Графы-расширители и их применения в теории кодирования

С. Б. Гашков

1. Введение

Графы-расширители (по английски экспандеры) — это очень интересный класс графов, изучение которых первыми начали московские математики М. С. Пинскер, Л. А. Бассалыго и Г. А. Маргулис в семидесятые годы прошлого века. За прошедшее время эти графы нашли много неожиданных применений, например в теории сложности вычислений и в теории кодирования. Загадочным образом они оказались также связаны с далекими от классической теории графов разделами современной математики, например, с теорией групп и теорией чисел, и являются в настоящее время предметом активных исследований, ведущихся (увы!) в основном зарубежными математиками. Библиография по этой теме насчитывает сотни публикаций. Читателю предлагается обзор некоторых из них, в основном связанных с теорией кодирования. Надеемся, что он будет способствовать активизации подобных исследований в нашей стране.

Часть доказательств упрощена по сравнению с оригинальными статьями.

1.1. Что такое графы-расширители

Графы-расширители были впервые введены М. С. Пинскером [1] (см. также [2, 3]).

Определить их можно следующим образом. Рассмотрим неориентированный граф G (возможно с петлями и параллельными ребрами), обычный или двудольный (все необходимые определения из теории графов можно найти, например, в [4]). Для произвольного множества S его вершин (взятого из одной доли, если он двудольный) обозначим ∂S множество ребер, выходящих из S в его дополнение, и назовем его *реберной границей*. Множество концов этих ребер обозначим $\Gamma(S)$ и назовем *вершинной границей*.

Назовем *реберным (вершинным) отношением расширения* графа G

число

$$h_e(G) = \min_{\{S:|S|\leqslant\frac{n}{2}\}} \frac{|\partial S|}{|S|}$$

и соответственно число

$$h(G) = \min_{\{S:|S|\leqslant\frac{n}{2}\}} \frac{|\Gamma(S)|}{|S|}\,.$$

Последовательность G_i d-регулярных графов[1)] с растущим числом вершин называется *семейством графов-расширителей*, если существует $\epsilon > 0$ такое, что для любого i справедливо неравенство $h(G_i) \geqslant \epsilon$ ($h_e(G_i) \geqslant \epsilon$, если рассматривается реберное расширение).

Это самое общее определение, но в разных статьях с разными целями даются разные его варианты, иногда для произвольных, но чаще для двудольных графов, регулярных с одной или с двух сторон (не обязательно равной мощности) и как правило для случая вершинного расширения; часто также специализируются участвующие в определении параметры.

1.2. Что такое линейные коды

Бинарным линейным кодом длины n называется любое множество двоичных векторов длины n такое, что результат покомпонентного сложения по модулю два любых двух его векторов всегда принадлежит коду. Число единиц в сумме двух векторов по модулю два называется *расстоянием* между этими векторами. Минимальное расстояние между разными векторами кода (*кодовыми словами*) называется *расстоянием кода*. Линейный код можно рассматривать как линейное пространство над полем из двух элементов. Размерность этого пространства есть *размерность кода*.

Линейные коды можно определить и над любым конечным полем. Пусть $GF(q)$ — конечное поле из q элементов[2)]. Назовем q-ичным линейным кодом C длины n и размерности k (сокращенно $[n,k]$-кодом) любое линейное k-мерное подпространство C пространства $GF(q)^n$ всех n-мерных векторов над полем $GF(q)$. Число $R = k/n$ называется *пропускной способностью* кода (для того, чтобы воспользоваться кодом, надо произвольное q-ичное слово длины k превратить в кодовое слово длины n с помощью подходящего кодирующего отображения $GF(q)^k \to C$.) *Расстоянием Хемминга* между векторами $x, y \in GF(q)^n$ называется число

[1)] Т.е. графов, у которых число ребер, выходящих из каждой вершины равно d.

[2)] Никаких существенных сведений из теории конечных полей далее не понадобится. Полезно, однако, знать, что порядок поля (число элементов в нем) есть степень простого числа, и каждое поле определяется по своему порядку однозначно (с точностью до изоморфизма), благодаря чему поле порядка q обычно обозначается $GF(q)$ (первая буква напоминает о первооткрывателе конечных полей Эваристе Галуа).

$\rho(x, y)$ координат, в которых эти вектора не совпадают. Легко видеть, что так определенное расстояние совпадает в случае $q = 2$ с введенным выше, и для него выполнено неравенство треугольника $\rho(x, y) \leqslant \rho(x, z) + \rho(y, z)$. Кодовым расстоянием называется, как и бинарном случае, минимальное расстояние между разными кодовыми векторами. Легко видеть, что кодовое расстояние линейного кода равно минимальному весу, который может иметь ненулевой кодовый вектор (*весом вектора* называется число его ненулевых координат). Отношение кодового расстояния к длине кода называется *относительным кодовым расстоянием*.

Если $[n, k]$-код имеет кодовое расстояние $d = 2t + 1$ (такие коды часто называются $[n, k, d]$-кодами), то он может исправлять вплоть до t ошибок. Действительно, если при передаче кодового слова x в нем произошло t ошибок, то мы получим искаженное слово x', для которого $\rho(x', x) = t$. По искаженному слову можно однозначно восстановить кодовое слово, так как, если из двух разных кодовых слов x, y получено одно и то же искаженное не более чем t ошибками слово $x' = y' = z$, то, согласно неравенству треугольника,

$$2t + 1 = d \leqslant \rho(x, y) \leqslant \rho(x, z) + \rho(y, z) \leqslant 2t,$$

а это невозможно. Восстановление кодового слова по искаженному слову называется *декодированием*. Очевидный алгоритм декодирования заключается в переборе всех слов, удаленных от искаженного слова на расстояние не более t, и проверке их на принадлежность данному коду. Вместо этого можно перебирать все кодовые слова и искать среди них ближайшее к данному искаженному слову. Для произвольного линейного кода эти алгоритмы могут работать слишком медленно. В теории кодирования разработаны множество методов эффективного построения линейных (и не только линейных) кодов с достаточно быстрыми алгоритмами кодирования и декодирования. Из-за ограниченности места для ознакомления с ними мы вынуждены отослать читателя к специальной литературе (например, [5]).

Однако пример одного важного в теоретическом и прикладном отношении кода можно описать в нескольких строчках. Это код Рида-Соломона над полем $GF(q)$ (сокращенно RS-код). Простейший вариант его построения следующий. Пусть $k < n \leqslant q$. Сопоставим каждому вектору $a \in GF(q)^k$ многочлен

$$a(x) = a_0 + a_1 x + \ldots + a_{k-1} x^{k-1} \in GF(q)[X]$$

степени $k - 1$ над полем $GF(q)$. Пусть $x_1, \ldots, x_n \in GF(q)$ различные элементы этого поля. Рассмотрим линейное отображение $l : GF(q)^k \to GF(q)^n$, определяемое равенством $l(a) = (a(x_1), \ldots, a(x_n)) \in GF(q)^n$. Образ $l(GF(q)^k) \subset GF(q)^n$ этого отображения — линейный код C, называе-

мый RS-кодом. В силу неравенства $n > k$ многочлен $a(x)$ степени $k-1$ однозначно восстанавливается по своим значениям в n точках, поэтому отображение $l : GF(q)^k \to C$ взаимно однозначно, значит, мощность кода C равна q^k, поэтому его размерность равна k. Кодовое расстояние $d(C) \geqslant n-k+1$, так как для любого ненулевого многочлена $a(x)$ вектор его значений $(a(x_1), \ldots, a(x_n)) \in C \subset GF(q)^n$ имеет вес, не меньший $n-k+1$, потому что ненулевой многочлен степени $k-1$ имеет не более $k-1$ корней. На самом деле $d(C) = n-k+1$, потому что согласно так называемой границе Синглтона[3)] для любого $[n,k]$-кода $d(C) \leqslant n-k+1$. Коды, лежащие на этой границе, называются *кодами с максимальным расстоянием*. Как видим, такими являются RS-коды[4)]. Доказательство границы Синглтона довольно просто и поэтому оставляется читателям (есть возможность его прочитать, например, в [5]). Но построение быстрых алгоритмов кодирования и декодирования совсем непросто и мы вынуждены опять отослать читателя, например, к [5].

Недостатком кодов Рида – Соломона является ограниченность их блоковой длины $n \leqslant q$. Далее будет описано построение некоторых классов линейных кодов с произвольной блоковой длиной, у которых и кодовое расстояние, и сложность алгоритма декодирования, и число исправляемых им ошибок растут линейно с ростом блоковой длины. Некоторые из этих кодов близки к границе Синглтона.

Методы построения этих кодов основаны на применении графов-расширителей и некоторых простых результатов линейной алгебры и не требуют знания теории кодирования.

