Алгебро-геометрические коды

Том Хохольд, Якобус Х. ван Линт и Рууд Пеликаан

1 Введение

Рассмотрим геометрический объект $\mathcal{X}$ с подмножеством $\mathcal{P}$, содержащим n точек, которые пронумерованы следующим образом: $P_1, \ldots, P_n$. Предположим, что имеется векторное пространство L над $\mathbb{F}_q$, состоящее из функций на $\mathcal{X}$ со значениями в $\mathbb{F}_q$. То есть, $f(P_i) \in \mathbb{F}_q$ для всех i и $f \in L$. Таким образом, получено отображение вычисления

$$ev_{\mathcal{P}} \colon L \longrightarrow \mathbb{F}_q^n,$$

которое определено так: $ev_{\mathcal{P}}(f) = (f(P_1), \ldots, f(P_n))$. Это отображение вычисления линейно, и потому его образ это линейный код. Этот образ и его дуальный — предметы изучения данной главы. Будут рассмотрены размерность и минимальное расстояние этих кодов и их дуальных. Будут обсуждены алгоритмы декодирования этих кодов.

Поскольку коды определены в такой общности, не так уж много можно сказать об их параметрах. Далее, $\mathcal{X}$ это подмножество афинного или проективного пространства, которое является общим множеством нулей некоторого заданного множества полиномов, называемое *многообразием.* $P_1, \ldots, P_n$будут рациональными точками $\mathcal{X}$, т. е. точками с координатами в $\mathbb{F}_q$. Функциями будут выступать полиномы или рациональные функции, т. е. отношения полиномов. Мы называем построенные выше коды *алгебро-геометрическими* кодами, если какая-либо теория, описывающая многообразие $\mathcal{X}$ даёт границы на размерность векторного пространства L и минимальное расстояние кода.

Классический пример описанной выше ситуации дают коды *Рида—Соломона (РС).* Здесь геометрическим объектом $\mathcal{X}$ является афинная прямая над $\mathbb{F}_q$, точками — n различных элементов $\mathbb{F}_q$ и L это векторное пространство полиномов степени не выше $k-1$ с коэффициентами в $\mathbb{F}_q$. Это векторное пространство имеет размерность k. Такие полиномы имеют не более $k-1$ нулей, потому кодовые слова содержат, как минимум, $n-k+1$ ненулевых элементов. Таким образом, этот код имеет параметры $[n, k, n-k+1]$, если $k \leqslant n$. Длина РС-кода, самое большее, q. Способ получения более длинных кодов это рассмотрение подкодов на подполях или кодов-следов РС-кодов. На этом пути получаются циклические коды.

Если мы возьмём в качестве геометрического объекта $\mathcal{X}$ афинное пространство размерности m над $\mathbb{F}_q$, в качестве множества $\mathcal{P}$ — все q^n точек этого афинного пространства, а в качестве векторного пространства все полиномы степени не выше r, мы получим коды *Рида—Маллера (РМ)* порядка r с m переменными над $\mathbb{F}_q$.

Каждое многообразие имеет размерность и многообразие размерности один называется *алгебраической кривой*. Если $\mathcal{X}$ это алгебраическая кривая над $\mathbb{F}_q$, $\mathcal{P}$ это n различных точек $\mathcal{X}$, которые определены над $\mathbb{F}_q$, и L — векторное пространство рациональных функций с предписанным поведением в нулях и полюсах, мы получаем *геометрические* *коды Гоппы*. Параметры этих кодов определены теоремой Римана—Роха и удовлетворяют следующей границе

$$k + d \geqslant n - 1 - g, \qquad \text{или, эквивалентно,} \qquad d \geqslant n + 1 - k - g,$$

где g это инвариант кривой, называемый *родом*. Лучшие коды получаются из кривых рода нуль. Фактически это расширенные обобщённые РС-коды. Они имеют длину, самое большее, $q + 1$ и потому не могут дать асимптотически хорошие последовательности кодов. Длина n РМ-кодов не ограничена, но k/n или d/n стремятся к нулю, если $n \to \infty$. *Информационная скорость (rate)* $R = k/n$ и $\delta = d/n$ геометрических кодов Гоппы удовлетворяет следующему неравенству

$$R + \delta \leqslant 1 - \frac{g-1}{n}.$$

Потому для хороших геометрических кодов Гоппы нужны кривые с малым родом и большим числом рациональных точек. Путём изучения числа рациональных точек на модулярных кривых над конечными полями было показано, что существуют асимптотически хорошие последовательности геометрических кодов Гоппы, удовлетворяющих границе *Цфасмана—Влэдуца—Цинка* (ЦВЦ)

$$R + \delta \leqslant 1 - \frac{1}{\sqrt{q} - 1}, \qquad \text{когда } q \text{ является квадратом.}$$

Эта граница лучше, чем граница *Варшамова—Гилберта (ВГ)*, когда $q \geqslant 49$. Впервые ВГ-граница могла бы быть улучшена.

