольким другим известным задачам. Среди них задача построения сигнатурных кодов для суммирующего канала множественного доступа, задача поиска фальшивых монет на точных весах, задача построения мультимедийных кодов, устойчивых к атакам коалиций (подробнее см. обзор [4]). Наконец, самая новая постановка задачи, весьма близкая к задаче нахождения носителя неизвестного вектора, – это

¹ Исследование выполнено при финансовой поддержке гранта правительства Испании TCO-RISEBLOCK (номер гранта PID2019-110224RB-I00, проект MINECO).

² Исследование выполнено при финансовой поддержке Российского научного фонда в рамках гранта РНФ 22-41-02028.

³ Исследование выполнено при финансовой поддержке Министерства образования Сингапура (Фонд академических исследований 2-го уровня, номера грантов MOE2019-T2-2-083 и MOE-T2EP20121-0007).

⁴ Исследование выполнено при финансовой поддержке совместной японско-российской научной программы Японского общества содействия развитию науки (JSPS) и Российского фонда фундаментальных исследований в рамках проекта JPJSBP 120204802.

коды для передачи информации по каналу множественного доступа, когда нужно правильно восстановить только переданную информацию, но не отправителей информации [7].

Напомним кратко постановку задачи сжатых измерений [8, 9]. Требуется найти неизвестный K -разреженный вектор $\mathbf{x} \in \mathbb{R}^n$ по m линейным измерениям, искаженным вектором шума $\mathbf{w} \in \mathbb{R}^m$, что формально можно описать как нахождение K -разреженного решения $\mathbf{x}$ следующего уравнения:

$$\mathbf{s} = H\mathbf{x}^T + \mathbf{w}, \quad (1)$$

где H – $(m \times n)$ -матрица линейных измерений. Нас же будет интересовать только носитель вектора-решения. Этой задаче также посвящено много публикаций (см. работы [10–12] и библиографию в них). В части работ рассматривается вероятностная модель шума (см. [10]), но тогда точное нахождение носителя невозможно. В данной статье мы исследуем задачу точного нахождения носителя неизвестного вектора в предположении, что величина шума $\mathbf{w}$ ограничена сверху некоторой известной величиной t_p в метрике ℓ_p .

Сформулируем основные результаты статьи. Явно строится семейство кодов, позволяющих точно находить носитель K -разреженного n -мерного вектора $\mathbf{x}$, у которого модули всех ненулевых координат примерно равны, с помощью m линейных измерений, т.е. находить $\text{supp}(\mathbf{x})$ из уравнения (1), где $m = O(K \log n)$ при фиксированном K и $n \rightarrow \infty$, а шум имеет порядок $t_p = O(m^{1/p})$. Для этого семейства кодов предлагается алгоритм декодирования с полиномиальной по Km сложностью, позволяющий реализовать половину соответствующей корректирующей способности. Параметры построенных кодов и их алгоритмов декодирования асимптотически оптимальны.

§ 2. Коды, позволяющие точно находить носитель разреженного “почти” двоичного вектора

Начнем со случая, когда априори известно, что все ненулевые координаты вектора $\mathbf{x}$ равны (это так, например, для модели 1 из [10]). Будем для простоты предполагать, что вектор $\mathbf{x}$ двоичный, т.е. все $x_i \in \{0, 1\}$. Условие, что матрица H позволяет точно найти носитель двоичного вектора $\mathbf{x}$ (т.е. и сам вектор) из уравнения (1) при дополнительном ограничении, что $\|\mathbf{w}\|_p < t_p$, равносильно тому, что для любых двух различных подмножеств $A, B \subset [n]$, таких что $|A|, |B| \leq K$, справедливо неравенство

$$\left\| \sum_{j \in A} \mathbf{h}_j - \sum_{j \in B} \mathbf{h}_j \right\|_p \geq 2t_p. \quad (2)$$

Далее вместо матрицы H мы будем рассматривать код $\mathcal{H}$, состоящий из столбцов этой матрицы.