2. Спектр графа и графы-расширители

Изучение реберного расширения обычных d-регулярных графов можно свести вершинному расширению двудольных право-d-регулярных и лево-2-регулярных графов, если заменить обычный граф $G = (V, E)$ на двудольный граф с долями V, E, в котором $v \in V$ и $e \in E$ соединяются ребром, если и только если в обычном графе G v и e инцидентны. Далее для обычных графов рассматривается реберное расширение.

2.1. Спектр графа

Сопоставим произвольному d-регулярному графу G его матрицу смежности вершин, т. е. матрицу, элементы которой $A(G)_{i,j}$ равны числу ребер, соединяющих i и j. Эта матрица действительная и симметрическая,

[3)] Это фамилия, а не термин!

[4)] Бинарные коды не достигают этой границы.

поэтому все ее собственные значения действительны. Обозначим их в убывающем порядке $\lambda_1 \geqslant \ldots \geqslant \lambda_n$. Заметим, что $Ae = de$, где $e = (1, \ldots, 1)$, потому что в каждой строке матрицы сумма чисел равна d, и очевидно для любого v

$$|(Av)_i| = \left| \sum_{i=1}^{n} A_{ij} v_j \right| \leqslant d \max |v_i| = d|v|,$$

поэтому для любого λ_j и $AV_j = \lambda_j V_j$ имеем $|AV_j| = |\lambda_j||V_j| \leqslant d|V_j|$, откуда $|\lambda_j| \leqslant d = \lambda_1$. Кратность числа λ_1 равна числу компонент связности графа. Поэтому граф связен тогда и только тогда, когда $\lambda_1 > \lambda_2$.

Граф является двудольным тогда и только тогда, когда $\lambda_n = -\lambda_1$. В этом случае все его собственные значения разбиваются на пары противоположных по знаку.

Действительно, рассматривая любую его компоненту связности и предполагая, что $Av = -dv$, имеем

$$d|v_i| = |(Av)_i| = \left| \sum_{j=1}^{n} A_{ij} v_j \right| \leqslant \sum_{j=1}^{n} A_{ij} |v_j|,$$

выбирая i так, чтобы $|v_i| = |v| = \max_j |v_j|$, получаем, что

$$d|v| = d|v_i| \leqslant \sum_{j=1}^{n} A_{ij} |v_j| \leqslant \sum_{j=1}^{n} A_{ij} |v| = d|v|,$$

откуда

$$d|v_i| = \sum_{j=1}^{n} A_{ij} |v_j| = \sum_{j \in \Gamma(i)} A_{ij} |v_j| \sum_{j=1}^{n} A_{ij} |v|,$$

значит, во всех вершинах графа, соседних с вершиной в которой v максимально по модулю, модуль v имеет то же значение. Двигаясь по вершинам компоненты связности, получаем, что на них везде модуль v одинаков. Рассматривая множества вершин, в которых значения v равны, получаем два непересекающихся подмножества рассматриваемой компоненты. Они являются долями двудольного графа, так как внутри них не бывает ребер, иначе для одной из вершин получилось бы неравенство

$$\pm d|v| = -dv_i = (Av)_i = \sum_{j=1}^{n} A_{ij} v_j = \sum_{j \in \Gamma(i)} A_{ij} \pm |v| > -dv_i.$$

Так как компоненты двудольны, то и весь граф двудолен. Если же он двудолен, то его матрица имеет вид

$$A = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}.$$

Если

$$\lambda v_i = (Av)_i = \sum_{j=1}^{n} A_{ij} v_j,$$

то, меняя знак у v на всех вершинах одной доли, получаем, что

$$-\lambda v_i = (Av)_i = \sum_{j=1}^{n} A_{ij} v_j,$$

т. е. $-\lambda$ — тоже собственное значение.

Разность $d - \lambda_2 = \lambda_1 - \lambda_2$ называется *спектральным зазором*. Все графы, имеющие положительный спектральный зазор, являются расширителями. Обозначим μ модуль следующего по абсолютной величине собственного значения после максимального значения d (по определению $\mu = \max\{|\lambda_2|, |\lambda_n|\}$, если граф не двудольный, и $\mu = \lambda_2$, если он двудольный).

В [6, 7] доказано[5)], что для любого d-регулярного n-вершинного графа справедливо неравенство $\mu \geqslant \sqrt{2d-1} - \epsilon_n$, где $\epsilon_n \to 1$.

Доказательство этого достаточно сложное. В [8] приведена более слабая оценка $\mu \geqslant \sqrt{d}(1 - \epsilon_n)$ со следующим простым доказательством.

Пусть A — матрица смежности вершин графа. Ясно, что диагональные элементы матрицы A^k есть число путей длины k, начинающихся и заканчивающихся в соответствующей вершине графа. В частности, диагональные элементы матрицы A^2 не меньше d, так как всегда есть d тривиальных циклов длины 2, проходящих через любую вершину. Поэтому след $\mathrm{tr}(A^2) \geqslant nd$. С другой стороны, матрица A^2 имеет собственные числа λ_i^2, поэтому

$$\mathrm{tr}(A^2) = \sum_{i=1}^{n} \lambda_i^2 = d^2 + \sum_{i=2}^{n} \lambda_i^2 \leqslant d^2 + \mu^2(n-1),$$

откуда $\mu^2 \geqslant d(n-d)/(n-1) = d(1 - (d-1)/(n-1)) = d(1-\epsilon_n)$.

В [6, 7, 9] d-регулярные графы, для которых $\mu = \sqrt{2d-1}$, и двудольные двусторонне d-регулярные графы, для которых $\lambda_2 = \sqrt{2d-1}$, названы *графами Рамануджана*. Другими словами, графами Рамануджана называются графы с наибольшим спектральным зазором. Эти графы достигают наименьшей границы $\mu = \sqrt{2d-1}$ для d-регулярных графов. Построены они были независимо в [6, 13] для случая $d = p + 1$, где p — простое. В [14] эти результаты были перенесены на случай $d = p^k + 1$. Для других d графы Рамануджана неизвестны.

Примеры этих графов строятся не очень сложно, но сложно доказывается их правильность. В [7] это занимает почти всю книгу[6)].

Согласно [6, 13] графом Рамануджана является граф Кэли для группы $PGL(2, \mathbb{Z}_q)$, т. е. проективной линейной группы матриц второго порядка над полем $\mathbb{Z}_q$ со специально выбранным $(p+1)$-элементным множеством S, где p — простое, не равное q.

Пусть G — группа и S — ее подмножество. *Граф Кэли* $C(G, S)$ — это ориентированный граф, имеющий G множеством вершин, в котором упорядоченная пара (g, h) является ребром, если $g = hs$, $s \in S$. Если S замкнуто относительно инвертирования, т. е. $s \in S \Rightarrow s^{-1} \in S$, то граф $C(G, S)$ можно рассматривать как обычный неориентированный граф. Очевидно, он будет $|S|$-регулярным. Если S — порождающее G множество, то он будет связным, так как для любых g, h найдутся $s_i \in S$, такие что $g = hs_1 \dots s_m$, и эти вершины связаны путем $h, hs_1, hs_1s_2, \dots, hs_1 \dots s_m = g$. Очевидно, верно и обратное. Очевидно также, что графы Кэли вершинно транзитивны, т. е.

[5)] Эта теорема принадлежит Алону и Боппане.

[6)] Про графы там только третья глава, но доказательство опирается на результаты первых двух глав о модулярных формах, в которых решается проблема Рамануджана, откуда видимо и произошло название графов.

любую вершину можно перевести в любую другую подходящим автоморфизмом графа (т. е. перестановкой вершин, переводящей рёбра в рёбра).

Группа $PGL(2,\mathbb{Z}_q)$ определяется так: берем группу $GL(2,\mathbb{Z}_q)$ невырожденных матриц второго порядка и вычисляем ее фактор-группу по подгруппе скалярных матриц, т. е. матриц, кратных единичной (эти матрицы образуют центр группы, так как коммутируют со всеми матрицами, а центр очевидно является нормальным делителем и фактор-группа действительно существует). Другими словами, матрицы разбиваются на классы эквивалентности по отношению пропорциональности матриц друг другу в обычном естественном смысле. Читатель может проверить, что $|PGL(2,\mathbb{Z}_q)| = q(q^2-1)$. Подробности см. [6, 7, 9].

Через второе по величине собственное число μ графа можно оценить среднюю степень его подграфа, порожденного данным подмножеством вершин S.

В [8] обобщенный вариант соответствующей леммы из [10] дан в таком виде: *для d-регулярного графа G и любых подмножеств S, T его вершин*

$$\left||E(S,T)| - \frac{d|S||T|}{n}\right| \leqslant \mu\sqrt{|S||T|},$$

где $|E(S,T)|$ — число ребер, идущих из S в T. При этом ребра, соединяющие вершины $S \cup T$, учитываются дважды, так как в [8] $E(S,T)$ определяется как множество ориентированных ребер, идущих из S в T (неориентированные ребра понимаются как пары ребер, ориентированных в разных направлениях).

