Алгебраический подход к дискретным косинусным и синусным преобразованиям и их быстрым алгоритмам*

Маркус Пушел и Жозе М. Ф. Моура
(перевод с английского М. И. Вашкевича)

Аннотация

Известно, что дискретное преобразование Фурье (ДПФ), используемое в цифровой обработке сигналов, можно рассматривать в рамках теории представления алгебр как матрицу разложения регулярного модуля $\mathbb{C}[\mathsf{Z}_n] = \mathbb{C}[x]/(x^n - 1)$. Такое описание дает глубокое понимание ДПФ, которое также можно использовать для синтеза быстрых алгоритмов и объяснения их структуры. В данной работе мы представляем алгебраическое описание важного класса дискретных косинусных и синусных преобразований как матриц разложения определенных регулярных модулей, связанных с четыремя видами полиномов Чебышева. Затем мы выведем большинство известных для них быстрых алгоритмов чисто алгебраическими средствами. Мы устанавливаем математический принцип, лежащий за каждым алгоритмом, и проливаем свет на его структуру. Наши результаты показывают, что абстрактная алгебра связана с цифровой обработкой сигналов гораздо сильнее, чем это осознавалось ранее.

Ключевые слова. Дискретное косинусное преобразование (ДКП), дискретное синусное преобразование (ДСП), дискретные тригонометрические преобразования, дискретное преобразование Фурье, БПФ, полиномиальное преобразование, быстрый алгоритм, полином Чебышева, представление алгебр, представление групп, симметрия.

1 Введение

Многие алгоритмы в цифровой обработке сигналов основаны на использовании линейных дискретных преобразований сигнала. Математически такое преобразование является умножением вектора на матрицу $a \mapsto M \cdot a$, где $a \in \mathbb{F}^n$ это дискретный сигнал, а $M \in \mathbb{F}^{n \times n}$ — преобразование над некоторым базовым полем $\mathbb{F}$. Мы будем рассматривать только случай, когда $\mathbb{F} = \mathbb{C}$. Решающим фактором применимости преобразования M является существование быстрого алгоритма, который позволяет снизить вычислительную сложность с $O(n^2)$ операций, которые необходимы для умножения вектора на матрицу, до $O(n \log n)$ (или менее) операций. Задача нахождения этих алгоритмов для различных преобразований явилась важнейшей темой многих исследваний и привела к появлению большого количества публикаций по обработке сигналов и математике.

В данной работе мы представляем алгебраический подход для класса из 16-и тригонометрических преобразований в рамках теории представления алгебр. Мы используем алгебраический метод для получения многих известных для них быстрых алгоритмов. Наши результаты дают представление о структуре и возможности существования данных алгоритмов и расширяют взаимосвязь обработки сигналов и алгебры, которая в настоящее время ограничена дискретным преобразованием Фурье.

1.1 Преобразования и алгоритмы

Наверное самым известным примером преобразования является дискретное преобразование Фурье (ДПФ), которое используется в гармоническом анализе для декомпозиции сигнала на его

*Перевод статьи M. Püshel and J. M. F. Moura *The Algebraic Approach to the Discrete Cosine and Sine Transforms and their Fast Algorithms*, SIAM Journal of Computing 2003, Vol. 32, No. 5, pp. 1280-1316.

частотные составляющие. Важные алгоритмы для ДПФ включая «быстрое преобразование Фурье» (БПФ) были найденные Кули и Тьюки [11] (его истоки обнаруживаются у Гаусса [23]), Рэйдером — алгоритм для последовательности простой длины [39], Виноградом [54] и другими. Обзор алгоритмов вычисления ДПФ содержится, например, в [49].

Другой важный класс преобразований состоит из 8 различных типов дискретных косинусных и синусных преобразований (ДКП и ДСП соответственно), также называемых дискретными тригонометрическими преобразованиями (ДТП). Областью их применения является сжатие видео и изображений [40]. Важные алгоритмы для тригонометрических преобразований были разработаны Ченом, Смитом и Фраликом [7], Вангом [52], Уипом и Рао [55, 56], Веттерли и Нуссбаумером [50], Ли [28], Фейгом [20], Чаном и Хо [6], Стейделом и Таше [46], Фейгом и Виноградом [21].

Каждый из этих алгоритмов был получен путем сложных манипуляций с элементами матрицы преобразования. Данные алгоритмы обладают определенной структурой, благодаря чему их можно кратко записать, как разложение матрицы преобразования в произведение слабозаполненных матриц, используя математические операторы. Например, алгоритм Кули-Тьюки можно записать так:

$$\mathrm{DFT}_{mn} = (\mathrm{DFT}_m \otimes \mathrm{I}_n) \cdot D \cdot (\mathrm{I}_m \otimes \mathrm{DFT}_m) \cdot P, \tag{1.1}$$

приведем также пример для ДКП-2

$$\text{DCT-2}_{2^n} = Q \cdot (\text{DCT-2}_n \oplus \text{DCT-4}_n) \cdot B. \tag{1.2}$$

Обозначения будут даны в параграфе §2; матрицы D, P, Q, B являются слабозаполненными. Оба алгоритма рекурсивны по своей природе.

1.2 Алгебраическая характеристика ДПФ

Известно, что ДПФ (размера n) может быть определено в строго алгебраических терминах как матрица разложения групповой алгебры $\mathbb{C}[\mathsf{Z}_n]$ циклической группы Z_n из n элементов,

$$\mathrm{DFT}_n : \mathbb{C}[\mathsf{Z}_n] \to \mathbb{C} \oplus \ldots \oplus \mathbb{C}, \tag{1.3}$$

или, равносильно, как матрица разложения алгебры $\mathbb{C}[x]/(x^n - 1)$,

$$\mathrm{DFT}_n : \mathbb{C}[x]/(x^n - 1) \to \mathbb{C}[x]/(x - \omega_n^0) \oplus \ldots \oplus \mathbb{C}[x]/(x - \omega_n^{n-1}), \tag{1.4}$$

с подходящим базисом в каждом случае. Данные декомпозиции являются конкретизацией теоремы Веддерберна для полупростой алгебры $\mathbb{C}[\mathsf{Z}_n] \cong \mathbb{C}[x]/(x^n - 1)$. Уравнения (1.3) и (1.4) показывают, что ДПФ несомненно является алгебраическим объектом и, следовательно, позволяет проникнуть в суть его использования в обработке сигналов. Более того, (1.3) и (1.4) можно использовать для получения и объяснения структуры быстрых алгоритмов ДПФ алгбераческими средствами, а не манипулированием матрицей ДПФ. В качестве примера скажем, что (1.1) возникает из пошаговой декомпозиции $\mathbb{C}[\mathsf{Z}_n]$, как это было показано Осландером, Фейгом и Виноградом [2] и Бетом [3].

Дав алгебраическую характеристику ДПФ, перед нами естественным образом возникает вопрос: возможно ли обобщить (1.3) и (1.4) на более широкий класс преобразований? И, в случае положительного ответа, можем ли мы использовать алгебраическую характеристику для получения и объяснения быстрых алгоритмов для данных преобразований?

1.3 За пределами ДПФ

В зависимости от интерпретации ДПФ в (1.3) и (1.4) можно выделить два направления обобщения.

Первое направление — обобщение (1.3) на случай произвольной конечной группы $G \neq \mathsf{Z}_n$ приводит к «Фурье-анализу на группах», который предоставляет обширный класс преобразований и теорию получения быстрых алгоритмов для них. Примеры важных результатов в этой области включают работы Бета [4], Клозена [9], Дьякониса и Рокмора [13], [43]. Хороший обзор данной

области можно найти в работе [10] и в обзорных статьях [29] и [44]. Однако, за несколькими исключениями, преобразования Фурье на группах не соответствуют преобразованиям, используемым при обработке сигналов. Эта проблема дала толчок дальнейшему обобщению, которое сделал Минквиц [31, 32], добавив $\mathbb{C}[G]$-модули, которые дают возможность представления произвольных перестановок. Говорят, что матрица разложения таких модулей обладает «симметрией». Минквиц обнаружил, что ДКП-3 обладает такой симметрией, и получил чисто алгебраическими средствами быстрый алгоритм для него. Данный подход в дальнейшем был развит Эгнером и Пушелом и привел к введению мономиального представления. Были разработаны теория декомпозиции [36, 37] и средства для анализа матрицы на наличие симметрии и автоматического получения факторизации матриц [17, 19, 15]. В [16] данные средства были с успехом применены к нескольким преобразованиям. Среди дискретных тригонометрических преобразований ДКП и ДСП третьего и четвертого типов обладают симметрией. Их можно факторизовать, используя данный метод.

Второе направление обобщения, связаное с распространением (1.4) на случай произвольных полиномов $p(x) \neq x^n - 1$ и произвольных базисов в $\mathbb{C}/p(x)$, приводит к классу «полиномиальных преобразований». При произвольном p и выбранном базисе $(1, x, \ldots, x^{n-1})$ получается матрица Вандермонда, для которой, как известно, существует слабозаполненная факторизация, см. [5]. Дрискол, Хили и Рокмор [14] разработали быстрый алгоритм для случая произвольного (приводимого) полинома p и базиса, состоящего из последовательности ортогональных полиномов. Независимо от них, Поттс, Стейдл и Таше представили численно стабильный вариант этого алгоритма [35]. В этой статье определено, что ДКП-1 является полиномиальным преобразованием. Стейдл и Таше [46] также выявили, что ДКП-3 является полиномиальным преобразованием и использовали это свойство для получения быстрого алгоритма. В различных контекстах ДКП и ДСП 1–4 типов были отнесены к полиномальным преобразованиям, в некоторых случаях это делалось после надлежащей нормализации [26].

Собирая все воедино, мы сталкиваемся со следующей ситуацией в отношении ДТП.

1) Существует 16 типов ДТП и большое количество публикаций по быстрым алгоритмам для них.
2) В обработке сигналов ДТП могут быть определены как матрицы составленные из собственных векторов некоторых линейных, инвариантых во времени процессов с заданными граничными условиями [33].
3) Показано, что четыре ДТП обладают групповой симметрией, и для каждого из них быстрый алгоритм выводится чисто алгебраическими средстами.
4) Доказано, что два ДТП являются полиномиальными преобразованиями (заметим, что это свойство не эквивалентно пункту 3). В одном случае это было использовано для получения быстрого алгоритма. Более того, обнаружено, что шесть ДТП после определенной нормализации также являются полиномиальными преобразованиями.

Данные обстоятельства формируют каркас для результатов, представленных в данной работе.

1.4 Алгебраическая характеристика ДТП

В настоящей работе мы дадим алгебраическую характеристику ДТП. Будет показано, что как и ДПФ, ДТП являются алгебраическими объектами. Далее мы используем алгебраическую структуру для получения и объяснения наиболее известных быстрых алгоритмов ДТП. Данные результаты расширяют нашу предыдущую работу [38].

В частности, мы представляем следующее:

1) Полную алгебраическую характеристику всех 16-и ДТП как масштабированных полиномиальных преобразований (дается определение обобщенному понятию полиномиального преобразования), которая возникает из полиномиальных алгебр $A = \mathbb{C}[x]/p(x)$ и A-модулей вида $f \cdot A$, где f — масштабирующая функция. Конструкция этих модулей определяется свойствами дискретных тригонометрических преобразований, как матриц составленных из собственных векторов некоторых линейных инвариантных во времени процессов с заданными граничными условиями. Таким образом, наше построение связывает область цифровой обработки сигналов с теорией представления алгебр. В качестве полиномов здесь естественным образом вступают в действие четыре вида полиномов Чебышева.

2) Полный обзор существующих быстрых алгоритмов и вывод формул для них с использованем чисто алгебраических средств, т.е. выполняя действия с модулями и алгебрами, а не с элементами матриц. Алгоритмы делятся на классы в зависимости от математических принципов, которые лежат в основе их построения. Примеры, основанные на действиях с A-модулем M и полиномом $p(x)$, связанным с ДТП, включают: (1) перевод одного ДТП в другое ДТП путем смены базиса в M; (2) рекурсивную декомпозицию, основанную на факторизации p; и (3) рекурсивную декомпозицию на основе декомпозиции p. Мы продолжаем наши исследования выводом поразительного свойства ДТП. Харрактеристика дискретных тригонометрических преобразований, как полиномиальных преобразований, т.е., в рамках полиномиальной алгебры и её представлений, естественным образом приводит нас к свойству групповой симметрии, т.е. к свойству, относящемуся к теории груп и их представлений. Мы обозначим два пути, которыми групповая симметрия проявляет себя: (1) путем расширения A-модуля M до подходящего $\mathbb{C}[G]$-модуля, где G — конечная группа; и (2) через некоторые подгруппы группы автоморфизмов A. Эти свойства симметрии приводят к алгоритмам, которые структурно отличаются от тех, что получены непосредственным методом (см. выше). Все способы, используемые для получения быстрых алгоритмов ДТП потенциально более общеприменимы.

Собранные воедино, наши результаты представляют собой всеобъемлющую концепцию, которая помещает все полученные ранее результаты по ДТП в единый контекст, связывая воедино свойства ДТП, известные из цифровой обработки сигналов, с алгебраическими свойствами и со структурой их быстрых алгоритмов.

1.5 Организация

Данная статья разделена на две части. Первая часть (§2–§6) представляет собой математический каркас и устанавливает алгебраические интерпретации дискретных тригонометрических преобразований (ДТП). Во второй части (§7–§10) различные алгебраические методы используются для получения и объяснения большинства известных быстрых алгоритмов ДТП.

Часть I. В §2 мы кратко описываем используемые обозначения и математические концепции. Полиномиальные преобразования и масштабированные полиномиальные преобразования вводятся в §3 вместе с их модульно-теоретической интерпретацией. В §4 мы представляем обобщение полиномов Чебышева, уделяя особое внимание их четырем разновидностям. Шестнадцать типов ДТП вводятся в параграфе §5, где определяются их свойства с точки зрения цифровой обработки сигналов. В §6, используя эти свойства, мы выводим для каждого ДТП, отвечающий ему модуль, который показывает, что ДТП является полиномиальным преобразованием.

Часть II. В §7 мы представим общий метод получения быстрых алгоритмов для полиномиальных преобразований и обсудим известные из литературы результаты. В §8 мы используем алгебраические свойства ДТП для получения и объяснения некоторых известных быстрых алгоритмов ДТП. Другие классы алгоритмов для ДТП объяснены в §9 на основе симметрии представлений групп. В §10 мы кратко обсуждаем алгоритмы, которые не охватываются предыдущими методами.

2 Обозначение и терминология

Будем использовать следующую терминологию и математические понятия.

Матрицы: Матрица $(n \times n)$ с элементом $a_{k,\ell}$ в k-ой строке и ℓ-ом столбце записывается как $[a_{k,\ell}]$. В большинстве случаем мы записываем диапазоны для k и ℓ в виде подстрочных индексов. Через $A \oplus B = \left[\begin{smallmatrix} A & \\ & B \end{smallmatrix}\right]$ мы обозначаем прямую сумму A и B. Если $A = [a_{k,\ell}]$, то $A \otimes B = [a_{k,\ell} \cdot B]$ обозначает тензорное или кронекерово произведение матриц A и B. Сопряжение записывается как $A^B = B^{-1} \cdot A \cdot B$. Мономиальная матрица имеет в точности один ненулевой элемент в каждой строке и в каждом столбце. Если σ перестановка (обычно записываемая в циклической нотации), то соответствующая ей матрица $(n \times n)$ обозначается $[\sigma, n]$ и содержит единицы в позициях $[i, \sigma(i)]$. Специальный случай $\sigma\colon i \mapsto ki \bmod n-1$, для $i = 0 \dots n-2$, и $n-1 \mapsto n-1$, когда $k \mid n$ называется «шаговой перестановкой» и записывается $[\sigma, n] = \mathrm{L}^n_k$. Диагональная матрица записывается

как $\mathrm{diag}(L)$, где L — список диагональных элементов. Мономиальная матрица обозначается как $[\sigma, L] = [\sigma, |L|] \cdot \mathrm{diag}(L)$.

Полиномы: Полиномы обозначаются строчными латинскими буквами, например $p(x)$, $q(x)$ и т.д. Для удобства обычно мы будем опускать аргумент. Полином называется приводимым, если его корни попарно различаются. Т.о., если $\deg(p) = n$, то

$$p(x) = \prod_{k=0}^{n-1}(x - \alpha_k), \quad \alpha_i \neq \alpha_j, \text{ для } i \neq j.$$

Алгоритмы: Если B это матрица $(n \times n)$, то под «быстрым алгоритмом для B» мы подразумеваем быстрый алгоритм вычисления произведения вектора на матрицу $x \mapsto B \cdot x$. Алгоритмы задаются факторизациями $B = B_1 \dots B_k$, где все B_i — слабозаполненные матрицы. Если мы обращаемся к арифметической стоимости алгоритма или арифметической сложности матрицы B, то имеем в виду число сложений и умножений на коэффициенты отличные от $1, -1$ (сравни с [5]).

Алгебры и модули: Предполагается, что читатель знаком с базовой теорией алгебр и модулей. Примерами вводных книг по данной тематике являются [12, 25]. Все алгебры в данной статье являются $\mathbb{C}$-алгебрами. В частности, мы рассмотрим полиномиальные алгебры $\mathbb{C}[x]$ и факторалгебры $\mathbb{C}[x]/p(x)$, где p — приводимый полином, и групповую алгебру $\mathbb{C}[G]$, где G — конечная группа. Каждая из алгебр $A = \mathbb{C}/p(x)$ либо $A = \mathbb{C}[G]$ является конечномерной и полупростой, а каждый конечномерный (левый либо правый) A-модуль может быть разложен в прямую сумму неприводимых подмодулей:

$$M \cong M_1 \oplus \dots M_k,$$

такое разложение называется декомпозицией Веддерберна модуля M. Если базисы в M и M_i, $i = 1, \dots k$ выбраны, то этот изоморфизм выражается в виде матрицы, которую мы будем называть *Веддерберновой матрицей*. Если другого не указано, будем считать M левым модулем. Если M имеет размерность n такую же, как и векторное пространство над $\mathbb{C}$, и выбран базис, то M отвечает матричное представление A, т.е. гомоморфизм

$$\phi\colon A \to \mathbb{C}^{n \times n}.$$

Декомпозиция Веддерберна модуля M эквивалентна декомпозиции ϕ в прямую сумму неприводимых представлений. В особом случае, когда $M \cong A$ (как A-модуль), M называют регулярным A-модулем и соответсвующее представление называют регулярным.

Аннулятор M в A определяется как

$$\mathrm{ann}_A(M) = \{a \in A \mid a \cdot m = 0, \quad \text{для всех } m \in M\};$$

и является двухсторонним идеалом в A. Если M это A-модуль, то M есть также $A/\mathrm{ann}_A(M)$-модуль.

3 Полиномиальные алгберы, модули и преобразования

В данном разделе вводятся полиномиальные преобразования и их алгебраическая интерпретация в виде матрицы разложения полиномиальной алгебры. Затем мы расширяем это определение на более общий класс «масштабированных» полиномиальных преобразований,что позволяет поместить тригонометрические преобразования в рамки теории алгебр.

3.1 Полиномиальные преобразования

Пусть

$$p(x) = \prod_{k=0}^{n-1}(x - \alpha_k)$$

приводимый полином. Тогда алгебра $A = \mathbb{C}/p(x)$ является полупростой и декомпозиция Веддерберна (регулярного модуля) $M = A$ задается Китайской теоремой об остатках

$$\mathbb{C}[x]/p(x) \cong \bigoplus_{k=0}^{n-1} \mathbb{C}[x]/(x - \alpha_k). \tag{3.1}$$

Мы хотим представить изоморфизм в (3.1) в виде матрицы.

