Formal uniqueness in Ewald sphere corrected single particle analysis

Thursday 8th April, 2021

Pär Kurlberg[*]   *Gustav Zickert*[†]

Abstract

In single particle analysis (SPA), the task is to recover the scattering potential of a macromolecular structure from cryo-electron microscope images of many copies of the structure in unknown orientations. The idealized, noise-free SPA inverse problem has been shown to be uniquely solvable — up to hand — when the forward model is based on the Ray-transform. More accurate forward models take the non-zero curvature of the Ewald sphere into account. We analyze an Ewald sphere corrected forward model for SPA and use the diffraction slice theorem to prove that the corresponding inverse problem is uniquely solvable, *including the hand of the structure.*

1   Introduction

A central problem in biology is to determine the structure of biological macromolecules. Single particle analysis (SPA) can be used to reveal the structure of molecules for which the traditional techniques of X-ray crystallography and nuclear magnetic resonance spectroscopy have proved hard to use [ACGS14, VBLVB17, Pep17].

SPA is an imaging technique in which a large number of ideally identical "particles" (e.g. protein molecules) in different relative orientations are imaged by a transmission electron microscope. The inverse problem of recovering the 3D structure of the particle from the resulting images is similar to other tomographic inverse problems such as X-ray computerized tomography, but what makes SPA rather special is that the projection directions are unknown.

Let SE(3) and SO(3) denote the special Euclidean and special orthogonal groups in three dimensions, respectively. We use the convention that SE(3) acts first by rotation and then by translation. More precisely, SE(3) and SO(3) act on a function $f: \mathbb{R}^3 \to \mathbb{R}$ via

$$R \cdot f(x) := f(R^{-1}(x)), \quad R \in \text{SO}(3) \tag{1}$$

$$(R, c) \cdot f(x) := f(R^{-1}(x - c)), \quad (R, c) \in \text{SE}(3) \tag{2}$$

If the center of mass of f is at the origin, which we will assume in this text, then c equals the center of mass of $(R, c) \cdot f$.

In the absence of noise, SPA data are often modeled as [CGPW15, BBS20]:

$$g_j = \text{PSF} * \mathcal{P}((R_j, c_j) \cdot f), \quad j \in J, \tag{3}$$

1

where $f\colon \mathbb{R}^3 \to \mathbb{R}_{\geq 0}$ represents the unknown particle, J is some index set, $g_j\colon \mathbb{R}^2 \to \mathbb{R}$ are the data, $(R_j, c_j) \in \mathrm{SE}(3)$ are unknown rigid body motions, $\mathcal{P}$ is a single-angle ray-transform:

$$\mathcal{P}(f)(x_1, x_2) := \int_{\mathbb{R}} f(x_1, x_2, x_3) dx_3, \tag{4}$$

and PSF, the point spread function, is a given convolution kernel.

Clearly, SPA data from (3) can not distinguish f from elements in the $\mathrm{SE}(3)$-orbit of f. In fact we may only ever hope to recover the $\mathrm{E}(3)$-orbit of f. In more practical terms, information of the *hand* of f (i.e. the orientation of f) is missing in data as in (3). To see why this is the case, let $O := \mathrm{diag}(1, 1, -1) \in \mathrm{O}(3)$, $R \in \mathrm{SO}(3)$ and $R_O = ORO \in \mathrm{SO}(3)$. Then

$$\mathcal{P}(R_O \cdot (O \cdot f)) = \mathcal{P}((R_O O) \cdot f) = \mathcal{P}((OR) \cdot f) = \mathcal{P}(O \cdot (R \cdot f)) = \mathcal{P}(R \cdot f), \tag{5}$$

where we used that $\mathcal{P}(O \cdot f) = \mathcal{P}(f)$. Hence, SPA data corresponding to f and $\{R_j\}_{j \in J}$, is identical to data corresponding to $O \cdot f$ and $\{(R_j)_O\}_{j \in J}$.

Apart from these obstructions to uniqueness, the SPA inverse problem of recovering the structure f from data given by (3) has been shown to be uniquely solvable using several different techniques. The method of common lines [VG86, VH87] is based on the fact that the Fourier transform of any two projection images will agree on at least one central line. For a generic triplet of projections this enables recovery of their mutual orientations (up to hand). In Kam's method [Kam80, BBSK+17] one computes correlations of the Fourier transform of the structure using averaged correlations of the Fourier transforms of the projection images. Uniqueness was studied from a probabilistic point of view in [Pan09], where the author proved that a pair of distinct objects induce mutually singular probability measures in the data space. Finally, there is the class of methods to which this paper belongs. It is based on relating moments of data to moments of the particle [Gon88, Sal90, BB00, LY07, Lam08]. While it has not been widely used in practical reconstructions, it is useful for rigorously analyzing uniqueness.

The aforementioned studies all assume that data are originating from line integrals as in (3). This *projection assumption* corresponds to the assumption of a flat Ewald sphere, which limits the reconstruction quality at high resolutions [Gla19]. More accurate image formation models that better account for the wave properties of the imaging electron therefore take the curvature of the Ewald sphere into account [DeR00, FÖ08, RH18, LYZJ10, WDG06, JK00, WCZ04, KKHC10, Voo14]. In frequency space, these models provide data on half-spheres, rather than on planes[1].

Prior work on structure recovery based on Ewald sphere corrected models has mainly focused on the development of algorithms. However, to the best of our knowledge, there does not exist any explicit precise statement of uniqueness in the literature.