В конце восьмидесятых, когда был обобщён алгоритм декодирования РС-кодов, начался активный период исследований алгоритмов декодирования АГ-кодов. РС-коды декодируются вплоть до половины минимального расстояния вначале нахождением позиций ошибок, как нулей полинома, известного под названием полинома локаторов ошибок. Если позиции ошибок известны и их число строго меньше, чем минимальное расстояние, то значения ошибок могут быть получены решением системы линейных уравнений, включающей *синдромы*. Эта идея была обобщена до функций локаторов ошибок на кривых. Итоговый *базовый алгоритм* декодирует вплоть до половины конструктивного кодового расстояния минус род. Метод, названный *голосованием неизвестных синдромов большинством*, даёт алгоритм, который декодирует до половины конструктивного кодового расстояния. Были также разработаны более быстрые алгоритмы декодирования с применением *линейных рекуррентных последовательностей нескольких переменных*. Это многомерное обобщение алгоритма *Берлекэмпа—Месси*.

Теория алгебро-геометрических кодов достаточно разработана и глубока. Рассмотрение алгебраических кривых (или, что то же, *полей алгебраических функций* от одной переменной) в самодостаточной форме лежит за пределами изложения в этой главе. Большая часть теории *модулярных кривых* необходимы, чтобы понять результат об асимптотически хороших последовательностях кодов на этих кривых. Сложность построения этих кодов полиномиальна, но из-за того, что степень полинома велика, всё ещё не удовлетворительна для практических реализаций.

Было сделано несколько попыток дать элементарное изложение этой темы. Это подразумевает более простой способ построения кодов, а также понимания и доказательства их свойств. В случае плоских кривых для подсчёта параметров кодов была использована теорема *Безу*, но для дуальных кодов по-прежнему нужна была теорема *Римана—Роха*. Голосование большинством для неизвестных синдромов даёт новую границу для минимального расстояния. Это было отправной точкой элементарного рассмотрения АГ-кодов и это же легло в основание основной части этой главы.

Более того, это дало своим результатом явное и простое описание асимптотически хороших последовательностей кодов над $\mathbb{F}_q$, когда q является квадратом. Таким образом, теория была радикально упрощена, но она по-прежнему нуждается в теории расширений Артина—Шрайера. Соответствующие коды пока ещё не имеют явного описания, но начало положено.

Наша цель в этой главе на обзор обширной литературы по АГ-кодам, но рассмотрение конструирования и декодирования этих кодов, которые могут быть изложены самодостаточным и элементарным способом.

Ключевая идея в нашем изложении — понятие *порядковой функции*. Эта идея хорошо известна в контексте *вычислительной алгебры* и *базисов Грёбнера*, где интенсивно используются редукционные порядки. Будут даны два других приложения порядковых функций: границы для минимального расстояния и декодирование.

В разделах 3–7 разрабатывается теория для класса кодов вычисления и их дуальных, включая все необходимые определения, теоремы и доказательства, которые используют только линейную алгебру и некоторые элементарные сведения о кольцах многочленов нескольких переменных в качестве бэкграунда.

Классов кодов вычисления и их дуальных содержит коды на многообразиях произвольной размерности и пересекает, таким образом, класс геометрических кодов Гоппы по множеству так называемых одноточечных кодов на кривых.

Часть, касающаяся асимптотически хороших последовательностей АГ-кодов, быть лишь намечена.

Раздел 2 содержит набросок стандартного писания алгебро-геометрических кодов. Раздел 3 знакомит с понятиями порядковых и весовых функций. В разделе 4 определяются и доказываются границы на минимальное расстояние кодов вычисления и их дуальных. Раздел 5 описывает специальные порядковые функции, которые называются весовыми функциями, и ассоциированные с ними полугруппы. Приведены свойства минимального кодового расстояния. Декодирование АГ-кодов рассматривается в разделе 6, где объясняются базовый алгоритм и схема голосования неизвестных синдромов большинством. Раздел 7 даёт быстрый алгоритм декодирования.

Ссылки не включены в основной текст, но каждый раздел заканчивается подразделом с названием Замечания, где как раз приводятся ссылки и немного истории.