Определение 3.1 (Полиномиальное преобразование). *Пусть $b = (p_0, \ldots, p_{n-1})$ есть базисные полиномы для $\mathbb{C}[x]/p(x)$, p — приводимый полином, $\alpha = (\alpha_0, \ldots, \alpha_{n-1})$ — корни p, также предположим, что в одномерных алгебрах $\mathbb{C}[x]/(x - \alpha_k)$ задан базис, состоящий из одного вектора $1 = x^0$. При данных условиях, изоморфизм (3.1) задается матрицей $(n \times n)$*

$$\mathcal{P}_{b,\alpha} = [p_\ell(\alpha_k)]_{k,\ell=0\ldots n-1}\,, \tag{3.2}$$

где k — индекс строки. Веддербернова матрица $\mathcal{P}_{b,\alpha}$ называется полиномиальным преобразованием *относительно полиномов b и выбранных точек α (заметим, что порядок следования базисных полиномов и выбранных точек имеет значение).*

Полиномиальное преобразование $\mathcal{P}_{b,\alpha}$ также можно характеризовать через представление ϕ, определяемое модулем $M = A$ с базисом b. Этот факт является предметом следующей леммы.

Лемма 3.2 *Используем введенные выше обозначения. Путь $p(x) = \prod_{k=0}^{n-1}(x - \alpha_k)$ приводимый полином, $A = \mathbb{C}[x]/p(x)$ и $M = A$ левый (регулярный) модуль с базисом $b = (p_0, \ldots, p_{n-1})$ и полиномиальным преобразованием $\mathcal{P}_{b,\alpha}$. Также пусть ϕ — представление, соответствующее A. Тогда*

(i) *$\mathcal{P}_{b,\alpha}^{-1}$ разлагает ϕ в прямую сумму неприводимых представлений. Точнее*

$$\mathcal{P}_{b,\alpha} \cdot \phi(q(x)) \cdot \mathcal{P}_{b,\alpha}^{-1} = \operatorname{diag}(q(\alpha_0), \ldots, q(\alpha_{n-1})), \quad \text{для } q(x) \in A.$$

Все подобные матрицы разложения имеют вид $\mathcal{P}_{b,\alpha}^{-1} \cdot D$, где D — диагональная обратимая матрица.

(ii) *$\mathcal{P}_{b,\alpha}^{T}$ разлагает ϕ в прямую сумму неприводимых представлений. Точнее*

$$(\mathcal{P}_{b,\alpha}^{T})^{-1} \cdot \phi(q(x)) \cdot \mathcal{P}_{b,\alpha}^{T} = \operatorname{diag}(q(\alpha_0), \ldots, q(\alpha_{n-1})), \quad \text{для } q(x) \in A.$$

Все подобные матрицы разложения имеют вид $\mathcal{P}_{b,\alpha}^{T} \cdot D$, где D — диагональная обратимая матрица.

Доказательство. (i) Матрица $\mathcal{P}_{b,\alpha}$ выражает базис b для модуля $M = A$ в базисе модуля $M' = \bigoplus_{k=0}^{n-1} \mathbb{C}[x]/(x - \alpha_k)$. Следовательно, представление ρ, определяемое модулем M', задается как $\rho = \phi^{\mathcal{P}_{b,\alpha}^{-1}}$. Так как $\mathbb{C}[x]/(x - \alpha_k)$ это подмодули (размерности 1) модуля M, то представление ρ имеет диагональный вид. Проекция $q(x)$ на $\mathbb{C}[x]/(x - \alpha_k)$ есть вычисление $q(\alpha_k)$. Из того, что для $q(x) = x$ все собственные значения $\phi(x)$ различны следует, что матрица разложения задается как $\mathcal{P}_{b,\alpha}^{-1} \cdot D$, где D — диагональная обратимая матрица; (ii) следует из (i), если применить транспонирование. ∎

Замечание. Представление ϕ^T возникает из правого регулярного модуля A.

Пример 3.3 *(Матрица Вандермонда). Пусть $A = M = \mathbb{C}[x]/p(x)$, где p приводимый полином. Рассмотрим специальный случай, когда $b = (x^0, x^1, \ldots, x^{n-1})$. В этом случае полиномиальное преобразование задается как*

$$\mathcal{P}_{b,\alpha} = [\alpha_k^\ell]_{k,\ell=0\ldots n-1},$$

что в точности совпадает с транспонированной матрицей Вандермонда [5].

Далее построим для A регулярное представление ϕ относительно базиса b. Так как алгебра A циклическая (образованная полиномом x), то достаточно определить образ элемента $x \in A$ под действием ϕ. Пусть $p(x) = \sum_{i=0}^{n} \eta_i \cdot x^i$. Тогда

$$x \cdot x^i = x^{i+1}, \quad i = 0, \dots n-2, \text{ и}$$
$$x \cdot x^{n-1} = x \equiv \sum_{i=0}^{n-1} -\eta_i \cdot x^i \bmod p(x).$$

Соответственно, получаем матрицу

$$\phi(x) = \begin{bmatrix} 0 & & & & -\eta_0 \\ 1 & 0 & & & -\eta_1 \\ & \ddots & \ddots & & \vdots \\ & & 1 & 0 & -\eta_{n-2} \\ & & & 1 & -\eta_{n-1} \end{bmatrix},$$

которая является транспонированной сопровождающей матрицей p. Используя лемму 3.2 получаем:

$$\phi(x)^{\mathcal{P}_{b,\alpha}^{-1}} = (\phi(x)^T)^{\mathcal{P}_{b,\alpha}^T} = \operatorname{diag}(\alpha_0, \dots, \alpha_{n-1}).$$

Пример 3.4 *(дискретное преобразование Фурье). Продолжим пример 3.3 и потребуем, чтобы* $p(x) = x^n - 1$, *что влечет за собой* $\alpha_k = e^{2\pi ik/n}$, $k = 0, \dots n-1$. *В этом случае транспонированная матрица Вандермонда совпадает с дискретным преобразованием Фурье (ДПФ) размера* n,

$$\mathrm{DFT}_n = \left[e^{2\pi ik\ell/n}\right]_{k,\ell=0\dots n-1}.$$

Это выражение отождествляет ДПФ с полиномиальным преобразованием. Соответствующее представление ϕ *отображает* x *в циклический сдвиг*

$$\phi(x) = \begin{bmatrix} 0 & & & & 1 \\ 1 & 0 & & & 0 \\ & \ddots & \ddots & & \vdots \\ & & 1 & 0 & 0 \\ & & & 1 & 0 \end{bmatrix},$$

и в соответствии с леммой 3.2 и поскольку ДПФ является симметричным, получаем

$$(\phi(x)^T)^{\mathrm{DFT}_n} = \operatorname{diag}_{k=0}^{n-1}(e^{2\pi ik/n}).$$

3.2 Масштабированное полиномиальное преобразование.

В §3.1 было введено полиномиальное преобразование как Веддербернова матрица регулярного A-модуля M, где $A = M = \mathbb{C}[x]$. Для того, чтобы охватить все дискретные тригонометрические преобразования (ДТП) в рамках алгебраического подхода, мы должны отчасти изменить понятие полиномиального преобразования. Коротко говоря, мы будем рассматривать *масштабированные* полиномиальные преобразования. Последние возникают, когда полиномы p_ℓ в (3.2) заменяются на $f \cdot p_\ell$, где f — комплекснозначная функция. Каждое масштабированное полиномиальное преобразование связано с регулярным модулем $M \cong A$, где M может быть $\neq A$. Начнем со следующего определения.

Определение 3.5 (Масштабированное полиномиальное преобразование). *Пусть* $\mathbb{C}[x]/p(x)$, b, α *имеют тот же смысл, что и в определении 3.1. Далее, пусть* f *— комплекснозначная функция, удовлетворяющая условию* $f(\alpha_k) \neq 0$, $k = 0 \dots n-1$. *Определим* масштабированное *полиномиальное преобразование относительно масштабирующей функции* f, *базиса* b *и выбранных точек* α *как*

$$\mathcal{P}_{f \cdot b,\alpha} = [(f \cdot p_\ell)(\alpha_k)]_{k,\ell=0\dots n-1}. \tag{3.3}$$

С масшиабированным полиномиальным преобразованием $\mathcal{P}_{f\cdot b,\alpha}$ можно связать регулярный модуль следующим образом. Векторное пространство $f \cdot \mathbb{C}[x] = \{f \cdot q \mid q \in \mathbb{C}[x]\}$ естественным образом становится $\mathbb{C}[x]$-модулем по определению

$$r \cdot (f \cdot q) = f \cdot rq, \quad \text{для } r \in \mathbb{C}[x].$$

Пусть p — приводимый полином с корнями $\alpha = (\alpha_0, \ldots, \alpha_{n-1})$ и $A = \mathbb{C}[x]/p$. Тогда $\mathbb{C}[x]\cdot(f\cdot p)$ есть подмодуль $f \cdot \mathbb{C}[x]$ и мы можем построить фактор-модуль $M = f \cdot \mathbb{C}[x]/(\mathbb{C}[x] \cdot (f \cdot p))$. Мы будем записывать это кратко $M = f \cdot A$. Аннулятор модуля M $\mathrm{ann}_{\mathbb{C}[x]}(M) = \mathbb{C}[x] \cdot p$, и, следовательно, M являются A-модулем и если $b = (p_0, \ldots, p_{n-1})$ базис для A, тогда $f \cdot b = (f \cdot b_0, \ldots, f \cdot p_{n-1})$ является базисом для M.

Резюмируем свойства модуля $M = f\cdot A$ и полиномиального преобразования $\mathcal{P}_{f\cdot b,\alpha}$ в следующей лемме.

Лемма 3.6 *Пусть $A = \mathbb{C}[x]/p(x)$, $b = (p_0, \ldots, p_{n-1})$ базис в A и p приводимый полином с корнями $\alpha = (\alpha_0, \ldots, \alpha_{n-1})$. Допустим, что f определена как и выше, и $f(\alpha_k) \neq 0$, $k = 0 \ldots n-1$. Далее, пусть $M = f \cdot A$ с базисом $f \cdot b$ как было определено ранее. Тогда верны следующие утверждения.*

(i) *M является регулярным A-модулем.*
(ii) *Регулярное представление ϕ для A, определяемое модулем M и базисом $f \cdot b$ эквивалентно регулярному представлению для A, определяемому модулем A и базисом b.*
(iii) $\mathcal{P}_{f\cdot b,\alpha} = \mathrm{diag}(f(\alpha_0), \ldots, f(\alpha_{n-1})) \cdot \mathcal{P}_{b,\alpha}$.
(iv) *Представление ϕ диагонализируется преобразованием $\mathcal{P}^{-1}_{f\cdot b,\alpha}$, а представление ϕ^T — преобразованием $\mathcal{P}^{T}_{f\cdot b,\alpha}$.*

Доказательство. (i) следует из того, что $f \not\equiv 0$, $p_i \mapsto f \cdot p_i$, $i = 0 \ldots n-1$ определяет изоморфизм $A \to f \cdot A$; (ii) очевидно; (iii) следует прямо из определений (3.2) и (3.3); (iv) следует из (iii) и леммы 3.2. ∎

Замечание. Масштабированное полиномиальное преобразование $\mathcal{P}_{f\cdot b,\alpha}$ не является Веддерверновой матрицей модуля $f \cdot A$ с базисом $f \cdot b$. Отображение $f \cdot p_k \mapsto p_k$, $k = 0 \ldots n-1$ определяет изоморфизм из $f \cdot A$ в A. Соответствующая матрица (относительно базиса $f \cdot b$ и b) является единичной.

4 Полиномы Чебышева

В данном разделе мы рассмотрим класс обобщенных полиномов Чебышева и их свойства, которые будем использовать на протяжении всей статьи. Начнем с классического случая.

4.1 Классический случай

Классические полиномы Чебышева (первого рода) T_n задаются рекуррентным соотношением

$$T_0(x) = 1,\ T_1(x) = x, \quad T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), \quad n \geq 2. \tag{4.1}$$

T_n является полиномом степени n и его можно записать в аналитическом виде

$$T_n(x) = \cos n\theta,\ \cos\theta = x, \quad \text{для } x \in (-1,\, 1). \tag{4.2}$$

Рекуррентное соотношение (4.1) симметрично, n может пробегать значения в обоих направлениях, таким образом можно получить полиномы Чебышева для отрицательных n. Выполняя это, мы получаем свойство симметрии $T_{-n} = T_n$, которое можно видеть из (4.2). Последовательность $\{T_n \mid n > 0\}$ ортогональна на интервале $(-1,\, 1)$ относительно весовой функции $w(x) = (1-x^2)^{-1/2}$, т.е.

$$\int_{-1}^{1} T_n(x)T_m(x)w(x)dx = 0, \quad \text{для } n \neq m.$$

Таблица 4.1. T_n и U_n для $-2 \le n \le 3$.

n	-2	-1	0	1	2	3
T_n	$2x^2-1$	x	1	x	$2x^2-1$	$4x^3-3x$
U_n	-1	0	1	$2x$	$4x^2-1$	$8x^3-4x$

Из выражения (4.2) мы также легко можем получить корни T_n

$$\cos\frac{(k+1/2)\pi}{n}, \quad k = 0 \dots n-1.$$

Используя рекуррентное соотношение (4.1) с измененными начальными условиями получаем полиномы Чебышева второго рода

$$U_0(x) = 1,\ U_1(x) = 2x, \quad U_n(x) = 2xU_{n-1}(x) - U_{n-2}(x),\ n \ge 2.$$

В аналитическом виде U_n задаются как

$$U_n(x) = \frac{\sin(n+1)\theta}{\sin\theta},\ \cos\theta = x, \quad \text{для } x \in (-1,\ 1),$$

отсюда получаем $U_{-1} = 0$ и симметрию $U_{-n-2} = -U_n$. Для $-2 \le n \le 3$ полиномы T_n, U_n приведены в таблице 4.1.

Законченное введение в теорию полиномов Чебышева и ортогональных полиномов вообще можно найти в книгах Чихара, Жего и Ривлина [8, 42, 48].

4.2 Обобщенные полиномы Чебышева

Рассмотрим множество $\mathcal{C}$ последовательностей полиномов $\{P_n \mid n \in \mathbb{Z}\}$, которые удовлетворяют трехчленной рекурсии

$$P_n(x) = 2xP_{n-1}(x) - P_{n-2}(x). \tag{4.3}$$

Будем называть каждую такую последовательность полиномами Чебышева.

Лемма 4.1 *Пусть $\{P_n \mid n \in \mathbb{Z}\}$ последовательность полиномов Чебышева (для простоты опустим аргумент x). Тогда верны следующие свойства*

(i) $P_n = P_1 \cdot U_{n-1} - P_0 \cdot U_{n-2}$.
(ii) $T_k \cdot P_n = (P_{n+k} + P_{n-k})/2$.

Доказательство. Очевидно, что последовательность P_n однозначно определяется начальными условиями P_0 и P_1. Если P_0, P_1 порождают последовательность P_n, а Q — произвольный полином, тогда $Q \cdot P_0$, $Q \cdot P_1$ порождают последовательность $Q \cdot P_n$. Далее если P'_0, P'_1 порождают последовательность P'_n, то $P_0 + P'_0$, $P_1 + P'_1$ порождают последовательность $P_n + P'_n$.

(i) Первыми рассмотрим начальные полиномы 0, 1 и 1, 0 и получим (сравни с таблицей 4.1)

$$\begin{aligned} P_0 = 1, \quad P_1 = 0\colon &\quad P_n = -U_{n-2} \\ P_0 = 0, \quad P_1 = 1\colon &\quad P_n = -U_{n-1}. \end{aligned}$$

Учитывая предыдущее замечание это доказывает (i).

(ii) Доказывается индукцией по k. Для $k = 0$ это тривиально, для $k = 1$ это определяется рекурсией (4.3) для P_n. Далее мы имеем

$$\begin{aligned} T_{k+1} \cdot P_n &= (2xT_k - T_{k-1}) \cdot P_n \\ &= 2x(P_{n+k} + P_{n-k}/2 - (P_{n+k-1} + P_{n-k+1})/2) \\ &= (P_{n+k+1} + P_{n-k-1})/2. \end{aligned}$$

Таблица 4.2. Четыре рода полиномов Чебышева. Индекс для корней $k = 0 \dots n-1$.

	$n = 0,\ 1$	Аналитический вид	Симметрия	Корни	Весовая функция $w(x)$
T_n	$1,\ x$	$\cos(n\theta)$	$T_{-n} = T_n$	$\cos\frac{(k+\frac{1}{2})\pi}{n}$	$\frac{1}{(1-x^2)^{1/2}}$
U_n	$1,\ 2x$	$\frac{\sin(n+1)\theta}{\sin\theta}$	$U_{-n} = -U_{n-2}$	$\cos\frac{(k+1)\pi}{n+1}$	$(1-x^2)^{1/2}$
V_n	$1,\ 2x-1$	$\frac{\cos(n+\frac{1}{2})\theta}{\cos\frac{1}{2}\theta}$	$V_{-n} = V_{n-1}$	$\cos\frac{(k+\frac{1}{2})\pi}{n+\frac{1}{2}}$	$\frac{(1+x)^{1/2}}{(1-x)^{1/2}}$
W_n	$1,\ 2x+1$	$\frac{\sin(n+\frac{1}{2})\theta}{\sin\frac{1}{2}\theta}$	$W_{-n} = -W_{n-1}$	$\cos\frac{(k+1)\pi}{n+\frac{1}{2}}$	$\frac{(1-x)^{1/2}}{(1+x)^{1/2}}$

На последнем шаге мы использовали (4.3). ∎

Замечание. (1) Оба утверждения в лемме 4.1 остаются в силе, когда полиномы P_n и, следовательно, P_0, P_1 являются произвольными комплексно-значными функциями. (2) Лемма 4.1 показывает, что $\mathcal{C}$ — свободный $\mathbb{C}[x]$-модуль, ранг которого равен двум.

В частности, из всего множества $\mathcal{C}$ нам будут интересны четыре последовательности полиномов T_n, U_n, V_n и W_n, которые получаются при различных начальных условиях. Последовательности T_n и U_n — это полиномы первого и второго рода, которые были введены ранее. Все четыре последовательности можно записать в аналитическом виде, они имеют простые свойства симметрии и ортогональны относительно весовой функции $w(x)$ на интервале $(-1,\ 1)$. Корни для данных полиномов можно получить из их выражений в аналитическом виде. Данные свойства сведены в таблице 4.2. Результаты, относящиеся к V_n и W_n можно найти в [8, стр. 37–39].

В дальнейшем нам понадобятся следующие свойства полиномов Чебышева.

Лемма 4.2 *Следующие выражения верны для всех* $m,\ n \in \mathbb{Z}$.

(i) $T_{nm} = T_n(T_m) = T_m(T_n)$.
(ii) $U_{mn-1} = U_{m-1}(T_n)U_{n-1}$.
(iii) $U_{2m} = V_m \cdot W_m$.
(iv) $W_n(x) = (-1)^n V_n(-x)$.
(v) $T_n(1) = 1$.