Определение 1. Двоичный код $\mathcal{H} = \{\mathbf{h}_1, \dots, \mathbf{h}_n\} \subset \{0, 1\}^m$ называется $(K, t_p)_2$ -кодом, если для любых двух различных кодовых подмножеств $F, G \subset \mathcal{H}$, таких что $|F|, |G| \leq K$, выполнено

$$\left\| \sum_{\mathbf{h} \in F} \mathbf{h} - \sum_{\mathbf{h} \in G} \mathbf{h} \right\|_p \geq 2t_p. \quad (3)$$

Отметим, что это определение уже появлялось ранее в [2] в рамках задачи о построении мультимедийного кода, устойчивого к атаке усреднения коалициями мощности не более K и к ошибкам ограниченной евклидовой длины. На самом деле,

задача о мультимедийных кодах, устойчивых к общей линейной атаке K -коалиций, почти эквивалентна задаче о нахождении носителя у *произвольного* K -разреженного вектора, а задача о мультимедийных кодах, устойчивых к атаке усреднения, близка к рассматриваемой задаче о нахождении *двоичного* вектора. Отметим только, что в постановке задачи о мультимедийных кодах накладывается дополнительное ограничение, что координаты вектора $\mathbf{x}$ – это вероятности, т.е. все $x_i \geq 0$ и $\sum_i x_i = 1$ (подробнее о сравнении этих задач см. [2, 4]). Более того, приводимая ниже конструкция $(K, t_p)_2$ -кодов – это по существу конструкция из [2], у которой расширена область применения (не только двоичные векторы) и, главное, предложен алгоритм декодирования. Мы приведем эту конструкцию для полноты изложения, а также потому, что она осталась незамеченной в теории сжатых измерений.

Рассмотрим два двоичных линейных кода C и V длины n и m соответственно, при этом коды не только исправляют K и T ошибок соответственно, но и известны эффективные алгоритмы их исправления. Пусть H' – какая-то проверочная $(r \times n)$ -матрица кода C , позволяющая эффективно исправить K ошибок, т.е. найти единственное двоичное решение $\mathbf{z}$ веса Хэмминга $\text{wt}(\mathbf{z}) \leq K$ соответствующего синдромного уравнения $H'\mathbf{z}^T = \mathbf{s}$. Код C будем называть *внутренним* кодом (как мы сейчас увидим, у предлагаемой конструкции есть определенное сходство с каскадной конструкцией).

Код V , называемый *внешним*, – это двоичный линейный код длины m с r информационными символами, и пусть $\theta: \mathbb{F}_2^r \rightarrow V$ – его систематическое кодирование. Мы предполагаем, что код V исправляет T ошибок с помощью алгоритма декодирования $\psi: \mathbb{F}_2^m \rightarrow V$.

Код $\mathcal{H}(C, V) = \{\mathbf{h}_1, \dots, \mathbf{h}_n\}$ строится следующим образом. Все n столбцов матрицы H' кодируются с помощью отображения θ , превращаясь в n двоичных векторов $\mathbf{h}_1, \dots, \mathbf{h}_n$ длины m , где $\mathbf{h}_j = \theta(\mathbf{h}'_j)$.

Лемма 1. Код $\mathcal{H}(C, V)$ является $(K, t_p)_2$ -кодом с $2t_p = (2T + 1)^{1/p}$.

Доказательство. Рассмотрим два различных подмножества $A, B \subset [n]$, таких что $|A|, |B| \leq K$, соответствующие им векторы

$$\mathbf{h}_A := \sum_{j \in A} \mathbf{h}_j = \sum_{j \in A} \theta(\mathbf{h}'_j), \quad \mathbf{h}_B := \sum_{j \in B} \mathbf{h}_j = \sum_{j \in B} \theta(\mathbf{h}'_j)$$

и эти же векторы, взятые по модулю 2, а именно $\mathbf{h}_A \bmod 2$ и $\mathbf{h}_B \bmod 2$. Тогда в силу линейности отображения кодирования θ имеем

$$\mathbf{h}_A \bmod 2 = \sum_{j \in A} \theta(\mathbf{h}'_j) \bmod 2 = \theta\left(\sum_{j \in A} \oplus \mathbf{h}'_j\right).$$

Аналогично

$$\mathbf{h}_B \bmod 2 = \theta\left(\sum_{j \in B} \oplus \mathbf{h}'_j\right).$$