Обозначения …

2 Коды на кривых

Коды Рида—Соломона могут быть определены при помощи рассмотрения точек с координатами а $\mathbb{F}_q$ на проективной прямой. Кодовые слова определяются рассмотрением рациональных функций с полюсом ограниченного порядка в заданной точке и взятием значения этих функций в заданных точках как координатах. Классические коды Гоппы определяются вычислением вычетов некоторых функций в данных точках. множество функций ограничено требованиями к нулям и полюсам. Именно эти две идеи мы обобщим в этом разделе. Мы должны изучить алгебраические кривые, найти способ описать ограничения на множество функций, которые мы используем, и обобщить понятие вычета. Мы описываем два класса дуальных между сообой кодов. Наконец, мы рассматрвиаем асимтотически хорошие коды на кривых.

В этом разделе теория намечена лишь в общих чертах и большинство доказательств опущено.

2.1 Алгебраические кривые

Далее $\mathbb{F}$ это алгебраически замкнутое поле. В наших приложениях $\mathbb{F}$ будет алгебраическим замыканием $\mathbb{F}_q$. $\mathbb{A}^n$ будет обозначать n-мерное афинное порстранство с координатами $x_1, x_2, \ldots, x_n$. Аналогично, $\mathbb{P}^n$ будет n-мерным проективным пространством с однородными координатами $x_0, x_1, \ldots, x_n$. Сначала мы обсудим афинный случай. Ситуация в проективном пространстве немного более сложная.

В пространстве $\mathbb{A}^n$ *алгебраические множества* это множества нулей идеалов I кольца $\mathbb{F}[X_1, X_2, \ldots, X_n]$, то есть

$$B = V(I) = \{(x_1, x_2, \ldots, x_n) \in \mathbb{A}^n \mid F(x_1, x_2, \ldots, x_n) = 0 \text{ для всех } F \in I\}.$$

Мы всегда предполагаем, что I *радикален*, это означает, что $F \in I$, если $F^n \in I$ для некоторого $n \in \mathbb{N}_0$, и, таким образом, следуя теореме Гильберта о нулях, I содержит *все* полиномы, зануляющиеся в B. Алгебраическое множество B называется *неприводимым*, если B не может быть записано в виде объединения двух собственных алгебраических подмножеств B. Идеал I называется простым, если $F \in I$ или $G \in I$ для всех F, G, таких что $FG \in I$. Множество $V(I)$ неприводимо тогда и только тогда, когда I простой идеал.

...

2.2 Теорема Безу

Теперь мы рассмотрим пересечение кривой и гиперповерхности в $\mathbb{P}^n$. мы предполагаем, что читатель знаком с тем фактом, что полином от одной переменной степени m имеет не более m корней. Если поле алгебраически замкнуто и если нули считать с кратностями, тогда число нулей равно m. Сейчас мы сформулируем теорему, известную как *теорема Безу*, которая является обобщением этих фактов для полиномов нескольких переменных.

Степень проективной кривой это максимальное число точек в пересечении с гиперповерхностью, не содержащей этой кривой. Таким образом, степень проективной плоской кривой равна степени определяющего её уравнения.

Мы рассмотрим только пересечение неприводимой несингулярной проективной кривой $\mathcal{X}$ степени l и гиперповерхности $\mathcal{Y}$, определённой уравнением $G = 0$ степени m. мы предполагаем, что $\mathcal{X}$ не содержится в $\mathcal{Y}$.

Определение 2.22. Пусть P это точка $\mathcal{X}$. Пусть H это однородная линейная форма, такая что $H(P) \neq 0$. Пусть h это класс эквивалентности H по модулю идеала, определяющего $\mathcal{X}$. Тогда *кратность пересечения* $I(P; \mathcal{X}, \mathcal{Y})$ множеств $\mathcal{X}$ и $\mathcal{Y}$в точке P определена как $v_P(g/h^m)$.

Это определение не зависит от выбора H, так как h/h' единица кольца $\mathcal{O}_P$ для любого другого выбора линейной формы H', ненулевой в P.

Теорема 2.23. *Пусть $\mathcal{X}$ это неприводимая несингулярная проективная кривая степени l и $\mathcal{Y}$ — гиперповерхность степени m в $\mathbb{P}^n$, такая что $\mathcal{X}$ не содержится в $\mathcal{Y}$. Тогда они пересекаются точно в lm точках (считая с кратностями).*

Мы не доказываем эту теорему. Если $\mathbb{F}$ не является алгебраически замкнутым или кривые афинны, тогда они пересекаются, самое большее, в lm точках.

Мы упомянем два следствия этой теоремы.