Доказательтсво. Следует из выражений для полиномов в аналитическом виде (таблица 4.2) и тригонометрических тождеств. ∎

Завершим данный раздел формулировкой интересных свойств для четырех видов полиномов Чебышева. Пусть P_n — любой из T_n, U_n, V_n, W_n. Тогда, применяя известные тригонометрические тождества, $P_n - P_{n-2}$, $P_n - P_{n-1}$, $P_n + P_{n+1}$ можно выразить используя эти же полиномы. В частности, это позволяет нам определить их корни используя таблицу 4.2. Полный набор тождеств дан в таблице 4.3. Выражения во втором столбце тривиальны и введены для того, чтобы согласовать таблицу с последующим изложением. Для примера во второй строке первого столбца показано, что $U_n - U_{n-2} = 2T_n$.

Замечание. (1) Лишь в немногих литературных источниках все четыре рода полиномов Чебышева появлялись вместе, см. [30]. (2) Используя аналитический вид полиномов T_n, U_n и расширяя их определение на случай рациональных $n \in \mathbb{Q}$, мы можем записать $V_n = T_{n+1/2}T_{1/2}$ и $W_n = U_{n-1/2}/U_{-1/2}$. (3) Полный обзор факторизаций T_n и U_n над $\mathbb{Q}$ дан в [41].

5 16 типов дискретных тригонометрических преобразований

Первое дискретное косинусное преобразование было введено Ахмедом, Натараяном и Рао [1]. Полный же набор из 8 типов дискретных косинусных и синусных преобразований был впервые

Таблица 4.3. Тождества для полиномов Чебышева; P_n заменяется на T_n, U_n, V_n или W_n для получения столбца 1, 2, 3 или 4, соответственно.

	$P_n - P_{n-2}$	P_n	$P_n - P_{n-1}$	$P_n + P_{n+1}$
T_n	$2(x^2-1)U_{n-2}$	T_n	$(x-1)W_{n-1}$	$(x+1)V_{n-1}$
U_n	$2T_n$	U_n	V_n	W_n
V_n	$2(x-1)W_n$	V_n	$2(x-1)U_{n-1}$	$2T_n$
W_n	$2(x+1)V_n$	W_n	$2T_n$	$2(x+1)U_{n-1}$

представлен в работе Ванга и Хунта [53]. Мы будем называть их дискретными тригонометрическими преобразованиями. Каждое из преобразований задается $(n \times n)$-матрицей M, $n \geq 0$, на которую слева умнажается вектор a: $a \mapsto M \cdot a$. К примеру, чтобы сослаться на ДКП второго типа мы будем использовать символьное обозначение DCT-2, DST-7$_n$ для обозначения дискретного синусного преобразования седьмого типа размера n. Если рассматривается произвольное дискретное тригонометрическое преобразование, мы будем писать DTT или DTT$_n$. В данных обозначениях первое ДКП, введенное в 1974 году было второго типа.

В таблице 5.1 определены 16 типов дискретных косинусных и синусных преобразований путем задания элементов их матриц в позициях (k, ℓ), где k — номер столбца, ℓ — номер строки и $k, \ell = 0 \ldots n-1$. Легко заметить, что все элементы матриц являются синусами или косинусами вида $\cos r\pi$ или $\sin r\pi$, где r — некоторое рациональное число. Таким образом, числа во всех позициях матрицы являются элементами соответствующего циклотомического поля над $\mathbb{Q}$. Определения, данные в таблице 5.1, относятся к *немасштабированным* версиям ДКП и ДСП, которые будут рассмотрены в данной статье. *Масштабированные* версии ДКП и ДСП являются ортонормированными и получаются из немасштабированных преобразований умножением первой и/или последней строки и/или столбца на $1/\sqrt{2}$, что делает матрицу ортогональной. Также, чтобы получить ортонормированную матрицу, необходимо умножить её на соответствующий коэффициент. Например, элементы ортонормированной версии DCT-2 равны

$$\sqrt{\frac{2}{n}} \cdot c_k \cdot \cos k\left(\ell + \frac{1}{2}\right)\frac{\pi}{n}, \qquad k, \ell = 0 \ldots n-1,$$

где $c_k = \sqrt{1/2}$ для $k = 0$ и $c_k = 1$ иначе. Для удобства читателя масштабированные и ортонормированные версии ДКП и ДСП приведены в приложении (таблица A.1). Заметим, что множество всех ДТП замкнуто относительно операции транспонирования. Из таблицы 5.1 легко заметить, что ДКП и ДСП 1-го, 4-го, 5-го и 8-го типов симметричны, а преобразования 2-го и 3-го типа и 6-го и 7-го типов переходят друг в друга при транспонировании.

Все 16 типов ДКП и ДСП возникают как матрицы, составленные из собственных векторов определенных трехдиагональных матриц размера $(n \times n)$ [47, 45], которые можно записать в форме

$$B(\beta_1, \beta_2, \beta_3, \beta_4) = \frac{1}{2} \cdot \begin{bmatrix} \beta_1 & \beta_2 & & & & & \\ 1 & 0 & 1 & & & & \\ & 1 & 0 & 1 & & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & 1 & 0 & 1 \\ & & & & & \beta_3 & \beta_4 \end{bmatrix} \tag{5.1}$$

Внутренняя структура $B(\beta_1, \beta_2, \beta_3, \beta_4)$ соответствует уравнению

$$a_k = \frac{1}{2}(a_{k-1} + a_{k+1}), \quad 1 \leq k \leq n-2. \tag{5.2}$$

Переменные β_1, β_2 определяют выбор левых граничных условий, которые показывают как в уравнении (5.2) выбирается a_{-1} при $k = 0$. Рассматриваются четыре различных левых граничных

Таблица 5.1. Восемь типов ДКП и ДСП размера n.

	ДКП	ДСП
Тип 1	$\cos k\ell\frac{\pi}{n-1}$	$\sin(k+1)(\ell+1)\frac{\pi}{n+1}$
Тип 2	$\cos k(\ell+\frac{1}{2})\frac{\pi}{n}$	$\sin(k+1)(\ell+\frac{1}{2})\frac{\pi}{n}$
Тип 3	$\cos(k+\frac{1}{2})\ell\frac{\pi}{n}$	$\sin(k+\frac{1}{2})(\ell+1)\frac{\pi}{n}$
Тип 4	$\cos(k+\frac{1}{2})(\ell+\frac{1}{2})\frac{\pi}{n}$	$\sin(k+\frac{1}{2})(\ell+\frac{1}{2})\frac{\pi}{n}$
Тип 5	$\cos k\ell\frac{\pi}{n-\frac{1}{2}}$	$\sin(k+1)(\ell+1)\frac{\pi}{n+\frac{1}{2}}$
Тип 6	$\cos k(\ell+\frac{1}{2})\frac{\pi}{n-\frac{1}{2}}$	$\sin(k+1)(\ell+\frac{1}{2})\frac{\pi}{n+\frac{1}{2}}$
Тип 7	$\cos(k+\frac{1}{2})\ell\frac{\pi}{n-\frac{1}{2}}$	$\sin(k+\frac{1}{2})(\ell+1)\frac{\pi}{n+\frac{1}{2}}$
Тип 8	$\cos(k+\frac{1}{2})(\ell+\frac{1}{2})\frac{\pi}{n+\frac{1}{2}}$	$\sin(k+\frac{1}{2})(\ell+\frac{1}{2})\frac{\pi}{n-\frac{1}{2}}$

условия: $a_{-1}=a_1$, $a_{-1}=0$, $a_{-1}=a_0$, $a_{-1}=-a_0$. Например, выбор $a_{-1}=a_1$ приводит к тому, что $\beta_1=0$, $\beta_2=2$. Таким же образом, β_3, и β_4 определяют правые граничные условия, которые зависят от выбора a_n в (5.2) при $k=n-1$. Правые граничные условия являются зеркальными отображениями левых граничных условий: $a_n=a_{n-2}$, $a_n=0$, $a_n=a_{n-1}$, $a_n=-a_{n-1}$. Весь набор значений β_1, β_2, β_3, β_4 для 16-и возможных комбинаций граничных условий даны в таблице 5.2. Если граничные условия заданы, $\beta_1,\beta_2,\beta_3,\beta_4$ выбраны и $a=(a_0,\ldots,a_{n-1})^T$, тогда (5.2) для $k=0\ldots n-1$ можно записать как

$$a=B(\beta_1,\beta_2,\beta_3,\beta_4)\cdot a.$$

Замечания. (1) Матрицы $B(\cdot)$ из (5.1) соответствуют линейным инвариантным во времени процессам с граничными условиями [33, 47]. (2) Граничное условие $a_{-1}=0$ и $a_{-1}=-a_0$ есть дискретная версия граничных условий Дирихле; $a_{-1}=a_1$ и $a_{-1}=a_0$ — дискретная версия условий фон Неймана. Аналогично для правых граничных условий [33, 47].

Шестнадцать ДТП соответсвуют различным выборам граничных условий, как показано в таблице 5.2 [47]. Если числа $\beta_1,\beta_2,\beta_3,\beta_4$ (и следовательно правые и левые граничные условия) выбраны из строк k и ℓ таблицы 5.2 соответственно, то соответствующая матрица $B(\beta_1,\beta_2,\beta_3,\beta_4)$ диагонализируется транспонированной матрицей ДТП размера n, которая определена в k-ой строке ℓ-го столбца таблицы 5.3.

Пример 5.1 *В качестве примера выберем левое граничное условие $a_{-1}=a_0$ и правое граничное условие $a_n=a_{n-1}$ и получим $\beta_1=\beta_2=\beta_3=\beta_4=1$. Матрица $(n\times n)$*

$$B(1,1,1,1)=\begin{bmatrix}1&1&&&\\1&0&1&&\\&.&.&.&\\&&1&0&1\\&&&1&1\end{bmatrix}, \tag{5.3}$$

диагонализируется матрицей $\text{DCT-2}_n^T=\text{DCT-3}_n$, *т.е.* $B(1,1,1,1)^{\text{DCT-3}_n}$ *имеет диагональный вид.*

Замечания. (1) ДТП типов 5–8 таже называют «нечетными». (2) В работе [47] рассматриваются матрицы вида $2\text{I}-2B(\cdot)$, а не $B(\cdot)$, которые также приводят к эквивалентным диагонализирующим свойствам. Также определения ДТП транспонированы относительно наших опередений. Мы выбрали оригинальные [53] и общеупотребимые определения.

Таблица 5.2. Значения $\beta_1, \beta_2, \beta_3, \beta_4$ для определения левых и правых г.у.

Левые граничные условия	β_1	β_2
$a_{-1} = a_1$	0	2
$a_{-1} = 0$	0	1
$a_{-1} = a_0$	1	1
$a_{-1} = -a_0$	-1	1

Правые граничные условия	β_3	β_4
$a_n = a_{n-2}$	2	0
$a_n = 0$	1	0
$a_n = a_{n-1}$	1	1
$a_n = -a_{n-1}$	1	-1

Таблица 5.3. Правые и левые граничные условия, отвечающие различным ДКП и ДСП

	$a_n = a_{n-2}$	$a_n = 0$	$a_n = a_{n-1}$	$a_n = -a_{n-1}$
$a_{-1} = a_1$	DCT-1	DCT-3	DCT-5	DCT-7
$a_{-1} = 0$	DST-1	DST-1	DST-7	DST-5
$a_{-1} = a_0$	DCT-6	DCT-8	DCT-2	DCT-4
$a_{-1} = -a_0$	DST-8	DST-6	DST-4	DST-2

6 Алгебраические характеристики ДТП

В данном разделе путем построения соответствующего модуля и базиса мы покажем, что все 16 ДТП являются масштабированными полиномиальными преобразованиями (§5). Чтобы связать данное ДТП и его диагонализирующее свойство (т.е. ассоциированную матрицу $B(\beta_1, \beta_2, \beta_3, \beta_4)$ см. §5) с алгебраической теорией, мы будем строить модуль с базисом b, который имеет представление ϕ, такое что

$$\phi^T(x) = B(\beta_1, \beta_2, \beta_3, \beta_4).$$

Другими словами действие операции x (посредством умножения) на b выражается матрицей $B(\beta_1, \beta_2, \beta_3, \beta_4)$. Лемма 3.6, (iv) обосновывает соответствие между ДТП и модулем, построенным таким образом.

Следующая трехшаговая процедура описывает построение модуля и базиса в нем. Предположим, что задано ДТП с ассоциированой матрицей $B(\cdot)$.

1. *Внутреняя структура* (§6.1): Определяется последовательность полиномов, которая формирует внутреннюю стуктуру $B(\cdot)$, т.е. $\ldots, \frac{1}{2}, 0, \frac{1}{2}, \ldots$ в каждом столбце. На этом шаге вводятся в действие обобщенные полиномы Чебышева.

2. *Левые граничные условия* (§6.2): Фиксируются левые граничные условия. Данный шаг соответствует определению начальных условия для полиномов Чебышева, т.е. задается определенная последовательность полиномов Чебышева.

3. *Правые граничные условия* (§6.3): Фиксируются правые граничные условия, что соответствует выбору подходящего полинома p для модуля (и алгебры) $\mathbb{C}[x]/p$.

6.1 Внутренняя структура

Вначале рассмотрим n-мерный модуль, который обладает структурой, заданной в (5.2). Переписывая уравнение (4.3) в измененной форме, получаем

$$x \cdot P_k = \frac{1}{2}(P_{k-1} + P_{k+1}), \tag{6.1}$$

откуда видно, что ему отвечает регулярный модуль $A = \mathbb{C}[x]/p$, $\deg(p) = n$, если в нем выбран базис $b = (P_0, \ldots, P_{n-1})$, где P_k обобщенные полиномы Чебышева. Другими словами, представление $\phi(x)$, отвечающее модулю A с базисом b, будет иметь внутреннюю структуру, схожую с матрицей, заданной в (5.1).

6.2 Левые граничные условия

Есть четыре различных левых граничных условия (г.у.), связанных с ДТП (см. таблицу 5.3)

$$a_{-1} = a_1,\ a_{-1} = 0,\ a_{-1} = a_0,\ a_{-1} = -a_0. \tag{6.2}$$

Они применяются в граничном случае при $k = 0$ в (5.2). Эквивалентное поведение проявляется в (6.1), если мы выберем четыре специальных вида полиномов Чебышева T_k, U_k, V_k, W_k, введенных в таблице 4.2. Свойства симметрии этих полиномов (сравни с таблицей 4.2) соответствуют левым г.у. в (6.2),

$$T_{-1} = T_1,\ U_{-1} = 0,\ V_{-1} = V_0,\ W_{-1} = -W_0,$$

соответственно. Для примера, *каждый* регулярный модуль $\mathbb{C}[x]/p$ с базисом $(T_0, \ldots, T_{n-1})$ служит носителем левого г.у. $a_{-1} = a_1$.

6.3 Правые граничные условия

Четыре правых граничных условия, связанных с ДТП, зеркально отражают левые г.у. (см. таблицу 5.3)

$$a_n = a_{n-2},\ a_n = 0,\ a_n = a_{n-1},\ a_n = -a_{n-1}. \tag{6.3}$$

Правые г.у. определяются выбором p в $\mathbb{C}[x]/p$. Для примера, чтобы опередлить правое г.у. $a_n = a_{n-2}$, выберем $p = P_n - P_{n-2}$, где $P \in \{T, U, V, W\}$. Таким образом, выбор p относительно (6.3) определяется как

$$P_n - P_{n-2},\ P_n,\ P_n - P_{n-1},\ P_n + P_{n-1}, \tag{6.4}$$

соответственно. Чтобы опередлить корни p в данном случае, и, следовательно, декомпозицию модуля A и связанного с ним полиномиального преобразования, неоходимо обратиться к таблице 4.3, которая охватывает все случаи (6.4) для $P \in \{T, U, V, W\}$.

6.4 Резюме

Прежде чем дать интерпретацию дискретных тригонометрических преобразований как масштабированных полиномиальных преобразований, полезнее будет рассмотреть пример.

Пример 6.1 *(ДСП-3). Выберем левое г.у. $a_{-1} = 0$, что влечет за собой выбор базиса $b = (U_0, \ldots, U_{n-1})$. В качестве правого г.у. выберем $a_n = a_{n-2}$, что приводит к $p = U_n - U_{n-2} = 2T_n$ (см. таблицу 4.3). Декомпозиция регулярного модуля $A = \mathbb{C}[x]/T_n$ (константу 2 можно отбросить) определяется корнями T_n, которые равны $\alpha = (\cos\frac{1}{2}\pi/n, \ldots, \cos(n - \frac{1}{2})\pi/n)$ (см. таблицу 4.2), т.е.*

$$A = \mathbb{C}[x]/(U_n - U_{n-2}) = \mathbb{C}[x]/T_n = \bigoplus_{k=0}^{n-1} \mathbb{C}[x]/(x - \cos(k + \frac{1}{2})\pi/n).$$

Полиномиальное преобразование, выполняющее декомпозицию, задается как

$$\begin{aligned}\mathcal{P}_{b,\alpha} =& [U_\ell(\cos(k + 1/2)\pi/n)]_{k,\ell=0\ldots n-1} \\ =& \left[\frac{\sin(\ell+1)(k+1/2)\pi/n}{\sin(k+1/2)\pi/n}\right]_{k,\ell=0\ldots n-1} \\ =& \operatorname{diag}_{k=0}^{n-1}\left(\frac{1}{\sin(k+1/2)\pi/n}\right) \cdot \text{DST-3}_n,\end{aligned}$$

Таблица 6.1. Обзор ДТП и модулей, связанных с ними. Левые г.у. и правые г.у. даны в первом столбце (значение a_{-1}) и первой строке, соответственно. Данное ДТП_n связанно с модулем $f \cdot \mathbb{C}[x]/Q_n$, где полином Q_n показан под ДТП, а масштабирующая функция f приведена во втором столбце. Базис для $\mathbb{C}[x]/Q_n$ задан в третьем столбце.

			$a_n - a_{n-2}$	a_n	$a_n - a_{n-1}$	$a_n + a_{n-1}$
a_1	1	T_ℓ	**DCT-1**	**DCT-3**	**DCT-5**	**DCT-7**
			$2(x^2-1)U_{n-2}$	T_n	$(x-1)W_{n-1}$	$(x+1)V_{n-1}$
0	$\sin\theta$	U_ℓ	**DST-3**	**DST-1**	**DST-7**	**DST-5**
			$2T_n$	U_n	V_n	W_n
a_0	$\cos\frac{1}{2}\theta$	V_ℓ	**DCT-6**	**DCT-8**	**DCT-2**	**DCT-4**
			$2(x^2-1)W_{n-1}$	V_n	$2(x-1)U_{n-1}$	$2T_n$
$-a_0$	$\sin\frac{1}{2}\theta$	W_ℓ	**DST-8**	**DST-6**	**DST-4**	**DST-2**
			$2(x^2+1)V_{n-1}$	W_n	$2T_n$	$2(x+1)U_{n-1}$

это показывает, что DST-3$_n$ *является масштабированным полиномиальным преобразованием*

$$\text{DST-3}_n = \mathcal{P}_{f\cdot b,\alpha}, \qquad f = \sin\theta,$$

связанным с модулем $f \cdot A$ *и базисом* $f \cdot b$.