In this paper we focus on the noise-free and continuous setting. (The latter means that we do not introduce any discretization of the detector). In this setting we provide a mathematically rigorous proof that accounting for a non-zero curvature of the Ewald sphere renders the SPA inverse problem *generically uniquely solvable.*

Our theorem treats a continuum limit where all $\mathrm{SO}(3)$ elements are realized in the data (this assumption can be relaxed: the proof is valid with minor modifications if the set of rotations in data is only assumed to be a *countable dense subset of SO(3)*). The theorem also assumes that all particles are located at the same distance from the detector, that the real- and imaginary parts are related by a non-zero constant amplitude contrast ratio, that f has compact support and that the first few moments of f satisfy some generically valid conditions (c.f. Assumption 3.3). However, our

---

[1] Some models [WCZ04, Voo14] yield Fourier space data on paraboloids rather than on spheres.

result is valid for any defocus and we assume neither that orientations are known a priori, nor that the particle belongs to a certain symmetry group, nor that multiple projections were recorded with varying defocus.

Since the intersection of two half-spheres is a circular segment, one could, perhaps, envision a uniqueness result based on finding the common curve associated to each pair of projection images. Orientation estimation based on such common curves has been considered in X-ray free electron laser imaging [BT11], see also [RG17] for a discussion of the common curve problem in the context of cryoEM.

The basic strategy here is to use the method of moments. Since the model we work with is more amenable for analysis in frequency space, we compute moments by differentiating the Fourier transform of data. A key element is to find the second order moments of f by looking at extremal values of the second order moments of data, a technique which was also used in [Sal90].

The paper is organized as follows. In section 2 we define our Ewald sphere corrected forward model. In section 3 we state and prove our uniqueness theorem. Finally, in section 4 we give a conclusion.

2   Forward model

We start out with the assumption that the unknown structure is described by a complex-valued and compactly supported scattering potential $f^{\mathbb{C}} \in L^2(\mathbb{R}^3; \mathbb{C})$ whose real- and imaginary part respectively determine the elastic- and inelastic scattering properties of the specimen. As mentioned in the introduction we assume that there exists a *known* constant $Q \in \mathbb{R}_{>0}$, called the amplitude contrast ratio, such that real- and imaginary parts f and f^{Im} of $f^{\mathbb{C}}$ are related via

$$f^{\mathrm{Im}} = Qf. \tag{6}$$

The incoming electron is a spatial monochromatic plane wave $u^{\mathrm{in}}(x) := e^{-ikx_3}$ with wave number $k > 0$ traveling in the direction of $-e_3$. Its interaction with the specimen is modeled by the Schrödinger equation. Stationary solutions to the latter are given as solutions to a Helmholtz equation. In biological applications the specimen is typically weakly scattering and in this case the Born approximation, which linearizes the dependence on the scattering potential of the (approximate) solution to the Helmholtz equation, is applicable [FÖ08, KL11]. Under the Born approximation, which we use in this paper, the interaction of u^{in} with $f^{\mathbb{C}}$ results in an outgoing wave u^{out} at the specimen exit plane $\Pi_{\mathrm{out}} := e_3^{\perp} \cong \mathbb{R}^2$. The scattering operator $\mathcal{T}^{\mathrm{sc}}$ maps $f^{\mathbb{C}}$ to the corresponding u^{out}:

$$\mathcal{T}^{\mathrm{sc}}\left(f^{\mathbb{C}}\right)(x) := u^{\mathrm{out}}(x) = 1 + k^{-1}\mathcal{U}\left(f^{\mathbb{C}}\right)(x), \quad x \in \mathbb{R}^2 \tag{7}$$

where $\mathcal{U}$ is the so-called propagation operator [NW01, FÖ08] that is defined as

$$\mathcal{U}\left(f^{\mathbb{C}}\right) := \frac{k}{u^{\mathrm{in}}} G_k * \left(u^{\mathrm{in}} f^{\mathbb{C}}\right)\Big|_{e_3^{\perp}}. \tag{8}$$

Here $*$ denotes convolution on $\mathbb{R}^3$ and

$$G_k(x) := \frac{1}{4\pi}\frac{e^{ik|x|}}{|x|}, \quad x \in \mathbb{R}^3 \setminus \{0\}. \tag{9}$$

The outgoing wave then passes through the optics of the electron microscope. The effect of the optics is here modelled by the convolution[2]

$$u^{\mathrm{out}} \mapsto h * u^{\mathrm{out}}, \tag{10}$$

_____

[2] Formally, we define convolution by point-wise multiplication in frequency space.
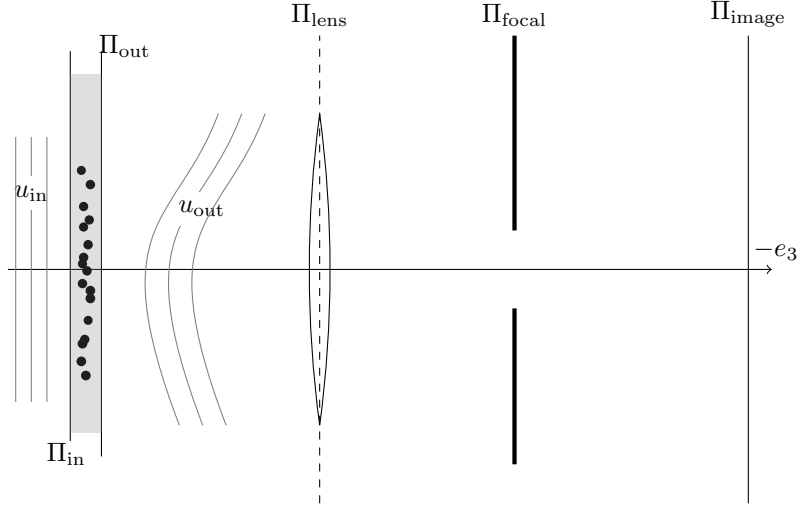
Fig. 1: Schematic representation of the SPA set-up. For the uniqueness theorem we assume that the relative shifts along the optical axis are small enough to be considered equal to zero.

