Тема проекта

Разработка метода “экстра-компонент” для быстрого вычисления дискретных преобразований

Состав коллектива: Терехов Андрей Валерьевич, ИВМиМГ СО РАН, д.ф.-м.н.

Научное содержание работы:

При расчёте интегральных и дискретных преобразований необходимо иметь в своем распоряжении быстрые алгоритмы умножения векторов на матрицы, элементы которых задаются как значения специальных функций (Чебышева, Лежандра, Лагерра, сферические и т.д.). На данный момент существующие быстрые алгоритмы на порядки уступают в экономичности процедуре быстрого преобразования Фурье. С целью сокращения этого разрыва в проекте поставлена задача разработать высокоэффективный общий подход для вычисления матрично-векторных произведений. В итоге была предложена серия новых быстрых методов и алгоритмов, отличающихся простотой структуры и допускающих эффективную программную реализацию для современных микропроцессоров.

Современное состояние проблемы (на момент начала работы).

Одним из факторов, сделавших дискретное преобразование Фурье крайне востребованным методом численного анализа, является существование быстрого алгоритма дискретного преобразования Фурье (FFT), который позволяет сократить вычислительные затраты с $O(N^2)$ до $O(N\log N)$ в задачах расчёта матрично-векторного произведений вида

$$y = \mathcal{F}x, \quad \mathcal{F} \in \mathbb{C}^{N\times N}, \mathrm{x}, \mathrm{y} \in$$

где $\mathcal{F}$ – матрица Фурье преобразования, такая что $\mathcal{F}\mathcal{F}^* = E$, а E – единичная матрица. Это привело к настоящему прорыву в разработке как методов математического моделирования, так и методов цифровой обработки и анализа сигналов. Закономерно возникает вопрос, можно ли создать экономичные процедуры не только для тригонометрического базиса, но и для классических ортогональных многочленов: Чебышева, Лежандра, Гегенбауэра, Якоби, Лагерра и Эрмита, а также для других систем функций, таких как сферические гармоники. Рассмотрим некоторые существующие подходы для конструирования таких быстрых алгоритмов.

Из всех классических ортогональных многочленов наиболее простой является задача построения экономичного метода для вычисления дискретного преобразования Чебышева, которое может быть вычислено посредством быстрого дискретного косинус преобразования (DCT), реализованного на основе алгоритма FFT. В свою очередь, на основе быстрого преобразования Чебышева можно построить экономичный алгоритм для многочленов Лежандра, а в более общем случае для многочленов Гегенбауера и Якоби. Различные подходы для пересчёта коэффициентов разложения при замене базиса предложены в многочисленных публикациях. На основе принципа "разделяй и властвуй" был предложен новый быстрый алгоритм для вычисления преобразования Лагерра.

Решена проблема реализации быстрого преобразования Эрмита, где предложен быстрый алгоритм, имеющий вычислительную сложность $O(N\log^2 N)$ операций. Математическое моделирование климата, а также решение задач: прогноза погоды, геофизики, астрономии, а также дифференциальных уравнений в частных производных – это неполный перечень приложения сферических функций, роль которых для сферической системы координат аналогична роли тригонометрического базиса ряда Фурье для декартовой системы координат. Для этого случая, как и для классических ортогональных многочленов, были разработаны быстрые алгоритмы, которые позволяют сократить общий объём вычислений с $O(N^4)$ до $O(N^2\log N)$. Более общие алгоритмы основаны идее предварительного сжатия исходной матрицы, например, посредством вейвлет преобразования, локального косинус преобразования или основе "butterfly" алгоритм. Как правило, предварительное сжатие матрицы требует значительных вычислений, однако в последующем это позволяет сократить общее время счёта при многократном вычислении матрично-векторного произведения.

Подробное описание работы, включая используемые алгоритмы.

Рассмотрим многочлены Чебышева $T_n\colon \mathbb{R} \to \mathbb{R}$ первого рода, заданные на сегменте $\Omega = [-1,1]$

$T_n(x) = \arccos\big(n\cos(nx)\big), x \in \Omega,$ (1)

которые образуют ортонормированный базис в $L_{2,\omega}(\Omega)$

$$\int_{-1}^{1} T_n(x)T_m(x)\omega(x)dx = \frac{c_n\pi}{2}\delta_{mn},$$

где , $c_0 = 2$, и $c_n = 1$ для $n \geq 1$, and δ_{mn} – Кронекеровская дельта функция.