Эта лемма в [8] названа смешанной леммой о расширителях. В [11] доказано следующее ее обращение.

Если G — d-регулярный граф, для которого всегда выполнено неравенство

$$\left||E(S,T)| - \frac{d|S||T|}{n}\right| \leqslant \rho\sqrt{|S||T|}$$

при некоторой положительной константе ρ, то $\mu(G) = O(\rho(1+\log(\frac{d}{\rho})))$, причем оценка точная.

Из леммы о расширителях следует, что *для n-вершинного d-регулярного расширителя с $\mu < d/2$ и любого множества вершин S, $|S| \leqslant n/2$, степень вершинного расширения*

$$h(S) = \frac{|\Gamma(S)|}{|S|} \geqslant 1 - \frac{\mu}{d} - \frac{|S|}{n},$$

т. е. число соседей $|\Gamma(S)|$ у множества S не меньше $|S|(1-\mu/d-|S|/n)$.

Действительно, число ребер $|E(S,S)| = 2|E(S)|$ удовлетворяет неравенству

$$|E(S,S)| \leqslant \frac{d|S|^2}{n} + \mu|S|,$$

откуда

$$|E(S,S)|/|S| \leqslant \frac{d|S|}{n} + \mu,$$

значит, число ребер, идущих из S вовне его, равно

$$|E(S,\overline{S})|/|S| \geqslant d\left(1-\frac{|S|}{n}\right)-\mu,$$

так как $E(S,S)+E(S,\overline{S})=E(S,V)=d|S|$ в силу регулярности графа. Но очевидно $|\Gamma(S)| \geqslant E(S,\overline{S})/d$, потому что из каждого соседа выходит не более d ребер, идущих в S. Поэтому

$$|\Gamma(S)| \geqslant E(S,\overline{S})/d = |S||E(S,\overline{S})|/(d|S|) \geqslant |S|\left(1-\frac{|S|}{n}-\frac{\mu}{d}\right).$$

В [12] доказан следующий вариант леммы о расширителях.

Пусть G — двусторонне d-регулярный двудольный граф с n вершинами в каждой доле. Пусть S,T — подмножества вершин разных долей. Средняя степень $d_{S,T}$ подграфа с долями S,T равна $|2E(S,T)|/(|S|+|T|)$. Тогда

$$d_{S,T} \leqslant \frac{d}{n}\frac{2|S||T|}{|S|+|T|}+\mu-\frac{\mu}{n}\frac{|S|^2+|T|^2}{|S|+|T|}.$$

Эквивалентная формулировка такова:

$$E(S,T) \leqslant \frac{d|S||T|}{n}+\frac{\mu(|S|+|T|)}{2}-\frac{\mu(|S|^2+|T|^2)}{2n}.$$

Ниже будет доказано ее обобщение.

3. Графы-расширители и коды

Далее пойдет речь о применении графов-расширителей в теории кодирования, точнее в построении линейных кодов с большим относительным расстоянием, большой пропускной способностью и линейной сложностью декодирования. Эта тематика тесно связана с так называемыми кодами с низкой плотностью проверок на четность и каскадными кодами, получившими важные промышленные применения в самое последнее время. Существенный вклад в эту теорию еще в семидесятых годах внесли московские кодировщики (см., например, [15, 16]).

3.1. Регулярные (c,d) графы

Так называются двудольные графы с долями $M,N,|M|=m,|N|=n$ у которых из левой доли M выходит по c ребер из каждой вершины, а из правой доли N — по d ребер. Очевидно $cm=dn$. В [19] при изучении кодов Таннера, о которых будет речь дальше, была получена следующая лемма.

Пусть G указанный выше граф и $S \subset M, N \subset N$. Обозначим $E(S,T)$ множество ребер из S в T. Тогда

$$\left||E(S,T)| - \frac{d|S||T|}{m}\right| \leqslant \frac{\mu}{2}\left(|S| + |T| - \frac{|S|^2}{m} - \frac{|T|^2}{n}\right).$$

При $c = d$ очевидно $m = n$ и из этого неравенства получается неравенство, приведенное в конце раздела 2.1.

На самом деле можно доказать более сильное неравенство.

$$\left||E(S,T)| - \frac{d|S||T|}{m}\right| \leqslant \mu\left(\left(|S| - \frac{|S|^2}{m}\right)\left(|T| - \frac{|T|^2}{n}\right)\right)^{1/2}.$$

Из него с помощью неравенства $\sqrt{ab} \leqslant (a+b)/2$ вытекает сформулированное выше неравенство [19].

Докажем его. Матрица графа имеет вид

$$A(G) = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix},$$

где A есть (m,n)-матрица, такая, что $a_{i,j} = 1$ если и только если вершина $i \in M$ соединяется ребром с вершиной $j \in N$. Сначала проверим, что у этой матрицы максимальное собственное значение равно $\sqrt{cd}$ и собственный вектор (ненормированный) равен

$$(\sqrt{c}, \ldots, \sqrt{c}, \sqrt{d}, \ldots, \sqrt{d}),$$

где вначале идут m равных чисел, а потом n равных чисел. Для любого вектора v длины $n+m$ и для $i \leqslant m$ очевидно

$$|(A(G)v)_i| = \left|\sum_{j=1}^{n} A_{i,j} v_{m+j}\right| \leqslant c \max_j |v_{m+j}|,$$

так как

$$\sum_{j=1}^{n} A_{i,j} = c.$$

Аналогично для $i > m$

$$|(A(G)v)_i| = \left|\sum_{j=1}^{m} A_{j,i-m} v_j\right| \leqslant d \max_{j \leqslant m} |v_j|,$$

так как

$$\sum_{j=1}^{m} A_{ji} = d.$$

Поэтому для любого λ_k и $AV_k = \lambda_k V_k$ имеем

$$|\lambda_k| \max_{i \leqslant m} |V_{k,i}| = \max_{i \leqslant m} |(AV_k)_i| \leqslant c \max_{i \leqslant n} |V_{k,m+i}|$$

и аналогично

$$|\lambda_k| \max_{i > m} |V_{k,i}| = \max_{i > m} |(AV_k)_i| \leqslant d \max_{i > m} |V_{k,i-m}|$$

откуда

$$|\lambda_k| \leqslant \frac{c \max_{i \leqslant n} |V_{k,m+i}|}{\max_{i \leqslant n} |V_{k,i}|},$$

и аналогично

$$|\lambda_k| \leqslant \frac{d \max_{i > m} |V_{k,i-m}|}{\max_{i > m} |V_{k,i}|}.$$

Перемножая эти неравенства и выполняя очевидное сокращение дробей, получаем что $|\lambda_k|^2 \leqslant cd$. Вычисляя для вектора

$$v = (\underbrace{\sqrt{c}, \ldots, \sqrt{c}}_{m}, \underbrace{\sqrt{d}, \ldots, \sqrt{d}}_{n})$$

вектор Av, рассматривая по отдельности случаи $i \leqslant m$ и $i > m$, получаем, что в первом случае

$$|(A(G)v)_i| = \left| \sum_{j=1}^{n} A_{i,j} v_{m+j} \right| = c\sqrt{d} = \sqrt{cd}\sqrt{c} = \sqrt{cd} v_i,$$

и аналогично во втором случае. Поэтому v — собственный вектор для значения $\sqrt{cd}$. Аналогично проверяется, что противоположному собственному значению $-\sqrt{cd}$ отвечает вектор

$$v = (\underbrace{\sqrt{c}, \ldots, \sqrt{c}}_{m}, \underbrace{-\sqrt{d}, \ldots, -\sqrt{d}}_{n}).$$

Пусть 1_S — характеристический вектор множества S, т. е. он равен 1 на вершинах из S и нулю вне их. Аналогично вводится вектор 1_T.