Далее построим представление ϕ, *отвечающее модулю* A *и базису* b. *Выполняя построение получаем* $x \cdot U_0 = \frac{1}{2}U_1$, $x \cdot U_\ell = \frac{1}{2}(U_{\ell-1} + U_{\ell+1})$ *для* $\ell = 1 \ldots n-2$ *и* $x \cdot U_{n-1} = U_{n-2}$ *(в* A*). Имеем*

$$\phi(x) = \frac{1}{2}\begin{bmatrix} 0 & 1 & & & \\ 1 & 0 & 1 & & \\ & . & . & . & \\ & & 1 & 0 & 2 \\ & & & 1 & 0 \end{bmatrix}.$$

Лемма 3.6 показывает, что $\phi(x)^T$ *диагонализируется* $\text{DST-3}_n^T = \text{DST-2}_n$, *а именно*

$$\phi^T(x)^{\text{DST-2}_n} = \text{diag}(\cos\frac{1}{2}\pi/n, \ldots, \cos(n-\frac{1}{2})\pi/n).$$

Используя обозначения из §5, $\phi(x)^T = B(0,1,2,0)$, *что соответствует* DST-3, *как и ожидалось(см. таблицы 5.2 и 5.3).*

Замечания. (1) Любопытно, что левые и правые г.у. по внешнему виду задаются независимо (начальные условия и фактор-полином). В §8.1 мы увидим, что это конструкция может быть пересмотрена. (2) Заметим, что граничные условия, соответсвующие *левому* модулю влияют на первый и последний столбец *левого* представления $\phi(x)$. Лемма 3.2 показывает, что *правое* представление ϕ^T раскладывается транспонированной версией соответствующего ДТП, что подтверждает тот факт, что г.у. влияют на первую и последнюю *строку* матрицы $B(\cdot)$ (см. §5). (3) Полиномиально определенные правые г.у. в примере 6.1 можно записать двумя способами $U_n - U_{n-1} = 2T_n$ (см. таблицу 4.3). Левая часть выражения определяет г.у., а правая часть задает декомпозицию $\mathbb{C}[x]/T_n$, которая соответствует корням T_n.

Полное соответствие между ДТП и модулями дано в теореме 6.2. Чтобы дать удобное представление, и поскольку данная информация будет интенсивно использоваться в дальнейшем, мы объединили таблицу 4.3 с таблицей 5.3 и соответствующими масштабирующими функциями в таблицу 6.1.

Теорема 6.2 *Определим четыре масштабирующие функции* $f_1 = 1$, $f_2 = \sin\theta$, $f_3 = \cos\frac{1}{2}\theta$ *и* $f_4 = \sin\frac{1}{2}\theta$, *где* $\cos\theta = x$. *Выберем комбинацию левых и правых г.у. с индексами* i, j *из таблицы 6.1,* $i, j = 1 \dots 4$, *и пусть* DTT_n *— соответствующее ДТП. Обозначим полином указаный под* DTT *в таблице 6.1 через* Q_n, *а его корни через* $\alpha = (\alpha_0, \dots, \alpha_{n-1})$. *Выберем базис* $b = (P_0, \dots, P_{n-1})$ *в* $A = \mathbb{C}[x]/Q_n$, *где* $P = T, U, V, W$ *для* $i = 1, 2, 3, 4$, *соответственно. Тогда*

(i) DTT_n *является масштабированным полиномиальным преобразованием*

$$\mathrm{DTT}_n = \mathcal{P}_{f_i \cdot b, \alpha}$$

связанным с модулем $f_i \cdot A$ *и абзисом* $f_i \cdot b$.

(ii) *Если* ϕ *— представление, отвечающее* A *с базисом* b, *тогда* $\phi(x)^T$ *есть матрица* $B(\cdot)$ *из (5.1), задаваемая выбранными левыми и правыми г.у.*

(iii) *Матрица* $\phi(x)^T$ *диагонализируется* DTT_n^T, *а именно*

$$(\mathrm{DTT}_n^T)^{-1} \cdot \phi(x)^T \cdot \mathrm{DTT}_n^T = \mathrm{diag}(\alpha_0, \dots, \alpha_{n-1}),$$

что означает, что DTT_n^T *является матрицей разложения (правого) регулярного представления* ϕ^T *модуля* A.

Доказательство. Выводится при помощи вычислений, по аналогии с примером 6.1 для всех 16-и случаев. ∎

Замечания. (1) Теорема 6.2 показывает, что ДТП является полиномиальным преобразованием (т.е. не масштабированным) тогда и только тогда, когда оно содержится в первой строке таблицы 6.1. Для DCT-1 и DCT-3 этот факт был установлен в [36] и [46], соответственно. (2) Слабозаполненные матрицы $B(\cdot)$ возникают как образы $T_1 = x$ при (правом) представлении ϕ^T соответствующего модуля. Используя лемму 4.1, (ii) можно вычислить образы $\phi^T(T_k)$, $k = 0 \dots n-1$, которые должны оказаться слабозаполненными матрицами. Это делает $(T_0, \dots, T_{n-1})$ естественным базисом для алгебры A во всех 16 случаях. Образ $\phi^T(a)$ (или $\phi(a)$) произвольного элемента $a = \sum a_k T_k \in A$ имеет структуру, которую обычно называют Теплиц + Ганкель.

С алгебраической характеристикой дискретных тригонометрических преобразований, которая дана в теореме 6.2, мы выходим на позиции, с которых можно вывести и объяснить многие быстрые алгоритмы ДТП, известные из литературы. Это будет предметом всего последующего изложения.

7 Быстрые алгоритмы полиномиальных преобразований

Быстрые алгоритмы для умножения вектора на матрицу для полиномиального преобразования $z \mapsto \mathcal{P}_{b,\alpha} \cdot z$ или, что эквивалентно, разложение $\mathcal{P}_{b,\alpha}$ в произведение слабозаполненных матриц явились предметом нескольких работ. В [36] DCT-3 и действительная и мнимая часть ДПФ были идентифицированы как полиномиальные преобразования, что использовалось для их факторизации. В [14] и [35] алгоритмы со сложностью $O(n \log^2 n)$ получены для случая, когда b — произвольная последовательность ортогональных полиномов и α — список различных точек (в которых производится вычисление). Использование этих результатов совмесно с теоремой 6.2 показывает, что сложность вычисления DTT_n равна $O(n \log^2 n)$. Однако быстрые алгоритмы, извесные из литературы и те, что обсуждаются в дальнейшем показывают, что на самом деле сложность равна $O(n \log n)$.

В данном разделе мы представляем два общих метода, которые можно использовать для факторизации полиномиального преобразования $\mathcal{P}_{b,\alpha}$, связанного с регулярным модулем $\mathbb{C}[x]/p$. Они применимы в случаях

1. $p(x) = q(x) \cdot r(x)$ (факторизация p)
2. $p(x) = q(r(x))$ (декомпозиция p)

Важно знать, когда результирующие матрицы-сомножители являются слабозаполненными. Это подразумевает их быстрое вычисление. Заметим, что возможно факторизовать $\mathcal{P}_{b,\alpha}$ если

3. $p(x) = q(x) \otimes r(x)$ (p является тензорным произведением),

однако мы опустим этот случай, поскольку он не применим к ДТП. Из леммы 3.6 следует, что отыскание быстрых алгоритмов для $\mathcal{P}_{b,\alpha}$ и $\mathcal{P}_{f.b,\alpha}$ выполняются аналогично.

На протяжении всего раздела p — приводимый полином, а α — список его корней.

7.1 Прямая сумма

Прямой путь получения быстрого полиномиального преобразования заключается в рекурсивном разложении полинома p, с использованием того факта, что если $p = q \cdot r$, то

$$\mathbb{C}[x]/p = \mathbb{C}[x]/q \oplus \mathbb{C}[x]/r. \tag{7.1}$$

Это сводит проблему вычисления полиномиального преобразования к вычислению двух полиномиальных преобразований меньшего размера.

Лемма 7.1 *Путь $p = q \cdot r$, предположим, что p, q и r имеют корни α, β и γ соответственно. Пусть также b, c и d — базисы для $\mathbb{C}[x]/p$, $\mathbb{C}[x]/q$ и $\mathbb{C}[x]/r$, соответственно. Тогда*

$$\mathcal{P}_{b,\alpha} = P \cdot (\mathcal{P}_{c,\beta} \oplus \mathcal{P}_{d,\gamma}) \cdot B,$$

где B является матрицей смены базиса $b \to (c, d)$ (конкатенация) соответствущая (7.1) и P — матрица перестановки, выполняющая отображение $(\beta, \gamma) \mapsto \alpha$.

Доказательство. Следует из определения B и P. ∎

Очевидно, что декомпозиция из леммы 7.1 полезна для быстрых алгоритмов только если B слабозаполненная матрица или для неё имеется быстрый алгоритм. Например, быстрый алгоритм для матрицы Вандермонда основан на том, что B в этом случае имеет структуру теплицевой матрицы, которая допускает вычисление с $O(n \log n)$ арифметическими операциями.

7.2 Декомпозиция

Более интересные факториации полиномиальных преобразований получаются, если возможно выполнить *декопозицию* p на два полинома, $p(x) = q(r(x))$. Нам потребуется следующая лемма.

Лемма 7.2 *Пусть p — приводимый полином степени n с корнями $(\alpha_0, \ldots, \alpha_{n-1})$. Предположим, что $p(x) = q(r(x))$, q имеет степень k и r имеет степень ℓ. Тогда для каждого корня β полинома q имеется ровно ℓ корней α_m полинома p таких, что $r(\alpha_m) = \beta$.*

Доказательтсво. Пусть α_m — это корень полинома p. Тогда $0 = p(\alpha_m) = q(r(\alpha_m))$. Тогда r отображает $n = k\ell$ корней p на k корней полинома q. Если β — это один из k корней q, тогда уравнение $r(\alpha_m) = \beta$ имеет максимум $\deg(r) = \ell$ решений α_m и, следовательно, ровно ℓ решений. ∎

Как и в лемме 7.2 пусть степени p, q, r равны соответственно n, k, ℓ и $n = k\ell$. Выберем базис $c = (q_0, \ldots, q_{k-1})$ в $\mathbb{C}[x]/q$ и базис $d = (r_0, \ldots, r_{\ell-1})$ в $\mathbb{C}[x]/r$. Тогда

$$\begin{aligned} b' = (\ & r_0 \cdot q_0(r), \ldots, r_0 \cdot q_{k-1}(r), \\ & r_1 \cdot q_0(r), \ldots, r_1 \cdot q_{k-1}(r), \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & r_{\ell-1} \cdot q_0(r), \ldots, r_{\ell-1} \cdot q_{k-1}(r)\ ) \end{aligned}$$

является базисом для $\mathbb{C}[x]/p$. Используя более короткую запись $p_{j,i,m} = (r_j \cdot q_i(r))(\alpha_m)$, соответствующее полиномиальное преобразование задается как

$$\mathcal{P}_{b',\alpha} = \begin{bmatrix} p_{0,0,0} & \cdots & p_{0,k-1,0} & \cdots & p_{\ell-1,0,0} & \cdots & p_{\ell-1,k-1,0} \\ p_{0,0,1} & \cdots & p_{0,k-1,1} & \cdots & p_{\ell-1,0,1} & \cdots & p_{\ell-1,k-1,1} \\ \ldots\ldots & \ldots & \ldots\ldots & \ldots & \ldots\ldots & \ldots & \ldots\ldots \\ p_{0,0,n-1} & \cdots & p_{0,k-1,n-1} & \cdots & p_{\ell-1,0,n-1} & \cdots & p_{\ell-1,k-1,n-1} \end{bmatrix}.$$

В силу леммы 7.2 для каждого i ровно n чисел $q_i(r(\alpha_m))$, $m = 0 \dots n-1$ будет делится на k групп, равных ℓ. Используя перестановку P из α получим α', тогда $r(\alpha_{i+jk}) = \beta_i$, $i = 0 \dots k-1$, $j = 0 \dots \ell - 1$ т.е.

$$\mathcal{P}_{b',\alpha'} = P \cdot \mathcal{P}_{b',\alpha}.$$

Теперь $\mathcal{P}_{b',\alpha'}$ обнаруживает следующую блочную структуру

$$\begin{aligned} \mathcal{P}_{b',\alpha'} &= [D_{h,j} \cdot \mathcal{P}_{c,\beta}]_{h,j=0\dots\ell-1}, \qquad \text{где} \\ D_{h,j} &= \operatorname{diag}(r_j(\alpha'_{hk}), r_j(\alpha'_{hk+1}), \dots, r_j(\alpha'_{hk+k-1})) \end{aligned}$$

Тогда мы можем записать $\mathcal{P}_{b',\alpha'}$ как

$$\mathcal{P}_{b',\alpha'} = [D_{h,j}]_{h,j=0\dots\ell-1} \cdot (I_\ell \otimes \mathcal{P}_{c,\beta}).$$

Так как $D_{h,j}$ диагональная, $h, j = 0 \dots \ell - 1$ матрица $[D_{h,j}]$ состоит из k блоков $(\ell \times \ell)$ с шагом k. Тогда,

$$[D_{h,j}]^{\mathrm{L}^n_\ell}$$

есть прямая сумма матриц $(\ell \times \ell)$, которые также являются полиномиальными преобразованиями. Учитывая, что $(\mathrm{L}^n_\ell)^{-1} = \mathrm{L}^n_k$ получаем следующую теорему.

Теорема 7.3 *Будем использовать обозначения, введенные выше. Тогда*

$$\mathcal{P}_{b,\alpha} = P \cdot \left(\bigoplus_{i=0}^{k-1} \mathcal{P}_{d,\overline{\alpha}_i} \right)^{\mathrm{L}^n_k} \cdot (\mathrm{I}_\ell \otimes \mathcal{P}_{c,\beta}) \cdot B,$$

где B — матрица, задающая смену базиса $b \to b'$, P — матрица перестановки, и

$$\overline{\alpha}_i = (\alpha'_{0\cdot k+i}, \alpha'_{1\cdot k+i}, \dots, \alpha'_{(\ell-1)\cdot k+i}).$$

Как и в лемме 7.1 значимость этой факторизации при получения быстрого алгоритма для $\mathcal{P}_{b,\alpha}$ зависит от матрицы смены базиса B.

Теорему 7.3 можно интерпретировать как обобщение БПФ Кули-Тьюки, как это станет понятно из следующего примера.

Пример 7.4 (БПФ формата 4). *Рассмотрим случай $p(x) = x^4 - 1 = (x^2)^2 - 1$ при этом $q(x) = x^2 - 1$ и $r(x) = x^2$. В качестве базиса выберем $b = (1, x, x^2, x^3)$ и $c = d = (1, x)$. Корни p задаются списком $\alpha = (1, i, -1, -i)$, а корни q равны $\beta = (1, -1)$. Это такая же ситуация, как и в примере 3.4, где $\mathcal{P}_{b,\alpha} = \mathrm{DFT}_4$, а $\mathcal{P}_{c,\beta} = \mathrm{DFT}_2$. Поскольку $r(1) = r(-1)$ и $r(i) = r(-i)$, то $\alpha' = \alpha$. Далее, $b' = (1, x^2, x, x^3)$ и, следовательно, $B = [(2, 3), 4] = \mathrm{L}^4_2$. Остается вычислить $\mathcal{P}_{d,\overline{\alpha}_0}$, $\mathcal{P}_{d,\overline{\alpha}_1}$:*

$$\mathcal{P}_{d,\overline{\alpha}_0} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \mathrm{DFT}_2, \text{ и } \mathcal{P}_{d,\overline{\alpha}_1} = \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} = \mathrm{DFT}_2 \cdot \operatorname{diag}(1, i).$$

В результате получаем БПФ размера 4,

$$\mathrm{DFT}_4 = (\mathrm{DFT}_2 \otimes \mathrm{I}_2) \cdot \operatorname{diag}(1, 1, 1, i) \cdot (\mathrm{I}_2 \otimes \mathrm{DFT}_2) \cdot \mathrm{L}^4_2.$$

Замечания. Будет полезнее дать алгебраическую интерпретацию теоремы 7.3 используя введенные выше обозначения. Из равенства $p(x) = q(r(x))$ следует, что $A' = \mathbb{C}[r(x)]/p(x) = \mathbb{C}[y]/q(y)$ $(y = r(x))$ является подалгеброй алгебры $A = \mathbb{C}[x]/p(x)$. Имеем

$$A \cong r_0 \cdot A' \oplus \dots \oplus r_{\ell-1} \cdot A'$$

векторное пространство, т.е. $d = (r_0, \dots, r_{\ell-1})$ есть пересечение A' в A. Похожим образом это выполняется для $\mathbb{C}[G]$-модуля (G — группа)[12, стр. 73]. Мы можем построить индуцированный модуль

$$A \otimes_{A'} A' = (r_0 \otimes A') \oplus \dots (r_{\ell-1} \otimes A'),$$

с базисом b'. Модули A и $A \otimes_{A'} A'$ изоморфны и имеют матрицу смены базиса B. Следовательно, теорема 7.3 (для полиномиальных алгебр) эквивалентна теореме 3.33 (для групповой алгебры разрешимых групп) в [37]. Они совпадают для случая $\mathbb{C}[Z_n] \cong \mathbb{C}[x]/(x^n - 1)$ (Z_n циклическая группа порядка n), и приводят к алгоритмам БПФ Кули-Тьюки (см. пример 7.4).

8 Получение быстрых ДТП посредством декомпозиции полиномиальных преобразований

В данном разделе мы получим и поясним несколько различных типов рекурсивных алгоритмов для ДТП прямо из их алгебраической интерпретации. В отличии от выводов, которые даются в литературе, мы не выполняем действий над элементами матриц, наоборот мы получаем алгоритм напрямую из модуля, лежащего в основе данного преобразования. Это делает вывод проще, прозрачнее и позволяет проникнуть в математическую суть и структуру алгоритма.

Алгоритмы представленные в этом разделе можно сгруппировать в следующие категории.

1. *Перевод* (§8.1): Одно ДТП переводится в другое ДТП, используя слабозаполненные матрицы. Выделяются два различных метода.
2. *Прямая сумма* (§8.2): Разложение ДТП в прямую сумму двух ДТП меньшего размера, используя слабозаполненные матрицы. Эти алгоритмы обусловлены леммой 7.1.
3. *Редукция* (§8.3): Декомпозиция ДТП в ДТП того же типа, используя слабозаполненные матрицы. Эти алгоритмы обусловленны теоремой 7.3.

Важно отметить, что мы всегда можем получить из любого заданного быстрого алгоритма новый быстрый алгоритм путем непосредственного применения таких операций, как транспонирование или инверсия, так как они совместимы с $\otimes$ и $\oplus$. Например, факторизацию

$$\text{DCT-2}_n = P \cdot (\text{DCT-2}_{n/2} \oplus \text{DCT-4}_{n/2}) \cdot B$$

можно транспонировать, чтобы получить

$$\text{DCT-3}_n = B^T \cdot (\text{DCT-3}_{n/2} \oplus \text{DCT-4}_{n/2}) \cdot P^T,$$

поскольку $\text{DCT-2}^T = \text{DCT-3}$, а DCT-4 симметрично. Более того, всегда возможно делать локальные изменения в данных формулах. Пусть, например, Q и R являются перестановками. Тогда

$$Q \cdot (\mathrm{I}_n \otimes \mathrm{DFT}_2) \cdot R = (QP^{-1}) \cdot (\mathrm{I}_n \otimes \mathrm{DFT}_2) \cdot (PR)$$

для любой перестановки P, переставляющей блоки (2×2) (всего имеется $n!$ таких перестановок P). Мы рассмотрим алгоритмы, которые можно преобразовать друг в друга подобными «алгебраически эквивалентными» манипуляциями. Сравнение алгоритмов, выведенных в данной работе, с теми, что известны из литературы, необходимо выполнять «по модулю» таких равенств, хотя во многих случаях соответствие будет полным.