Пусть задана функция $f(x)\colon \Omega \to \mathbb{R}$ и существует интеграл

$$\| f \|^2_{L_{2,\omega}(\Omega)} = \int_{-1}^{1} \omega(x)|f(x)|^2 dx,$$

тогда справедливо представление вида

$$f(x) = \sum_{k=0}^{\infty} \tilde{f}_k T_k(x), \;\; x \in \Omega, \tag{2}$$

$$\tilde{f}_k = \frac{p_k}{\pi}\int_{-1}^{1} f(x)T_k(x)\frac{dx}{\sqrt{1-x^2}} = \frac{p_k}{\pi}\int_{0}^{\pi} \cos(k\theta)\, f(\theta)d\theta, \tag{3}$$

где $p_0 = 1$, и $p_k = 2$ для $k > 0$.

Считая, что значения $\tilde{f}_k$ известны, рассмотрим проблему вычисления частичной суммы для (2), которую запишем в виде матрично-векторного произведения $\mathbf{F} = A\tilde{\mathbf{F}}$, где

$A \coloneqq \big(\cos\big(m\arccos(x_n)\big)\big)_{n,m=0}^{N,M} \in \mathbb{R}^{(N+1)\times(M+1)},$

(4)

$$\mathbf{F} = \big(f(x_n)\big)_{n=0}^{N},$$

$\big(\tilde{f}_m\big)_{m=0}^{M}$.

Для Чебышевского набора узлов и $M = N$ умножение вектора на матрицу (4) может быть вычислено за $O(N\log N)$ арифметических действий посредством дискретного косинус преобразования (DCT). В общем случае для произвольного распределения узлов $x_k \in \Omega$ вычислительные затраты составят $O(NM)$ операций, поэтому будет предложен новый алгоритм, позволяющий вычислить $A\tilde{\mathbf{F}}$ и $A^{\mathrm{T}}\mathbf{F}$ за $O\big(N\log N + M\rho(\varepsilon)\big)$ и $O\big(M\log M + N\rho(\varepsilon)\big)$ арифметических действий, соответственно, где $\rho(\varepsilon) \leq 25$ – величина зависящая от требуемой точности.

Рассмотрим первую строку матрицы A

$$A_{1\cdot} = [1, \cos(\theta_0), \cos(2\theta_0), \dots, \cos(M\theta_0)], \tag{5}$$

где $\theta_n = \arccos(x_n)$. Пример последовательности $A_{1\cdot}$ приведен на рис. 1a, а модули соответствующих Фурье-компонент $\mathcal{F}A_{1\cdot}^{\mathrm{T}}$ представлены на рис. 1b (см. кривая "Kaiser 0"). Резкое изменение значений функции на границах приводит к тому, что абсолютные значения всех коэффициентов ряда Фурье становятся отличными от нуля, т.е. спектр не является локализованным или хотя бы конечным.

Для исключения этого нежелательного эффекта можно использовать, например, весовую функцию Кайзера

(6)

Здесь I_0 – это одифицированная функция Бесселя, $\zeta \geq$
0 определяется форму функции Бесселя. На рис. 2 изображена весовая функция Кайзера во временной и спектральной областей для различных значений параметра ζ.

Теперь рассмотрим для строки (5) преобразование $\ddot{A}_{1\cdot} = \mathcal{F}W_M A_{1\cdot}^{\mathrm{T}}$, где диагональная матрица W_M определена как

$$W_M = \mathrm{diag}\left\{w_0^{\zeta,M}, w_1^{\zeta,M}, \dots, w_M^{\zeta,M}\right\} \in \mathbb{R}^{M+1,M+1}.$$

Абсолютные значения для компонент спектра в логарифмическом масштабе представлены на рис. 1b. Видно, что с ростом параметра ζ спектр для (5) в значительной степени локализован, так что для некоторого $\varepsilon > 0$ элементы матрицы, удовлетворяющие условию $|\tilde{a}_{1j}| < \varepsilon \max_j |\tilde{a}_{1j}|$, можно считать нулевыми. Применив аналогичное преобразование для всех строк матрицы A, получаем сжатую матрицу:

$$\ddot{A} = AW_M\mathcal{F}^*, \tag{7}$$

для которой значительное число элементов можно игнорировать в силу их относительной малости (рис. 3).