where the point spread function h is defined by its Fourier transform:

$$\mathcal{F}[h](\xi) := A(\xi)e^{i\chi(\xi)}, \quad \chi(\xi) := a|\xi|^2 + b|\xi|^4. \tag{11}$$

In the above, the aperture function A is the indicator function of some ball and

$$a := \frac{\Delta z}{2k} \in \mathbb{R} \tag{12}$$

$$b := -\frac{C_s}{4k^3} \in \mathbb{R}, \tag{13}$$

where C_s is a constant that encodes spherical aberrations of the optics and Δz is the defocus. We have assumed perfectly coherent illumination for simplicity. Finally, the wave forms an intensity distribution in the detector plane. This leads to the following total non-linear model for the measured intensity:

$$\mathcal{I}_{\text{non-lin}}\left(f^{\mathbb{C}}\right)(x) := \left| h * \left\{ 1 + k^{-1}\mathcal{U}\left(f^{\mathbb{C}}\right) \right\}(x) \right|^2, x \in \mathbb{R}^2. \tag{14}$$

Next, we assume that the quadratic term is negligible and study instead the linearized model

$$\mathcal{I}_0(f^{\mathbb{C}}) := k^{-1}h * \mathcal{U}\left(f^{\mathbb{C}}\right) + k^{-1}\overline{h} * \overline{\mathcal{U}(f^{\mathbb{C}})}. \tag{15}$$

With assumption (6) in place we define our final intensity operator $\mathcal{I}$ that acts directly on a real-valued potential via

$$\mathcal{I}(f) := \mathcal{I}_0((1 + iQ)f) = k^{-1}(1 + iQ)h * \mathcal{U}(f) + k^{-1}\overline{(1 + iQ)h} * \overline{\mathcal{U}(f)}. \tag{16}$$

Remark 2.1. It can be shown [FÖ08, NW01] that, point-wise in frequency space, one has

$$2 \cdot \mathcal{U}(f) \xrightarrow{k \to \infty} i\mathcal{P}(f), \tag{17}$$

so for large k the Born approximation reproduces the classical model:

$$\mathcal{I}(f) \approx k^{-1}\text{Re}\left((i - Q)h\right) * \mathcal{P}(f) = -k^{-1}\text{PSF}_Q * \mathcal{P}(f), \tag{18}$$

$$\mathcal{F}[\text{PSF}_Q](\xi) := 1_{B(0,r)}(\xi)\left[Q\cos\chi(\xi) + \sin\chi(\xi)\right]. \tag{19}$$

So far we have defined the forward model corresponding to the electron microscope image of a single particle. In single particle analysis, the complete data is a collection

$$\{\mathcal{I}\left((R_j, c_j) \cdot f\right) \mid j \in J\} \tag{20}$$

of rotated and translated copies of f, where the R_j's and c_j's are unknown and J is some index set. A schematic drawing of the set-up is given in Fig. 1.

3   A uniqueness result

3.1   Some preparatory definitions, lemmas and assumptions

The moments of a function $f \colon \mathbb{R}^n \to \mathbb{C}$ are defined as

$$m_\alpha f := \int_{\mathbb{R}^n} x^\alpha f dx, \quad \alpha \in \mathbb{N}^n, \tag{21}$$

where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ is a multi-index, and $x^\alpha := \prod_i x_i^{\alpha_i}$. The order of a multi-index α is the natural number $|\alpha|$ defined by $|\alpha| := \sum_i \alpha_i$. We define the matrix

$$\Lambda(f) := \begin{pmatrix} m_{200}f & m_{110}f & m_{101}f \\ m_{110}f & m_{020}f & m_{011}f \\ m_{101}f & m_{011}f & m_{002}f \end{pmatrix}. \tag{22}$$

For $f \colon \mathbb{R}^n \to \mathbb{C}$ we use the following version of the Fourier transform:

$$\mathcal{F}(f)(\xi) := \hat{f}(\xi) := \int_{\mathbb{R}^n} f(x) e^{-i\xi \cdot x} dx. \tag{23}$$

In order to carry out our analysis in frequency space, we need the following two basic lemmas.

Lemma 3.1.

$$(-i)^{|\alpha|} m_\alpha f = \partial^\alpha \hat{f}(0), \quad \alpha \in \mathbb{N}^n. \tag{24}$$

Lemma 3.2.

$$\mathcal{F}\left[(R, c) \cdot f\right](\xi) = e^{-ic \cdot \xi} \left(R \cdot \hat{f}\right)(\xi), \quad (R, c) \in SE(3). \tag{25}$$

The space of functions that may serve as real-valued scattering potentials is

$$X_0 := \left\{f \in L^2(\mathbb{R}^3) \mid f \text{ has compact support and } f \geq 0\right\}. \tag{26}$$

We will require the following mild generic asymmetry assumptions on the scattering potential.

Assumption 3.3.

1. $\Lambda(f)$ has three distinct (real) eigenvalues.

2. In a coordinate-system where $\Lambda(f)$ is diagonal, neither of the third order moments m_{300} and m_{210} vanish.

Finally, let

$$X := \{f \in X_0 \mid f \text{ satisfies Assumption 3.3}\}. \tag{27}$$

3.2   Theorem statement and proof

Theorem 3.4. *Consider the data*

$$y := \{ \mathcal{I}\left((R_j, c_j) \cdot f\right) \mid (R_j, c_j) \in SE(3), j \in J \}, \qquad (28)$$

for some index set J. If $\{R_j \mid j \in J\} = SO(3)$ and there exits a known positive constant c_0 such that $\forall j \colon c_j \cdot e_3 = c_0$, then y determines $f \in X$ uniquely up to a rigid body motion.

Remark 3.5. As already mentioned, a minor modification of the argument shows that the conclusion holds for any countable set J such that $\{R_j \mid j \in J\}$ is dense in $SO(3)$. We leave the details to the interested reader.

