

Н. А. Аношина¹, Д. В. Сорокин¹

МЕТОД СОВМЕЩЕНИЯ ИЗОБРАЖЕНИЙ КРИО-ЭМ ОДИНОЧНЫХ ЧАСТИЦ НА ОСНОВЕ ПРЕОБРАЗОВАНИЯ ФУРЬЕ-БЕССЕЛЯ И БЫСТРОГО ПРОЕКЦИОННОГО МЕТОДА С ИСПОЛЬЗОВАНИЕМ ФУНКЦИЙ ЛАГЕРРА*

Введение

За последнее время появились технологии, позволяющие строить трехмерные модели высокого разрешения молекул биологических образцов для подробного изучения их строения. Одной из таких технологий является криоэлектронная микроскопия (крио-ЭМ) одиночных частиц [1]. Данная технология использует криогенные условия, где образец предварительно замораживается до температуры жидкого этана (-182°) для последующего анализа в электронном микроскопе. Так, исследуемый образец находится в естественной среде, ничем не окрашенный и никак не зафиксированный. Однако, в результате взаимодействия с пучком электронов образец повреждается, и из-за этого полученное изображение имеет очень низкое отношение сигнала к шуму (англ. *SNR*).

Криоэлектронная микроскопия одиночных частиц состоит из следующих шагов (Рис. 1). Сначала, предварительно обеспечив криогенные условия, с помощью электронного микроскопа получают изображение. Такое изображение содержит много проекций молекул исследуемого образца, расположенных хаотично в плоскости изображения (Рис. 2(а)). Для того, чтобы по полученным проекциям частиц построить трехмерную модель молекулы, нужно выделить все проекции на изображении. Обычно это делается с помощью компьютерных методов, которые работают в автоматическом или полуавтоматическом режиме [2]. Далее, используя методы кластеризации, выделенные проекции группируют по классам таким образом, что внутри одной группы находятся изображения проекций, которые одинаково ориентированы в 3Д пространстве и отличаются друг от друга поворотом

¹Лаборатория математических методов обработки изображений, факультет Вычислительной математики и кибернетики, МГУ имени М.В. Ломоносова, e-mail: n.anoshina@cs.msu.ru, dsorokin@cs.msu.ru

*Исследование выполнено за счет гранта Российского научного фонда №22-21-00125.

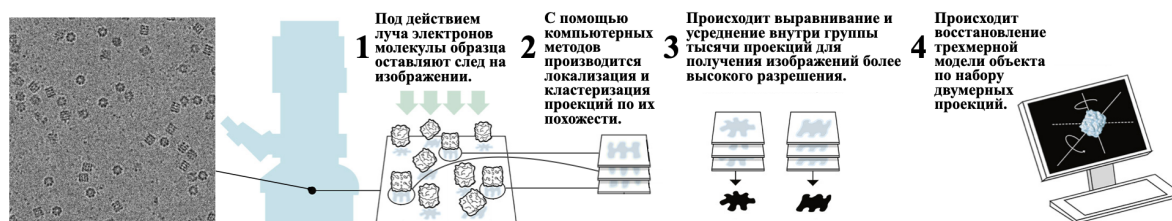


Рис. 1. Криоэлектронная микроскопия одиночных частиц: общий алгоритм работы.

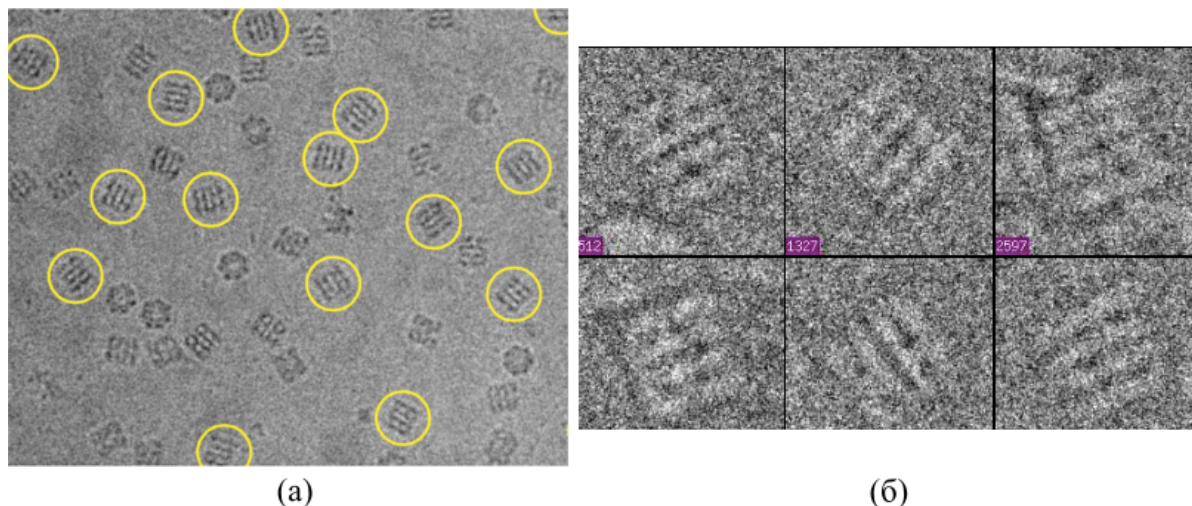


Рис. 2. а) Пример изображения криоэлектронной микроскопии одиночных частиц, желтыми кругами отмечены локализованные проекции частиц одного класса. б) Увеличенные изображения частиц одного класса с различной ориентацией.

и смещением в плоскости проекции (Рис. 2(б)). Следующий шаг – приведение всех проекций одной группы к единой ориентации с помощью двумерных методов совмещения и их усреднение. Это позволяет повысить отношение сигнала к шуму и улучшить разрешение проекций. Количество проекций, которые необходимо выровнять, составляет сотни, а иногда и тысячи изображений. Данный процесс классификации, выравнивания и усреднения проекций является итеративным, поскольку заранее неизвестно количество классов – различных ориентаций частицы. На каждой итерации вычисляется метрика схожести между проекцией-кандидатом со схожей ориентацией и текущей усредненной версией группы проекций. Затем усредненные проекции используются для построения первичной трехмерной модели. Процедура выравнивания проекций внутри группы, их усреднение и классификация является очень важным этапом в построении первичной трехмерной модели высокого разрешения.

За последние годы было разработано множество различных методов, позволяющих улучшить отдельные этапы технологии крио-ЭМ одиночных частиц [2]. Данная работа посвящена разработке методов выравнивания изображений отдельных проекций для последующего усреднения. Математически, данная задача является задачей *жесткого совмещения двумерных изображений*.

