

Enhancing scattering images for orientation recovery with diffusion map

Martin Winter,¹ Ulf Saalmann,^{1,*} and Jan M. Rost^{1,2}

¹Max-Planck-Institut für Physik komplexer Systeme, Nöthnitzer Str. 38, 01187 Dresden, Germany

²PULSE Institute, Stanford University and SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA

*us@pks.mpg.de

Abstract: We explore the possibility for orientation recovery in single-molecule coherent diffractive imaging with diffusion map. This algorithm approximates the Laplace-Beltrami operator, which we diagonalize with a metric that corresponds to the mapping of Euler angles onto scattering images. While suitable for images of objects with specific properties we show why this approach fails for realistic molecules. We introduce a modification of the form factor in the scattering images which facilitates the orientation recovery and should be suitable for all recovery algorithms based on the distance of individual images.

© 2016 Optical Society of America

OCIS codes: (100.4995) Pattern recognition, metrics; (290.5825) Scattering theory; (340.7440) X-ray imaging.

References and links

1. H. N. Chapman, A. Barty, M. J. Bogan, S. Boutet, M. Frank, S. P. Hau-Riege, S. Marchesini, B. W. Woods, S. Bajt, W. H. Benner, R. A. London, E. Plönjes, M. Kuhlmann, R. Treusch, S. Dusterer, T. Tschentscher, J. R. Schneider, E. Spiller, T. Möller, C. Bostedt, M. Hoener, D. A. Shapiro, K. O. Hodgson, D. van der Spoel, F. Burmeister, M. Bergh, C. Caleman, G. Huldt, M. M. Seibert, F. R. N. C. Maia, R. W. Lee, A. Szöke, N. Timneanu, and J. Hajdu, “Femtosecond diffractive imaging with a soft X-ray free-electron laser,” *Nat. Phys.* **2**, 839–843 (2006).
2. T. R. M. Barends, L. Foucar, S. Botha, R. B. Doak, R. L. Shoeman, K. Nass, J. E. Koglin, G. J. Williams, S. Boutet, M. Messerschmidt, and I. Schlichting, “De novo protein crystal structure determination from X-ray free-electron laser data,” *Nature* **505**, 244–247 (2014).
3. R. A. Kirian, T. A. White, J. M. Holton, H. N. Chapman, P. Fromme, A. Barty, L. Lomb, A. Aquila, F. R. N. C. Maia, A. V. Martin, R. Fromme, X. Wang, M. S. Hunter, K. E. Schmidt, and J. C. H. Spence, “Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals,” *Acta Crystallogr. A* **67**, 131–140 (2011).
4. J. C. H. Spence, U. Weierstall, and H. N. Chapman, “X-ray lasers for structural and dynamic biology,” *Rep. Prog. Phys.* **75**, 102601 (2012).
5. F. Filsinger, G. Meijer, H. Stapelfeldt, H. N. Chapman, and J. Küpper, “State- and conformer-selected beams of aligned and oriented molecules for ultrafast diffraction studies,” *Phys. Chem. Chem. Phys.* **13**, 2076–2087 (2011).