8.1 Перевод ДТП

В данном разделе мы обсудим и дадим вывод разреженных отношений между различными типами ДТП. Мы говорим, что DTT_n находится в разреженном отношении к DTT'_n в случае если DTT_n можно получить из DTT'_n используя $O(n)$ операций. Приведем пример разреженного отношения

$$\mathrm{DTT}_n = B_n \cdot \mathrm{DTT}'_n \cdot C_n,$$

где B_n, C_n — слабозаполненные матрицы ($O(n)$ ненулевых элементов). Данные отношения важны для быстрых алгоритмов. Если быстрый алгоритм для DTT_n задан и DTT_n с DTT'_n находятся в разреженном отношении, тогда мы получаем быстрый алгоритм для DTT'_n и наоборот.

Изучая таблицу 6.1 мы находим, что пары преобразований в транспонированных позициях (i, j) и (j, i), $i, j = 1 \ldots 4$ обладают одной и той же ассоциированной алгеброй (т.е. одинаковым полиномом Q_n). Например, DCT-5 и DCT-6 возникают из $\mathbb{C}[x]/(x-1)W_{n-1}$ с различными базисами. Это приводит к концепции дуальности (или двойственности), которая вводится в следующем определении.

Определение 8.1 (Дуальность). *Назовем пару* DTT_n *и* DTT'_n *дуальными, если левые г.у.* DTT_n *соответствуют правым г.у.* DTT'_n *и наоборот. Это эквивалентно тому, что* DTT_n *и* DTT'_n *стоят в транспонированных позициях* (i, j) *и* (j, i) *таблицы 6.1. Если* $i = j$ $\mathrm{DTT}_n = \mathrm{DTT}'_n$ *называется самодвойственным.*

Мы покажем, как дуальность можно использовать для вывода разреженных отношений между преобразованиями.

В §6 мы получали модуль для заданной пары г.у. фиксируя (1) базовую последовательность полиномов Чебышева P_n в зависимости от левого г.у., и (2) в зависимоти от правого г.у. полином p в $\mathbb{C}[x]/p$. Т.к. рекуррентное соотношение (4.3) для полиномов Чебышева симметрично, то это можно выполнить в обратном порядке. Мы проиллюстрируем это на паре DCT-3 и DST-3. DST-3 имеет левые г.у. $a_{-1} = 0$, которые фиксируют базовую последовательность U_ℓ, также для него заданы правые г.у. $a_n = a_{n-2}$, которые фиксируются $p = U_n - U_{n-2} = 0$. Альтернативно можно реализовать те же граничные условия последовательностью T_ℓ, $\ell = -n + 1 \dots 0$. Теперь правые г.у. задаются как $a_{-1} = a_1$, т.е. $T_1 = T_{-1}$, что соответствует $U_n - U_{n-2} = 0$. Левые г.у. фиксируются $p = T_{-n} = T_n = 0$. Соответствие между направленными вперед U_ℓ и направленными в обратную сторону T_ℓ приведены ниже, где вертикальные линии обозначают границы.

$$\begin{array}{l|ccc|l} 0 = U_{-1} & U_0, & \dots, & U_{n-1} & U_n = U_{n-2} \\ 0 = T_{-n} & T_{-(n-1)}, & \dots, & T_0 & T_{-1} = T_1, \end{array}$$

Другими словами, при использовании $T_{-\ell} = T_\ell$, два базиса $(U_0, \dots, U_{n-1})$ и $(T_{n-1}, \dots, T_0)$ задают идентичные представления $A = \mathbb{C}[x]/T_n$. Следовательно, соответствующие полиномиальные преобразования должны быть масштабированными версиями друг друга. И, более того, если через α_k обозначить корни T_n, то используя простые тригонометрические соотношения, получаем

$$\begin{aligned} T_{n-1-\ell}(\alpha_k) &= \cos(n - 1 - \ell)(k + \frac{1}{2})\pi/n \\ &= (-1)^k \cdot \sin(\ell + 1)(k + \frac{1}{2})\pi/n, \end{aligned}$$

и следовательно, используя определения DST-3 и DCT-3 (таблица 5.1),

$$\mathrm{diag}_{k=0}^{n-1}((-1)^k) \cdot \text{DST-3}_n = \text{DCT-3}_n \cdot \mathrm{J}_n, \tag{8.1}$$

где J_n матрица с единицами на побочной диагонали, т.е. матрица перестановки $i \leftrightarrow n - i$, $i = 0 \dots n-1$. Подобные вычисления для всех пар дуальных преобразований приводят к следующему результату.

Теорема 8.2 (Перевод посредством дуальности). *Путь* DTT_n *и* DTT'_n *пара дуальных преобразований. Тогда*

$$\mathrm{diag}_{k=0}^{n-1}((-1)^k) \cdot \mathrm{DTT}_n = \mathrm{DTT}'_n \cdot \mathrm{J}_n.$$

В частности дуальные DTT_n *имеют одинаковую арифметическую сложность.*

Второй класс разреженных отношений можно получить в некоторых случаях путем подходящей смены базиса, что будет объяснено далее. Возвращаясь к таблице 6.1, мы видим, что 16 ДТП разделены на 4 группы по 4 преобразования каждая в зависимости от полинома Q_n, который по существу совпадает с одним из полиномов Чебышева T_n, U_n, V_n, W_n. Например, на главной диагонали в таблице 6.1 находятся ДТП, относящиеся к «U-группе», которая полностью состоит из самодвойственных преобразований. Все преобразования из других групп состоят из двух пар дуальных ДТП. Для каждого из двух ДТП внутри одной группы соответствующие алгебры $\mathbb{C}[x]/Q_n$ по существу равны. Разница заключается только в базисе, выбранном для модуля. Следовательно, становится возможным вывести разреженное отношение производя подходящую смену базиса. Мы иллюстрируем эту мысль на следующих двух примерах.

Пример 8.3 (DCT-3 и DST-3). *Рассмотрим снова пару преобразований* DCT-3$_n$ *и* DST-3$_n$. *Используя таблицу 6.1 мы видим, что оба преобразования соответствуют одной и той же алгебре, но с разными базисами,*

$$\text{DCT-3}_n \leftrightarrow \mathbb{C}[x]/T_n,\ b = (T_0, \dots, T_{n-1}),$$
$$\text{DST-3}_n \leftrightarrow \mathbb{C}[x]/T_n,\ b' = (U_0, \dots, U_{n-1}),$$

также DCT-3$_n$ $= [T_\ell(\alpha_k)]$, *и* DST-3$_n$ $= D \cdot [U_\ell(\alpha_k)]$, *где* α_k *корни* T_n, *а* $D = \operatorname{diag}_{k=0}^{n-1}(\sin(k + \frac{1}{2})\pi/n)$ *определяется масштабирующей функцией. Для вычисления матрицы смены базиса* B *для* $b \to b'$ *мы используем тот факт, что* $T_\ell = \frac{1}{2}(U_\ell - U_{\ell-2})$ *(вторая строка, первого столбца в таблице 4.3), получаем*

$$B = \frac{1}{2} \cdot \begin{bmatrix} 2 & 0 & -1 & & & \\ & 1 & 0 & -1 & & \\ & & . & . & . & \\ & & & 1 & 0 & -1 \\ & & & & 1 & 0 \\ & & & & & 1 \end{bmatrix}.$$

Таким образом, $[T_\ell(\alpha_k)] = [U_\ell(\alpha_k)] \cdot B$, *и, следовательно*

$$D \cdot \text{DCT-3}_n = \text{DST-3}_n \cdot B.$$

Заметим, что это тождество отличается от того, которое возникает из дуальности DCT-3$_n$ *и* DST-3$_n$ *(теорема 8.2).*

Пример 8.4 (DCT-1 и DST-2). *Совершим перевод* DCT-1$_{n+1}$ *в* DST-2$_n$. *Воспользуемся вновь таблицей 6.1, откуда получим связанные с данными преобразованиями алгебры и базисы:*

$$\text{DCT-1}_{n+1} \leftrightarrow \mathbb{C}[x]/(x^2-1)U_{n-1},\ b = (T_0, \dots, T_n),$$
$$\text{DST-2}_n \leftrightarrow \mathbb{C}[x]/(x+1)U_{n-1},\ b' = (W_0, \dots, W_{n-1}).$$

Заметим, что мы выбрали размер $n+1$ *и* n, *соответственно, для того, чтобы получить сравнимые алгебры. Мы имеем следующее* DCT-1$_{n+1}$ $= [T_\ell(\alpha_k)]$ *и* DST-2$_n$ $= D \cdot [W_\ell(\alpha_k)]$, *где* $\alpha_k = \cos k\pi/n$, $k = 0 \dots n$ *это корни* $(x^2-1)U_{n-1}$ *(вторая строка в таблице 4.2). Для* DST-2$_n$, $\alpha_0 = 1$ *пропускается. Масштабирующая матрица* $D = \operatorname{diag}_{k=0}^{n-1}(\sin(k+1)\pi/2n)$ *(теорема 6.2). Вычислим матрицу смены базиса* B *для*

$$\mathbb{C}[x]/(x^2-1)U_{n-1} \cong \mathbb{C}[x]/(x-1) \oplus \mathbb{C}[x]/(x+1)U_{n-1}.$$

Базисами являются $b, (1)$ *и* b', *соответственно. Используя* $T_\ell = \frac{1}{2}(W_\ell - W_{\ell-1})$ *(4-я строка, 3-й столбец в таблице 4.3) и* $T_n = \frac{1}{2}(W_n - W_{n-1}) = -W_{n-1} \bmod (x+1)U_{n-1}$ *(следует из таблицы 4.3,* $(x+1)U_{n-1} = \frac{1}{2}(W_n + W_{n-1})$*) получаем*

$$B = \frac{1}{2} \cdot \begin{bmatrix} 2 & 2 & 2 & . & . & 2 \\ 2 & -1 & & & & \\ & 1 & -1 & & & \\ & & . & . & & \\ & & & . & -1 & \\ & & & & 1 & -2 \end{bmatrix}.$$

Единицы в первой строке получаются поскольку $T_\ell(1) = 1$ *(лемма 4.2). Получаем* $[T_\ell(\alpha_k)] = \mathrm{I}_1 \oplus [W_\ell(\alpha_k)] \cdot B$ *и, следовательно,*

$$(\mathrm{I}_1 \oplus D) \cdot \text{DCT-1}_{n+1} = (\mathrm{I}_1 \oplus \text{DST-2}_n) \cdot B.$$

Таким образом, мы подходим к следующей теореме.

Теорема 8.5 (Перевод путем смены базиса). *Все ДТП 1–4 типов находятся в разреженном отношении. Все ДТП 5–8 типов находятся в разреженном отношении.*

Доказательство. Вычислениями, схожими с примерами 8.3 и 8.4 показывается, что все ДТП 1-го и 2-го типа («U-группа») находятся в разреженном отношении, также все ДТП 3-го и 4-го типа («T-группа») находятся в разреженном отношении. Путем транспонирования можно получить разреженные отношения для ДТП 2-го и 4-го типа и, следовательно, для всех ДТП 1–4 типа, что доказывает первое утверждение. Второе утверждение доказывается аналогично. ■

Замечания. (1) В стороне от определения 8.1 есть другая, более очевидная форма дуальности среди ДТП: DTT и DTT′ дуальны, если $\mathrm{DTT}^T = \mathrm{DTT}'$. В настоящее время мы не имеем алгебраичского объяснения для такой дуальности (или двойственности). (2) Заметим, что «разреженное отношение» не определяет отношения равенства. Любые две матрицы (одного размера) можно преобразовать друг в друга используя достаточно длинную последовательность слабозаполненных матриц.

8.2 Прямая сумма: построение быстрых алгоритмов путем факторизации полиномов

В данном разделе мы выведем рекурсивные алгоритмы для всех ДТП U-группы, т.е. для ДКП и ДСП 1-го и 2-го типа. Все алгоритмы основанны на факторизациях полиномов U_n, которая дана в лемме 4.2, (ii) и (iii).

В качестве примера рассмотрим DCT-2_n, где $n = 2m$. Обратившись к таблице 6.1, находим соответствующую алгебру $\mathbb{C}[x]/(x-1)U_{n-1}$ с базисом $b = (V_0, \ldots, V_{n-1})$. Лемма 4.2, (ii) задает факторизацию $U_{2m-1} = 2 \cdot U_{m-1} \cdot T_m$, которая приводит к изоморфизму

$$\mathbb{C}[x]/(x-1)U_{2m-1} \cong \mathbb{C}[x]/(x-1)U_{m-1} \oplus \mathbb{C}[x]/T_m. \tag{8.2}$$

Выберем для слагаемых базисы b, b' и b' соответственно ($b' = (V_0, \ldots, V_{m-1})$). Корнями $(x-1)U_{n-1}$ являются числа $\cos k\pi/n$, где $k = 0 \ldots n-1$. Следовательно, первое слагаемое в (8.2) вбирает все корни с четными k, а второе слагаемое корни с нечетным k (сравни с таблицей 4.2). Теперь декомпозиция DCT-2_n выполняется согласно лемме 7.1. Для вычисления матрицы смены базиса B в (8.2) нам необходимы тождества

$$V_{m+k} \equiv V_{m-k-1} \bmod (x-1)U_{m-1}, \text{ и}$$
$$V_{m+k} \equiv -V_{m-k-1} \bmod T_m,$$

которые можно доказать по индукции. Воспользуемся тем, что $(x-1)U_{m-1} = V_m - V_{m-1}$, $T_m = V_m + V_{m-1}$, и (4.3), получаем

$$B = \begin{bmatrix} 1 & & & & 1 \\ & . & & . & \\ & & 1 & 1 & \\ 1 & & & & 1 \\ & . & & . & \\ & & 1 & -1 & \end{bmatrix} = \begin{bmatrix} \mathrm{I}_m & \mathrm{J}_m \\ \mathrm{I}_m & -\mathrm{J}_m \end{bmatrix} = (\mathrm{DFT}_2 \otimes \mathrm{I}_m)(\mathrm{I}_m \oplus \mathrm{J}_m).$$

Слагаемые в (8.2) раскладываются рекурсивно с использованием DCT-2_m и DCT-4_m. Получаемые одноразмерные слагаемые приводятся к каноническому виду шаговой перестановкой L^n_m (см. §2). Так как DCT-2 и DCT-4 имеют одну и ту же масштабирующую функцию, мы получаем

$$\text{DCT-}2_{2m} = \mathrm{L}^{2m}_m \cdot (\text{DCT-}2_m \oplus \text{DCT-}4_m) \cdot B.$$

Кроме DCT-2 похожие выводы можно сделать для трех оставшихся преобразований из U-группы DCT-1, DST-1 и DST-2, воспользовавшись факторизациями $U_{2m-1} = 2U_{m-1}T_m$ и $U_{2m} = V_m W_m$.

Полный набор тождеств можно построить используя два типа блочных матриц и два типа матриц перестановки. Блочные матрицы задают смену базиса

$$B_{2m} = \begin{bmatrix} \mathrm{I}_m & \mathrm{J}_m \\ \mathrm{I}_m & -\mathrm{J}_m \end{bmatrix}, \quad B_{2m+1} = \begin{bmatrix} \mathrm{I}_m & 0 & \mathrm{J}_m \\ 0 & 1 & 0 \\ \mathrm{I}_m & 0 & -\mathrm{J}_m \end{bmatrix},$$

матрицы перестановки переупорядочивают неприводимые модули

$$P_{2m} = \mathrm{L}_m^{2m}$$
$$P_{2m+1} \colon i \to (m+1)i \bmod 2m+1, \quad i = 0 \dots 2m.$$

Теорема 8.6 *Следующие рекурсивные алгоритмы для* DTT *основаны на факторизации* $U_{2m-1} = 2 \cdot U_{m-1} \cdot T_m$. *Мы также отметим, где они впервые появляются в литературе (насколько нам это извесно).*

(i) $\text{DCT-1}_{2m+1} = P_{2m+1} \cdot (\text{DCT-1}_{m+1} \oplus \text{DCT-3}_m) \cdot B_{2m+1}$ [27],
(ii) $\text{DST-1}_{2m-1} = P_{2m-1} \cdot (\text{DST-3}_m \oplus \text{DST-1}_{m-1}) \cdot B_{2m-1}$ [53].
(iii) $\text{DCT-2}_{2m} = P_{2m} \cdot (\text{DCT-2}_m \oplus \text{DCT-4}_m) \cdot B_{2m}$ [7].
(iv) $\text{DST-2}_{2m} = P_{2m} \cdot (\text{DST-4}_m \oplus \text{DST-2}_m) \cdot B_{2m}$ [52].

Теорема 8.6 дополняется декомпозициями из следующей теоремы. Мы не нашли их в литературе.

Теорема 8.7 *Следующие рекурсивные алгоритмы для ДТП основаны на факторизации* $U_{2m} = V_m W_m$.

(i) $\text{DCT-1}_{2m} = P_{2m} \cdot (\text{DCT-5}_m \oplus \text{DCT-7}_m) \cdot B_{2m}$.
(ii) $\text{DST-1}_{2m} = P_{2m} \cdot (\text{DST-7}_m \oplus \text{DST-5}_m) \cdot B_{2m}$.
(iii) $\text{DCT-2}_{2m+1} = P_{2m+1} \cdot (\text{DCT-6}_{m+1} \oplus \text{DCT-8}_m) \cdot B_{2m+1}$.
(iv) $\text{DST-2}_{2m+1} = P_{2m+1} \cdot (\text{DST-8}_{m+1} \oplus \text{DST-6}_m) \cdot B_{2m+1}$.

Замечания. (1) Транспонирование декомпозиций из теорем 8.6 и 8.7 приводят к алгоритмам ДТП 3-го типа. (2) Теорема 8.6 показывает почему DCT-1 и DST-1, как правило, имеют длины $2^k + 1$ и $2^k - 1$, соответственно. Доступные алгоритмы более эффективны, поскольку они представляют собой не просто слабозаполненную факторизацию ДТП типов 5–8. (3) Теорема 8.5 и тоерема 8.6 в объединении задают полный набор алгоритмов для ДТП типов 1–4, длины равной степени 2 (для ДТП первого типа длина отличается на единицу, см. замечание (2)). (4) Возможно вывести алгоритмы для более общего случая $n = k\ell$, используя факторизацию из леммы 4.2, (ii).

8.3 Редукция: быстрые алгоритмы посредством декомпозиции полиномов

В данном разделе мы выведем алгоритмы, основанные на декомпозиции полинома T_n из леммы 4.2, (i). Данное свойство декомпозиции позволяет выполнить декомпозицию всех ДТП из T-группы, т.е. ДКП и ДСП 3-го и 4-го типа, используя теорему 7.3.