Принимая во внимание свойство ортогональности $\mathcal{F}\mathcal{F}^* = \mathcal{F}^*\mathcal{F} = E$, запишем произведение $\mathbf{F} = A\hat{\mathbf{F}}$ как

$$\mathbf{F} = A\underbrace{W_M\mathcal{F}^*\mathcal{F}W_M^{-1}}_{E}\hat{\mathbf{F}} = \ddot{A}\mathcal{F}W_M^{-1}\hat{\mathbf{F}}, \qquad (8)$$

а произведение $\hat{\mathbf{F}} = A^{\mathrm{T}}\mathbf{F}$ как

$$\hat{\mathbf{F}} = \underbrace{W_N^{-1}\mathcal{F}^*\mathcal{F}W_N}_{E}A^{\mathrm{T}}\mathbf{F} = W_N^{-1}\mathcal{F}\ddot{A}^{\mathrm{T}}\mathbf{F}. \qquad (9)$$

Таким образом, предварительно вычислив матрицы $\ddot{A}$ или $\ddot{A}^{\mathrm{T}}$, обладающие компактным портретом, можно экономично рассчитать искомые матрично-векторные произведения. Для каждого вектора потребуется выполнить однократное Фурье преобразование и умножение на сжатую матрицу $\ddot{A}$ или $\ddot{A}^{\mathrm{T}}$. Умножение на диагональную матрицу W_M увеличивает общий объём вычислений незначительно.

Метод экстра-компонент отличается от существующих простой программной реализацией и высокой эффективностью. На подготовительном этапе, выполняемом однократно, исходная матрица посредством Фурье преобразования приводится к матрице с разреженным портретом. Далее на вычислительном этапе матрично-векторное произведение может быть вычислено достаточно быстро в силу игнорирования относительно близких к нулю элементов сжатой матрицы. Один из ключевых факторов, влияющих на производительность – это векторизация вычислений, которая не всегда может быть эффективно использована в рамках существующих алгоритмов. Для современных микропроцессоров векторные инструкции требуют, чтобы данные в оперативной памяти были непрерывны и одновременно выровнены по определённым значениям адресов. Структура представленных в проекте алгоритмов позволяет удовлетворить этим ограничениям для векторных вычислений, тем самым гарантируется высокий уровень производительности для широкого диапазона размерности преобразований. Как следствие, если в теоретическом плане оценки числа операций метода экстра-компонент не лучше оценок для существующих подходов, то реальная производительность может быть даже выше, чем производительность высокооптимизированных библиотечных процедур, что продемонстрировано на примере классического преобразования Фурье (Рис. 4).

Полученные результаты.

Результаты, полученные в ходе реализации проекта, соответствуют высокому мировому уровню. В проекте предложен новый эмпирический подход для построения быстрых алгоритмов вычисления матрично-векторных произведений, связанных с интегральными и дискретными преобразованиями. Изначально цель данной работы состояла в разработке быстрого алгоритма вычисления прямого и обратного преобразований Лагерра, однако, предлагаемый метод позволил на несколько порядков сократить время счёта и для других преобразований, таких как Чебышева, Лежандра, Гегенбауера, Якоби и даже для классического преобразования Фурье. Для последнего предложен новый способ, позволяющий сократить время вычислений для неоптимальных размерностей. Идея состоит в том, чтобы расширить исходную матрицу до некоторой оптимальной размерности с точки зрения производительности, тем самым сократив почти на порядок итоговое время счёта. Также предварительные расчёты показали, что метод экстра-компонент является весьма перспективным для решения довольно сложной проблемы о разложении функции в ряд по сферическим функциям. Для решения дифференциальных уравнений на современных суперкомпьютерах многообещающей выглядит комбинация рассмотренных в статье быстрых методов и параллельного алгоритма дихотомии, который можно трактовать как быстрый метод умножения матрицы обратной к блочно-трёхдиагональной матрице.