6. P. Ho, D. Starodub, D. K. Saldin, V. L. Shneerson, A. Ourmazd, and R. Santra, "Molecular structure determination from X-ray scattering patterns of laser-aligned symmetric-top molecules," *J. Chem. Phys.* **131**, 131101 (2009).
7. S. Pabst, P. J. Ho, and R. Santra, "Computational studies of X-ray scattering from three-dimensionally-aligned asymmetric-top molecules," *Phys. Rev. A* **81**, 043425 (2010).
8. J. Küpper, S. Stern, L. Holmegaard, F. Filsinger, A. Rouzée, A. Rudenko, P. Johnsson, A. V. Martin, M. Adolph, A. Aquila, S. Bajt, A. Barty, C. Bostedt, J. Bozek, C. Caleman, R. Coffee, N. Coppola, T. Delmas, S. Epp, B. Erk, L. Foucar, T. Gorkhover, L. Gumprecht, A. Hartmann, R. Hartmann, G. Hauser, P. Holl, A. Hömke, N. Kimmel, F. Krasniqi, K.-U. Kühnel, J. Maurer, M. Messerschmidt, R. Moshhammer, C. Reich, B. Rudek, R. Santra, I. Schlichting, C. Schmidt, S. Schorb, J. Schulz, H. Soltau, J. C. Spence, D. Starodub, L. Strüder, J. Thøgersen, M. J. Vrakking, G. Weidenspointner, T. A. White, C. Wunderer, G. Meijer, J. Ullrich, H. Stapelfeldt, D. Rolles, and H. N. Chapman, "X-ray diffraction from isolated and strongly aligned gas-phase molecules with a free-electron laser," *Phys. Rev. Lett.* **112**, 083002 (2014).
9. D. K. Saldin, H. C. Poon, M. J. Bogan, S. Marchesini, D. A. Shapiro, R. A. Kirian, U. Weierstall, and J. C. H. Spence, "New light on disordered ensembles: Ab initio structure determination of one particle from scattering fluctuations of many copies," *Phys. Rev. Lett.* **106**, 115501 (2011).
10. H.-C. Poon, P. Schwander, M. Uddin, and D. K. Saldin, "Fiber diffraction without fibers," *Phys. Rev. Lett.* **110**, 265505 (2013).
11. H. Liu, B. K. Poon, D. K. Saldin, J. C. H. Spence, and P. H. Zwart, "Three-dimensional single-particle imaging using angular correlations from X-ray laser data," *Acta Cryst. A* **69**, 365–373 (2013).
12. J. J. Donatelli, P. H. Zwart, and J. A. Sethian, "Iterative phasing for fluctuation X-ray scattering," *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10286–10291 (2015).
13. R. R. Coifman, Y. Shkolnisky, F. J. Sigworth, and A. Singer, "Reference free structure determination through eigenvectors of center of mass operators," *Appl. Comput. Harmon. Anal.* **28**, 296–312 (2010).
14. V. L. Shneerson, A. Ourmazd, and D. K. Saldin, "Crystallography without crystals. I. the common-line method for assembling a three-dimensional diffraction volume from single-particle scattering," *Acta Crystallogr. A* **64**, 303–315 (2008).
15. N. D. Loh and V. Elser, "Reconstruction algorithm for single-particle diffraction imaging experiments," *Phys. Rev. E* **80**, 026705 (2009).
16. Z. Kam, "Determination of macromolecular structure in solution by spatial correlation of scattering fluctuations," *Macromolecules* **10**, 927–934 (1977).
17. D. K. Saldin, V. L. Shneerson, R. Fung, and A. Ourmazd, "Structure of isolated biomolecules obtained from ultrashort x-ray pulses: Exploiting the symmetry of random orientations," *J. Phys.: Condens. Matter* **21**, 134014 (2009).
18. V. Elser, "Strategies for processing diffraction data from randomly oriented particles," *Ultramicroscopy* **111**, 788–792 (2011).
19. M. Walczak, "Bayesian orientation estimate from single molecule X-ray scattering data," Master's thesis, Georg-August-Universität Göttingen (2010).
20. M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.* **15**, 1373–1396 (2003).
21. D. Giannakis, P. Schwander, and A. Ourmazd, "The symmetries of image formation by scattering. I. Theoretical framework," *Opt. Express* **20**, 12799–12826 (2012).
22. P. Schwander, D. Giannakis, C. H. Yoon, and A. Ourmazd, "The symmetries of image formation by scattering. II. Applications," *Opt. Express* **20**, 12827–12849 (2012).
23. J. de laPorte, B. M. Herbst, W. Hereman, and S. J. van der Walt, "An introduction to diffusion maps," Tech. rep., Proceedings of the 19th Symposium of the Pattern Recognition Association of South Africa (PRASA 2008) (2008).
24. J. Jost, *Riemannian Geometry and Geometric Analysis* (Springer, 2011).
25. V. Devanathan, *Angular Momentum Techniques in Quantum Mechanics* (Springer, 2002).
26. S. Stenholm, "Quantum theory of electromagnetic fields interacting with atoms and molecules," *Phys. Rep.* **6**, 1–121 (1973).