Proof. To fix the position and pose of f, we assume that

1. f has center of mass in the origin, which implies that $\nabla \widehat{f}(0) = 0$.

2. $\Lambda(f) = \mathrm{diag}(m_{200}, m_{020}, m_{002})$, with $m_{002} < m_{020} < m_{200}$.

3. $m_{300}, m_{210} > 0$.

Since f is compactly supported, its Fourier transform is analytic and therefore uniquely determined by its Taylor series coefficients. The basic idea of the proof is to show that (up to a rigid body motion) these coefficients can be uniquely recovered from y via a system of polynomial equations whose unknowns are given by the coefficients along with the particle translations and rotations. The uniqueness of the solution to the system of equations is established by induction over the order of the Taylor coefficients. Once a coefficient order is fixed, we are able to reduce our system to the one that corresponds to the Ray transform, and the latter is known to have a unique solution. Unless otherwise stated, in the proof below, "... is known" should be interpreted as "... is uniquely determined by y up to a rigid body motion".

Step 1: Fourier transform data and apply the diffraction slice theorem

For ξ small enough so that the aperture function satisfies $A(\xi) = 1$, the Fourier transform of (15) is given by

$$\mathcal{F}\left[\mathcal{I}_0(f^{\mathbb{C}}) \right] = e^{i\chi} \mathcal{F}\left[\mathcal{U}\left(f^{\mathbb{C}} \right) \right] + e^{-i\chi} \mathcal{F}\left[\overline{\mathcal{U}(f^{\mathbb{C}})} \right]. \qquad (29)$$

Assuming a constant amplitude contrast ratio (i.e. $f^{\mathbb{C}} = (1 + iQ)f$), (29) can be expressed as

$$\omega_1 \mathcal{F}\left[\mathcal{U}\left(f \right) \right] + \overline{\omega_1} \mathcal{F}\left[\overline{\mathcal{U}(f)} \right], \qquad (30)$$

where $\omega_1 := (1 + iQ)e^{i\chi}$. We further rewrite this using the identity $\overline{\mathcal{F}\left[h \right]}(\xi) = \mathcal{F}\left[\bar{h} \right](-\xi)$ as

$$\omega_1 \mathcal{F}\left[\mathcal{U}\left(f \right) \right] + \overline{\omega_1} \overline{\mathcal{F}\left[\mathcal{U}(f) \right]}(-\cdot). \qquad (31)$$

The diffraction slice theorem [NW01] provides us with the following expression for $\mathcal{F}\left[\mathcal{U}(f) \right]$:

$$\mathcal{F}\left[\mathcal{U}(f) \right](\xi) = \phi(\xi) \widehat{f}\left(\gamma^+(\xi) \right), \qquad (32)$$

where

$$\phi(\xi) := \frac{i}{2} \frac{k}{k - \gamma_3(\xi)} \qquad (33)$$

and γ^+ is the map that lifts the plane to a half-sphere:

$$\gamma^+(\xi) := (\xi_1, \xi_2, \gamma_3(\xi)) \tag{34}$$

$$\gamma_3(\xi) := k - \sqrt{k^2 - |\xi|^2} \tag{35}$$

Next, again using the identity $\overline{\mathcal{F}[h]}(\xi) = \mathcal{F}\left[\bar{h}\right](-\xi)$, we compute:

$$\overline{\mathcal{F}[\mathcal{U}(f)]}(-\xi) = \bar{\phi}(-\xi)\overline{\widehat{f}(\gamma^+(-\xi))} = \bar{\phi}(\xi)\widehat{f}\left(-\gamma^+(-\xi)\right) = \bar{\phi}(\xi)\widehat{f}\left(\gamma^-(\xi)\right), \tag{36}$$

where γ^- maps the plane to a half-sphere that extends downwards. More precisely, it is defined as

$$\gamma^-(\xi) := (\xi_1, \xi_2, -\gamma_3(\xi)). \tag{37}$$

Hence, from (31) it follows that

$$\mathcal{F}\left[\mathcal{I}_0\left(f^{\mathbb{C}}\right)\right] = \omega_2\widehat{f}\circ\gamma^+ + \overline{\omega_2}\widehat{f}\circ\gamma^-, \tag{38}$$

where

$$\omega_2 := -\frac{1}{k}\omega_1\phi = \frac{1}{2}(Q-i)e^{i\chi}\frac{1}{k-\gamma_3}. \tag{39}$$

Hence we obtain the Fourier space model

$$2\left(k-\gamma_3\right)\mathcal{F}\left[\mathcal{I}_0\left(f^{\mathbb{C}}\right)\right] = ze^{i\chi}\widehat{f}\circ\gamma^+ + \bar{z}e^{-i\chi}\widehat{f}\circ\gamma^-, \tag{40}$$

where

$$z := (Q-i). \tag{41}$$

Step 2: Handle in-plane translations

For $(R, c) \in \mathrm{SE}(3)$ we introduce the notation

$$h_{R,c}^{(1)}(\xi) := 2\left(k-\gamma_3\right)\mathcal{F}\left[\mathcal{I}((R,c)\cdot f)\right](\xi) \tag{42}$$

$$= ze^{i\chi}\mathcal{F}\left[(R,c)\cdot f\right]\circ\gamma^+ + \bar{z}e^{-i\chi}\mathcal{F}\left[(R,c)\cdot f\right]\circ\gamma^- \tag{43}$$

$$= ze^{i\chi}e^{-ic\cdot\gamma^+}\left(R\cdot\widehat{f}\right)\circ\gamma^+ + \bar{z}e^{-i\chi}e^{-ic\cdot\gamma^-}\left(R\cdot\widehat{f}\right)\circ\gamma^- \tag{44}$$

$$= e^{-i(c_1,c_2)\cdot\xi}\left[ze^{i\chi}e^{-ic_0\gamma_3}\left(R\cdot\widehat{f}\right)\circ\gamma^+ + \bar{z}e^{-i\chi}e^{+ic_0\gamma_3}\left(R\cdot\widehat{f}\right)\circ\gamma^-\right]. \tag{45}$$

(To simplify the notation we suppress the dependence on ξ and simply write $\gamma^\pm$, γ_3, ξ and $\widehat{f}$.)