Двоичные векторы $\sum_{j \in A} \oplus \mathbf{h}'_j$ и $\sum_{j \in B} \oplus \mathbf{h}'_j$ различны, так как это две различные суммы из K или менее столбцов проверочной матрицы кода, исправляющего K ошибок. Поэтому $\mathbf{h}_A \bmod 2$ и $\mathbf{h}_B \bmod 2$ – это два различных слова кода V , и тем самым, они различаются как минимум в $2T + 1$ координатах. Тем более, целочисленные векторы $\mathbf{h}_A$ и $\mathbf{h}_B$ различаются как минимум в $2T + 1$ координатах, и следовательно,

$$\|\mathbf{h}_A - \mathbf{h}_B\|_p^p \geq 2T + 1. \quad \blacktriangle$$

Вернемся к общей задаче нахождения носителя *произвольного* K -разреженного вектора $\mathbf{x}$ в присутствии шума $\mathbf{w}$ ограниченной ℓ_p -длины. Заметим, что если не наложить дополнительного ограничения, что значения ненулевых координат вектора $\mathbf{x}$ не могут быть слишком малы, то точное нахождение $\text{supp}(\mathbf{x})$ невозможно (см. [3]). Действительно, рассмотрим в качестве примера простейший случай, когда $K = 2$, и два вектора $\mathbf{x} = (1, \varepsilon, 0, \dots, 0)$ и $\mathbf{x}' = (1, 0, \varepsilon, 0, \dots, 0)$. Тогда вектор ошибки $\mathbf{w} = (0, -\varepsilon, \varepsilon, 0, \dots, 0)$ переводит вектор $\mathbf{x}$ в вектор $\mathbf{x}'$, но при этом длина вектора ошибки мала.

Теорема 1. *Если неизвестный K -разреженный вектор $\mathbf{x}$ “почти” двоичный, т.е. все его ненулевые координаты лежат в диапазоне $[1 - \tau, 1 + \tau]$, $0 < \tau$, то код $\mathcal{H}(C, V)$ позволяет точно найти носитель вектора, если длина шума $\mathbf{w}$ в метрике ℓ_p меньше $\Delta/2$, где $\Delta = (2T + 1)^{1/p} - 2K\tau m^{1/p}$.*

Доказательство. Нужно доказать, что для любых двух “почти” двоичных K -разреженных векторов $\mathbf{x}, \mathbf{y}$ с различными носителями $A, B \subset [n]$, такими что $|A|, |B| \leq K$, справедливо неравенство

$$\left\| \sum_{j \in A} x_j \mathbf{h}_j - \sum_{j \in B} y_j \mathbf{h}_j \right\|_p \geq \Delta. \quad (4)$$

Представим $x_j = 1 + \alpha_j$, $y_j = 1 + \beta_j$, где $|\alpha_j|, |\beta_j| \leq \tau$. Тогда

$$\begin{aligned} \left\| \sum_{j \in A} x_j \mathbf{h}_j - \sum_{j \in B} y_j \mathbf{h}_j \right\|_p &\geq \left\| \sum_{j \in A} \mathbf{h}_j - \sum_{j \in B} \mathbf{h}_j \right\|_p - \left\| \sum_{j \in A} \alpha_j \mathbf{h}_j - \sum_{j \in B} \beta_j \mathbf{h}_j \right\|_p \\ &\geq (2T + 1)^{1/p} - (|A| + |B|)\tau \max \|\mathbf{h}_j\|_p \geq \Delta. \quad \blacktriangle \end{aligned}$$

§ 3. Коды, позволяющие находить носитель разреженного вектора, у которого модули всех ненулевых координат примерно равны

То, что коды конструкции, предложенной в [2] и описанной выше, позволяют находить носитель не только двоичного, но и почти двоичного разреженного вектора, было отмечено в [5], но без алгоритма декодирования. Отметим, что эта конструкция кодов была впервые предложена Т. Эриксоном и В.И. Левенштейном в [13] как конструкция сигнатурных кодов для канала множественного доступа, суммирующего по модулю 2, и после этого она переоткрывалась для разных других задач, например, для задачи построения кодов, исправляющих ошибки в канале и синдрома [14].