Существуют различные подходы совмещения двумерных изображений [3]. Среди них можно выделить классические методы совмещения, которые используют математические операции, такие как вычисление преобразования Фурье или функции корреляции, решение системы уравнений и др., и методы машинного обучения, которые хорошо зарекомендовали себя в последнее время. К классическим относятся методы, основанные на поиске параметров совмещения (угол поворота и двумерный вектор смещения) в пространственной области, что в большинстве случаев представляет из себя вычислительно затратный итеративный поиск. Также существует группа методов, которые выполняют поиск параметров совмещения в частотном пространстве и используют различные модификации преобразования Фурье [4–11]. Помимо этого, существуют методы, разработанные специально для криоэлектронной микроскопии одиночных частиц, которые решают задачу выравнивания совместно с задачей кластеризации частиц [12, 13]. Метод, представленный в [13], использует итеративный поиск параметров совмещения, и из-за этого является очень времязатратным. Некоторые методы совмещения реализованы в рамках комплексного программного обеспечения (ПО), которые содержат процедуры выравнивания частиц и их усреднение. Примерами таких программных комплексов являются CRYOSPARC [14], RELION [15] и FREALIGN [16]. Наиболее популярным является RELION [15], где авторы используют метод максимального правдоподобия для решения задачи двумерной классификации и выравнивания проекций.

Методы машинного обучения и применение нейронных сетей набирают популярность во многих областях. В крио-ЭМ одиночных частиц методы машинного обучения активно используются для различных подзадач, таких как, например, выделение частиц [17], построение 3Д модели частицы [18] или уменьшения количества шума на изображениях крио-ЭМ [19]. Однако, для выравнивания и классификации частиц принято использовать аналитические методы, поскольку основная цель разработки новых методов в области крио-ЭМ – это увеличение скорости обработки данных и усовершенствование процесса классификации частиц.

В данной работе представлен метод совмещения двумерных изображений, который может быть эффективно применен для выравнивания проекций внутри одного класса в криоэлектронной микроскопии одиночных частиц. Предложенный подход основан на вычислении функции кросс-корреляции между парой изображений и использует преобразование Фурье-Бесселя для ее вычисления. Поиск угла поворота и вектора смещения происходит одновременно. Предложенный подход является улучшением метода [11], где повышение эффективности совмещения достигается за счет применения проекционного метода с использованием функций Лагерра для вычисления преобразования Фурье-Бесселя. Проекционный метод также дополнительно ускорен с помощью быстрого алгоритма расчета проекционных коэффициентов на основе квадратуры Гаусса-Лагерра в вычислении преобразования Ганкеля. Применение проекционного метода существенно ускоряет метод по сравнению с [11], а применение квадратуры Гаусса-Лагерра в расчете коэффициентов разложения также дает дополнительное ускорение проекционного метода. Предложенный метод с применением быстрого алгоритма расчета коэффициентов и без него был протестирован на синтетических и реальных данных. Результаты исследований на наборах данных с различным SNR показали, что все версии метода имеют устойчивость к шуму и позволяют получать точные результаты даже при низких показателях SNR . Для демонстрации эффективности метода и его ускорения было проведено сравнение предложенного метода совмещения с использованием проекционного метода вычисления преобразования Фурье-Бесселя, а также с использованием быстрого метода расчета проекционных коэффициентов, и со стандартной версией из [11]. Помимо этого, предложенный подход был сравнен с корреляционным методом, использующим пространство Фурье для вычисления корреляции отдельно для поиска угла поворота и вектора смещения [7].

Постановка задачи

Для выравнивания проекций частиц в крио-ЭМ используют следующую постановку задачи. Дана пара изображений, одно из которых называется фиксированным f , другое подвижным g . Требуется найти такое преобразование $M_{\theta,x,y}$:

$$g \circ M_{\theta,x,y} \approx f,$$

где θ – искомый угол поворота, (x,y) – искомый вектор смещения. Здесь $g \circ M_{\theta,x,y}$ обозначает результат преобразования изображения g . Преобразование $M_{\theta,x,y}$ является аффинным и представлено в виде композиции матрицы поворота на угол θ и матрицы смещения на вектор (x,y) :

$$M_{\theta,x,y} = T_{x,y} \circ R_{\theta}.$$

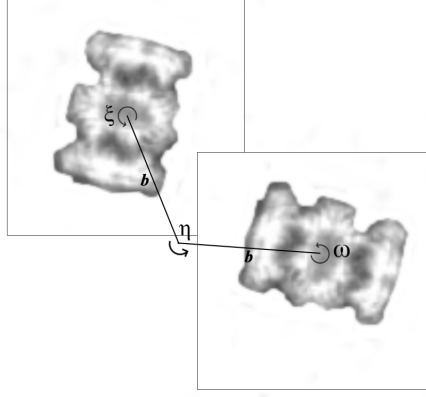


Рис. 3. Параметры искомого преобразования, заданного тройкой углов (ξ, η, ω) .

Обычно, в качестве фиксированного изображения выступает всегда одно и то же эталонное изображение, которым может быть либо усредненная на предыдущих итерациях проекция, либо сгенерированное с помощью проецирования трехмерной модели на плоскость изображение, либо просто случайно выбранное изображение из группы проекций, которые необходимо привести к одному виду.

Предложенный метод представляет $M_{\theta,x,y}$ через два вращения и смещение вдоль фиксированной оси:

$$M_{\theta,x,y} = \hat{M}_{\varphi,\rho,\psi} = R_{\varphi} \circ T_{\rho,0} \circ R_{\psi} = T_{\rho \cos \varphi, \rho \sin \varphi} \circ R_{\varphi+\psi} = M_{\varphi+\psi, \rho \cos \varphi, \rho \sin \varphi}.$$

Далее, перейдя к композиции двух преобразований $\hat{M} = \hat{M}_1 \circ \hat{M}_2$, где $\hat{M}_{\varphi_i, \rho_i, \psi_i}$, $i = 1, 2$ и зафиксировав величину смещения $\rho_1 = \rho_2 = b$ (b выбирается как половина максимально возможного смещения), можно перейти к заданию преобразования M через тройку углов (ξ, η, ω) (Рис. 3). Подробности данного перехода даны в работе [11].

Таким образом, для вычисления жесткого преобразования между двумя изображениями f и g достаточно задать параметр b и найти тройку (ξ, η, ω) . Ниже описано, каким образом можно найти данные углы с помощью вычисления корреляционной функции на основе преобразований Фурье-Бесселя.