В качестве примера рассмотрим DCT-3_n для $n = 2m$. Используя таблицу 6.1 находим соответствующую алгебру $\mathbb{C}[x]/T_n$ с базисом $b = (T_0, \dots, T_{n-1})$. Для получения быстрого алгоритма воспользуемся декомпозицией $T_{2m} = T_m(T_2)$ (лемма 4.2). Следуя теореме 7.3 и ее доказательству, мы выберем базис $c = (T_0, \dots, T_{m-1})$ и $d = (T_0, T_1)$ для $\mathbb{C}[x]/T_m$ и $\mathbb{C}[x]/T_2$, соответственно. Мы получим новый базис

$$b' = (T_0, T_2, \dots, T_{2m-2}, T_1, (T_1 + T_3)/2, \dots, (T_{2m-3} + T_{2m-1})/2).$$

Таким образом, матрица смены базиса имеет вид

$$B = \begin{bmatrix} 1 & & & 0 & & & \\ 0 & 0 & & 1 & \frac{1}{2} & & \\ 0 & 1 & & 0 & 0 & & \\ 0 & 0 & & 0 & \frac{1}{2} & \frac{1}{2} & \\ & \vdots & & & & \vdots & \\ & & & & & & \frac{1}{2} \\ & & 1 & & & & 0 \\ & & 0 & & & & \frac{1}{2} \end{bmatrix},$$

а матрица смены базиса $b \to b'$ задается как B^{-1}. Корнями T_n являются числа $\alpha_k = \cos(k+\frac{1}{2})\pi/n$, причем $\alpha_k = -\alpha_{n-1-k}$, и $T_2(\alpha_k) = T_2(\alpha_{n-1-k})$. Таким образом, перестановка P в теореме 7.3 имеет вид $P = (\mathrm{I}_m \oplus \mathrm{J}_m)$. Далее

$$\mathcal{P}_{d,\overline{\alpha}_i} = \begin{bmatrix} T_0(\alpha_i) & T_1(\alpha_i) \\ T_0(\alpha_{n-1-i}) & T_1(\alpha_{n-1-i}) \end{bmatrix} = \begin{bmatrix} 1 & \alpha_i \\ 1 & -\alpha_i \end{bmatrix} = \mathrm{DFT}_2 \cdot \mathrm{diag}(1, \alpha_i),$$

для $i = 0 \dots m$. Это можно использовать для получения $(\bigoplus_{i=0}^{m-1} \mathcal{P}_{d,\overline{\alpha}_i})^{\mathrm{L}_m^{2m}} = (\mathrm{DFT}_2 \otimes \mathrm{I}_m) \cdot (\mathrm{I}_m \oplus \mathrm{diag}_{i=0}^{m-1}(\alpha_i))$ и

$$\text{DCT-3}_{2m} = P \cdot (\mathrm{DFT}_2 \otimes \mathrm{I}_m) \cdot (\mathrm{I}_m \oplus \mathrm{diag}_{i=0}^{m-1}(\alpha_i)) \cdot (\mathrm{I}_2 \otimes \text{DCT-3}_m) \cdot B^{-1}.$$

Дальнейшее упрощение выражения можно получить, если записать $B = C \cdot (\mathrm{I}_m \oplus \frac{1}{2}\mathrm{I}_m)$ и заметить, что $\mathrm{I}_m \oplus 2\mathrm{I}_m$ перестановочно с $\mathrm{I}_2 \otimes \text{DCT-3}_m$. Для простоты положим $D = \mathrm{diag}_{i=0}^{m-1}(\alpha_i)$ и получим

$$\text{DCT-3}_{2m} = P \cdot (\mathrm{DFT}_2 \otimes \mathrm{I}_m) \cdot (\mathrm{I}_m \oplus 2D) \cdot (\mathrm{I}_2 \otimes \text{DCT-3}_m) \cdot C^{-1}. \tag{8.3}$$

Выражение (8.3) является хорошим примером, чтобы изучить эффект транспонирования и инверсии при выводе новых алгоритмов из уже существующих. Транспонирование (8.3) выполняется непосредственно и приводит к

$$\text{DCT-2}_{2m} = C^{-T} \cdot (\mathrm{I}_2 \otimes \text{DCT-2}_m) \cdot (\mathrm{I}_m \oplus 2D) \cdot (\mathrm{DFT}_2 \otimes \mathrm{I}_m) \cdot P. \tag{8.4}$$

Для инверсии (8.3) нам необходимо знать, что $\text{DCT-3}_n^{-1} = \frac{2}{n} \cdot \mathrm{diag}(\frac{1}{2}, 1, \dots, 1) \cdot \text{DCT-2}_m)$. После упрощений получаем

$$\text{DCT-2}_{2m} = C_1 \cdot (\mathrm{I}_2 \otimes \text{DCT-2}_m) \cdot (\mathrm{I}_m \oplus (2D)^{-1}) \cdot (\mathrm{DFT}_2 \otimes \mathrm{I}_m) \cdot P, \tag{8.5}$$

где C_1 получается из C, если в позиции $(2, m+1)$ число два заменить на единицу. Для реализации умножения на C_1 требуются только операции сложения. Транспонируя (8.5) (или, что эквивалентно, инвертируя (8.4)) получаем снова алгоритм для DCT-3.

$$\text{DCT-3}_{2m} = P \cdot (\mathrm{DFT}_2 \otimes \mathrm{I}_m) \cdot (\mathrm{I}_m \oplus (2D)^{-1}) \cdot (\mathrm{I}_2 \otimes \text{DCT-3}_m) \cdot C_1^T. \tag{8.6}$$

Уравнения (8.5) и (8.6) очень похожи на (8.4) и (8.3), отличаясь только инвертированными элементами в диагональной матрице (средний множитель), что экономит одно умножение на 2-ку в матрице смены базиса (C_1 относительно C). Более впечатляет то, что сложения в C_1 и C_1^T можно выполнять параллельно (т.е. критический путь равен 1), что не верно для C^{-1} и C^{-T}.

Каждое из уравнений (8.3)–(8.6) появлялось в литературе в следующих источниках (8.3) в [24], (8.4) в [24], (8.5) в [57] (транспонированные определения ДТП), (8.6) в [28] и в [56] (транспонированные определения ДТП).

Заметим, что (8.4) также можно получить применяя вначале теорему 8.6, (iii) и затем переводя полученное DCT-4, используя теорему 8.5.

Похожие вычисления для других ДТП из T-группы приводят к следующему результату.

Теорема 8.8 *Пусть $n = 2m$. Все ДТП из T-группы имеют быстрые рекурсивные алгоритмы следующего вида:*

$$\mathrm{DTT}_{2m} = P \cdot (\mathrm{DFT}_2 \otimes \mathrm{I}_m) \cdot (\mathrm{I}_m \oplus D) \cdot (\mathrm{I}_2 \otimes \mathrm{DTT}_m) \cdot B,$$

где P — матрица перестановки, D — диагональная матрица, B — слабозаполненная матрица. Данная факторизация основана на равенстве $T_{2m} = T_m(T_2)$, а конкретные значения P и B можно получить используя теорему 7.3.

Для DST-3 факторизацию можно также найти в [56]. Факторизации для DCT-4 и DST-4 в литературе не появлялись. Они менее эффективны в отношении арифметической сложности.

Замечания. Возможно вывести рекурсивный алгоритм, основанный на $T_{nm} = T_n(T_m)$, используя теорему 7.3. Задачей для больших m является дальнейшая декомпозиция матрицы $\mathcal{P}_{d,\overline{\alpha}_i}$, получаемой в теореме 7.3.

9 Быстрые алгоритмы ДТП, получаемые посредством групповой симметрии

В данном разделе мы выведем быстые алгоритмы для ДТП, которые основываются на «групповой симметрии» (англ. group symmetries) в смысле, который определен ниже. В случаях, когда они возникают, эти симметрии являются прямым следствием свойств ДТП из теоремы 6.2. Мы обозначим два пути, которыми групповые симметрии приходят в действие.

1. *Расширение* (§9.2): Как расширение групповой алгебры $A = \mathbb{C}[x]/p$ связанной с ДТП.

2. *Автоморфизмы* (§9.3): Как подгруппы автоморфизмов группы A. Данные симметрии приводят к алгоритмам, которые существенно отличаются от выведенных в §8.

Для удобства читателя, мы предлагаем краткий обзор по факторизации матриц на основе групповой симметрии. В дальнейшем мы будем придерживаться подхода на основе «представлений» вместо эквавалетной «модульной» точки зрения.

9.1 Факторизация матриц, основанная на групповой симметрии

Факторизация матриц, основанная на групповой симметрии, имеет свои истоки в работах [31, 32] и затем была обобщена в [15, 36, 37, 19] до вида, который приводится здесь. В [18] данная методика была успешно применена к нескольким дискретным преобразованиям, что дало толчок исследованиям, представленным в данной работе. В силу ограниченности пространства мы даем лишь краткий обзор и отсылаем за деталями к цитируемой литературе.

Везде в дальнейшем G есть конечная разрешимая группа. Все представления G (или, что эквивалентно, $\mathbb{C}[G]$) в последующем образуются из *правого* G-модуля. Весь подход основан на следующем понятии симметрии.

Определение 9.1 *Пусть B — произвольная комплексная матрица. Пара представлений (ϕ_1, ϕ_2) группы G называется* симметрией *B, если*

$$\phi_1(g) \cdot B = B \cdot \phi_2(g), \qquad \text{для } g \in G.$$

G называют группой симметрии *B.*

Если B имеет симметрию, мы можем разложить её на множители в соответствии с рис. 9.1, где выбраны матрицы A_1, A_2, которые раскладывают ϕ_1, ϕ_2 в прямую сумму неприводимых представлений ρ_1 и ρ_2. Далее вычислим матрицу

$$D = A_1^{-1} \cdot B \cdot A_2$$

так, чтобы диаграмма стала коммутативной. Получаем разложение на множители

$$B = A_1 \cdot D \cdot A_2^{-1}.$$

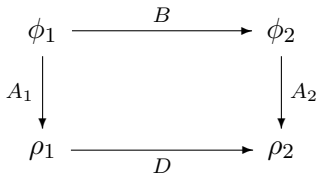

Рис. 9.1. Разложение на множетили (факторизация) матрицы B с симметрией (ϕ_1, ϕ_2)

Таблица 9.1. Типы симметрий, которые могут быть использованы для факторизации B

Симметрия «mon»–«mon»	ϕ_1 и ϕ_2 мономиальные представления
Симметрия «mon»–«irred»	ϕ_1 мономиальное представление, ϕ_2 переставленная прямая сумма неприводимых представлений
Симметрия «irred»–«mon»	ϕ_2 мономиальное представление, ϕ_1 переставленная прямая сумма неприводимых представлений

Матрица D слабозаполнена, поскольку является матрицей сопряжения двух приведенных представлений ρ_1, ρ_2 (следствие леммы Шура [12]). Это означает, что факторизация B пригодна для построения быстрого алгоритма для B только если A_i слабозаполнены или могут сами быть представлены, как произведение слабозаполненных матриц. Это возможно, по крайней мере, в следующих случаях: (1) ϕ_i переставленная прямая сумма неприводимых представлений, т.е. $\phi_i = \rho_i^P$, где P — матрица перестановки. В этом случае мы говорим, что ϕ_i имеет тип «irred». В этом случае $A_i = P^{-1}$. (2) ϕ_i является мономиальным представлением. (Представление является мономиальным, если все его образы являются мономиальными матрицами, т.е. имеют ровно один ненулевой элемент в каждой строке и каждом столбце.) В этом случае мы говорим, что ϕ_i имеет тип «mon». Матрицу разложения A_i можно представить в виде произведения слабозаполненных матриц, используя алгоритм [37]. Коротко говоря, данный алгоритм переводит мономиальное представление в индукцию, которая пошагово раскладывается по композиционным рядам, используя оперделенные рекурсивные формулы, схожие с теми, что представленны в теореме 7.3.

В зависимости от типов ϕ_i мы получаем 3 типа симметрии, показанные в таблице 9.1. Мы не приводим тип «irred»–«irred» поскольку он требует чтобы B уже была слабозаполненной.

Алгоритмы для поиска симметрии [19] и алгоритм [37] пошаговой декомпозиции мономиальных представлений были реализованы как часть системы GAP [22] в виде пакета AREP [17, 16] по конструктивной теории представлений групп. Таким образом, AREP может найти данные факторизации автоматически и его можно использовать как средство для исследования разложимости матрицы в произведение слабозаполненных матриц, т.е. для поиска быстрых алгоритмов.

В оставшейся части раздела мы покажем, что симметрия «mon»–«irred», как и симметрия «mon»–«mon» имеется среди ДТП. Мы также обсудим структуру получающихся быстрых алгоритмов.

9.2 Алгоритмы, получаемые при расширении до групповых алгебр

В данном разделе мы покажем ДТП, обладающие симметрией «mon»–«irred», которую можно использовать для получения быстрых алгоритмов. В [18] показано, что ровно четыре ДТП обладают симметрией «mon»–«irred» с симметрией группы диэдра во всех случаях. Речь идет о ДКП и ДСП 3-го и 4-го типов. Мы собираемся вывести и объяснить данные симметрии. Заметим, что мы будем иметь дело с правыми представлениями (возникающими из правых модулей), чтобы удовлетворить определению симметрии 9.1. Правое представление является транспонированным левым представлением.

Начнем с общего случая полиномиального преобразования. Как обычно, пусть b — это базис $A = \mathbb{C}[x]/p$ и α — список корней полинома p. Далее пусть также f является масштабирующей функцией. Если ϕ — это *правое* представление регулярного модуля A (или, что тоже самое, модуля

$f \cdot A$), тогда, по лемме 3.6

$$\phi \cdot \mathcal{P}^T_{f\cdot b,\alpha} = \mathcal{P}^T_{f\cdot b,\alpha} \cdot \rho,$$

где ρ — прямая сумма одномерных неприводимых представлений A. Если ϕ можно расширить до представления $\overline{\phi}$ групповой алгебры $\mathbb{C}[G]$ конечной группы G, тогда ρ расширяется до переставленной прямой суммы неприводимых представлений $\mathbb{C}[G]$ (при расширении одноразмерные неприводимые представления в ρ — не обязательно соседние — могут объединятся в неприводимые представления в $\mathbb{C}[G]$ больших размерностей). Другими словами, $\mathcal{P}^T_{f\cdot b,\alpha}$ раскладывает $\overline{\phi}$ с точностью до перестановки. Мы получаем следующий результат.

Лемма 9.2 *Будем использовать обозначения, введеные ранее. Если правое регулярное представление ϕ модуля $A = \mathbb{C}[x]/p$ можно расширить до представления $\overline{\phi}$ групповой алгебры $\mathbb{C}[G]$, где G конечна, тогда*

$$\overline{\phi} \cdot \mathcal{P}^T_{f\cdot b,\alpha} = \mathcal{P}^T_{f\cdot b,\alpha} \cdot \overline{\rho},$$

где $\overline{\rho}$ — переставленная прямая сумма неприводимых представлений $\mathbb{C}[G]$. В частности, если $\overline{\phi}$ мономиальное представление, тогда $\mathcal{P}^T_{f\cdot b,\alpha}$ обладает симметрией «mon»–«irred», а $\mathcal{P}_{f\cdot b,\alpha}$ симметрией «irred»–«mon», оба с группой симметрий G.

Применим лемму 9.2 для определения ДТП, которые обладают симметрией «mon»–«irred». Рассмотрим фиксированное ДТП, с которым связано регулярным модулем ϕ. ϕ можно расширить до мономиального представления тогда и только тогда, когда все образы $\phi(g)$, для $g \in A$ можно записать в виде линейной комбинации мономиальных матриц. Поскольку алгебра A циклическая, достаточно рассмотреть образы генератора $\phi(x)$, которые задаются соответствующей матрицей $B(\cdot)$ в 5.1.

Теорема 9.3 *Четыре преобразования* DCT_n *и* DST_n *3-го и 4-го типа, при $n > 0$ единственные ДТП, обладающие симметрией «mon»–«irred» (ϕ, ρ). Обозначим через $D_{2k} = \langle \sigma, \tau | \sigma^2 = \tau^2 = (\sigma\tau)^k = 1 \rangle$ группу диэдра с числом элементов $2k$. Далее, пусть для четных n*

$$\pi_1 = (1,2)(3,4)\dots(n-1,n), \text{ и } \pi_2 = (2,3)(4,5)\dots(n-2,n-1),$$

а для нечетных n,

$$\pi_1 = (1,2)(3,4)\dots(n-2,n-1), \text{ и } \pi_2 = (2,3)(4,5)\dots(n-1,n),$$

(показано в виде перестановки на $\{1,\dots,n\}$). D_{2n} является группой симметрий для $\mathrm{DCT\text{-}3}_n$ *и* $\mathrm{DST\text{-}3}_n$*, а D_{4n} — для* $\mathrm{DCT\text{-}4}_n$ *и* $\mathrm{DST\text{-}4}_n$*. Соответствующие мономиальные представления ϕ задаются для четных n как*

$$\begin{aligned}
\mathrm{DCT\text{-}3}_n &: \sigma \mapsto [\pi_1, n],\ \rho \mapsto [\pi_2, n],\\
\mathrm{DST\text{-}3}_n &: \sigma \mapsto [\pi_1, n],\ \rho \mapsto [\pi_2, (-1,1,\dots,1,-1)],\\
\mathrm{DCT\text{-}4}_n &: \sigma \mapsto [\pi_1, n],\ \rho \mapsto [\pi_2, (1,\dots,1,-1)],\\
\mathrm{DST\text{-}4}_n &: \sigma \mapsto [\pi_1, n],\ \rho \mapsto [\pi_2, (-1,1,\dots,1)],
\end{aligned}$$

а для нечетных n как

$$\begin{aligned}
\mathrm{DCT\text{-}3}_n &: \sigma \mapsto [\pi_1, n],\ \rho \mapsto [\pi_2, n],\\
\mathrm{DST\text{-}3}_n &: \sigma \mapsto [\pi_1, (1,\dots,1,-1)],\ \rho \mapsto [\pi_2, (-1,1,\dots,1)],\\
\mathrm{DCT\text{-}4}_n &: \sigma \mapsto [\pi_1, (1,\dots,1,-1)],\ \rho \mapsto [\pi_2, n],\\
\mathrm{DST\text{-}4}_n &: \sigma \mapsto [\pi_1, n],\ \rho \mapsto [\pi_2, (-1,1,\dots,1)].
\end{aligned}$$

Доказательтсво. Для 16-ти ДТП и связанных с ними представлениями ϕ, рассмотрим матрицы $\phi(x) = B(\beta_1, \beta_2, \beta_3, \beta_4)$ с β_i заданными в таблице 5.2. В силу их структуры, $B(\cdot)$ можно записать как линейную комбинацию мономиальных матриц тогда и только тогда, когда она представима в виде суммы двух мономиальных матриц. Предположим, что $\beta_1 = 0$. Записывая $B(0, \ldots)$ как сумму двух матриц M_1 и M_2, необходимо выполнение условия, что обе матрицы имеют ненулевой элемент в позиции $(1, 2)$. Поскольку элемент $(3, 2)$ матрицы $B(0, \ldots)$ также ненулевой, данная декомпозиция невозможна. Аналогично, декомпозиция при $\beta_4 = 0$. В оставшихся 4-х случаях декомпозиция возможно и приводит к желаемым результатам. Мы приведем один случай в качестве примера. Легко проверить, что

$$B(1, 1, 1, 1) = [\pi_1, n] + [\pi_2, n].$$

Перестановки π_1, π_2 являются инволюциями и, следовательно, порождают группу диэдра D_{2m}. Число m есть порядок произведения $\pi_1\pi_2$, здесь равен n. По теореме 6.2 и таблице 5.2 $B(1, 1, 1, 1)$ диагонализируется DCT-3_n, что доказывает результат. Остальные три случая выводятся аналогично. ∎

Замечание. Теорема 9.3 объясняет симметрию, найденную в [18].

Мы хотим коротко очертить процедуру декомпозиции для DCT-4. За деталями рекомендуем читателю обратиться к работам [18, 19, 37].