Предложенная конструкция очевидно ограничена применимостью только к векторам, у которых все координаты одного знака. Сейчас мы опишем более общую конструкцию, позволяющую находить разреженный вектор (а следовательно, и его носитель), все ненулевые координаты которого равны -1 или $+1$. Матрица измерений H будет троичной, т.е. $h_{i,j} \in \{-1, 0, +1\}$. То, что матрица H позволяет однозначно найти троичный вектор из уравнения (1) при дополнительном ограничении, что $\|\mathbf{w}\|_p < t_p$, означает, что для любых двух различных K -разреженных векторов $\mathbf{a}, \mathbf{b} \in \{-1, 0, +1\}^m$ справедливо неравенство

$$\|H\mathbf{a}^T - H\mathbf{b}^T\|_p \geq 2t_p. \quad (5)$$

Как и выше, вместо матрицы H мы будем рассматривать *троичный* код $\mathcal{H}$, состоящий из столбцов этой матрицы.

Определение 2. Троичный код $\mathcal{H} = \{\mathbf{h}_1, \dots, \mathbf{h}_n\} \subset \{-1, 0, +1\}^m$ называется $(K, t_p)_3$ -кодом, если для любых двух различных K -разреженных векторов $\mathbf{a}, \mathbf{b} \in$

$\in \{-1, 0, +1\}^m$ выполнено

$$\left\| \sum_{j=1}^n a_j \mathbf{h}_j - \sum_{j=1}^n b_j \mathbf{h}_j \right\|_p \geq 2t_p. \quad (6)$$

Построение троичных $(K, t_p)_3$ -кодов дословно повторяет построение их двоичных аналогов. А именно, выбираются два троичных линейных кода C и V длины n и m с известными эффективными алгоритмами исправления K и T ошибок соответственно. Пусть H' – какая-то проверочная $(r \times n)$ -матрица кода C , позволяющая эффективно исправить K ошибок, т.е. найти единственное троичное решение $\mathbf{z}$ веса Хэмминга $\text{wt}(\mathbf{z}) \leq K$ соответствующего синдромного уравнения $H' \mathbf{z}^T = \mathbf{s}$ (в поле $\mathbb{F}_3$). Код V представляет собой троичный линейный код длины m с r информационных символами, и пусть $\theta: \mathbb{F}_3^r \rightarrow V$ – его систематическое кодирование. Мы предполагаем, что код V исправляет T ошибок с помощью алгоритма декодирования $\psi: \mathbb{F}_3^m \rightarrow V$. Код C будем называть *внутренним* кодом, а код V – *внешним*.

Код $\mathcal{H}(C, V) = \{\mathbf{h}_1, \dots, \mathbf{h}_n\}$ состоит из n троичных векторов длины m , полученных применением кодирования θ к столбцам матрицы H' , т.е. $\mathbf{h}_j = \theta(\mathbf{h}'_j)$.

Лемма 2. Код $\mathcal{H}(C, V)$ является $(K, t_p)_3$ -кодом с $2t_p = (2T + 1)^{1/p}$.

Доказательство. Рассмотрим два произвольных различных K -разреженных троичных вектора $\mathbf{a}, \mathbf{b} \in \{-1, 0, +1\}^m$, соответствующие им векторы

$$\mathbf{h}_a := \sum_{j=1}^n a_j \mathbf{h}_j = \theta \left(\sum_{j=1}^n a_j \mathbf{h}'_j \right), \quad \mathbf{h}_b := \sum_{j=1}^n b_j \mathbf{h}_j = \theta \left(\sum_{j=1}^n b_j \mathbf{h}'_j \right)$$

и эти же векторы, взятые по модулю 3, а именно $\mathbf{h}_a \bmod 3$ и $\mathbf{h}_b \bmod 3$. Тогда в силу линейности отображения кодирования θ имеем

$$\mathbf{h}_a \bmod 3 = \theta \left(\sum_{j=1}^n a_j \mathbf{h}'_j \right) \bmod 3 = \theta \left(\sum_{j=1}^n a_j \mathbf{h}'_j \bmod 3 \right).$$

Аналогично

$$\mathbf{h}_b \bmod 3 = \theta \left(\sum_{j=1}^n b_j \mathbf{h}'_j \bmod 3 \right).$$