Преобразования Ганкеля и Фурье-Бесселя

Преобразование Ганкеля порядка α определяется следующим образом:

$$H_{\alpha}[z(x)] = \tilde{z}_{\alpha}(u) = \int_0^{\infty} z(x) J_{\alpha}(ux) x dx, H_{\alpha} : L_2[0, \infty) \rightarrow L_2[0, \infty], \quad (1)$$

где J_{α} — функция Бесселя первого рода порядка $\alpha \geq 0$.

Представим функцию f в полярных координатах $f = f(r, \lambda)$, тогда преобразованием Фурье-Бесселя функции f называется [20]:

$$F_m(x) = \int_0^{\infty} \hat{f}_m(u) J_m(ux) u du = H_m[\hat{f}_m(u)], \forall x \geq 0, \quad (2)$$

где $\hat{f}_m(u)$ – преобразование Фурье функции f вдоль угловой оси полярных координат для фиксированного радиуса u :

$$\hat{f}_m(u) = \frac{1}{2\pi} \int_0^{2\pi} f(u, \theta) e^{-im\theta} d\theta. \quad (3)$$

Метод совмещения на основе преобразования Фурье-Бесселя

Ниже кратко описан метод совмещения на основе преобразования Фурье-Бесселя, более подробное его изложение можно найти в [11]. Пусть $f(r, \lambda)$ и $g(r, \lambda)$ – заданные в полярных координатах исходные изображения, которые необходимо совместить, $F_m(x)$ и $G_m(x)$ – преобразования Фурье-Бесселя для f и g соответственно. Тогда функция корреляции от трех углов определяется так:

$$c(\xi, \eta, \omega) = 2\pi \sum_{m, h, m'} e^{i(m\xi + h(\eta + \varepsilon) + m'\omega)} \times \int_0^\infty J_{m-h}(bx) J_{h-m'}(bx) F_m(x) \overline{G_{m'}(x)} x dx, \quad (4)$$

где ε – маленький угол для того, чтобы избежать повторяющихся значений в семплировании угла η . Авторы предлагают зафиксировать это значение равным $\pi/(2k)$. Затем производится замена индексов и переменных для удобства вычислений:

$$h = h_1 + m', \quad m = m_1 + h = m_1 + h_1 + m', \\ \eta' = \xi + \eta, \quad \omega' = \xi + \eta + \omega.$$

В результате функция корреляции $c(\xi, \eta, \omega)$ преобразуется в функцию $c'(\xi, \eta', \omega')$, преобразование Фурье которой выражается следующим образом:

$$\hat{T}(m_1, h_1, m') = 2\pi e^{i(h_1 + m')\varepsilon} \times \int_0^\infty F_{m_1 + h_1 + m'}(x) \overline{G_{m'}(x)} J_{m_1}(bx) J_{h_1}(bx) x dx. \quad (5)$$

Так, вычислив обратное преобразование Фурье для (5) получим функцию корреляции, положение максимального значения которой соответствует искомым параметрам. Эффективность данного метода в том, что процедура совмещения пары изображений сведена к однократному вычислению обратного преобразования Фурье. Также в формуле (5) некоторые значения под интегралом, а именно преобразования Фурье-Бесселя $F_{m_1 + h_1 + m'}(x)$, $\overline{G_{m'}(x)}$ и функции Бесселя $J_{m_1}(bx)$, $J_{h_1}(bx)$ могут быть вычислены однократно и затем переиспользоваться.

Для того, чтобы определить количество точек для вычисления интегралов в (2), (3) и (5), а также области изменения порядков функций Бесселя m_1 , h_1 и m' нужно задать следующие параметры:

- Радиус изображения, т.е. радиус круговой области, внутри которой полностью содержится объект интереса на обоих изображениях.

- Среднее количество пикселей на точку семплирования.
- Значение максимально возможного расстояния, на которое может быть сдвинут объект.

Все остальные значения, используемые в дискретизации интегралов, могут быть выведены из представленных [11].

Проекционный метод с использованием функций Лагерра для преобразования Фурье-Бесселя

Ниже предложен способ ускорения метода совмещения на основе преобразования Фурье-Бесселя, описанного в предыдущем разделе, путем применения проекционного метода с использованием функций Лагерра для вычисления преобразования Ганкеля в (2).

Построим проекционный метод вычисления преобразования Ганкеля (1) [21], используя систему собственных функций данного преобразования – функций Лагерра [22]. Нормированные функции Лагерра определены следующим образом:

$$\begin{aligned}\psi_n^\alpha(x) &= \frac{1}{\sqrt{n!\Gamma(n+\alpha+1)}} x^{\alpha/2} e^{-x/2} L_n^\alpha(x), \\ L_n^\alpha(x) &= (-1)^n x^{-\alpha} e^x \frac{d}{dx^n} (x^{n+\alpha} e^{-x}),\end{aligned}\tag{6}$$

где $L_n^\alpha(x)$ – полином Лагерра, $n = 0, 1, \dots$

Для их вычисления можно использовать рекуррентное соотношение:

$$\begin{aligned}\psi_{n+1}^\alpha(x) &= \frac{(x - \alpha - 2n - 1)}{\sqrt{(n+1)(n+\alpha+1)}} \psi_n^\alpha(x) - \\ &\quad - \sqrt{\frac{n(n+\alpha)}{(n+1)(n+\alpha+1)}} \psi_{n-1}^\alpha(x), \quad n = 0, 1, \dots, \\ \psi_0^\alpha(x) &= \frac{1}{\sqrt{\Gamma(\alpha+1)}} x^{\alpha/2} e^{-x/2}, \quad \psi_{-1}^\alpha(x) \equiv 0.\end{aligned}$$

Функции Лагерра $\psi_n^\alpha(x^2)$ являются собственными функциями преобразования Ганкеля на полупрямой:

$$H_\alpha [\psi_n^\alpha(x^2)] = \int_0^\infty \psi_n^\alpha(x^2) J_\alpha(ux) x dx = (-1)^n \psi_n^\alpha(u^2), \quad n = 0, 1, \dots$$

Перейдем к ортонормированной системе функций $\{\hat{\psi}_n^\alpha(x)\}$: $\hat{\psi}_n^\alpha(x) = \sqrt{2x} \psi_n^\alpha(x^2)$. Пусть

$$\hat{H}_\alpha [z(x)] = \int_0^\infty z(x) J_\alpha(ux) \sqrt{ux} dx,$$

тогда

$$\begin{aligned}H_\alpha [z(x)] &= \int_0^\infty z(x) J_\alpha(ux) x dx = \\ &= \frac{1}{\sqrt{u}} \int_0^\infty (z(x) \sqrt{x}) J_\alpha(ux) \sqrt{ux} dx = \frac{1}{\sqrt{u}} \hat{H}_\alpha [z(x) \sqrt{x}],\end{aligned}\tag{7}$$

