Экспандеры: конструкции и приложения.

Жалобы на ошибки, опечатки и непонятные места в этом тексте можно посылать Андрею Ромащенко по адресу andrei.romashchenko@gmail.com

11 марта 2019 г.

Оглавление

Глава 1

Введение

Экспандерами, или расширяющими графами, называют класс разреженных графов (графов с относительно небольшим числом рёбер), обладающих замечательными комбинаторными свойствами — *сильной связности*, *вершинного и рёберного расширения*, *быстрого перемешивания* и т.д.

Понятие экспандера первые возникло в начале 1970-х годов в работах М.С. Пинскера и Л.А. Бассалыго. Интерес к экспандерам возрос в 1990-х, когда появились многочисленные приложения экспандеров в самых разных областях математики и информатики. Экспандеры оказались связаны одновременно и с вполне абстрактным областям математики (аддитивная комбинаторика, теория чисел, теория представлений), и с теорией сложности вычислений (вероятностно проверяемые доказательства, оценка сложности аппроксимации, методы дерандомизации), и с прикладными инженерными задачами (например, помехоустойчивое кодирование).

В этой книге мы рассматриваем экспандер прежде всего как инструмент в теоретической информатике. Мы обсуждаем разные варианты определения экспандера (для однородных и двудольных графов, комбинаторные и спектральные) и их взаимосвязь, описываем несколько методов эффективного построения экспандеров и рассматриваем различные примеры использования экспандеров в теории сложности вычислений и теории кодирования.

Мы излагаем результаты, для понимания которых и не требуются знания, выходящие за пределы университетского курса математики. Несмотря на то, что вся книга посвящена особому классу графов, никаких специальных знаний по теории графов от читателя не требуется — достаточно лишь знать, что такое граф. Мы предполагаем, что читатель знаком со стандартными университетскими курсами линейной алгебры и теории вероятностей. Желательны также начальные знания по теории сложности вычислений (понятие эффективной вычислимости, вычисления с ограничением на используемые время и память, сложностные классы детерминированных и вероятностных алгоритмов) и теории кодирования (линейные

коды). В отдельных главах используются понятия из курса общей алгебры (группы, конечные поля, представления и характеры конечных групп) и анализа (преобразование Фурье). Таким образом, материал этой книги доступен студентам старших курсов, имеющим базовую математическую подготовку.

1.1 Как организована эта книга

Главы, отмеченные звёздочкой, при первом чтении можно пропускать. Главы 7 и 8 (о разных применениях экспандеров) можно читать независимо друг от друга.

1.2 Чего в этой книге нет

Мы не претендуем на исчерпывающий обзор всех важных результатов об экспандерах или даже применений экспандеров в теоретической информатике; отбор материала отражает субъективные вкусы автора. Прекрасный обзор теории экспандеров, написанный с точки зрения computer scientists, можно найти в [1].

Мы не касаемся использования экспандеров в математике. Читателю–математику мы рекомедуем обзор [2].

Наконец, мы лишь коротко упоминаем один из самых удивительных успехов теории экспандеров — эффективное построение графов Рамануджана. Изучение данной темы требует по-настоящему серьезной математической подготовки. Подробное изложение этого вопроса можно найти в [3].

В этой книге мы не обсуждаем другие многочисленые типы комбинаторных объектов, близкие по своим свойствам к экспандерам — extractors, dispersers, randomness conductors, hitting set generators (мы приводим английские названия, поскольку русская терминология в этой области науки ещё не сложилась).

1.3 Учебная литература по экспандерам

Для computer scientists: кроме специализированного обзора [1] мы рекомендуем посвященные экспанерам главы в [5] и [6] и материалы на homepage Луки Тревисана (`http://www.eecs.berkeley.edu/~luca/`). Для математиков: обзор [2], монография [3], а также заметки по курсу лекций [4]. Студенты всех специальностей найдут для себя что-то интересное в [30].

1.4 Используемые обозначения

Для неориентрированных графов мы используем обозначение $G = (V, E)$, где V есть множество вершин, а E — множество рёбер. При этом мы до-

пускаем графы с петлями и с кратными (параллельными) рёбрами[1]. Для двудольных графов мы иногда используем обозначение $G = (R, L, E)$, чтобы подчёркнуть, что множество вершин разбито на два непересекающихся класса L и R («левая» и «правая» доли соответственно); E по-прежнему обозначает множество рёбер (у каждого ребра один из концов принадлежит L, а другой R).

Если A и B являются подмножествами (возможно, пересекающимися) вершин графа, мы обозначаем $E(A, B)$ множество рёбер, у которых один из концов принадлежит A, а второй B. Также мы обозначаем $\Gamma(v)$ множество соседей вершины v (множество всех вершин w, соединённых с v ребром). Аналогичное обозначение используется для множеств вершин: если A есть подмножество вершин графа, то $\Gamma(A)$ обозначает множество всех соседей A, т.е.,

$$\Gamma(A) = \bigcup_{v \in A} \Gamma(v).$$

Векторы-строки мы обозначаем $\mathbf{x} = (x_1, \ldots, x_n)$. Транспонирование матрицы M мы обозначаем $M^\perp$; в частности, если $\mathbf{x} = (x_1, \ldots, x_n)$, то соответствующий вектор-столбец обозначается

$$\mathbf{x}^\perp = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

1.5 Исправление ошибок и опечаток.

Автор благодарит внимательных читателей, обнаруживших многочисленные ошибки в ранних вариантах этих заметок: Н. Верещагина, М. Вялого, А. Козачинского, Е. Намаконова.

Наверняка в тексте остались неисправленные ошибки. Жалобы на замеченные опечатки, ошибки и непонятные места можно посылать автору на электронный адрес `andrei.romashchenko@gmail.com`.

[1] Для графа с кратными рёбрами правильнее называть E не множеством, а *мультимножеством* рёбер, поскольку для каждой пары вершин в E может содержаться больше одного ребра с данными концами.

Глава 2

Комбинаторные определения экспандеров

В этой главе мы рассмотрим базовые определения экспандеров и докажем существование графов, удовлетворяющих этим определениям.

2.1 Однородные экспандеры.

Начнём мы с самого простого варианта определения экспандера. Мы определим экспандер как однородный граф со свойством вершинного расширения (потребуем, чтобы у каждого не сишком большого множества вершин графа имелось достаточно много соседей). Сформулируем определение более точно.

Определение 1 *Граф $G = (V, E)$ называется однородным комбинаторным (n, d, ε)-экспандером (расширяющим графом), если $|V| = n$ (в графе n вершин), степени всех вершин равны d (допускаются кратные ребра и петли), и выполняется следующее свойство* вершинного расширения*: для любого множества $S \subset V$, $|S| \leqslant n/2$ множество соседей S достаточно велико: $|\Gamma(S)| > (1 + \varepsilon)|A|$.*

Замечание 1: Чем больше значение ε в определении 1, тем более сильное свойство требуется от графа.

Замечание 2: Степень вершины графа — это число рёбер, для которых данная вершина является концом. Это определение распространяется и на петли (ребра, у которых концы совпадают). Таким образом, если некоторой вершине инцидентны d_1 рёбер, не являющихся петлями, и ещё d_2 петель, то степень этой вершины равна $d_1 + d_2$ (каждая петля учитывается с кратностью один, как и всякое другое ребро).

Теорема 1 *Пусть ε – некоторое положительное число меньшее 1. Тогда для всех достаточно больших четных d и всех n существует однородный (n, d, ε)-экспандер.*

Доказательство: Мы выберем граф случайно и покажем, что с положительной (и даже довольно близкой к 1) вероятностью такой граф оказывается экспандером. Отсюда будет следовать, что экспандеры существуют.

Прежде всего, нам нужно уточнить, что означает *случайный выбор графа*. Другими словами, нужно зафиксировать распределение вероятностей на графах. Мы выберем случайно $d/2$ перестановок π_i на множестве вершин графа,

$$\pi_i \ : \ \{1, \ldots, n\} \to \{1, \ldots, n\}, \ i = 1, \ldots, d/2.$$

(Каждая перестановка π_i выбирается среди $n!$ равновероятных вариантов; при этом все $d/2$ перестановок выбираются незамисимо друг от друга.) Ребрами графа будем считать все (неупорядоченные) пары вершин $\{v, \pi_i(v)\}$. Таким образом, из каждой вершины v выходит $d/2$ рёбер $\{v, \pi_i(v)\}$ и ещё $d/2$ рёбер $\{v, \pi_i^{-1}(v)\}$. У перестановок могут быть неподвижные точки (перестановка может оставлять некоторые вершины на месте), так что в случае $v = \pi_i(v)$ мы получаем петлю — ребро, оба конца которой совпадают с v. Чтобы степень каждой вершины была равна d, мы будем учитывать каждую петлю дважды.

Отметим, что в случайно выбранном графе с положительной вероятностью появляются кратные рёбра (поскольку одно и то же ребро $\{v, \pi_i(v)\}$ может получаться из нескольких перестановок π_i).

Теперь оценим вероятность того, что полученный в результате граф *не* окажется экспандером. Согласно определению, граф не является экспандером, если найдется множество вершин S (состоящее из не более, чем $n/2$ вершин), все соседи которого лежат в некотором множестве T, состоящем из $\lfloor (1+\varepsilon)|S| \rfloor$ вершин.

Зафиксируем некоторые множества вершин S и T. Зафиксируем номер перестановки π_i. Вероятность того, что для каждой вершины $v \in S$ второй конец ребра $\{v, \pi_i(v)\}$ попадёт в T, равна

$$\frac{|T|}{n} \cdot \frac{|T|-1}{n-1} \cdot \ldots \cdot \frac{|T|-|S|+1}{n-|S|+1} \leqslant \left(\frac{|T|}{n}\right)^{|S|}.$$

Поскольку мы выбираем $d/2$ перестановок независимо, вероятность того, что данное событие произойдёт для всех i, не превосходит $\left(\frac{|T|}{n}\right)^{(d/2)\cdot|S|}$. Таким образом,

$$\text{Prob[свойство экспандерности графа нарушено]} \leqslant \sum_{S,T} \left(\frac{|T|}{n}\right)^{(d/2)\cdot|S|},$$

где суммирование происходит по всем множествам вершин S размера не более $n/2$ и по всем множествам T размера $\lfloor (1+\varepsilon)|S| \rfloor$.

На самом деле интересующая нас вероятностью ещё меньше — чтобы свойство экспандерности нарушилось, для каждой вершины $v \in S$ все рёбра вида $\{v, \pi_i^{-1}(v)\}$ также должны попасть в T. Но мы не будем этого учитывать; рёбер вида $\{v, \pi_i(v)\}$ уже достаточно, чтобы получить нужную нам оценку на вероятность «неприятного» события.

Оценим интересующую нас сумму:

$$\sum_{S,T} \left(\frac{|T|}{n}\right)^{(d/2)\cdot|S|} \leqslant \sum_{s=1}^{n/2} C_n^s \cdot C_n^{(1+\varepsilon)s} \cdot \left(\frac{(1+\varepsilon)s}{n}\right)^{sd/2}. \tag{2.1}$$

Каждый биномиальный коэффициент C_n^k можно оценить сверху величиной $\left(\frac{ne}{k}\right)^k$ (см. упражнение 2 ниже). Таким образом, сумма (2.1) не превосходит

$$\sum_{s=1}^{n/2} \left(\frac{ne}{s}\right)^s \cdot \left(\frac{ne}{(1+\varepsilon)s}\right)^{(1+\varepsilon)s} \cdot \left(\frac{(1+\varepsilon)s}{n}\right)^{sd/2} =$$

$$= \sum_{s=1}^{n/2} \left[(s/n)^{d/2-2-\varepsilon} \cdot (1+\varepsilon)^{d/2} \cdot \frac{e^{1+\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right]^s. \tag{2.2}$$

Остаётся заметить, что $s \leqslant n/2$, а $1+\varepsilon < 2$. Таким образом, можно подобрать такое $d = d(\varepsilon)$, чтобы выражение в квадратных скобках в правой части (2.2) было меньше $1/2$ при всех значениях s. Следовательно, сумма (2.1) меньше единицы. А это и означает, что с положительной вероятностью случайный граф является (n, d, ε)-экспандером. Теорема доказана.

Упражнение 1 *Оцените асимптотику зависимости $d = d(\varepsilon)$ в теореме 1: насколько большой должна быть степень графа, чтобы гарантировать существование экспандера с данным параметром расширения ε?*

Упражнение 2 *Докажите для биномиальных коэффициентов оценку*

$$C_n^k \leqslant \left(\frac{ne}{k}\right)^k,$$

где e — основание натурального логарифма.

Упражнение 3 *Докажите следующие утверждения:*

(а) Вероятность того, что в случайно выбранной (по равномерному распределению) перестановке $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ нет ни одной неподвижной точки, стремится к $1/e$ при $n \to \infty$;

(б) Утверждение теоремы 1 выполнено для графов без петель: для любого $\varepsilon < 1$, для всех достаточно больших четных d и всех n существует однородный комбинаторный (n, d, ε)-экспандер без петель.

Следующее упражнение показывает, что $d = 3$ есть минимальная степень графа, для которой определение однородного комбинаторного экспандера имеет смысл.

Упражнение 4 *(а) Докажите, что для некоторого $\varepsilon > 0$ и всех достаточно больших чётных n существует однородный комбинаторный $(n, 3, \varepsilon)$-экспандер.*

(б) Докажите, что для всякого $\varepsilon > 0$ найдётся такое n_0, что при $n > n_0$ однородных комбинаторных $(n, 2, \varepsilon)$-экспандеров не существует.

Замечание: Не следует воспринимать определение 1 догматически — в некоторых случаях может оказаться удобно его немного подправить. В стандартном определении требуется, чтобы свойство расширения выполнялось для множеств, содержащих не более 50% от всех вершин графа. Выбор границы $n/2$ в определении достаточно произволен и не существенен для построения теории экспандеров. Для приложений иногда бывает удобнее потребовать, чтобы свойство расширения выполнялось лишь для достаточно малых множеств A (например, для множеств, содержащих не более 1% всех вершин графа) или, напротив, даже для достаточно больших множеств (например, для всех множеств, содержащих не более 99% всех вершин графа).

Упражнение 5 *Докажите, что для любого целого $d \geqslant 3$, любого $\delta > 0$ найдётся такое $\rho > 0$, что всех достаточно больших n, для большинства d-регулярных графов с n вершинами*

$$\min_{S \subset V,\ |S| \leqslant \rho n} \frac{|\Gamma(S)|}{|S|} \geqslant d - 1 - \delta$$

Объясните, почему оценку $(d - 1 - \delta)$ в правой части неравенства нельзя заменить на величну $d - \delta$.

2.2 Коэфициенты вершинного и рёберного расширение графа

В нашем основном определении 1 на странице 6 требуется, чтобы граф обладал свойством «вершинного расширения». Введём числовую характеристику графа — коэффициент вершинного расширения, который показывает, насколько хорошими «экспандерными» свойствами обладает данный граф.

Определение 2 *Будем называть* коэффициентом вершинного расширения *графа $G = (V, E)$ число*

$$h_V(G) = \min_{0 < |S| \leqslant |V|/2} \frac{|\Gamma(S) \backslash S|}{|S|}$$

(*минимум берётся по множествам S, содержащим не более половины вершин графа*).

Заметим, что у каждого однородного (n, d, ε)-экспандера коэффициент вершинного расширения не меньше ε. С другой стороны, если к d-однородному графу без петель с n вершинами и с коэффициентом вершинного расширения ε добавить петли (в каждой из вершин), мы получим однородный $(n, d+1, \varepsilon)$-экспандер. Также можно рассмотреть свойство *рёберного расширения* графа:

Определение 3 *Будем называть* коэффициентом рёберного расширения *графа* $G = (V, E)$ *число*

$$h_E(G) = \min_{0<|S|\leqslant|V|/2} \frac{|E(S, \bar{S})|}{d|S|}$$

(*снова минимум берётся по всем множествам, содержащим не более половины вершин графа*).

Большое значение коэффициента рёберного расширения означает, что для любого множества вершин S достаточно большая доля рёбер, выходящих из вершин этого множества, приходят в вершины вне S (так сказать, «торчат наружу»).

Упражнение 6 *Докажите, что для любого d-однородного графа G*

$$h_V(G) \geqslant h_E(G).$$

Упражнение 7 *Пусть некоторый d-однородный граф с n вершинами имеет коэффициент рёберного расширения ε. Добавим к каждой вершине графа петлю. Докажите, что получившийся в результате граф будет $(n, d+1, \varepsilon)$-экспандером в смысле определения 1.*

Насколько сильным свойством рёберного расширения может обладать граф? Из определения ясно, что коэффициент рёберного расширения любого графа заведомо меньше единицы. Однако, как оказывается, единица — слишком оптимистичная оценка для $h_E(G)$. Следующее утверждение показывает, что коэффициент рёберного расширения не может быть намного больше $1/2$.

Утверждение 1 *Для любого d-регулярного графа с n вершинами*

$$h_E(G) \leqslant \frac{1}{2} + O(1/n).$$

Доказательство: Пусть число вершин n чётно. Мы покажем, что в любом d-регулярном графе найдется множество вершин S из $n/2$ вершин, для которого значение $|E(S, \bar{S})|$ сравнительно невелико. Для этого мы возьмем *случайное* множество из $n/2$ вершин и посчитаем среднее значение величины $\frac{|E(S,\bar{S})|}{|S|}$.

Каждая вершина попадает в S с вероятностью $1/2$. У каждого ребра вероятность того, что один его конец попадет в S, а второй не попадет, равна

$$\frac{n/2}{n-1} = 1/2 + O(1/n).$$

Следовательно, математическое ожидание $|E(S, \bar{S})|$ равно общему числу рёбер в графе, умноженному на $1/2 + O(1/n)$.

Можно заключить, что найдется хотя бы одно множество вершин S размера $n/2$, у которого число ребер, пересекающих границу S, не превосходит $dn/4 + O(d)$. Это и означает, что коэффициент рёберного расширение графа не превосходит $1/2 + O(1/n)$. Таким образом, для чётных n утверждение доказано.

Упражнение 8 *Докажите утверждение 1 для нечётных n.*

Замечание: Можно доказать, что в достаточно больших графах коэффициент вершинного расширения должен быть строго меньше $1/2$. Точнее, существует такая константа $C > 0$, что для всех достаточно больших n в каждом d-регулярном графе с n вершинами найдётся множество S для которого

$$\frac{|E(S, \bar{S})|}{d|S|} \leqslant \frac{1}{2} - \frac{C}{\sqrt{d}},$$

см. [23]. Ниже мы покажем, что существуют экспандеры, у которых коэффициент рёберного расширения приближается к данной границе: существуют сколь угодно большие графы G степени d, у которых $h_E(G) = \frac{1}{2} - O(1/\sqrt{d})$.

2.3 Двудольные экспандеры.

В этой главе мы введём ещё одно важное определение расширяющего графа — двудольный экспандер.

Определение 4 *Двудольный граф $G = (L, R, E)$ (L и R — левая и правая доли графов, E — множество рёбер) называется двудольным $(n, m, d, k, \varepsilon)$-экспандером, если $|L| = n$, $|R| = m$, степень всех вершин в левой доле L равна d, и выполняется следующее свойство* расширения*: для любого множества $S \subset L$, $|S| \leqslant k$ множество соседей (соседи S лежат в R) достаточно велико: $|\Gamma(S)| > (1 - \varepsilon)d|S|$.*

Замечание: Чем меньше значение ε в этом определении, тем сильнее требование к графу. В приложениях как правило используют двудольные экспандеры с $\varepsilon < 1/2$. А для применения в теории кодирования (для построения экспандерных кодов) часто требуются двудольные экспандеры с ещё меньшими значениями ε.

Теорема 2 *Пусть ε – некоторое положительное число. Тогда для любых n и $k \leqslant n$ найдется $d = O(\log n)$ и $m = O(dk)$ такие, что существует двудольный $(n, m, d, k, \varepsilon)$-экспандер.*

Замечание: Константы в $O(\cdot)$-обозначения в этой теореме зависят от ε.

Доказательство: Выберем граф случайно. Это значит, что для каждой вершины в L мы случайно и независимо выбираем d соседей в R (таким образом, разрешаются кратные рёбра). Покажем, что с большой вероятностью такой граф оказывается экспандером.

Граф *не* является экспандером, если некоторое множество вершин из левой доли графа $S \subset L$ (размера не более k) имеет не больше $(1-\varepsilon)d|S|$ соседей. Другими словами, все соседи S содержатся в некотором подмножестве правой доли графа $T \subset R$, состоящем из $(1-\varepsilon)d|S|$ вершин.

Поскольку при случайном выборе графа мы проводим все nd рёбер случайно и независимо, то для каждого ребра вероятность того, что его правый конец окажется в фиксированном множестве T, равна $|T|/m$. Следовательно,

$$\text{Prob[свойство экспандерности графа нарушено]} \leqslant \sum_{S,T} \left(\frac{|T|}{m}\right)^{sd},$$

где суммирование происходит по всем множествам $S \subset L$ размера не более k и по всем множествам $T \subset R$ размера $(1-\varepsilon)d|S|$. Оценим данную сумму сверху:

$$\sum_{s=1}^{k} C_n^s \cdot C_m^{(1-\varepsilon)sd} \cdot \left(\frac{(1-\varepsilon)sd}{m}\right)^{sd} \tag{2.3}$$

Оценивая биномиальные коэффициенты, получаем, что сумма не превосходит

$$\sum_{s=1}^{k} \left(\frac{ne}{s}\right)^{s} \cdot \left(\frac{me}{(1-\varepsilon)sd}\right)^{(1-\varepsilon)sd} \cdot \left(\frac{(1-\varepsilon)sd}{m}\right)^{sd} \leqslant$$

$$\leqslant \sum_{s=1}^{k} \left[\frac{ne}{s} \cdot \left(\frac{e^{(1-\varepsilon)/\varepsilon}(1-\varepsilon)sd}{m}\right)^{\varepsilon d}\right]^{s}. \tag{2.4}$$

Положим $m := Const \cdot kd$ (с достаточно большим значением $Const$), чтобы для всех возможных s выполнялось неравенство $\frac{e^{(1-\varepsilon)/\varepsilon}(1-\varepsilon)sd}{m} \leqslant 1/2$. Тогда выражение в квадратных скобках в правой части (2.4) не превосходит $ne \cdot (1/2)^{\varepsilon d}$. Остаётся выбрать d большим $\frac{1}{\varepsilon}\log(2en)$, и мы получаем

$$ne \cdot (1/2)^{\varepsilon d} < 1/2.$$

Таким образом, для выбранных значений параметров суммы (2.3) и (2.4) не превосходят 1. Это и означает, что с положительной вероятностью случайный двудольный граф является $(n, m, k, d, \varepsilon)$-экспандером. Теорема доказана.

Упражнение 9 *Докажите несколько более сильный вариант теоремы 2: при заданных параметрах n и k ($k \leqslant n$) существует существует двудольный $(n, m, d, k, \varepsilon)$-экспандер с $d = O(\log \frac{n}{k})$ и $m = O(\log dk)$ (усиление состоит в более точной оценке для степени графа d). Указание: выражение в квадратных скобках в правой части (2.4) можно оценить более точно.*

Упражнение 10 *Оцените асимптотику зависимости d и m от ε в теореме 2: как зависят степень графа и размер правой доли от параметра расширения ε?*

2.4 Замечание об эффективных конструкциях

Когда мы говорим о экспандерных свойствах графа — вершинном или рёберном расширении, различных вариантах свойства «сильной связности» или «быстром перемешивании» (мы обсудим эти свойства в следующей главе) — возможные различные постановки вопроса:

1. Типичные свойства графа: каковы свойства «типичного», случайно выбранного графа? Например, что можно утверждать про коэффициент вершинного расширения для 99% графов степени d с n вершинами?

2. Экстремальные свойства графов: насколько сильными экспандерными свойствами может обладать граф? Например, насколько большим может быть коэффициент вершинного расширения для графа степени d с n вершинами?

3. Явные примеры экспандеров: для каких «конкретных», «явно описанных» графов можно оценить их эскадренные свойства? Нередко бывает проще показать, что некоторое комбинаторное свойство выполнено для 99% графов с n вершинам, чем доказать это же свойство для какого-то конкретного графа с простым описанием (например, заданного простой алгебрической формулой).

4. Алгоритмически эффективные конструкции: требуется найти экспандер, для которого не просто имеется «явное описание», но который может можно построить с помощью быстрого алгоритма.

Приведённые выше доказательства Теоремы 1 и Теоремы 2 неконструктивны. Эти рассуждения показывают, что экспандеры с заданными параметрами существуют и, более того, *большинство* графов являются такими экспандерами. Однако эти доказательства не дают способа предъявить хотя бы один из экспандеров явно. Разумеется, мы можем перебрать все графы с заданным числом вершин и найти среди них экспандер. Но такой перебор потребует экспоненциального (от числа вершин) времени. Хуже того, даже для одного графа на n вершинах прямая проверка определения экспандера требует экспоненциального перебора (нужно перебрать все подмножества вершин и для каждого из них подсчитать число соседей).

В приложения (в теории сложности вычислений и в теории кодирования) как правило требуются алгоритмически эффективные конструкции экспандеров. При этом эффективность может пониматься в двух разных смыслах.

Конструкции эффективные в слабом смысле: экспандеры с n вершинами, которые можно построить за время $\mathrm{poly}(n)$. В данном случае «построить» граф означает, что мы должны предъявить некоторое стандартное описание этого графа (скажем, матрицу смежности или список всех его рёбер).

Конструкции эффективные в сильном смысле: экспандеры с $N = 2^{\Theta(n)}$ вершинами, простые операции с которыми можно производить за время $\mathrm{poly}(n)$. В таком графе каждая вершина задаётся индексом из $\Theta(n)$ битов; мы требуем, чтобы по индексу вершины можно было найти список всех её соседей (точнее, список *индексов* всех её соседей) за время $\mathrm{poly}(n)$. (Тут стоит напомнить, что мы интересуемся *разреженными* графами, в которых у каждой вершины сравнительно небольшое число соседей. Так что для каждой вершины размер списка её соседей будет очень коротким; трудность лишь в том, чтобы научиться эти списки быстро вычислять.)

Поиск эффективных конструкций экспандеров с параметрами, близкими к оптимальным, является одной из главных задач теории экспандеров. В главах 4 и 5 мы изучим несколько таких конструкций, основанных на разных математических идеях.

2.5 Исторический комментарий

Определение экспандера, неконструктивное доказательство существование экспандеров и первые примеры примеры их применений появилось в начале 1970-х в работах сотрудников московского Института проблем передачи информации Л.А. Бассалыго и М.С. Пинскера, [26, 27]. Другой математик из ИППИ Г.А. Маргулис дал первое конструктивное доказательство существование экспандера [28]. Стоит также отметить, что граф со свойством сильной связности (довольно близким к ставшему стандартным определению экспандера) использовался ещё в 1960-х в работе Барздиня и Колмогорова [25].

Глава 3

Спектральные экспандеры.

В этой главе мы введём определение *спектрального экспандера* и изучим его связь с определением комбинаторного экспандера из главы 2.

3.1 Матрица графа.

Граф с n вершинами описывается *матрицей смежности* M размерности $n \times n$, в которой элемент m_{ij} равно числу рёбер, соединяющих i-ую и j-ую вершины графа. Эта матрица симметрична. Если граф однородный и степень каждой вершины равна d, то сумма чисел в каждой строке и каждом столбце матрицы равна d.

Различные свойства графа удобно описывать в терминах этой матрицы:

- (i, j)-й элемент матрицы M^k есть число путей длины k, идущих из вершины i в вершину j;
- если разделить матрицу M на d, то получится матрица, у которой сумма любой строки и любого столбца равна 1. Умножение на эту матрицу описывает случайное блуждание: если $\mathbf{p} = (p_1, \ldots, p_n)^\perp$ есть вектор-столбец, состоящий из вероятностей, описывающих некоторое распределение на вершинах графа, то вектор $(\frac{1}{d} M\mathbf{p})$ задает распределение через один шаг случайного блуждания (мы выбираем случайную вершину v согласно распределению $\mathbf{p}$ и переходим к её соседу, выбрав случайно одно из d рёбер, выходящих из v).

Последнее наблюдение показывает, что случайное блуждание по графу (из текущей вершины мы равновероятно переходим по одному из рёбер в соседнюю вершину, затем в соседнюю вершину к соседней, и так далее) связано со степенями матрицы M/d: чем ближе эти степени к матрице равномерного перемешивания (в которой все элементы равны $1/n$), тем более равномерно распределен результат случайного блуждания.

Изучать степени матрицы естественно в собственном базисе. Мы увидим, что в терминах собственных чисел матрицы M выражаются многие

комбинаторные свойства графа. Для начала сделаем несколько простых наблюдений:

- матрица M симметрична и потому имеет ортогональный собственный базис над вещественным полем, с вещественными собственными значениями;
- поскольку сумма всех чисел в каждой строке равна d, вектор-столбец $(1, 1, \ldots, 1)^\perp$ является собственным вектором и имеет собственное значение d;
- все собственные значения не превосходят d по модулю: поскольку суммы элементов во всех строках матрицы равны d, то максимум модулей координат собственного вектора при умножении на M увеличивается не более чем в d раз;
- если граф состоит из нескольких компонент связности, то имеется несколько собственных векторов с собственным значением d (для вершин одних компонентах связности берём единицы, для других нули);
- напротив, если граф связен, то собственный вектор со значением d единственный: возьмём максимальную по модулю координату этого вектора (вершину графа), она равна среднему по соседям, и потому во всех соседях должно быть то же значение; то же верно для соседей соседей, и т.д.;
- для двудольного графа имеется собственный вектор со значением $-d$: надо в одной доле взять единицы, а в другой минус единицы;
- если имеется собственный вектор со значением $-d$, то граф имеет двудольную связную компоненту (возьмём максимальную по модулю координату собственного вектора; в её соседях будет то же число с противоположным знаком, и так далее: связная компонента этой вершины делится на две доли).

Упражнение 11 *Найдете все собственный числа полного графа с n вершинами (а) без петель и (б) с петлями.*

Упражнение 12 *Пусть спектр графа G (с n вершинами и m рёбрами, без петель и кратных рёбер) состоит из чисел $\lambda_1, \ldots, \lambda_n$.*

(а) Чему равна сумма $\sum_{i=1}^{n} \lambda_i$?

(б) Чему равна сумма $\sum_{i=1}^{n} \lambda_i^2$?

(в) Выразите через собственные числа графа число треугольников (циклов длины 3) в G.

Упражнение 13 *Докажите, что спектр регулярного графа симметричен относительно нуля тогда и только тогда, когда граф является двудольным.*

Напомним также, что максимальное по модулю собственное число всякой симметричной матрицы M равно максимуму абсолютной величины *отношения Рэлея* по всем ненулевым векторам:

$$|\lambda_1| = \max_{\mathbf{x} \neq 0} \frac{|\mathbf{x} M \mathbf{x}^\perp|}{||\mathbf{x}||^2}.$$

Далее, если $\mathbf{x}_1$ есть собственный вектор, соответствующий собственному значению λ_1, то модуль второго (по абсолютной величине) собственное число матрицы может быть определено как максимум модуля отношения Рэлея по всем векторам, ортогональным $\mathbf{x}_1$:

$$|\lambda_2| = \max_{\mathbf{x} \perp \mathbf{x}_1} \frac{|\mathbf{x} M \mathbf{x}^\perp|}{||\mathbf{x}||^2}.$$

В частности, если первый собственный вектор имеет вид $\mathbf{x}_1 = (1, 1, \ldots, 1)$, то

$$|\lambda_2| = \max_{\mathbf{x} \,:\, x_1 + \ldots + x_n = 0} \frac{|\mathbf{x} M \mathbf{x}^\perp|}{||\mathbf{x}||^2}.$$

Аналогичное утверждение верно и для собственных чисел матрицы, упорядоченных по убыванию (не по абсолютной величине):

Упражнение 14 *Пусть λ_i, $i = 1, \ldots, n$ —собственные числа симметричной матрицы M, расположенные в порядке убывания,*

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n,$$

и $\mathbf{x}_i$ соответствующие им собственные векторы. Докажите, что для $i = 2, \ldots, n$

$$\lambda_i = \max_{\mathbf{x} \perp \mathbf{x}_1, \ldots, \mathbf{x}_{i-1}} \frac{\mathbf{x} M \mathbf{x}^\perp}{||\mathbf{x}||^2}.$$

Упражнение 15 *Если A, B являются симметричными стохастическими матрицами (все матричные элементы неотрицательны, и сумма элементов в каждом столюце равна единице), то $\lambda(A + B) \leqslant \lambda(A) + \lambda(B)$ (где $\lambda(M)$ обозначает второе по абсолютной величине собственное число матрицы).*

3.2 Спектральный зазор и свойство перемешивания

В этой главе мы определим спектральный экспандер как однородный граф с достаточно большим *спектральным зазором*. Мы изучим простейшие свойства спектральных экспандеров и покажем, что спектральный зазор связан с некоторыми комбинаторными свойствами графа (свойством «быстрого перемешивания»).

Определение 5 *Регулярный граф степени d с n вершинами, у которого все собственные числа кроме одного по абсолютной величине не превосходят γd, будем называть* спектральным (n, d, γ)-экспандером.

Если $\gamma = 0$, это означает, что матрица графа является матрицей полного перемешивания (в каждой клетке матрицы стоит элемент $1/n$). Такое возможно лишь в случае $d = n$ (такой матрицей обладает полный граф с петлями). Если же значение γ положительно, но сравнительно мало, это означает, что граф в том или ином смысле обладает свойствами «хорошего перемешивания». Далее мы докажем два утверждения, несколько разным способом формализующие это соображение.

Лемма 1 (о перемешивании – expander mixing lemma) *Для произвольных (возможно, пересекающихся) множества вершин A и B в спектральном (n, d, γ)-экспандере число рёбер, ведущих из A в B, выполняется неравенство*

$$\left| |E(A,B)| - \frac{d \cdot |A| \cdot |B|}{n} \right| \leqslant \gamma d \sqrt{|A| \cdot |B|}.$$

Замечание 1: Леммой о перемешивании удобно пользоваться, когда множества вершин A и B достаточно велики (каждое из них занимает некоторую фиксированную долю среди n вершин графа), а параметр γ очень мал (значительно меньше, чем доли $|A|/n$ и $|B|/n$). Из леммы следует, что число рёбер между парой таких множеств A, B достаточно велико, т.е., спектральный экспандер обладает определённым свойством «сильной связности».

Замечание 2: Лемма о перемешивании и говорит, в частности, что один шаг случайного блуждания в спектральном экспандере (с достаточно малым значением параметра γ) довольно быстро приближает любое начальное распределение к равномерному. Чтобы заметить это, перепишем её утверждение в менее симметричном виде

$$\left| \frac{|E(A,B)|}{d|A|} - \frac{|B|}{n} \right| \leqslant \gamma \sqrt{\frac{|B|}{|A|}}.$$

Для интерпретации этого неравенства удобно рассмотреть равномерное распределения вероятностей на подмножестве вершин A. Мы выбираем случайную вершину из A и делаем один шаг случайного блуждания (переходим по случайно выбранному ребру в соседа данной вершины). Какова вероятность того, что в результате мы попадем в одну из вершин множества B? Мы можем утверждать, что эта вероятность равна $\frac{|E(A,B)|}{d|A|}$. Если бы итоговое распределение оказалось равномерным, то данная вероятность была бы равна доле множества B во всём графе, т.е., $|B|/n$. Хотя на самом деле получаемое распределение и не является равномерным, но вероятность попасть из случайной вершина A в какую-нибудь вершину множества B отличается от «идеальной» вероятности $|B|/n$ ненамного — не более, чем на $\gamma\sqrt{\frac{|B|}{|A|}}$.

Доказательство леммы о перемешивании: Обозначим $\mathbf{1}_A$ и $\mathbf{1}_B$ характеристические векторы множеств A и B (i-ая координата соответствующего вектора равен единице, если i-ая вершина графа принадлежит A или B соответственно; значение координаты равно нулю в противном случае). Заметим, что сумма квадратов координат вектора $\mathbf{1}_A$ есть в точности число элементов в множестве A; аналогично, сумма квадратов координат вектора $\mathbf{1}_B$ есть в точности число элементов в множестве B. Следовательно, евклидова норма этих векторов есть квадратный корень из числа элементов в A и B соответственно,

$$||\mathbf{1}_A||_2 = \sqrt{|A|},\ ||\mathbf{1}_B||_2 = \sqrt{|B|}.$$

Если M — матрица графа, то число рёбер, ведущих из A в B равно

$$|E(A,B)| = \mathbf{1}_A M \mathbf{1}_B^{\perp} \tag{3.1}$$

Мы должны оценить эту величину, используя определение спектрального экспандера.