Важнейшие результаты, полученные при реализации Проекта

Разработан новые класс методов для вычисления матрично-векторных произведений в случае, когда элементы матрицы являются значениями специальных функций.

Иллюстрации, визуализация результатов.

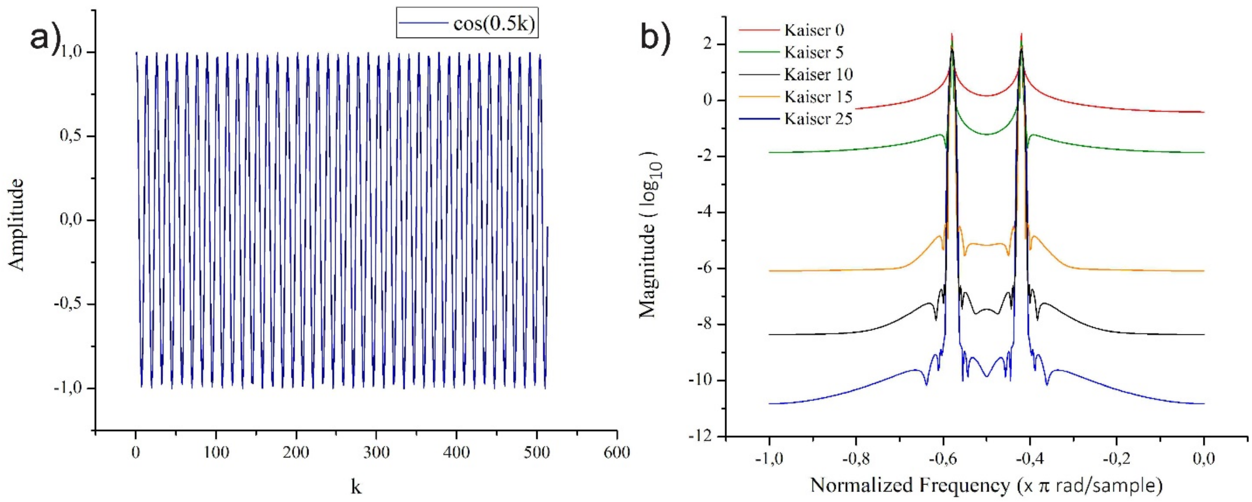

Рисунок 1: a) График строки $A_{1\cdot}$, где отдельные точки соединены линиями; b) абсолютные значения коэффициентов соответствующего ряда Фурье для различных весовых функций

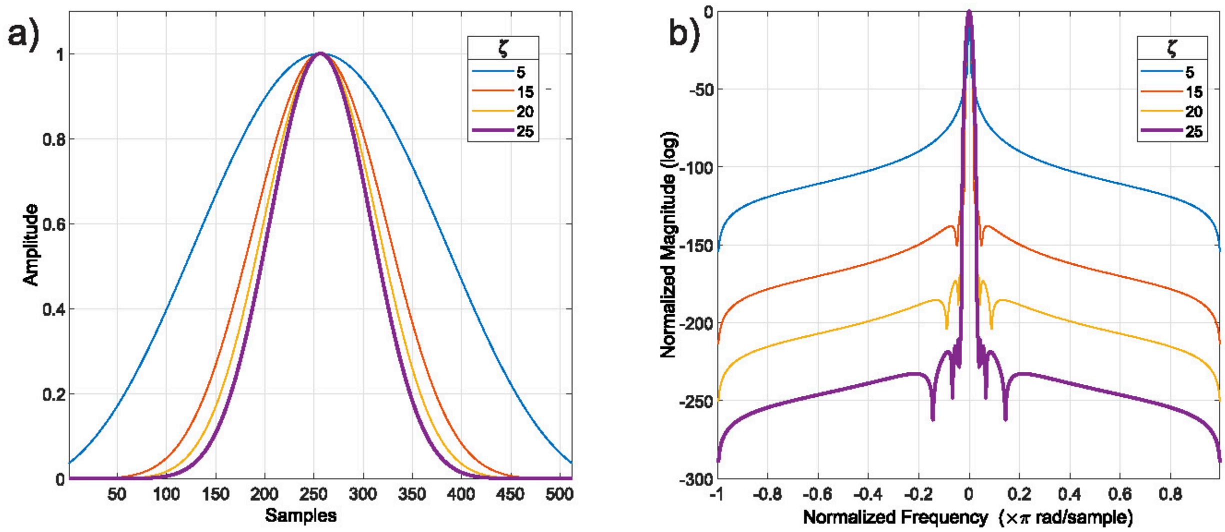

Рис. 2: Весовая функция Кайзера для различных значений параметра ζ, a) временная область и b) частотная область