27. J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.* **21**, 2758–2769 (1982).
 28. C. H. Yoon, M. Barthelmess, R. J. Bean, F. Capotondi, R. A. Kirian, M. Kiskinova, E. Pedersoli, L. Raimondi, F. Stellato, F. Wang, and H. N. Chapman, "Conformation sequence recovery of a non-periodic object from a diffraction-before-destruction experiment," *Opt. Express* **22**, 8085–8093 (2014).
 29. J. Ham, D. D. Lee, S. Mika, and B. Schölkopf, "A kernel view of the dimensionality reduction of manifolds," in "Proceedings of the twenty-first international conference on machine learning," (ACM, 2004), pp. 47.
-

1. Introduction

X-ray free electron lasers (XFELs) hold the promise to image in a single laser-shot a large macromolecule or even a cell. Nano-scale images of specially prepared samples whose orientation with respect to the detector is known were already reconstructed in the beginning of XFELs with a soft-XFEL [1] sampling images from many single laser-shots. One of the biggest successes in XFEL imaging so far have been the characterization of biological molecules in the form of nano-crystals through serial femtosecond crystallography [2], where multiple copies of the molecule in the crystal with fixed relative orientation create the enhancement of the single shot diffractive image necessary for reconstruction [3, 4]. Also, the a priori orientation of dipolar molecules to be exposed to the X-ray flashes has made progress [5–8]. Further, the reconstruction of icosahedral and helical samples has been achieved [9, 10] and the overall structure of artificial and natural nano-scale objects has been determined successfully [11, 12] by exploiting correlations of images from different orientations.

In general, however, single-molecule coherent diffractive imaging without crystallization or prior orientation of the molecules in a sample has been proven to be very difficult since the orientation of a molecule in the gas phase varies from shot to shot and the averaging of images from similar orientations requires a reliable orientation recovery of each image.

A universal approach to classify orientations is the method of common lines. From the position of these lines in the scattering images the orientation can be calculated [13]. However, due to the low number of scattered photons, these lines cannot be found. For this method to be successful a photon flux 1000 times higher than currently available is needed [14]. Several proposals for reconstruction addressing the problem of these faint single-shot images, have been made: The EMC method proposed by Elser [15] maximizes a logarithmic probability function of the scattered intensity. Furthermore, it is possible to calculate the orientation from the correlations of the scattering images [16, 17], unfortunately this method does not work in three dimensions [18]. In another proposal, a model of the molecule is fitted to its scattering images by rotating parts of it [19]. Finally, a number of so called embedding strategies try to retrieve the orientation a posteriori by first determining how different the images are. This amounts to identifying a parameterization of the rotation (low-dimensional set of parameters) in the high dimensional set of many diffraction patterns.

In this paper we will introduce a systematic enhancement of the diffractive scattering images which optimizes the "distance" between images to be specified below and

facilitates therefore the orientation retrieval for all embedding algorithms which work with a distance measure of images. To illustrate the difficulties, motivate, formulate and apply the enhancement, we will use in the following explicitly diffusion map [20], as proposed recently [21, 22].

2. Diffusion map

Diffusion map is an algorithm which uncovers the low-dimensional structure (here 3 Euler angles) of a high-dimensional data set (in our case a set of scattering images), often referred to as latent manifolds [23]. The basic idea is to define so-called “diffusion distances” between two members X_i and X_j of the data set by setting up a diffusion matrix describing the probability of transitions between all members in the data set. The probabilities are obtained from the heat kernel, i. e., the transition from $X_i \rightarrow X_j$ is

$$P_{i \rightarrow j} = e^{-D_{ij}^2/D^2} \quad \text{with} \quad D_{ij} \equiv |X_i - X_j|, \quad (1)$$

where D is a parameter which corresponds to a typical distance in the data set. One should choose D large enough to prevent gaps in the data set and small enough to suppress transition for large distances D_{ij} and prevent short-cuts. D_{ij} is the Euclidean distance between the elements i and j of the data set. For scattering images composed of M pixels this is the distance between two M -dimensional vectors containing the intensities as vector components. The diffusion matrix with properly defined diagonal elements is also referred to as graph Laplacian [13].

Having set up the diffusion matrix one can describe a diffusion process starting from X_i and X_j , respectively, and compare the corresponding contributions. The similarity of the contributions is the “diffusion distance”. The decisive point in the dimensionality reduction is that in order to describe the diffusion process it is sufficient to determine the lowest eigenvalues and corresponding eigenvectors of the graph Laplacian.

Diffusion map can be applied to a set of scattering images, defined in Eqs. (3) below, in order to retrieve their orientations [21, 22]. The elements of the distance matrix D_{ij} , which is the essential ingredient of the graph Laplacian, are obtained from the Euclidean distance of two vectors whose components are given by scattering intensities per pixel. Having set up the graph Laplacian, its eigenvector of the lowest eigenvalues are used to define a rotation matrix for each image. This procedure has been described in detail before [21].