Пусть v_i — вектора ортонормированного базиса, состоящего из собственных векторов матрицы графа A. Вектор, соответствующий числу $\lambda_1 = \sqrt{cd}$, после нормировки будет равен

$$v_1 = \frac{1}{\sqrt{2cm}} (\underbrace{\sqrt{c}, \ldots, \sqrt{c}}_{m}, \underbrace{\sqrt{d}, \ldots, \sqrt{d}}_{n}),$$

так как норма исходного вектора равна $cm + dn = 2cm$. Аналогично, вектор, соответствующий числу $\lambda_{n+m} = -\sqrt{cd}$, равен

$$v_{n+m} = \frac{1}{\sqrt{2cm}} (\underbrace{\sqrt{c}, \ldots, \sqrt{c}}_{m}, \underbrace{-\sqrt{d}, \ldots, -\sqrt{d}}_{n}).$$

Рассмотрим вектора

$$1_S - \alpha_1 v_1 - \alpha_{n+m} v_{n+m}, 1_T - \beta_1 v_1 - \beta_{n+m} v_{n+m}.$$

У первого из них координаты при $i \in S$ равны

$$1 - \frac{|S|}{m},$$

при $i \notin S, i \leqslant m$ равны

$$-\frac{|S|}{m},$$

при $i > m$ равны 0. Поэтому его евклидова норма равна

$$\|1_S - \alpha_1 v_1 - \alpha_{n+m} v_{n+m}\|_2 = \left(\left(1 - \frac{|S|}{m}\right)^2 |S| + (m - |S|) \frac{|S|^2}{m^2} \right)^{1/2} = \left(|S| - \frac{|S|^2}{m} \right)^{1/2}.$$

Аналогично, евклидова норма вектора $1_T - \beta_1 v_1 - \beta_{n+m} v_{n+m}$ равна

$$\left(|T| - \frac{|T|^2}{n}\right)^{1/2}.$$

Очевидно их координаты в базисе $\{v_1, \dots, v_{n+m}\}$ равны

$$(0, \alpha_2, \dots, \alpha_{n+m-1}, 0), (0, \beta_2, \dots, \beta_{n+m-1}, 0),$$

поэтому, используя неравенство Коши-Буняковского, имеем

$$\sum_{i=2}^{n+m-1} |\alpha_i \beta_i| \leqslant \left(\sum_{i=2}^{n+m-1} |\alpha_i|^2\right)^{1/2} \left(\sum_{i=2}^{n+m-1} |\beta_i|^2\right)^{1/2} =$$
$$= \|1_S - \alpha_1 v_1 - \alpha_{n+m} v_{n+m}\|_2 \|1_T - \beta_1 v_1 - \beta_{n+m} v_{n+m}\|_2 =$$
$$= \left(\left(|S| - \frac{|S|^2}{m}\right)\left(|T| - \frac{|T|^2}{n}\right)\right)^{1/2}.$$

Отсюда

$$\left||E(S,T)| - \frac{d|S||T|}{m}\right| \leqslant \mu \left(\sum_{i=2}^{n+m-1} |\alpha_i|^2\right)^{1/2} \left(\sum_{i=2}^{n+m-1} |\beta_i|^2\right)^{1/2} =$$
$$= \mu \|1_S - \alpha_1 v_1 - \alpha_{n+m} v_{n+m}\|_2 \|1_T - \beta_1 v_1 - \beta_{n+m} v_{n+m}\|_2 =$$
$$= \mu \left(\left(|S| - \frac{|S|^2}{m}\right)\left(|T| - \frac{|T|^2}{n}\right)\right)^{1/2}.$$

3.2. Коды Таннера

Они были введены Таннером в [17], но получили популярность после статьи [18]. Результаты [18] улучшены в [12], и далее в [20, 21]. Результаты [12] перенесены на случай (c,d)-регулярных двудольных графов в [19].

Дадим определение кода Таннера. Это будет линейный код над произвольным полем $GF(q)$. Пусть G есть (c,d)-регулярный двудольный граф с долями M, N мощностей m, n и множеством ребер E мощности $e = cm = dn$. С каждой вершиной $v \in M$ ассоциируем линейный $[c, r_1 c, d_1 = \delta_1 c]$-код C_1, где c — его блоковая длина (длина кодовых слов), $r_1 c$ — его размерность как линейного пространства (r_1 — это пропускная способность), а d_1 — его расстояние (δ_1 — относительное минимальное расстояние). Аналогично с каждой вершиной $v \in N$ ассоциируем линейный $[d, r_2 d, d_2 = \delta_1 d]$-код C_2. Занумеруем ребра и каждому из них сопоставим символ $x_i, 1 \leqslant i \leqslant e$. Символы, соответствующие ребрам, выходящим из вершины $v \in M$ обозначим $x_{v(1)}, \dots, x_{v(c)}$. Из них составим вектор $E(v) = (x_{v(1)}, \dots, x_{v(c)})$. Аналогично для любой $u \in M$ определим вектор $E(u) = (x_{u(1)}, \dots, x_{u(d)})$.

Код Таннера $\mathcal{C}(G, C_1, C_2)$, *ассоциированный с графом* G *и кодами* C_i, определяется как линейный код с блоковой длиной $e = mc = nd$, определяемый условиями $\forall v \in M\ E(v) \in C_1$, $\forall u \in N\ E(u) \in C_2$. Другими словами, он состоит из всех векторов, у которых проекции на множества координат $E(v)$, $v \in M$, $E(u)$, $u \in N$ (т. е. соответствующие подвектора) принадлежат кодам C_i соответствующей длины. Так как проекция линейной комбинации векторов равна линейной комбинации проекций, то код Таннера линеен.

Оценим его размерность. Если в его определении оставить только ограничения, соответствующие вершинам доли M, т. е. рассмотреть код определяемый условиями $\forall v \in M\ E(v) \in C_1$, то его размерность будет равна $r_1 mc = r_1 e$, так как подвектора $E(v)$ попарно не имеют общих координат и указанный набор условий определяет код, который можно рассматривать как прямое произведение m экземпляров кода C_1.

Аналогично, код, определяемый условиями $\forall u \in N\ E(u) \in C_2$, имеет размерность $r_2 e$.

Код $\mathcal{C}(G, C_1, C_2)$ равен пересечению этих кодов. Известна лемма о том, что пересечение подпространств L_i размерности D_i пространства L размерности D имеет размерность не меньше $D_1 + D_2 - D$.

Применяя эту лемму, получаем, что размерность кода $\mathcal{C}(G, C_1, C_2)$ не меньше $r_1 e + r_2 e - e = (r_1 + r_2 - 1)e$.

Что касается расстояния кода Таннера, то оценки [18] улучшены в [12] и обобщены в [19]. Последний результат таков: *если* $d_i > \mu(G)/2$, *где* $\mu(G)$ — *второе по величине собственное число графа* G, *то*

$$d(\mathcal{C}(G, C_1, C_2)) \geqslant \frac{m}{d}\left(d_1 d_2 - \frac{\mu(G)}{2}(d_1 + d_2)\right),$$

а если в терминах относительных расстояний, то

$$\delta(\mathcal{C}(G, C_1, C_2)) = \frac{d(\mathcal{C}(G, C_1, C_2))}{e} \geqslant \frac{1}{cd}\left(d_1 d_2 - \frac{\mu(G)}{2}(d_1 + d_2)\right) =$$

$$= \delta_1 \delta_2 - \frac{\mu(G)}{2}\left(\frac{\delta_1}{d} + \frac{\delta_2}{c}\right).$$

3.3. Нижняя граница для минимального расстояния кода Таннера

Сначала получим следствие границы для плотности ребер в двудольном графе

$$\left||E(S,T)| - \frac{d|S||T|}{m}\right| \leqslant \mu\left(\left(|S| - \frac{|S|^2}{m}\right)\left(|T| - \frac{|T|^2}{n}\right)\right)^{1/2}$$

в более удобном и компактном виде. Введем обозначения $s = |S|/m$, $t = |T|/n$ для относительного числа вершин в S, T, $e = mc = nd = \sqrt{mncd}$ для числа ребер в графе G, $\gamma(G) = \mu(G)/\sqrt{cd}$ для нормализованного второго собственного значения (очевидно $0 < \gamma(G) < 1$). Тогда для относительного числа ребер справедливо неравенство

$$\left|\frac{|E(S,T)|}{e} - st\right| = \left|\frac{|E(S,T)|}{nd} - st\right| = \left|\frac{|E(S,T)|}{e} - \frac{|S||T|}{mn}\right| \leqslant$$
$$\leqslant \frac{\mu}{\sqrt{cd}}\left(\left(\frac{|S|}{m} - \frac{|S|^2}{m^2}\right)\left(\frac{|T|}{n} - \frac{|T|^2}{n^2}\right)\right)^{1/2} =$$
$$= \frac{\mu}{\sqrt{dc}}\sqrt{(s-s^2)(t-t^2)} = \gamma\sqrt{(s-s^2)(t-t^2)}.$$

Так как очевидно при $0 < s, t \leqslant 1$

$$\sqrt{(s-s^2)(t-t^2)} = \sqrt{st}\sqrt{(1-s)(1-t)} \leqslant \sqrt{st}(1-\sqrt{st}) = \sqrt{st} - st,$$

потому что

$$\sqrt{(1-s)(1-t)} + \sqrt{st} \leqslant \frac{1-s+1-t}{2} + \frac{s+t}{2} = 1,$$

то из предыдущего неравенства следует, что

$$\left|\frac{|E(S,T)|}{e} - st\right| \leqslant \gamma\sqrt{(s-s^2)(t-t^2)} \leqslant \gamma(\sqrt{st} - st),$$

в частности

$$\frac{|E(S,T)|}{e} \leqslant st + \gamma(\sqrt{st} - st) = (1-\gamma)st + \gamma\sqrt{st}.$$

В конце раздела 3.2 было обещано доказать, что если $d_i > \mu/2$, то

$$d(\mathcal{C}(G, C_1, C_2)) \geqslant \frac{m}{d}\left(d_1 d_2 - \frac{\mu}{2}(d_1 + d_2)\right).$$

Докажем более сильное неравенство

$$\frac{d(\mathcal{C}(G, C_1, C_2))}{e} \geqslant D = \frac{1}{cd}\frac{d_1 d_2 - \mu\sqrt{d_1 d_2}}{1-\gamma},$$

где $e = mc = nd$ — число ребер в G.