Set $a_\alpha := \partial^\alpha\widehat{f}(0)/\alpha!$. Since f has compact support, it follows from the Paley-Wiener theorem (see e.g. [Hör03, Theorem 7.1.14]) that $\widehat{f}$ can be extended to an entire function on $\mathbb{C}^n$, and in particular the following Taylor expansions in $\xi \in \mathbb{R}^2$ are well-defined.

$$\gamma_3 = \frac{1}{2k}|\xi|^2 + \mathcal{O}\left(|\xi|^4\right) \tag{46}$$

$$e^{i(\pm\chi(\xi)\mp c_0\gamma_3)} = 1 \pm i\left(a - \frac{c_0}{2k}\right)|\xi|^2 + \mathcal{O}\left(|\xi|^4\right) \tag{47}$$

$$\widehat{f} = \widehat{f}(0) + \sum_{|\alpha|=2,3} a_\alpha\xi^\alpha + \mathcal{O}\left(|\xi|^4\right) \tag{48}$$

Hence (45) implies that

$$h_{R,c}^{(1)}(\xi) = (z + \overline{z})\widehat{f}(0) - i(z + \overline{z})\widehat{f}(0)(c_1, c_2) \cdot \xi + \mathcal{O}\left(|\xi|^2\right). \tag{49}$$

We note that since Q is assumed to be known, $\widehat{f}(0)$ is known from the zeroth order term of (49). Therefore, (c_1, c_2) can be read of from the first order term of (49). Thus $((c_j)_1, (c_j)_2)_{j \in J}$ is known, hence the set

$$H_1 := \left\{ h_R^{(2)}(\xi) \mid R \in \text{SO}(3) \right\} \tag{50}$$

is known, where

$$h_R^{(2)}(\xi) := z e^{i\chi} e^{-ic_0\gamma_3} \left(R \cdot \widehat{f} \right) \circ \gamma^+ + \overline{z} e^{-i\chi} e^{+ic_0\gamma_3} \left(R \cdot \widehat{f} \right) \circ \gamma^-. \tag{51}$$

Step 3: Extract a one-parameter subset of data

Note that

$$h_R^{(2)}(\xi) = z \left(1 + i \left(a - \frac{c_0}{2k} \right) |\xi|^2 \right) \left(\widehat{f}(0) + \sum_{|\alpha|=2,3} a_\alpha (R^{-1}\gamma^+(\xi))^\alpha \right) \tag{52}$$

$$+ \overline{z} \left(1 - i \left(a - \frac{c_0}{2k} \right) |\xi|^2 \right) \left(\widehat{f}(0) + \sum_{|\alpha|=2,3} a_\alpha (R^{-1}\gamma^-(\xi))^\alpha \right) + \tag{53}$$

$$+ \mathcal{O}\left(|\xi|^4\right) \tag{54}$$

$$= (z + \overline{z})\widehat{f}(0) + C|\xi|^2 \tag{55}$$

$$+ z \left(a_{200}(R^{-1}\gamma^+(\xi))_1^2 + a_{020}(R^{-1}\gamma^+(\xi))_2^2 + a_{002}(R^{-1}\gamma^+(\xi))_3^2 \right) \tag{56}$$

$$+ \overline{z} \left(a_{200}(R^{-1}\gamma^-(\xi))_1^2 + a_{020}(R^{-1}\gamma^-(\xi))_2^2 + a_{002}(R^{-1}\gamma^-(\xi))_3^2 \right) \tag{57}$$

$$+ z \sum_{|\alpha|=3} a_\alpha (R^{-1}\gamma^+(\xi))^\alpha + \overline{z} \sum_{|\alpha|=3} a_\alpha (R^{-1}\gamma^-(\xi))^\alpha + \mathcal{O}\left(|\xi|^4\right), \tag{58}$$

where $C := 2\widehat{f}(0)\text{Re}\left(zi\left(a - \frac{c_0}{2k} \right) \right)$

Since $\widehat{f}(0)$ is known, C is known. Now if

$$R^{-1} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix}, \tag{59}$$

then

$$\left[R^{-1}\gamma^\pm(\xi) \right]_j^2 = \left(R_{j1}\xi_1 + R_{j2}\xi_2 \pm R_{j3}\frac{1}{2k}(\xi_1^2 + \xi_2^2) \right)^2 + \mathcal{O}(|\xi|^3) \tag{60}$$

$$= R_{j1}^2\xi_1^2 + 2R_{j1}R_{j2}\xi_1\xi_2 + R_{j2}^2\xi_2^2 + \mathcal{O}(|\xi|^3), \tag{61}$$

and therefore the ξ_1^2-term in the expansion of $h_R^{(2)}$ is

$$(z + \overline{z})\left(a_{200}R_{11}^2 + a_{020}R_{21}^2 + a_{002}R_{31}^2 \right) + C. \tag{62}$$

Note that Lemma 3.1 and $m_{200}f > m_{020}f > m_{002}f$ together imply that $a_{200} < a_{020} < a_{002}$, so the ξ_1^2-term has a minimal value of $(z + \overline{z})a_{200} + C$, which is achieved exactly when $R_{11} = \pm 1, R_{21} = R_{31} = 0$. The maximal value is $(z + \overline{z})a_{002} + C$. Minimizing the ξ_1^2-term thus forces R^{-1} to have the form

$$R^{-1} = R_{S_1,\theta}^{-1} := \begin{pmatrix} S_1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \tag{63}$$

for some $\theta \in [0, 2\pi)$ and $S_1 \in \{-1, 1\}$.