Пусть $\mathbf{e}_1, \ldots, \mathbf{e}_n$ — ортонормированный собственный базис матрицы M заданного графа, а $\lambda_1, \ldots, \lambda_n$ – соответствующие собственные числа. Будем считать, что собственные числа упорядочены по убыванию абсолютной величины:

$$|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_n|.$$

При этом

$$\mathbf{e}_1 = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1),$$

а $\lambda_1 = d$, и $|\lambda_i| \leqslant \gamma d$ для $i > 1$. Разложим векторы $\mathbf{1}_A$ и $\mathbf{1}_B$ по собственному базису: $\mathbf{1}_A = \sum a_i \mathbf{e}_i$, $\mathbf{1}_B = \sum b_i \mathbf{e}_i$. Получаем

$$|E(A,B)| = \mathbf{1}_A M \mathbf{1}_B^{\perp} = (\sum_{i=1}^{n} a_i \mathbf{e}_i) M (\sum_{i=1}^{n} b_i \mathbf{e}_i)^{\perp}$$

Выделим первый член из суммы (3.1):

$$|E(A,B)| = d\frac{|A|}{\sqrt{n}} \cdot \frac{|B|}{\sqrt{n}} + \sum_{i=2}^{n} \lambda_i a_i b_i$$

Остается оценить сумму всех остальных членов этой суммы.

$$\begin{aligned}\left||E(A,B)| - \frac{d \cdot |A| \cdot |B|}{n}\right| &= |\sum_{i=2}^{n} \lambda_i a_i b_i| \leqslant \gamma d |\sum_{i=1}^{n} a_i b_i| \\ &\leqslant \gamma d \cdot ||\mathbf{1}_A||_2 \cdot ||\mathbf{1}_B||_2 = \gamma d \cdot \sqrt{|A||B|},\end{aligned}$$

и лемма доказана.

В следующем утверждение мы с другой стороны посмотрим на замечание 2 к лемме о перемешивании (стр. 18). Мы снова начнем с равномерного

распределения на подмножестве вершинах графа A и сделаем один шаг случайного блуждания. Мы покажим, что получающееся в результате распределение вероятностей на вершинах графа довольно близко к равномерному. Точнее, мы оценим L_2-норму «погрешности» — разности между полученным нами распределением и настоящим равномерным распределением на графе.

Утверждение 2 *В спектральном (n, d, γ)-экспандере для любого множества вершин A выполняется неравенство*

$$\sum_v \left(|E(v, A)| - \frac{d|A|}{n} \right)^2 \leqslant (\gamma d)^2 \frac{|A|(n - |A|)}{n}$$

(*здесь $|E(v, A)|$ обозначает число рёбер, ведущих из вершины v во множество A; сумма по всем вершинам графа*).

Доказательство: Обозначим $a = \frac{|A|}{n}$ (доля, которую множество A занимает среди всех вершин графа). Заметим, что выражение в левой части доказываемого неравенства есть квадрат нормы вектора

$$\begin{pmatrix} |E(v_1, A)| - da \\ |E(v_2, A)| - da \\ \cdots \\ |E(v_n, A)| - da \end{pmatrix} = \begin{pmatrix} |E(v_1, A)| \\ |E(v_2, A)| \\ \cdots \\ |E(v_n, A)| \end{pmatrix} - d \cdot \begin{pmatrix} a \\ a \\ \cdots \\ a \end{pmatrix}.$$

В этой разности уменьшаемое

$$\mathbf{x} = \begin{pmatrix} |E(v_1, A)| \\ |E(v_2, A)| \\ \vdots \\ |E(v_n, A)| \end{pmatrix},$$

есть вектор, в котором i-ая координата равна числу рёбер, ведущих из i-ой вершины графа в множество A, а вычитаемое

$$\mathbf{y} = d \cdot \begin{pmatrix} a \\ a \\ \vdots \\ a \end{pmatrix}$$

есть проекция $\mathbf{x}$ на направление $\mathbf{1} = (1, \ldots, 1)^\perp$. В самом деле, в разности $\mathbf{x} - \mathbf{y}$ сумма координат равна нулю; это значит, что $\mathbf{x} - \mathbf{y}$ ортогонален $\mathbf{1}$.

Таким образом, мы хотим оценить квадрат нормы вектора $\mathbf{x} - \mathbf{y}$, принадлежащего подпространству векторов с нулевой суммой координат. В этом подпространстве все собственные числа матрицы графа M по модулю не превосходят γd, так что при умножении на M норма вектора увеличивается не более, чем в (γd) раз.

Обозначим $\mathbf{1}_A$ вектор из $\{0,1\}^n$, у которого единицы стоят в позициях, соответствующих элементам множества A и нули во всех остальных позициях. Тогда $\mathbf{x} = M \cdot \mathbf{1}_A$ и $\mathbf{y} = M \cdot (a \cdot \mathbf{1}_A)$. Таким образом,

$$\mathbf{x} - \mathbf{y} = M \cdot \mathbf{1}_A - M \cdot (a \cdot \mathbf{1}_A) = M \cdot (\mathbf{1}_A - (a \cdot \mathbf{1}_A)),$$

и

$$||M \cdot (\mathbf{1}_A - (a \cdot \mathbf{1}_A))||^2 \leqslant (\gamma d)^2 \cdot ||\mathbf{1}_A - (a \cdot \mathbf{1}_A)||^2.$$

Наконец, нетрудно подсчитать квадрат нормы $\mathbf{1}_A - a \cdot \mathbf{1}_A$; в этом векторе в an координатах стоит число $1-a$ и в оставшихся $(1-a)n$ координатах стоит число $-a$. Поэтому $||\mathbf{1}_A - a\mathbf{1}_A||^2 = a(1-a)^2 n + (1-a)a^2 n = a(1-a)n$. В итоге мы получаем

$$\left\| \begin{pmatrix} |E(v_1, A)| - da \\ |E(v_2, A)| - da \\ \vdots \\ |E(v_n, A)| - da \end{pmatrix} \right\|^2 \leqslant (\gamma d)^2 a(1-a)n.$$

Упражнение 16 *Выведите* Лемму о перемешивании *из Утверждения 2.*

Упражнение 17 *Вершины спектрального (n, d, γ)-экспандера нельзя раскрасить менее чем в $1/\gamma$ цветов так, чтобы никакие смежные вершины не были покрашены в один цвет.*

Упражнение 18 *Пусть граф G является спектральным (n, d, γ)-экспандером, целое число $k \leqslant 1/\gamma$ является делителем n, и вершины графа раскрашены в k цветов так, что каждый из цветов использован ровно для n/k вершин. Докажите, что найдётся хотя бы одна вершина, среди соседей которой встречаются все k цветов.*

3.3 Лапласиан графа

Пусть $G = (V, E)$ — граф с n вершинами (не обязательно однородный) и $\mathbf{x} = (x_1, \ldots, x_n)$ распределение весов на вершинах графа. Рассмотрим меру «неоднородности» данного распределения:

$$\mathrm{Lap}(\mathbf{x}) = \sum_{\{u,v\} \in E} (x_u - x_v)^2$$

(для каждого рёбра мы берём квадрат разности весов, сопоставленных его весам). Эта функция $\mathbf{x}$ называется *лапласианом* графа. Лапласиан графа с n вершинами определен для любого $\mathbf{x} \in \mathbb{R}^n$.

Лапласиан – это квадратичная форма, матрицу которую легко описать:

Утверждение 3 *Для всякого графа G*

$$\mathrm{Lap}(\mathbf{x}) = \mathbf{x}(\mathrm{Deg}(G) - M)\mathbf{x}^{\perp},$$

где M — матрица смежности графа, а $\mathrm{Deg}(G)$ — матрица степеней графа (в i-ой клетке диагонали стоит степень i-ой вершины графа, а во всех клетках вне диагонали стоят нули). В частности, для однородного графа, в котором все вершины имеют степень d,

$$\mathrm{Lap}(\mathbf{x}) = \mathbf{x}(d \cdot I - M)\mathbf{x}^{\perp},$$

где I — единичная матрица размера $n \times n$.

Доказательство: Прежде всего заметим, что добавление или удаление петель не меняет ни величину $Lap(\mathbf{x})$, ни значение $\mathbf{x}^{\perp}(Deg - M)\mathbf{x}$. Так что без ограничение общности можно считать, что в графе G петель нет. А для графов без петель нетрудно зметить, что

$$\mathrm{Lap}(\mathbf{x}) = \sum_{\{u,v\}\in E} (x_u^2 + x_v^2 - 2x_u x_v) = \sum_{v\in V} (deg(v) \cdot x_v^2) - 2 \sum_{\{u,v\}\in E} x_u x_v = \\ = \mathbf{x} \cdot Deg \cdot \mathbf{x}^{\perp} - \mathbf{x} \cdot M \cdot \mathbf{x}^{\perp}.$$

Утверждение доказано.

Далее мы будем предполагать, что граф является однородным и каждая вершина имеет степень d. Если спектр матрицы графа M состоит из чисел

$$d = \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n,$$

то спектр лапласиана состоит из чисел

$$d - \lambda_n \geqslant d - \lambda_{n-1} \geqslant \ldots \geqslant d - \lambda_1 = 0.$$

Мы уже знаем, что спектр M лежит в интервале $[-d, d]$, причём (i) кратность собственного числа d равна числу компонент связности, и (ii) кратность собственного числа $-d$ равна числу двудольных компонент связности. Следовательно, собственные числа лапласиана лежат в интервале $[0, 2d]$, причём кратность собственного числа 0 равна числу компопнет связности, а кратность собственного числа 2 есть число двудольных компонент связности.

Эти свойства спектра можно усмотреть непосредственно из определения лапласиана. Во-первых, ясно, что $\mathrm{Lap}(\mathbf{x})$ является неотрицательно определенной функцией; следовательно, все собственные числа неотрицательны. Далее, $\mathrm{Lap}(x_1, \ldots, x_n) = 0$, если и только если внутри каждой компоненты связности значения x_i постоянны (для каждого ребра значения, сопоставленные его концам, одинаковы). Наконец,

$$\mathrm{Lap}(\mathbf{x}) = \sum_{\{u,v\}\in E} (x_u - x_v)^2 \leqslant \sum_{\{u,v\}\in E} (x_u^2 + x_v^2) = 2d \sum_{v\in E} x_v^2.$$

Это значит, что каждое собственное число матрицы лапласиана $L := d \cdot I - M$ не превосходит $2d$; причем собственный вектор $\mathbf{x}$ соответствует собственному значению $2d$, если и только если для каждого ребра (i, j) соответствующие значения x_i и x_j противоположны. Размерность подпространства таких $\mathbf{x}$ равна числу двудольных компонент графа.

Упражнение 19 *Обозначим L лапласиан d-регулярного графа $G = (V, E)$, где число вершин $n = |V|$. Обозначим $0 = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$ собственные числа лапласиана. Пусть $\lambda_2 = 0$, а $\lambda_{10} > 0$. Обозначим $\mathbf{e}_2$ собственный вектор, соответствующий собственному числу λ_2. Докажите, что среди координат $\mathbf{e}_2$ встречается не более 9 различных значений.*

3.4 От спектрального экспандера к комбинаторному

В этой главе мы изучим связь между спектральным и комбинаторным определениями экспандера. Мы покажем, что всякий спектральный экспандер является однородным комбинаторным экспандером (и чем больше зазор между первым и вторым собственным числом у спектрального экспандера, тем более сильные свойства рёберного и вершинного расширения мы можем для гарантировать для этого графа).

Теорема 3 *Пусть граф G содержит n вершин, степень каждой вершины равна d и спектр матрицы графа состоит из чисел*

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n.$$

Тогда для любого множества вершин S (непустого и не совпадающего со множеством всех вершин) $\frac{|E(S,\bar{S})|}{\frac{1}{n} \cdot |S| \cdot |\bar{S}|} \geqslant d - \lambda_2$.

Замечание 1: В данном случае λ_2 – это второе в порядке убывания (не по абсолютной величине!) собственное значение графа.

Замечание 2: Если $\lambda_2 = d$, то в графе больше одной компоненты связности. Выбрав в качестве множества S одну из компонент связности, мы получим $|E(S, \bar{S})| = 0$ и $|\Gamma(S)| \leqslant |S|$. Так что в графе с нулевым зазором между первым и вторым собственным числом коэффициенты рёберного и вершинного расширения также равны нулю.

Из Теоремы 3 немедленно получаем связь спектрального и комбинаторного определения экспандера:

Следствие 1 (спектральный зазор $\Longrightarrow$ рёберное расширение) *Для всякого спектрального (n, d, γ)-экспандера*

$$\min_{|S| \leqslant n/2} \frac{|E(S, \bar{S})|}{|S|} \geqslant \frac{d(1-\gamma)}{2},$$

так что $h_E(G) \geqslant \frac{1-\gamma}{2}$.

Следствие 2 (спектральный зазор $\Longrightarrow$ вершинное расширение) *Для всякого спектрального (n,d,γ)-экспандера*

$$\min_{|S|\leqslant n/2}\frac{|\Gamma(S)\backslash S|}{|S|}\geqslant\frac{(1-\gamma)}{2},$$

m.e., $h_V(G)\geqslant\frac{1-\gamma}{2}$.

Следствие 3 (спектральный зазор $\Longrightarrow$ однородный экспандер) *Если в спектральном (n,d,γ)-экспандере без петель в каждой вершине добавить петлю, мы получим однородный комбинаторный $(n,d+1,\frac{1-\gamma}{2})$-экспандер.*

Первое доказательство теоремы о рёберном расширении: Пусть S – некоторое множество вершин графа (непустое и не совпадающие с множеством всех вершин). Обозначим $\mathbf{x}\in\{0,1\}^n$ характеристическую функцию этого множества ($x_i=1$, если i-ая вершина графа принадлежит S, и $x_i=0$ иначе). Тогда

$$\frac{|E(S,\bar{S})|}{|S||\bar{S}|}=\frac{\sum\limits_{\{u,v\}\in E}|x_u-x_v|}{\sum\limits_{\{u,v\}}|x_u-x_v|}$$

(в числителе сумма берётся по всем неупорядоченным парам точек, соединённых ребром, а в знаменателе по произвольным неупорядоченным парам вершин). Поскольку каждое значение $|x_i-x_j|$ есть ноль или единица, данное выражение можно переписать в виде

$$\frac{|E(A,\bar{A})|}{|A||\bar{A}|}=\frac{\sum\limits_{\{u,v\}\in E}(x_u-x_v)^2}{\sum\limits_{\{u,v\}}(x_u-x_v)^2}=\frac{\mathbf{x}^\perp L\mathbf{x}}{\sum\limits_{\{u,v\}}(x_u-x_v)^2},\tag{3.2}$$

где L есть матица лапласиана (см. предыдущий параграф).

Напомним, что второе (снизу) собственное число лапласиана (равное $d-\lambda_2$) равно минимуму отношения Рэлея по всем векторам, ортогональным первому собственному вектору $\mathbf{1}=(1,\ldots,1)$:

$$d-\lambda_2=\min_{\mathbf{y}\perp\mathbf{1}}\frac{\mathbf{y}L\mathbf{y}^\perp}{||\mathbf{y}||^2}=\min_{\mathbf{y}\perp\mathbf{1}}\frac{\mathbf{y}L\mathbf{y}^\perp}{\sum\limits_v y_v^2}\tag{3.3}$$

Выражения (3.2) и (3.3) выглядят похожими. Чтобы сделать аналогию между этими отношениями более очевидной, преобразуем знаменатель (3.3). При условии $\mathbf{y}\perp\mathbf{1}$ (т.е., сумма координат $\mathbf{y}$ равна нулю) имеем

$$\sum_{\{u,v\}}(y_u-y_v)^2=\sum_{\{u,v\}}(y_u^2+y_v^2-2y_uy_v)=2n\sum_v y_v^2-(\sum_v y_v)^2=2n\sum_v y_v^2.$$

Таким образом, (3.3) можно переписать в виде

$$d-\lambda_2=\min_{\mathbf{y}\perp\mathbf{1}}\frac{\mathbf{y}^\perp L\mathbf{y}}{\frac{1}{2n}\cdot\sum\limits_{\{u,v\}}(y_u-y_v)^2}\tag{3.4}$$

Заметим также, что сдвигая вектор $\mathbf{y}$ (прибавляя одно и то же число к каждой компоненте вектора), мы не изменим ни числитель, ни знаменатель (3.4). Следовательно, можно переписать (3.4) в виде

$$d - \lambda_2 = \min_{\mathbf{y} \nparallel \mathbf{1}} \frac{\mathbf{y}^\perp L \mathbf{y}}{\frac{1}{2n} \cdot \sum\limits_{\{u,v\}} (y_u - y_v)^2} \tag{3.5}$$

(условие $\mathbf{y} \nparallel \mathbf{1}$ означает, что знаменатель (3.5) не обращается в ноль). Сравнивая (3.2) и (3.5), заключаем, что

$$d - \lambda_2 \leqslant \frac{|E(A, \bar{A})|}{\frac{1}{n} \cdot |A||\bar{A}|}$$

для любого множеств вершин A, для которого правая часть неравенства имеет смысл (т.е., A и $\bar{A}$ непусты). Теорема доказана.

Второе доказательство теоремы о рёберном расширении: Пусть S есть множество вершин графа (размера не более $n/2$). Обозначим $\mathbf{1}_S$ и $\mathbf{1}_{\bar{S}}$ характеристические векторы самого множества S и его дополнения. Рассмотрим вектор

$$\mathbf{f} = |\bar{S}|\mathbf{1}_S - |S|\mathbf{1}_{\bar{S}}$$

сумма координат которого равна нулю. Его норма

$$||\mathbf{f}||^2 = |\bar{S}|^2 \cdot |S| + |S|^2 \cdot |\bar{S}| = |S| \cdot |\bar{S}| \cdot n$$

Далее мы подсчитаем значение $\mathbf{f}M\mathbf{f}^\perp = \sum_{i,j} m_{ij} f_i f_j$. Рассмотрим вклад каждого ребра графа в эту сумму. Если ребро не является петлёй (ребро с концами (i, j), где $i \neq j$), то его вклад состоит из учтённого дважды произведения $f_i f_j$, что равняется

- $2|\bar{S}|^2$, если оба конца ребра лежат в S,
- $2|S|^2$, если оба конца ребра лежат в $\bar{S}$,
- $2|S| \cdot |\bar{S}|$ со знаком минус, если один конец ребра лежит в S, а другой в $\bar{S}$.

Если же концы рёбра совпадают (ребро является петлей с концами (i, i)), то его вклад в сумму вклад сумму $\sum_{i,j} m_{ij} f_i f_j$ состоит из единственного члена f_i^2. Это число равно

- $|\bar{S}|^2$, если i-ая вершина лежит в S,
- $|S|^2$, если i-ая вершина лежит в $\bar{S}$.

Таким образом, сумма $\mathbf{f}M\mathbf{f}^\perp = \sum\limits_{i,j} m_{ij} f_i f_j$ представляется в следующем виде:

$$\mathbf{f}M\mathbf{f}^\perp = 2|E(S, S)| \cdot |\bar{S}|^2 + 2|E(\bar{S}, \bar{S})| \cdot |S|^2 - 2|E(S, \bar{S})| \cdot |S| \cdot |\bar{S}|.$$

Поскольку из каждой вершины выходит по d рёбер, величину $2|E(S,S)|$ можно заменить на $d|S| - |E(S,\bar{S})|$, а величину $2|E(\bar{S},\bar{S})|$ можно заменить на $d|\bar{S}| - |E(S,\bar{S})|$. Теперь нетрудно подсчитать, что

$$\mathbf{f}M\mathbf{f}^{\perp} = dn|S| \cdot |\bar{S}| - |E(S,\bar{S})|n^2.$$

Разложим вектор $\mathbf{f}$ по векторам ортонормированного собственного базиса матрицы графа:

$$\mathbf{f} = (\mathbf{f},\mathbf{e}_2) \cdot \mathbf{e}_2 + \ldots + (\mathbf{f},\mathbf{e}_n) \cdot \mathbf{e}_n$$

(первый коэффициент в разложении $\mathbf{f}$ по собственному базису равен нулю, поскольку $\mathbf{f}$ ортогонален $\mathbf{e}_1 = \frac{1}{\sqrt{n}}(1,\ldots,1)$). Заметим, что

$$\mathbf{f}M\mathbf{f}^{\perp} = \sum_{i \geqslant 2} \lambda_i f_i^2 \leqslant \lambda_2 ||\mathbf{f}||^2.$$

Сравнивая два полученных выражения для $\mathbf{f}M\mathbf{f}^{\perp}$, получаем

$$dn|S| \cdot |\bar{S}| - n^2 \cdot |E(S,\bar{S})| \leqslant \lambda_2 \cdot n|S| \cdot |\bar{S}|,$$

что и требовалось доказать.

3.5 От комбинаторного экспандера к спектральному*

Теорема 4 *Пусть G является однородным графом степени d с n вершинами, и спектр этого графа состоит из чисел*

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n.$$

Тогда $\min\limits_{|A| \leqslant n/2} \frac{|E(A,\bar{A})|}{|A|} \leqslant \sqrt{2d(d-\lambda_2)}$.

Замечание: Для наглядности можно соединить оценки для коэффициента рёберного расширения из Следствия 2 и из Теоремы 4:

$$\frac{d-\lambda_2}{2} \leqslant h_E(G) \leqslant \sqrt{2d(d-\lambda_2)}.$$

Доказательство: Обозначим $\mathbf{e}_i$ собственные векторы матрицы графа, соответствующие собственным числам λ_i. Как обычно, мы можем считать, что эти векторы попарно ортогональны, причём $\mathbf{e}_1 = (1,\ldots,1)$.

Нас будет интересовать второй собственный вектор $\mathbf{e}_2$. Поскольку он ортогонален $\mathbf{e}_1$, сумма координат $\mathbf{e}_2$ равна нулю.

Мы хотим показать, что если зазор между d и λ_2 мал, то в графе найдется множество со сравнительно малым коэффициентом рёберного расширения. Другими словами, в графе есть сравнительно небольшой *разрез*: множество вершин графа можно так разделить на две части $V = A \cup \overline{A}$, что

число рёбер, соединяющих A и $\overline{A}$ не превосходит $|A| \cdot \sqrt{2d(d-\lambda_2)}$. Такой разрез мы построим с помощью собственного вектора $\mathbf{e}_2$.

Без ограничения общности будем считать, что не более половины координат вектора $\mathbf{e}_2$ неотрицательны. Будем также считать, что вершины графа v_i пронумерованы так им образом, что координаты $\mathbf{e}_2$ идут в порядке невозрастания:

$$(\mathbf{e}_2)_1 \geqslant (\mathbf{e}_2)_2 \geqslant \ldots \geqslant (\mathbf{e}_2)_n .$$

Заменим отрицательные координаты этого вектора на нули, а положительные (которых, напомним, не больше $n/2$) оставим прежними. Обозначим полученный вектор $\mathbf{f} = (f_1, \ldots, f_n)$. Формально координаты нового вектора определены по правилу

$$f_i := \max\{0, (\mathbf{e}_2)_i\}.$$

В дальнейшем мы будем изучать две меры «неоднородности» вектора $\mathbf{f}$. Первая из них — это уже знакомый нам лапласиан вектора $\mathbf{f} = (f_1, \ldots, f_n)$,

$$Lap(\mathbf{f}) = \sum_{\{i,j\}\in E} |f_i - f_j|^2$$

(см. с. 21). Вторая мера неоднородности похожа на лапласиан, но вместо суммы *квадратов разностей* координат вектора f (для всех пар координат, соответствующих вершинам, соединённым ребром) мы просуммируем абсолютные величины *разностей квадратов* координат вектора f (также для каждой пары вершин, соединённых ребром). Эту величину мы обозначим

$$\mathcal{V}(\mathbf{f}) := \sum_{\{i,j\}\in E} |f_i^2 - f_j^2|. \tag{3.6}$$

Сначала мы оценим коэффициента рёберного расширения графа через значение $\mathcal{V}(\mathbf{f})$, а затем покажем, как $\mathcal{V}(\mathbf{f})$ связано со спектральным зазором.

Лемма 2 $\mathcal{V}(\mathbf{f}) \geqslant h_E(G) \cdot ||\mathbf{f}||^2$.

Доказательство леммы: Напомним, что мы считаем вершины графа v_i пронумерованными таким образом, что

$$f_1 \geqslant f_2 \geqslant \ldots \geqslant f_n.$$

Это соглашение позволяет нам избавиться от модуля в определение (3.6):

$$\begin{aligned}\mathcal{V}(\mathbf{f}) = \sum_{\{i,j\}\in E, i<j} f_i^2 - f_j^2 = \sum_{k=i}^{j-1}(f_k^2 - f_{k+1}^2) = \\ = \sum_{k=1}^{n-1}(f_k^2 - f_{k+1}^2)\cdot |E(\{v_1,\ldots,v_k\},\{v_{k+1},\ldots,v_n\})| \\ = \sum_{k\ :\ f_i>0}(f_k^2 - f_{k+1}^2)\cdot |E(\{v_1,\ldots,v_k\},\{v_{k+1},\ldots,v_n\})| \end{aligned} \tag{3.7}$$

Напомним, что у вектора $\mathbf{f}$ не более половины координат не равны нулю. Это значит, что в сумме (3.7) ненулевой вклад дают только слагаемые для $k \leqslant n/2$.

По определению коэффициента рёберного расширения для всех $k \leqslant n/2$ мы можем оценить множитель $|E(\{v_1, \ldots, v_k\}, \{v_{k+1}, \ldots, v_n\})|$ (количество рёбер, у которых один конец имеет номер не больше k, а другой — строго больше k) числом $h_E(G) \cdot k$. Получаем

$$\mathcal{V}(\mathbf{f}) = \sum_{\{i,j\}\in E, i<j} f_i^2 - f_j^2 \geqslant h_E(G) \sum_{k=1}^{n-1} (f_k^2 - f_{k+1}^2) \cdot k = \\ = h_E(G) \sum_{f_i>0} f_i^2 = h_E(G) \cdot ||\mathbf{f}||^2,$$

и лемма доказана.

Замечание: Из доказательства Леммы 2 видно, что если значение $\mathcal{V}(\mathbf{f})$ достаточно мало, то среди разбиений графа вида

$$V = \{v_1, \ldots, v_k\} \sqcup \{v_{k+1}, \ldots, v_n\}$$

найдётся хотя бы один, пересекающий сравнительно мало рёбер.

Лемма 2 позволяет нам оценить рёберное расширение графа с помощью $\mathcal{V}(\mathbf{f})$. Теперь мы покажем, как значение $\mathcal{V}(\mathbf{f})$ связано с более стандартной величной $Lap(\mathbf{f})$.

Лемма 3 $\mathcal{V}(\mathbf{f}) \leqslant \sqrt{2d} \cdot \sqrt{Lap(\mathbf{f})} \cdot ||\mathbf{f}||$.

Доказательство леммы: Применим неравенство Коши–Буняковского:

$$\sum_{\{i,j\}\in E} |f_i^2 - f_j^2| \leqslant \sum_{\{i,j\}\in E} |f_i - f_j| \cdot |f_i + f_j| \leqslant \sqrt{\sum_{\{i,j\}\in E} (f_i - f_j)^2} \cdot \sqrt{\sum_{\{i,j\}\in E} (f_i + f_j)^2}$$

Первый из двух сомножителей, полученных в правой части, равен корню из $Lap(\mathbf{f})$. А второй можно оценить как

$$\sqrt{\sum_{\{i,j\}\in E} (f_i + f_j)^2} \leqslant \sqrt{2 \sum_{\{i,j\}\in E} (f_i^2 + f_j^2)} = \sqrt{2d \cdot ||\mathbf{f}||},$$

и лемма доказана.

Остаётся связать лапласиан $\mathbf{f}$ и спектральный зазор графа.

Лемма 4 *Для лапласиана вектора* $\mathbf{f}$ *выполнено неравенство*

$$Lap(\mathbf{f}) \leqslant (d - \lambda_2)||\mathbf{f}||^2.$$

Доказательство леммы: Для самого вектора $\mathbf{f}$ лапласиан вычислить трудно, но мы знаем значение лапласиана на каждом из собственных векторов графа. В частности, из Утверждения 3 мы немедленно получаем

$$L\mathbf{e}_2^\perp = (d \cdot I - M)\mathbf{e}_2^\perp = (d - \lambda_2)\mathbf{e}_2^\perp$$

(где L обозначает матрицу лапласиана).

Теперь сравним $L\mathbf{e}_2^{\perp}$ и $L\mathbf{f}^{\perp}$. Заметим, что для координат i, в которых $f_i = (\mathbf{e}_2)_i > 0$, значение $(L\mathbf{f}^{\perp})_i$ может быть только меньше, чем $(L\mathbf{e}_2^{\perp})_i$. Следовательно,

$$\left((d \cdot I - M)\mathbf{f}^{\perp}\right)_i \leqslant (d - \lambda) f_i.$$

С другой стороны, координаты, в которых $f_i = 0$, не дают никакого вклада в сумму $\sum f_i \cdot (L\mathbf{f})_i^{\perp}$. Таким образом,

$$Lap(\mathbf{f}) = \sum_{i=1}^{n} f_i \cdot (L\mathbf{f})_i^{\perp} = \sum_{i\ :\ f_i > 0} (\mathbf{f})_i \cdot \left((d \cdot I - M)\mathbf{f}^{\perp}\right)_i \leqslant (d - \lambda_2)||\mathbf{f}||^2,$$

и лемма доказана.

Соединяя утверждения Лемм 2, 3 и 4, мы получаем утверждение теоремы.

3.6 Оценка снизу для спектрального зазора

Мы уже знаем, что чем больше зазор между первым и вторым собственным числом, тем более сильные свойства перемешивания и расширения мы можем гарантировать для такого графа. Возникает вопрос: насколько большим можно сделать этот зазор? В этой главе мы установим границы возможного — мы покажем, что спектральный зазор невозможно сделать слишком большим. Другими словами, в d-регулярном графе модуль второго собственного числа не может быть слишком маленьким. Лучшее, на что мы можем надеяться — это второе собственное число примерно равное $2\sqrt{d-1}$. Сформулируем это утверждение более точно.

Теорема 5 *Для любого числа d, в d-регулярных графах с n вершинами второе по абсолютной величине собственное число ограничено снизу: если спектр графа состоит из чисел $|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_n|$, то*

$$|\lambda_2| \geqslant 2\sqrt{d-1} - o(1)$$

при $n \to \infty$.

Доказательство: Обозначим $\lambda(G)$ модуль второго по абсолютное величине собственное число d-регулярного графа G. Чтобы вычислить $\lambda(G)$, рассмотрим степень графа (l-ая степень графа G есть граф с тем же множеством вершин; ребрами же становятся пути длины l в исходном графе). Матрица l-ой степени графа есть l-ая степень матрицы исходного графа; собственные числа при этом тоже возводятся в l-ую степень.

Мы рассмотрим некоторую чётную степень графа $l = 2k$. Для того, чтобы оценить второе собственное значение симметричной матрицы, нужно ограничить соответствующую ей квадратичную форму на ортогональное

дополнение к первому собственному вектору $\mathbf{e} = (1, \ldots, 1)$ и взять её максимум на единичном шаре. Таким образом,

$$\lambda(G)^{2k} = \lambda(G^{2k}) \geqslant \frac{\mathbf{f} M^{2k} \mathbf{f}^{\perp}}{||\mathbf{f}||^2}$$

для любого вектора $\mathbf{f}$ с нулевой суммой координат. В качестве $\mathbf{f}$ мы берём вектор вида

$$\mathbf{f} = (0, 0, \ldots, 0, 1, 0, \ldots, 0, -1, 0, \ldots)$$

У этого вектора ровно две ненулевые координаты (i-ая и j-ая). Вершины i и j мы выбираем так, чтобы расстояние между ними было максимальным возможным (т.е., равно диаметру графа G).

Для выбранного вектора $\mathbf{f}$ имеем

$$\begin{aligned} \mathbf{f} M^{2k} \mathbf{f}^{\perp} \quad = \quad & [\text{число } (2k)\text{-путей из } i \text{ в } i] + \\ & + [\text{число } (2k)\text{-путей из } j \text{ в } j] - 2 \cdot [\text{число } (2k)\text{-путей из } i \text{ в } j] \end{aligned}$$

Теперь выбираем $k = \lfloor \frac{diameter(G)-1}{2} \rfloor$; поскольку расстояние между i и j больше $2k$, число $(2k)$-путей из i в j равно нулю. Остаётся оценить число циклов с началами и концами в i и в j. Мы оценим число циклов в графе G снизу через число циклов в дереве степени d. Такие циклы соответствуют циклам в G, которые можно «стянуть» в вершину по рёбрам графа. Таким образом,

$$\lambda(G)^{2k} \geqslant \left[\begin{array}{l} \text{число путей длины } 2k \text{ с началом и} \\ \text{концом в корне дерева степени } d \end{array} \right].$$

Теперь нам нужно подсчитать число циклов длины $2k$ в дереве степени d (с фиксированным началом и концом). В таком цикле $2k$ рёбер делятся на шаги, на которых мы удаляемся от корня, и шаги, на которых мы приближаемся к корню; каждый раз, когда мы делаем шаг в сторону от корня, мы выбираем одно ребро из $(d-1)$; когда мы делаем шаг по направлению к корню, у нас есть единственная возможность. Число способов разделить $2k$ шагов на шаги, на которых мы приближаемся к корню, и шаги, на которых мы удаляемся от корня (число правильных скобочных структур, составленных из k пар скобок) равно k-ому числу Каталана C_k. Таким образом, число интересующих нас циклов не меньше

$$C_k \cdot (d-1)^k = \frac{C_{2k}^k}{k+1} \cdot (d-1)^k = \frac{2^{2k}}{\mathrm{poly}(k)} \cdot (d-1)^k.$$

Следовательно,

$$\lambda(G) \geqslant 2\sqrt{d-1} \left(\frac{1}{\mathrm{poly}(k)} \right)^{1/2k}.$$

Остаётся заметить, что k (диаметр графа, деленный пополам) стремится к бесконечности при росте числа вершин графа n. Так что выражение $\left(\frac{1}{\mathrm{poly}(k)} \right)^{1/2k}$ можно заменить на $(1 + o(1))$.

3.7 Спектральные экспандеры: теорема о существовании

Мы изучили разные свойства спектральных экспандеров, однако до сих пор не задавались вопросом о существовании таких графов. Напомним, что мы хотели бы иметь графы, у которых второе по абсолютной величине собственное число мало по сравнению с d. В этой главе мы докажем утверждение о существовании таких графов, хотя и не в самой сильной форме. (Теоремы о существовании, которую мы докажем, достаточно для большинства приложений). В следующем параграфе мы обсудим более сильный вариант этого утверждения, требующий более сложного доказательства.

Теорема 6 *Пусть $\gamma > 0$ — произвольное число. Тогда для достаточно больших d существует граф с $n = d^4$ вершинами степени d, у которого все собственные числа, кроме первого d, не превосходят по модулю γd.*

Доказательство: Мы докажем, что при определённом соотношении между числом вершин и числом рёбер *почти все* однородные графы обладают таким свойством. Слова «почти все» здесь означают, что при некотором естественном распределении вероятностей случайно выбранный граф оказывается спектральным экспандером (с нужными нам параметрами) с вероятностью близкой к единице.

Прежде всего опишем распределение вероятностей, которое мы будем использовать. Оно будет отличаться от распределения, использованного в доказательстве Теоремы 1. Мы будем считать, что n (число вершин) чётно. Если n чётно, мы имеем право рассмотреть на n вершинах *совершенные паросочетания*. (Совершенное паросочетание есть такой набор из $n/2$ рёбер, что каждая из n вершин является концом ровно одного ребра. Другими словами, совершенное паросочетание на n вершинах есть граф степени 1, состоящий из $n/2$ рёбер.) Мы выбираем на n вершинах d случайных паросочетаний $P_1, \ldots, P_d$ (каждое из d паросочетантий выбирается равномерно; все d выборов делаются независимо). Объединением выбранных паросочетаний мы и будем считать графом G. Отметим, что в таком графе не может быт петель, однако могут быть кратные рёбра (поскольку одно и то же ребро может входить в несколько паросочетаний).