Пример 9.4 (DCT-4). *Рассмотрим* DCT-4 *размера* $n = 2^k$. *Согласно теореме 9.3 матрица* $B = \text{DCT-4}_{2^k}$ *имеет симметрию «mon»–«irred»* (ϕ, ρ) *с групповой симметрии диэдра* $D_{2^{k+2}}$. *Мы следуем диаграмме на рис. 9.1. Алгоритм декомпозиции пошагово разложит* ϕ *в композиционный ряд*

$$D_{2^{k+2}} \geq D_{2^{k+1}} \geq \cdots \geq D_{2^2},$$

используя рекурсивную формулу для индукции представлений. Заметим, что последнее представление D_{2^2} *не раскладывается, поскольку оно имеет размерность равную единице. Это приводит к факторизованной матрице разложения* A_1 *для* ϕ_1. *Представление* ρ *— переставленная прямая сумма неприводимых представлений, и поэтому оно может быть разложено матрицей перестановки* A_2. *Корректирующая матрица* D *вычисляется как* $D = A_1^{-1} \cdot B \cdot A_2$, *тогда получаем*

$$\text{DCT-4}_{2^k} = A_1^{-1} \cdot B \cdot A_2.$$

В качестве примера дадим факторизацию DCT-4_8, которая ищется автоматически системой AREP,

$$\begin{aligned} \text{DCT-4}_8 = \quad & [(1,2,8)(3,6,5), (1,-1,1,1,1,-1,1,1)] \cdot \\ & (\mathrm{I}_2 \otimes ((\mathrm{I}_2) \oplus \tfrac{1}{\sqrt{2}} \cdot \mathrm{DFT}_2) \cdot [(3,4),4] \cdot (\mathrm{DFT}_2 \otimes \mathrm{I}_2))) \cdot \\ & [(1,3)(2,4)(5,7)(6,8),8] \cdot (\mathrm{I}_4 \otimes \mathrm{R}_{\frac{15}{8}\pi}) \oplus \mathrm{R}_{\frac{11}{8}\pi} \cdot \\ & (\mathrm{DFT}_2 \otimes \mathrm{I}_4) \cdot [(3,5,7)(4,6,8),8] \cdot \\ & \tfrac{1}{2} \cdot (\mathrm{R}_{\frac{31}{32}\pi}) \oplus \mathrm{R}_{\frac{19}{32}\pi} \oplus \mathrm{R}_{\frac{27}{32}\pi}) \oplus \mathrm{R}_{\frac{23}{32}\pi}) \cdot \\ & [(1,8,5,6,3,2)(4,7),8]. \end{aligned} \tag{9.1}$$

Факторизованная матрица A_1 задается строками 1–4, матрица D — строкой 5, и матрица A_2^{-1} — строкой 6.

Мы видим, что факторизация в (9.1) содержит матрицы поворота

$$R_a = \begin{bmatrix} \cos(a) & \sin(a) \\ -\sin(a) & \cos(a) \end{bmatrix},$$

которых не было в алгоритмах, полученных в §8. Обобщение (для любых $n = 2^k$) для этих алгоритмов можно найти в [7] (скорректированы в [51, 52]). Комбинируя этот алгоритм с теоремой 8.2, (iii) можно получить факторизацию DCT-2 и, следовательно, для DCT-3 на матрицы

поворотов [7]. Полученный алгоритм совпадает с тем, что выведен из симметрии «mon»–«irred» для DCT-3.

Заметим, что алгоритмы, возникающие из симметрии «mon»–«irred», появляются только в неитеративной форме в литературе, т.е. матрица преобразования полностью факторизуется (как в (9.1)) и а не сводятся к преобразованиям меньшего размера. Причина в процедуре декомпозиции (рис. 9.1), поскольку не B, а A_1 раскладывается рекурсивно.

Замечания. Поразительно, что алгоритм для DCT-3_{2^k}, происходящий из симметрии «mon»–«irred», и алгоритм из теоремы 8.8 имеют одинаковую арифметическую сложность [7, 28, 55].

9.3 Алгоритмы, вознкающие из групп автоморфизмов

В §9.2 мы показали как в некоторых случаях симметрия «mon»–«irred» ДТП может быть выведена из его интерпретации как (масштабированного) полиномиального преобразования. В дальнейшем мы покажем, что абсолютно другой тим симметрии «mon»–«mon» также встречается среди ДТП. Данный тип симметрии, если он имеется в ДТП, возникает из группы автоморфизмов алгебры, связанной с этим преобразованием. Все модули в данном разделе являются правыми модулями.

Введем следующие обозначения. Пусть $A = \mathbb{C}[x]/p$. Автоморфизмы A будем обозначать буквами g, h. Мы умножаем автоморфизмы слева направо, т.е. gh означает, что сначала применяется g, а затем h. Это соответствует применению автоморфизмов справа, т.е. если $q \in A$, выражение q^g означает образ элемента q под действием g. Если ϕ есть представление A, и g — автоморфизм, тогда $\phi^g \colon q \mapsto \phi(q^g)$ определяет другое представление A. В соответствии с принятыми обозначениями, $(\phi^g)^h = \phi^{gh}$.

Возможный источник симметрии «mon»–«mon» для полиномиального преобразования $\mathcal{P}_{b,\alpha}$ определяется в следующей теореме.

Теорема 9.5 *Пусть $A = \mathbb{C}[x]/p$ регулярный модуль с базисом b. При этом полином p приводим и имеет корни $\alpha = (\alpha_0, \ldots, \alpha_{n-1})$. Обозначим через ϕ (правое регулярное) представление, отвечающее A и b. Предположим, что A имеет группу автоморфизмов G, такую, что для каждого $g \in G$ существует мономиальная матрица M_g*

$$\phi^g = \phi^{M_g^{-1}}. \tag{9.2}$$

Тогда $\mathcal{P}_{b,\alpha}^T$ обладает симметрией (χ, ψ) с группой симметрий $\overline{G} \cong \langle M_g | g \in G \rangle$. Здесь $G \cong \overline{G}/N$, где $N \trianglelefteq \overline{G}$ обозначает нормальную подгруппу, определяемую как

$$\begin{aligned} g' \in N &\Leftrightarrow \phi(g) \cdot \chi(g') = \chi(g') \cdot \phi(q), \qquad \text{для всех } q \in A \\ &\Leftrightarrow \chi(g') \in \phi(A). \end{aligned}$$

Если D произвольная диагональная обратимая матрица, тогда $(D \cdot \mathcal{P}_{b,\alpha}^T)^T = \mathcal{P}_{b,\alpha}^T \cdot D$ имеет такую же симметрию «mon»–«mon», как и $\mathcal{P}_{b,\alpha}^T$.

Доказательтсво. Заметим, что множество $S = \{M_g \mid g \in G\}$ не является группой, поскольку для каждого g имеется много возможных вариантов для M_g, например, все $a \cdot M_g$, где $a \in \mathbb{C}$. И наоборот, каждая $M_g \in S$ единственным образом определяет автоморфизм A и из $\phi^g = \phi^h$ вытекает, что $g = h$. Теперь изменим ситуацию на противоположную и определим $\gamma \colon S \to G$, $M_g \mapsto g$. Пусть $\overline{G} = \langle S \rangle$ (группа с образующей S). Тогда γ можно расширить до гомоморфизма $\overline{\gamma} \colon \overline{G} \to G$, поскольку $M, M' \in S$, и, используя (9.2),

$$\phi^{\overline{\gamma}(MM')} = \phi^{(MM')^{-1}} = \left(\phi^{M'^{-1}}\right)^{M^{-1}} = \left(\phi^{\overline{\gamma}(M')}\right)^{M^{-1}} = \left(\phi^{M^{-1}}\right)^{\overline{\gamma}(M')} = \phi^{\overline{\gamma}(M)\overline{\gamma}(M')}.$$

По определению, отображение $\overline{\gamma}$ сюрьективно и ядро $\overline{\gamma}$ задается как $N = \{M \mid \phi = \phi^M\}$, и, следовательно, $G \cong \overline{G}/N$. Поскольку $M \in N$ подразумевает, что M коммутативна с каждым $\phi(g)$, $g \in A$, $M \in \phi(A)$. Рассмотрение $\overline{G}$, как мономиального представления χ, само по себе доказывает все утверждение относительно $\overline{G}$.

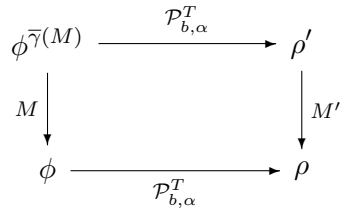

Рис. 9.2. Конструкция «mon»–«mon» симметрии для $\mathcal{P}^T_{b,\alpha}$.

Остается показать, что $\mathcal{P}^T_{b,\alpha}$ обладает симметрией «mon»–«mon» (χ, ψ). Для этого выберем произвольную матрицу $M = \chi(M)$ из $\overline{G}$. Представление ϕ раскладывается матрицей $\mathcal{P}^T_{b,\alpha}$ в прямую сумму неприводимых представлений ρ (см. лемму 3.2). Таким образом, $\phi^{\overline{\gamma}(M)}$ также раскладывается матрицей $\mathcal{P}^T_{b,\alpha}$ в прямую сумму неприводимых представлений ρ'. На рис. 9.2 показана однозначно определяемая матрица M' такая, что

$$M \cdot \mathcal{P}^T_{b,\alpha} = \mathcal{P}^T_{b,\alpha} \cdot M'.$$

Матрица M' является мономиальной, поскольку она переводит ρ' в ρ. Выбор $\psi(M) = M'$ определяет мономиальное представление $\overline{G}$ и показывает, что $\mathcal{P}^T_{b,\alpha}$ имеет симметрию «mon»–«mon» (χ, ψ).

Если D произвольная диагональная обратимая матрица, тогда $\rho^D = \rho$ и $\rho'^D = \rho'$, поскольку все неприводимые слагаемые имеют размерность 1. Поэтому мы можем заменить на рис. 9.2 $\mathcal{P}^T_{b,\alpha}$ на $\mathcal{P}^T_{b,\alpha} \cdot D$, получая такую же симметрию «mon»–«mon». Это завершает доказательство. ■

Замечания. (1) $\mathcal{P}^T_{b,\alpha}$ обладает симметрией «mon»–«mon» (χ, ψ) тогда и только тогда, когда $\mathcal{P}_{b,\alpha}$ обладает симметрией «mon»–«mon» (χ^T, ψ^T). (2) Последнее утверждение в теореме 9.5 показывает, что мы можем применить его к масштабированному полиномиальному преобразованию, и следовательно к ДТП.

В оставшейся части раздела мы используем теорему 9.5 для того, чтобы вывести симметрию «mon»–«mon» для ДТП, чьи транспонированные версии принадлежат T-группе, т.е. те для которых $\mathbb{C}[x]/T_n$ является ассоциированной алгеброй (см. таблицу 6.1) для частного случая $n = 2^m$. Полное исследование всех ДТП произвольной длинны потребовало бы дополнительного места для изложения.

Нам необходима подходящая гуппа автоморфизмов для $A = \mathbb{C}[x]/T_n$.

Лемма 9.6 *Пусть $n = 2^m$ и $A = \mathbb{C}[x]/T_n$. Каждое отображение*

$$g_k \colon T_1 \mapsto T_k, \text{ и } g_{-k} \colon T_1 \mapsto -T_k, \qquad 1 \le k \le n, \ k \text{ нечетное},$$

определяет автоморфизм алгебры A. Множество G_n всех таких $g_{\pm k}$ есть циклическая группа порядка n.

Доказательство. Прежде чем начать доказательство рассмотрим последовательность T_k, $k \ge 0$ в A. Следующие два уравнения позволяют произвести сокращение каждого T_k по модулю T_n.

$$\begin{aligned} 0 \equiv T_n T_{n-k} = \tfrac{1}{2}(T_{2n-k} + T_k) \quad &\Rightarrow T_k \equiv -T_{2n-k}, \quad \text{и} \\ 0 \equiv T_n T_{n+k} = \tfrac{1}{2}(T_{2n+k} + T_k) \quad &\Rightarrow T_k \equiv -T_{2n+k}. \end{aligned} \tag{9.3}$$

Последнее уравнение также показывает, что $T_k \equiv T_{k+4n}$, т.е. последовательность T_k, $k \ge 1$ имеет период $4n$ в A. Используя (9.3) мы можем вычислить сокращенные T_k, $k = 0 \dots 4n-1$, как

$$T_0 \dots T_{n-1} \mid 0 \ -T_{n-1} \dots -T_1 \mid -T_0 \dots -T_{n-1} \mid 0 \ T_{n-1} \dots T_1 \mid, \tag{9.4}$$

где вертикальные линии показывают точки отражения, кратные n.

Теперь приступим к доказательству леммы 9.6. Пусть $n = 2^m$. Мы часто будем использовать то, что T_n четная функция, а также, что T_k, при нечетном k — нечетная функция. Заметим, что $g_{\pm k}$ отображает $T_\ell = T_\ell(T_1) \mapsto T_\ell(\pm T_k)$.

(1) Отображение $g_{\pm k}$ является гомоморфизмом, поскольку $T_n(\pm T_k) = T_n(T_k) = T_k(T_n) = 0$ (лемма 4.2, (i)), т.е. определяющее уравнение $T_n = 0$ в A сохраняется. (2) Отображение $g_{\pm k}$ обратимо; G_n является группой. Пусть g_k задано и k нечетно. Выберем ℓ так, что $k\ell \equiv 1 \mod 4n$. Отображение $T_1 \mapsto T_\ell$ является обратным для g_k, т.к. $T_{k\ell} \equiv T_1$ (см. начало этого доказательства). Также $T_1 \mapsto -T_\ell$ является обратным для g_{-k}. Используя (9.3), мы можем сократить $T_\ell \equiv T_{\ell'}$ или $\equiv -T_{\ell'}$ для соответствующего нечетноего $\ell' < n$. Это показывает, что G_n замкнута относительно инверсии. Также $g_{\pm k}g_{\pm \ell}\colon T_1 \mapsto \pm T_\ell(\pm T_k) = \mp T_{k\ell}$, что можно сократить аналогичным образом. Тем самым подтверждается, что G_n является группой. (3) Группа G_n является циклической. Для $n = 2$, g_{-1} имеет порядок 2. Для $n = 4$, g_3 имеет порядок 4 ($T_3(T_3) = T_9 \equiv -T_1$). Для $n > 4$ покажем, что g_5 имеет порядок n. Рассмотрим (9.4), заметим, что g_5^e это единица тогда и только тогда, когда $5^e \equiv \pm 1 \mod 4n$. Поскольку 5^e никогда не $\equiv -1$ и имеет порядок $n \mod 4n$ ($n = 2^m$), то предположение доказано. ∎

Далее мы используем группу автоморфизмом G_n (лемма 9.6) и теорему 9.5 для получения «mon»–«mon» симметрии для всех ДТП из T-группы.

Теорема 9.7 *Пусть $n = 2^m \geq 4$ и G_n имеют тот же смысл, что и в лемме 9.6. Преобразования* DCT-2$_n$, DST-2$_n$, DCT-4$_n$ *и* DST-4$_n$ *обладают «mon»–«mon» симметрией (χ, ψ) с матрицей, порожденной группой автоморфизмов G_n алгебры $\mathbb{C}[x]/T_n$ (см. теорему 9.5). Все ненулевые элементы данной матрицы имеют значения ± 1. Обозначим через $Z_n = \langle \sigma \mid \sigma^n = 1 \rangle$ циклическую группу порядка n. Группа симметрий для* DCT-2$_n$, DST-2$_n$ *есть Z_n, а для* DCT-4$_n$, DST-4$_n$ *— Z_{2n}. Соответствующее мономиальное представление χ задается как*

$$\begin{aligned}
\text{DCT-2}_n\colon &\quad \sigma \mapsto (T_i \mapsto T_{ki} \mod T_n)^{-1},\\
\text{DST-2}_n\colon &\quad \sigma \mapsto (U_i \mapsto U_{k-1+ki} \mod T_n)^{-1},\\
\text{DCT-4}_n\colon &\quad \sigma \mapsto (V_i \mapsto V_{(k-1)/2+ki} \mod T_n)^{-1},\\
\text{DST-4}_n\colon &\quad \sigma \mapsto (W_i \mapsto W_{(k-1)/2+ki} \mod T_n)^{-1},
\end{aligned} \tag{9.5}$$

(9.6)

где $i = 0 \dots n-1$ и $k = 3$ для $n = 4$, и $k = 5$ для $n \geq 8$.

Доказательство. Пусть $g_k \in G_n$, т. е. $T_1^{g_k} = T_k$. Рассмотрим первый случай DCT-2$_n$ = DCT-3$_n^T$ с ассоциированной алгеброй $A = \mathbb{C}[x]/T_n$ и $b = (T_0, \dots, T_{n-1})$, т. е. DCT-2$_n$ производит декомпозицию правого регулярного представления A, (теорема 6.2). Следуя теореме 9.5, нам необходимо найти мономиальную матрицу смены базиса $M_{g_k}\colon b \to b'$, такую, что T_1 действует на b, как $T_1^{g_k} = T_k$ на b'. Это возможно при $b' = (T_{k\cdot 0}, T_{k\cdot 1}, \dots, T_{k\cdot n-1})$ поскольку, используя лемму 4.1, (ii)

T_1 в b			T_k в b'		
$T_1 \cdot T_0$	$=$	T_1	$T_1 \cdot T_{k\cdot 0}$	$=$	$T_{k\cdot 1}$
$T_1 \cdot T_i$	$=$	$(T_{i-1} + T_{i+1})/2$	$T_1 \cdot T_{ki}$	$=$	$(T_{i-1} + T_{i+1})/2$
$T_1 \cdot T_{n-1}$	$=$	$T_{n-1}/2$	$T_1 \cdot T_{k(n-1)}$	$=$	$T_{k(n-1)}/2$

где $i = 2 \dots n-2$, в последней строке мы использовали то, что $T_n \equiv 0$, и, следовательно, $T_{kn} = T_k(T_n) \equiv 0$, поскольку T_k нечетная функция. Смена базиса $b \to b'$ задается матрицей $M_k\colon T_i \mapsto T_{ki}$, $i = 0 \dots n-1$, и значит

$$\phi^{g_k} = \phi^{M_k}.$$

(Заметим, что рассматриваются правые представления, где ϕ сопряжено с $\phi^{M^{-1}}$ матрицей смены базиса M.) Как и в доказательстве леммы 9.6 мы видим, что M_k является мономиальной, поскольку каждый T_{ki} можно сократить до подходящего $\pm T_\ell \mod T_n$, $0 \leq \ell \leq n-1$. Теорема 9.5 устанавливает симметрию «mon»–«mon» для (χ, ψ). Остается показать, что группа симметрий циклическая и её порядок равен n. Для этого нам необходима последовательность T_ℓ, $\ell \geq 0$, сокращенная по $\mod T_n$, заданная в первой строке таблицы 9.2. Мы видим, что $M_k^e = \mathrm{I}_n$ тогда и только тогда, когда $T_{k^e i} \equiv T_i \mod T_n$ ($i = 0 \dots n-1$) и $k^e \equiv \pm 1 \mod 4n$. Как и в доказательстве леммы 9.6, это показывает, что максимальный порядок группы $e = n$ получается при $k = 3$, если $n = 4$ и $k = 5$ для $n \geq 8$.

Таблица 9.2. Последовательности полиномов $P_0, \ldots, P_{4n-1}$ (один период), приведенные по модулю T_n, для $P \in \{T, U, V, W\}$. Вертикальные линии отделяют группы из n элементов.