This means that the set

$$H_2 := \left\{ h_{R_{S_1}, \theta}^{(2)}(\xi) \mid \theta \in [0, 2\pi), S_1 \in \{-1, 1\} \right\} \tag{64}$$

is known.

Now we take a closer look at the third-order expansion of $h_{R_{S_1}, \theta}^{(2)}$:

$$h_{R_{S_1}, \theta}^{(2)}(\xi) = (z + \bar{z})\widehat{f}(0) + ((z + \bar{z})a_{200} + C) \xi_1^2 \tag{65}$$

$$+ \left((z + \bar{z})a_{020} \cos^2(\theta) + (z + \bar{z})a_{002} \sin^2(\theta) + C\right) \xi_2^2 \tag{66}$$

$$+ \Big((z + \bar{z})a_{210} \cos(\theta) + (z + \bar{z})a_{201} \sin(\theta) \tag{67}$$

$$+ \frac{z - \bar{z}}{k}(a_{002} - a_{020}) \cos(\theta) \sin(\theta) \Big) \xi_1^2 \xi_2 \tag{68}$$

$$+ (z + \bar{z})S_1 a_{300}\xi_1^3 + \beta_1(S_1, \theta)\xi_1\xi_2^2 + \beta_2(S_1, \theta)\xi_2^3 + \mathcal{O}\left(|\xi|^4\right), \tag{69}$$

for some functions $\beta_1, \beta_2 : \{-1, 1\} \times [0, 2\pi) \to i\mathbb{R}$. Observe that the sign of the imaginary part of the ξ_1^3-coefficient equals S_1. Hence, with $R_\theta := R_{1, \theta}$, the following set is known:

$$H_3 := \left\{ h_{R_\theta}^{(2)}(\xi) \mid \theta \in [0, 2\pi) \right\}. \tag{70}$$

Step 4: Compute the second order moments of f

In what follows, unless otherwise is stated, "coefficient" refers to a Taylor coefficient of $h_{R_\theta}^{(2)}$. Note first that a_{200} is determined by the ξ_1^2 coefficient. Likewise, a_{020} and a_{002} are known, since the minimal and maximal values (recall that the ξ_1^2 coefficient is minimal) of the ξ_2^2 coefficients equals $a_{020} + C$ and $a_{002} + C$, respectively (note that C was determined in step 3).

Step 5: Compute some third order moments of f

The ξ_2^2-coefficient is minimized exactly when $\cos(\theta) = \pm 1$ and $\sin(\theta) = 0$. Hence, an examination of the $\xi_1^2\xi_2$-coefficient shows that $|a_{210}|$ is known. Thus a_{210} is known, since the sign of a_{210} was fixed in the beginning of this proof. Similarly, by maximizing the ξ_2^2-coefficient, we can determine $|a_{201}|$.

Step 6: Recover the hand of f along with some rotations

The ξ_2^2-coefficient determines $\cos^2(\theta)$, hence each θ is determined up to a most four possibilities. More precisely, each θ is determined up to sign and a shift by π.

We will now see that all small θ may be determined. There exists a known positive constant $\epsilon = \epsilon(z, a_{020}, a_{002}, a_{210}, |a_{201}|) < \pi/2$ such that if $\min\{|\theta|, |\theta - \pi|\} < \epsilon$, then the sign of the imaginary part of the $\xi_1^2\xi_2$-coefficient[3] equals the sign of its pure cosine term. Let

$$\sigma(\theta) := (z + \bar{z})a_{020} \cos^2(\theta) + (z + \bar{z})a_{002} \sin^2(\theta) + C \tag{71}$$

and for $0 < \theta < \epsilon$ consider the known set

$$y_1(\theta) := \Big\{ h \in H_3 \ \Big| \ \text{The } \xi_2^2\text{-coefficent of } h \text{ equals } \sigma(\theta) \text{ and the} \tag{72}$$

$$\text{imag. part of the } \xi_1^2\xi_2\text{-coefficent of } h \text{ is negative} \Big\}. \tag{73}$$

---

[3] Note that the $\xi_1^2\xi_2$-coefficient is purely imaginary.

Recall that $m_{210} > 0$ by assumption 3 in the beginning of this proof. It then follows from Lemma 3.1 and the discussion above that for any φ and any $\theta \in (0, \epsilon)$:

$$h_{R_\varphi}^{(2)} \in y_1(\theta) \Rightarrow \cos(\varphi) > 0, \tag{74}$$

so we have that

$$y_1(\theta) = \left\{ h_{R_\theta}^{(2)}, h_{R_{-\theta}}^{(2)} \right\}, 0 < \theta < \epsilon. \tag{75}$$

To remove the final sign-ambiguity, we study the $\xi_1^2 \xi_2$-coefficient. Note that $|\sin(\theta)| = \sin(|\theta|)$ is known for $|\theta| < \epsilon$, since $|\theta|$ is known. We fix some $\theta_0 \in (0, \epsilon)$ and further fix some $h_0 \in y_1(\theta_0)$. Let h_{21} denote the $\xi_1^2 \xi_2$-coefficient of h_0 and let

$$D := \frac{|h_{21} - \cos(\theta_0)(z + \bar{z})a_{210}|}{|\sin(\theta_0)|}. \tag{76}$$

Note that

$$D = |A + B|, \tag{77}$$

$$A = (z + \bar{z})a_{201}, \tag{78}$$

$$B = \frac{z - \bar{z}}{k}(a_{002} - a_{020})\cos(\theta_0). \tag{79}$$

In the above $|A|, B$ and D are known and $|A|, B \neq 0$[4]. Thus A is known, so a_{201} is known.