Indeed, the Laplacian for a given data set is an approximate representation of the full Laplacian, usually referred to as Laplace-Beltrami operator (LBO). In the limit of a data set with infinitely many elements it converges to the LBO [13]. Apart from the parameter D in Eq. (1) it is the coverage of the low-dimensional manifold, i. e., the number and the distribution of the underlying parameters, which matters. In order to exclude any influence of this distribution we will in the following study the LBO directly and not the graph Laplacian. As an additional benefit the numerical diagonalization is considerably cheaper than for the graph Laplacian. For the latter one has to calculate all distances D_{ij} , which is quite expensive for 10^6 images each consisting of $M=10^5$ pixels, just to give a concrete example.

3. Metric and Laplace-Beltrami operator

3.1. General

The Laplace-Beltrami operator (LBO) reads [24]

$$\hat{L}(\vec{\gamma}) = \frac{1}{J(\vec{\gamma})} \sum_{\mu, \nu=1}^3 \partial_{\mu} J(\vec{\gamma}) [g_{\mu\nu}(\vec{\gamma})]^{-1} \partial_{\nu} \quad \text{with} \quad J(\vec{\gamma}) \equiv \sqrt{\det |g_{\mu\nu}(\vec{\gamma})|} \quad (2)$$

with the three Euler angles $\vec{\gamma} = \{\varphi, \theta, \xi\}$ and the partial derivatives $\partial_{\nu} \equiv \frac{\partial}{\partial \gamma_{\nu}}$. The metric $g_{\mu\nu}(\vec{\gamma})$ measures how much scattering images change if the Euler angles are changed. Clearly, the change depends on the actual orientation given by $\vec{\gamma}$.

In the following the metric for scattering images is considered. A single scattering image for a molecule, approximated as a set of N individual atoms, is specified by the atomic positions \vec{r}_j and the atomic form factors $f_j(k)$ as

$$S_{\vec{\gamma}}(\vec{k}) = \sum_{jl}^N f_{jl}(k) e^{i\vec{k} \underline{R}_{\vec{\gamma}} \vec{r}_{jl}}, \quad (3a)$$

$$f_{jl}(k) = f_j(k) f_l(k), \quad \vec{r}_{jl} = \vec{r}_j - \vec{r}_l. \quad (3b)$$

The rotation matrix $\underline{R}_{\vec{\gamma}}$, cf. Eq. (6) for an explicit expression, defines the molecule's orientation. The scattered wave vector is given by $\vec{k} = k_1 \vec{e}_1 + k_2 \vec{e}_2$, with \vec{e}_1 and \vec{e}_2 being two orthogonal unit vectors in the detector plane, and $k = |\vec{k}|$.

The (squared) Euclidean distance between two images $S_{\vec{\gamma}}(\vec{k})$ and $S_{\vec{\gamma}'}(\vec{k})$ is given as the integral

$$D_{ij}^2 = \int_{\Omega} d^2k [S_{\vec{\gamma}}(\vec{k}) - S_{\vec{\gamma}'}(\vec{k})]^2, \quad (4)$$

over the image area Ω in k -space, which is defined by the incoming wavelength and the size of the detector. Typically Ω implies $|k_x|, |k_y| \leq k_{\max}$. The distance between two infinitesimally close images at orientation $\vec{\gamma}$ in 2nd order of the Euler angles reads $D_{\delta\vec{\gamma}}(\vec{\gamma}) = \delta\vec{\gamma} \underline{g}(\vec{\gamma}) \delta\vec{\gamma}$ and defines the metric for $\mu, \nu = 1 \dots 3$

$$g_{\mu\nu}(\vec{\gamma}) = \frac{1}{2} \int d^2k [\partial_{\mu} S_{\vec{\gamma}}(\vec{k})] [\partial_{\nu} S_{\vec{\gamma}}(\vec{k})]. \quad (5)$$

Assuming Gaussian electron densities and $k_{\max} \rightarrow \infty$ one can perform the k -integration analytically which leads to an explicit expression for the metric $g_{\mu\nu}$, given in Eq. (15) of appendix A. With the metric $g_{\mu\nu}$ the LBO (2) is uniquely defined.