где H_α – оператор преобразования Ганкеля на полупрямой (1), а функции $\hat{\psi}_n^\alpha(x)$ являются собственными функциями оператора $\hat{H}_\alpha$:

$$\begin{aligned}\hat{H}_\alpha[\hat{\psi}_n^\alpha(x)] &= \int_0^\infty \hat{\psi}_n^\alpha(x) J_\alpha(ux) \sqrt{ux} dx = \\ &= \int_0^\infty \sqrt{2x} \psi_n^\alpha(x^2) J_\alpha(ux) \sqrt{ux} dx = \sqrt{2u} H_\alpha[\psi_n^\alpha(u^2)] = \\ &= \sqrt{2u} (-1)^n \psi_n^\alpha(u^2) = (-1)^n \hat{\psi}_n^\alpha(u).\end{aligned}\quad (8)$$

Таким образом, для вычисления преобразования Фурье-Бесселя, являющегося преобразованием Ганкеля от функции $\hat{f}_m(u)$ (2), достаточно выполнить разложение данной функции в ряд по функциям Лагерра:

$$\begin{aligned}(\hat{f}_m(u) \sqrt{u}) &= \sum_{n=0}^{N-1} c_n \hat{\psi}_n^m(u), \\ c_n &= \int_0^\infty (\hat{f}_m(u) \sqrt{u}) \hat{\psi}_n^m(u) du, \quad n = 0, \dots, N-1,\end{aligned}\quad (9)$$

а далее, используя уравнения (2), (7), (8) и вычисленные коэффициенты разложения c_n , можно получить преобразование Фурье-Бесселя:

$$\begin{aligned}F_m(x) &= H_m[\hat{f}_m(u)] = \frac{1}{\sqrt{x}} H_m[\hat{f}_m(u) \sqrt{u}] = \\ &= \frac{1}{\sqrt{x}} H_m \left[\sum_{n=0}^{N-1} c_n \hat{\psi}_n^m(u) \right] = \frac{1}{\sqrt{x}} \sum_{n=0}^{N-1} c_n H_m[\hat{\psi}_n^m(u)] = \\ &= \frac{1}{\sqrt{x}} \sum_{n=0}^{N-1} (-1)^n c_n \hat{\psi}_n^m(x).\end{aligned}\quad (10)$$

Ключевым шагом данного метода является выбор параметра N , задающего количество функций Лагерра, используемых в разложении, от которого зависит точность восстановления преобразования Фурье-Бесселя. Помимо этого, необходимо задать параметр масштабирования сетки a : $\bar{x}_i = ax_i$. Параметр масштабирования задается таким образом, чтобы функции Лагерра помещались целиком на заданной сетке, а именно имели пики соответственно параметрам n и m и ближе к концу заданной сетки затухали к нулевому значению.

Быстрый алгоритм расчета проекционных коэффициентов

В данной работе предлагается использовать быстрый метод [21], основанный на применении квадратуры Гаусса-Лагерра, для вычисления проекционных коэффициентов c_n . В формуле расчета коэффициентов (9) возникают уравнения вида:

$$c_n = \int_0^\infty \hat{f}(u) \hat{\psi}_n^m(u) du,$$

где

$$\hat{f}(u) = (\hat{f}_m(u) \sqrt{u}).$$

Преобразуем уравнение:

$$\begin{aligned} c_n &\approx \int_0^\infty \hat{f}(u) \hat{\psi}_n^m(u) du = \int_0^\infty \hat{f}(u) \sqrt{2u} \psi_n^m(u^2) du = \\ &= \int_0^\infty \hat{f}(u) (2u)^{-\frac{1}{2}} \psi_n^m(u^2) d(u^2). \end{aligned} \quad (11)$$

Сделаем следующую замену переменных $u^2 = t$ и $\beta_n^m = \sqrt{n!(n+m+1)}$, тогда

$$\begin{aligned} \int_0^\infty \hat{f}(u) (2u)^{-\frac{1}{2}} \psi_n^m(u^2) d(u^2) &= \frac{1}{\sqrt{2}} \int_0^\infty \hat{f}(\sqrt{t}) t^{-\frac{1}{4}} \psi_n^m(t) dt = \\ &= \frac{1}{\sqrt{2}} \int_0^\infty t^m e^{-t} \left(\hat{f}(\sqrt{t}) \frac{1}{\beta_n^m} t^{-\frac{m}{2}-\frac{1}{4}} e^{\frac{t}{2}} L_n^m(t) \right) dt. \end{aligned} \quad (12)$$

Принимая во внимание (11), уравнение (12) может быть преобразовано так:

$$c_n \approx \frac{1}{\sqrt{2}} \int_0^\infty t^m e^{-t} y(t) dt, \quad (13)$$

где

$$y(t) = \hat{f}(\sqrt{t}) \frac{1}{\beta_n^m} t^{-\frac{m}{2}-\frac{1}{4}} e^{\frac{t}{2}} L_n^m(t). \quad (14)$$

Интеграл (13) может быть вычислен с помощью квадратуры Гаусса-Лагерра [23]:

$$\int_0^\infty t^m e^{-t} y(t) dt \approx \sum_{k=1}^M A_k f(t_k) + R_M, \quad (15)$$

где

$$A_k = \frac{(\beta_M^m)^2 t_k}{(L_{M+1}^m(t_k))^2}, \quad (16)$$

а t_k – нули полинома Лагерра L_N^m , M – нужное количество нулей полинома и R_M – остаток. В данной работе был использован метод Эберса [24] для нахождения нулей полинома L_M^m . Такой выбор обусловлен тем, что применение стандартных вычислительных методов, например, метод Ньютона или метод деления отрезка пополам и т.д., для данной задачи затруднено в связи с плотным расположением нулей полинома Лагерра для больших M ($M > 100$). Стандартные методы находят не все нули или иногда просто расходятся. Значения коэффициентов A_k для $\alpha = 0$ и различных значений t_k и M представлены в таблице 1 [23].

Однако, если использовать (16) для вычисления коэффициентов A_k , то это приведет к уменьшению точности аппроксимации и увеличению вычислительной сложности алгоритма.