Обозначая относительное расстояние $d(\mathcal{C}(G, C_1, C_2))/e$ через δ и используя обозначения для относительных расстояний $\delta_1 = d_1/c$, $\delta_2 = d_2/d$, можно переписать это неравенство как

$$\delta \geqslant \frac{\delta_1\delta_2 - \gamma\sqrt{\delta_1\delta_2}}{1-\gamma},$$

где $0 < \gamma = \mu/\sqrt{cd} < 1$. В частном случае $c = d$ это неравенство получено в [21]. Заметим, что при $\gamma \to 0$

$$\delta \geqslant \frac{\delta_1\delta_2 - \gamma\sqrt{\delta_1\delta_2}}{1-\gamma} = \delta_1\delta_2\left(1 - \gamma((\delta_1\delta_2)^{-1/2} - 1) - O(\gamma^2)\right).$$

Так как код Таннера линеен, то достаточно взять ненулевое кодовое слово $\mathbf{x} \in \mathcal{C}(G, C_1, C_2)$, рассмотреть соответствующее ему множество ребер $X = \{e_i \in E : x_i \neq 0\}$

и доказать, что $|X|/e \geqslant D$. Рассмотрим минимальный подграф, содержащий множество X, т. е. возьмем множества вершин

$$S = \{u \in M : \exists v \in N\ (u, v) \in X\} \subset M, T = \{v \in N : \exists u \in N\ (v, u) \in X\} \subset N,$$

и рассмотрим порожденный множеством вершин $S \cup T$ подграф с множеством ребер $E(S,T)$, $X \subset E(S,T)$. Очевидно $|X| \leqslant |E(S,T)|$. Так как $\mathbf{x}$ — кодовый вектор, то по определению кода Таннера его подвектора, образованные координатами из множеств $E(v)$, $v \in S$, принадлежат коду C_1, а подвектора, образованные координатами из множеств $E(u)$, $u \in T$, принадлежат коду C_2, причем они не равны нулю в силу определения S, T. Поэтому подвектора, образованные координатами из множеств $E(v)$, $v \in S$, содержат не менее $d_1 = \delta_1 c$ ненулевых координат, значит, общее число ненулевых координат в $\mathbf{x}$, т. е. $|X|$, не меньше $d_1|S| = \delta_1 cms = \delta_1 es$. Аналогично $|X| \geqslant \delta_2 et$. Отсюда очевидно имеем

$$\frac{|X|}{e} \geqslant \sqrt{\delta_1\delta_2 st}.$$

В начале этого раздела было доказано, что

$$\frac{|X|}{e} \leqslant \frac{|E(S,T)|}{e} \leqslant (1-\gamma)st + \gamma\sqrt{st}.$$

Поэтому

$$\sqrt{\delta_1\delta_2 st} \leqslant \frac{|X|}{e} \leqslant (1-\gamma)st + \gamma\sqrt{st},$$

откуда $\sqrt{st}(\sqrt{\delta_1\delta_2} - \gamma) \leqslant (1-\gamma)st$, значит

$$\frac{\sqrt{\delta_1\delta_2} - \gamma}{1-\gamma} \leqslant \sqrt{st},$$

следовательно

$$D = \frac{\delta_1\delta_2 - \gamma\sqrt{\delta_1\delta_2}}{1-\gamma} \leqslant \sqrt{\delta_1\delta_2 st} \leqslant \frac{|X|}{e}\ .$$

3.4. Декодирование кода Таннера

Согласно определению кода Таннера $\mathcal{C}(G, C_1, C_2)$ в (c,d)-регулярном графе G множество ребер E разбито двумя способами на подмножества

$$E = \bigcup_{i=1}^{m} E_{v_i} = \bigcup_{j=1}^{n} E_{u_j}.$$

Так как код линейный, то достаточно рассмотреть случай, когда в нулевом кодовом слове произошли ошибки и получилось ненулевое слово $y \in GF(q)^e$, $e = mc = nd$. Итеративный алгоритм устранения ошибок выглядит следующим образом.

Для каждой вершины $v \in M$ берем подвектор $E(v) = (x_{v(1)}, \ldots, x_{v(c)})$ и, так как он принадлежит $[c, r_1 c, \delta_1 c]$-коду C_1, то применяем алгоритм декодирования для этого кода. Эти процедуры декодирования для всех $v \in M$ можно делать параллельно. В случае, если вес подвектора $E(v)$

меньше $d_1/2$, то в результате декодирования получается правильное слово, т. е. нулевое, в противном случае может получиться неправильное слово, т. е. ненулевое. Из полученных подслов составляем слово $z \in GF(q)^e$ — результат первого шага алгоритма декодирования кода Таннера. К полученному слову применяем аналогичную параллельную процедуру, декодирующую подслова $E(u)$, $u \in N$, соответствующие вершинам правой доли N графа G. В результате получим слово $w \in GF(q)^e$. Если $w = z$, то алгоритм заканчивает работу и выдает ответ $w = z$. Если нет, то к слову w применяем опять процедуру параллельного декодирования вначале всех подслов $E(v), v \in M$, а потом — всех подслов $E(u), u \in N$ и т. д.

Можно показать, что при любом α, $0 < \alpha < 1$, таком, что $\sqrt{\delta_1\delta_2} > 2\gamma/\alpha$, этот алгоритм заканчивает работу за

$$\frac{\log_2(mn)}{2\log_2\frac{1}{\alpha}} + O_{\mathcal{C},\alpha}(1),\ \mathcal{C} = \mathcal{C}(G, C_1, C_2)$$

шагов и правильно декодирует любое слово $y \in GF(q)^e$ с числом ненулевых символов не более

$$e\frac{\alpha\delta_1\delta_2 - 2\gamma\sqrt{\delta_1\delta_2}}{4(1-\gamma)}\,.$$

Мультипликативные константы в этих оценках зависят от алгоритмов декодирования кодов C_1, C_2 и определяются величинами d_i, c, d. Числа m, n можно выбирать при этом сколь угодно большими.

Заметим, что при $\gamma \to 0$

$$\frac{\alpha\delta_1\delta_2 - 2\gamma\sqrt{\delta_1\delta_2}}{4(1-\gamma)} = \frac{\alpha}{4}\delta_1\delta_2\left(1 - \gamma\left(\frac{2}{\alpha}(\delta_1\delta_2)^{-1/2} - 1\right) - O(\gamma^2)\right).$$

Ограничение на вес вектора Y (число исправляемых ошибок) имеет вид

$$\frac{\alpha\delta_1\delta_2 - 2\gamma\sqrt{\delta_1\delta_2}}{4(1-\gamma)} \geqslant \frac{|Y|}{e}.$$

Для расстояния кода Таннера выше была получена оценка

$$\frac{d(\mathcal{C}(G, C_1, C_2))}{e} \geqslant \frac{\delta_1\delta_2 - \gamma\sqrt{\delta_1\delta_2}}{1-\gamma}\,.$$

Указанное выше ограничение чуть более, чем в четыре раза хуже. Если взять в качестве G граф с $\gamma(G) \to 0$ (например, при $c = d$ — граф Рамануджана), то асимптотически отношение расстояния кода к числу исправляемых этим алгоритмом ошибок будет равно четырем. Оценка для числа итераций имеет вид

$$k = \frac{\log_2\left(\frac{2|Y|}{\sqrt{d_1 d_2}}\right)}{\log_2\frac{1}{\alpha}} \leqslant \frac{\log_2\left(\sqrt{mn}\frac{\alpha\sqrt{\delta_1\delta_2} - 2\gamma}{2(1-\gamma)}\right)}{\log_2\frac{1}{\alpha}} = \frac{\log_2(mn)}{2\log_2\frac{1}{\alpha}} + O_{\mathcal{C},\alpha}(1).$$

3.5. Конкретный пример кода Таннера

В [12] предложен следующий пример кода Таннера. В качестве G берем граф Рамануджана (построенный методом [6, 7]). У этого графа $m = n = q$ — простое число, $c = d = p + 1$, $p = 223$, согласно [7] q следует выбрать так, чтобы p не было квадратом по модулю q. Соответствующий граф имеет $q(q^2 - 1)$ вершин и является двудольным и

$$\gamma(G) = \mu(G)/(p+1) = 2\sqrt{p}/(p+1) = \sqrt{223}/112 < 15/112.$$

В качестве кодов C_i берем $[224, 115, 30]$-BCH-код[7)]. Тогда относительное расстояние

$$\frac{d(\mathcal{C}(G, C_1, C_2))}{e} = \delta \geqslant \frac{\delta_1\delta_2 - \gamma\sqrt{\delta_1\delta_2}}{1-\gamma} = (30/224)\frac{30/224 - \gamma}{1-\gamma} > 9 \cdot 10^{-5},$$

а пропускная способность (т. е. размерность, деленная на длину блока) не меньше $2 \cdot 115/224 - 1 = 6/224$. Блоковая длина $(p+1)q(q^2-1) > 2.5 \cdot 10^9$. Условие $\sqrt{\delta_1\delta_2} > 2\gamma$ для него выполняется.