Мы обозначаем собственные числа полученного графа λ_i и считаем, что

$$d = |\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_n|.$$

Теперь нам нужно оценить λ_2 — второе (по абсолютной величине) собственное число этого графа. При возведении матрицы в степень (мы выберем десятую степень) все собственные числа возводятся в ту же степень и след матрицы станет равным

$$\lambda_1^{10} + \lambda_2^{10} + \ldots + \lambda_n^{10}.$$

Первое слагаемое равно d^{10}; если для какой-то матрицы вся сумма близка к d^{10}, то все слагаемые кроме первого малы. А существование такой матрицы будет доказано, если мы убедимся что *среднее* значение следа матрицы

M^{10} (для матрицы графа, выбранного случайно описанным выше способом) близко к d^{10}.

В нашем распределении вероятностей все вершины графа равноправны. След M^{10} равен сумме диагональных элементов, поэтому его среднее значение равно среднему значению одного элемента, умноженному на n. А среднее значение одного элемента равно среднему числу путей длины 10, начинающихся и заканчивающихся в данной вершине. Так что нам нужно доказать, что среднее число таких путей чуть больше, чем d^{10}/n.

Подсчёт удобно интерпретировать в терминах вероятностей. Будем считать, что помимо d паросочетаний $P_1, \ldots, P_d$ (напомним, что каждое из которых выбирается независимо, причём все паросочетания равновероятны) мы отдельно (и независимо) выбираем набор из 10 чисел $\omega = (\omega_1, \ldots, \omega_{10})$, каждое число от 1 до d. После этого мы (для фиксированной вершины графа) строим путь длины 10, выходящий из этой вершины. На первом шаге он идёт вдоль паросочетания P_{ω_1}, на втором — вдоль P_{ω_2}, и так далее. Нас интересует вероятность того, что после 10 шагов мы вернёмся в исходную точку. Точнее, мы хотим показать, что она равна $\frac{1}{n} \cdot (1 + o(1))$.

Поменяем порядок усреднения. Если усреднять сначала по выбору ω_i, то получается число петель длины 10 (делённое на d^{10}), которое затем можно усреднять по выбору P_i. Мы же проведём усреднение в другом порядке: сначала для каждого фиксированного набора ω_i мы усредняем по всем графам, и лишь потом усредняем по всевозможным наборам ω_i.

Все наборы $\omega = \omega_1, \ldots, \omega_{10}$ делятся на три категории:

1. гарантированно приводящие в исходную точку (независимо от выбора P_d); к этой категории относятся наборы, в которые после сокращений идущих подряд равных чисел ничего не остаётся;

2. наборы, состоящие из десяти разных чисел.

3. наборы, которые сокращаются не полностью, но в которых присутствуют равные числа (мы идём несколько раз по одному и тому же паросочетанию, но не обязательно подряд).

Для каждого из этих трёх типов наборов ω мы оцениваем количество таких наборов, а также (для каждого фиксированного набора) вероятность получить замкнутый путь при случайном выборе паросочетаний.

1. Количество наборов первого типа не превосходит $O(d^5)$. В самом деле, есть некоторое (фиксированное, так как длина цепочки — число 10 — фиксировано) число способов сокращения, и для каждого способа сокращения имеется не более d^5 способов его реализации (пять сокращающихся пар). Для каждого такого ω вероятность получить замкнутый путь равна 1 — какой бы граф мы не выбрали, путь с пометками ω обязательно приведет в исходную вершину.

2. Наборы без повторений составляют большинство из общего числа d^{10} (при достаточно больших значениях d). При этом вероятность того,

что на последнем шаге цикл замкнётся, и мы вернемся в исходную вершину, не превосходит $1/(n-1) = \frac{1}{n} \cdot (1 + o(1))$, поскольку последнее ω_{10} число в наборе ранее не встречалось и соответствующее ω_{10} паросочетание независимо с предыдущими 9 шагами нашего пути.

3. Количество наборов второго типа есть $O(d^9)$, где константа в O-обозначении соответствует числу возможных пар позиций, где происходит совпадение (то есть $C_{10}^2 = 10 \cdot 9/2$).

   Докажем, что вероятность вернуться в исходную точку для такого набора есть $O(1/n)$. Разобьём это событие на случаи в зависимости от того, когда путь в первый раз возвращается в уже пройденную вершину и того, какой эта вершина была по счёту. Разных случаев снова будет не больше C_{10}^2, так что достаточно рассмотреть вероятность одного из них.

   В момент перед назначенным возвращением в уже пройденную вершину уже фиксированы некоторые рёбра некоторых паросочетаний (те, что использованы в пути), а следующее ребро (по которому мы должны попасть в уже посещённую вершину) ещё не фиксировано. Поэтому для его конца остаётся не менее $n - 10$ вариантов, и вероятность выбрать один из них не больше $1/(n-10) = O(1/n)$.

Осталось сложить оценки вероятности из трёх разобранных случаев. Получаем для среднего числа замкнутых путей оценку сверху

$$O\left(\frac{d^5}{d^{10}}\right) \cdot 1 + \frac{1}{n} \cdot (1 + o(1)) + O\left(\frac{d^9}{d^{10}}\right) \cdot O\left(\frac{1}{n}\right).$$

Если $n = d^4$, то второй член в этой сумме будет основным, и потому среднее значение следа есть

$$n \cdot d^{10} \cdot \frac{1}{n}(1 + o(1)) = d^{10}(1 + o(1)).$$

Следовательно, существуют и даже образуют большинство графы, у которых след десятой степени матрицы близок к d^{10} и потому все собственные числа (кроме первого) равны $o(d)$. Таким образом, теорема доказана.

Упражнение 20 *Докажите аналогичное утверждение для $n = d^8$.*

Замечание: Как мы покажем в следующей главе, некоторое обобщение метода из доказательства Теоремы 6 позволяет доказать значительно более сильную оценку для спектрального зазора в случайном графе. Однако при первом чтении параграф 3.8 можно пропустить, поскольку Теоремы 6 и утверждения из Упражнения 20 достаточно для всех применений, которые нам потребуются в этой книге.

3.8 Усиление спектральной оценки для случайного графа*

В этой главе мы покажем, как улучшить оценку из параграфа 3.7 и доказать существование d-регулярных графов со вторым собственным числом $O(d^{3/4})$ (вместо доказанной в предыдущей главе оценки $o(d)$).

Теорема 7 *Для всякого d и всех достаточно больших n существует спектральный экспандер с параметрами $(n, d, O(1/d^{1/4}))$ (d-регулярный граф, второе собственное число которого по модулю не превосходит $O(d^{3/4})$).*

Доказательство: Мы докажем теорему для чётных n (доказательство для нечётных n оставляется читателю в качестве упражнения). Случайный граф с n вершинами мы выбираем так же, как и в доказательстве Теоремы 6. Мы независимо выбираем d случайных (по равномерной мере) паросочетаний на n вершинах и объединим их в один граф. В результате получается граф, в котором каждая вершина имеет степень d (в графе могут быть кратные ребра, но не может быть петель). Матрицу полученного графа обозначим M. Нужно доказать, что с большой вероятностью второе (по абсолютной величине) собственное число этой матрицы равно $O(d^{3/4})$.

Мы будем оценивать среднее значение следа M^{2k} для матрицы M случайно выбранного графа при подходящем k, которое подберём позже. (Мы возводим M непременно в чётную степень; это нужно для того, чтобы степени всех собственных чисел стали положительными.) Удобно заранее поделить матрицу на d, чтобы первое собственное число стало равным единице. Мы оценим второе собственное число графа $\lambda(G)$ через след M^{2k}:

$$1 + (\lambda/d)^{2k} \leqslant \operatorname{tr}[(M/d)^{2k}] \tag{3.8}$$

Из этого неравенства видно, что нам достаточно доказать существование графа, для которого правая часть в данном неравенстве мала. Для этого мы, как и раньше, оцениваем сверху среднее значение правой части.

Математическое ожидание следа равно сумме математических ожиданий диагональных элементов. Все они одинаковы, поэтому правая часть (3.8) равна n, умноженному на вероятность вернуться из данной точки в себя после k случайных шагов по нашему случайному графу. Таким образом, мы должны показать, что эта вероятность лишь немного превосходит $1/n$.

Вероятностное пространство является произведением двух независимых выборов. Во-первых, мы выбираем d независимых паросочетаний $P_1, \ldots, P_d$ на n вершинах (эти паросочетания и образуют граф G). Во-вторых, мы выбираем случайное слово длины $2k$ в алфавите из d букв $P_1, \ldots, P_d$ (каждом шаге блуждания по графу мы идем по ребру, которое получилось из одного из d паросочетаний P_i). Нас интересует следующее событие: начав с фиксированной вершины и проходя по рёбрам выбранных паросочетаний в выбранном порядке, мыв итоге возвращаемся в исходную вершину.

Оценим эту вероятность сначала для каждого $(2k)$-буквенного слова отдельно, а потом усредним по всем словам. Прежде всего мы заметим, что

идущие таком слове две одинаковые бувы подряд можно сократить (мы идем по одному и тому же ребру сначала в одну сторону, а потом обратно). Среди слов есть такие, от которых после выполнения сокращений ничего не остаётся; для таких слов интересующая нас вероятность равна 1. Другая простая ситуация: если после сокращения остаётся слово, в котором никакое паросочетание P_i не встречается дважды, то вероятность равна $1/(n-1)$: на последнем шаге условная вероятность вернуться в начало (при любом развитии событий на предыдущих шагах) равна $1/(n-1)$, так как предыдущие шаги никак не ограничивают последнее паросочетание.

Мы увидим, что для большинства слов вероятность вернуться в исходную вершину близка к $1/n$. Это большинство образуют *регулярные* слова. Чтобы определить, будет ли слово регулярным, представим себе его написанным по кругу и сократим (в том числе в точке контакта начала и конца). Если останется непустое слово, не являющееся степенью (не имеющее периода, меньшего длины цикла), то исходное слово будем называть регулярным. Нерегулярные слова, таким образом, после сокращений имеют вид XY^iX^{-1}, где Y — несократимое слово (слово, в котором нигде не встречаются две одинаковые буквы подряд).

Лемма 5 *Доля нерегулярных слов не больше* $O(k^2(9/d)^k)$.

Доказательство леммы 5: Нам нужно оценить число слов, которые после сокращения имеют вид XY^iX^{-1}.

Чтобы описать нерегулярное слово, нужно задать буквы в соответствующих X и Y. При этом требуется задать не более k букв (буквы X парны к буквам X^{-1}, а буквы в Y^i входят в $i \geqslant 2$ копиях). Таким образом, при фиксированных длинах X и Y у нас есть не более d^k способа выбрать слово XY^iX^{-1}.

Остаётся для каждой пары X, Y подсчитать число схем сокращения — число нерегулярных слов, которые после сокращения приводятся к виду XY^iX^{-1}. Будем сокращать исходное слово длины $2k$ слева направо (добавляем очередную букву и сокращаем её с предыдущей, если сокращается). Схему сокращения опишем символически: если добавленная буква впоследствии сократится, изображаем её левой скобкой, а ту, с которой она сократится, правой. Буквы, которые так и не сократятся, изобразим звёздочками. Таких последовательностей из скобок и звёздочек существует не больше, чем 3^{2k}. Наконец, к такой последовательности остаётся добавить длины слов X и Y, каждое не более $O(k)$.

Получатся, что число нерегулярных словне превосходит $d^k \cdot 3^{2k} \cdot O(k^2)$. Делим эту величину на общее число всех слов длины $2k$ в алфавите из d символов (т.е., на d^{2k}) и получаем $O(k^2(9/d)^k)$. Лемма 5 доказана.

Упражнение 21 *Докажите, что последовательность и скобок и звёздочек в схеме сокращения можно однозначно восстановить, если знать, где стоят левые скобки. (Это наблюдение позволяет усилить оценку в Лемме 5 до* $O(k^2(4/d)k)$.)

Лемма 6 *Для любого регулярного слова W вероятность того, что задаваемое W преобразование оставляет данную точку на месте (для случайных паросочетаний $P_1, \ldots, P_d$), не превышает $1/(n-2k) + O(k^4/n^2)$.*

Доказательство леммы 6: Нам будет удобно представлять себе вероятностный процесс следующим образом: вместо того, чтобы выбирать случайные паросочетания заранее, будем определять их постепенно по мере чтения слова W, оставляя не фиксированными те части паросочетаний $P_1, \ldots, P_d$, которые пока не понадобились. В каждый момент этого процесса для каждого паросочетания P_i уже определённая часть представляет собой набор рёбер (пар вершин графа), причём каждая вершина используется не более одного раза. Когда мы переходим к следующей букве слова W, может оказаться, что очередное ребро уже определено имеющейся частью перестановок (*вынужденный* ход), а может оказаться, что нет (*свободный* ход). В последнем случае мы доопределяем нужную перестановку, соединяя вершину (равновероятно) со всеми оставшимися кандидатами.

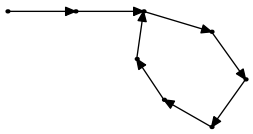
Будем говорить, что происходит *совпадение*, когда на свободном ходу мы попадаем в уже пройденную вершину. Заметим, что первое попадание в уже пройденную вершину всегда будет совпадением (в предыдущую вершину мы попали впервые, и потому из неё все ходы будут свободными, кроме возвратного, который невозможен, так как W несократимо). Поэтому интересующая нас вероятность разбивается на два случая:

- мы вернулись в исходную вершину, при этом произошло ровно одно совпадение;
- мы вернулись в исходную точку, при этом произошло не менее двух совпадений.

Второй случай разбивается на $O(k^2)$ вариантов в зависимости от моментов совпадений (мест в слове W, где они произошли). Вероятность каждого варианта не более $1/(n-2k)^2$. В самом деле, условная вероятность совпадения при фиксированной предыстории (пути до него) не больше $k/(n-2k)$, так как мы доопределяем перестановку в новой точке и имеется не менее $n-2k$ равновероятных вариантов, из которых не более k успехов. Таким образом вероятность второго случая не превосходит $O(k^4/n^2)$ (можно считать, что $k \ll n$, иначе оценка тривиальна).

Осталось показать, что в первом случае вероятность совпадения не превосходит $1/(n-2k)$. Это следует из того, что совпадение произойдёт в заранее известном месте, а именно, при движении по последней букве слова Y в разложении $W = XYX^{-1}$. Более того, до этого момента все пройденные вершины будут различны.

Сейчас мы это докажем, и при этом нам придётся воспользоваться непериодичностью слова Y (см. определение регулярного слова). Посмотрим на момент, когда мы впервые попадаем в уже пройденную вершину.



Этот момент будет совпадением, поэтому достаточно доказать, что это случается в конце слова Y. В самом деле:

- после этого новые вершины в пути уже не появятся, так как из новых вершин нужно вернуться в старые, и это будет совпадением;
- свободных ходов больше не будет, так как это было бы вторым совпадением;
- вынужденных ходов из каждой вершины (за исключением точки ветвения) не более двух, при этом один невозможен (возвращение по только что пройденному ребру означало бы, что слово W сократимо);
- поэтому движение определяется почти однозначно и состоит из нескольких циклов плюс возврат в исходную вершину (по предположению мы туда возвращаемся);
- однозначно определяются не только ходы, но и соответствующие буквы слова, поэтому разбиение на цикл и отросток совпадает с разложением $W = XYX^{-1}$;
- поэтому мы можем сделать только один оборот (иначе Y было бы периодично).

Лемма 6 доказана.

Подведём итог. Мы получили суммарную оценку для среднее суммы всех собственных чисел матрицы $(M/d)^{2k}$

$$n \cdot \left(\frac{1}{n-2k} + O(k^4/n^2) + O(k^2(9/d)^k) \right) = \\ = n \cdot \left(\frac{1}{n} + O(k^4/n^2) + O(k^2(9/d)^k) \right)$$

Остаётся подобрать параметр k. Выберем его так, чтобы $(9/d)^k = 1/n^2$, то есть $k = 2\log_{d/9} n$. Тогда третье слагаемое в скобках поглощается вторым. Значит, математическое ожидание следа матрицы $(M/d)^{2k}$ не превосходит

$$1 + O(k^4/n).$$

Соответственно, среднее значение второго собственного числа этой матрицы не превосходит $O(k^4/n)$ (напомним, что мы возвели матрицу в чётную степень, так что все собственные числа положительны). Теперь мы можем

заключить, что найдётся хотя бы одна матрица M, у которой второе собственное число не превосходит $\sqrt[2k]{\frac{O(k^4)}{n}}$. (На самом деле это условие выполнено не только для хотя бы одной матрицы, но для большинства матриц M.) Вспоминая о выборе k (условие $n^2 = (d/9)^k$), получаем оценку для второго собственного числа

$$\sqrt[2k]{\frac{\mathrm{poly}(k)}{n}} = \frac{O(1)}{d^{1/4}}.$$

Теорема доказана.

Замечания:

1. Мы доказали существование *хотя бы одного* графа с оценкой $O(d^{3/4})$ на второе собственное число. Тоже самое рассуждение вместе с неравенством Чебышёва показывает, что *большинство* графов по указанному распределению обладает этим свойством.

Можно доказать, что это распределение достаточно близко к равномерному распределению на множестве всех $2d$-регулярных графов.

2. Мы рассматривали графы чётной степени; если требуется построить граф нечётной степени, можно добавить по петле к каждой вершине, отчего собственные числа увеличатся на единицу.

3. Теорема о том, что у большинства графов второе собственное число имеет порядок $O(d^{3/4})$ была доказано Бродером и Шамиром (Broder–Shamir, [14]). Намного более сложные рассуждения (Joel Friedman, [13]) позволяют доказать, что на самом деле у большинства d-регулярных графов второе собственное число близко к $2\sqrt{d-1}$. В тоже время, Теорема 5 показывает, что второе собственное число не может быть меньше $2\sqrt{d-1} - o(1)$ (для больших n).

3.9 Случайное блуждание на экспандерах

Мы уже отмечали, что спектральные экспандеры обладают свойствами, которые можно назвать свойством хорошего «перемешивания». Даже один шаг случайного блуждания на спектральном экспандере заметно приближает исходное распределение вероятностей на вершинах к равномерному (см. Лемму о перемешивании на с. 18 и Утверждение 2 на с. 20). В этой главе мы подробнее изучим случайное блуждание на спектральном экспандере, состоящее из нескольких шагов. Для начала мы продемонстрируем основной технический приём данной главы на простом примере.

Утверждение 4 *Пусть граф G является спектральным (n, d, γ)-экспандером без петель и A — некоторое множество вершин графа, состоящее из αn вершин. Тогда число рёбер в индуцированном A подграфе (число рёбер, оба конца которых принадлежат A) не превосходит*

$$\frac{nd}{2}(\alpha^2 + \gamma\alpha(1-\alpha)).$$

Замечание 1: Если выбирать ребро графа случайно, то для каждого из его концов вероятность попасть в A равна α. Если бы эти события для двух концов ребра были бы независимы друг от друга, то вероятность того, что ребро попало в индуцированные A подграф, равнялась бы α^2. Конечно же, на самом деле эти события зависимы. Доказываемое утверждение гласит, что для экспандера данную вероятность можно оценить величиной $(\alpha^2 + \gamma\alpha(1-\alpha))$.

Замечание 2: Мы уже знаем, что спектральный экспандер обладает свойством хорошего рёберного расширения: если $|A|$ не слишком велико, то найдется достаточно много рёбер, ведущих из множества A в его дополнение. По существу, доказываемое утверждение говорит то же самое, но в других терминах: найдется не слишком много рёбер, у которых оба конца принадлежат A.

Доказательство: Рассмотрим вектор $\mathbf{f} = (f_1, \ldots, f_n)$, где $f_i = 1$, если i-ая вершина графа G принадлежит A, и $f_i = 0$ иначе. Если M матрица графа, то число рёбер, оба конца которых принадлежат A, можно вычислить как $\frac{1}{2}(\mathbf{f}M\mathbf{f}^\perp)$. В самом деле,

$$\mathbf{f}M\mathbf{f}^\perp = \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij} f_i f_j,$$

и каждое ребро графа $\{i, j\}$ даёт в эту сумму вклад, равный 2 (каждое ребро считается дважды).

Пусть $\mathbf{e}_1, \ldots, \mathbf{e}_n$ – собственный ортонормированный базис матрицы M и $\lambda_1, \lambda_2, \ldots, \lambda_n$ соответствующие собственные числа. Мы знаем, что $e_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1)$ и $\lambda_1 = d$, и $|\lambda_i| \leqslant \gamma d$ для всех $i > 1$.

Разложим вектор $\mathbf{f}$ в сумму двух ортогональных векторов: $\mathbf{f} = \mathbf{f}_\| + \mathbf{f}_\perp$, где $\mathbf{f}_\|$ есть проекция $\mathbf{f}$ на собственный вектор $\mathbf{e}_1$, и $\mathbf{f}_\perp$ есть проекция f на пространство, порождённое $\mathbf{e}_2, \ldots, \mathbf{e}_n$.

Заметим, что $\mathbf{f}_\| = c\mathbf{e}_1$, где коэффициент c равен скалярному произведению $\mathbf{f}$ и $\mathbf{e}_1$. Другими словами,

$$\mathbf{f}_\| = \frac{f_1 + \ldots + f_n}{n} \cdot (1, \ldots, 1). \tag{3.9}$$

Теперь вернемся к произведению $\mathbf{f}M\mathbf{f}^\perp$ и оценим его сверху.

$$\mathbf{f}M\mathbf{f}^\perp = \mathbf{f}_\| M\mathbf{f}_\| + \mathbf{f}_\perp M\mathbf{f}_\perp^\perp = \lambda_1||f_\||^2 + \sum_{i=1}^{n}\lambda_i(\mathbf{f}_\perp, e_i)^2 \leqslant$$

$$\leqslant d||\mathbf{f}_\||^2 + (\gamma d)\sum_{i=1}^{n}(\mathbf{f}_\perp, e_i)^2 \leqslant d||\mathbf{f}_\||^2 + (\gamma d)||\mathbf{f}_\perp||^2. \tag{3.10}$$

Поскольку нормы векторов $\mathbf{f}_\|$ и $\mathbf{f}_\perp$ не превосходят нормы $\mathbf{f}$, мы заключаем, что

$$\mathbf{f}M\mathbf{f}^\perp \leqslant (1+\gamma)d||\mathbf{f}||^2.$$

Напомним, что среди координат вектора $\mathbf{f}$ имеется αn единиц и $(1-\alpha)$ нулей. Следовательно, квадрат нормы $||\mathbf{f}||^2 = \alpha n$. Следовательно,

$$\mathbf{f}M\mathbf{f}^{\perp} \leqslant \alpha(1+\gamma)dn,$$

и число рёбер в индуцированном подграфе не превосходит $\frac{\alpha(1+\gamma)dn}{2}$. Это неравенство уже само по себе является неплохой оценкой. Но мы обещали доказать немного более сильное утверждение.

Где можно усилить оценку? Мы очень грубо оценили величины $||\mathbf{f}_{\parallel}||^2$ и $||\mathbf{f}_{\perp}||^2$ как $||\mathbf{f}||^2$. На самом же деле по теореме Пифагора мы имеем

$$||\mathbf{f}_{\parallel}||^2 + ||\mathbf{f}_{\perp}||^2 = ||\mathbf{f}||^2.$$

Таким образом, (3.10) можно более аккуратно оценить как

$$d||\mathbf{f}_{\parallel}||^2 + (\gamma d)||\mathbf{f}_{\perp}||^2 \leqslant ||\mathbf{f}_{\parallel}||^2 + (\gamma d)(||\mathbf{f}||^2 - ||\mathbf{f}_{\parallel}||^2) = (\gamma d)||\mathbf{f}||^2 + (1-\gamma)d||\mathbf{f}_{\parallel}||^2.$$

Остаётся вспомнить равенство (3.9), из которого следует

$$||\mathbf{f}_{\parallel}||^2 = \frac{(f_1 + \ldots + f_n)^2}{n} = \alpha^2 n.$$

Таким образом, мы получаем, что $\mathbf{f}M\mathbf{f}^{\perp}$ не превосходит

$$(\gamma d)\alpha n + (1-\gamma)d\alpha^2 n = (\alpha^2 + \gamma\alpha(1-\alpha)) \cdot (dn).$$

Число рёбер в индуцированном подграфе не превосходит половины от этой величины, и утверждение доказано.

Теорема 8 *Пусть граф G является спектральным (n, d, γ)-экспандером с матрицей M и $\mathbf{f} = (f_1, \ldots, f_n)$ некоторый вектор. Тогда*

$$\mathbf{f}M\mathbf{f}^{\perp} \leqslant \gamma d||f||^2 + \frac{d(1-\gamma)}{n}\left(\sum f_i\right)^2.$$

Доказательство: Рассуждение по существу повторяет доказательство утверждения 4. Обозначим $\mathbf{e}_1, \ldots, \mathbf{e}_n$ ортонормированный собственный базис для M и $\lambda_1, \ldots, \lambda_n$ соответствующие собственные числа. (При этом, как обычно, мы можем полагать $\mathbf{e}_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1)$ и $\lambda_1 = d$). Далее, разложим вектор $\mathbf{f}$ в сумму $\mathbf{f} = \mathbf{f}_{\parallel} + \mathbf{f}_{\perp}$, где $\mathbf{f}_{\parallel}$ параллелен, $\mathbf{f}_{\perp}$ перпендикулярен базисному вектору $\mathbf{e}_1$.

Заметим, что

$$\mathbf{f}_{\parallel} = (\mathbf{f}, \mathbf{e}_1) \cdot \mathbf{e}_1 = \frac{\sum f_i}{\sqrt{n}}(1, \ldots, 1).$$

Далее,

$$\mathbf{f}M\mathbf{f}^{\perp} = \mathbf{f}_{\parallel}M\mathbf{f}_{\parallel}^{\perp} + \mathbf{f}_{\perp}M\mathbf{f}_{\perp}^{\perp}.$$

Поскольку $\mathbf{f}_{\perp}$ лежит в подпространстве собственных векторов, собственные числа которых не превосходят γd, мы получаем

$$\mathbf{f}M\mathbf{f}^{\perp} \leqslant d||\mathbf{f}_{\parallel}||^2 + (\gamma d)||\mathbf{f}_{\perp}||^2.$$

По теореме Пифагора мы имеем $||\mathbf{f}||^2 = ||\mathbf{f}_\parallel||^2 + ||\mathbf{f}_\perp||^2$. Следовательно,

$$\mathbf{f}M\mathbf{f}^\perp \leqslant d||\mathbf{f}_\parallel||^2 + (\gamma d)(||\mathbf{f}||^2 - ||\mathbf{f}_\parallel||^2).$$

Остаётся заметить, что $||\mathbf{f}_\parallel||^2 = \frac{(\sum f_i)^2}{n}$, и теорема доказана.

Даже если граф G однороден (все вершины имеют степень d), его индуцированный подграф может быть неоднородным. Однако матрица индуцированного подграфа симметрична, а значит, имеет собственный базис. Все собственные числа подграфа по абсолютной величине не превосходят максимума отношения Рэлея

$$\frac{|\mathbf{f}M\mathbf{f}^\perp|}{||\mathbf{f}||^2}$$

среди всех ненулевых векторов $\mathbf{f}$, сосредоточенных на вершинах подграфа (координаты f для вершин, не принадлежащих A, должны быть равны нулю). Из теоремы 8 вытекает следующее следствие.

Следствие 4 *Пусть граф G является спектральным (n, d, γ)-экспандером без петель и A — некоторое множество вершин графа, состоящее из αn вершин. Тогда все собственные числа индуцированного подграфа на вершинах A не превосходят*

$$(\gamma + \alpha(1 - \gamma))d.$$

Теперь мы готовы доказать несколько утверждений о блуждании на экспандере.

Утверждение 5 *Пусть граф G является спектральным (n, d, γ)-экспандером без петель и A — некоторое множество вершин графа, состоящее из αn вершин. Рассмотрим случайное блуждание по графу*

$$x_0 - x_1 - \ldots - x_k,$$

где вершина x_0 выбирается случайно (по равномерному распределению), а затем на каждом шаге $i = 1, \ldots, k$ следующая вершина x_i выбирается случайно (также равномерно) среди всех соседей x_{i-1}. Тогда

$$\mathrm{Prob}[x_i \in A \text{ для всех } i] \leqslant (\alpha + \gamma - \alpha\gamma)^k.$$

Доказательство: Общее число путей $x_0 - x_1 - \ldots - x_k$, в графе G равно nd^k (имеется n вариантов для выбора первой вершины x_0 и по d вариантов для выбора каждого из k шагов). Нужно подсчитать, сколько из этих путей полностью лежат в A.

Обозначим $\mathbf{f}^{(i)} = (f_1^{(i)}, \ldots, f_n^{(i)})$ такой вектор, где $f_j^{(i)}$ есть число путей длины i, проходящих только по вершинам A и заканчивающимся в j-ой вершине графа G. В частности, в векторе $\mathbf{f}^{(0)}$ в позициях вершин A стоят единицы, а в позициях вершин вне A стоят нули.

Каждый следующий вектор $\mathbf{f}^{(i+1)}$ получается из $\mathbf{f}^{(i)}$ умножением на матрицу индуцированного подграфа. Поскольку собственные числа матрицы этого графа не превосходят $(\gamma d + \alpha(1-\gamma))d$, мы получаем

$$||\mathbf{f}^{(k)}|| \leqslant ((\gamma d + \alpha(1-\gamma))d)^k ||\mathbf{f}^{(0)}|| = ((\gamma d + \alpha(1-\gamma))d)^k \cdot \sqrt{\alpha n}.$$

Применяем неравенство Коши и получаем, что сумма координат (L_1-норма) вектора $\mathbf{f}^k$ не превосходит

$$\sqrt{n} \cdot ||\mathbf{f}^{(k)}|| \leqslant ((\gamma + \alpha(1-\gamma)))^k \cdot (d^k n).$$

Утверждение доказано.

Замечание 1: Рассмотрим случайное блуждание по экспандеру, состоящее из $k = rs$ шагов,

$$x_0 - x_1 - \ldots - x_{rs}.$$

Как и раньше, вершина x_0 выбирается случайно (по равномерному распределению), а затем на каждом шаге $i = 1, \ldots, k$ следующая вершина x_i выбирается случайно (также равномерно) среди всех соседей x_{i-1}. Будем интересоваться вероятностью того, что все вершины с номерами, кратными s (т.е., $x_0, x_s, x_{2s}, \ldots, x_{rs}$) попали в множество A. Вероятность этого события оценивается следующим образом:

$$\text{Prob}[x_i \in A \text{ для всех } i \text{ кратных } s] \leqslant (\alpha + \gamma^s(1-\alpha))^r \leqslant (\alpha + \gamma(1-\alpha))^r.$$

В самом деле, нужно применить Утверждение 5 не к исходному графу G, а к его s-ой степени (к графу на n вершинах, рёбрами которого являются пути длины s в G).

Замечание 2: Снова рассмотрим случайное блуждание по экспандеру, состоящее из k шагов,

$$x_0 - x_1 - \ldots - x_k.$$

На этот раз оценим вероятностью того, что в множество A попали все вершины с номерами $x_{i_1}, x_{i_1+i_2}, x_{i_1+i_2+i_3}, \ldots, x_{i_1+i_2+\ldots+i_r}$. Вероятность этого события оценивается сверху

$$(\alpha + \gamma^{i_1}(1-\alpha)) \cdot (\alpha + \gamma^{i_1}(1-\alpha)) \cdot \ldots \cdot (\alpha + \gamma^{i_r}(1-\alpha)) \leqslant (\alpha + \gamma(1-\alpha)^r.$$

В самом деле, в рассуждение из замечания 1 легко переносится на случай, когда шаги между «контролируемыми» номерами шагов x_j неодинаковы. Таким образом, мы доказали следующий результат:

Утверждение 6 *Пусть граф G является спектральным (n, d, γ)-экспандером без петель и A — некоторое множество вершин графа, состоящее из αn вершин. Рассмотрим случайное блуждание по графу*

$$x_0 - x_1 - \ldots - x_k,$$

где вершина x_0 выбирается случайно и равномерно, а затем на каждом шаге $i = 1, \ldots, k$ следующая вершина x_i выбирается случайно равномерно среди всех соседей x_{i-1}.

Пусть $I \subset \{0, \ldots, t\}$ некоторое подмножество номеров шагов. Тогда

$$\mathrm{Prob}[x_i \in A \text{ для всех } i \in N] \leqslant (\alpha + \gamma - \alpha\gamma)^{|I|-1}.$$

Упражнение 22 *Для распределения вероятностей $\mathbf{p} = (p_1, \ldots, p_n)$ рассмотрим три варианта меры «неопределённости»:*

(а) энтропия Шеннона $H(\mathbf{p}) = \sum\limits_{p_i \neq 0} p_i \log \frac{1}{p_i}$,

(б) энтропия Реньи $H_2(\mathbf{p}) = -\log\left(\sum\limits_{i=1}^{n} p_i^2\right)$,

(в) min-энтропия $H_\infty(\mathbf{p}) = \log\left(\min\limits_{p_i > 0} \frac{1}{p_i}\right)$.

Докажите, что при умножение вектора распределения $\mathbf{p}$ на любую дважды стохастическую матрицу M (все матричные элементы M неотрицательны; сумма элементов в каждом столбце и в каждой строке равна 1) величина каждого из этих трёх видов энтропий не уменьшается.

Упражнение 23 *Пусть $G = (V, E)$ является спектральным (n, d, γ)-экспандером, и $A \subset E$ — некоторое множество его рёбер. Выберем случайное ребро из A, а затем случайно выберем один из концов этого ребра. Затем сделаем i шагов случайного блуждания по графу. Покажите, что вероятность того, что последнее ребро в данном случайно выбранном пути принадлежит A, не превосходит $\frac{|A|}{|E|} + \gamma^{i-1}$.*

Глава 4

Рекурсивные конструкции экспандеров

4.1 Классические произведения графов

Напомним, что мы уже встречались с некоторым способом умножения графа на самого себя. Возведением графа G в степень k называется следующая операция: мы сохраняем прежнее множество вершин, а рёбрами в новом графе считаем все пути длины k в исходном графе. Результат возведения графа в степень k мы обозначаем G^k. Если исходный граф был однородным степени d, то его k-ая степень будет также однородным графом, но степени d^k. Если M — матрица исходного графа, то матрицей его k-ой степени будет будет M^k (обычное возведение матрицы в степень k). При этом, разумеется, собственные векторы матрицы сохраняются, а собственные числа также возводятся в степень k.

Нетрудно также определить тензороное произведение графов. Для произвольных графов $G_1 = (V_1, E_1)$ и $G_2 = (V_2, E_2)$ назовём их *тензороным произведением* граф $G = (V, E)$, в котором множество вершин $V = V_1 \times V_2$ (каждая вершина в новом графе есть пара вершин исходных графов), а рёбрами соединяются все такие пары (v_1, v_2) и (v_1', v_2'), для которых

$$\{v_1, v_1'\} \in E_1 \text{ и } \{v_2, v_2'\} \in E_2.$$

Тензорное произведение графов G_1 и G_2 мы обозначаем $G_1 \otimes G_2$.

Упражнение 24 *Объясните, как из матриц графов G_1 и G_2 получить матрицу тензорного произведения $G_1 \otimes G_2$.*

Упражнение 25 *Пусть спектр графа G_1 состоит из чисел*

$$\lambda_{1,1}, \lambda_{1,2}, \ldots, \lambda_{1,n},$$

а спектр граф G_2 состоит из чисел

$$\lambda_{2,1}, \lambda_{2,2}, \ldots, \lambda_{2,m}.$$

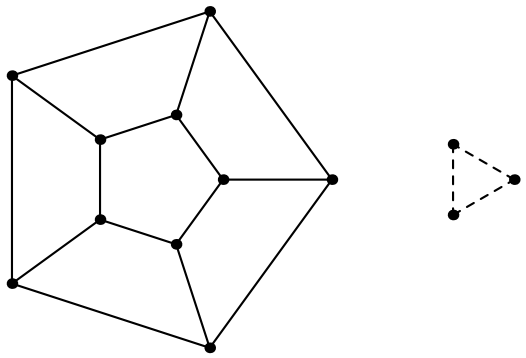
Докажите, что спектр $G_1 \otimes G_2$ состоит из всевозможных произведений $(\lambda_{1,i} \cdot \lambda_{2,j})$.