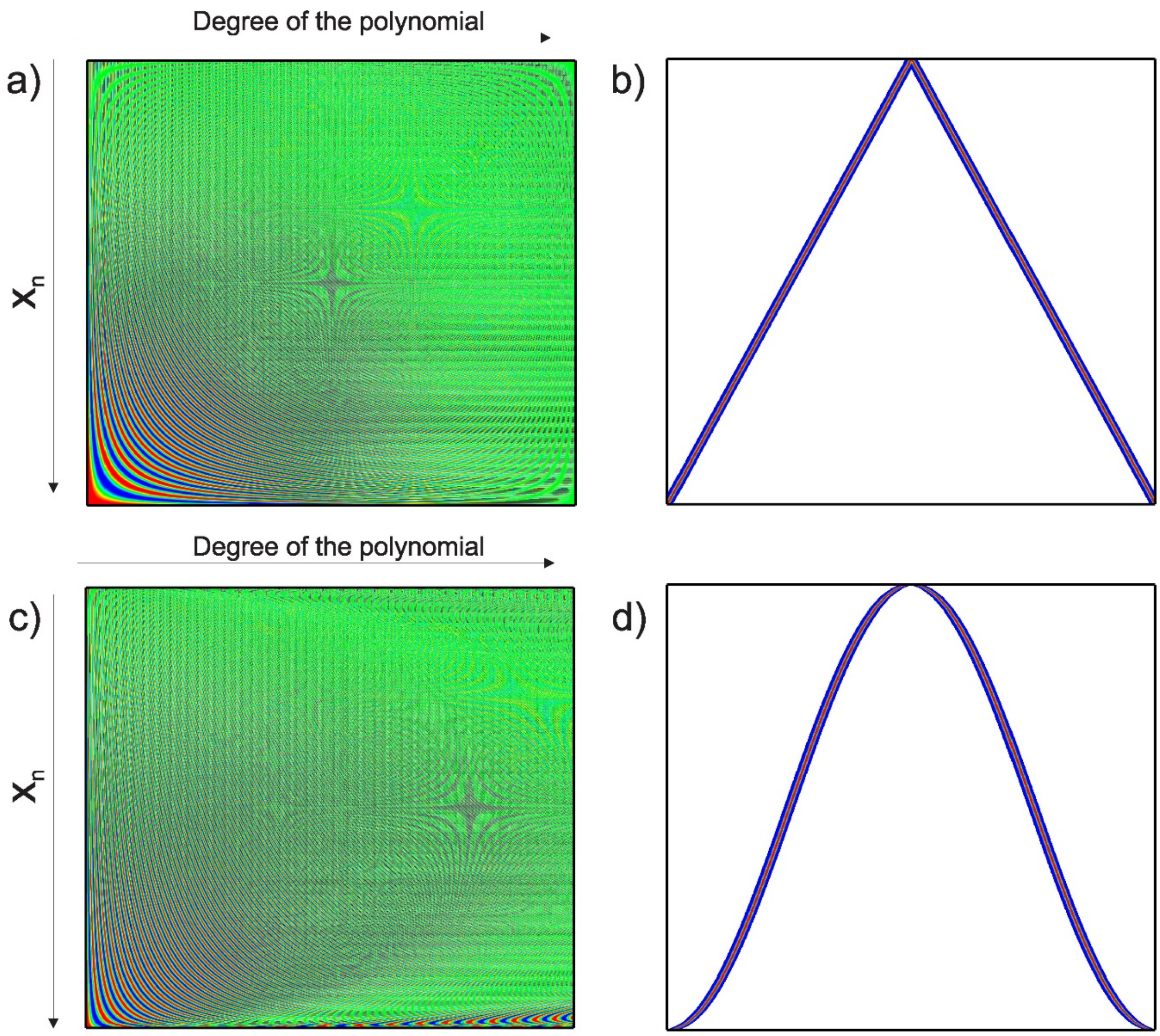

Рис. 3: Матрица $A_e \in \mathbb{R}^{512\times512}$ для (a) Чебышевских узлов и (b) равномерно распределённых