4. Экспандерные коды с почти минимальным расстоянием

В [20] показано, что *для любой пропускной способности $R \in (0, 1]$ и $\epsilon < \epsilon_0$ можно построить бесконечное семейство кодов с пропускной способностью не менее R над алфавитом размера $2^{O(\log(1/\epsilon)/(R\epsilon^4))}$, и относительным минимальным расстоянием не меньше $1 - R - \epsilon$. Кодирование осуществляется в линейное время со сложностью $(1/\epsilon)^{O(1)}$ на символ. Декодировать можно долю ошибок, не меньшую $(1 - R - \epsilon)/2$ от длины кода, в линейное время со сложностью $(1/\epsilon)^{O(1)}$ на символ.*

Рот и Скачек в [21] показали, что в этом результате можно улучшить границу для мощности алфавита до $2^{O(\log(1/\epsilon)/\epsilon^3)}$.

Опишем их конструкцию чуть в более общем виде. Вспомним конструкцию кода Таннера $\mathcal{C}(G, C_1, C_2))$ над алфавитом $F = GF(q)$. Здесь C_1 линейный $[c, r_1c, d_1 = \delta_1 c]$-код, где c — его блоковая длина, ассоциированный с каждой вершиной $v \in M$, и с каждой вершиной $u \in N$ ассоциируем линейный $[d, r_2d, d_2 = \delta_1 d]$-код C_2. Занумеруем ребра и каждому из них сопоставим элемент $x_i \in F, 1 \leqslant i \leqslant e$. Элементы, соответствующие ребрам, выходящим из вершины $v \in M$ обозначим $x_{v(1)}, \ldots, x_{v(c)}$. Из них составим вектор $E(v) = (x_{v(1)}, \ldots, x_{v(c)})$. Аналогично для любой $u \in N$ определим вектор $E(u) = (x_{u(1)}, \ldots, x_{u(d)})$. Код Таннера $\mathcal{C}(G, C_1, C_2)$ над полем $GF(q)$, ассоциированный с графом G и кодами C_i, определяется как линейный код с блоковой длиной $e = mc = nd$, определяемый условиями $\forall v \in M\ E(v) \in C_1$, $\forall u \in N\ E(u) \in C_2$.

Пусть $\Phi = F^{r_1}$ — новый алфавит. Фиксируем биективное линейное отображение кодирования $\mathcal{E}$ из r_1-мерного пространства $\Phi = F^{r_1}$ над полем F в кодовое пространство C_1 той же размерности над тем же полем и

[7)] О кодах Боуза-Чоудхури-Хоквингема, см., например, [5].

рассмотрим биективное линейное отображение $\varphi_{\mathcal{E}} : \mathcal{C} \to \Phi^m$, определяемое равенством

$$\varphi_{\mathcal{E}}(c) = (\mathcal{E}^{-1}(E(v)) : v \in M) \in \Phi^m.$$

Образ кода $\mathcal{C}$ при этом отображение является линейным кодом $\mathcal{C}_\Phi$ над алфавитом Φ с блоковой длиной m. Размерность кода $\mathcal{C}(G, C_1, C_2)$, как было показано в разделе 3.2, не меньше $(r_1 + r_2 - 1)e = (r_1 + r_2 - 1)cm$. Размерность кода $\mathcal{C}_\Phi$ над алфавитом F такая же. Блоковая длина над алфавитом F равна $r_1 cm$, поэтому пропускная способность над алфавитом F не меньше

$$(r_1 + r_2 - 1)e/(r_1 cm) = (r_1 + r_2 - 1)/r_1 = 1 - 1/r_1 + r_2/r_1.$$

Получим нижнюю оценку для относительного расстояния кода $\mathcal{C}_\Phi$ над алфавитом Φ. Очевидно относительное расстояние равно $d(\mathcal{C}_\Phi)/m$. Докажем, что

$$\frac{d(\mathcal{C}_\Phi)}{m} \geqslant \frac{\delta_2 - \gamma\sqrt{s/t}}{1 - \gamma}\,.$$

Так как код $\mathcal{C}_\Phi$ линеен, то достаточно взять произвольное ненулевое кодовое слово

$$\mathbf{c} = (\mathcal{E}^{-1}(E(v)) : v \in M) \in \mathcal{C}_\Phi \subset \Phi^m,$$

и оценить снизу его вес, т. е. число ненулевых символов из Φ в нем. Рассмотрим соответствующее ему кодовое слово

$$\mathbf{x} = (E(v) : v \in M) \in \mathcal{C}(G, C_1, C_2) \subset F^e,$$

и возьмем соответствующее ему множество ребер $X = \{e_i \in E : x_i \neq 0\}$. Вес слова $\mathbf{c} \in \Phi^m$ очевидно совпадает с числом ненулевых векторов $E(v)$, $v \in M$, составляющих вектор $\mathbf{x} \in \mathcal{C}(G, C_1, C_2)$. Рассмотрим минимальный подграф, содержащий множество X, т. е. возьмем множества вершин

$$S = \{u \in M : \exists v \in N\ (u, v) \in X\} \subset M,$$
$$T = \{v \in N : \exists u \in N\ (v, u) \in X\} \subset N,$$

и рассмотрим порожденный множеством вершин $S \cup T$ подграф с множеством ребер $E(S, T)$, $X \subset E(S, T)$. Очевидно $|X| \leqslant |E(S, T)|$. Так как $\mathbf{x}$ — кодовый вектор, то по определению кода $\mathcal{C}_\Phi$ его подвектора, образованные координатами из множеств $E(v)$, $v \in S$, принадлежат коду C_1, а подвектора, образованные координатами из множеств $E(u)$, $u \in T$, принадлежат коду C_2, причем они не равны нулю в силу определения S, T, к тому же число ненулевых векторов $E(v)$, $v \in M$, в точности равно $|S|$, и число ненулевых векторов $E(u)$, $u \in N$, в точности равно $|T|$. Поэтому вес слова $\mathbf{c} \in \Phi^m$ равен $|S|$, откуда для оценки снизу $d(\mathcal{C}_\Phi)/m$ достаточно оценить снизу $|S|/m = s$.

Так как подвектора, образованные координатами из множеств $E(v)$, $v \in S$, содержат не менее $d_1 = \delta_1 c$ ненулевых координат (это кодовые слова кода C_1), значит общее число ненулевых координат в $\mathbf{x}$, т. е. $|X|$, не меньше $d_1|S| = \delta_1 cms = \delta_1 es$. Аналогично $|X| \geqslant \delta_2 et$, где $t = |T|/n$. Отсюда имеем

$$\frac{|X|}{e} \geqslant max\{\delta_1 s, \delta_2 t\}.$$

В разделе 3.3 было доказано, что

$$\frac{|X|}{e} \leqslant \frac{|E(S,T)|}{e} \leqslant (1-\gamma)st + \gamma\sqrt{st}, \frac{\sqrt{\delta_1\delta_2} - \gamma}{1-\gamma} \leqslant \sqrt{st},$$

следовательно при $s/t \geqslant \delta_2/\delta_1$

$$\frac{\delta_2 - \gamma\sqrt{\frac{\delta_2}{\delta_1}}}{1-\gamma} = \frac{\sqrt{\delta_1\delta_2} - \gamma}{1-\gamma}\sqrt{\frac{\delta_2}{\delta_1}} \leqslant \sqrt{st}\sqrt{s/t} = s,$$

а при $s/t < \delta_2/\delta_1$ очевидно

$$\delta_2 t \leqslant \frac{|X|}{e} \leqslant (1-\gamma)st + \gamma\sqrt{st},$$

откуда $\delta_2 t - \gamma\sqrt{st} \leqslant (1-\gamma)st$, значит, и в этом случае

$$\frac{\delta_2 - \gamma\sqrt{\frac{\delta_2}{\delta_1}}}{1-\gamma} \leqslant \frac{\delta_2 - \gamma\sqrt{s/t}}{1-\gamma} \leqslant s.$$