Далее в этой главе мы определим более сложные операции на графах — зигзаг-произведение и подстановочное произведение. Используя эти операции (вместе с тензорным произведением и обычным возведением в степень) мы сможем строить экспандеры сколь угода большого размера из маленьких «строительных блоков».

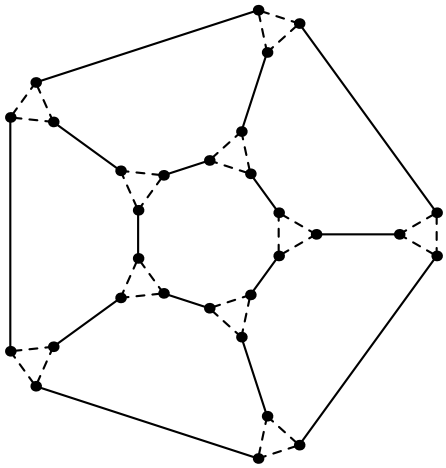
4.2 Зигзаг-произведение графов

В этой главе мы изучим метод, позволяющий получать экспандеры с хорошими параметрами с помощью особого «произведения» графов. Это произведение позволяют «собирать» большие спектральные экспандеры из маленьких блоков (а подходящего вида маленькие блоки, которые и сами должны быть эспандерами, мы можем найти перебором.)

Пусть даны два графа $G(n, D)$ и $H(D, d)$. Запись в скобках указывает параметры: число вершин и степень каждой вершины (одинаковую для всех вершин). Пусть при этом число вершин второго графа равно степени первого (как в примере на рисунке).



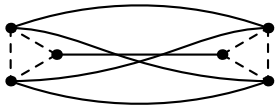
Мы определим *загзаг-произведение* этих графов. Для этого каждую вершину первого графа заменим маленькой копией второго графа, прикрепив рёбра первого графа к вершинам второго. (В маленьком графе как раз нужное число вершин.) Обратим внимание, что в прикреплении есть произвол — конкретный выбор соответствия в каждой вершине не играет роли. Получится граф с nD вершинами и рёбрами двух типов — большими рёбрами, унаследованными из из первого графа, и малыми , унаследованными из второго. (На рисунке — сплошные и пунктирные линии соответственно.)



Зигзаг-произведение (zig-zag product): Вершины у зигзаг-произведения будут те же, но рёбра совсем другие. В качестве рёбер нового графа мы берём все пути длины 3 вида

[пунктирное ребро] – [сплошное ребро] – [пунктирное ребро].

Другими словами, каждое сплошное ребро порождает d^2 рёбер зигзаг-произведения (соединяющих пары пунктир-соседей концов сплошного ребра), как показано на рисунке (рёбра зигзаг-произведения показаны кривыми линиями):



Легко видеть, что все вершины зигзаг-произведения имеют степень d^2 (каждое из двух пунктирных рёбер можно выбрать d способами).

Опишем матрицу графа, полученного в результате зигзаг-произведения. Для этого мы рассмотрим две матрицы размера $nD \times nD$ (координаты соответствуют вершинам графа). Первая матрица $\tilde{H}$ есть матрица графа с рёбрами из пунктирных линий, вторая матрица $\tilde{G}$ соответствует графу с теми же вершинами, но с рёбрами из сплошных линий. (Обозначения показывают, что эти матрицы происходят из соответственно первого и второго графов, участвующих в зигзаг-произведении). В $\tilde{H}$ каждый столбец и каждая строка содержат d единиц, а в $\tilde{G}$ — только одну единицу. Отметим, что $\tilde{G}$ задаёт перестановку на множестве вершин, и её норма равна единице. Ясно, что матрицей зигзаг-произведения будет произведение матриц $\tilde{H}\tilde{G}\tilde{H}$.

Далее мы докажем оценки для второго собственного числа зигзаг-произведения.

4.3 Первая спектральная оценка для зигзаг-произведения

Докажем, что зигзаг-произведение двух графов, у которых малы вторые (по абсолютной величине) собственные числа, тоже имеет небольшое второе собственное число.

Теорема 9 *Зигзаг-произведение спектрального (n, D, α)-экспандера G и спектрального (D, d, β)-экспандера H является спектральным экспандером с параметрами $(nD, d^2, \leqslant \alpha + \beta + \beta^2)$.*

Замечание. Можно улучшить оценку для второго собственного числа до $\alpha+\beta$, но мы ограничимся доказательством более слабого (и более простого) утверждения.

Доказательство: Чтобы оценить второе собственное значение симметричной матрицы, надо ограничить квадратичную форму на ортогональное дополнение к первому собственному вектору $\mathbf{e}_0 = (1, \ldots, 1)$ и взять её максимум на единичном шаре. Другими словами, модуль второго собственного числа матрицы $\tilde{H}\tilde{G}\tilde{H}$ есть максимум выражения

$$|\mathbf{f}\tilde{H}\tilde{G}\tilde{H}\,\mathbf{f}^{\perp}|$$

по всем векторам $\mathbf{f}$ длины nD, имеющим единичную длину и ортогональных e_0 (то есть имеющих нулевую сумму координат). Чтобы оценить это выражение, разложим $\mathbf{f}$ в сумму $\mathbf{f} = \mathbf{g} + \mathbf{h}$ следующим образом: координаты $\mathbf{g}$ одинаковы в каждой из «облаков» (копий графа H), а для $\mathbf{h}$ на каждой копии графа H сумма координат равна нулю. Чтобы оценить значение квадратичной формы на произвольном векторе $\mathbf{f} \perp \mathbf{e}_0$, мы отдельно изучим действие матриц $\tilde{G}$ и $\tilde{H}$ на векторы $\mathbf{g}$ и $\mathbf{h}$:

(a) $\tilde{H}\mathbf{g} = d \cdot \mathbf{g}$, поскольку внутри каждой копии H веса (координаты) вектора $\mathbf{g}$ одинаковы, и каждый из них распространяется в d соседей в том же облаке.

(b) $\|\tilde{H}\mathbf{h}\| \leqslant \beta d \cdot \|\mathbf{h}\|$, поскольку вектор $\mathbf{h}$ в каждой копии H ортогонален первому собственному вектору матрицы графа H, а все остальные собственные векторы этой матрицы не превосходят (по модулю) βd. (Если в каждой компоненте норма оператора не превосходит βd, то это верно и для всего оператора.)

(c) $|\mathbf{g}\tilde{G}\mathbf{g}^{\perp}| \leqslant \alpha\|\mathbf{g}\|^2$; в самом деле, квадратичная форма в левой части содержит один ненулевой член для каждого ребра, позаимствованного из графа G (теперь это ребро соединяет две вершины из разных облаков). Поэтому выражение внутри знака модуля из левой части равно $(\hat{\mathbf{g}})M(G)\hat{\mathbf{g}}^{\perp}$, где $M(G)$ — матрица смежности графа G, а $\hat{\mathbf{g}}$ получается из $\mathbf{g}$ склеиванием равных значений в каждой компоненте. При этом сумма координат в $\hat{\mathbf{g}}$ (как и в исходном векторе $\mathbf{f}$) равна нулю. Поэтому данное выражение (по предположению о графе G) оценивается как $\alpha m\|\hat{\mathbf{g}}\|^2$, что равно как раз $\alpha\|\mathbf{g}\|^2$.

Теперь мы можем оценить искомое выражение:

$$|\mathbf{f}\tilde{H}\tilde{G}\tilde{H}\mathbf{f}^{\perp}| = |(\mathbf{g}+\mathbf{h})\tilde{H}\tilde{G}\tilde{H}(\mathbf{g}+\mathbf{h})^{\perp}| \leqslant$$
$$\leqslant |\mathbf{g}\tilde{H}\tilde{G}\tilde{H}\mathbf{g}^{\perp}| + 2|\mathbf{g}\tilde{H}\tilde{G}\tilde{H}\mathbf{h}^{\perp}| + |\mathbf{h}\tilde{H}\tilde{G}\tilde{H}\mathbf{h}^{\perp}|.$$

Первое из трёх слагаемых равно $d^2|\mathbf{g}\tilde{G}\mathbf{g}|$ по свойству (a) и потому не превосходит $\alpha d^2\|\mathbf{g}\|^2$ по (c). Второе слагаемое не превосходит $2\cdot(\beta d)\cdot d\cdot\|\mathbf{g}\|\cdot\|\mathbf{h}\|$ (матрица $\tilde{G}$ есть матрица перестановки и сохраняет норму). Наконец, третье слагаемое не превосходит $(\beta d)^2\|\mathbf{h}\|^2$ по аналогичным причинам. В этих оценках можно заменить $\|\mathbf{g}\|$ и $\|\mathbf{h}\|$ на $\|\mathbf{f}\|$ и получить

$$(\alpha + 2\beta + \beta^2)\|\mathbf{f}\|^2.$$

Это уже довольно хорошая оценка, но мы можем её усилить и избавиться от двойки перед коэффициентом β. Для этого заметим, что по неравенству Коши–Буняковского $2\|\mathbf{g}\|\cdot\|\mathbf{h}\| \leqslant \|\mathbf{h}\|^2 + \|\mathbf{g}\|^2$. Следовательно,

$$\begin{aligned}\alpha d^2\|\mathbf{g}\|^2 + 2\cdot(\beta d)\cdot d\cdot\|\mathbf{g}\|\cdot\|\mathbf{h}\| + (\beta d)^2\|\mathbf{h}\|^2 \leqslant\\ \leqslant d^2\cdot(\alpha\|\mathbf{g}\|^2 + \beta\|\mathbf{g}\| + \beta\|\mathbf{h}\|^2 + (\beta d)^2\|\mathbf{h}\|^2) \leqslant\\ \leqslant (\alpha+\beta+\beta^2)d^2\cdot(\|\mathbf{g}\|^2 + \|\mathbf{h}\|^2).\end{aligned}$$

Остается вспомнить, что векторы $\mathbf{g}$ и $\mathbf{h}$ ортогональны друг другу, и по теореме Пифагора $\|\mathbf{g}\|^2 + \|\mathbf{h}\|^2 = \|\mathbf{f}\|^2$. Таким образом, мы получаем оценку

$$|\mathbf{f}\tilde{H}\tilde{G}\tilde{H}\mathbf{f}^{\perp}| \leqslant (\alpha+\beta+\beta^2)d^2\|\mathbf{f}\|^2,$$

и теорема доказана.

4.4 Две рекурсивные конструкции с зигзаг-произведением

Построим последовательность явно заданных графов одной и той же степени с растущим числом вершин, имеющих малые собственные числа. Основная идея: возводя матрицу в квадрат, мы не меняем число вершин графа и уменьшаем (возводим в квадрат) нормализованное (т.е. делённое на степень графа) второе собственное число. Зато мы увеличиваем (тоже возводим в квадрат) степень вершины. Но это можно скомпенсировать зигзаг-умножением на фиксированный граф H.

Опишем конструкцию более подробно. Зафиксируем граф $H(d^4, d, 1/10)$ для некоторого d (для достаточно больш*и*х d такой граф, как мы видели, существует). Затем построим последовательность графов $G_0, G_1, \ldots$, положив

- $G_0 = H^2$. Параметры этого экспандера: $(d^4, d^2, 1/100)$.

- G_{n+1} есть зигзаг-произведение G_n^2 и H. По индукции доказывается, что экспандер G_n имеет параметры $(d^{4n+4}, d^2, \leqslant 1/2)$. В самом деле, после возведения в квадрат получаем $(d^{4n+4}, d^4, 1/4)$, а умножение на $H(d^4, d, 1/10)$ даёт число вершин d^{4n+8}, степень каждой вершины d^2 и третий параметр
$$\frac{1}{4} + \frac{1}{10} + \frac{1}{10^2} < \frac{1}{2},$$
что завершает доказательство.

Мы получили конструкцию экспандера, конторая является эффективной в «слабом» смысле — такие графы можно строить за время полиномиально зависящее от числа вершин. В приложениях нам могут понадобиться экспандеры эффективные в более сильном смысле — графы, для которых по номеру вершины можно эффективно найти список номеров её соседей (см. обсуждение на с. 13).

Чтобы более точно определить, что такое эффективная конструкция графа $G = (V, E)$, мы произвольным образом зафиксируем для каждой вершины графа нумерацию инцидентных ей рёбер. Будем называть *функцией вращения* отображение
$$N \ : \ \langle x, i \rangle \to y,$$
которое сопоставляет вершине графа $x \in V$ и номеру i (не превосходящему степени вершины x) вершину $y \in V$, которая является i-ым соседом x (выйдя из x по i-ому ребру, мы попадём в вершину $y = N(x, i)$).

Если число вершин в графе не превосходит 2^n, а степень каждой вершины равна 2^d, то каждую вершину можно задавать n-битным индексом, а номер выходящего из вершины ребра, соответственно, d-битным индексом. Таким образом, функцию вращения можно понимать как отображение
$$N \ : \ \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^n.$$

Сильная эффективность означает, что время вычисления $N(x, i)$ полиномиально зависит от длины аргументов, т.е., от от логарифма числа вершин и от логарифма степени графа. (Для сравнения: в «слабо эффективиной» конструкции графа функция вращения будет вычислимой за время полиномиально зависящее от самого числа вершин графа, а не от его логарифма).

Как происходит вычисление функции N в построенной нами последовательности графов? Номер ребра представляет собой пару чисел, каждое от 1 до d. Вершина графа G_{n+1} представляет собой пару: одна вершина G_n^2 (=вершина G_n) и одна вершина H. Движение по ребру: сначала идём по ребру H, попадаем в какую-то вершину H (в диапазоне $1 \ldots d^4$), воспринимаем её как пару чисел в диапазоне $1 \ldots d^2$, рекурсивно идём по двум рёбрам графа G_n и затем делаем ещё ход в H. Таким образом, вычисление N для графа G_{n+1} использует два вызова аналогичного вычисления для G_n, что приводит к экспоненциальной оценке по n, и полиномиальной вычислимости функции N не получается.

Однако можно модифицировать конструкцию, используя не предыдущий граф G_n, а граф с половинным индексом. Для начала выберем граф H с параметрами $(d^8, d, 1/10)$, а затем построим последовательность

$$\begin{aligned} G_0&\colon (1, d^2, 1/2)\\ G_1&\colon (d^8, d^2, 1/2)\\ G_2&\colon (d^{16}, d^2, 1/2)\\ &\dots\\ G_n&\colon (d^{8n}, d^2, 1/2)\\ &\dots \end{aligned}$$

Начальные графы G_0 и G_1 построить легко (G_0 — это граф, состоящий из единственной вершины и d^2 петель, граф G_1 можно получить из H размножением рёбер в d раз, что не меняет собственных чисел). Затем можно воспользоваться рекуррентной формулой

$$G_n = (G_{\lfloor (n-1)/2 \rfloor} \otimes G_{\lceil (n-1)/2 \rceil})^2 ⓩ H,$$

где $\otimes$ обозначает тензорное произведение, а ⓩ — зигзаг-произведение. Тензорное произведение в скобках имеет параметры $(d^{8(n-1)}, d^4, 1/2)$, после возведения в квадрат получается $(d^{8(n-1)}, d^8, 1/4)$ и после зигзаг произведение $(d^{8n}, d^2, 1/2)$ (в силу того же вычисления с $1/4$ и $1/10$, что и раньше).

Преимущество новой конструкции в том, что при вычислении функции вращения два рекурсивных вызова относятся к половинным значениям n; глубина рекурсии теперь стала логарифмической по n, и общее время вычисления полиномиально по n.

4.5 Аффинная плоскость как экспандер

Все явные последовательности экспандеров, которые мы строили, формировались вокруг «затравочного» экспандера H с подходящими параметрами, который мы находили перебором (длина перебора не зависела от n, так что мы имели право использовать его в полиномиальном по n алгоритме). Сейчас мы опишем алгебраическую конструкцию, которая позволяет строить нужный нам затравочный граф без перебора.

Пусть q – простое число. Рассмотрим граф AP_q, вершинами которого являются все пары $(a, b) \in \mathbb{Z}^2$, а рёбрами соединены такие вершины (a, b), (c, d), что

$$ac = b + d \ mod \ q$$

Полезно представлять себе пару (a, b) кук точку на аффинной плоскости над $\mathbb{Z}_q$, а (c, d) — как прямую, задаваемую уравнением $y = cx - d$, которая проходит через эту точку.

Таком образом, граф состоит из q^2 вершин, и степень каждой вершины равна q. Покажем, что второе по абсолютной величине собственное число графа равно $\sqrt{q}$.

Можно заранее догадаться, что данный граф обладает хорошими свойствами перемешивания. В самом деле, вторая степень этого графа (соответствующая блужданию по графу AP_q по путям длины два) очень близка к полному перемешиванию. Поэтому удобно произвести спектральный анализ не для самого AP_q, а для его квадрата.

Обозначим M матрицу графа AP_q. Будем считать, что вершины (a, b) нумеруются сначала по первой, а потом по второй координате. Таким образом, матрица M состоит из q^2 квадратных блоков размера $q \times q$; в каждом таком блоке (i-ом по горизонтали, j-ом по вертикали) ребра соответствуют переходу из вершин вида $(i, *)$ в вершины $(j, *)$.

Матрицу M^2 легко выписать в явном виде. Действительно, M^2 описывает пути длины 2 на AP_q. Если $i \neq j$, то есть ровно один такой путь из (i, k) в (j, l) (поскольку на плоскости есть ровно одна прямая, которая проходит через точки (i, k) и (j, l)). Если $k \neq l$, то из (i, k) в (i, l) нет путей длины два (мы не рассматриваем вертикальные прямые на плоскости). Наконец, из для каждой вершины (i, k) имеется q циклов длины два.

Таким образом, матрица M^2 имеет вид

$$\begin{pmatrix} qI & J & J & \dots & J \\ J & qI & J & \dots & J \\ \dots & \dots & \dots & \dots & \dots \\ J & J & J & \dots & qI \end{pmatrix}$$

где I — диагональная единичная матрица $q \times q$, а J_q — матрица $q \times q$, в которой на всех местах стоят единицы.

В тензорных обозначениях это можно записать так:

$$M^2 = I_{q\times q} \otimes (qI_{q\times q}) + (J_{q\times q} - I_{q\times q}) \otimes J_{q\times q}.$$

У матрицы $I_{q\times q}$ все собственные числа равны единице; у $J_{q\times q}$ есть собственное число 1 кратности один и собственное число 0 кратности $(q-1)$. Несложный подсчёт показывает, что у M^2 спектр состоит из чисел q^2 (кратность 1), 0 (кратность $(q-1)$) и q (кратность $q(q-1)$). Следовательно, у самой матрицы M второе собственное число равно $\sqrt{q}$.

Зафиксируем простое число q и рассмотрим следующую последовательность графов:

$$\begin{aligned} AP^1 &= AP_q \otimes AP_i \\ AP^{k+1} &= AP^k \textcircled{z} AP_q \end{aligned}$$

По свойству зигзаг-произведения, AP^k является спектральным экспандером с параметрами $(q^{2(k+1)}, q^2, O(\frac{i}{\sqrt{q}}))$. Таком образом, при $k = 7$ (для достаточно больших q) мы получаем граф, который можно брать в качестве графа H в нашей основной конструкции явно заданных экспандеров.

4.6 Вторая спектральная оценка для зигзаг-произведения

В этой главе мы докажем оценку для второго собственного числа спектрального произведения в предположении, что в исходных графах второе собственное число *хотя бы немного* отделено от первого.

Теорема 10 *Зигзаг-произведение спектрального* $(n, D, 1-\alpha)$*-экспандера* G *и спектрального* $(D, d, 1-\beta)$*-экспандера* H *является спектральным экспандером с параметрами* $(nD, d^2, \leqslant 1-\alpha\beta^2)$.

Замечание: Если α и β достаточно малы, то оценка из теоремы 9 становится бессмысленной, поскольку сумма $(1-\alpha)+(1-\beta)+(1-\beta)^2$ будет больше единицы.

Лемма 7 *Пусть* A *— матрица блуждания по некоторому графу (первое собственное число равно* 1*), и все остальные собственные числа по модулю не превосходят* $1-\delta$*. Тогда* A *можно представить в виде* $(1-\delta)B + \delta J$*, где* B *— матрица с нормой не больше* 1*, а* J *— матрица полного перемешивания (все матричные элементы равны* $1/[$*число вершин*$]$*).*

Доказательство: вычитая из A матрицу δJ, мы уменьшаем первое собственное число (единицу) на δ, а остальные не трогаем, так что все собственные числа становятся не больше $1-\delta$ по модулю. Таким образом, у разности $(A-\delta J)$ норма не превосходит $(1-\delta)$, и лемма доказана.

Доказательство теоремы: Как и в доказательстве теоремы 9, мы представим матрицу полученного зигзаг-произведения в виде $\tilde{H}\tilde{G}\tilde{H}$, где матрица $\tilde{H}$ (размера $nD \times ND$) есть матрица графа, составленного из рёбер, соответствующих графу H (граф из пунктирных линий на стр. 46), а $\tilde{G}$ есть матрица из рёбер, соответствующих графу G (граф с теми же вершинами, но с рёбрами из сплошных линий на стр. 46). Напоним, что в $\tilde{H}$ каждый столбец и каждая строка содержат d единиц, а в $\tilde{G}$ — ровно одну единицу. Важно помнить, что матрица $\tilde{G}$ задаёт перестановку на множестве вершин, и её норма равна единице.

Воспользуемся леммой 7 и представим матрицу $\tilde{H}$ в виде $\beta\tilde{J}+(1-\beta)B$, где $\tilde{J}$ есть сумма матриц «полного перемешивания» для каждого из n «облаков» нашего зигзаг-произведения, а B — некоторая матрица с нормой, не превосходящей 1. Таким образом, матрица графа оказывается представлена а виде

$$\tilde{H}\tilde{G}\tilde{H} = (\beta\tilde{J}+(1-\beta)B)\cdot\tilde{G}\cdot(\beta\tilde{J}+(1-\beta)B).$$

Раскрывая скобки, преобразуем матрицу графа в сумму

$$\beta^2\tilde{J}\cdot\tilde{G}\cdot\tilde{J} + (1-\beta^2)C,$$

где C — некоторая матрица с нормой, не превосходящей 1.

Теперь рассмотрим более внимательно произведение $\tilde{J} \cdot \tilde{G} \cdot \tilde{J}$. Эта матрица соответствует следующему трёхшаговому блужданию по графу: начав с некоторой вершины v, мы переходим к случайно (равномерно) выбранной вершине v' в том же облаке (умножение на $\tilde{J}$), затем от v' переходим по «сплошному» ребру в вершину v'' в соседнем облаке (умножение на $\tilde{G}$), и затем ещё раз переходим к некоторой v'' — случайно выбранной соседке v' по облаку (ещё одно умножение на $\tilde{J}$). Заметим, что описанное блуждание совпадает можно описать умножением на матрицу $J \otimes G$. В самом деле: мы переходим из текущего облака в соседнее по случайно выбранному сплошному ребру, а затем выбираем случайную координату внутри нового облака. Но второе собственное число матрицы $J \otimes G$ легко вычислить: оно равно второму собственному числу G, т.е., $(1-\alpha)D$. Следовательно, второе собственное число $\tilde{H}\tilde{G}\tilde{H}$ не превосходит

$$\beta^2(1-\alpha) + (1-\beta^2) = 1 - \alpha\beta^2,$$

и теорема доказана.

4.7 Подстановочное произведение*.

В перегрaфе 4.2 мы определили зигзаг-произведение на графах, которое оказалось полезным для получения явных конструкций экспандеров. В этой главе мы введём ещё два аналогичных вида произведения на графах.

Для графа G (с n вершинами, степени D) и графа H (с D вершинами, степени d) мы определим простое и сбалансированное *подстановочное произведение* (как и в определении зигзаг-произвдения, важно, что степень первого графа равна числу вершин во втором графе). Начало конструкции совершенно аналогично определению зигзаг-произведения: мы заменим каждую вершину графа G копией графа H, прикрепив рёбра первого графа к вершинам второго. При этом у нас получится граф с nD вершинами и рёбрами двух типов — большими из первого графа и малыми из второго (см. рисунок на стр. 46: рёбра первого типа показаны сплошными, а рёбра второго типа — пунктирные линии.)

Простое подстановочное произведение (replacement product): Построенный выше граф в точности и есть простое подстановочное произведение G и H (и «сплошные» и «пунктирные» рёбра на равных правах включаются в новый граф). В полученном графе nD вершин (n «облаков» по D вершин в каждой); степень каждой вершине равна $d+1$ (из каждой вершины выходит одно сплошное и d пунктирных рёбер).

Сбалансированное подстановочное произведение (balanced replacement product): Отличие состоит лишь в том, что мы берём каждое сплошное ребро с кратностью d. таким образом, в полученном графе-произведении из каждой вершины выходит $2d$ рёбер: d сплошных и d пунктирных.

Оценим второе собственное число для сбалансированного подстановочного произведения. Мы будем оценивать не их малость второго собственного числа, а его отделённость от единицы.

Теорема 11 *Пусть графы G и H являются спектральными экспандерами с параметрами $(n, D, 1-\alpha)$ и $(D, d, 1-\beta)$ соответственно. Тогда их сбалансированное подстановочное произведение $G\textcircled{r}H$ имеет параметры $(nD, 2d, 1-\alpha\beta^2/24)$.*

Доказательство: Удобно описывать происходящее в терминах блужданий (соответствующих нормализованным матрицам, полученных делением матрицы смежности на степень графа). Блуждание по взвешенному произведению является полусуммой двух блужданий: *локального*, где мы движемся внутри одном «облаке» в соответствии с матрицей графа H, и *глобального*, где мы движемся по рёбрам графа G (а выбор ребра определяется текущей H-координатой: вершин в графе H как раз столько, сколько рёбер в G, и мы предполагаем, что фиксировано какое-то соответствие). Таким образом, матрицу блуждания можно записать как

$$U = \frac{1}{2}\hat{G} + \frac{1}{2}\hat{H},$$

где $\hat{G}$ и $\hat{H}$ — матрицы перехода по «локальным» и «глобальным» рёбрам. Чтобы оценить второе собственное число U, достаточно оценить второе собственное число $U^3 = (\frac{1}{2}\hat{G} + \frac{1}{2}\hat{H})^3$ и доказать, что оно не больше $1 - \varepsilon\delta^2/8$ (а затем воспользоваться неравенством Бернулли).

В разложении для U^3 будет восемь членов. Все эти члены имеют два инвариантных подпространства: одномерное — векторы, у которых все координаты равны (все восемь членов на этом подпространстве единичные, и при каждом стоит коэффициент 1/8), и ортогональное к нему (векторы, сумма координат которых равна нулю), где максимальное собственное значение (и тем самым норма ограничения на это подпространство) и есть интересующий нас параметр. Если бы мы доказали, что для одного из этих восьми произведений второе собственное число не больше $1 - \alpha\beta^2$, то это бы гарантировало, что для U^3 это второе собственное число не больше $1 - \alpha\beta^2/8$, поскольку у оставшихся семи произведений норма не больше 1.

Какое из восьми слагаемых выбрать? Кажется, что наилучшие шансы на перемешивание у $\hat{H}\hat{G}\hat{H}$ (сначала перемешиваем внутри облака с помощью графа H, потом идём по ребру большого графа, потом перемешиваем внутри другом облаке — как в зизаг-произведении). Если бы перемешивание внутри облака было полным (переход в случайную точку облака), то такой переход был бы переходом в случайную вершину случайной соседнем облаке. Соответствующее преобразование является тензорным произведением G и полного перемешивания, и потому имеет второе собственное число $1 - \varepsilon$, как у G.

Это много лучше, чем нам нужно (мы получили $1-\alpha$ вместо $1-\alpha\beta^2/8$), что и не удивительно: мы волюнтаристски заменили перемешивание вдоль

H полным перемешиванием. Но поскольку переход по ребрам H не является полным перемешиванием, нам придётся несколько усложнить рассуждение. (При этом вместо $1-\alpha$ мы получим для второго собственного числа более скромную оценку $1-\alpha\beta^2/8$).

Применим Лемму 7 к блужданию по графу H и разложим его в сумму $(1-\beta)A+\beta B$. Это разложение можно провести в каждом облаке и получить разложение $\hat{H} = (1-\beta)\hat{A} + \beta\hat{B}$, где $\hat{A}$ — некоторый оператор с нормой не больше 1, а B — то самое полное перемешивание внутри облаков, о котором мы говорили выше.

Повторяем прежнее рассуждение. Теперь в разложении

$$U^3 = \left(\frac{1}{2}\hat{G} + \frac{1-\beta}{2}\hat{A} + \frac{\beta}{2}\hat{B}\right)^3$$

будет уже не 8, а 27 слагаемых. В одном из этих слагаемых (а именно, в $\hat{B}\hat{G}\hat{B}$) второе собственное значение не превосходит $1-\alpha$, а остальные представляют собой операторы с нормой не больше 1 с некоторыми скалярными коэффициентами (сумма коэффициентов при этих 27 слагаемых равна единице). Поэтому второе собственное значение U не больше $1-\alpha\beta^2/8$. Теорема Доказана

Применим полученные оценки чтобы описать ещё одну явную конструкцию спектральных экспандеров. Прежде всего мы выберем спектральный экспандер H с параметрами $(d^{50}, d/2, 1/100)$, а также спектральные экспандеры G_1 и G_2 с параметрами $(d^{100}, d, <1/2)$ и $(d^{200}, d, <1/2)$ соответственно (такие графы существуют для всех достаточно больших d). Далее построим последовательность

$$G_n = \left(G_{\lfloor (n-1)/2\rfloor} \otimes G_{\lceil (n-1)/2\rceil}\right)^{50} \textcircled{r} H,$$

Упражнение 26 *Проверьте, что G_n является спектральным экспандером с параметрами $(d^{100n}, d, <1-1/50)$; функция вращения для такого графа G_n вычисляется за полиномиальное от n время.*

4.8 Комбинаторная оценка для подстановочного произведения*

В этом разделе мы покажем, что экспандерные свойства (рёберное расширение) сбалансированного подстановочного произведения можно оценить с помощью прямого комбинаторного рассуждения, без использования спектральной техники.

Теорема 12 *Пусть даны два однородных графа:*

- *граф G степени D с n вершинами, с коэффициентом рёберного расширения $h_E(G) \geqslant \alpha$,*

- *граф H степени d с D вершин, с коэффициентом рёберного расширения $h_E(G) \geqslant \beta$.*

Тогда сбалансированное подстановочное произведение этих графов $G\textcircled{r}H$ будет графом степени $2d$ с nD вершинами и с коэффициентом рёберного расширения не меньше $\frac{1}{64}\alpha^2\beta$.

Доказательство: Рассмотрим в построенном графе $G\textcircled{r}H$ произвольное множество вершин S размера не более $nD/2$ (т.е. не более половины всех вершин графа). Напомним, что в конструкции подстановочного произведения возникают два типа рёбр — «сплошные» (они проводятся между облаками по D вершин) и «пунктирные» (внутри каждого облака). Мы хотим показать, что из множества S в его дополнение $\bar{S} = V\backslash S$ ведёт не менее $\frac{1}{64}\alpha^2\beta \cdot (2d)|S|$ рёбер. Мы докажем даже немного более сильное утверждение: нужное нам число рёбер из S в $\bar{S}$ можно набрать, рассматривая либо только сплошные рёбра, либо только пунктирные рёбер.

По определению подстановочного произведение множество всех вершин $G\textcircled{r}H$ состоит из n «облаков» по D вершин в каждом. Обозначим эти облака $V_1, \ldots, V_n$, и

$$S_i := S \cap V_i.$$

Назовём облако V_i *насыщенным* для S, если $|S_i| \geqslant (1 - \frac{\alpha}{4})D$; в противном случае назовём облако *ненасыщенным*. Разделим S на две части: множество $S_{\text{насыщ}}$ (вершины, лежашие в насыщенных облаках) и $S_{\text{ненасыщ}}$ (вершины, лежашие в ненасыщенных облаках).

Первый случай: предположим, что $|S_{\text{ненасыщ}}| \geqslant \frac{\alpha}{8}|S|$ и оценим число рёбер в $E(S, \bar{S})$ в этом предположении. В этом случае мы подсчитаем число пунктирных рёбер, которые ведут из множества S в его дополнение (внутри одних только ненасыщенных облаков — этого будет достаточно, чтобы получить нужную оценку).

Каждое из облаков V_i с пунктирными облаками образует граф, изоморфный H. В ненасыщенном облаке V_i множество S_i занимает не более $(1-\frac{\alpha}{4})D$ вершин, а значит,

$$|V_i\backslash S_i| \geqslant \frac{\alpha}{4}D > \frac{\alpha}{4}|S_i|.$$

Следовательно, число рёбер $E(S_i, V_i\backslash S_i)$ не может быть меньше, чем

$$h_E \cdot d \min\{|S_i|, |V_i\backslash S_i|\} \geqslant \beta \cdot d \cdot \left(\frac{\alpha}{4}|S_i|\right).$$

Суммируя это неравенство по всем ненасыщенным облакам, получим

$$|E(S, \bar{S})| \geqslant \frac{\alpha\beta d}{4} \cdot |S_{\text{ненасыщ}}|.$$

Поскольку мы предположили, что

$$|S_{\text{ненасыщ}}| \geqslant \frac{\alpha}{8}|S|,$$

получаем $|E(S,\bar{S})| \geqslant \frac{\alpha^2\beta}{64} \cdot 2d|S|$, что и требовалось.

Второй случай: теперь предположим, что $|S_{\text{ненасыщ}}| < \frac{\alpha}{8}|S|$ и, соответственно, $|S_{\text{насыщ}}| > (1-\frac{\alpha}{8})|S|$. На этот раз оценим число *сплошных* рёбер, выходящих из вершин S в *насыщенных облаках* и ведущих в вершины $\bar{S}$ в *ненасыщенных облаках*. Разумеется, в графе могут быть и другие рёбра, ведущие из S в $\bar{S}$. Но мы покажем, что одних только рёбер указанного вида найдётся достаточно много.

Из определения насыщенного облака следует, что

$$[\text{число насыщенных облаков}] \leqslant \frac{|S_{\text{насыщ}}|}{(1-\frac{\alpha}{4})D} \leqslant \frac{|S_{\text{насыщ}}|}{\frac{3}{4}D}.$$

Поскольку мы предполагаем, что S содержит не более половины всех вершин графа (т.е. не более $Dn/2$), получаем

$$[\text{число насыщенных облаков}] \leqslant \frac{Dn/2}{\frac{3}{4}D} \leqslant \frac{2}{3}n,$$

и, соответственно,

$$[\text{число ненасыщенных облаков}] \geqslant \frac{1}{3}n.$$

Это значит, что

$$[\text{число ненасыщенных облаков}] \geqslant \frac{1}{2} \cdot [\text{число насыщенных облаков}].$$

Теперь воспользуемся тем, что в исходном графе G коэффициент рёберного расширения равен α. Заключаем, что число сплошных рёбер (с учётом кратности), ведущих из насыщенных облаков в ненасыщенные, не меньше

$$\alpha dD \cdot \min\{[\text{число насыщенных облаков}], [\text{число ненасыщенных облаков}]\} > \\ > \frac{1}{2}\alpha dD \cdot [\text{число насыщенных облаков}].$$

Однако не все эти рёбра ведут из S в $\bar{S}$; нужно исключить из подсчёта два сорта рёбер, которые нам не подходят:

- во-первых, вычтем число сплошных рёбер, у которых один из концов лежит в насыщенном облаке V_i, но не в интересующем нас множестве S_i, а в его дополнении $V_i \backslash S_i$; таких рёбер заведомо не больше

$$\sum_{\text{насыщенные облака } V_i} d \cdot |V_i \backslash S_i| \leqslant \frac{\alpha dD}{4} \cdot [\text{число насыщенных облаков}];$$

- во-вторых, вычтем число сплошных рёбер, у которых один из концов лежит в ненасыщенном облаке V_i, но при этом попадает в множество $S_i \subset V_i$; таких рёбер заведомо не больше

$$d \cdot |S_{\text{ненасыщ}}| < d \cdot \frac{\alpha}{8}|S| \leqslant d \cdot \frac{\alpha}{8} \cdot \frac{|S_{\text{насыщ}}|}{1 - \frac{\alpha}{8}} \leqslant$$
$$\leqslant \frac{\alpha d D}{7} \cdot [\text{число насыщенных облаков}].$$

После вычитания остаётся не меньше

$$(\frac{1}{2} - \frac{1}{4} - \frac{1}{7}) \cdot \alpha d D \cdot [\text{число насыщенных облаков}] \geqslant$$
$$\geqslant \frac{3}{28} \cdot \alpha d D \cdot [\text{число насыщенных облаков}]$$

сплошных рёбер, которые ведут из *вершин S в насыщенных облаках* в *вершины в дополнении S в ненасыщенных облаках*. Понятно, что все эти рёбра заведомо лежат в $E(S, \bar{S})$, и их число не может быть меньше

$$\frac{3\alpha d D}{28} \cdot \frac{|S_{\text{насыщ}}|}{D} > \frac{3\alpha d}{28} \cdot (1 - \frac{\alpha}{8})|S|,$$

что с большим запасом больше нужной нам оценки $\frac{1}{64}\alpha^2\beta \cdot 2d|S|$.