3.2. Reference metric

To see whether the metric of a particular molecule and the corresponding LBO will allow for a classification by means of diffusion map we have to compare the eigenvectors of the 9 lowest eigenvalues with those from a reference LBO [21]. An obvious and

generic choice for this metric is the mapping from Euler angles to the rotation matrix. With the *zyz*-convention for Euler angles the rotation matrix reads

$$\underline{R}_{\vec{\gamma}} = \underline{R}_{\xi}^{(z)} \underline{R}_{\theta}^{(y)} \underline{R}_{\phi}^{(z)} \quad (6a)$$

$$\underline{R}_{\alpha}^{(z)} = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \underline{R}_{\alpha}^{(y)} = \begin{pmatrix} \cos \alpha & 0 & -\sin \alpha \\ 0 & 1 & 0 \\ \sin \alpha & 0 & \cos \alpha \end{pmatrix} \quad (6b)$$

Then, the reference metric is given analytically

$$g_{\mu\nu}^{[R]} = \begin{pmatrix} 1 & 0 & \cos \theta \\ 0 & 1 & 0 \\ \cos \theta & 0 & 1 \end{pmatrix}. \quad (7)$$

The corresponding LBO is the Hamilton operator for spherically symmetric rigid rotor [25]. Its eigenfunctions are the Wigner *D*-matrices $D_{mm'}^j(\vec{\gamma})$ with the eigenvalues $\varepsilon_{mm'}^j = j(j+1)$.

Since the scattering problem implies a projection onto a 2-dimensional detector plane (spanned by \vec{e}_1 and \vec{e}_2), one might consider a second reference metric for the matrix $\underline{P}\underline{R}_{\vec{\gamma}}$ with \underline{P} accounting for this projection. E. g., if the detector lies in the *xy*-plane the projection matrix is $\underline{P} = ((1, 0, 0), (0, 1, 0), (0, 0, 0))$, which corresponds to the metric

$$g_{\mu\nu}^{[P]} = \begin{pmatrix} \frac{1}{4}[3 + \cos(2\theta)] & 0 & \cos \theta \\ 0 & \frac{1}{2} & 0 \\ \cos \theta & 0 & 1 \end{pmatrix}. \quad (8)$$

The LBO based on $g_{\mu\nu}^{[P]}$, which incorporates the projection step, is also diagonal in the Wigner *D*-matrix basis, but its eigenvalues $\tilde{\varepsilon}_{jmm'} = j(j+1) - m'^2$ are no longer degenerated with respect to *m* and *m'*.

This suggests to diagonalize the LBO with the molecule-specific metric $g_{\mu\nu}$ in the basis of Wigner *D*-matrices and calculate the principal angles [26] between the 9 lowest eigenvectors (excluding the lowest one for *j*=0) and the 9 functions $D_{mm'}^1(\vec{\gamma})$, since these 9 eigenvectors are sufficient [21].

3.3. Analysis

Before quantifying the LBO we would like to give an intuitive explanation why the metric differs considerably from both reference metrics (7) and (8), respectively. To this end we separate out in the metric the term which depends on the difference in atomic vectors $P_{\mu\nu}$ (with the exact expression given in Eq. (15) in appendix A)

$$g_{\mu\nu} \sim \sum_{jlmn}^N e^{-\kappa_{jlmn}^2(\vec{r}_{jl} - \vec{r}_{mn})^2} \left[\vec{r}_{jl} \underline{A}_{\mu\nu} \vec{r}_{mn} + P_{\mu\nu}(\vec{r}_{jl} - \vec{r}_{mn}) \right], \quad (9)$$

where the indices *jlmn* run over all *N* atoms in the molecule and the rotated and projected coordinates are \vec{r}_{jl} . The κ_{jlmn} , which are inverse to the extension of the contributing electron shells, determine the effective range over which the scattering image

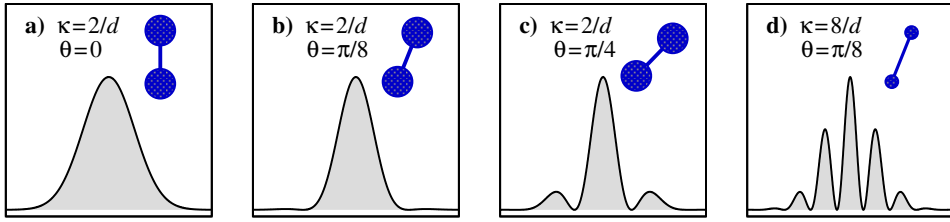


Fig. 1. Sketches of a diatomic molecule consisting of two atoms with a form factor $f_{\kappa}(k) = \exp(-k^2/[4\kappa]^2)$ and an inter-atomic distance d at different orientations given by θ , the angle between intra-molecular axis and normal of the detector plane. For each orientation a cross section of the corresponding scattering image $S_{\kappa d}(k) = f_{\kappa}^2(k)[1 + \cos(kd \sin \theta)]$ for weak (a–c) and strong (d) localizations is shown.