DCT-3: $T_0 \ldots T_{n-1}$	$0\ -T_{n-1} \ldots -T_1$	$-T_0 \ldots -T_{n-1}$	$0\ T_{n-1} \ldots T_1$
DST-3: $U_0 \ldots U_{n-1}$	$U_{n-2} \ldots -U_0\ 0$	$-U_0 \ldots -U_{n-1}$	$-U_{n-2} \ldots U_0\ 0$
DCT-4: $V_0 \ldots V_{n-1}$	$-V_{n-1} \ldots -V_0$	$-V_0 \ldots -V_{n-1}$	$V_{n-1} \ldots V_0$
DST-4: $W_0 \ldots W_{n-1}$	$W_{n-1} \ldots W_0$	$-W_0 \ldots -W_{n-1}$	$-W_{n-1} \ldots -W_0$

Доказательство для оставшихся трех случаев проводится аналогично. Базис модуля b заменяется на $b = (P_0, \ldots, P_{n-1})$, где $P = U, V, W$ соответственно.

Надлежащая матрица M_k смены базиса $b \to b'$ соответствует автоморфизму g_k, который задается строками 2–4 выражения (9.5) (без инверсии). Действие T_k в b' можно обосновать, используя лемму 4.1, (ii). Чтобы определить порядок M_k нам необходимо, в каждом случае последовательность P_ℓ сокращенная по модулю $\bmod T_n$ заданная в таблице 9.2 (каждая из этих последовательностей имеет период $4n$). Удивительно, но получается, что для DCT-4 и DST-4 группа симметрий вдвое больше G_n. Рассмотрим пример DCT-4_n. Обозначим через $V_{a_{e,i}}$ образ V_i над M_k^e, $i = 0 \ldots n-1$. Получаем рекурсию со следующим решением:

$$a_{0,i} = i, \quad a_{e,i} = (k-1)/2 + k \cdot a_{e-1,i} \quad \Rightarrow \quad a_{e,i} = (k^e - 1)/2 + ik^e.$$

Используя третью строку таблицы 9.2, получаем:

$$\begin{aligned}
& V_{(k^e-1)/2ik^e} \equiv V_i \mod T_n \quad (i = 0 \ldots n-1) \\
\Leftrightarrow\ & (k^e-1)/2 + ik^e \equiv i \text{ или } -i-1 \mod 4n \quad (i = 0 \ldots n-1) \\
\Leftrightarrow\ & k^e(2i+1) \equiv \pm(2i+1) \mod 8n \quad (i = 0 \ldots n-1) \\
\Leftrightarrow\ & k^e \equiv \pm 1 \mod 8n,
\end{aligned}$$

что показывает, что максимальный порядок $e = 2n$ получается при $k = 3$, если $n = 4$, и $k = 5$ для $n \geq 8$. ∎

Завершим данный раздел небольшим примером.

Пример 9.8 (DCT-4_4) *По теореме 9.7* DCT-4_4 *обладает «mon»–«mon» симметрией* (χ, ψ) *с циклической группой* $Z_8 = \langle \sigma \rangle$. *Образ* $\chi(\sigma)$ *определяется инверсией* $V_i \mapsto V_{1+3i} \mod T_4$, $i = 0 \ldots 3$. *Используя таблицу 9.2, получаем* $V_4 \equiv -V_3$, $V_4 \equiv -V_0$, $V_{10} \equiv -V_2$, *и, следовательно,*

$$\chi(\sigma) = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \quad \psi(\sigma) = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}.$$

Матрица $\psi(\sigma)$ *вычислена с использованием AREP. Обе матрицы имеют порядок, равный 8.*

«Mon»–«mon» симметрия DCT-4_4, показанная в примере 9.8, использовалась в [21] для вывода быстрого алгоритма (симметрия определялась другим образом), с применением теоремы 8.6, (iii) для получения быстрого алгоритма для DCT-2_8. Вывод производился согласно диаграмме на рисунке 9.1, но мономиальные представления ϕ_1, ϕ_2 раскладывались над $\mathbb{Q}$. Это перенесло все нерациональные операции в корректирующую матрицу D.

Замечание. Используя AREP мы установили, что (до определенного размера) все 16 типов ДТП обладают «mon»–«mon» симметрией для любого n.

10 Другие быстрые алгоритмы

Алгебраические методы, представленные в §8–§9, объясняют большинство извесных из литературы алгоритмов. Однако, один класс алгоритмов пока не может быть объяснен указанными

методами. Мы коротко рассмотрим эти алгоритмы, чтобы сделать обзор алгоритмов ДТП всесторонним.

Вкратце, существует возможность вычислить ДТП, встраивая его матрицу в матрицу преобразования большего размера, которую можно эффективно вычислить. В качестве примера рассмотрим первый алгоритм, предложенный для DCT-2_n $= [\cos k(\ell + \frac{1}{2})\pi/n]$ в [1]. Если мы определим DFT как

$$\mathrm{DFT}_n = [e^{2\pi ik\ell/n}]_{k,\ell=0\ldots n-1},$$

мы можем легко вывести, что

$$\mathrm{Re}\left(\mathrm{diag}_{k=0}^{2n-1}(e^{\pi ik/2n}) \cdot \mathrm{DFT}_{2n}\right) = [\cos k(\ell + \frac{1}{2})\pi/n]_{k,\ell=0\ldots 2n-1},$$

где $\mathrm{Re}(M)$ означает действительную часть матрицы M. Это показывает, что DCT-2_n можно вычислить при дополнении входного вектора x длины n нулями до длины $2n$. Тогда первые n значений выходного вектора являются результатом.

Подобные конструкции допустимы при вычислении любого ДТП посредством ДПФ соответствующего размера. Это показывает, что арифметическая сложность каждого ДТП равна $O(n \log n)$, вне зависимости от размера n. В частности, это верно и для ДТП типов 5–8, для которых нет других алгоритмов в литературе.

Встраивание в другие преобразования также возможны. Например, теорема 8.7 позволяет включить ДТП типов 5–8 в ДТП 1-го и 2-го типа.

11 Резюме

Дана полная характеристика всех 16-и типов ДТП, как масштабированных полиномиальных преобразований, соответствующих надлежащему A-модулю M с базисом b, где $A = \mathbb{C}[x]/p(x)$, $M = f \cdot A$ с масштабирующей функцией f и последовательностью полиномов Чебышева b (теорема 6.2). Каждое ДТП единственным образом определяется своими алгебраическими свойствами.

Далее, используя алгебраические характеристики, выведено большинство известных из литературы быстрых ДТП и выявлены математические принципы, на которых они основаны. В частности, были получены:

1. *Алгоритмы, получаемые при непосредственной манипуляции/декомпозиции модуля M* (§8): (1) Перевод одного ДТП в другое, с использованием дуальности (теорема 8.2); (2) Перевод ДТП с использованием смены базиса (теорема 8.5); (3) Декомпозиция путем разложения полинома на множители (теоремы 8.6 и 8.7); (4) Декомпозиция с использованием декомпозиции полиномов (теорема 8.8).
2. *Алгоритмы, основанные на групповой симметрии* (§9): (1) Декопозиция с использованием симметрии «mon»–«irred» (теорема 9.3); (2) Декопозиция с использованием симметрии «mon»–«mon» (теорема 9.5)
3. *Алгоритмы со встраиванием* (§10).

Наши результаты показывают, что связь между цифровой обработкой сигналов и теорией представления алгебр уходит за пределы ДПФ. Вопрос, который остается, касается того, каковы пределы этой связи и как можно расширить её, чтобы включить другие преобразования и их алгоритмы и как эту связь можно использовать для приложений обработки сигналов. Мы хотели бы завершить статью следующим вопросом: *До какой степени обработка сигналов алгебраична?*

A Ортонормированные ДКП и ДСП

Таблица A.1 содержит ортонормированные версии всех 16-и ДТП.

Список литературы

[1] N. Ahmed, T. Natarajan, and K. R. Rao, *Discrete Cosine Transform*, IEEE Trans. on Computers, C-23(1974), pp.90–93.

Таблица A.1. Определение ортонормированных версии ДКП и ДСП; $a_{k,\ell}$ это элемент в k-ой строке ℓ-того столбца соответствующего немасштабированного ДТП заданного в таблице 5.1. Все матрицы имеют размер $(n \times n)$ с индексами по строкам $k = 0 \dots n-1$ и индексами по столбцам $\ell = 0 \dots n-1$. Масштабирующие множители для строк/столбов задаются как: $c_i = 1/\sqrt{2}$ для $i = 0$ и $= 1$ иначе. $d_i = 1/\sqrt{2}$ для $i = n-1$ и $= 1$ иначе.

	ДКП	ДСП
Тип 1	$\sqrt{\frac{2}{n-1}} \cdot c_k c_\ell d_k d_\ell \cdot a_{k,\ell}$	$\sqrt{\frac{2}{n-1}} \cdot a_{k,\ell}$
Тип 2	$\sqrt{\frac{2}{n}} \cdot c_k \cdot a_{k,\ell}$	$\sqrt{\frac{2}{n}} \cdot c_k \cdot a_{k,\ell}$
Тип 3	$\sqrt{\frac{2}{n}} \cdot c_\ell \cdot a_{k,\ell}$	$\sqrt{\frac{2}{n}} \cdot c_\ell \cdot a_{k,\ell}$
Тип 4	$\sqrt{\frac{2}{n}} \cdot a_{k,\ell}$	$\sqrt{\frac{2}{n}} \cdot a_{k,\ell}$
Тип 5	$\sqrt{\frac{2}{n-1/2}} \cdot c_k c_\ell \cdot a_{k,\ell}$	$\sqrt{\frac{2}{n-1/2}} \cdot a_{k,\ell}$
Тип 6	$\sqrt{\frac{2}{n-1/2}} \cdot c_k d_\ell \cdot a_{k,\ell}$	$\sqrt{\frac{2}{n-1/2}} \cdot a_{k,\ell}$
Тип 7	$\sqrt{\frac{2}{n-1/2}} \cdot d_k c_\ell \cdot a_{k,\ell}$	$\sqrt{\frac{2}{n-1/2}} \cdot a_{k,\ell}$
Тип 8	$\sqrt{\frac{2}{n-1/2}} \cdot a_{k,\ell}$	$\sqrt{\frac{2}{n-1/2}} \cdot d_k d_\ell \cdot a_{k,\ell}$

[2] L. Auslander, E. Feig, and S. Winograd, *Abelian Semi-simple Algebras and Algorithms for the Discrete Fourier Transform*, Advances in Applied Mathematics, 5 (1984), pp. 31–55.

[3] T. Beth, *Verfahren der Schnellen Fourier transformations*, Teubner, 1984.

[4] T. Beth, *On the computational complexity of the general discrete Fourier transform*, Theoretical Computer Science, 51 (1987), pp.331–339.

[5] P. Bürgisser, M. Clausen, and M. A. Shokrollahi, *Algebraic Complexity Theory,* Springer, 1997.

[6] S. Chan and K. Ho, *Direct Methods for computing discrete sinusoidal transforms,* IEE Proceedings, 137 (1990), pp. 433–442.

[7] W.-H. Chen, C. Smith, and S. Fraclick, *A fast computational Algorithm for the Discrete Cosine Transform,* IEEE Trans. on Communications, COM-25 (1977), pp. 1004–1009.

[8] T. S. Chihara, *An Introduction to Orthogonal Polynomials*, Gordon and Breach, 1978.

[9] M. Clausen, *Beiträge zum Entwurf schneller Spektraltransformationen (Habilitationsschrift),* Univ. Karlsruhe, 1988.

[10] M. Clausen and U. Baum, *Fast Fourier Transforms*, BI-Wiss.-Verl., 1993.

[11] J. W. Cooley and J. W. Tukey, *An Algorithm for the Machine Calculation of Complex Fourier Series*, Math. of Computation, 19(1965), pp. 297–301.

[12] W. C. Curtis and I. Reiner, *Representation Theory of Finite Groups*, Interscience, 1962.

[13] P. Diaconis and D. Rockmore, *Efficient computation of the Fourier transform on finite groups,* Amer. Math. Soc., 3(2) (1990), pp. 297–301.

[14] P. Driscoll, M. Healy Jr., and D. N. Rockmore, *Fast Discrete Polynomial Transforms with Application to Data Analysis for Distance Transitive Graphs,* SIAM Journal Computation, 26 (1997), pp. 1066–1099.

[15] S. Egner, *Zur Algorithmischen Zerlegungstheorie Linearer Transformationen mit Symmetrie,* PhD thesis, Universität Karlsruhe, Informatik, 1997.

[16] S. Egner, J. Johson, D. Padua, M. P üschel, and J. Xiong, *Automatic Derivation and Implementation and Implementation of Signal Processing Algorithms*, ACM SIGSAM Bulletin Communications in Computer Algebra, 35 (2001), pp. 1–19.

[17] S. Egner and M. Püshel, *AREP—Constructive Representation Theory and Fast Signal Transforms*, GAP share package, 1998.

[18] S. Egner and M. Püshel, *Automatic generation of fast discrete signal transforms*, IEEE Trans. on Signal Processing, 49 (2001), pp. 1992–2002.

[19] S. Egner and M. Püshel, *Symmetry-Based Matrix Factorization,* Journal of Symbolic Computation, (2002). To appear.

[20] E. Feig, *A fast scaled-DCT algorithm,* in SPIE Processings, vol. 1244, 1990, pp. 2–13.

[21] E. Feig and S. Winograd, *Fast Algorithms for the Discrete Cosine Transforms,* IEEE Trans. on Signal Processing, 40 (1992), pp. 2174–2193.

[22] The GAP Team, *GAP—Groups, Algorithms, and Programming*, University of St. Andrews, Scotland, 1997, `http://www-gap.dcs.st-and.ac.uk/~gap/.`

[23] M. Heideman, D. Johnson, and C. Burrus, *Gauss and History of the Fast Fourier Transform*, Archive for History of Exact Sciences, 34 (1985), pp. 265–277.

[24] H. Hou *A fast recursive algorithm for computing the discrete cosine transform*, IEEE Trans. on Acoustic, Speech, and Signal Processing, ASSP-35 (1987), pp. 1455–1461.

[25] N. Jacobson, *Basic Algebra II*, W. H. Freeman and Co., 1980.

[26] T. Kailath and V. Olshevsky, *Displacement structure approach to discrete trigonomentic transform based preconditioners of G. Strang and T. Chan type,* Calcolo, 30 (1996), pp. 191–208.

[27] H. Kitajima, *A Symmetric cosine transform*, IEEE Trans. on Computers, C-29 (1980), pp. 317–323.

[28] B. Lee, *A New Algorithm to Compute the Discrete Cosine Transform,* IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-32 (1984), pp. 1243–1245.

[29] D. Maslen and D. Rockmore, *Generalized FFTs – A survey of some recent results,* in Processings of IMACS Workshop in Groups and Computation, vol. 28, 1995, pp. 182–238.

[30] J. C. Mason *Chebyshev polynomials of the second, third and fourth kind in approximation, indefinite integration, and integral transforms* Journal of Computational and Applied Mathematics, 49 (1993), pp. 169–178.

[31] T. Minkwitz, *Algorithmensynthese für lineare Systeme mit Symmetrie*, PhD thesis, Universität Karlsruhe, Informatik, 1993.

[32] T. Minkwitz, *Algorithms Explained by Symmetry*, Lecture Notes on Computer Science, 900 (1995), pp. 169–178.

[33] J. M. .F. Moura and M. .G. S. Bruno, *DCT/DST and Gauss-Markov Fields: Conditions for Equivalence*, IEEE Trans. on Signal Processing, 46 (1998), pp. 2571–2574.

[34] D. Potts, G. Steidl, *Optimal trigonometric preconditioners for nonsymmetric Toeplitz system,* Linear Algebra Application, 281 (1998), pp. 265–292.

[35] D. Potts, G. Steidl, and M. Tasche, *Fast Algorithms for Discrete Polynomial Transforms,* Mathmatics and Computation, 67 (1998), pp. 1577–1590.

[36] M. Püschel, *Konstruktive Darstellungstheorie und Algorithmengenerierung,* PhD thesis, Universität Karlsruhe, Informatik, 1998. Also available in English as Tech. Rep. Drexel-MCS-1999-1, Drexel University, Philadelphia.

[37] M. Püschel, *Decomposing Monomial Representations of Solvable Groups,* Journal of Symbolic Computation, 34 (2002), pp. 561–596.

[38] M. Püschel and J. M. .F. Moura, *The Discrete Trigonometric Transforms and Their Fast Algorithms: An Algebraic Symmetry Approach,* in Proc. 10th IEEE DSP Workshop, 2002.

[39] C. M. Rader, *Discrete Fourier Transforms When the Number of Data Samples is Prime,* Proceedings of the IEEE, 56 (1968), pp. 1107–1108.

[40] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Application,* Academic Press, 1990.

[41] M. O. Rayes, V. Trevisan, and P. S. Wang, *Factorization of Chebyshev Polynomials*, Tech. Report ICM-199802-0001, Kent State University, 1998.

[42] T. J. Rivlin, *The Chebyshev Polynomials*, Wiley Interscience, 1974.

[43] D. Rockmore, *Efficient computation of Fourier invertion for finite groups,* Assoc. Comp. Mach., 41(1) (1994), pp.31–66.

[44] D. Rockmore, *Some applications of generalized FFTs*, in Processing of DIMAS Workshop in Groups and Computation, vol. 28, 1995, pp. 329–370.

[45] V. Sánchez, P. Garcá, A. M. Peinado, J. C. Segura, and A. J. Rubio, *Diagonalizing Properties of the Discrete Cosine Transform*, IEEE Trans. on Signal Processing, 43 (1995), pp. 2631–2641.

[46] G. Steidl and M. Tasche, *A Polynomial Approach to Fast Algorithms for Discrete Fourier-Cosine and Fourier-Sine Transforms,* Mathematics of Computation, 56 (1991), pp. 281–296.

[47] G. Strang, *The Discrete Cosine Transform*, SIAM Review, 41 (1999), pp 135–147.

[48] G. Szegö, *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. Publ., 3rd ed., 1967.

[49] R. Tolimieri, M. An and C. Lu, *Algorithms for Discrete Fourier Transforms and Convolution*, Spronger, 2nd ed., 1997.

[50] M. Vetterli and H. Nussbaumer, *Simple FFT and DCT Algorithms with reduced Number of Operations,* Signal Processing, 6 (1984), pp. 267–278.

[51] Z. Wang, *Reconsideration of "A Fast Computational Algorithm for the Discrete Cosine Transform"*, IEEE Tran. on Communications, COM-31 (1983), pp. 121–123.

[52] Z. Wang, *Fast Algorithms for the Discrete W Transform and for the Discrete Fourier Transform,* IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-32 (1984), pp. 803–816.

[53] Z. Wang and B. Hunt, *The Discrete W Transform,* Applied Mathematics and Computation, 16 (1985), pp. 19–48.

[54] S. Winograd, *Arithmetic Complexity of Computation*, Siam, 1980.

[55] P. Yip and K. Rao, *A Fast Computational Algorithms for a Family of Discrete Sine and Cosine Transforms,* Circuits, Systems, and Signal Processing, 3 (1984), pp. 304–307.

[56] P. Yip and K. Rao, *Fast Decimation-In-Time Algorithms for a Family of Discrete Sine and Cosine Transforms,* Circuits, Systems, and Signal Processing, 3 (1984), pp. 387–408.

[57] P. Yip and K. Rao, *The Decimation-In-Frequency Algorithms for a Family of Discrete Sine and Cosine Transforms,* Circuits, Systems, and Signal Processing, 3 (1988), pp. 3–19.