Глава 5

Экспандеры на группах

5.1 Графы Кэли: определение и примеры

В этой главе мы определим графы Кэли и приведём примеры спектрального анализа таких графов.

Пусть G — произвольная группа, а $S \subset G$ — симметричное множество элементов группы (если $h \in S$, то $h^{-1} \in S$). Графом Кэли (G, S) называется граф, вершинами которого являются все элементы группы G; вершины v и w соединяются (неориентированным) ребром, если $v = wh$ для некоторого $h \in S$. Поскольку множество S симметрично, данное определение корректно (если $v = wh$ для некоторого $h \in S$, то $w = vh^{-1}$).

Из определения немедленно следует, что степень каждой вершины в графе Кэли равна числу элементов в S. При этом в графе Кэли не может быть кратных рёбер. Петли в графе Кэли имеются (причем одновременно у всех вершин), если единичный элемент группы принадлежит S.

Пример 1. G — произвольная группа, $S = G$. Графом Кэли (G, S) будет полный граф с $|G|$ вершинами (с петлями).

Пример 2. $G = \mathbb{Z}_n$ (группа вычетов по модулю n с операцией сложения), $S = \{1, -1\}$. Графом Кэли (G, S) будет цикл длины n.

Пример 3. $G = \mathbb{Z}_2^k$; S состоит из естественных образующих группы: $S = \{e_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0),\ i = 1, \ldots, n\}$ (у e_i единица стоит в позиции номер i; остальные координаты нулевые). Графом Кэли (G, S) будет граф рёбер n-мерного гиперкуба.

Для спектрального анализа графа Кэли полезно рассмотреть неприводимые представления группы G. Для конечных абелевых групп нужно изучить *характеры* G.

Определение 6 *Характерами группы G называют гомоморфизмы в мультипликативную группу комплексных чисел $\xi : G \to \mathbb{C}^*$.*

Напоминание из курса алгебры (свойства характеров):

- У каждой группы есть *тривиальный* характер ξ, тождественно равный единице.
- Если ξ_1, ξ_2 характеры абелевой G, то их покомпонентное произведение $\xi_1(g)\xi_2(g)$ тоже является характером; комплексное сопряжения характера также будет характером.
- У всякой конечной абелевой группы, состоящей из n элементов, имеется ровно n характеров.
- Если ξ характер группы G, то

$$\sum_{g\in G} \xi(g) = \begin{cases} |G|, & \text{если } \xi \text{ тривиален,} \\ 0, & \text{иначе.} \end{cases}$$

- Если ξ_1 и ξ_2 — два характера G, то

$$\sum_{g\in G} \xi_1(g) \cdot \overline{\xi_2(g)} = \begin{cases} |G|, & \text{если } \xi_1 \text{ и } \xi_2 \text{ совпадают,} \\ 0, & \text{иначе,} \end{cases}$$

  т.е., характеры группы попарно ортогональны.

Теорема 13 *Пусть $G = \{a_1, \ldots, a_n\}$ — конечная абелева группа, $S \subset G$ — её симметричное подмножество, ξ — один из характеров группы. Тогда вектор $(\xi(a_1), \ldots, \xi(a_n))$ является собственным для матрицы графа Кэли (G, S); соответствующее этому вектору собственное число равно*

$$\sum_{h\in S} \xi(h).$$

Замечание: В данном случае мы рассматриваем матрицу графа как линейный оператор над полем комплексных чисел. Так что координаты собственных векторов, которые мы вычислим, могу быть, вообще говоря, комплексными. Однако все собственные числа окажутся действительными числами. Нам это известен заранее — собственные числа всякой матрицы графа обязаны быть действительными числами, поскольку такая матрица симметрична. Для графов Кэли данный факт можно независимо доказать с помощью Теоремы 13. В самом деле, поскольку множество S симметрично, для каждого характера ξ сумма $\sum\limits_{h\in S} \xi(h)$ является сопряженной к самой себе, т.е., будет действительным числом.

Доказательство теоремы: Подействуем на вектор $(\xi(a_1), \ldots, \xi(a_n))$ матрицей M графа Кэли. Вычислим значение в i-ой координате полученного в результате вектора. Понятно, что там должна стоять сумма величин $\xi(a_j)$ по всем a_j, которые соединены ребром с a_i. Это значит, что a_j получается из a_i умножением на некоторый элемент из S. Таким образом, i-ая координата $M \cdot (\xi(a_1), \ldots, \xi(a_n))^{\perp}$ равна

$$\sum_{h\in S} \xi(a_i h) = \xi(a_i) \cdot \sum_{h\in S} \xi(h)$$

Тем самым, теорема доказана.

Теперь мы применим доказанную теорему, чтобы найти спектр нескольких графов Кэли.

Возвращение к Примеру 2. Рассмотрим более подробно граф из Примера 2 на стр. 59. Характер группы $\mathbb{Z}_n$ однозначно определяется его значением на элементе 1 (при этом характер должен отобразить элементы группы в корни из единицы степени n). Таким образом, мы получаем n характеров ξ_k, определяемых условием

$$\xi_k(1) = e^{2\pi ki/n}.$$

Соответственно, собственные числа графа равны

$$\lambda_k = \xi_k(1) + \xi_k(-1) = 2\cos(2\pi k/n),$$

$k = 0, 1, \ldots, n-1$.

Мы видим, что у графа имеется собственное число $\lambda_0 = 2$; это неудивительно — цикл является графом степени два, так что число 2 должно быть его собственным числом. Если n чётно, то $\lambda_{n/2} = 2\cos(\pi) = -2$; это согласуется с тем, что цикл четной длины является двудольным графом.

Если же n нечётно, то второе по абсолютной величине собственное число будет равно $\lambda_{\frac{n-1}{2}} = \lambda_{\frac{n+1}{2}}$. Отметим, что зазор между первым и вторым собственным числом невелик — второе по абсолютной величине собственное число равно $2 \cdot (1 - O(1/n^2))$.

Возвращение к Примеру 3. Теперь изучим граф из Примера 3 на стр. 59. Характеры группы $\mathbb{Z}_2^n$ однозначно определяются значениями на образующих элементах группы e_i (причём $\xi(e_i)$ может быть равно 1 или -1). В данном случае характеры естественно индексировать строками из n битов $b_1 \ldots b_n$; мы имеем 2^n различных характеров $\xi_{b_1 \ldots b_n}$:

$$\xi_{b_1 \ldots b_n}(a_1, \ldots, a_n) = \prod_{i=1}^{n} (-1)^{a_i b_i}$$

Собственные числа этого графа Кэли равны

$$\lambda_{b_1 \ldots b_n} = [\text{число нулей в строке } b_1 \ldots b_n] - [\text{число единиц в строке } b_1 \ldots b_n]$$

т.е., собственными числами будут значения $n, n-2, n-4, \ldots, -n$ (кратности собственных чисел будут равны соответствующему биномиальному коэффициенту).

Мы видим, что максимальное собственное число данного графа равно степени графа n (степень каждой вершины равна n); имеется собственное число $-n$ (граф двудольный); следующее по абсолютной величине собственное число равно $n \cdot (1 - 2/n)$.

Упражнение 27 *Опишите графы Кэли для следующих групп:*

(а) $G = \mathbb{Z}^6$, $S = \{3\}$;
(б) $G = \mathbb{Z}^6$, $S = \{2, -2\}$;
(в) G *есть группа симметрий квадрата, а* S *состоит из двух элементов: симметрии относительно вертикальной оси квадрата и симметрией относительно главной диагонали.*

Упражнение 28 *Покажите, что оценка коэффициента вершинного расширения для спектрального экспандера* $h_V(G) \geqslant \frac{d-\lambda(G)}{2}$ *из Следствия 2 (стр. 24) является точной (в общем случае константу* 2 *нельзя уменьшить). Указание: подберите граф Кэли, для которого данная оценка оказывается точной.*

5.2 Линейное пространство как экспандер*

В этом разделе мы пишем конструкцию, напоминающую экспандер на плоскости из раздела **??**. Мы покажем, как построить экспандер из конечномерного линейного пространства над конечным полем.

Пусть $\mathbb{F}$ есть поле из $q = 2^t$ элементов. Рассмотрим d-мерное линейное пространство $\mathbb{F}^d$ над этим полем. Точки этого пространства будут вершинами графа. Нам будет удобно представлять эти точки в координатном виде, как $(a_0, a_1, \ldots, a_{d-1}) \in \mathbb{F}^d$.

Мы соединяем рёбрами каждую вершину $(a_0, a_1, \ldots, a_{d-1})$ со всеми вершинами вида

$$(a_0, a_1, \ldots, a_{d-1}) + (y, xy, x^2y, \ldots, x^{d-1}y).$$

Таким образом, степень каждой вершина равна q^2, и выходящие из каждой вершины рёбра естественным образом индексируются парами $(x, y) \in \mathbb{F}^2$. Заметим, что определение симметрично (не нужно отдельно учитывать рёбра обратные к уже описанным). Обозначим описанный граф $LS(q, d)$.

Теорема 14 *Граф* $LS(q, d)$ *является спектральным* $(q^d, q^2, (d-1)/q)$*-экспандером.*

Доказательство: Заметим, что описанный граф является графом Кэли. Группой в данном случае будет линейное пространство $\mathbb{F}^d$ с обычной операцией сложения, а в качестве S мы возьмём множество всех векторов

$$\mathbf{s}_{x,y} = (y, xy, x^2y, \ldots, x^{d-1}y),\ x, y \in \mathbb{F}.$$

Снова отметим, что данное S симметрично (над полем характеристики 2 каждый вектор x обратен сам себе).

Зафиксируем произвольное линейное отображение,

$$L\ : \mathbb{F} \to \mathbb{Z}_2$$

не равное тождественно нулю (такое отображение существует для любого поля характеристики 2). Теперь мы можем описать q^n попарно ортогональных характеров аддитивной группы $\mathbb{F}^d$. Характеры данной группы мы будем индексировать наборами $(c_0, c_1, \ldots, c_{d-1})$ из $\mathbb{F}^d$:

$$\xi_{c_0,\ldots,c_{d-1}}(x_0, \ldots, x_{d-1}) := (-1)^{L(\sum_i c_i x_i)}$$

(ср. с характерами группы $\mathbb{Z}_2^n$ из примера 3 на с. 59). Отметим, что значениями всех характеров данной группы могут быть только ± 1 (это свойство было нетрудно предсказать заранее, поскольку каждый элемент группы обратен самому себе).

С помощью теоремы 13 мы можем описать собственные векторы матрицы графа $LS(q, d)$

$$\mathbf{v}_{c_0,\ldots,c_{d-1}} = \big(\xi_{c_0,\ldots,c_{d-1}}(\mathbf{x})\big)_{\mathbf{a}\in\mathbb{F}^d} \tag{5.1}$$

(координаты собственных векторов, как и вершины графа, нумеруются элементами $\mathbb{F}^d$), а также собственными числа матрицы

$$\lambda_{c_0,\ldots,c_{d-1}} = \sum_{(x,y)\in\mathbb{F}^2} \xi_{c_0,\ldots,c_{d-1}}(y, xy, x^2y, \ldots, x^{d-1}y).$$

Перепишем выражения для собственных векторов в более явном виде:

$$\lambda_{c_0,\ldots,c_{d-1}} = \sum_{(x,y)\in\mathbb{F}^2} (-1)^{L(y\cdot p_{\mathbf{c}}(x))}, \text{ где } p_{\mathbf{c}}(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_{d-1}x^{d-1}.$$

Заметим, что

$$\lambda_{c_0,\ldots,c_{d-1}} = \sum_{(x,y)\ :\ p_{\mathbf{c}}(x)=0} (-1)^{L(y\cdot p_{\mathbf{c}}(x))} + \sum_{(x,y)\ :\ p_{\mathbf{c}}(x)\neq 0} (-1)^{L(y\cdot p_{\mathbf{c}}(x))} \tag{5.2}$$

Если $p_{\mathbf{c}}(x) = 0$, то значение $(-1)^{L(y\cdot p_{\mathbf{c}}(x))} = 1$ для всех y, так что каждое такое значение x даёт вклад в сумму, равный q. Если же $p_{\mathbf{c}}(x) \neq 0$, то произведение $y \cdot p_{\mathbf{c}}(x)$ пробегает все q элементов поля, и среди соответствующих $(-1)^{L(y\cdot p_{\mathbf{c}}(x))}$ встречается равное число $+1$ и -1. Таким образом, общий вклад этих слагаемых в сумму (5.2) равен нулю.

Таким образом, $\lambda_{c_0,\ldots,c_{d-1}} = q \cdot [\text{число нулей многочлена } p_{\mathbf{c}}(x)]$. Тривиальный многочлен (все коэффициенты которого равны нулю) тождественно равен нулю во всех точках поля; это соответствует тому, что

$$\lambda_{0,\ldots,0} = q^2.$$

У всех остальных многочленов $p_{\mathbf{c}}(x)$ число нулей не превосходит $d-1$. Следовательно, все собственные числа кроме одного не превосходят $(d-1)q$. Теорема доказана.

Упражнение 29 *Дайте прямое доказательство того, что векторы (5.1) попарно ортогональны и являются собственными векторами матрицы графа.*

Упражнение 30 *Докажите, что для любого $\varepsilon > 0$, для всякого целого r и всех достаточно больших $q = 2^t$ граф $LS(q, r)$ имеет коэффициент рёберного расширения не меньше $\frac{1}{2} - \varepsilon$.*

Граф $LS(q = 2^t, d)$ даёт пример простой и алгоритмически эффективной конструкции экспандера с хорошей оценкой для второго собственного числа. К сожалению, у этого графа степень не ограничена: если мы хотим поддерживать для второго собственного числа оценку $(r-1)/q < \delta$ (для некоторой константы δ), то с ростом числа вершин приходится увеличивать значение q, а значит и степень графа q^2.

Однако из графов $LS(q, d)$ можно построить экспандеры с ограниченной степенью, если воспользоваться подстановочным произведением. При этом не требуется сложная рекурсивная конструкция — подстановочное произведение достаточно применить лишь дважды! Ниже мы опишем данную конструкцию.

Пусть $q = 2^t$, $r = \Theta(q^4)$, и $n = q^{4r}$ (обратим внимание, что для выбранных параметров $q = O(\sqrt{n})$). Мы опишем явную (в сильном смысле) конструкцию экспандера с n вершинами, со степенью $O(1)$ и коэффициентом рёберного расширения не меньше некоторого $\delta > 0$. Конструкция будет использовать в качестве «строительных блоков» следующую тройку графов:

- G_1: граф степени 3 с q^2 вершинами, с коэффициентом рёберного расширения $> \delta'$ для некоторой абсолютной константы $\delta' > 0$ (мы знаем, что такой экспандер существует, см. упражнение **??** на с. **??**, и найти такой граф можно перебором за время $q^{O(q^2)} = \mathrm{poly}(n)$),
- G_2: граф $LS(q, 6)$; в этом графе q^6 вершин, степень равна q^2, коэффициент рёберного расширения не меньше $1/4$,
- G_3: граф $LS(q^4, r-8)$; в этом графе q^{4r-8} вершин, степень равна q^8, коэффициент рёберного расширения не меньше $1/4$.

Теперь рассмотрим граф

$$G := G_3 \textcircled{r} (G_2 \textcircled{r} G_1).$$

Из определения сбалансированного подстановочного произведения следует, что мы получили граф с $n = q^{4r}$ вершинами степени 12. Теорема 12 гарантирует, что коэффициентом рёберного расширения полученного графа не меньше некоторого числа $\delta > 0$ (не зависящего от n).

Упражнение 31 *Докажите, что построенный граф является спектральным $(n = q^{4r}, 12, < \gamma)$-экспандером для некоторой константы $\gamma > 0$, не зависящей от n.*

5.3 Графы Рамануджана*

Напомним, что в любом регулярном графе степени d второе по абсолютной величине собственное число не может быть меньше $2\sqrt{d-1} - o(1)$, см. параграф 3.6. В то же вермя, для большинства графов второе собственное значение очень близко к этой границе (см. обсуждение в конце параграфа 3.8). Понятно, что d-регулярные графы, которые у которых второе по абсолютной величине собственное число ниже границы $2\sqrt{d-1}$, заслуживают особого внимания — это экспандеры с максимальным возможным спектральным зазаором. Такие графы называют *графами Рамануджана*.

Любоцкий, Сарнак, Филлипс и Маргулис указали явную (и алгоритмически эффективную) конструкцию графов Кэли, являющихся графами Рамануджана. Таким образом, была получена эффективная (в «сильном» смысле) конструкция спектрального экспандера практически с наилучшими возможными параметрами. Ниже мы опишем эту конструкцию (без доказательства оценки для второго собственного числа).

Пусть p и q простые числа, $p = 1 \ mod \ 4$ и $q = 1 \ mod \ 4$. В качестве группы G возьмём $PGL(2, \mathbb{Z}/q\mathbb{Z})$, т.е., невырожденные матрицы 2×2 над полем вычетов по модулю q, профакторизованные по отношению пропорциональности (с обычной операцией матричного умножения).

Далее мы зададим в этой группе симметричное множество S. Выберем такое целое i, что $i^2 = -1 \ mod \ q$. Можно доказать, что имеется ровно $(p+1)$ целочисленное решение уравнения

$$a_0^2 + a_1^2 + a_2^2 + a_3^2 = p$$

такое, что a_0 положительно и нечётно, а a_1, a_2, a_3 чётны. Каждой такой четвёрке сопоставим матрицу

$$A = \begin{pmatrix} a_0 + ia_1 & a_2 + ia_3 \\ -a_2 + ia_3 & a_0 - ia_1 \end{pmatrix}$$

Эти матрицы и образуют множество S.

Нетрудно понять, что граф Кэли (G, S) состоит из $\Theta(q^3)$ вершин, и степень каждой вершины равна $(p + 1)$. Свойства данного графа зависят от соотношения p и q. Рассмотрим случай, когда p является квадратичным вычетом по модулю q. Тогда полученный граф Кэли состоит из двух связных компонент (в одной компоненте лежат матрицы, определитель которых является квадратичным вычетом, в другой — матрицы с определителем, являющимся квадратичным невычетом по модулю q). Обозначим $X^{p,q}$ связную компоненту полученного графа. Можно доказать, что у $X^{p,q}$ второе по абсолютной величине собственное число не превосходит $2\sqrt{p}$, т.е. мы получили граф Рамануджана. Однако доказательство этого факта непросто и использует сложную алгебраическую технику, см. [3].

5.4 Экспандер Маргулиса*

В этой главе мы изложим ещё одну явную конструкцию однородного экспандера. Граф в этой конструкции определяется чрезвычайно просто, и его удобно использовать на практике. Данная конструкция является незначительной модификацией исторически первой явной конструкции экспандера, предложенной Г.А. Маргулисом в начале 1970-х, [28]. Однако доказательство оценки второго собственного числа значительно отличается от оригинального доказательства Маргулиса.

Предлагаемая техника доказательства представляет самостоятельный интерес. Она использует преобразование Фурье. Идея доказательства была предложена Габбером и Галилом [21], а затем упрощена в [22] . Мы следуем изложению доказательства из [1]. К сожалению, не смотря на все упрощения, доказательство содержит некоторый «магический трюк», который затрудняет перенос этого рассуждения на другие конструкции экспандеров.

5.4.1 Метод преобразования Фурье (напоминание)

Напомним, что *характерами* группы называют гомоморфизмы из этой группы в мультипликативную группу комплексных чисел. В этой главе нас будут интересовать характеры группы $\mathbb{Z}_n^2$, т.е., отображения

$$\xi \ : \ \mathbb{Z}_n^2 \to \mathbb{C}$$

такие, что

$$\xi(a_1, a_2) \cdot \xi(b_1, b_2) = \xi(a_1 + b_1 \mod n, a_2 + b_2 \mod n)$$

для любых (a_1, a_2) и (b_1, b_2). Для группы $\mathbb{Z}_n^2$ существует ровно n^2 характеров (столько же, сколько элементов в группе), и эти характеры имеют простое описание. А именно, для каждой пары чисел k_1, k_2 (от 0 до $n-1$) имеется характер ξ_{k_1,k_2}, задаваемый формулой

$$\xi_{k_1,k_2}(x, y) = e^{2\pi k_1 x i/n} \cdot e^{2\pi k_2 y i/n}.$$

Введём на функциях $f\ \mathbb{Z}_n^2 \to \mathbb{C}$ скалярное произведение,

$$\langle f, g \rangle := \sum_{\mathbf{v} \in \mathbb{Z}_n^2} f(\mathbf{v}) \cdot \overline{g(\mathbf{v})}$$

(здесь черта обозначает комплексное сопряжение). Напомним, что в смысле данного скалярного произведения все характеры группы $\mathbb{Z}_n^2$ попарно ортогональны. Отсюда следует, что характеры образуют базис в линейном пространстве комплекснозначных функций на $\mathbb{Z}_n^2$. Точнее, имеет место следующее утверждение.

Утверждение 7 *Всякая функция $f \ : \ \mathbb{Z}_n^2 \to \mathbb{C}$ однозначно представляется в виде*

$$f(\mathbf{v}) = \sum_{(k_1,x_2) \in \mathbb{Z}_n^2} \hat{f}(k_1, k_2) \xi_{k_1,k_2}(\mathbf{v}),$$

где $\hat{f}(k_1, k_2)$ *— некоторые комплексные коэффициенты.*

Коэффициенты данного представления можно вычислить по формуле

$$\hat{f}(k_1, k_2) = \langle \xi_{k_1,k_2}, f \rangle = \frac{1}{n^2} \sum_{(z_1,z_2)} \xi_{k_1,k_2}(z_1, z_2) \cdot \overline{f(z_1, z_2)}.$$

Набор коэффициентов $\hat{f}(k_1, k_2)$ из Утверждения 7 можно рассматривать как функцию их $\mathbb{Z}_n^2$ в $\mathbb{C}$. Эту функцию называют *преобразованием Фурье* исходной функции f.

Известно, что для преобразования Фурье выполняются следующие свойства:

(а) $\sum_{k_1,k_2} f(k_1, k_2) = 0$, если и только если $\hat{f}(0, 0) = 0$,

(б) $\langle f, g, \rangle = \frac{1}{n^2} \langle \hat{f}, \hat{g} \rangle$,

(в) $\sum\limits_{k_1,k_2} |f(k_1, k_2)|^2 = \frac{1}{n^2} \sum\limits_{k_1,k_2} |\hat{f}(k_1, k_2)|^2$,

(г) $f(x_1, x_2) = \frac{1}{n^2} \sum\limits_{k_1,k_2} \hat{f}(k_1, k_2) e^{-2\pi k_1 x_1 i/n - 2\pi k_2 x_2 i/n}$,

(д) пусть $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ некоторая матрица линейного преобразования и $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ вектор сдвига, $g(x_1, x_2) = f\left(A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \mathbf{b}\right)$; тогда преобразование Фурье данной функции g можно вычислить по формуле

$$\hat{g}(y_1, y_2) = e^{\frac{2\pi i}{n} \langle A^{-1}\mathbf{b}, (y_1,y_2) \rangle} \cdot \hat{f}\left(\left(A^{-1}\right)^{\perp} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right)$$

Упражнение 32 *Докажите свойства преобразования Фурье (а-д) самостоятельно (либо вспомните эти доказательства, если вы изучали преобразование Фурье в курсе анализа).*

5.4.2 Применение преобразования Фурье для оценки спектрального зазора

Вернёмся к построению экспандера. Определим граф следующим образом. В качестве множества вершин возьмём множество $V = \mathbb{Z}_n \times \mathbb{Z}_n$ (граф будет содержать n^2 вершин). Чтобы описать рёбра графа, определим матрицы

$$T_1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \; T_2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

и векторы сдвига

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Из каждой вершины $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ проведём четыре рёбра в вершины, получающиеся с помощью преобразований $T_1\mathbf{v}$, $T_2\mathbf{v}$, $T_1\mathbf{v}+\mathbf{e}_1$, $T_2\mathbf{v}+\mathbf{e}_2$ и ещё четыре ребра, получающиеся обратными преобразованиями. Таким образом, степень каждой вершины равна 8. (Отметим, что при достаточно больших n в этом графе не будет кратных рёбер.)

Оказывается, что определённый выше граф G является спектральным $(n^2, 8, < 5\sqrt{2})$-экспандером. Доказательство этого результата можно найти в [21]. Мы докажем несколько более слабое утверждение:

Теорема 15 *Существует такое число $\gamma < 8$, что при всех достаточно больших n для построенного однородного графа G_n (степени 8, с n^2 вершинами) выполнено $\lambda(G) \leqslant \gamma$.*

Таким образом, мы не указываем точное значение спектрального зазора и лишь утверждаем, что он (для всех значений n) от нуля некоторой константой. Из доказательства, которое мы приведём ниже, можно извлечь некоторое конкретное γ, однако мы не будем проделывать это вычисление и предоставим его читателю в качестве упражнения.

Доказательство: Чтобы оценить второе собственное число матрицы графа M, нужно оценить отношение Рэлея. Мы должны доказать, что

$$\min_{\mathbf{v}\perp(1,\ldots,1)} \frac{\mathbf{v}M\mathbf{v}^{\perp}}{||\mathbf{v}||^2} \leqslant \gamma$$

для некоторого $\gamma < 8$. Для нашего графа это утверждение можно переформулировать следующим образом. Мы должны доказать, что для всех отображений

$$f \ : \ \mathbb{Z}_n^2 \to \mathbb{R},$$

удовлетворяющих условию $\sum\limits_{\mathbf{v}\in\mathbb{Z}_n^2} f(\mathbf{v}) = 0$, выполнено неравенство

$$\sum_{\substack{(\mathbf{v},\mathbf{w}) \ : \ \mathbf{v} \text{ и } \mathbf{w} \\ \text{соединены ребром}}} f(\mathbf{v})f(\mathbf{w}) \ \leqslant \gamma \sum_{\mathbf{v}\in\mathbf{Z}_n^2} f^2(v)$$

(сумма в левой части берётся по всем *упорядоченным* парам вершин $\mathbf{v}$ и $\mathbf{w}$, соединённых в ребром). Поделим это неравенство пополам (перестанем считать каждое ребро дважды) и переформулируем интересующее нас неравенство в виде следующей леммы.

Лемма 8 *Если* $\sum\limits_{\mathbf{v}\in\mathbb{Z}_n^2} f(\mathbf{v}) = 0$, *то*

$$\sum_{\mathbf{v}\in\mathbf{Z}_n^2} f(\mathbf{v})[f(T_1\mathbf{v}) + f(T_1\mathbf{v}+\mathbf{e}_1) + f(T_2\mathbf{v}) + f(T_2\mathbf{v}+\mathbf{e}_2)] \leqslant \frac{\gamma}{2} \cdot \sum_{\mathbf{v}\in\mathbf{Z}_n^2} f^2(\mathbf{v}).$$

Доказательство леммы: Обозначим $\hat{f}(v_1, v_2)$ преобразование Фурье функции f. Условие $\sum f(\mathbf{v}) = 0$ означает, что $\hat{f}(0,0) = 0$. Неравенство, которое мы хотим доказать, после преобразование Фурье превращается в

$$\sum_{\mathbf{z}\in\mathbf{Z}_n^2} \overline{\hat{f}(\mathbf{z})} \cdot [\hat{f}(T_2^{-1}\mathbf{z})(1 + e^{-2\pi z_1 i/n}) + \hat{f}(T_1^{-1}\mathbf{z})(1 + e^{-2\pi z_2 i/n})] \leqslant \frac{\gamma}{2} \cdot \sum_{\mathbf{z}\in\mathbf{Z}_n^2} |\hat{f}(\mathbf{z})|^2.$$

Чтобы несколько упростить громоздки выкладки, обозначим $G(\mathbf{z}) := |\hat{f}(\mathbf{z})|$. напомним также, что

$$\left|1 + e^{-2\pi t i/n}\right| = 2\left|\cos\frac{\pi t}{n}\right|$$

(мы воспользуемся этим равенством для $t = z_1$ и $t = z_2$).

Ещё раз переформулируем утверждение леммы. Мы хотим показать, что для любой функции

$$G \ : \ \mathbb{Z}_n^2 \to \mathbb{R},$$

удовлетворяющей условию $G(0,0) = 0$, выполняется неравенство

$$2\sum_{\mathbf{z}\in\mathbf{Z}_n^2} G(\mathbf{z}) \cdot \left[G(T_2^{-1}\mathbf{z}) \cdot |\cos\frac{\pi z_1}{n}| + G(T_1^{-1}\mathbf{z}) \cdot |\cos\frac{\pi z_2}{n}|\right] \leqslant \frac{\gamma}{2} \cdot \sum_{\mathbf{z}\in\mathbf{Z}_n^2} G(\mathbf{z})^2. \tag{5.3}$$

Если заменить $|\cos\frac{\pi z_1}{n}|$ и $|\cos\frac{\pi z_2}{n}|$ на единицу, то неравенство (5.3) превратится в

$$\sum_{\mathbf{z}\in\mathbf{Z}_n^2} G(\mathbf{z}) \cdot \left[G(T_2^{-1}\mathbf{z}) + G(T_1^{-1}\mathbf{z})\right] \leqslant \frac{\gamma}{4} \cdot \sum_{\mathbf{z}\in\mathbf{Z}_n^2} G(\mathbf{z})^2.$$

Это неравенство очевидно верно при $\gamma = 8$ (неравенство Коши–Буняковского). Но нам нужно доказать (5.3) для некоторого $\gamma < 8$. Это значит, что замена модуля косинусов на единицу была слишком грубой ценкой. Нам нужно существенно использовать то, что значения косинусов в некоторых точках намного меньше единицы.

Идея доказательства основана на элементарном арифметическом неравенстве: для любых действительных чисел A, B, τ выполнено неравенство

$$2AB \leqslant \tau A^2 + \frac{1}{\tau}B^2. \tag{5.4}$$

Мы мы будем применять это неравенство для каждого произведения вида $G(\mathbf{z}) \cdot G(T_2^{-1}\mathbf{z})$ или $G(\mathbf{z}) \cdot G(T_1^{-1}\mathbf{z})$ из левой части (5.3). Полагая в (5.4) $A = G(\mathbf{v})$ и $B = G(\mathbf{w})$ мы будем получать

$$2G(\mathbf{v}) \cdot G(\mathbf{w}) \leqslant \tau(\mathbf{v}, \mathbf{w})G^2(\mathbf{v}) + \tau(\mathbf{w}, \mathbf{v})G^2(\mathbf{w}).$$

Главный трюк состоит в выборе подходящего τ. Мы будем выбирать разные τ для разных произведений вида $\mathbf{v}) \cdot G(\mathbf{w})$. Таким образом, τ можно

считать функцией от пары $(\mathbf{v}, \mathbf{w})$. Нам потребуется, чтобы данная функция $\tau : V^2 \to \mathbb{R}$ обладала свойством $\tau(\mathbf{v}, \mathbf{w}) = \tau(\mathbf{w}, \mathbf{v})^{-1}$. (Разумеется, на диагональных элементах такая функция обязана быть равной единице, $\tau(\mathbf{v}, \mathbf{v}) = 1$.) Отложим на некоторое время вопрос о выборе τ и покажем, как с её помощью можно упростить неравенство (5.3).

Многократно воспользовавшись (5.4), мы заменим левую часть (5.3) на сумму

$$\sum_{\mathbf{z}\in\mathbf{Z}_n^2} [\tau(\mathbf{z}, T_2^{-1}\mathbf{z})G^2(\mathbf{z}) + \tau(T_2^{-1}\mathbf{z}, \mathbf{z})G^2(T_2^{-1}\mathbf{z})] \cdot |\cos\frac{\pi z_1}{n}| + [\tau(\mathbf{z}, T_1^{-1}\mathbf{z})G^2(\mathbf{z}) + \tau(T_1^{-1}\mathbf{z}, \mathbf{z})G^2(T_1^{-1}\mathbf{z})] \cdot |\cos\frac{\pi z_2}{n}|,$$

которую можно переписать в виде

$$\sum_{\mathbf{z}\in\mathbf{Z}_n^2} G^2(\mathbf{z}) \left(|\cos\frac{\pi z_1}{n}| \cdot [\tau(\mathbf{z}, T_2\mathbf{z}) + \tau(\mathbf{z}, T_2^{-1}\mathbf{z})] + |\cos\frac{\pi z_2}{n}| \cdot [\tau(\mathbf{z}, T_1\mathbf{z}) + \tau(\mathbf{z}, T_1^{-1}\mathbf{z})] \right).$$

Теперь для доказательства леммы нам остаётся показать, что для некоторого $\gamma < 8$

$$\sum_{\mathbf{z}\in\mathbf{Z}_n^2} G^2(\mathbf{z}) \left(|\cos\frac{\pi z_1}{n}| \cdot [\tau(\mathbf{z}, T_2\mathbf{z}) + \tau(\mathbf{z}, T_2^{-1}\mathbf{z})] + |\cos\frac{\pi z_2}{n}| \cdot [\tau(\mathbf{z}, T_1\mathbf{z}) + \tau(\mathbf{z}, T_1^{-1}\mathbf{z})] \right) \leqslant \frac{\gamma}{2} \cdot \sum_{\mathbf{z}\in\mathbf{Z}_n^2} G^2(\mathbf{z}).$$

Мы хотим добиться того, чтобы для каждого $\mathbf{z}$ выражение

$$\left(|\cos\frac{\pi z_1}{n}| \cdot [\tau(\mathbf{z}, T_2\mathbf{z}) + \tau(\mathbf{z}, T_2^{-1}\mathbf{z})] + |\cos\frac{\pi z_2}{n}| \cdot [\tau(\mathbf{z}, T_1\mathbf{z}) + \tau(\mathbf{z}, T_1^{-1}\mathbf{z})] \right) \tag{5.5}$$

оказалось меньше 4 (точнее, меньше некоторой независящей от n константы $\frac{\gamma}{2}$, которая в свою очередь меньше 4).

В данном случае удобно представлять набор остатков по модулю n в виде

$$\mathbf{Z}_n = \{-n/2 + 1, \ldots, -1, 0, 1, \ldots, n/2\}$$

(для нечётных n нужно добавить округление для концов интервала). Таким образом, мы каждый из двух аргументов τ есть целочисленная точка в квадрате со стороной длины n и с центром в точке $(0, 0)$.

Для точек $\mathbf{z} = (z_1, z_2)$, достаточно далёких от начала координат, хотя бы одно из значений $|\cos\frac{\pi z_1}{n}|$, $|\cos\frac{\pi z_1}{n}|$ будет отделено от нуля, и беспокоиться не о чем (если хотя бы одна из компонет $\mathbf{v}$, $\mathbf{w}$ достаточно далека от точки $(0, 0)$, можно положить $\gamma(\mathbf{v}, \mathbf{w}) = 1$). Но для точек $\mathbf{z}$ в окрестности нуля значения обоих косинусов становятся близки к единице. Поэтому для того, чтобы сумма (5.5) была отделена от 4, нужно удачно подобрать функцию τ.

Мы переходим к ключевому моменту доказательства.

Определение А: Будем называть *близкой окрестностью нуля* множество всех таких $\mathbf{v} = (v_1, v_2)$, что

$$|v_1| + |v_2| < n/10$$

(граница в $1/10$ от n выбрана произвольно и не является ни в каком смысле оптимальной).