Перепишем (15), принимая во внимание (14):

$$\begin{aligned}
\int_0^\infty t^m e^{-t} y(t) dt &\approx \sum_{k=1}^M A_k f(t_k) + R_M = \\
&= \sum_{k=1}^M \frac{(\beta_M^m)^2 t_k \hat{f}(\sqrt{t_k}) t_k^{-\frac{m}{2}-\frac{1}{4}} e^{\frac{t_k}{2}} L_n^m(t_k)}{(L_{M+1}^m(t_k))^2 \beta_n^m} + R_M = \\
&= \sum_{k=1}^M \frac{\hat{f}(\sqrt{t_k}) t_k^{\frac{3}{4}} \psi_n^m(t_k)}{\left(\frac{\beta_{M+1}^m}{\beta_n^m}\right)^2 (\psi_{M+1}^m(t_k))^2} + R_M = \\
&= \sum_{k=1}^M \frac{\hat{f}(\sqrt{t_k}) t_k^{\frac{3}{4}} \psi_n^m(t_k)}{(M+1)(M+m+1) (\psi_{M+1}^m(t_k))^2} + R_M.
\end{aligned}$$

Тогда коэффициенты разложения c_n могут быть эффективно вычислены с помощью следующей формулы:

$$c_n \approx \frac{1}{\sqrt{2}M(M+m)} \sum_{k=1}^M \hat{f}(\sqrt{t_k}) \mu_{M+1}^n(t_k), \quad (17)$$

где

$$\mu_M^n(t_k) = \frac{t_k^{\frac{3}{4}} \psi_n^m(t_k)}{(\psi_M^m(t_k))^2}.$$

Значения $\hat{f}(\sqrt{t_k})$ могут быть вычислены с помощью линейной интерполяции для достаточно гладкой функции $\hat{f}$, поскольку узлы сетки t_k расположены достаточно разреженно.

Для того, чтобы ускорить время работы метода, поиск нулей полинома можно выполнить заранее, поскольку эти данные определяются двумя параметрами, а именно количество нулей M и порядок m полинома Лагерра L_n^m . Также нахождение нулей для полиномов больших порядков может занять значительное время. В данной работе нули полинома L_n^m были вычислены предварительно и сохранены в файл.

Количество операций

Главным преимуществом вычисления преобразования Фурье-Бесселя с помощью проекционного метода в данной работе является существенное ускорение вычислений. В рассуждениях ниже предполагается, что значения функций Бесселя и функций Лагерра вычислены заранее на необходимых сетках и переиспользуются. Так, при вычислении $F_m(x)$ по формуле (2) необходимо выполнить $O(K^2)$ умножений, где K это количество точек в сетке по координатам u и x (считаем его одинаковым). При вычислении $F_m(x)$ проекционным методом по формуле (10), необходимо посчитать N коэффициентов разложения c_n для вычисления каждого из которых требуется порядка

$M = 1$	
$t_1 = 1.000000000000$	$A_1 = 1.000000000000$
$M = 2$	
$t_1 = 0.585786437627$	$A_1 = 0.853553390593$
$t_2 = 3.414213562373$	$A_2 = 0.146446609407$
$M = 7$	
$t_1 = 0.193043676560$	$A_1 = 0.409318951701$
$t_2 = 1.026664895339$	$A_2 = 0.421831277862$
$t_3 = 2.567876744951$	$A_3 = 0.147126348658$
$t_4 = 4.900353084526$	$A_4 = 0.(1)206335144687$
$t_5 = 8.182153444563$	$A_5 = 0.(2)107401014328$
$t_6 = 12.734180291798$	$A_6 = 0.(4)158654643486$
$t_7 = 19.395727862263$	$A_7 = 0.(7)317031547900$

Табл. 1. Значения коэффициентов A_k для $\alpha = 0$.

$O(K)$ умножений, а также посчитать взвешенную сумму из (10). Таким образом, общее количество операций имеет порядок $O(N * K)$, что дает существенное ускорение, при значении N существенно меньшем чем K . В экспериментах, проведенных в данной работе $N = 50$, $K = 256$, а общее ускорение для совмещения пары изображений составило 3.75 раза.

Использование алгоритма быстрого расчета проекционных коэффициентов по формуле (17) с помощью квадратуры Гаусса-Лагерра позволяет дополнительно ускорить вычисления. В сравнении со стандартным проекционным методом, количество операций, необходимое для вычисления c_n с помощью квадратурной формулы на равномерной сетке, пропорционально количеству точек сетки и имеет порядок $O(M)$, в данном случае сетке из нулей полинома t_k . Если вычислить $\mu_M^n(t_k)$ заранее, то общее количество операций для вычисления $F_m(x)$ имеет порядок $O(N)$. Количество нулей M может быть подобрано исходя из необходимой точности аппроксимации и времени выполнения. В экспериментах, проведенных в данной работе $N = 50$, $K = 256$, $M = 2N$, а общее ускорение для совмещения пары изображений составило 5 раз в сравнении со стандартным подходом и 1.43 раза в сравнении с предложенным методом, использующим проекционный метод для вычисления $F_m(x)$. Ускорение было вычислено на паре изображений из реального набора данных, каждое изображение имеет размер 128×128 .

В этой работе предложенные методы были сравнены с корреляционным подходом из работы [7]. Данный метод использует порядка $O(P \log P)$ операций для совмещения пары изображений, где P задает количество пикселей одного из совмещаемых изображений.

Результаты на синтетических данных

Для того, чтобы численно оценить работу метода, были созданы 2 набора синтетических данных с заранее известными искомыми параметрами, а именно, углом поворота и двумерным вектором смещения. Наборы данных отличаются трехмерными моделями, проекции которых были использованы при создании набора. Для генерации набора данных был использован метод, описанный в [25]. Для того, чтобы смоделировать различное количество шума, к сгенерированному набору данных был добавлен Пуассоновский шум соответственно следующим значениям отношения SNR : [0.1, 0.5, 1.0, 2.0]. Отношение сигнала к шуму определено следующим образом:

$$SNR = \frac{\mu_{\text{изображение}}}{\sigma_{\text{шум}}},$$

где $\mu_{\text{изображение}}$ – среднее значение пикселей изображения без шума, $\sigma_{\text{шум}}$ – стандартное отклонение уровня шума. Для выравнивания проекций было выбрано эталонное изображение, равное первому кадру последовательности. Было проведено сравнение стандартного метода из [11], предложенного метода на основе преобразования Фурье-Бесселя и его вычисления проекционным методом, а также его ускоренной версии с помощью применения быстрого алгоритма расчета проекционных коэффициентов, основанного на квадратурной формуле Гаусса-Лагерра. Помимо этих методов, было проведено сравнение с корреляционным методом из [7]. Для тестирования работы методов были вычислены ошибки угла поворота и вектора смещения. Ошибка угла поворота определена как $e_{\theta} = |\theta_{gt} - \theta|$ в градусах, где θ_{gt} – заданный угол поворота, θ – найденный угол поворота. Ошибка вектора смещения вычислена так $e_d = \|d_{gt} - d\|$, где $d_{gt} = (x_{gt}, y_{gt})$ – заданный вектор смещения и $d = (x, y)$ – найденный вектор смещения. Ошибка угла поворота, кратная 180 градусам, не была учтена, поскольку выбранные проекции частицы являются симметричными и трудноразличимыми. Описанные ошибки были вычислены для каждого из значений уровня шума для двух наборов данных. Результаты проведенных экспериментов приведены на Рис. 4 и в таблице 2. Также на Рис. 5 приведен график зависимости ошибок найденных параметров от уровня шума (соответствующего значения SNR). По результатам экспериментов можно сделать вывод, что предложенный проекционный метод не уступает по качеству методу на основе преобразования Фурье-Бесселя и имеет более высокую скорость вычисления (общее ускорение в среднем в 3.75 раз). Использование быстрого метода вычисления проекционных коэффициентов также не сильно увеличивает ошибку совмещения, в то время как ускоряет проекционный метод (общее ускорение в среднем в 1.43 раз). В сравнении с корреляционным методом, предложенные подходы более устойчивы к большему уровню шума, что соответствует

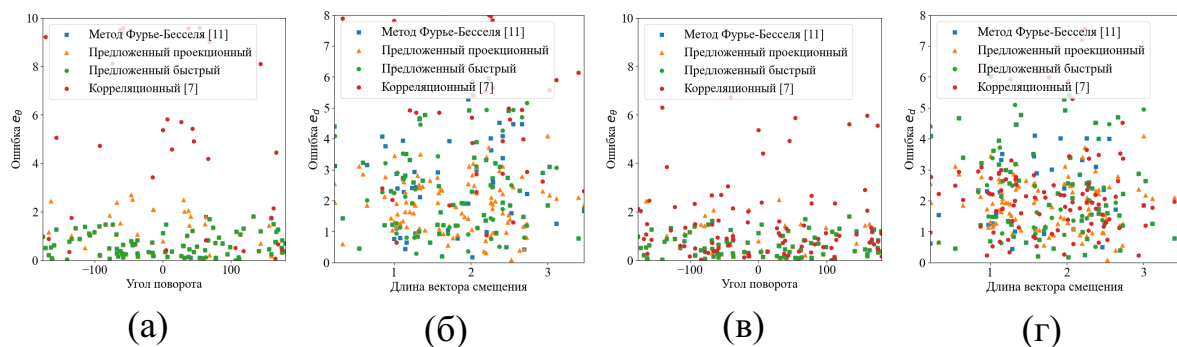


Рис. 4. Ошибки найденных параметров для разного уровня шума $SNR = 0.1$ (слева) и $SNR = 1.0$ (справа): угла поворота e_θ (на графиках а и в) и длины вектора смещения e_d (на графиках б и г) для предложенного проекционного метода, предложенного быстрого проекционного метода, исходного метода Фурье-Бесселя [11] и корреляционного [7].

низкому отношению сигнала к шуму ($SNR=0.1$), и имеют меньшую ошибку в нахождении угла поворота. Однако для вектора смещения при высоком отношении сигнала к шуму, а именно при $SNR=2.0$, ошибки предложенных методов, как и метода на основе Фурье-Бесселя, больше в сравнении с корреляционным методом. Это может быть связано с тем, что количество точек в вычислении трехмерной функции корреляции по осям, соответствующим углам, отвечающим за смещение объекта, в три раза меньше, чем количество точек для параметра, соответствующего углу поворота. Также для Модели 2 применение быстрого алгоритма несколько ухудшает результаты по углу поворота и вектору смещения для данных с высоким уровнем шума. Это может быть связано с неточным выбором параметров, необходимых для работы быстрого метода, таких как, например, количество нулей полинома L_n^m . Итак, предложенные методы существенно ускоряют базовый алгоритм из [11], не уступая при этом в качестве совмещения.

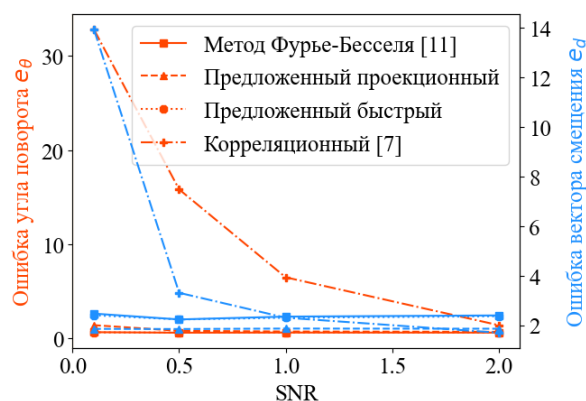


Рис. 5. Ошибки угла поворота e_θ и длины вектора смещения e_d для разных значений SNR .

SNR	Методы	Модель 1		Модель 2	
		Средняя ошибка угла поворота e_θ	Средняя ошибка вектора смещения e_d	Средняя ошибка угла поворота e_θ	Средняя ошибка вектора смещения e_d
0.1	Предложенный проекционный метод	0.63	2.45	0.56	2.58
	Предложенный быстрый метод	0.65	2.47	5.00	5.52
	Метод Фурье-Бесселя [11]	1.4	1.85	0.58	2.05
	Корреляционный метод [7]	32.82	13.94	46.7	20.38
0.5	Предложенный проекционный метод	0.61	2.25	0.59	2.67
	Предложенный быстрый метод	0.61	2.23	0.61	2.94
	Метод Фурье-Бесселя [11]	0.8	1.84	0.59	1.95
	Корреляционный метод [7]	15.82	3.30	2.37	2.2
1.0	Предложенный проекционный метод	0.62	2.33	0.57	2.74
	Предложенный быстрый метод	0.62	2.34	0.63	2.84
	Метод Фурье-Бесселя [11]	0.72	1.86	0.56	1.97
	Корреляционный метод [7]	6.48	2.29	1.22	2.06
2.0	Предложенный проекционный метод	0.61	2.39	0.58	2.76
	Предложенный быстрый метод	0.61	2.39	0.58	2.81
	Метод Фурье-Бесселя [11]	0.69	1.85	0.58	1.92
	Корреляционный метод [7]	1.4	1.69	1.07	1.84

Табл. 2. Сравнение работы методов на синтетических данных с разным уровнем шума, определенные различными значениями SNR .