For $\theta \in (0, \epsilon)$ and $h \in y_1(\theta)$, let $h_{21}(\theta)$ denote the $\xi_1^2 \xi_2$-coefficient of h. Then

$$h_{21}(\theta) - \cos(\theta)(z + \bar{z})a_{210} = \left(\pm \sin(\theta)\right) \left((z + \bar{z})a_{201} + \frac{z - \bar{z}}{k}(a_{002} - a_{020})\cos(\theta)\right). \tag{80}$$

Hence $\sin(\theta)$ is known for $0 < \theta < \epsilon$ (the right factor in the RHS of (80) might vanish for some θ: in this case we shrink ϵ accordingly.). Since $\cos(\theta)$ was already known, θ is known. More precisely, for $\theta \in (0, \epsilon)$ we now know

$$h_{R_\theta}^{(2)}(\xi) = z e^{i\chi} e^{-ic_0\gamma_3} \left(R_\theta \cdot \widehat{f}\right) \circ \gamma^+ + \bar{z} e^{-i\chi} e^{+ic_0\gamma_3} \left(R_\theta \cdot \widehat{f}\right) \circ \gamma^-. \tag{81}$$

Step 7: Recover all moments of f

Since

$$\left[R_\theta^{-1}\gamma^\pm(\xi)\right]_1^i \left[R_\theta^{-1}\gamma^\pm(\xi)\right]_2^j \left[R_\theta^{-1}\gamma^\pm(\xi)\right]_3^k \tag{82}$$

equals

$$\xi_1^i \left(\cos(\theta)\xi_2 + \mathcal{O}\left(|\xi|^2\right)\right)^j \left(\sin(\theta)\xi_2 + \mathcal{O}\left(|\xi|^2\right)\right)^k \tag{83}$$

$$= \xi_1^i \cos^j(\theta) \sin^k(\theta)\xi_2^{j+k} + \mathcal{O}\left(|\xi|^{i+j+k+1}\right), \tag{84}$$

it follows that the Taylor series for $h_{R_\theta}^{(2)}$ admits the decomposition

$$(z + \bar{z})^{-1} h_{R_\theta}^{(2)} = \sum_{i,j,k} a_{ijk}\xi_1^i \cos^j(\theta) \sin^k(\theta)\xi_2^{j+k} + \sum_{r,s} b_{rs}(\theta)\xi_1^r \xi_2^s, \tag{85}$$

where a_{ijk} denotes the ijk:th Taylor coefficient of $\widehat{f}$ and b_{rs} only depends on a_{ijk} for i, j, k such that $i + j + k < r + s$.

---

[4] Recall that $Q > 0$ by assumption, so $z + \bar{z} \neq 0$.

Hence, if c_{ij} is defined as the $\xi_1^i \xi_2^j$-coefficient in the known Taylor series of $(z + \bar{z})^{-1} h_{R_\theta}^{(2)}$, then

$$\sum_{k=0}^{j} a_{ik(j-k)} \cos^k(\theta) \sin^{j-k}(\theta) = c_{ij} - b_{ij}(\theta), \tag{86}$$

where $b_{ij}(\theta)$ depends only on a_{rst} for r, s, t such that $r + s + t < i + j$.

Fix a moment-order $m \in \mathbb{N}$ and assume that a_{ijk} has been determined for all i, j, k such that $i + j + k < m$. Moreover fix i and j such that $i + j = m$ and let $\theta_1, \ldots, \theta_{j+1} \in (0, \epsilon)$ be distinct angles. We consider the linear system of $j + 1$ equations in the $j + 1$ variables $\left(a_{ik(j-k)}\right)_{k=0}^{j}$ that results from inserting those angles in (86):

$$\sum_{k=0}^{j} a_{ik(j-k)} \cos^k(\theta_\ell) \sin^{j-k}(\theta_\ell) = c_{ij} - b_{ij}(\theta_\ell), \quad \ell = 1, \ldots, j+1. \tag{87}$$

The system of equations (87) is known (c.f. [MKW96]) to be uniquely solvable for the variables $\left(a_{ik(j-k)}\right)_{k=0}^{j}$.[5]

Consequently, a_{ijk} is known for all i, j, k such that $i + j + k = m$, and by induction a_{ijk} is known for any $i, j, k \in \mathbb{N}$. Since $\hat{f}$ can be extended to an entire function, it is uniquely determined by its Taylor series coefficients $(a_{ijk})_{i,j,k=0}^{\infty}$. Hence $\hat{f}$ is known, which implies that f is known. This concludes the proof.

$\square$

4   Conclusion

Existing uniqueness results for SPA applies to a forward model that is based on the ray-transform. A more accurate diffraction tomographic model that accounts for the curvature of the Ewald sphere is offered in the Born approximation framework. Although models based on the ray-transform have been used successfully for SPA structure recovery, the attainable resolution is ultimately limited by the incorrect assumption of a flat Ewald sphere.

We extend the classical method of moments, which is based on relating moments of data to moments of the unknown 3D structure, from the ray-transform model to a model based on the Born approximation. Through the diffraction slice theorem this allows us to prove that the SPA inverse problem based on an Ewald sphere corrected model is uniquely solvable, including the hand of the structure.

We believe that several sharpenings and extensions of the theorem are possible, e.g. it seems likely that some of the assumptions (e.g. on the set of needed rotations and translations) can be relaxed.

Acknowledgements

References

[ACGS14]   David Agard, Yifan Cheng, Robert M Glaeser, and Sriram Subramaniam. Single-particle cryo-electron microscopy (cryo-em): progress, challenges, and perspectives for further improvement. In *Advances in imaging and electron physics*, volume 185, pages 113–137. Elsevier, 2014.