узлов. Абсолютные значения элементов сжатой матрицы $A_e W_{512} \mathcal{F}^* \in \mathbb{C}^{512\times512}$ для (c) Чебышевских узлов и (d) равномерно распределённых узлов

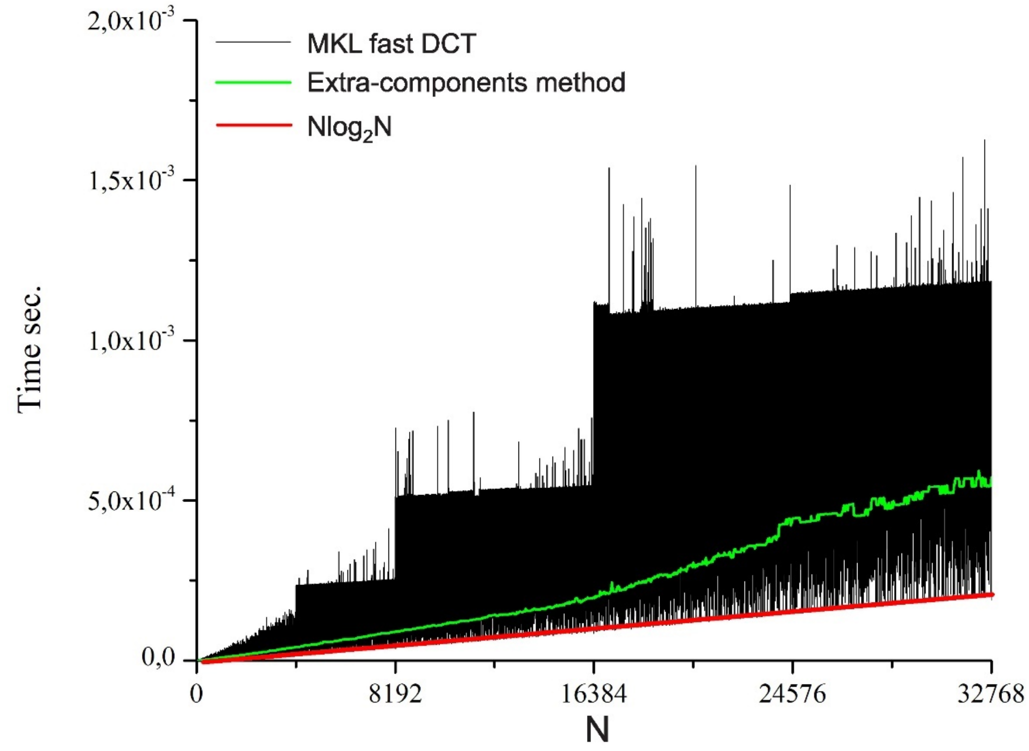

Рис. 4: Зависимость времени счёта вычисления дискретного косинус-преобразования от размерности N

Эффект от использования кластера в достижении целей работы.

Значительное сокращение времени расчетов, доступ к большим объемам оперативной памяти.

Перечень публикаций, содержащих результаты работы (если есть).

1. Andrew V. Terekhov, Generating the Laguerre expansion coefficients by solving a one-dimensional transport equation. Numerical Algorithms 89, 303–322 (2022), https://doi.org/10.1007/s11075-021-01115-8.
2. Andrew V. Terekhov, A divide-and-conquer algorithm for seismic data approximation by the Laguerre series//*J. Phys.: Conf. Ser.* 2099 012062, 2021. https://doi.org/10.1088/1742-6596/2099/1/012062
3. Terekhov, A.V. An extra-component method for evaluating fast matrix-vector multiplication with special functions. Numer Algor (2022). https://doi.org/10.1007/s11075-022-01383-y