decays. It turns out, that in general $P_{\mu\nu}$ is the part which modifies the metric in such a way that a successful recovery becomes impossible. Its influence can be suppressed for $\vec{r}_{jl} = \vec{r}_{mn}$ since $P_{\mu\nu}(0) = 0$, cf. Eq. (15c). This can be effectively achieved through localised Gaussians for $\kappa_{jlmn} \rightarrow \infty$. In this limit the Gaussians in (9) approaches a Kronecker symbol $\delta_{jm}\delta_{ln}$ and the metric simplifies to

$$\lim_{\kappa \rightarrow \infty} g_{\mu\nu}(\vec{\gamma}) \sim \sum_{jl} \vec{r}_{jl} \Delta_{\mu\nu}(\vec{\gamma}) \vec{r}_{jl}. \quad (10)$$

The sum (10) resembles the metric of the differences from the projected coordinates for point-like electron densities which is generally better suited for orientation recovery.

This may be illustrated, e.g., with Platonic solids. For their projected coordinates, the metric (10) converges to the reference metric $g_{\mu\nu}^{[P]}$ from Eq. (8). For other sets of coordinates the eigenvalues and -vectors of the metric (10) cannot be calculated analytically, but their lowest eigenfunctions are also linear combinations of Wigner D -matrices. This changes if the width of one dimension of the coordinates is about three times the width of the other two dimensions (prolate deformation). In this case the rotation around the long axis does not change the projected distances of the coordinates and at least one angle out of the three Euler angles is lost. Then only the orientation of the long axis can be reconstructed.

3.4. The diatomic molecule as an example

There is an intuitive interpretation for the expressions (9) and (10), respectively, considering a diatomic molecule. Scattering images (more precise cross sections of those images) are shown in Fig. 1 for different values of θ , the angle between the intra-molecular axis and the normal of the detector plane. For increasingly diffuse electron densities (realised by decreasing κ), the density projected onto the detector will overlap for almost normal orientation ($\theta \ll \pi/2$). Consequently, the resulting scattering image, cf. Fig. 1(b), will be very similar to the one where the intra-molecular axis is normal to the detector plane, cf. Fig. 1(a). This similarity is responsible for the term P in Eq. (9). For larger angles, cf. Fig. 1(c), interference oscillations start to emerge. Most impor-

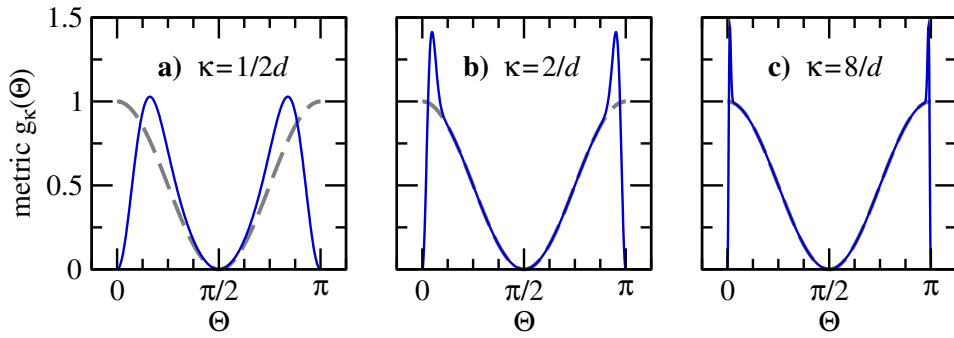


Fig. 2. Metric $g_{\kappa}(\theta)$ of the diatomic molecule (blue-solid line) as given in Eq. (11) for three ratios of κ/d in comparison with the reference metric g_{ref} (gray-dashed) from Eq. (12).

tantly, these oscillations develop already for *smaller* angles if the electron density is *stronger localized* at the atoms as shown in Fig. 1(d).