Результаты на реальных данных

Мы протестировали предложенные методы на реальных данных и сравнили их с базовым подходом из [11], а также с корреляционным методом из [7]. Набор данных из проекций реальных частиц был взят из банка данных электронной микроскопии (EMDB [26]). Данный набор состоит из 50 проекций Icosahedral Bacillus stearotherophilus pyruvate dehydrogenase E2 core [27], которые имеют одинаковую ориентацию и их необходимо выровнять (т.е. привести к одному виду). На Рис. 6(а) представлены несколько проекций из набора данных до совмещения. Мы взяли 1-ую, 8-ую, 15-ую, 22-ую, 29-ую, 36-ую, 43-ью и 50-ую проекцию из набора для иллюстрации. Все проекции в наборе были выровнены в соответствии с проекцией на первом кадре в наборе. Ввиду отсутствия искомым параметров было проведено визуальное сравнение работы методов. На Рис. 6(б) изображен результат совмещения корреляционным методом, базовым методом на основе преобразования Фурье-Бесселя (Рис. 6(в)), предложенным проекционным методом ((Рис. 6(г))) и быстрым методом (Рис. 6(д)). Все проекции были выровнены относительно первого изображения в наборе (соответствует крайнему левому кадру). Крайнее правое изображение в каждом ряду соответствует усредненной проекции по всему набору из 50 изображений после процедуры совмещения.

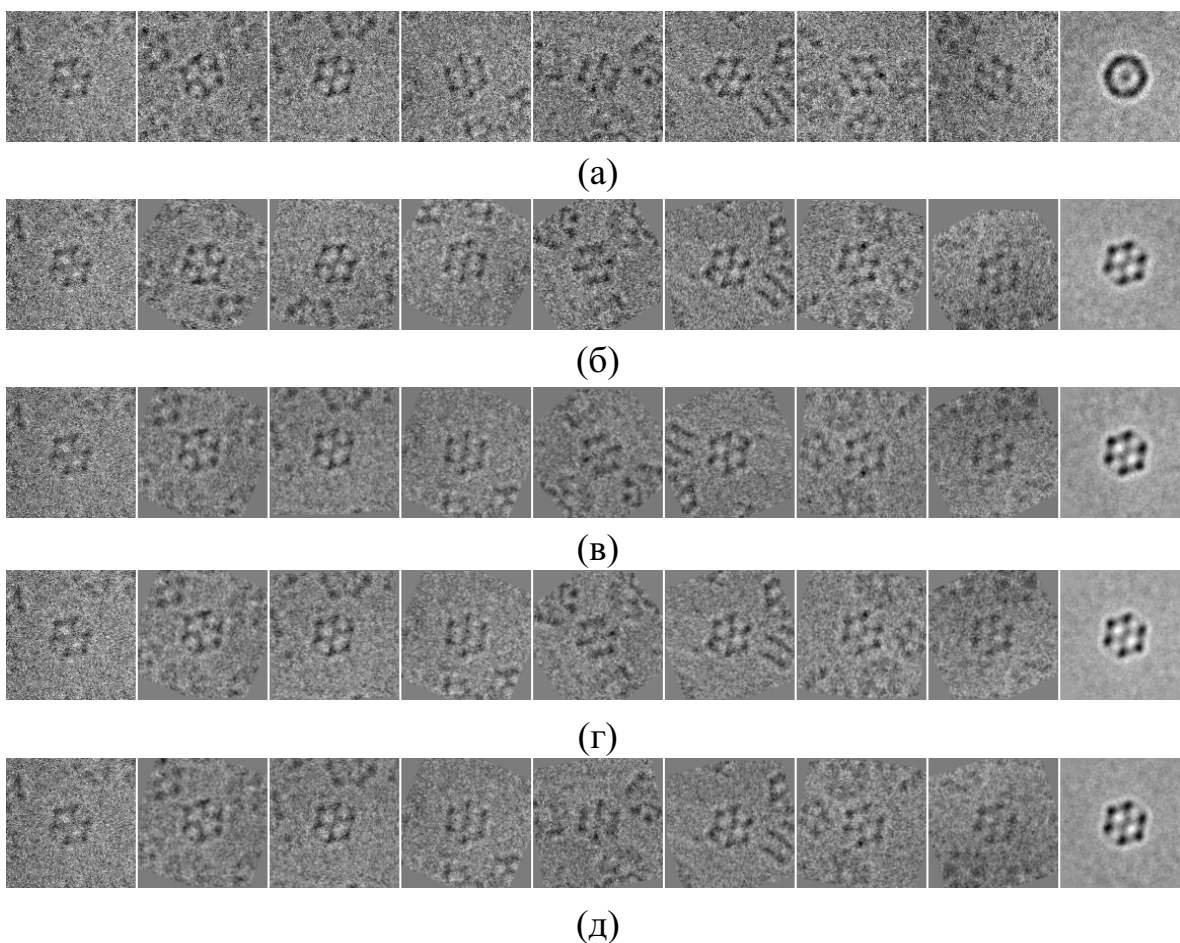


Рис. 6. Применение методов совмещения к реальному набору данных из базы EMDb. (а) Подмножество проекций из набора до совмещения. (б) Результат работы корреляционного метода [7]. (в) Результат работы базового метода на основе преобразования Фурье-Бесселя [11]. (г) Результат работы предложенного метода на основе преобразования Фурье-Бесселя и его вычисления с помощью проекционного метода. (д) Результат работы предложенного быстрого метода.

Анализируя результат усреднения проекций после выравнивания набора, можно сделать вывод, что результаты всех методов сравнимы и совмещение этими методами несомненно улучшает качество усредненного изображения и уточняет внутренние детали частицы. Относительно времени работы предложенный быстрый метод продемонстрировал наиболее высокую скорость вычислений в сравнении с методом [11] и предложенным проекционным методом. Сравнение времени работы корреляционного метода и предложенных подходов затруднительно провести в связи с использованием различных языков программирования и применения разных вычислительных библиотек для реализации методов.

В текущей реализации предложенные подходы выдают более точный результат, при этом корреляционный метод работает немного быстрее.