Определение Б: Будем говорить, что пара $\mathbf{w} = (w_1, w_2)$ *предшествует* паре $\mathbf{v} = (v_1, v_2)$ (обозначение $\mathbf{w} \prec \mathbf{v}$), если

$$\begin{cases} |w_1| & \leqslant & |v_1|, \\ |w_2| & \leqslant & |v_2| \end{cases}$$

и хотя бы одно их этих двух неравенств является строгим.

Теперь мы готовы определить функцию $\tau(\mathbf{v}, \mathbf{w})$:

- если $\mathbf{v}$ лежит в близкой окрестности нуля и $\mathbf{w} \prec \mathbf{v}$, то положим $\tau(\mathbf{v}, \mathbf{w}) = \tau_0$,
- если $\mathbf{w}$ лежит в близкой окрестности нуля и $\mathbf{v} \prec \mathbf{w}$, то положим $\tau(\mathbf{v}, \mathbf{w}) = 1/\tau_0$,
- во всех остальных случаях положим $\tau(\mathbf{v}, \mathbf{w}) = 1$.

Какими могут быть значения τ в сумме (5.5) в случае, когда (z_1, z_2) лежит в близкой окрестности нуля? Нетрудно проверить, что в этом случае из четырёх значений τ либо два равны τ_0, а два других равны $1/\tau_0$, либо три равны $1/\tau_0$, и только одно равно τ_0. Положив $\tau_0 = 5/4$, мы получаем в обоих случаях сумму меньше 4. Лемма доказана.

Глава 6

Эффективные конструкции двудольных экспандеров*

Напомним, что в Теореме 2 мы неконструктивно доказали существование двудольных экспандеров. При этом для любого $\varepsilon > 0$, любых целых N и $K \leqslant N$ мы получали экспандер с параметрами $(N, M, D, K, \varepsilon)$, где

$$D = \Theta(\log N) \text{ и } M = \Theta(DK) \tag{6.1}$$

(константы в $O(\cdot)$ зависят от ε).

Известные методы не позволяют строить экспандеры с такими параметрами эффективно. Однако существует несколько явных конструкций, которые позволяют за полиномиальное получить двудольные экспандеры с параметрами, довольно близкими к оценке (6.1). В этой главе мы подробно рассмотрим одну из таких конструкций и коротко упомянем ещё одну.

6.1 Конструкция на основе кода Варди–Парвареша

В этом параграфе мы опишем конструкцию двудольного экспандера из [19]. Данный метод для каждого $\alpha > 0$ позволяет получить алгоритм, который для любого $\varepsilon > 0$ и $\forall N, K \leqslant N$ за полиномиальное (по N) время находит некоторый двудольный экспандер с параметрами $(N, M, D, K, \varepsilon)$, где

$$D = O\left(((\log N)(\log K))^{1+\frac{1}{\alpha}}\right) \text{ и } M = D^2 \cdot K^{1+\alpha} \tag{6.2}$$

(здесь также константа в обозначении $O(\cdot)$ зависит от ε). Видно, что в (6.2) и степень графа, и число вершин в правой доли несколько избыточно по сравнению с (6.1). Но, варьируя параметр α, мы можем по своему желанию перераспределять эту избыточность между значениями D и M.

Приведём пример. Если мы хотим, чтобы свойство расширения выполнялось для множеств размера до $K = N^{0.1}$, то неконструктивное доказательство гарантирует существование экспандера с $D = O(\log N)$ и $M = O(N^{0.1} \log n)$, а предлагаемая эффективная конструкция позволяет получить экспандер с $D = O(\mathrm{poly}(\log N))$ и, скажем, $m = O(N^{0.10001}\,\mathrm{poly}(\log N))$. Подчеркнём ещё раз, что предлагаемая конструкция работает для любых (сколь угодно малых) $\varepsilon > 0$.

Конструкция экспандера, которую мы сейчас опишем, задаётся следующим набором параметров:

- конечное поле $\mathbb{F}_q$ (число $q = |\mathbb{F}_q|$ есть степень простого числа),
- натуральное число n, неприводимый многочлен $h(y)$ степени n над полем $\mathbb{F}_q$,
- наутральные числа m и t.

Ниже мы обсудим, как именно следует выбирать эти параметры, чтобы получить граф с нужными нам свойствами. Теперь перейдем к описанию конструкции — объясним, как будет устроен граф.

Левая доля графа: Будем отождествлять вершины левой доли графа с многочленами $p(x)$ степени не выше n над полем $\mathbb{F}_q$. (Не обязательно считать вершинами *все* такие многочлены — можно взять только некоторые из них.) Понятно, что число вершин в левой доле графа не превосходит q^n.

Правая доля графа: Будем отождествлять вершины правой доли графа с $\mathbb{F}_q^{m+1}$. Таким образом, число вершин в правой доле графа равно q^{m+1}.

Рёбра графа: Из каждой вершине левой доли графа будет выходить q рёбер; рёбра каждой вершины будет удобно индексировать элементами поля $\mathbb{F}_q$. Правый конец каждого такого ребра мы будем вычислять с помощью отображения

$$N \;:\; \mathbb{F}_q^n \times \mathbb{F}_q \to \mathbb{F}_q^{m+1},$$

т.е., $N(v, i)$ есть правый конец i-го ребра, выходящего из вершины v левой доли графа. Самая важная часть конструкции — это, разумеется, описание данной функции N. Напомним, что первый аргумент функции N мы понимаем как многочлен степени не выше n, а второй аргумент как элемент поля $\mathbb{F}_q$. Итак, если $p(x)$ многочлен, а y некоторый элемент поля, мы должны определить $N(p, y)$. Чтобы упростить запись, введём обозначения

$$p_j(x) := p^{t^j}(x) \mod h(x) \tag{6.3}$$

(возводим многочлен $p(x)$ в степень t^j и приводим по модулю $h(x)$). Теперь: используя введённые обозначения, мы готовы определить отображение N:

$$N(p, y) := [y, p(y), p_1(y), \ldots, p_{m-1}(y)]$$

(мы находим степени многочлена $p(y)$, приводя их по модулю h, а затем вычисляем все эти многочлены в точке y).

Замечание: Отображение N полезно представлять себе как *кодирование*. При этом коэффициенты многочлена $p(y)$ играют роль «сообщения», а набор значений $N(p, y)$ для всевозможных y играет роль «кодового слова». Данная конструкция возникла в работе Варди и Парвареша [18] как обобщение классического кода Рида–Соломона. В [19] было замечено, что взяв код Варди–Парвареша с необычными значениями параметров, мы получим экспандер. Код Варди–Парвареша обладает некоторыми замечательными свойствами — этот код допускает эффективное *декодирование списком* (см. [20]). По существу, именно это свойство декодирования списком и гарантирует, что построенный граф окажется хорошим экспандером. Но формально мы не будем опираться на какие-либо свойства кода Варди–Парвареша — мы дадим независимое доказательство нужных нам комбинаторных свойств отображения N.

Конструкция графа полностью описана. Остаётся доказать, что полученный граф обладает нужным нам свойством расширения. Сначала мы докажем техническое утверждение, а затем убедимся, что при правильном выборе параметров эта техническая оценка даёт нам требуемый экспандер.

Утверждение 8 *Если A некоторое множество вершин левой доли построенного графа, состоящее из не более, чем t^m вершин, то множество его соседей достаточно велико:*

$$|\Gamma(A)| \geqslant (q - (n-1)(t-1)m)|A|.$$

Доказательство: Обозначим $B := \Gamma(A)$ — множество вершин левой доли графа, являющихся соседями A, и $K = |A|$. Нам нужно доказать, что B состоит из не менее, чем из κK вершин, где

$$\kappa = q - (n-1)(t-1)m.$$

Предположим противное: пусть $|B| < \kappa K$

Напомним, что элементы множества B (как и все вершины правой доли графа) есть наборы из $(m+1)$ элементов поля $\mathbb{F}_q$. Мы подберём многочлен $Q(y, y_1, \ldots, y_m)$ который равен нулю на каждом наборе $(c_0, c_1, \ldots, c_m)$ из B. При этом многочлен мы будем искать в виде

$$Q(y, y_1, \ldots, y_m) = \sum_{j=0}^{K-1} \sum_{i=0}^{\kappa-1} c_{ij} y^i R_j(y_1, \ldots, y_m).$$

Здесь $R_j(y_1, \ldots, y_m)$ обозначает многочлен m переменных

$$R_j(y_1, \ldots, y_m) = y_1^{j_0} \cdot y_2^{j_1} \cdot \ldots \cdot y_m^{j_{m-1}},$$

где $j = j_0 + j_1 t + \ldots + j_{m-1} t^{m-1}$. Можно ли найти нетривиальный (не тождественно нулевой) многочлен указанного вида, обнуляющийся на каждом

элементе множества B? Каждое условие равенство многочлена нулю в одной точке множества B есть линейное уравнение для набора коэффициентов c_{ij}. Таких уравнений будет столько же, сколько элементов в B, т.е., по нашему предположению, меньше, чем κK. При этом число коэффициентов, задающих многочлен Q, равно в в точности κK. Таким образом, мы имеем систему линейных уравнений, в которой число уравнений меньше числа неизвестных. Такая система обязательно имеет ненулевые решения.

Среди всех возможных Q указанного вида, равных нулю во всех точках B, мы выберем многочлен с наименьшей степенью по переменной y (точнее, выберем один многочлен среди всех многочленов с наименьшей степенью по y). Перепишем выбранный нами $Q(y, y_1, \ldots, y_m)$ в виде

$$Q(y, y_1, \ldots, y_m) = \sum_{j=0}^{K-1} q_j(y) S_j(y_1, \ldots, y_m).$$

(Здесь q_j и S_j — некоторые многочлены одной и m переменных соответственно.) Заметим, что хотя бы один из многочленов q_j не делится на $h(y)$ (в противном случае можно было бы поделить Q на h и понизить степень многочлена по переменной y).

Теперь возьмём произвольную вершину из множества A; ей соответствует некоторый многочлен $p(y)$. Подставим в Q вместо $y_1, \ldots, y_m$ многочлены $p_j(y)$, определённые равенствами (6.3). Заметим, что полученный в результате многочлен одной переменной y

$$Q(y, p(y), p_1(y), \ldots, p_{m-1}(y))$$

равен нулю для любого $y \in \mathbb{F}_q$. При этом степень данного многочлена не превосходит $K - 1 + (n-1)(t-1)m$, что меньше q. Это значит, что данный многочлен тождественно равен нулю — все его коэффициенты равны нулю. Это, в свою очередь, означает, что многочлен

$$Q(y, p(y), p(y)^{\perp}, \ldots, p(y)^{t^{m-1}}) \tag{6.4}$$

делится на многочлен $h(y)$.

Теперь посмотрим на это утверждение как на уравнение в поле $\mathbb{F}_q/h(y)$ (в поле разложения многочлена $h(y)$, т.е., в поле размера q^n, элементами которых являются многочлены над $\mathbb{F}_q$, приведённые по модулю $h(y)$). Рассмотрим в этом поле многочлен

$$\begin{aligned} Q^*(z) &:= Q(y, z, z^{\perp}, \ldots, z^{t^m}) \mod h(y) = \\ &= \sum_{j=0}^{K-1} (p_j(y) \mod h) \cdot R_j(z, z^{\perp}, \ldots, z^{t^{m-1}}) = \\ &= \sum_{j=0}^{K-1} (p_j(y) \mod h)\, z^j \end{aligned}$$

(последнее равенство вытекает из определения многочлена R_j, в котором мы использовали разложение j по степеням t). Важно, что этот многочлен

не является тривиальным (не все его коэффициенты равны нулю в в поле $\mathbb{F}_q/h(y)$), и степень этого многочлена строго меньше K.

Наше замечание о том, что многочлен (6.4) делится на $h(y)$, можно переформулировать так: для каждого из многочленов $p(y)$, соответствующих вершинам графа из множества A, в поле $\mathbb{F}_q/h(y)$ выполнено равенство $Q^*(p) = 0$. Таким образом, Q^* имеет не меньше $|A|$ нулей. Но число нулей многочлена не может быть больше его степени. Получаем $|A| < K$, что противоречит выбору K. Теорема доказана.

Теперь, подбирая подходящие значения параметров, мы получим нужный нам экспандер.

Теорема 16 *Для любого $\alpha > 0$ и $\varepsilon > 0$ существует алгоритм, который для любого N и любого $K \leqslant N$ находит за полиномиальное по N время некоторый двудольный экспандер с параметрами $(N, M, D, K, \varepsilon)$, где*

$$D = O\left(((\log N)(\log K))^{1+\frac{1}{\alpha}}\right) \text{ и } M = O\left(D^2 \cdot K^{1+\alpha}\right).$$

Доказательство: Воспользуемся конструкцией графа из Утверждения 8 со следующими значениями параметров:

- $n = \log N$ (такой выбор гарантирует, что $q^n \geqslant N$)
- $t = \left\lceil \left(\frac{2n \log K}{\varepsilon}\right)^{1/\alpha} \right\rceil$
- в качестве q можно взять любую степень простого числа (например, степень двойки) из интервала $\frac{1}{2}t^{1+\alpha} < q \leqslant t^{1+\alpha}$
- $m = \left\lceil \frac{\log K}{\log t} \right\rceil$ (такой выбор m гарантирует, что $t^m \geqslant K$)

В результате мы получаем граф, в котором все вершины левой доли имеют степень q, и для любого множества вершин левой доли A, если $|A| \leqslant K$, то

$$|\Gamma(A)| \geqslant (q - (n-1)(t-1)m)|A|.$$

Для выбранных значений параметров получаем, что число соседей A не может быть меньше

$$q - (n-1)(t-1)m|A| > (1-\varepsilon)q.$$

Таким образом, мы построили экспандер с требуемым свойством расширения.

Описанная конструкция эффективна (матрицу графа можно выписать за время $\mathrm{poly}(N)$) поскольку все арифметические операции в поле и поиск неприводимого многочлена h можно произвести за полиномиальное по n время.

Упражнение 33 *(а) Проверьте, что для выбранных значений параметров выполнено нужное нам условие $q - (n-1)(t-1)m > (1-\varepsilon)q$.*

(б) Проверьте, что для выбранных значений параметров выполняются равенства $D = O\left(((\log N)(\log K))^{1+\frac{1}{\alpha}}\right)$ и $M = O\left(D^2 \cdot K^{1+\alpha}\right)$.

6.2 Конструкция с зигзаг-произведением

Параметры двудольного экспандера из конструкции их параграфа 6 далеки от оптимальных в случае, когда значения параметров N и K (размер левой доли графа и максимальный размер множества, для которого должно быть выполнено свойство расширения) близки (скажем, $N/K = const$). Но как раз для таких значений параметров хорошие оценки даёт конструкция из работы [24]. Эта конструкция использует зигзаг-произведение, перенесённое на (неоднородные) двудольные граф. Она позволяет для любых фиксированных $\varepsilon > 0$ и $t > 1$ и для всех натуральных n эффективно строить двудольный экспандер с параметрами

$$(N = 2^n, M = 2^{n-t}, D = 2^d, K = 2^k, \varepsilon).$$

где $d = O(\log t)$ и $k = n - t - d - O(1)$. Мы не приводим доказательства этого результата и отсылаем заинтересованного читателя к оригинальной статье [24].

Глава 7

Применение экспандеров: вероятностные алгоритмы

Мы в этой главе мы применим экспандеры для улучшения качества работы вероятностного алгоритма. Пусть имеется полиномиальный вероятностный алгоритм $\mathcal{A}$ с некоторой вероятностью ошибки $\delta < 1/2$. Будем предполагать, что данный алгоритм использует $k = poly(n)$ случайных битов на входах длины n. Далее мы опишем общий способ, позволяющий уменьшить вероятность ошибки такого алгоритма и при этом (а) не потерять полиномиальности времени работы, и (б) сравнительно «экономно» расходовать случайные биты.

Нивное решение состоит в том, чтобы параллельно запустить t копий имеющиего алгоритма на независимо выбранных значениях датчика случайных битов, а затем из полученных t результатов выбрать наиболее часто случающийся. У нового алгоритма вероятностью ошибки не будет превосходить c^t для некоторого $c < 1$. Таким образом, сделав число итераций t достаточно большим, можно сделать вероятность ошибки меньше любого наперёд заданного числа. Более того, положив $t = \Theta(n)$, можно даже сделать вероятность ошибки экспоненциально убывающей (с ростом длины входа n). При этом время работы алгоритма будет оставаться полиномиальным. Очевидным недостатком этого подхода является рост числа используемых случайных битов — их число умножается на t.

Мы покажем, что существует альтернатива простому повторению исходного алгоритма на независимых наборах случайных битов. Мы будем генерировать с помощью экспандеров «псевдослучайные» биты. Набор псевдослучайных битов можно будет получать из короткой «затравки» — небольшого набора настоящих случайных битов. При этом полученные псевдослучайные биты, как мы увидим, можно использовать для параллельного запуска многих копий вероятностного алгоритма, (почти) как если бы они были по-настоящему случайными и независимыми.

7.1 Уменьшение вероятности ошибки алгоритма без увеличения числа случайных битов

В этой главе мы рассмотрим самый простой способ получения «псевдослучайных» битов с помощью экспандера. Мы покажем, как уменьшить вероятности ошибки вероятностного алгоритма *без увеличения числа используемых случайных битов*. Мы ограничимся рассмотрением алгоритмов с односторонней ошибкой. Напомним стандартное определение класса задач, для которых существует полиномиальный вероятностный алгоритм с односторонней ошибкой.

Определение 7 *Язык L принадлежит сложностному классу* RP*, если существует полиномиальный алгоритм $\mathcal{A}$ такой что*

1. *если $x \in L$, то для случайно выбранного набора битов $r \in \{0,1\}^{\mathrm{poly}(n)}$ $\mathrm{Prob}_r \mathcal{A}(x,r) \geqslant 1/2$,*
2. *если $x \notin L$, то $\mathcal{A}(x,r) = 0$ для всех $r \in \{0,1\}^{\mathrm{poly}(n)}$.*

Покажем, что для любого $\delta > 0$ всякий полиномиальный вероятностный алгоритм $\mathcal{A}$ можно переделать в другой полиномиальный вероятностный алгоритм $\mathcal{A}'$ так, чтобы вероятность ошибки уменьшилась с $1/2$ до δ, а число используемых случайных битов при этом не изменится.

Пусть исходный алгоритм использует $k = k(n)$ случайных битов для вычислений на входах длины n. Зафиксируем однородный $(2^k, d, \varepsilon)$-экспандер G для некоторого $\varepsilon > 0$. Индекс (номер) каждой вершины в этом графе записывается последовательностью из k нулей и единиц. Таким образом, мы можем отождествить вершины G и наборы из k битов.

Новый алгоритм действует следующим образом: выбирается случайная вершина v графа (для этого требуется k случайных битов); затем исходный алгоритм $\mathcal{A}$ последовательно запускается на всех d наборах случайных битов, соответствующих соседям вершины v в графе. Если все полученные ответы равны 0, новый алгоритм также возвращает 0; если же получен хотя бы один положительный ответ, то алгоритм возвращает 1.

Покажем, что у нового алгоритма вероятность ошибки не превосходит $\frac{1}{2(1+\varepsilon)}$. Обозначим $B = B(x)$ множество всех *плохих* (для данного x) вершин графа — множество таких вершин w из правой доли графа, которые соответствуют неверному ответу старого алгоритма на входе x. Аналогично, обозначим $C = C(x)$ множество таких вершин v графа, которые для которых новый алгоритм даёт неверный ответ на входе x. Очевидно, C состоит из вершин, все соседи которых лежат в B.

Предположим, что C содержит не менее $\frac{1}{2(1+\varepsilon)}$ вершин. Произвольным образом выберем из множества C некоторое подмножество, состоящее *ровно* из $\frac{1}{2(1+\varepsilon)}$ вершин и назовём его C'. Из определения экспандера следует, что

$$|\Gamma(C)| > (1+\varepsilon)|C'| = n/2.$$

Это противоречит тому, что все соседи C' лежат в B.

Таким образом, мы построили алгоритм, в котором ошибка снизилась с $\frac{1}{2}$ до $\frac{1}{2(1+\varepsilon)}$. Покажем, как понизить вероятность ошибки ещё больше. Зададимся некоторым числом t и построим алгоритм, вероятность ошибки которого меньше $\frac{1}{2(1+\varepsilon)^t}$. В новом алгоритме мы выбираем в графе случайную вершину v (для по-прежнему этого требуется k случайных битов); затем запускаем исходный алгоритм $\mathcal{A}$ на всех наборах случайных битов, соответствующих вершинам w графа, в которые можно попасть из v за t шагов (таких вершин заведомо не более d^t). Если все полученные ответы равны 0, новый алгоритм также возвращает 0; в противном случае возвращается 1.

Оценим вероятность ошибки нового алгоритма. Снова обозначим $C = C(x)$ множество таких вершин v графа, которые для которых новый алгоритм даёт неверный ответ на входе x. Предположим, что C содержит не менее $\frac{1}{2(1+\varepsilon)^t}$ вершин. Выберем среди них подмножество, состоящее из ровно $\frac{1}{2(1+\varepsilon)^t}$ вершин и назовём его C'. Из определения экспандера следует, что

$$|\underbrace{\Gamma(\Gamma(\Gamma(\ldots\Gamma(}_{t\text{ итераций}}C')\ldots)))| > (1+\varepsilon)^t|C'| = n/2$$

Это противоречит тому, что все цепочки из t рёбер, начинающиеся вершиной из C', обязаны заканчиваться вершиной из B.

Выбирая параметр t достаточно большим, мы получим алгоритм с вероятностью ошибки менее $\frac{1}{2(1+\varepsilon)^t} < \delta$. При этом значение t зависит от желаемой вероятности ошибки δ, но не зависит от размера входа n.

Остаётся обсудить время работы построенного алгоритма. Мы используем старый алгоритм как «чёрный ящик» и вызываем его (на разных наборах случайных битов) d^t раз. Поскольку d и t – некоторые константы (не зависящие от входа алгоритма), и исходный алгоритм $\mathcal{A}$ работал за полиномиальное время, можно заключить, что все требуемые вызовы выполняются за полиномиальное время.

Однако кроме нескольких вызовов старого алгоритма нам требуется производить манипуляции с графом G — нужно уметь быстро находить всех соседей заданной вершины графа. Чтобы иметь возможность делать это за полиномиальное время, нам нужна *явная* конструкция экспандера. Более того, нам нужен экспандер *явный в сильном смысле*: размер графа экспоненциально растёт с увеличением k, и нам необходим алгоритм, который по заданному номеру вершины w за время $\mathrm{poly}(k)$ находит список номеров всех соседей w.

7.2 Блуждание на экспандере как генератор псевдослучайных битов: вероятностные алгоритмы с односторонней ошибкой.

Мы в этой главе мы покажем, как добиться экспоненциального уменьшения вероятности ошибки алгоритма, используя сравнительно небольшое число случайных битов. Здесь мы по-прежнему рассматриваем только алгоритмы с односторонней ошибкой. Мы предполагаем, что если алгоритм на входах длины n использует $k = \text{poly}(n)$ случайных битов. Мы считаем, что алгоритм выдает ответ 1, то он заведомо правильный, а выдаваемый ответ 0 может быть ошибочным. При этом для любого входа вероятность ошибки ограничена некоторым $\delta < 1$.

Мы хотим сделать вероятность ошибки экспоненциально убывающей функцией от n. Как и раньше, наивное решение состоит в том, чтобы параллельно запустить t копий исходного алгоритма на независимых значениях датчика случайных битов. (Считаем, что $t = \Theta(n)$.) Если в произведенных параллельных вычислениях хотя бы один из полученных результатов окажется раным 1, то в качестве ответа нужно выдать 1; если же все t результатов равны 0, то в качестве ответа нужно выдать 0. Нетрудно видеть, что вероятность ошибки после t-кратного повторения уменьшилась с δ до δ^t. Однако и число используемых случайных битов выросло в t раз. Далее мы покажем, как добиться экспоненциального уменьшения вероятности ошибки, используя значительно меньше случайных битов.

Построим спектральный $(2^k, d, \gamma)$-экспандер (здесь k — число случайных битов, которое требовалось исходному вероятностному алгоритму). Выберем случайно вершину графа x_0, а затем сделаем t шагов случайного блуждания по графу, $x_0 - x_1 - \ldots - x_t$. Затем запустим $t + 1$ копию старого алгоритма, используя индексы вершин $x_0, x_1, \ldots, x_t$ как наборы случайных битов. Как и прежде, если в произведенных параллельных вычислениях хотя бы один из полученных результатов окажется раным 1, то в качестве ответа нужно выдать 1; если же все t результатов равны 0, то в качестве ответа выдаём 0.

Если исходный алгоритм был полиномиальным, то и новый алгоритм будет работать за полиномиальное время. Разумеется, нужно, чтобы конструкция используемого $(2^k, d, \gamma)$-экспандера была явной в сильном смысле (по номеру вершины требуется эффективно находить номера её соседей).

Сколько же случайных битов использует новый алгоритм? Чтобы задать на графе путь длины t, нам нужно $k + O(t)$ случайных битов. Это много лучше, чем tn случайных битов, которые были нужны для наивного t-кратного повторения исходного алгоритма. При этом согласно Утверждению 5 вероятность ошибки нового алгоритма будет не больше $(\delta + \gamma - \delta\gamma)^t$. Если взять γ достаточно малым (нопомним, что мы умеем строить спектральные экспандеры для сколь угодно малого парамтера $\gamma > 0$), то вероятность ошибки будет меньше c^t для некоторого $c < 1$, т.е., вероятность ошибки экспоненциально убывает с ростом t.

7.3 Блуждание на экспандере как генератор псевдослучайных битов: вероятностные алгоритмы с двусторонней ошибкой.

Рассмотрим теперь более общий случай — будем считать, что исходный вероятностный алгоритм может выдавать как положительные, так и отрицательные ложные ответы. При этом для любого входа вероятность ошибки ограничена некоторым $\delta < 1/2$.

Как и в случае односторонней ошибки, мы можем последовательно запустить исходный алгоритм t раз на независимых значениях датчика случайных битов. Затем выберем из t полученных экземпляров ответа самый часто встречающийся. С ростом t вероятность ошибки в итоговом ответе будет экспоненциально убывать. Однако число используемых случайных битов также увеличивается в t раз.

Как и в предыдущей главе, мы можем добиться экспоненциального убывания вероятности ошибки, более экономно расходуя случайные биты. Снова рассмотрим спектральный $(2^k, d, \gamma)$-экспандер (где k — число случайных битов, которое требовалось исходному вероятностному алгоритму). Сделаем t шагов случайного блуждания по графу, $x_0 - x_1 - \ldots - x_t$. Затем запустим $t+1$ копию старого алгоритма, используя индексы вершин $x_0, x_1, \ldots, x_t$ как наборы случайных битов. Среди полученных ответов выбереть самый часто встречающийся и объявим его результатом работы нового алгоритма.

Как и раньше, случайное блуждание задается $k + O(t)$ случайными битами. А Утверждение 6 позволяет оценить вероятность ошибки нового алгоритма. Она не превосходит

$$\sum_{I \subset \{0,\ldots,t\},\ |I|>t/2} (\delta + \gamma - \delta\gamma)^{|I|-1} \leqslant 2^{t+1}(\delta + \gamma - \delta\gamma)^{(t-1)/2},$$

т.е., для достаточно малых γ вероятность ошибки будет экспоненциально убывать с ростом t.

7.4 Алгоритм проверки связности графа с использованием логарифмической памятью

Мы уже видели, что зигзаг-произведение позволяет «собирать» из маленьких экспандеров сколь угодно большие экспандеры с ограниченной степенью и достаточно малым вторым собственным числом. Теперь мы рассмотрим ещё одно замечательное одно применение этих операций. Мы докажем теорему о дерандомизации одного из самых знаменитых вероятностных алгоритмов — алгоритма проверки s-t-связности в неориентированном графе (задача UPATH) с логарифмической памятью.

Задача UPATH: Задан неориентированный граф $G = (V, E)$, в котором выделены две вершины $s, t \in V$. Требуется выяснить, есть ли в графе путь из

вершины s в вершину t.

Теорема 17 *Задача* UPATH *может быть решена вероятностным алгоритмом с логарифмической памятью.*

Вероятностный алгоритм для решения задачи UPATH устроен очень просто: нужно сделать $N = \mathrm{poly}(|V|)$ (выбор полинома мы уточним чуть позже) шагов случайного блуждания по графу, начав с вершины s. Если за N шагов нам удастся побывать в вершине t, мы точно знаем, что в графе есть путь из s в t. В противном случае мы полагаем, что такого пути нет.

В каждый момент работы алгоритма нам требуется помнить номер текущего шага блуждания (от 1 до N) и номер вершины, в которой мы в данным момент находимся. Для хранения этой информации достаточно памяти размера $O(\log |V|)$.

Ясно, что если пути из s в t нет, то алгоритм выдаст правильный ответ. Остаётся оценить вероятность другой ошибки: путь из s в t существует, но за N шагов блуждания мы его не обнаружим. Без ограничения общности можно считать, что граф регулярен и недвудолен (мы всегда можем добиться этого, добавив в граф некоторое количество петель). Далее покажем, что при случайном блуждании по связному однородному и недвудольному графу распределение вероятностей на вершинах быстро приближается к однородному. Ключевое свойство графа:

Лемма 9 *В d-регулярном однородном и недвудольном графе с n вершинами щель между первым и вторым по абсолютной величине собственными числами не может быть меньше $1/\mathrm{poly}(n)$, т.е.*

$$\lambda/d \geqslant 1 - \Theta(1/n^c)$$

для некоторой константы c (не зависящей ни от n, ни от d).

Доказательство леммы: Прежде всего, без ограничения общности можно считать, что граф связен (если это не так, мы перейдём рассмотрению одной связной компоненты).

Далее, если у графа есть отрицательные собственные числа, мы перейдём от исходного графа G к его квадрату G^2. При возведении в квадрат все собственные числа также возведутся в квадрат и станут положительными (щель между перым и вторым по модулю собственным числом также измениться в полином раз — умножится на $O(d)$). Важно отметить, что поскольку исходный граф G не был двудольным, в его квадрате максимальное собственное число имеет кратность 1 (связный недвудольный граф при возведении в квадрат остаётся связным).

Таким образом, остаётся доказать лемму для связного графа, у которого все собственные значения положительны. Обозначим $\mathbf{f} = (f_1, \ldots, f_n)$ собственный вектор, соответствующий второму собственному числу G^2 (он ортогонален первому собственному вектору $(1, 1, \ldots, 1)$, т.е., $\sum f_i = 0$).

Всегда можно считать, что норма $\mathbf{f}$ равна единице. Тогда найдётся координата i такая, что $|f_i| \geqslant 1/\sqrt{n}$. Предположим для определённости, что f_i положительно. Поскольку сумма всех координат $\bar{f}$ равна нулю, то найдётся и координата j, для которой $f_j \leqslant 0$.

Рассмотрим в графе кратчайший путь из i-ой вершины в j-ую:

$$f_i - \ldots - f_j.$$

В этом пути найдётся хотя бы одно ребро $f_l - f_m$, для которого

$$|f_l - f_m| \geqslant |f_i - f_j|/n \geqslant \frac{1}{n\sqrt{n}}$$

Итак, мы нашли в графе такую пару вершин, соединённых ребром, что разница $|f_l - f_m|$ не меньше $1/n^{1.5}$.

Теперь вычислим лапласиан графа: просуммируем $(f_l - f_m)^2$ по всем рёбрам (l, m) графа (см. параграф 3.3). При этом каждое ребро мы считаем по одному разу:

$$\sum_{\{l,m\}\in E} (f_l - f_m)^2 = \sum_{\{l,m\}\in E} (f_l^2 + f_m^2 - 2f_l f_m) = d\sum_{s=1}^{n} f_s^2 - \mathbf{f}M\mathbf{f}^{\perp} = 2d^2 - 2\lambda$$

(здесь M, как обычно, обозначает матрицу графа, d — его степень). Напомним, что данное равенство верно независимо от того, если в графе петли. Данная сумма снизу ограничена $(f_s - f_l)^2 \geqslant \frac{1}{n^3}$. Следовательно, разность $d - \lambda$ ограничена снизу $\Theta(1/n^3)$. Лемма доказана.

С помощью этой леммы мы докажем корректность работы нашего алгоритма. Обозначим $\bar{p}(i)$ распределение вероятностей на вершинах после i шагов случайного блуждания по графу (распределение $\bar{p}(0)$ сосредоточено в единственной вершине s). Пусть обозначим равномерное распределение $\bar{u} = (\frac{1}{n}, \ldots \frac{1}{n})$ на вершинах компоненты связности s, и разложим $\bar{p}(i)$ в сумму $\bar{u}$ и некоторого вектора из его ортогонального дополнения:

$$\bar{p}(i) = \bar{u} + \bar{q}(i),$$

где сумма координат вектора $\bar{q}(i)$ равна нулю. Если M — нормализованная матрица графа, то $\bar{q}(i+1) = M\bar{q}(i)$. На подпространстве векторов с нулевой суммой норма линейного оператора M равна (нормализованному) второму собственному числу графа; по лемме это число не может быть больше $1 - \Theta(1/n^c)$, где n есть число вершин в компоненте связности вершины t. Следовательно, на каждом шаге норма $\bar{q}(i)$ уменьшается по крайней мере в $(1 - \Theta(1/n^c))$ раз, и через $\mathrm{poly}(n)$ шагов распределение $\bar{p}(i)$ станет очень близко к равномерному (на компоненте связности графа). Таким образом, если s и t принадлежат одной компоненте связности, то вероятность попасть через $\mathrm{poly}(n)$ шагов в вершину t будет близка к $1/n$. Если же увеличить число шагов ещё в полином раз, то вероятность хотя бы раз побывать в t станет близка к единице. Теорема доказана.

Далее мы покажем, как дерандомизовать алгоритм случайного блуждание на графе без значительного увеличения используемой памяти. Более формально, мы докажем следующую теорему.

Теорема 18 *Задача* UPATH *может быть решена* детерминированным *алгоритмом с логарифмической памятью.*

Прежде чем доказывать теорему, заметим, что мы уже умеем решать на логарифмической памяти задачу UPATH для $(n, d, 0.99)$-экспандеров. В самом деле, мы знаем, что диаметр такого экспандера равен $O(\log n)$. Мы можем перебрать все пути длины $C \log n$ с началом в вершине s и проверить, ведёт ли хотя бы один из них в t; такая проверка очевидно требует лишь логарифмической памяти (и полиномиального времени).

Чтобы решить задачу для произвольного графа G, мы превратим его в экспандер с помощью и зигзаг-произведения

Доказательство теоремы: Мы предполагаем, что нам задан (в виде оракула) неориентированный граф G с n вершинами, без петель и параллельных рёбер. Далее мы построим на основе G несколько «воображаемых» графов; мы сможем моделировать блуждание по каждому из этих воображаемых графов с помощью исходного оракула и дополнительной памяти размера $O(\log n)$.

Воображаемый граф G': заменим в исходном графе каждую вершину v_i степени $d_i > 3$ на цикл длины d_i; рёбра, входившие ранее в данную вершину мы по одному присоединим к вершинам этого цикла. Таким образом, в графе G' степень каждой вершины не превосходит 3. Обозначим через n' число вершин в G' (это число не превосходит $\mathrm{poly}(n)$).

Воображаемый граф G'': Добавим к каждой из вершин G_1 нужное число петель так, чтобы получился D-регулярный граф для некоторого $D = d^4$; целое число d мы выберем так, чтобы существовал спектральный экспандер H с параметрами $(D = d^4, d, < 0.01)$.

Воображаемые графы G_i : $G_0 = G''$; каждый следующий граф G_{i+1} определяется рекурсивно:

$$G_{i+1} = (G_i ⓩ H)^2.$$

При этом каждый граф G_i будет экспандером с параметрами

$$(n' \cdot D^i, d^4, < 1 - \varepsilon_i)$$

для некоторого ε_i.