This interpretation can be quantified in terms of the metric of a diatomic molecule. With atoms at $\vec{r}_1 = \{+\sin\theta, 0, -\cos\theta\}d/2$ and $\vec{r}_2 = \{-\sin\theta, 0, +\cos\theta\}d/2$ this metric reads

$$g_{\kappa}(\theta) = \cos^2\theta \left((8[\kappa d]^2 \sin^2\theta - 1)e^{-4[\kappa d]^2 \sin^2\theta} + 1 \right). \quad (11)$$

Metrics for three values of κ are shown in Fig. 2, where b) and c) are the metrics from Fig. 1. The reference metric in this case is

$$g^{[R]}(\theta) = \cos^2\theta, \quad (12)$$

since the projected distance is proportional to $\sin\theta$. Looking at the extrema for the metric for the scattering images (11), namely $g_{\kappa \gg 1}(\theta) \propto \cos^2\theta$ and $g_{\kappa \ll 1}(\theta) \propto \cos^2\theta \sin^2\theta$ it becomes clear that only tightly bound electron, i. e. large κ , will allow for a successful orientation. The deviation for finite κ emerge at $\theta \approx 0$ and $\theta \approx \pi$ as can be seen in Fig. 2(b). This corresponds to the case when the molecular axis is parallel to the detector normal. For very small values of κ , cf. Fig. 2(c), the deviation spreads over the whole range $\theta = 0 \dots \pi$, preventing a successful orientation recovery.

4. Recovery of the recovery

4.1. Application to Chignolin

In a realistic molecule the extension of the electron density is comparable to the interatomic distances and thus not small. Therefore situations as sketched in Fig. 1(b) will occur frequently. This explains that the principal angles for Chignolin are so large (4 out of the 9 are about 90°) that the 2D scattering images, which are cuts through the 3D Fourier space, cannot be placed at the right positions and the calculation of a 3D scattering image fails. The inspection of the metric indicates a solution for this problem. Since one cannot localize electrons in a molecule, one has to modify the scattering

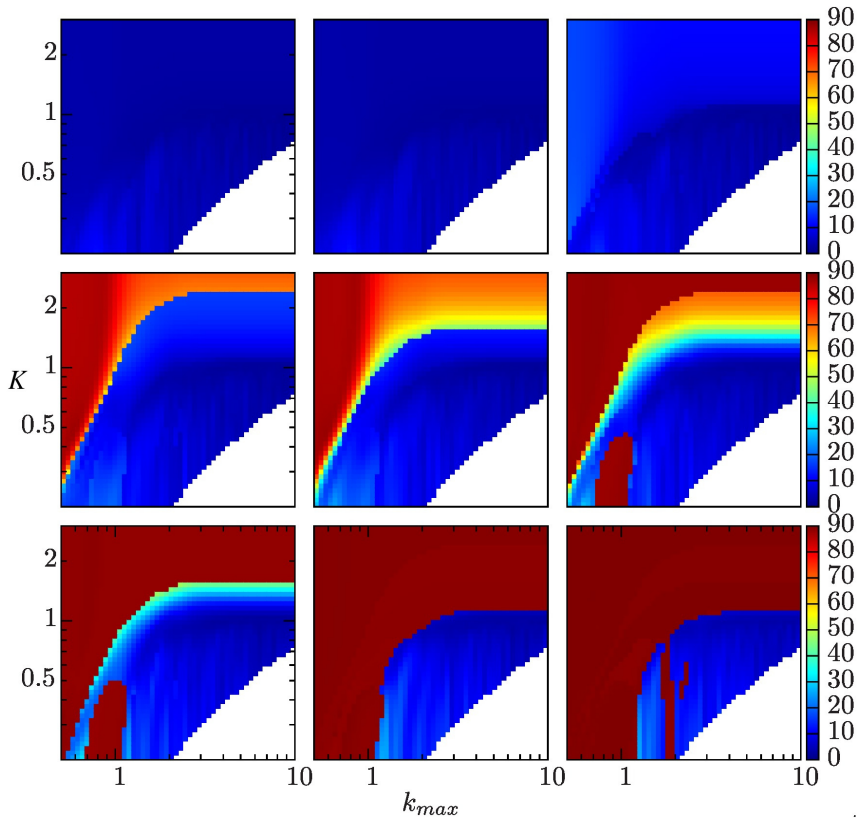


Fig. 3. Angles (colorcoded) between the subspaces spanned by the Wigner D -matrices and the eigenfunctions of the Laplace-Beltrami operator of Chignolin as function of maximal transfer momentum k_{\max} and the parameter K in the image-modification factor $C_K(k)$ given in Eq. (13).