Троичные векторы $\sum_{j=1}^n a_j \mathbf{h}'_j \bmod 3$ и $\sum_{j=1}^n b_j \mathbf{h}'_j \bmod 3$ различны, так как это две различные линейные комбинации над полем $\mathbb{F}_3$ из K или менее столбцов проверочной матрицы кода, исправляющего K ошибок. Поэтому $\mathbf{h}_a \bmod 3$ и $\mathbf{h}_b \bmod 3$ – это два различных слова кода V , и тем самым, они различаются как минимум в $2T + 1$ координатах. Тем более, целочисленные векторы $\mathbf{h}_a$ и $\mathbf{h}_b$ различаются как минимум в $2T + 1$ координатах, и следовательно, $\|\mathbf{h}_a - \mathbf{h}_b\|_p^p \geq 2T + 1$. $\blacktriangle$

Обобщим теперь теорему 1, рассмотрев вместо “почти” двоичных разреженных векторов разреженные векторы, у которых модули всех ненулевых координат почти равны.

Теорема 2. Если для всех ненулевых координат K -разреженного вектора $\mathbf{x}$ справедливо $|x_j - a_j| < \tau$, $0 < \tau$, $a_j \in \{-1, +1\}$, то троичный код $\mathcal{H}(C, V)$ позволяет точно найти носитель вектора $\mathbf{x}$, если длина шума $\mathbf{w}$ в метрике ℓ_p меньше $\Delta/2$, где $\Delta = (2T + 1)^{1/p} - 2K\tau m^{1/p}$.

Доказательство. Нужно доказать, что для любых двух K -разреженных векторов $\mathbf{x}, \mathbf{y}$, чьи ненулевые координаты удовлетворяют неравенствам $|x_j - a_j| < \tau$, $|y_j - b_j| < \tau$, где все $a_j, b_j \in \{-1, +1\}$, с различными носителями $A, B \subset [n]$ справедливо неравенство

$$\left\| \sum_{j \in A} x_j \mathbf{h}_j - \sum_{j \in B} y_j \mathbf{h}_j \right\|_p \geq \Delta. \quad (7)$$

Представим $x_j = a_j + \alpha_j$, $y_j = b_j + \beta_j$, где $|\alpha_j|, |\beta_j| < \tau$. Тогда

$$\begin{aligned} \left\| \sum_{j \in A} x_j \mathbf{h}_j - \sum_{j \in B} y_j \mathbf{h}_j \right\|_p &\geq \left\| \sum_{j \in A} a_j \mathbf{h}_j - \sum_{j \in B} b_j \mathbf{h}_j \right\|_p - \left\| \sum_{j \in A} \alpha_j \mathbf{h}_j - \sum_{j \in B} \beta_j \mathbf{h}_j \right\|_p \\ &\geq (2T + 1)^{1/p} - (|A| + |B|)\tau \max \|\mathbf{h}_j\|_p \geq \Delta. \quad \blacktriangle \end{aligned}$$

§ 4. Декодирование

Начнем с алгоритма нахождения K -разреженного двоичного вектора $\mathbf{x}$ при условии, что норма шума удовлетворяет условию $\|\mathbf{w}\|_1 < 0,5(T + 1)$. В качестве предварительного (нулевого) шага декодирования выполняется аналог жесткого приема применительно к “принятому” вектору-синдрому $\mathbf{s}$ из уравнения (1), т.е. каждая координата s_i заменяется на ближайшее целое число $\hat{s}_i$. Очевидно, $\hat{s}_i \neq (H\mathbf{x}^T)_i$ только в том случае, если

$$|s_i - (H\mathbf{x}^T)_i| = |w_i| \geq 1/2.$$

Следовательно, расстояние Хэмминга удовлетворяет неравенству

$$d_H(\hat{\mathbf{s}}, H\mathbf{x}^T) \leq 2\|\mathbf{w}\|_1.$$

Тем более,

$$d_H(\hat{\mathbf{s}} \bmod 2, H\mathbf{x}^T \bmod 2) \leq 2\|\mathbf{w}\|_1,$$

и если $2\|\mathbf{w}\|_1 < T + 1$, то число “ошибок” в векторе $\hat{\mathbf{s}} \bmod 2$, т.е. отличий от вектора $H\mathbf{x}^T \bmod 2$, не более T , и алгоритм декодирования ψ , примененный к вектору $\hat{\mathbf{s}} \bmod 2$, выдаст $H\mathbf{x}^T \bmod 2$. В свою очередь, первые r символов $H\mathbf{x}^T \bmod 2$ – это синдром $H'\mathbf{x}^T \bmod 2$, поскольку кодирование θ является систематическим. Тогда искомый вектор $\mathbf{x}$ находится алгоритмом декодирования кода C .