Нас интересует значение параметров ε_i (насколько хорошими экспандерами будет построенные графы). Оказывается, что на каждом шаге значение ε будет увеличиваться почти вдвое. В самом деле, произведение $G_i ⓩ H$ будет спектральным экспандером с параметрами $(nD, d^2, 1-(0.99)^2\varepsilon_i)$ (свойство зигзаг-произведения, Теорема 10). Затем мы берём вторую степень этого графа, и все собственные числа возводятся в квадрат. Для малых x имеем $(1-x)^2 \approx 1 - 2x$; таком образом, если ε_i достаточно мало, то

$$\varepsilon_{i+1} \approx 2 \cdot 0.99^2 \varepsilon_i \approx 2\varepsilon_i.$$

Применяя Лемму 9 к графу G_0, заключаем, что $\varepsilon_0 \geqslant \Omega(1/(n')^c)$. Далее для каждого следующего G_i значение ε_i становится почти в два раза больше. Следовательно, для $k = O(\log n)$ граф G_i оказывается экспандером, у которого нормализованное второе собственное число по крайней мере не превосходит 0.99.

Вершины G_i получаются как тензорное произведение вершины графа G'' и i копий вершин графа H. Зигзаг-произведение устроено так, что вопрос о существовании пути из s в t в исходном графе G эквивалентен вопросу о существовании пути в G_i из вершин, у которых первая тензорная компонента равна s, в вершины, у которых первая тензорная компонента рана t. Поскольку для $k = \Theta(\log n)$ у графа G_k нормализованное второе собственное число не превосходит 0.99, мы можем проверить данное свойство, перебрав все пути логарифмической длины.

Остаётся заметить, что моделирование блуждания по графу G_k моделируется на логарифмической памяти. В самом деле, для хранения номера вершины G_k нам нужно хранить набор из $(k+1)$ компонент; самая первое содержит некоторый номер вершины $G_0 = G''$, а каждая следующая — номер одной из вершин H. Ребро в каждом графе G_{i+1} есть путь длины 2 в графе $(G_i ⓩ H)$. Остаётся понять, как организовать рекурсивный вызов для моделирования одного шага по ребру $(G_i ⓩ H)$. Мы предлагаем читателю убедиться, что организация данной рекурсивной процедуры требует лишь $O(1)$ ячеек памяти на каждую компоненту $i = 1, \ldots, k = \Theta(\log n)$. Таким образом, моделирование одного шага блуждания на «воображаемом» графе G_k требует памяти $O(\log n)$.

Упражнение 34 *Опишите подробнее рекурсивный алгоритм, моделирующий один шаг блуждания на графе G_k с использованием памяти $O(\log n)$.*

Глава 8

Применение экспандеров: коды на графах

Напомним, что (двоичным) *кодом* называется набор слов $\mathcal{C} \subset \{0,1\}^n$. Элементы $\mathcal{C}$ называют *кодовыми словами*, число n называют *длиной кодового слова*, а $M = |\mathcal{C}|$ — *объёмом кода*. С помощью кода объема S можно пересылать наборы их $s = \lfloor \log S \rfloor$ битов (сопоставив разным наборам из m битов разные кодовые слова). *Минимальным расстоянием кода* называют наименьшее хэмминговское расстояние между парой кодовых слов. Если кодовое расстояние равно r, то говорят, что данный код позволяет исправлять $\left\lfloor \frac{r-1}{2} \right\rfloor$ ошибок. (Если в кодовом слове инвертировать до $\left\lfloor \frac{r-1}{2} \right\rfloor$ битов, мы можем однозначно восстановить исходное слово). Таким образом, если мы хотим исправлять ошибки в доле δ от всех битов кодового слова, то необходимо, чтобы минимальное расстояние кода удовлетворяло неравенству $r \geqslant 2\delta n + 1$.

Таким образом, всякий код характеризуется параметрами n (длина кода), m (чило передаваемых битов) и r (минимальное расстояние между кодовыми словами). Отношение s/n называют *скоростью кода* (это отношение является свого рода коэффициентом полезного действия кода — оно показывает, сколько «полезных» битов удаётся переслать в расчёте на один бит кодового слова). Задача теории кодирования состоит в поиске кодов с оптимальным соотношением параметров: при фиксированных n и s максимизировать r или при фиксированных n и r максимизировать s. Отдельная (и с практической точки зрения очень важная) задача — построение кодов с эффективными алгоритмами декодирования. Такой алгоритм должен восстанавливть кодовое слово, если в нём искажено не слишком большое число битов.

Важным классом кодов с разными замечательными свойствами являются *линейные* коды — такие коды, для которых множество кодовых слов $\mathcal{C} \subset \{0,1\}^n$ оказывается линейным подпространством в $\mathbb{F}_2^n$. Для линейных кодов минимальное расстояние между кодовыми словами равняется мини-

мальному возможному числу единиц в ненулевом кодовом слове.

Подробнее об основных понятиях теории кодирования можно прочитать в кратких курсах лекций [31], [20] либо (более подробно) в классической монографии [32].

В этой главе мы покажем, как с помощью экспандеров строить линейные коды. Эти коды будут обладающие достаточно хорошим соотношением параметров; кроме того, мы предъявим для этих кодов быстрые алгоритмы декодирования.

8.1 Коды на двудольном экспандере

В этой главе мы опишем простейшую конструкцию линейного кода на экспандере. Напомним, что линейный код с длиной кодового слова n задаётся его проверочной матрицей H (слово $x \in \{0,1\}^n$ является кодовым словом, если и только если $Hx^{\perp} = 0$). Другими словами, чтобы описать линейный код, мы должны задать систему линейных уравнений для переменных $x_1, \ldots, x_n$ над полем из двух элементов; решения этой системы и будут кодовыми словами.

Зафиксируем некоторый двудольный $(n, m, k, d, \varepsilon)$-экспандер. Сопоставим переменные $x_1, \ldots, x_n$ вершинам в левой доле графа. Вершинам из правой доли будут соответствовать уравнения. А именно, каждой вершине v из правой доли G мы сопоставляем уравнение

$$x_{i_1} + \ldots + x_{i_s} = 0 \ (mod\ 2),$$

где $x_{i_1}, \ldots, x_{i_s}$ – это список вершин, соединённых рёбрами с v.

При этом нам будет нужно, чтобы для параметров $(n, m, k, d, \varepsilon)$-экспандера выполнялись следующие соотношения:

$m < cn$ для некоторого $c < 1$,
$k > 2\delta n$, где δ есть доля исправляемых ошибок,
$\varepsilon < 1/2$.

Число уравнений равно m, так что размерность пространства решений не меньше $n - m$. Это значит, что в нашем коде будет не менее $2^{n-m} = 2^{\Omega(n)}$ кодовых слов.

Остаётся доказать, что данный код действительно исправляет δn ошибок. Для этого нужно проверить, что расстояние между любыми кодовыми словами больше $k = 2\delta n$. Для линейного кода нужное нам условие эквивалентно тому, что в каждом ненулевом кодовом слове должно быть более k единиц.

Предположим противное: пусть существует некоторое ненулевое кодовое слово $x_1 \ldots x_n$ (решение системы линейных уравнений, соответствующих графу G), в котором менее k единиц. Обозначим A множество вершин из левой доли графа, соответствующих единицам в данной последовательности битов. Поскольку $|A| < k$, можно применить определение экспандера:

число соседей A достаточно велико,

$$|\Gamma(A)| > (1-\varepsilon)d|A|.$$

Из A выходит ровно $d|A|$ рёбер. Оценим среднее (по всем вершинам $v \in \Gamma(A)$) число ребер, которое приходит из A в v. Это число не превосходит

$$\frac{d|A|}{(1-\varepsilon)d|S|} = 1/(1-\varepsilon) < 2$$

Итак, среднее число ребер, соединяющих вершину из $\Gamma(A)$ с множеством A, больше нуля и меньше двух. Это значит, что хотя бы у одной вершины v из правой доли есть *ровно один* сосед из A. Но в таком случае уравнение, соответствующее v, не выполняется на наборе $x_1 \dots x_n$. Значит, набор битов с менее чем k единицами не может быть кодовым словом.

Для построения экспандерных кодов нужны экспандеры *явные в слабом смысле*: нам нужно научиться строить неоднородные экспандеры с n вершинами в левой доле (и подходящими значениями других параметров) за время, полиномиально зависящее от n.

8.2 Экспандерные коды: параллельный алгоритм декодирования

Пусть $G = (L, R, E)$ — двудольный экспандер с параметрами $(n, m, d, k, \varepsilon)$. Построим на нём линейный код, как было описано в параграфе 8.1 — каждой вершине в левой доле графа сопоставляется бит кодового слова, а каждой вершине правой доли графа сопоставляется контрольная сумма; набор битов считается кодовым словом, если все его контрольный суммы равны нулю.

В параграфе 8.1 мы показали, что в таком коде расстояние между кодовыми словами не меньше k. Это значит, что если в кодовом слове искажены (инвертированы) менее $k/2$ битов, то мы можем исправить внесённые ошибки и восстановить исходное кодовое слово. Однако наивный алгоритм исправления ошибок (перебор всех возможных способов инвертировать $< k/2$ битов в слове) требует огромного перебора. Главное достоинство экспандерных кодов — это существование быстрых алгоритмы декодирования.

Далее мы рассмотрим несколько таких алгоритмов. Нлавное достоинство этих алгоритмов в том, что они работают очень быстро. Мы начнём с самого простого паралелльного алгоритма декодирования.

Однофазный параллельный алгоритм декодирования экспандерного кода.

Вход алгоритма: набор битов $x_1, \dots, x_n$, приписанных вершинам левой доли графа.

1. Для каждой вершины $w \in R$ вычислить соответствующую контрольную сумму
$$c_w := \bigoplus_{v \in L \ : \ (v,w) \in E} x_v$$

2. Если все контрольные суммы равны 0, закончить работу, выдав текущий набор битов $x_1, \ldots, x_n$.

3. Для каждой вершины $v \in L$ инвертировать бит x_v, если более половины контрольных сумм, включающих x_v не равны 0, т.е.
$$\text{число вершин } w \in R \text{ таких, что } (v,w) \in E \text{ и } c_w = 1, \text{ больше } d/2$$

4. Вернуться к пункту 1.

Замечание: Вычисления в пунктах 1 и 3 данного алгоритма можно выполнять параллельно для всех вершин графа.

Теорема 19 *Если $\varepsilon < 1/8$ и исходный набор битов $\bar{x} = x_1, \ldots, x_n$ отличается от некоторого кодового слова $\bar{x}' = x'_1, \ldots, x'_n$ в не более, чем $k/2$ позициях, то через $O(\log n)$ итераций описанный алгоритм остановится и выдаст в качестве результата кодовое слово $\bar{x}'$.*

Доказательство теоремы будет использовать следующее определение.

Определение 8 *Пусть $G = (L, R, E)$ — двудольный граф, и $A \subset L$ некоторое множество вершин левой доли этого графа. Будем называть вершина правой доли графа $w \in R$* уединённым соседом *множества A, если существует ровно одна вершина $v \in A$, соединённая ребром с v. Других соседи A будем называть* неуединёнными.

Прежде всего докажем несложную комбинаторную лемму:

Лемма 10 (об уединённых соседях) *Пусть граф $G = (L, R, E)$ является двудольным экспандером с параметрами $(n, m, d, k, \varepsilon)$ (без кратных рёбер), и $A \subset L$ — некоторое множество вершин левой доли графа, $|A| \leqslant k$. Тогда число уединённых соседей A (в правой доле графа R) не меньше $(1 - 2\varepsilon)d|A|$.*

Доказательство леммы: Обозначим U множество всех уединённых соседей A. Из A выходит $d|A|$ рёбер. При этом $|U|$ из них приходят в вершины, являющиеся уединёнными соседями A в правой доле (по одному ребру в каждого уединённого соседа). А все остальные рёбра приходят в *неуединённых* соседей (в каждого неуединённого соседа A приходит не меньше двух рёбер). Таким образом, мы получаем
$$|\Gamma(A)| \leqslant |U| + \frac{d|A| - |U|}{2} = \frac{1}{2}d|A| + \frac{1}{2}|U|.$$

С другой стороны, по определению экспандера мы имеем

$$|\Gamma(A)| > (1-\varepsilon)d|A|.$$

Следовательно,

$$(1-\varepsilon)d|A| < \frac{1}{2}d|A| + \frac{1}{2}|U|,$$

и $|U| > (1-2\varepsilon)d|A|$. Лемма доказана.

Доказательство Теоремы 19: Пусть $\bar{x} = (x_1, \ldots, x_n)$ — текущий набор битов, приписанных вершинам левой части графа. Мы предполагаем, что $\bar{x}$ не более, чем в $k/2$ позициях отличается от некоторого кодового слова $\bar{y} = (y_1, \ldots, y_n)$. Покажем, что после очередной итерации алгоритма расстояние между новым набором битов $\bar{x}' = (x'_1, \ldots, x'_n)$ и кодовым словом $\bar{y}$ сократится не менее, чем в c раз для некоторой константы $c > 1$. (Из этого свойства алгоритма немедленно следует, что через $O(\log n)$ итерации расстояние станет равно нулю, т.е., текущий набор битов превратиться в нужное нам кодовое слово $\bar{y}$.)

Обозначим $A \subset \{1, \ldots, n\}$ множество позиций, в которых текущее $\bar{x} = (x_1, \ldots, x_n)$ отличается от кодового слова $\bar{y} = (y_1, \ldots, y_n)$. После окончания одной итерации алгоритма $\bar{x} = (x_1, \ldots, x_n)$ превратится в некоторый набор битов $\bar{x}' = (x'_1, \ldots, x'_n)$; мы обозначим $A' \subset \{1, \ldots, n\}$ множество позиций, в которых новый набор битов будет отличаться от $\bar{y}$.

Изучим соотношение между A и A' более подробно. Для этого разделим A' на две части: положим

$$A' = B \cup C,$$

где $B \subset A$ и $C \subset \{1, \ldots, n\} \backslash A$. Другими словами, B состоит из позиций, в которых сохранятся исходные «неправильные» биты, а C состоит из позиций, в которых сначала стояли «правильные» биты (такие же, как в $\bar{y}$), но после применения одного шага алгоритма декодирования эти биты стали «неправильными» (отличающимися от битов в $\bar{y}$).

Оценим отдельно размеры B и C. Прежде всего отметим, что по Лемме 10 число уединённых соседей A не меньше $(1-2\varepsilon)d|A|$. Во всех этих соседях контрольные суммы заведомо равны 1. (Если вершина $w \in R$ не соединена ребром ни с одной вершиной из A, то её контрольная сумма равна 0, как у контрольной суммы кодового слова $\bar{y}$. Если же вершина w является неуединённым соседом A, то мы не можем точно сказать, будет ли её контрольная сумма равна 0 или 1 — это зависит от чётности числа вершин в A, соединённых ребром с w). Следовательно, не более $2\varepsilon d|A|$ рёбер ведут из A в некоторого неуединённого соседа.

Это наблюдение позволяет нам оценить размер B. В самом деле, «неправильный» (принадлежащий A) бит x_i остаётся не инвертированным (т.е., $i \in B$), только если хотя бы половина (хотя бы $d/2$ из d) его контрольных сумм равна нулю. А это значит, что хотя бы половина соседей данной вершины являются неуединёнными соседями A. Таким образом,

$$|B| \leqslant \frac{2\varepsilon d|A|}{d/2} = 4\varepsilon|A|.$$

Теперь оценим размер множества вершин C. Для этого нам потребуется рассмотреть множество соседей объединения $A \cup C$. Во-первых, среди соседей этого объединения встречаются все соседи A (коих не больше $d|A|$). Во-вторых, среди этих соседей встречаются также вершины $w \in R$, соединнные ребром с C, но не соединённые с A. Но вершин второго типа не может быть очень много. В самом деле, у каждой вершины C более половины соседей имеют единичную контрольную сумму; такие вершины обязаны быть соседями A (точнее, соседями *нечётного* числа вершин из A). Таким образом, каждая вершина из C может иметь не больше $d/2$ соседей, не покрытых множеством $\Gamma(A)$. Следовательно,

$$|\Gamma(A \cup C)| \leqslant d|A| + \frac{1}{2}d|C|$$

Теперь предположим, что объединение $A \cup C$ не очень велико (содержит не более k вершин). Тогда к $A \cup C$ можно применить свойство расширения экспандера. Получаем

$$d(1-\varepsilon)(|A| + |C|) < |\Gamma(A \cup C)| \leqslant d|A| + \frac{1}{2}dC,$$

откуда вытекает

$$|C| \leqslant \frac{\varepsilon}{\frac{1}{2} - \varepsilon}|A|.$$

Что делать, если в $A \cup C$ содержится больше k вершин? Просто выбросим из C «лишние» вершины — обозначим C' произвольное подмножество C, состоящее из ровно $k - |A|$ вершин. Затем применим к $A \cup C'$ приведенное выше рассуждение и получим

$$|C'| < \frac{\varepsilon}{\frac{1}{2} - \varepsilon}|A|.$$

Но тогда

$$|A \cup C'| < (1 + \frac{\varepsilon}{\frac{1}{2} - \varepsilon})|A| < \frac{|A|}{1 - 2\varepsilon} < k,$$

что противоречит выбору C'.

Теперь мы можем объединить полученные оценки для B и C. При $\varepsilon < 1/8$ получаем

$$|A'| = |B \cup C| < 4\varepsilon|A| + \frac{\varepsilon}{\frac{1}{2} - \varepsilon}|A| < \frac{6\varepsilon}{1 - 2\varepsilon}|A| < |A|.$$

Это значит, что число «неправильных» битов среди x_i на каждой итерации алгоритма уменьшается в константу раз. Понятно, что через $O(\log n)$ шагов ни одной ошибки не останется. Теорема доказана.

Замечание: Каждая итерация описанного алгоритма состоит из $O(n)$ операций. Таким образом, если использовать этот алгоритм без распараллеливания, его выполнение потребует времени $O(n \log n)$. В следующем параграфе мы опишем модификацию данного алгоритма, работающую за время $O(n)$.

Упражнение 35 *Докажите, что для экспандерного кода, построенного на экспандере с параметрами* $(n, m, d, k, <1/8)$*, кодовое расстояние не меньше* $\frac{3}{2}k$ *(что лучше оценки* k*, которую мы доказали в параграфе 8.1).*

8.3 Экспандерные коды: последовательный алгоритм декодирования

В этой главе мы опишим последовательный алгоритм декодирования экспандерного кода из параграфе 8.1, работающий за линейное время.

Однофазный параллельный алгоритм декодирования экспандерного кода.

Вход алгоритма: набор битов $\bar{x} = (x_1, \ldots, x_n)$, приписанных вершинам левой доли графа.

1. Для каждой вершины $w \in R$ вычислить соответствующую контрольную сумму
2. Для каждой вершины $v \in L$ вычислить число s_v - количество соответствующих ей контрольных сумм, равных 1.
3. Пока не все контрольные суммы равны 0, повторять следующую процедуру
   - 3.1. выбрать вершину $u \in L$, для которой более половины контрольных сумм равны 1;
   - 3.2. инвертировать бит x_u;
   - 3.3. скорректировать значение контрольных сумм для всех вершин $w \in R$, связанных с u;
   - 3.4. скорректировать число s_v, для затронутых вершин из левой доли графа.
4. Выдать в качестве результата набор текущих значений x_i.

Упражнение 36 *Покажите, что каждую итерацию шага 3 описанного алгоритма можно выполнять за время* $O(1)$*. (Указание: Нужно организовать хранение самого графа и чисел* s_v *таким образом, чтобы на каждом шаге было просто обновлять значения* s_v*.)*

Теорема 20 *Если* $\varepsilon < 1/4$ *и исходный набор битов* $\bar{x} = x_1, \ldots, x_n$ *отличается от некоторого кодового слова* $\bar{x}' = x'_1, \ldots, x'_n$ *в не более, чем* $k/2$ *позициях, то через* $O(n)$ *итераций описанный алгоритм остановится и выдаст в качестве результата кодовое слово* $\bar{x}'$*.*

Следствие 5 *Если* $\varepsilon < 1/4$ *и исходный набор битов* $\bar{x} = x_1, \ldots, x_n$ *отличается от некоторого кодового слова* $\bar{x}' = x'_1, \ldots, x'_n$ *в не более, чем* $k/2$ *позициях, то описанный алгоритм исправляет ошибки в* $\bar{x}$ *за время* $O(n)$*.*

Доказательство теоремы 20: Заметим, что на каждой итерации шага 3 описанного алгоритма число ненулевых контрольных сумм убывает. Следовательно, алгоритм остановится не более, чем через m шагов (напомним, что $m < n$). Чтобы доказать корректность алгоритма, мы должны установить следующий факт:

Пока не все контрольные суммы равны 0 *найдётся хотя бы одна вершина в левой доле графа, для которой более половины контрольных сумм раны* 1 (*таким образом, шаг 3.1 всего удается выполнить*).

Докажем это свойство в предположении, что число искаженных битов x_i (хэмминговское расстояние между текущим набором $\bar{x}$ и кодовым словом $\bar{y}$) не превосходит k. Обозначим через A_i множество таких вершин левой доли графа, которым на i-ой итерации шага 3 нашего алгоритма приписаны «неправильные» значения битов (т.е., $x_i \neq y_i$). Из Леммы 10 следует, что у A_i найдется не меньше

$$(1 - 2\varepsilon)d|A_i| > d|A_i|/2$$

уединённых соседей. Это значит, что *в среднем* у вершины из A_i имеется более $d/2$ уединённых соседей. Поскольку для каждого уединённого соседа A_i соответствующая контрольная сумма равна 1, мы заключаем, что пока A_i непусто, найдётся хотя бы одна вершина $v \in A_i$, для которой более половины контрольных сумм ненулевые.

Закончено ли доказательство теоремы? Не совсем. Наше рассуждение опиралось на предположение, что множество вершин A_i с «неправильными» битами не превосходит k. В начале работы алгоритма это так (более того, по условию теоремы для исходного множества ошибок выполнено $|A_0| < k/2$). Однако на некоторых итерациях алгоритма размер текущего множества A_i вполне может возрастать. Не перевалит ли он через k? К счастью, это не возможно. Мы знаем, что на каждой итерации алгоритма уменьшается другой важный параметр — число ненулевых контрольных сумм. При этом в начале работы алгоритма количество ненулевых контрольных сумм заведомо не превосходит $d|A_0|$ (все соседи исходного множества «неправильных» битов). Далее, на каждом шаге число единичных контрольных сумм остаётся не меньше, чем число уединённых соседей, т.е., не меньше

$$(1 - 2\varepsilon)d|A_i| \geqslant \frac{1}{2}d|A_i|$$

(мы снова используем Лемму 10). Следовательно,

$$\frac{1}{2}d|A_i| \leqslant [\text{число ненулевых контрольных сумм на } i\text{-ом шаге}] \leqslant d|A_0|.$$

Это значит, что размер A_i может вырасти, но не более, чем вдвое по сравнению с первоначальным, и число «неправильных» битов никогда не превосходит границы $2|A_0| = k$. Теперь теорема полностью доказана.

8.4 Экспандерные коды: двухфазное декодирование*

Напомним, что в главе 8.1 мы установили, что код исправляющий ошибки можно построить на $(n, m, d, k, \varepsilon)$-экспандерах в предположении, что $\varepsilon < 1/2$. В тоже время, алгоритмы декодирования из главы 8.2 и главы 8.3 применимы только для экспандеров с более сильным свойством — с параметром расширения $\varepsilon < 1/8$ и $\varepsilon < 1/4$ соответственно. Однако, чем меньше параметр ε, тем труднее построить граф с требуемым свойством расширения. Поэтому возникает вопрос: можно ли организовать быстрое исправление ошибок для кодов на экспандерах с более слабыми параметрами? Оказывается, что это возможно для экспандеров с ε превышающими границу $1/4$, если применять несколько более сложный способ исправления ошибок. В этой главе мы рассмотрим алгоритм декодирования, работающий за линейное время для кодов на $(n, m, d, k, \varepsilon)$-экспандерах с $\varepsilon < 1/3$.

Замечание: В [15] показано, что рассматриваемый ниже алгоритм работает даже для экспандеров с параметром ε чуть больше $1/3$ (точнее, для $\varepsilon = \frac{1}{3} + \frac{6}{d}$). Остаётся неизвестным, можно ли быстро декодировать экспандерные коды на графах с большими ε.

Работа алгоритм декодирования, который мы сейчас опишем, состоит из двух фаз. В первой фазе мы выявляем и «стираем» все биты слова, которые вызывают сомнения. При этом окажется, что мы сотрём все биты, в которых случились ошибки (а также некоторые биты, значения которых на самом деле правильные). Во второй фазе алгоритма мы восстановим правильные значения всех стертых битов.

Первая фаза алгоритма декодирования (стирание подозрительных битов).

Вход алгоритма: набор битов $x_1, \ldots, x_n$, приписанных вершинам левой доли графа.

1. Выбираем пороговое значение $t := (1 - 2\varepsilon)d$.

2. Инициализация:

   $A_0 := \varnothing$,
   $B_0 :=$ множество всех вершин из правой доли графа с ненулевой контрольной суммой

3. Пока можно найти $v \in L \backslash A_i$, у которой $\geqslant t$ соседей лежат в B_i,

   3.1. $A_{i+1} := A_i \cup \{v\}$,

   3.2. $B_{i+1} := B_i \cup \Gamma(v)$

Упражнение 37 *Покажите, что хранение данных можно организовтаь таким образом, что описанная* первая фаза *алгоритма декодирования будет выполнена за время $O(n)$.*

Лемма 11 *Пусть исходный набор битов $\bar{x} = (x_1, \ldots, x_n)$ отличается от некоторого кодового слова $\bar{y} = (y_1, \ldots, y_n)$ не более, чем в $k/(1-3\varepsilon)$ позициях. Обозначим A_{final} множество A_i в момент остановки* первой фазы *алгоритма декодирования.*

(а) Все позиции i, в которых биты x_i и y_i различаются, входят в A_{final}.

(б) Множество A_{final} содержит не более k элементов.

Доказательство леммы: (а) Обозначим $E \subset L$ множество всех вершин из левой доли графа для которых биты x_i и y_i различаются (позиции, в которых произошли «ошибки» в кодовом слове). Разделим это множество на две части:

$$E_{good} = E \cap A_{final},\ E_{bad} := E \backslash A_{final}.$$

Предположим, что E_{bad} непусто. Применим к этому множеству Лемму 10:

$$|\Gamma(E_{bad})| \geqslant (1-2\varepsilon)d|E_{bad}|.$$

Следовательно, найдётся вершина $v \in C_{bad}$, у которой не меньше $t = (1-2\varepsilon)d$ соседей являются уединёнными. Каждый из этих соседей либо является уединённым соседом всего E (такой вершине соответствует ненулевая контрольная сумма, и она включается в B_0 на стадии инициализации), либо имеет соседей из E_{good} (такая вершина на одной из итераций алгоритма должна быть включена в B_{i+1} при выполнении шага 4.2.). В любом случае, у такой вершины v должно быть не меньше t соседей, лежащих B_{final}. Но тогда v должна была быть включена в A_{final}. Мы получили противоречие с определением множества E_{bad}.

(б) Заметим, что при инициализации в множество B_0 включается не более $d|E|$ вершин (каждый «ошибочный» бит влияет на значения не более, чем на d контрольных сумм). Далее, на каждой итерации алгоритма к множеству A_i добавляется по одной вершине, т.е., $|A_i| = i$. При этом к множеству B_i на каждой итерации добавляется не более $d - t = 2\varepsilon d$ новых контрольных сумм (для новой вершины v, которую мы на i-ом шаге добавляем к текущему множеству A_i, не меньше t соседей уже были включены в B_i). Следовательно,

$$|\Gamma(A_i)| \leqslant |B_i| \leqslant |B_0| + 2\varepsilon d \cdot i.$$

Предположим, что алгоритм делает не менее k итераций. Применим к множеству A_k свойство расширения:

$$(1-\varepsilon)d|A_k| < |\Gamma(A_k)| \leqslant |B_0| + 2\varepsilon d|A_k| \leqslant d|E| + 2\varepsilon d|A_k|.$$

Получаем

$$|A_k| \leqslant \frac{|E|}{1-3\varepsilon} < k.$$

Но это противоречит тому, что на i-ом шаге множество A_i состоит в точности из i вершин. (Отметим, что здесь мы воспользовались важным ограничением: число ошибок не превосходит $k/(1-3\varepsilon)$). Лемма доказана.

Применив первую фазу алгоритма декодирования, мы сотрём в исходном наборе битов $(x_1, \ldots, x_n)$ все «подозрительные» биты x_i, попавшие в A_{final}. Формально можно сказать, что мы заменяем значение некоторых битов x_i на специальный символ '?'. При этом Лемма 11 (а) гарантирует, что все «ошибочные» биты x_i будут «стёрты» (если в позиции i после завершения первой фазы декодирования стоит не вопросительный знак, а ноль или единица, мы можем быть уверены, что это правильное значение соответствующего бита кодового слова). Кроме того, Лемма 11 (б) говорит, что общее число стертых битов не превосходит k. Теперь мы можем перейти ко второй фазе алгоритма, которая позволит восстановить правильные значения битов кодового слова в «стертых» позициях.

Вторая фаза алгоритма декодирования (восстановление стертых битов).

Вход алгоритма: набор символов $x_1, \ldots, x_n$, приписанных вершинам левой доли графа, $x_i \in \{0, 1, ?\}$.

1. Инициализация: $C_0 :=$ множество всех соседей вершин, помеченных '?'
2. Пока можно найти $w \in C_i$, у которой ровно один сосед $v \in L$ помечен '?',
   - 2.1. меняем пометку x_v на 0 или 1 так, чтобы контрольная сумма в w стала равна нулю,
   - 2.2. $C_{i+1} := C_i \backslash \{w\}$

Упражнение 38 *Покажите, что описанную* вторую фазу *алгоритма декодирования можно выполнить за время $O(n)$.*

Упражнение 39 *Покажите, что все биты, восстановленные из символов '?' при выполнении* второй фазы *алгоритма, совпадают с битами исходного кодового слова.*

Лемма 12 *Вторая фаза алгоритма декодирования останавливается только тогда, когда C_i становится пустым (а все вершины левой доли оказываются помеченными нулями или единицами — пометки '?' исчезают).*

Доказательство леммы: Достаточно заметить, что пока есть множество вершин $v \in L$ с пометками '?' непусто, у этого множества есть хотя бы один уединённый сосед.

Таким образом, мы доказали следующий результат.

Теорема 21 *Пусть $\varepsilon < 1/3$, и на экспандере с параметрами $(n, m, d, k, \varepsilon)$ построен код как в параграфе 8.1. Тогда описанный двухфазный алгоритм декодирования позволяет исправлять до $k/(1-3\varepsilon)$ ошибок за линейное время.*

8.5 Код Земора

Рассмотрим сбалансированный двудольный граф $G = (L, R, E)$, у которого по m вершин в левой и правой доле, и степень каждой вершины равна d. В графе G имеется $2m$ вершин и $n = dm$ рёбер. Будем считать, что $m \times m$-матрица смежности данного графа совпадает с матрицей спектрального экспандера с параметрами (m, d, γ) для некоторого достаточно малого значения γ.

Далее мы построим с помощью такого графа линейный код с кодовыми словами длины n (параметры кода мы уточним ниже). Прежде всего, каждому из n *рёбер* графа припишем переменную x_i, $i = 1, \ldots, n$ (одни бит кодового слова). Далее каждой *вершине* графа мы припишем некоторое семейство линейных уравнений — уравнений для переменных x_i (над полем из двух элементов), сопоставленных рёбрам, выходящим из данной вершины. Совокупность всех таких уравнений (для всех вершин графа) и будет задавать пространство кодовых слов.

Для начала выберем *проверочную матрицу* H для линейного $[d, k', \alpha d]$-кода с максимальным возможным k'. По теорема Варшамова–Гилберта можно взять $k' \approx d(1 - h(2\delta))$. Матрица H имеет размер $(d - k') \times d$ (она состоит из $\approx (dh(\alpha))$ линейно независимых d-битовых строк.) Для фиксированного d такую матрицу можно найти перебором. Далее, сопоставим каждой вершине нашего графа систему уравнений, состоящую в умножении H на вектор-столбец, составленный из x_i, приписанных рёбрам, выходящим из данной вершины. Наборы битов $x_1, \ldots, x_n$, удовлетворяющие всем этим уравнениям и образуют кодовые слова кода Земора.

В данном коде каждой вершине приписано $k' \approx dh(\alpha)$ уравнений. Общее число уравнений, таким образом, составляет примерно $2mdh(\alpha)$. Это значит, что размерность пространства решений этой системы $\approx n - 2mdh(\alpha) = (1 - 2h(\alpha))n$. Таким образом, скорость кода равна $(1 - 2h(2\delta))$.

Остаётся понять, сколько ошибок позволяет исправлять данный код.

Утверждение 9 *Расстояние построенного линейного кода не меньше, чем* $\alpha(\alpha - \gamma)n$ *(*$\approx \alpha^2 n$ *при малом* γ*).*

Доказательство: Рассмотрим ненулевое кодовое слово $\bar{x} = (x_1 \ldots x_n)$ с минимальным числом единиц (минимальным весом). Обозначим E' множество рёбер, соответствующих ненулевым битам этого кодового слова. Назовём A и B множества вершин из левой и правой долей графа, в которые входит хотя бы одно ребро из E'. Поскольку набор битов $\bar{x}$ является кодовым словом, он удовлетворяет уравнениям, приписанным всем вершинам графа. Но матрица H устроена так, что она может аннулировать только такой d-битовый вектор, в котором есть хотя бы αd единиц (либо уж *все* биты вектора $\bar{x}$ равны нули). Следовательно, в каждую из вершин A и B входит не менее αd рёбер E'. Таким образом, $|E'| \geqslant \frac{\alpha}{2}(|A| + |B|)$ (в знаменателе стоит коэффициент 2, т.к. у ребра есть два конца — в левой и в правой доле).

Поскольку двудольный граф G построен из (m, d, γ)-экспандера, мы можем применить лемму о перемешивании:

$$|E'| \leqslant |E(A,B)| \leqslant \frac{d|A||B|}{m} + \gamma d\sqrt{|A||B|}$$

Получаем

$$\frac{\alpha}{2}(|A|+|B|) \leqslant \frac{|A||B|}{n} + \gamma\sqrt{|A||B|} \leqslant \frac{(|A|+|B|)^2}{4n} + \frac{\gamma}{2}(|A|+|B|)$$

(среднее геометрическое не превосходит среднего арифметического). Следовательно, $|A|+|B| \geqslant 2m(\alpha-\gamma)$. Вспомним, что $|E'| \geqslant \frac{\alpha}{2}(|A|+|B|)$. Получается, что число рёбер в E' не может быть меньше $\alpha(\alpha-\gamma)n$, и утверждение доказано.

Таким образом, мы научились строить линейный код с параметрами

$$(n, (1-2h(\alpha))n, \alpha(\alpha-\gamma)n)$$

Данная конструкция имеет смысл если $h(2\delta) < 1/2$; это условие выполнено для $\delta < 0.01$.

Как и для любого линейного кода, для кода Земора нетрудно описать полиномиальный алгоритм кодирования (достаточно применить метод Гаусса и выписать общее решение системы уравнений, задающих пространство кодовых слов). С процедурой декодирования искаженных кодовых слов дело обстоит не так просто. Разумеется, для всякого искаженного набора битов можно полным перебором найти ближайшее (в смысле хэммингвосокго расстояния) кодовое слово. Однако такой перебор требует экспоненциального по n времени. Нас же интересуют быстрые (полиномиальные по n) алгоритмы декодирования. Не известно общего способа быстро декодировать произвольные линейные коды (с исправлением ошибок). Однако для кода Земора быстрый (полиномиальный по n) алгоритм декодирования существует.

Поскольку кодовое расстояние данного кода не меньше $\alpha(\alpha-\gamma)n \approx \alpha^2 n$, можно было бы надеяться исправлять до $\approx \frac{1}{2}\alpha^2 n$ ошибок. Столько хорошего результата мы не добъёмся. Однако мы покажем, что можно быстро исправлять $\approx \frac{1}{4}\alpha^2 n$ ошибок. Более точно, для любого $\varepsilon > 0$ можно построить быстрый алгоритм, который восстанавливает кодовое слово, в котором испорчено (инвертировано) не более $\frac{(1-\varepsilon)\alpha(\alpha-\gamma)n}{4}$ битов.