Теперь можно показать, как для любого достаточно малого ϵ выбрать код $\mathcal{C}_\Phi$ с данной пропускной способностью R и минимальным относительным расстоянием $1 - R - \epsilon$. Для данного $\epsilon > 0$, выберем $8/\epsilon^3 > c = d \geqslant 4/\epsilon^3$, $2c \geqslant q > c$, так чтобы $c = p + 1$, p — простое число, простое $p_1 = O(p)$ выбираем так, чтобы p не было квадратом по модулю p_1. Тогда согласно [7] соответствующий граф Рамануджана имеет $n = p_1(p_1^2 - 1) = O(p_1^3) = = O(1/\epsilon^9)$ вершин и является двудольным и для него

$$\gamma = \gamma(G) = \mu(G)/(p+1) = 2\sqrt{p}/(p+1) < \epsilon^{3/2}.$$

Пусть $\delta_1 = \epsilon$, $\delta_2 = \delta$, возьмем любое $R = r_2 > 1 - \delta$, $r = r_1 > 1 - \delta_1 = 1 - \epsilon$. Коды C_i с этими параметрами можно выбрать среди RS-кодов. Тогда код $\mathcal{C}_\Phi$ имеет пропускную способность не меньше

$$1 - \frac{1}{r_1} + \frac{r_2}{r_1} > 1 - \frac{1}{1-\epsilon} + \frac{R}{1-\epsilon} = \frac{R-\epsilon}{1-\epsilon} > R - \epsilon,$$

и относительное расстояние не меньше

$$\frac{\delta - \gamma\sqrt{\frac{\delta}{\epsilon}}}{1-\gamma} > \delta - \gamma\sqrt{\frac{\delta}{\epsilon}} > \delta - \epsilon^{3/2}\sqrt{\frac{1}{\epsilon}} = \delta - \epsilon > 1 - R - \epsilon.$$

Таким образом, код $\mathcal{C}_\Phi$ близок к границе Синглтона при $\epsilon \to 0$. Кроме того, $|\Phi| = q^{rc} < q^c = 2^{O(\epsilon^{-3}\log 1/\epsilon)}$.

Однако уже при $\epsilon = 1/10$ все оценки выходят за пределы разумных. Какой-то смысл имеет, например, выбор $c = d = 114$, $q = 2^7$, тогда соответствующий граф Рамануджана при выборе $p = 113$, $p_1 = 137$ имеет $n = 137(137^2 - 1) = 2571218$ вершин и является двудольным и для него

$$\gamma = \gamma(G) = \mu(G)/(p+1) = 2\sqrt{p}/(p+1) = \sqrt{113}/57 = 0.186\ldots.$$

В качестве кода C_1 возьмем $[114, 57, 58]$-RS-код, для него $r_1 = 57/114 = 1/2$, $\delta_1 = 58/114 = 0.508\ldots$, а в качестве C_2 возьмем $[114, 75, 40]$-RS-код, для него $r_2 = 75/114$, $\delta_2 = 40/114 = 20/57$, тогда условие $\sqrt{\delta_1\delta_2} > 2\gamma$, достаточное для хорошей оценки скорости декодирования, выполнено. Тогда код $\mathcal{C}_\Phi$ имеет длину $n = 2571218$, пропускную способность не меньше

$$1 - \frac{1}{r_1} + \frac{r_2}{r_1} = -1 + 75/57 = 18/57 = 6/19,$$

и относительное расстояние не меньше

$$\frac{\delta_2 - \gamma\sqrt{\dfrac{\delta_2}{\delta_1}}}{1-\gamma} = \frac{\dfrac{20}{57} - \dfrac{\sqrt{113\cdot 20}}{57\sqrt{29}}}{1 - \dfrac{\sqrt{113}}{57}} = 0.24\ldots.$$

Мощность алфавита для этого кода $|\Phi| = q^{r_1 c} = 2^{7\cdot 57} = 2^{399}$, т. е. буквы имеют длину 399 бит.

Приведенный пример показывает, что описанные выше конструкции, весьма эффектные теоретически, к сожалению, пока не применимы на практике.

4.1. Декодирование кода Рота-Скачека

Оно похоже на декодирование кода Таннера. Согласно определению кода Таннера $\mathcal{C}(G, C_1, C_2)$ в (c, d)- регулярном графе G множество ребер E разбито двумя способами на подмножества

$$E = \bigcup_{i=1}^{m} E_{v_i} = \bigcup_{j=1}^{n} E_{u_j}.$$

Так как код линейный, то достаточно рассмотреть случай, когда в нулевом кодовом слове произошли ошибки и получилось ненулевое слово $y \in \Phi^m$. Итеративный алгоритм устранения ошибок выглядит следующим образом.

Для каждой вершины $v \in M$ берем $\mathcal{E}^{-1}(E(v)) \in \Phi$ — букву кодового слова, по ней вычисляем подвектор $E(v) = (x_{v(1)}, \ldots, x_{v(c)})$, и так как он принадлежит $[c, r_1 c, \delta_1 c]$ коду C_1, то применяем алгоритм декодирования для этого кода. Эти процедуры декодирования для всех $v \in M$ можно делать параллельно. В случае, если вес подвектора $E(v)$ меньше $d_1/2 = \delta_1 c$, то в результате декодирования получается правильное слово, т. е. нулевое, в противном случае может получиться ненулевое. Из полученных слов

составляем слово $z \in GF(q)^e$ — результат первого шага алгоритма. К полученному слову применяем аналогичную параллельную процедуру, декодирующую подслова $E(u)$, $u \in N$, соответствующие вершинам правой доли N графа G. В результате получим слово $w \in GF(q)^e$. Если $w = z$, то алгоритм заканчивает работу и выдает ответ $w = z$. Если нет, то к слову w применяем опять процедуру параллельного декодирования вначале всех подслов $E(v)$, $v \in M$, а потом — всех подслов $E(u)$, $u \in N$ и т. д. Покажем, что при $\sqrt{\delta_1\delta_2} > 2\gamma$ этот алгоритм заканчивает работу за

$$(\log \sqrt{mn})/\log(\delta_1\delta_2/4\gamma^2)) + O_{\delta_1,\delta_2,\gamma}(1)$$

итераций, выполняет $O(\sqrt{mn})$ раз процедуры декодирования кодов C_i и правильно декодирует любое слово $y \in \Phi^m$ с числом ненулевых символов не более

$$m\frac{\delta_2/2 - \gamma\sqrt{\delta_2/\delta_1}}{1-\gamma}\,.$$

Числа m, n, c, d можно выбирать при этом сколь угодно большими. Тогда можно выбрать граф так, что $\gamma \to 0$, при этом

$$\frac{\delta_2/2 - \gamma\sqrt{\delta_2/\delta_1}}{1-\gamma} \to \delta_2/2,$$

если δ_2/δ_1 ограничено.

Рассмотрим множество Y ребер графа G, соответствующих ненулевым координатам декодируемого вектора $y \in \Phi^m$. Выберем минимальные множества вершин $S \subset M$, $T \subset N$, такие, что $Y \subset E(S,T)$. По условию

$$s = |S|/m < \beta = \frac{\delta_2/2 - \gamma\sqrt{\delta_2/\delta_1}}{1-\gamma}\,.$$

Далее, как и в случае кода Таннера, рассмотрим множество $S_1 \subset S$, состоящее из вершин степени $\geqslant d_1/2$ в подграфе (S,T,Y) и множество W ребер, соответствующих ненулевым координатам вектора w. Тогда $W \subset E(S_1,T)$, так как $W \subset Y \subset E(S,T)$, а все рёбра, выходящие из вершин со степенями, меньшими $d_1/2$, не входят во множество W, так как после декодирования становятся нулевыми (при декодировании слова с весом меньшим $d_1/2$ из кода C_1 с расстоянием d_1 получается нулевое слово, так как оно есть ближайшее кодовое слово и удалено на расстояние меньше $d_1/2$). Рассмотрим в графе $E(S_1,T)$ множество $T_1 \subset T$, состоящее из вершин степени $< d_1/2$. Тогда по аналогичным соображениям имеем, что множество Z ребер, соответствующих ненулевым координатам вектора z, содержится в $E(S_1,T_1)$.