Рассмотрим теперь нахождение носителя “почти” двоичного K -разреженного вектора $\mathbf{x}$, ненулевые координаты которого лежат в диапазоне $[1 - \tau, 1 + \tau]$, где $0 < \tau < 1/2$. Представим $\mathbf{x} = \mathbf{a} + \boldsymbol{\alpha}$, где $\mathbf{a}$ – характеристический вектор множества $\text{supp}(\mathbf{x})$, и следовательно, вектор $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ представляет собой K -разреженный вектор, в котором по предположению $|\alpha_i| \leq \tau$ для всех i . Принятый “вектор-синдром” $\mathbf{s}$ представим в виде $\mathbf{s} = H(\mathbf{a} + \boldsymbol{\alpha})^T + \mathbf{w} = H\mathbf{a}^T + \mathbf{w}'$, где $\mathbf{w}' = \mathbf{w} + H\boldsymbol{\alpha}^T$, и следовательно, $\|\mathbf{w}'\|_1 \leq \|\mathbf{w}\|_1 + K\tau m$. Поэтому описанный выше алгоритм позволяет найти двоичный вектор $\mathbf{a}$ и, тем самым, носитель $\text{supp}(\mathbf{x})$, если для нормы “шума” $\mathbf{w}'$ выполнено условие

$$\|\mathbf{w}'\|_1 < 0,5(T + 1).$$

Эти рассуждения почти дословно переносятся на “троичный” случай, т.е. поиск носителя разреженного вектора, у которого модули всех ненулевых координат примерно равны.

§ 5. Выбор кодов и сложность реализации

Оценим сложность предложенной конструкции для двух наиболее популярных асимптотических режимов. Первый – когда K постоянно, а $n \rightarrow \infty$, что соответствует практическим случаям, когда неизвестный разреженный вектор действительно разреженный. В качестве внутреннего кода C длины n с проверочной $(r \times n)$ -матрицей H' возьмем коды БЧХ или классические неприводимые коды Гоппы, для которых $n = 2^{r/K} - 1$ или $n = 2^{r/K}$ соответственно.

Что касается внешнего кода V , то снова берем код Гоппы или код БЧХ, либо, если $m \gg 1$, то код из семейства “хороших” кодов, т.е. кодов с фиксированной кодовой скоростью и отношением кодового расстояния к длине кода, отличным от нуля, плюс дополнительно с полиномиальной сложностью реализации. Например, можно взять семейство кодов-расширителей [15, 16].

Нулевой шаг декодирования – округление координат синдрома до целых значений – имеет линейную по m сложность. Первый шаг декодирования – декодирование вектора $\hat{s}$ “внешним” кодом V – имеет полиномиальную по m сложность для кодов БЧХ или для кодов-расширителей. Второй этап – синдромное декодирование “внутреннего” кода C – для кодов БЧХ имеет сложность $\text{poly}(Kn)$. Таким образом, общая сложность равна $\text{poly}(Kn)$ для K – константы и $n \rightarrow \infty$.

Второй асимптотический режим – когда K растет линейно по n . Тогда в качестве внутреннего и внешнего кодов берутся коды-расширители, что дает линейную по n сложность нахождения носителя разреженного n -мерного вектора.

Важно отметить, что предлагаемые коды могут быть явно построены, более того, с полиномиальной или даже линейной по n сложностью, но случайное кодирование позволяет увеличить кодовую скорость в $\log K$ раз (см. [17]).

В заключение, следуя [10], где исследовалась аппроксимация носителя разреженного двоичного вектора в условиях гауссовского шума, рассмотрим следующий числовой пример: $K = 25$, $n = 2^{18}$. Возьмем в качестве внутреннего кода C неприводимый код Гоппы с $r = 25 \times 18 = 450$ битами проверки на четность, а в качестве внешнего кода V – также код Гоппы длины $m = 1020$, исправляющий $T = 57$ ошибок. Тогда полученная (1020×2^{18}) -матрица H способна точно найти носитель неизвестного двоичного вектора разреженности $K = 25$ из уравнения (1) при условии, что сумма абсолютных величин координат шума w меньше, чем 57,5. Более того, предложенный алгоритм декодирования найдет носитель при условии, что сумма абсолютных величин координат шума меньше, чем 28,75.