Процедура декодирования будет устроена следующим образом. Попеременно для всех вершин левой и правой доли графа мы производим локальную процедуру декодирования: для каждой вершины берём все входящие в неё рёбра; если значения d соответствующих переменных x_i не удовлетворяют проверочной матрице H, мы меняем их на ближайшее кодовое слов длины d (так, чтобы все контрольные суммы в данной вершине стали равны нулю). На каждом шаге мы применяем процедуру коррекции к вершинам в одной доле графа, поэтому каждое ребро может участвовать только в одной

такой процедуре, и коллизий не возникает (именно здесь существенно, что граф двудольный). Повторяем процедуру попеременно для левой и правой доли графа, пока все контрольные суммы не обнулятся.

Утверждение 10 *Пусть α достаточно мало ($h(\alpha) < 1/2$). Для произвольного $\varepsilon > 0$ и всех достаточно малых γ и всех n описанный выше итеративный алгоритм декодирования кода Земора исправляет $\frac{(1-\varepsilon)\alpha(\alpha-\gamma)n}{4}$ ошибок.*

Более точно, если $y = y_1 \dots y_n$ отличается от одного из кодовых слов (построенного выше кода) не более чем в $\frac{(1-\varepsilon)\alpha(\alpha-\gamma)n}{4}$ битах, то приведённый алгоритм декодирования сходится через $O(\log n)$ итераций (и выдаёт соответствующее кодовое слово).

Доказательство: Чтобы упростить обозначения, мы ограничимся случаем, когда исходное слово y близко к кодовому слову, состоящему из одних нулей (таким образом, в y не более $\frac{(1-\varepsilon)\alpha(\alpha-\gamma)n}{4}$ единиц).

На каждом шаге процесса декодирования каждое ребро графа помечено нулём или единицей (в самом начале единицами помечено не более $\frac{(1-\varepsilon)\alpha(\alpha-\gamma)n}{4}$ рёбер). Нам нужно доказать, что через $O(\log n)$ шагов все рёбра будут помечены нулями.

Рассмотрим пометки на рёбрах графа после i-ого шага процедуры декодирования (на нечётных шагах мы обрабатываем вершины левой доли графа; на чётных шагах — вершины правой доли). Обозначим A_i и B_i множество вершин в левой и правой долях графа соответственно, в которые после i-ой итерации входит хотя бы одно ребро с пометкой 1. Далее мы докажем, что существует такая константа $c < 1$, что

- для нечётных i (на i-ом шаге обрабатывались вершин левой доли) $|A_{i+1}| \leqslant c|B_i|$
- для чётных i (на i-ом шаге обрабатывались вершин правой доли) $|B_{i+1}| \leqslant c|A_i|$

Из этих неравенств немедленно следует доказываемое нами Утверждение — на каждом шаге число «обеспокоенных» единицами вершин (в той доле графа, где только что произошла очередная коррекция) уменьшается в константу раз. Данные неравенства мы докажем по индукции по i.

Прежде чем проводить индукции, сделаем простое наблюдение. Пусть A_1 есть число вершин правой доли графа, в которые входят единичные рёбра *после* первого шага декодирования (первый шаг декодирования мы применяем к вершинам в левой доле графа). Сравним $|A_1|$ с числом единичных рёбер, которые имелись в кодовом слове *до* 1-го шага декодирования. Если очередной шаг декодирования не обнулила все рёбра, входящие в некоторую вершину, это значит, что в неё входило не менее $\alpha d/2$ единичных рёбер. Таким образом,

$$|A_1| \leqslant \frac{[\text{число единичных рёбер, входивших в вершины } A_1 \text{ до 1-ого шага}]}{\delta d/2}$$

Следовательно,

$$|A_1| \leqslant \frac{\frac{1}{4}(1-\varepsilon)\frac{\alpha}{2}(\alpha-\gamma)n}{\alpha d/2} = \frac{1}{2}\cdot(1-\varepsilon)(\alpha-\gamma)\cdot\frac{n}{d}.$$

Далее по индукции мы покажем, что на каждом четном шаге $i = 2, 4, 6, \ldots$ $|B_i| \leqslant A_{i-1}$, а на на каждом четном шаге $i = 3, 5, 7, \ldots$ $|A_i| \leqslant B_{i-1}$. Таким образом, для каждого i мы можем считать, что по предположению индукции

$$|A_{i-1}| \leqslant \frac{1}{2}\cdot(1-\varepsilon)(\alpha-\gamma)\cdot\frac{n}{d}$$

(если $i-1$ нечетно, и на предыдущем шаге происходила коррекция в вершинах слева) или, соответственно,

$$|B_{i-1}| \leqslant \frac{1}{2}\cdot(1-\varepsilon)(\alpha-\gamma)\cdot\frac{n}{d}$$

(если $i-1$ четно, и на предыдущем шаге происходила коррекция в вершинах справа).

Теперь мы готовы сделать шаг индукции. Рассмотрим случай чётного i, когда происходит исправление ошибок в локальных кодах в вершинах правой доли (для чётных номеров рассуждения симметричны). Применим Лемму о перемешивании:

$$|E(A_i, B_{i+1})| \leqslant \frac{d|A_i|\cdot|B_{i+1}|}{m} + dm\sqrt{|A_i||B_{i+1}|} \leqslant \frac{d|A_i|\cdot|B_{i+1}|}{m} + d\gamma\frac{|A_i|+|B_{i+1}|}{2}$$

(второе неравенство есть переход от среднего геометрического к среднему арифметическому).

Оценим снизу $|E(A_i, B_{i+1})|$. Для того, чтобы некоторая вершина в правой доле после i-ой итерации попала в B_{i+1}, в неё (до текущей процедуры коррекции) должно входить не менее $\alpha d/2$ рёбер с единичными пометками. Правый конец каждого такого ребра по определению лежит в A_i.

Соединим вместе нижнюю и верхнюю оценку для $|E(A_i, B_{i+1})|$ и воспользуемся тем, что по предположению индукции $|A_i| \leqslant \frac{1}{2}\cdot(1-\varepsilon)(\alpha-\gamma)\cdot\frac{n}{d}$; наше неравенство можно переписать в виде

$$(\alpha d/2)|B_{i+1}| \leqslant \frac{1}{2}\frac{(1-\varepsilon)(\alpha-\gamma)\frac{n}{d}}{m}|B_{i+1}| + d\gamma\frac{|A_i|+|B_{i+1}|}{2}.$$

Вспомним, что $n = md$ и получим

$$\frac{|B_{i+1}|}{|A_i|} \leqslant \frac{\gamma/2}{\alpha/2 - \gamma/2 - \frac{1}{2}(1-\varepsilon)(\alpha-\gamma))}.$$

Если γ достаточно мало, то данное отношение меньше 1. Утверждение доказано.

8.6 Надёжные схемы из функциональных элементов*

В этом параграфе мы обсуждаем задачу построения надёжных схем из функциональных элементов. Мы предполагаем, что читатель знаком с понятием схемы из функциональных элементов, вычисляющей булеву функцию $f : \mathbb{B}^n \to \mathbb{B}$. Мы будем предполагать, что зафиксирован некоторый конечный *полный базис* булевых функций B, и каждой внутренней вершине схемы сопоставляется некоторая функция $g \in B$, причём арность g совпадает с входной степенью вершины (строго говоря, нужно ещё зафиксировать соответствие между входящими рёбрами и аргументами g). Входным вершинам схемы (вершинам с входной степенью 0) сопоставляются $x_1, \ldots, x_n$ (аргументы функции, которую должна вычислять схема.

Пусть задана схема из N функциональных элементов, вычисляющая некоторую функцию $f : \mathbb{B}^n \to \mathbb{B}$. Рассмотрим работу данной схемы *с ошибками*. Будем предполагать, что каждый из функциональных элементов независимо от других элементов (и от входов схемы) с некоторой вероятностью ε «портится», становится «неисправным». Будем называть данное распределение сбоев *ε-случайным*. При этом мы не предполагаем, что испорченные функциональные элементы *всегда* возвращают неверное значение (отрицание правильного результата вычислений для заданных аргументов). Мы считаем поведение испорченного элемента непредсказуемым — он может возвращать и правильные, и неправильные значения. Можно полагать, что все неисправные элементы схемы переходят во власть злонамеренного противника, который по своему произволу определяет их выходы. При этом выходы на остальных (исправных) функциональных элементах определяются по обычным правилам.

Определение 9 *Схема из функциональных элементов (ε, δ)-надёжно вычисляет функцию f, если для любого набора входных значений, при ε-случайном выборе элементов, в которых возникает неисправность, с вероятностью не менее $(1 - \delta)$ схема выдаёт правильное значение функции, как бы ни действовал противник.*

Теорема 22 *Для произвольного полного базиса булевых функций B, для всех достаточно малых ε найдётся $\delta = O(\varepsilon)$ такое, что всякая булева функция может быть вычислена (ε, δ)-надёжной схемой в данном базисе.*

Доказательство: Прежде всего заметим, что если теорема верна для одного полного базиса, то она обязана выполняться и для любого другого базиса, быть может с другими ε и δ (поскольку элементы одного базиса можно моделировать блоками, составленными из элементов другого базиса). Без ограничения общности мы можем считать, что наш базис состоит из всех булевых функций трёх аргументов. Мы покажем, что любую обычную булеву схему можно переделать в (ε, δ)-надёжную. Доказательство проведём индукцией по глубине формулы.

Итак, пусть выход (обычной) булевой схемы вычисляется применением функционального элемента b к тройке значений f_1, f_2, f_3. Каждое из значений f_1, f_2, f_3 в свою очередь вычисляются некоторыми подсхемами (быть может, пересекающимися). Глубины этих подсхем заведомо меньше, чем глубина всей схемы; поэтому мы можем считать, что для f_1, f_2, f_3 уже имеются (ε, δ)-надёжные схемы T_1, T_2, T_3. Если к выходам схем T_1, T_2, T_3 применить операцию b, то вероятность получить неверный ответ не превосходит $(3\delta + \varepsilon)$ (итоговый результат может оказаться неверным, если хотя бы одно из значений f_i вычислено неправильно или если неисправность возникла в самом элементе b). Назвоём построенную схему R. Чтобы уменьшить вероятность ошибки, мы изготовим три копии схемы R и применим к выходам этих трёх схем функцию большинства. Вероятность того, что и после этого мы получим ошиочный ответ, не превосходит

$$3(3\delta + \varepsilon)^2 + \varepsilon$$

(ошибка должна случиться хотя бы в двух из трех независимых копий схемы C либо в итоговом вычислении большинства). Для малых ε и подходящего $\delta = O(\varepsilon)$ получаем

$$3(3\delta + \varepsilon)^2 + \varepsilon \leqslant \delta$$

и теорема доказана.

Отметим, что приведённая конструкция может экспоненциально увеличить размер схемы, хотя её глубина увеличивается лишь в константу раз.

Упражнение 40 *Докажите, что для всех достаточно малых ε найдётся $\delta = O(\varepsilon)$ такое, что* функцию большинства

$$majority(x_1, \ldots, x_n) = \begin{cases} 1, & \text{если более половины } x_i \text{ равны } 1, \\ 0, & \text{иначе.} \end{cases}$$

можно вычислить (ε, δ)-надёжной схемой размера $\mathrm{poly}(n)$.

Далее мы докажем более сильный вариант теоремы фон Неймана:

Теорема 23 *Для произвольного полного базиса булевых функций B, для всех достаточно малых ε найдётся $\delta = O(\varepsilon)$ такое, что всякая булева схема C из N элементов может быть переделана (за время* $\mathrm{poly}(N)$*) в (ε, δ)-надёжную схему размера $O(N \log N)$.*

Доказательство: Прежде чем доказывать теорему, введём определение:

Определение 10 *Двудольный граф называется (k, d, α, β)-компрессором, если*

1. *в левой и правой долях графа содержится по k вершин;*
2. *степень каждой вершины равна d;*

3. *пусть A — произвольное множество вершин левой доли графа, и $|A| \leqslant \alpha k$; обозначим B множество таких вершин правой доли графа, у которых большинство соседей принадлежат A; тогда размер B не превосходит βk.*

Лемма 13 (о компрессоре) *Если $4\alpha(\gamma^2 + \alpha) < \beta < 1/2$, то матрица смежности спектрального (k, d, γ)-экспандера задаёт (k, d, α, β)-компрессор (двудольный граф с $2 \times k$ вершинами также задаётся матрицей $k \times k$).*

Отложим доказательство леммы и покажем, как она помогает доказать теорему. Зафиксируем некоторый параметр k (в последствии мы выберем $k = O(\log N)$). Долее, построим (k, d, α, β)-компрессор такой, что $\beta + \Gamma\varepsilon < \alpha$ (константа Γ не зависит от k и определяется соотношением числа d и базиса, над которым мы строим схему; подробнее значение Γ мы обсудим ниже).

Мы преобразовываем заданную нам схему C в эквивалентную ей (ε, δ)-надёжную схему C'. Для этого мы заменим каждый функциональный элемент на некоторый блок из $O(k)$ элементов (устройство такого блока мы сейчас опишем). Если в схеме C выход элемента номер i подавался на вход элементу номер j, то в новой схеме C' от блока номер i к блоку номер j будет идти 'кабель' из k проводов. В идеальной ситуации (когда нет ошибок) сигналы во всех проводах этого кабеля будут одинаковы; более того, это будет ровно тот сигнал, который проходил по соответствующему проводу в исходной схеме (при тех же значениях входов схемы).

Теперь опишем устройство блока, соответствующего одному из элементов схемы C. Мы объясним конструкцию на простейшем примере: пусть в C присутствовал функциональный элемент конъюнкция; наша задача — построить надёжный блок, успешно моделирующий этот функциональный элемент при *умеренном* количестве ошибок. В этот блок будут входить $2k$ сигналов (два кабеля по k проводов). Мы сводим соответствующие провода из этих кабелей (первый с первым, второй со вторым, и т.д.) и для каждой пары вычисляем конъюнкцию. Получаем k результирующих сигналов. Затем пропускаем эти сигналы через *корректор*: это схема с k входами и k выходами; каждый выход вычисляется как *большинство* среди некоторых d входов; а правило, по которому каждому из выходов сопоставляются d входов, есть (k, d, α, β)-компрессор. Отметим, что блок реализуется схемой глубины $O(1)$ и состоит из $O(k)$ функциональных элементов (константы зависят от выбора базиса).

С помощью оценки вероятности больших уклонений (неравенство Чернова) нетрудно показать, что если $k = \Omega(\log N)$, то с большой вероятностью ни в одном из N описанных блоков не случится больше $\Gamma\varepsilon k$ (число Γ определяется глубиной схемы-корректора, т.е. зависит от выбора базиса). В таком случае, если каждый из входных кабелей несет не более αk 'неправильных' сигналов (т.е. сигналов, отличных от значения в соответствующем проводе исходной схемы C), то и среди k выходных сигналов не более αk ошибочных. Действительно, перед применением *корректора* неправильные сигналы обоих входов складываются – их может стать 2α. Затем мы пропускаем сигналы через компрессор, и доля ошибок уменьшается до β. Наконец,

нужно учесть ещё $O(\varepsilon k)$ новых ошибок, которые могли случиться в самом блоке. Всего на выходе имеем долю ошибок $\beta + O(\varepsilon) < \alpha$.

Чтобы закончить конструкцию, нам нужно вычленить из k-жильного кабеля на выходе последнего блока *один* сингнал с ответом. Для этого нам нужно вычислить *большинство* среди значений этих k сигналов. Это можно делать разными способами; например, можно применить «экспоненциальную» конструкцию фон Нейманна (при вычислении функции большинства среди $O(\log N)$ значений данный метод даст схему размера $\mathrm{poly}(\log N)$, см. Упражнение выше).

Доказательство леммы о компрессоре: Пусть $\mathbf{e}_1, \ldots, \mathbf{e}_k$ — ортонормированный собственный базис матрицы M заданного (k, d, γ)-экспандера, а $\lambda_1, \ldots, \lambda_k$ – соответствующие собственные числа. Мы будем считать, что собственные числа упорядочены по убывания абсолютной величины. При этом

$$\mathbf{e}_1 = \frac{1}{\sqrt{k}}(1, 1, \ldots, 1),$$

а $\lambda_1 = d$ (по условию леммы остальные собственные числа по модулю не превосходят γk). Пусть A – некоторое множество вершин графа, и $|A| \leqslant \alpha k$. Обозначим $\mathbf{f} = (f_1, \ldots, f_k)$ характеристический вектор этого множества ($f_i = 1$ если и только если i-ая вершина графа принадлежит A). Ясно, что $||\mathbf{f}||^2 \leqslant \alpha k$. Оценим норму вектора $M\mathbf{f}$.

$$||M\mathbf{f}||^2 = (M\mathbf{f}, M\mathbf{f}) = (\mathbf{f}, M^2\mathbf{f}) = \sum_{i=1}^{k} \lambda_i^2 (\mathbf{f}, \mathbf{e}_i)^2 = \alpha^2 d^2 k + \sum_{i=2}^{k} \lambda_i^2 (\mathbf{f}, \mathbf{e}_i)^2$$

Поскольку все собственные числа кроме первого по модулю не превосходят γk, получаем

$$||M\mathbf{f}||^2 \leqslant \alpha^2 d^2 k + (\gamma d)^2 \sum_{i=2}^{k} (\mathbf{f}, e_i)^2 \leqslant \alpha^2 d^2 k + (\gamma d)^2 ||\mathbf{f}||^2 \leqslant (\alpha^2 d^2 + \alpha \gamma^2 d^2) k.$$

Далее, для выбранного A мы рассмотрим множество B, которое состоит из всех вершин графа, у которых не менее $d/2$ соседей лежат в A. Это значит, что B состоит из таких вершин $i = 1, \ldots, n$, что в i-ой координате вектора $\mathbf{f}' = (M\mathbf{f})$ стоит число не менее $d/2$. Получаем

$$|B| \leqslant \sum_{i=1}^{k} \left(\frac{f_i'}{d/2} \right)^2 \leqslant \frac{4}{d^2} ||M\mathbf{f}||^2. \leqslant 4(\alpha^2 + \alpha\gamma^2) k \leqslant \beta k$$

8.7 Структура данных для хранения множества

В это главе мы применим экспандеры для построения структуры данных, которая позволяет хранить сжатое описание множества и при этом очень

быстро обрабатывать запросы о принадлежности элемента к этому множеству. Строго говоря, эта задача не относится к теории кодирования. Но при её решении мы воспользуемся техникой подобной той, которую мы использовали ранее в этой главе при анализе экспандерных кодов.

Сформулируем интересующую нас задачу более точно. Пусть имеется некоторое множество S, все элементы которого принадлежат универсуму $U = \{1, \ldots, N\}$. Требуется построить такую структуру данных, с помощью которой можно очень быстро отвечать на вопросы вида «$x \stackrel{?}{\in} S$». При этом мы предполагаем, что размер хранимого множества $n = |S|$ много меньше, чем размер универсума (например, $n = \mathrm{poly}(\log N)$ или $n = N^{0.01}$).

На практике для организации структур данных типа *множество* в различных ситуациях применяют разные методы. Самое простое решение — просто хранить массив из N битов (по одному биту для каждого элемента универсума), с единицами в позициях, соответствующих элементам, принадлежащим множеству S, и нулями в позициях, не принадлежащих множеству. При таком способе хранения данных ответить на запрос «$x \stackrel{?}{\in} S$» очень просто, нужно лишь прочитать в массиве знаечние x-ого бита. Очевидный недостаток такого наивного способа хранения множества состоит в том, что размер массива должен быть равен размеру универсума (это слишком расточительно, если $n \ll N$). Другой наивный подход состоит в том, чтобы хранить список элементов множества S (точнее, список *индексов* всех элементов универсума, которые входят в S). Один элемент универсума задается $\log N$ битами, так что при данном способе хранения множества наша «база данных» будет состоять из $n \log N$ битов. Это значение близко к оптимальному. В самом деле, их универсума размера n множно выбрать подмножества из m элементов C_N^n способами. Следовательно, для описания такого множества нам нужно как минимум $\log C_N^N$ битов. Если n много меньше чем N, то

$$\log C_N^n = \Omega(n \log N).$$

Недостаток второго наивного способа хранения множества также очевиден — для обработки одного запроса «$x \stackrel{?}{\in} S$» нам потребуется прочитать из хранилища данных довольного много информации.

Итак, первое наивное решение задачи (хранения множества в виде битового вектора) позволяет организовать очень быструю обработку запросов, но требует избыточного размера используемой памяти. Второе наивное решение позволяет минимизировать размер хранимой структуры данных, но требует значительных усилий при обработке запросов. Нельзя ли объединить достоинства этих двух подоходов и избежать их недостатков? Оказывается, что существует (по крайней мере, теретически) способ, который позволяет хранить множество максимально экономно — в виде набора из $O(n \log N)$ битов, и при этом при обработке каждого запроса вида «$x \stackrel{?}{\in} S$» читать *только один бит* из базы данных[1]. При этом алгоритм, обраба-

[1]Описываемая в этой главе структура данных представляет в основном теоретический интерес и не применяется на практике. Одна из причин — до сих пор не известно

тывающий запросы, будет вероятностным. Это значит, что при обработке запроса $x \stackrel{?}{\in} A$ может допускаться ошибка; однако для каждого x из универсума вероятность ошибки будет меньше некоторого наперёд заданного δ. Предлагаемый способ хранения множеств был предложен Harry Buhrman, Peter Bro Miltersen, Jaikumar Radhakrishnan, Venkatesh Srinivasan в [16].

Теорема 24 *Для любого $\delta > 0$ существует вероятностный полиномиальный алгоритм $\mathcal{A}$ со следующим свойством. Для любого n-элементного множества*

$$S \subset \{1, \ldots, N\}$$

существует такой набор из $O(n \log N)$ битов $X = X(N, S)$, что для произвольного $x \in \{1, \ldots, N\}$ алгоритм $\mathcal{A}(x, n, N)$ запрашивает единственный бит x_j *из X, после чего даёт ответ на вопрос «принадлежит ли элемент x множеству S?» с вероятностью ошибки не больше δ.*

Далее мы опишем структуру данных, существование которой утверждается в Теореме 24. Основным ингридиентом конструкции будет двудольный экспандер.

Пусть граф $G = (L, R, E)$ является двудольным экспандером с параметрами $(N, m, d, k, \varepsilon)$, где размер левой доли N совпадает с размером универсума, граница расширяемого множества $k = 2n$ равна удвоенному размеру множества S, а $\varepsilon = \delta/3$ (треть от допустимой вероятности ошибки). Степень графа d и размер правой доли m мы подберм позднее (заранее скажем, что размер правой доли будет равен $O(|S| \log N)$).

Мы отождествляем вершины левой доли графа с элементами универсума. Соответственно, множество S — это некоторое n-элементное подмножество левой доли графа. Вершины правой доли графа будут соответствовать битам нашей «базы данных». Таким образом, каждой вершине правой доли экспандера будет приписан ноль или единица. Именно это сопоставление нулей и единиц вершинам правой доли графа и есть «закодирование» представление множества S.

Далее мы укажем правило разметки — правило, по которому вершинам правой доли приписываются нули и единицы. При этом мы сможем гарантировать, что для каждой вершины v из левой доли графа выполняется следующее *основное свойство разметки*:

- если $v \in S$, то не меньше $(1 - \delta)d$ соседей вершины v в правой доле графа помечены битом 1;
- если $v \notin S$, то не меньше $(1 - \delta)d$ соседей вершины v в правой доле графа помечены битом 0.

алгоритмически эффективных конструкций двудольных экспандеров с необходимыми свойствами. На практике для хранения множества используются другие методы, например, техника *двойного хэширования* (Fredman, Komlós, Szemerédi, [17]), *cuckoo hashing* и структуры данных, построенные на разного вида *сбалансированных деревьях*.

Теперь можно объяснить, как будет происходить обработка запросов $x \stackrel{?}{\in} S$. Для заданного x (точнее, для вершины левой доли графа, соответствующей элементу x из универсума) мы случайно выбираем выходящее из неё ребро и запрашиваем из базы данных бит, соответствующий правому концу этого ребра. Если пометка полученной вершины равна 1, то мы отвечаем, что x принадлежит S; если же пометка равна 0, то мы отвечаем, что x не принадлежит S. Сформулированное выше *основное свойство разметки* гарантирует, что для каждого x вероятность ошибочного ответа будет не больше δ.

Замечание 1: Экспандер, используемый в конструкции, не зависит от множества S. Точнее, выбор экспандера зависит только от параметров n, N, δ. От выбора множества S зависит лишь разметка правой доли экспандера нулями и единицами.

Замечание 2: Размеры экспандера, который используется в данной конструкции (если мы захотим задать граф списком рёбер) занчительно больше, чем размер универсума. Кажется, что это обесценивает наше стремление к минимизации рамера базы данных. Но нам не обязательно постоянно хранить весь экспандер; достаточно уметь его вычислять по заданным параметрам. Постоянно хранить требуется только разметку правой вершины графа нулями и единицами. А эта информация займёт лишь $O(n \log N)$ битов.

Замечание 3: Для того, чтобы данная конструкция имела какой-либо практический смысл, нам потребовалась бы явная (в сильном смысле) конструкция экспандера, в которой по индексу вершины левой доли и номеру выходящего из неё ребра можно эффективно вычислить индекс вершины в правой доле, являющейся вторым концом ребра. Кроме того, как мы увидим ниже, при построении разметки правой доли графа (сопоставление вершинам правой доли нулей и единиц) нам потребуется некоторое ещё более сильное свойство экспандера. В настоящее время не известно эффективных конструкций экспандеров с требуемыми свойствами и обладающих параметрами, близкими к оптимальным. Именно поэтому предлагаемая структура данных в настоящее время не имеет практических применений.

Для доказательства теоремы осталось объяснить, почему правую долю графа можно разметить нулями и единицами так, чтобы выполнялось нужное нам «основное свойство разметки». Мы воспользуемся следующей леммой.

Лемма 14 *Пусть $\delta > 0$ и граф $G = (L, R, E)$ является двудольным экспандером с параметрами $(m, s, d, k, \delta/3)$. Тогда для любого $S \subset L$ размера не более $k/2$ число вершин $v \in L \backslash S$ таких, что*

$$|\Gamma(v) \cap \Gamma(S)| \geqslant \delta d$$

не превосходит $|S|/2$.

Доказательство леммы: Пусть в $L\backslash S$ найдется множество T, состоящее из $|S|/2$ вершин, у каждой из которых не мнее δd соседей являются также и соседями S. Рассмотрим множество соседей объединения A и B:

$$|\Gamma(S \cup T)| \leqslant d|S| + (1-\delta)d|T| = (1-\delta/3)d(|S| + |T|)$$

(у каждой вершины S не более d соседей, а у каждой вершины T не более $(1-\delta)d$ соседей, не учтённых как соседи S). С другой стороны, по определению экспандера мы имеем

$$|\Gamma(S \cup T)| > (1-\delta/3)d|S \cup T|,$$

и мы получаем противоречие. Лемма доказана.

Теперь мы готовы построить нужную нам разметку на вершинах правой доли. Мы получим её с помощью несложного итеративного алгоритма.

Шаг 1: Пометим всех соседей множества S битом 1, а все остальные вершины правой доли пометим 0. Для такой разметки все запросы для $x \stackrel{?}{\in} S$ будут обрабатываться правильно (с нулевой вероятностью ошибки). Однако для некоторых x из дополнения S вероятность ошибки может оказаться большой (если слишком много соседей x помечено единицей, хотя сама вершина x не принадлежит S). Обозначим T множество всех таких «патологических» вершин:

$$T := \{x \in L\backslash S \ : \ |\Gamma(x) \cap \Gamma(S)| \geqslant \varepsilon d\}.$$

По Лемме 14 число таких вершин будет не больше $|S|/2$.

Шаг 2: Поменяем пометки для некоторых вершин правой доли — сделаем пометки всех соседей T равными нулю. После этого исправления запросы для всех $x \notin S$ будут обрабатываться корректно (теперь для каждой вершины из $L\backslash S$ число соседей с пометкой 1 стало во всяком случае меньше εd, так что вероятность ошибки при обработке запроса также меньше ε). Однако для некоторых $x \in S$ ситуация могла ухудшиться. В самом деле, после исправления некоторых пометок с 1 на 0 у некоторых вершин из S могли появиться соседи, помеченные нулем. Хуже того, у некоторых вершин из S доля таких соседей могло стать больше ε. Обозначим S' множество «патологических» вершин в новой разметке:

$$S' := \{x \in S \ : \ |\Gamma(x) \cap \Gamma(T)| \geqslant \varepsilon d\}.$$

Снова применим Лемму 14 и заключим, что таких вершин будет не больше $|T|/2$.

На шаге 3 мы поменяем пометки всех соседей S' на единицы. После этого для некоторого множества вершин $T' \subset T$ число соседей с пометкой 1 снова станет больше или равно εd. Далее, на шаге 4 мы поменяем пометки соседей B' на нули, но опять появится множество проблемных вершин $S'' \subset S'$, у которых слишком много соседей с пометкой 0, и т.д. На шагах

1,3,5,... данной процедуры мы будем изменять текущую разметку, помечая единицами всех соседей вершин множеств

$$S \supset S' \supset S'' \supset \ldots,$$

а на шагах с номерами $2, 4, 6, \ldots$ мы менять разметку, помечая нулями всех соседей вершин из некоторых множеств

$$T \supset T' \supset T'' \supset \ldots$$

соотвественно. Лемма 14 гарантирует, что каждое очередное $T^{(l)}$ как минимум в два раза меньше предыдущего $S^{(l)}$ и, соответственно, каждое следующее $S^{(l+1)}$ как минимум в два раза меньше предыдущего $T^{(l)}$. Таким образом, на каждом шаге число «проблемных» вершин становится вдвое меньше. Через $\log n$ итерации проблемных вершин не останется вовсе, и процедура смены пометок завершится. Полученная в итоге разметка правой доли графа будет обладать нужным нам свойством.

Напомним, что по Теореме 2 для заданных значений N (размер универсума), $k = 2|S|$ (удвоенный размер множества) и $\varepsilon = \delta/3$ (треть допустимой вероятности ошибки) существует двудольный экспандер с параметрами $(N, m = O(n \log N), d, k = 2n, \varepsilon = \delta/3)$. (Как обычно, в константе в $O(\cdot)$ скрыта зависимость от ε.) Это означает, что размер структуры данных, которую мы построим на таком экспандере (разметка нулями и единицами правой доли графа) будет равен $O(n \log N)$ битов. Теорема доказана.

Замечание: Теперь, когда конструкция полностью описана, становится понятно, какое дополнительное свойство экспандера нам нужно, чтобы требуемую разметку можно было построить эффективно. Нам требуется, чтобы по всякому множеству S вершин из левой доли графа (размера не более k) можно было бы эффективно найти множество всех вершин $x \in L \backslash S$ таких, что

$$|\Gamma(x) \cap \Gamma(S)| \geqslant \delta d.$$

Слово *эффективно* в данном случае означает, что по заданному множеству S мы можем найти список таких вершин за время $\mathrm{poly}(\log N, |S|)$. Можно показать, что таким замечательным свойством обладает экспандер из главы 6. Однако для экспандера из главы 6 размер правой доли оказывается равен $O(n \,\mathrm{poly}(\log N))$ вместо желаемой асимптотики $O(n \log N)$, которая достигается с экспандером из неконструктивного доказательства Теоремы 2.

Литература

[1] S. Hoory, N. Linial, A. Wigderson. Expander graphs and their applications. Bulletin of the AMS, vol. 43, Number 4, Oct. 2006, pp.439–561.

[2] Alexander Lubotzky. Expander Graphs in Pure and Applied Mathematics. Bull. Amer. Math. Soc. 49 (2012), 113-162.

[3] P. Sarnak. Some applications of modular forms. Cambridge University Press, 1990. *Русский перевод:* П. Сарнак. Модулярные формы и их приложения. Москва, Фазис, 1998.

[4] Emmanuel Kowalski. Expander graphs (lecture notes), ETH, 2012. `http://www.math.ethz.ch/~kowalski/expanders.html`

[5] S. Arora, B. Barak. Computational Complexity: A modern Approach. Draft version is available online: `http://www.cs.princeton.edu/theory/complexity/`

[6] N. Alon, J. H. Spencer. The Probabilistic Method. 2nd ed. Wiley-Interscience Publication. *Русский перевод:* Н. Алон, Дж. Спенсер. Вероятностный метод. Бином. Лаборатория знаний, 2007

[7] G. Zémor. On Expander codes. IEEE Trans. on Inf. Theory. 47(2), 835–837, 2001.

[8] O. Reingold. Undirected st-connectivity in log-space. In Proceedings of the 37th Annual ACM Symposium on Theory of Computing, pages 376–385, 2005.

[9] V. Guruswami. Error-correcting Codes and Expander Graphs. SIGACT News Complexity Theory Column 45, 2004.

[10] D. Spielman. Constructing error-correcting codes from expander graphs. In Emerging Applications of Number Theory, IMA volumes in mathematics and its applications, volume 109, 1996.

[11] E. Ben-Sasson, M. Sudan, S. Vadhan, A. Wigderson. Randomness-efficient low-degree tests and short PCPs via epsilon-biased sets. STOC 2003, 612–621.

[12] P. Gács. Book chapter on reliable computation.
http://www.cs.bu.edu/~gacs/papers/iv-eng.pdf

[13] Joel Friedman. A proof of Alon's second eigenvalue conjecture and related problems, Mem. Amer. Math. Soc. 195 (2008), no. 910, viii+100 pp.

[14] A. Broder and E. Shamir. On the second eigenvalue of random regular graphs. In Proc. 28th Symp. Foundations of Computer Sci., pages 286–294, 1987.

[15] Michael Viderman. Linear time decoding of regular expander codes. ACM Transactions on Computation Theory (TOCT), vol. 5, no. 3, 10, 2013.

[16] H. Buhrman, P. B. Miltersen, J. Radhakrishnan, and S. Venkatesh. Are Bitvectors Optimal? SIAM J. Comput., 31(6), 1723-1744, 2002.

[17] Fredman, M.L., Komlós, J., Szemerédi, E.: Storing a sparse table with $O(1)$ worst case access time. Journal of the Association for Computing Machinery 31(3), 538-544 (1984)

[18] Parvaresh, Farzad; Alexander Vardy. Correcting Errors Beyond the Guruswami-Sudan Radius in Polynomial Time. Proceedings of the 2005 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05): 285-294.

[19] Guruswami, V., Umans, C., Vadhan, S.: Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. Journal of the ACM 56(4) (2009)

[20] Александр Шень, Андрей Румянцев, Андрей Ромащенко, Заметки по теории кодирования, МЦНМО, 2011.

[21] O. Gabber and Z. Galil. Explicit constructions of linear-sized superconcentrators. J. Comput. System Sci., 22(3):407-420, 1981. Special issue dedicated to Michael Machtey.

[22] S. Jimbo and A. Maruoka. Expanders obtained from affine transformations. Combinatorica, 7(4):343–355, 1987.

[23] Noga Alon, On the edge-expansion of graphs. Combinatorics, Probability and Computing (1993) 11, 1-10.

[24] Michael Capalbo, Omer Reingold, Salil Vadhan, Avi Wigderson. Randomness conductors and constant-degree lossless expanders. Proceedings of the 34th annual ACM symposium on Theory of computing (2002), 659-668.

[25] Барздинь Я.М., Колмогоров А.Н. О реализации сетей в трехмерном пространстве. Проблемы кибернетики. 1967. Т. 19. с. 261–268.

[26] Л.А. Бассалыго, М.С. Пинскер. О сложности оптимальной неблокирующей коммутационной схемы без перестроения Пробл. передачи информ., 9:1 (1973), 84–87.

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

[28] Г.А. Маргулис. Явные конструкции расширителей. Пробл. передачи информ., 9:4 (1973), 71–80.

[29] С.Б. Гашков, Графы-расширители и их применения в теории кодирования. М: МЦНМО, 2009. с. 70–122. (Математическое просвещение, третья серия).

[30] Mike Krebs and Anthony Shaheen, Expander families and Cayley graphs: A beginner's guide, Oxford University Press, 2011.

[31] Михаил Вялый, Юрий Журавлев, Юрий Флеров. Дискретный анализ. Основы высшей алгебры.

[32] Ф.Дж. Мак-Вильямс, Н.Дж.А. Слоэн. Теория кодов, исправляющих ошибки. М.: Связь, 1979. 744 с.