---

[5] Note that the right hand side of (87) is known by assumption.

[BB00]   Samit Basu and Yoram Bresler. Uniqueness of tomography with unknown view angles. *IEEE Transactions on image processing*, 2000.

[BBS20]   Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Processing Magazine*, 37(2):58–76, 2020.

[BBSK⁺17] Afonso S Bandeira, Ben Blum-Smith, Joe Kileel, Amelia Perry, Jonathan Weed, and Alexander S Wein. Estimation under group actions: recovering orbits from invariants. *arXiv preprint arXiv:1712.10163*, 2017.

[BT11]   Gábor Bortel and Miklós Tegze. Common arc method for diffraction pattern orientation. *Acta Crystallographica Section A: Foundations of Crystallography*, 67(6):533–543, 2011.

[CGPW15]   Yifan Cheng, Nikolaus Grigorieff, Pawel A. Penczek, and Thomas Walz. A Primer to Single-Particle Cryo-Electron Microscopy. *Cell*, 161:438–499, 2015.

[DeR00]   David J DeRosier. Correction of high-resolution data for curvature of the ewald sphere. *Ultramicroscopy*, 81(2):83–98, 2000.

[FÖ08]   Duccio Fanelli and Ozan Öktem. Electron tomography: A short overview with an emphasis on the absorption potential model for the forward problem. *Inverse Problems*, 2008.

[Gla19]   Robert M Glaeser. How good can single-particle cryo-em become? what remains before it approaches its physical limits? *Annual review of biophysics*, 48:45–61, 2019.

[Gon88]   AB Goncharov. Integral geometry and three-dimensional reconstruction of randomly oriented identical particles from their electron microphotos. *Acta Applicandae Mathematica*, 11(3):199–211, 1988.

[Hör03]   Lars Hörmander. *The Analysis of Linear Partial Differential Operators I*. Springer, 2003.

[JK00]   Grant J Jensen and Roger D Kornberg. Defocus-gradient corrected backprojection. *Ultramicroscopy*, 84(1-2):57–64, 2000.

[Kam80]   Zvi Kam. The reconstruction of structure from electron micrographs of randomly oriented particles. In *Electron Microscopy at Molecular Dimensions*, pages 270–277. Springer, 1980.

[KKHC10]   Ivan G Kazantsev, Joanna Klukowska, Gabor T Herman, and Laslo Cernetic. Fully three-dimensional defocus-gradient corrected backprojection in cryoelectron microscopy. *Ultramicroscopy*, 110(9):1128–1142, 2010.

[KL11]   H Kohr and AK Louis. Fast and high-quality reconstruction in electron tomography based on an enhanced linear forward model. *Inverse Problems*, 27(4):045008, 2011.

[Lam08]   Lars Lamberg. Unique recovery of unknown projection orientations in three-dimensional tomography. *Inverse Problems and Imaging*, 2(4):547–575, 2008.

[LY07]   Lars Lamberg and Lauri Ylinen. Two-dimensional tomography with unknown view angles. *Inverse problems and imaging*, 2007.

[LYZJ10]   Peter A Leong, Xuekui Yu, Z Hong Zhou, and Grant J Jensen. Correcting for the ewald sphere in high-resolution single-particle reconstructions. In *Methods in enzymology*, volume 482, pages 369–380. Elsevier, 2010.

[MKW96]   Peyman Milanfar, William Clement Karl, and Alan S Willsky. A moment-based variational approach to tomographic reconstruction. *IEEE Transactions on Image Processing*, 5(3):459–470, 1996.

[NW01]   Frank Natterer and Frank Wübbeling. *Mathematical Methods in Image Reconstruction*. Society for Industrial and Applied Mathematics, 2001.

[Pan09]   Victor M Panaretos. On random tomography with unobservable projection angles. *The Annals of Statistics*, 37(6A):3272–3306, 2009.

[Pep17]   Mark Peplow. Cryo-electron microscopy makes waves in pharma labs, 2017.

[RG17]   Elias Riedel Gårding. Geometric algebra, conformal geometry and the common curves problem, 2017.

[RH18]   Christopher J Russo and Richard Henderson. Ewald sphere correction using a single side-band image processing algorithm. *Ultramicroscopy*, 187:26–33, 2018.

[Sal90]   David B Salzman. A method of general moments for orienting 2d projections of unknown 3d objects. *Computer vision, graphics, and image processing*, 50(2):129–156, 1990.

[VBLVB17] Catherine Vénien-Bryan, Zhuolun Li, Laurent Vuillard, and Jean Albert Boutin. Cryo-electron microscopy and x-ray crystallography: complementary approaches to structural biology and drug discovery. *Acta Crystallographica Section F: Structural Biology Communications*, 73(4):174–183, 2017.

[VG86]   B.K. Vainshtein and A.B. Goncharov. Determination of the spatial orientation of arbitrarily arranged identical particles of unknown structure from their projections. In *Dokl. Akad. Nauk SSSR*, volume 287, pages 1131–1134, 1986.

[VH87]   Marin Van Heel. Angular reconstitution: a posteriori assignment of projection directions for 3d reconstruction. *Ultramicroscopy*, 21(2):111–123, 1987.

[Voo14]   LM Voortman. Ctf correction in cryo-electron tomography. 2014.

[WCZ04]   Yi Wan, Wah Chiu, and Z Hong Zhou. Full contrast transfer function correction in 3d cryo-em reconstruction. In *2004 International Conference on Communications, Circuits and Systems (IEEE Cat. No. 04EX914)*, volume 2, pages 960–964. IEEE, 2004.

[WDG06]   Matthias Wolf, David J DeRosier, and Nikolaus Grigorieff. Ewald sphere correction for single-particle electron microscopy. *Ultramicroscopy*, 106(4-5):376–382, 2006.