Поэтому $\delta_1 s_1 \leqslant 2((1-\gamma)s_1 t + \gamma\sqrt{s_1 t})$, где $s_1 = |S_1|/m, t = |T|/n$, и аналогично $\delta_2 t_1 \leqslant 2((1-\gamma)s_1 t_1 + \gamma\sqrt{s_1 t_1})$, где $t_1 = |T_1|/n$, а также

$$\sqrt{\delta_1\delta_2}\sqrt{s_1 t_1} \leqslant 2((1-\gamma)st + \gamma\sqrt{st}).$$

Из $\delta_2 t_1 \leqslant 2((1-\gamma)s_1 t_1 + \gamma\sqrt{s_1 t_1})$ следует, что $\gamma\sqrt{s_1 t_1} \geqslant t_1(\delta_2/2 - (1-\gamma)s_1)$, откуда

$$\sqrt{s_1/t_1} \geqslant \frac{(\delta_2/2) - (1-\gamma)s_1}{\gamma} > \frac{(\delta_2/2) - (1-\gamma)\beta}{\gamma} = \sqrt{\delta_2/\delta_1}.$$

Значит, $\sqrt{t_1} \leqslant \sqrt{s_1}\sqrt{\delta_1/\delta_2} \leqslant \sqrt{\beta}\sqrt{\delta_1/\delta_2}$, поэтому

$$\sqrt{s_1 t_1} \leqslant \sqrt{s_1}\sqrt{\beta}\sqrt{\delta_1/\delta_2} \leqslant \beta\sqrt{\delta_1/\delta_2} = \frac{\sqrt{\delta_1\delta_2}/2 - \gamma}{1-\gamma}\,.$$

Выполняя второй шаг итерации, по множествам S_1, T_1 построим множества S_2, T_2 и т. д. Из доказанного выше следует, что

$$\sqrt{s_{i+1}t_{i+1}} \leqslant \frac{2((1-\gamma)s_i t_i + \gamma\sqrt{s_i t_i})}{\sqrt{\delta_1\delta_2}}\,.$$

Полагая для краткости $x_i = \sqrt{s_i t_i}$, имеем

$$x_{i+1} \leqslant f(x_i),\ f(x) = ax^2 + bx,\ a = \frac{2(1-\gamma)}{\sqrt{\delta_1\delta_2}}, b = \frac{2\gamma}{\sqrt{\delta_1\delta_2}}\,.$$

Так как положительный корень уравнения $f(x) = x$ равен

$$\frac{1-b}{a} = \frac{\sqrt{\delta_1\delta_2}/2 - \gamma}{1-\gamma},$$

и $0 < x_1 = \sqrt{s_1 t_1} < (1-b)/a$, то

$$x_{i+1} \leqslant f(x_i) < x_i < x_1 < \frac{1-b}{a},\ \frac{x_{i+1}}{x_i} \leqslant ax_i + b < q = ax_1 + b < 1,$$

откуда $x_{i+1} \leqslant x_1 q^i$, т. е. после каждой итерации алгоритма величина

$$\sqrt{st} = \frac{\sqrt{|S||T|}}{\sqrt{mn}}$$

убывает не медленней, чем в геометрической прогрессии.

Если $st > 0$, то очевидно $st = |S||T|/(mn) \geqslant 1/(mn)$, поэтому при $st < 1/mn$ на самом деле $st = 0$, значит, $|S| = 0$ или $|T| = 0$, поэтому $|Y| = 0$ и алгоритм заканчивает работу, построив из ненулевого вектора (содержащего ошибки) нулевой вектор. Отсюда вытекает для числа итераций оценка $\log_{1/q}\sqrt{mn} + 1$, где $1/q = 1/(ax_1 + b) > 1$. Чтобы ее улучшить, прологарифмируем неравенство $\dfrac{x_{i+1}}{x_i} \leqslant ax_i + b < ax_1 + b < 1$, тогда

$$\log\frac{x_{i+1}}{x_i} \leqslant \log(ax_i + b) = \log b + \log\left(1 + \frac{ax_i}{b}\right) \leqslant \log b + \frac{ax_i}{b} \leqslant \log b + \frac{ax_1}{b}q^{i-1},$$

откуда, суммируя, имеем $\log\dfrac{x_{i+1}}{x_1} \leqslant i\log b + \dfrac{ax_1}{b(1-q)}$, значит,

$$x_{i+1} \leqslant \frac{1-b}{a}e^{\frac{1-b}{b(1-q)}}b^i,$$

поэтому для числа итераций справедлива оценка

$$\nu = \log_{1/b}\frac{a\sqrt{mn}}{1-b} + \frac{1-b}{b(1-q)\log(1/b)} + 1.$$

Для числа применений алгоритмов декодирования C_i очевидно справедлива оценка

$$m\left(1 + \sum_{i=1}^{\nu} s_i\right) + n\left(1 + \sum_{i=1}^{\nu} t_i\right).$$

Применяя неравенства $\delta_1 s_{i+1} \leqslant 2((1-\gamma)s_i t_i + \gamma\sqrt{s_i t_i}), \delta_2 t_{i+1} \leqslant 2((1-\gamma)s_i t_i + \gamma\sqrt{s_i t_i})$, имеем оценку

$$m\left(1+\beta+\frac{2}{\delta_1}\sum_{i=1}^{\nu-1}((1-\gamma)s_i t_i+\gamma\sqrt{s_i t_i})\right)+n\left(1+\frac{2}{\delta_2}\sum_{i=1}^{\nu-1}((1-\gamma)s_i t_i+\gamma\sqrt{s_i t_i})\right)=$$
$$=O_{\delta_1,\delta_2,\beta}(n+m)\left((1-\gamma)\sum_{i=1}^{\nu-1}x_i^2+\gamma\sum_{i=1}^{\nu-1}x_i\right)=$$
$$=O_{\delta_1,\delta_2,\beta}(n+m)\left(\frac{\gamma}{a}e^{\frac{1-b}{b(1-q)}}+\frac{1-\gamma}{a(1+b)}e^{\frac{2-2b}{b(1-q)}}\right).$$

Так как суммарная сложность применения каждого из алгоритмов декодирования кодов C_i равна $O_{\delta_1,\delta_2,r_1,r_2,c,d}(1)$, то сложность декодирования кода Рота-Скачека равна

$$O_{\delta_1,\delta_2,r_1,r_2,c,d,\beta}(m)\left(\frac{\gamma}{a}e^{\frac{1-b}{b(1-q)}}+\frac{1-\gamma}{a(1+b)}e^{\frac{2-2b}{b(1-q)}}\right).$$

Автор благодарен М. Н. Вялому за конструктивную критику, в значительной мере способствовавшую улучшению текста статьи.

Список литературы

[1] Pinsker M. S. On the complexity of a concentrator. In 7th International Telegraffic Conference, pages 318/1-318/4, 1973.

[2] Бассалыго Л.А., Пинскер М.С. Сложность оптимальной неблокирующей коммутационной схемы. Проблемы передачи информации, 9(1):84-87, 1973.

[3] Маргулис Г.А. Точные конструкции расширителей. Проблемы передачи информации. 9(4):71-80, 1973.

[4] Ловас Л., Пламмер М. Прикладные задачи теории графов. М. Мир, 1998.

[5] Сидельников В.М. Теория кодирования. Физматлит, 2008.

[6] Lubotzky A., Philips R., Sarnak P. Ramanujan graphs. Combinatorica 8(3), 261-277 (1988).

[7] Сарнак П. Модулярные формы и их приложения. М. Фазис, 1998.

[8] Hoory S., Linial N., Wigderson A. Expaner graphs and their applications Bull. Amer. Math. Assoc. S-0273-0979(06), 01126-8 (2006).

[9] Davidoff G., Sarnak P., Valette A. Elementary number theory, group theory and Ramanujan graphs. Cambridge University Press, 2003.

[10] Alon N., Chung F.R.K. Explicit constructions of linear sized tolerant networks, Discr.Math, 72, 15-19(1988).

[11] Bilu Y., Linial N. Lifts, discrepancy and nearly optimal spectral gaps. Combinatorica, to appear.

[12] Zemor G. On expander codes. IEEE Trans.Inform. Theory, IT-47(2), 835-837 (2001).

[13] Маргулис Г.А. Точные теоретико-групповые конструкции комбинаторных схем и их приложения в построении расширителей и концентраторов. Проблемы передачи информации. 24(1), 39-46 (1988).

[14] Morgenstern M. Existence and explicit constructions of $(q + 1)$-regular Ramanujan graphs for every prime power q. J. Comb. Theory, Ser. B, 62, 44-62 (1994).

[15] Зяблов В.В. Оценка сложности построения бинарных линейных каскадных кодов Проблемы передачи информации. 7(1), 3-10 (1971).

[16] Dumer I. Concateneted codes and their multilevel generalizations. In Handbook of coding theory, v.2, North-Holland, 1988, pp.1911-1988.

[17] Tanner R.M. A recursive approach to low complexity codes. IEEE Trans. Inform. Theory, IT-27(5), 533-547 (1981).

[18] Sipser M. , Spielman D. A. Expander codes. IEEE Trans. Inform. Theory, 42(6, part 1):1710-1722, 1996.

[19] Janwa H., Lal A.K. On Tanner codes: minimum distance and decoding. AAECC13 (2002).

[20] Guruswami V., Indyk P. Linear time encodable/decodable codes with near-optimal rate, IEEE Trans. Inform. Theory, IT-51(10), 3393-3400 (2005).

[21] Roth R., Skachek V. Improved nearly-MDS expander codes, IEEE Trans. Inform. Theory, (2006).

С. Б. Гашков, механико-математический факультет МГУ