СПИСОК ЛИТЕРАТУРЫ

1. Egorova E., Fernandez M., Kabatiansky G., Lee M.H. Signature Codes for Weighted Noisy Adder Channel, Multimedia Fingerprinting and Compressed Sensing // Des. Codes Cryptogr. 2019. V. 87. № 2–3. P. 455–462. <https://doi.org/10.1007/s10623-018-0551-9>
2. Егорова Е.Е., Фернандес М., Кабатянский Г.А., Мiao 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>
3. 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>
4. 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>
5. Джанабекова А., Кабатянский Г.А., Камель И., Рабие Т.Ф. Неперекрывающиеся выпуклые многогранники с вершинами из булева куба и другие задачи теории кодирования

ния // Пробл. передачи информ. 2022. Т. 58. № 4. С. 50–61. <https://www.mathnet.ru/rus/ppi2383>

6. *Fernandez M., Kabatiansky G., Miao Y.* A Novel Support Recovery Algorithms and Its Applications to Multiple-Access Channels // Proc. 2022 IEEE Int. Multi-Conf. on Engineering, Computer and Information Sciences (SIBIRCON). Yekaterinburg, Russian Federation. Nov. 11–13, 2022. P. 170–173. <https://doi.org/10.1109/SIBIRCON56155.2022.10017094>
7. *Polyanskiy Y.* A Perspective on Massive Random-Access // Proc. 2017 IEEE Int. Symp. on Information Theory (ISIT'2017). Aachen, Germany. June 25–30, 2017. P. 2523–2527. <https://doi.org/10.1109/ISIT.2017.8006984>
8. *Donoho D.L.* Compressed Sensing // IEEE Trans. Inform. Theory. 2006. V. 52. № 4. P. 1289–1306. <https://doi.org/10.1109/TIT.2006.871582>
9. *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>
10. *Gkagkos M., Pradhan A.K., Amalladinne V., Narayanan K., Chamberland J-F., Georgiades 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>
11. *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>
12. *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>
13. *Ericson T., Levenshtein V.I.* Superimposed Codes in the Hamming Space // IEEE Trans. Inform. Theory. 1994. V. 40. № 6. P. 1882–1893. <https://doi.org/10.1109/18.340463>
14. *Влэдуц С.Г., Кабатянский Г.А., Ломаков В.В.* Об исправлении ошибок при искажениях в канале и синдроме // Пробл. передачи информ. 2015. Т. 51. № 2. С. 50–56. <http://mi.mathnet.ru/ppi2169>
15. *Sipser M., Spielman D.A.* Expander Codes // IEEE Trans. Inform. Theory. 1996. V. 42. № 6. Part 1. P. 1710–1722. <https://doi.org/10.1109/18.556667>
16. *Spielman D.* Linear-Time Encodable and Decodable Error-Correcting Codes // IEEE Trans. Inform. Theory. 1996. V. 42. № 6. P. 1723–1731. <https://doi.org/10.1109/18.556668>
17. *Vorobyev I.* Complete Traceability Multimedia Fingerprinting Codes Resistant to Averaging Attack and Adversarial Noise with Optimal Rate // Des. Codes Cryptogr. 2023. V. 4. № 4. P. 1183–1191. <https://doi.org/10.1007/s10623-022-01144-x>

Фернандес Марсель

Политехнический университет Каталонии,

Барселона, Испания

marcelf@entel.upc.edu

Кабатянский Григорий Анатольевич

Сколковский институт науки и технологий (Сколтех), Москва

g.kabatyansky@skoltech.ru

Круглик Станислав Александрович

Наньянский технологический университет, Сингапур

stanislav.kruglik@ntu.edu.sg

Мяо Ин

Университет Цукубы, Цукуба, префектура Ибараки, Япония

miao@sk.tsukuba.ac.jp

Поступила в редакцию

30.12.2022

После доработки

21.02.2023

Принята к публикации

21.02.2023