Заключение

В данной работе были предложены методы совмещения на основе преобразования Фурье-Бесселя и его вычисления с помощью проекционного метода вычисления преобразования Ганкеля, а также с помощью быстрого алгоритма вычислений проекционных коэффициентов на основе квадратурной формулы Гаусса-Лагерра. Предложенные подходы были разработаны для совмещения двумерных проекций частиц криоэлектронной микроскопии одиночных частиц. Предложенные методы являются улучшением метода [11] с точки зрения вычислительной сложности, при этом описанные подходы не уступают в качестве совмещения. Ускорение достигается при применении проекционного метода для вычисления преобразований Фурье-Бесселя, а также при дополнительном использовании быстрого алгоритма расчета проекционных коэффициентов. Нахождение искоемых параметров сводится к однократному вычислению трехмерной корреляции в Фурье-пространстве и вычислению обратного преобразования Фурье для перехода в координатное пространство. Результаты проведенных экспериментов на синтетических и реальных данных продемонстрировали эффективность разработанных подходов и их устойчивость к большому уровню шума, характерному для изображений крио-ЭМ.

Литература

1. A primer to single-particle cryo-electron microscopy / Yifan Cheng, Nikolaus Grigorieff, Pawel A Penczek, Thomas Walz // *Cell*. — 2015. — Vol. 161, no. 3. — Pp. 438–449.
2. Vilas Jose Luis, Carazo Jose Maria, Sorzano Carlos Oscar S. Emerging Themes in CryoEM- Single Particle Analysis Image Processing // *Chemical Reviews*. — 2022. — Vol. 122, no. 17. — Pp. 13915–13951.
3. Joyeux Laurent, Penczek Pawel A. Efficiency of 2D alignment methods // *Ultramicroscopy*. — 2002. — Vol. 92, no. 2. — Pp. 33–46.
4. Penczek Pawel, Radermacher Michael, Frank Joachim. Three-dimensional reconstruction of single particles embedded in ice // *Ultramicroscopy*. — 1992. — Vol. 40, no. 1. — Pp. 33–53.
5. Wilson Cyrus A, Theriot Julie A. A correlation-based approach to calculate rotation and translation of moving cells // *IEEE Transactions on Image Processing*. — 2006. — Vol. 15, no. 7. — Pp. 1939–1951.
6. Guizar-Sicairos Manuel, Thurman Samuel T, Fienup James R. Efficient subpixel image registration algorithms // *Optics letters*. — 2008. — Vol. 33, no. 2. — Pp. 156–158.
7. Anoshina Nadezhda A, Krylov Andrey S, Sorokin Dmitry V. Correlation-based 2D registration method for single particle cryo-EM images // 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA) / IEEE. — 2017. — Pp. 1–6.
8. Wang Xiangwen, Lu Yonggang, Liu Jiaxuan. A fast image alignment approach for 2D classification of cryo-EM images using spectral clustering // *Current Issues in Molecular Biology*. — 2021. — Vol. 43, no. 3. — Pp. 1652–1668.
9. Yang Zhengfan, Penczek Pawel A. Cryo-EM image alignment based on nonuniform fast Fourier transform // *Ultramicroscopy*. — 2008. — Vol. 108, no. 9. — Pp. 959–969.
10. Cong Yao, Kovacs Julio A, Wriggers Willy. 2D fast rotational matching for image processing of biophysical data // *Journal of structural biology*. — 2003. — Vol. 144, no. 1-2. — Pp. 51–60.
11. Kovacs Julio A, Abagyan Ruben, Yeager Mark. Fast bessell matching // *Journal of Computational and Theoretical Nanoscience*. — 2007. — Vol. 4, no. 1. — Pp. 84–95.
12. Zhao Zhizhen, Singer Amit. Rotationally invariant image representation for viewing direction classification in cryo-EM // *Journal of structural biology*. — 2014. — Vol. 186, no. 1. — Pp. 153–166.

13. Iterative stable alignment and clustering of 2D transmission electron microscope images / Zhengfan Yang, Jia Fang, Johnathan Chittuluru et al. // *Structure*. — 2012. — Vol. 20, no. 2. — Pp. 237–247.
14. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination / Ali Punjani, John L Rubinstein, David J Fleet, Marcus A Brubaker // *Nature methods*. — 2017. — Vol. 14, no. 3. — Pp. 290–296.
15. Scheres Sjoers HW. RELION: implementation of a Bayesian approach to cryo-EM structure determination // *Journal of structural biology*. — 2012. — Vol. 180, no. 3. — Pp. 519–530.
16. Grigorieff Nikolaus. FREALIGN: high-resolution refinement of single particle structures // *Journal of structural biology*. — 2007. — Vol. 157, no. 1. — Pp. 117–125.
17. DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM / Feng Wang, Huichao Gong, Gaochao Liu et al. // *Journal of structural biology*. — 2016. — Vol. 195, no. 3. — Pp. 325–336.
18. CryoGAN: A new reconstruction paradigm for single-particle cryo-EM via deep adversarial learning / Harshit Gupta, Michael T McCann, Laurene Donati, Michael Unser // *IEEE Transactions on Computational Imaging*. — 2021. — Vol. 7. — Pp. 759–774.
19. Enhancing the signal-to-noise ratio and generating contrast for cryo-EM images with convolutional neural networks / Eugene Palovcak, Daniel Asarnow, Melody G Campbell et al. // *IUCrJ*. — 2020. — Vol. 7, no. 6. — Pp. 1142–1150.
20. Vilenkin Naum IAkovlevich. Special functions and the theory of group representations. — American Mathematical Soc., 1978. — Vol. 22.
21. Sorokin DV, Krylov AS. Laguerre projection method for finite Hankel transform of arbitrary order // *Moscow University Computational Mathematics and Cybernetics*. — 2010. — Vol. 34. — Pp. 149–156.
22. Sorokin DV, Krylov AS. A projection local image descriptor // *Pattern Recognition and Image Analysis*. — 2012. — Vol. 22. — Pp. 380–385.
23. Крылов Владимир Иванович. Приближенное вычисление интегралов. — М.: Наука, 1967. — С. 116–147.
24. Aberth Oliver. Iteration methods for finding all zeros of a polynomial simultaneously // *Mathematics of computation*. — 1973. — Vol. 27, no. 122. — Pp. 339–344.

25. Сагиндыков Тамерлан Бехмухамбетович, Сорокин Дмитрий Васильевич, Аношина Надежда Алексеевна. Метод синтеза реалистичных 2d и 3d изображений крио-электронной микроскопии // *Программирование*. — 2018. — № 4. — С. 46–54.
26. Electron Microscopy Data Bank. — <https://www.ebi.ac.uk/emdb/>.
27. Rosenthal Peter B, Henderson Richard. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy // *Journal of Molecular Biology*. — 2003. — October. — Vol. 333, no. 4. — Pp. 721–745.