image. Indeed, there is a possibility to improve the metric, which can be considered as an artificial localization of electrons. Multiplying the scattering images by

$$C_K(k) = e^{k^2/[4K]^2} \quad (13)$$

artificially enlarges the κ_{jlmn} as required by (10). The form of the modification factor (13) is motivated by a molecule consisting of just one type of atoms with a form factor f , for which the scattering image would read $S_{\vec{\gamma}}(\vec{k}) = f^2(k) \sum_{jl} e^{i\vec{k}R_{\vec{\gamma}}\vec{r}_{jl}}$. In this case (13) would exactly compensate the prefactor in the scattering image. In general, the multiplication by $C_K(k)$ enhances the large-angle part of the scattering image.

We support this claim with the protein Chignolin (studied before in this context [22]). Results as a function of the maximal scattered wave vector k_{\max} (whereby $k_{\max} \approx 3.2$ corresponds to a photon frequency of 12 keV) and the parameter K in the modification factor $C_K(k)$ from Eq. (13) are shown in Fig. 3. There we have considered sufficiently high photon numbers to avoid effects from shot noise. The first parameter was chosen to have different resolution of the electronic density in real space. As just discussed, a

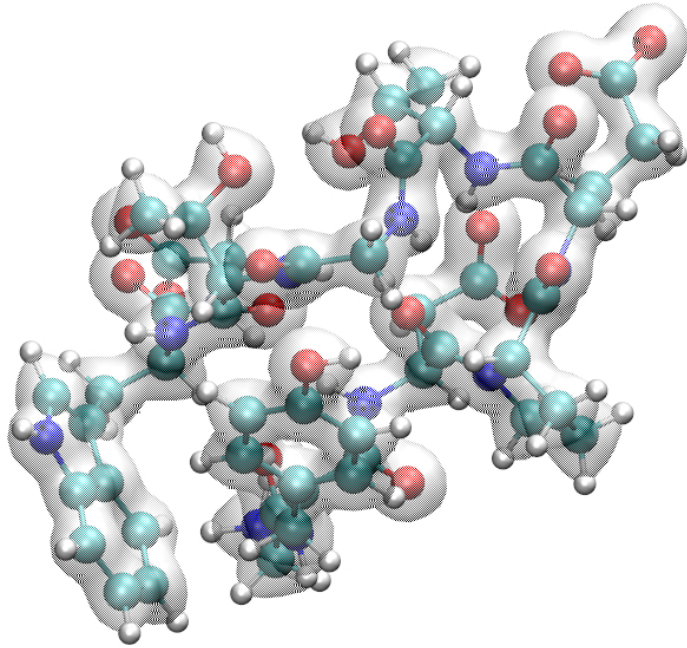


Fig. 4. Equipotential surfaces of the reconstructed electron density (note that the density around the hydrogen atoms does not show since it is smaller than the chosen equipotential value) and ball-and-stick-model of Chignolin.

more diffuse form factor corresponds to a more localized atomic electron density. If the resolution in real space given by $2\pi/k_{\max}$ is not sufficient to resolve a narrow electron density then the image density appears to be much wider and $P_{\mu\nu}$ in Eq. (9) does not vanish.

It can be seen clearly that this procedure offers a huge improvement over the unmodified version. For $k_{\max} > 2.5$ there is always a K that results in a success of the orientation classification. For $k_{\max} < 2$ this success can not be guaranteed, because the resolution is too small. In the white area in Fig. 3 the metric could not be calculated, since $C_K(k)$ becomes too large for $|\vec{k}| \approx k_{\max}$.

4.2. Reconstruction of the molecular electron density

Once the orientation recovery was successful using the modified diffraction images, reconstruction can begin with the oriented, original images. For the construction of the 3D Fourier image, 10^7 scattering images in different orientations are used. The normal of the detector plane is initially always along the z -axis and its grid-points are denoted by \vec{k}_i . It is important to note that when the molecule is rotated by $\underline{R}_{\vec{\gamma}}$ from (6), the grid-points \vec{k}_i on the detector plane have to be rotated by $\underline{R}_{\vec{\gamma}}^{-1}$, in order to lie on the right spot in the 3D Fourier-space (active vs. passive rotations). The procedure for forming the 3D scattering image is described in detail in appendix B.

For the phase-inversion the hybrid input-output algorithm [27] with a fixed support

is chosen. This algorithm calculates the phase of the 3D scattering image $S(\vec{K})$ by successive projections on the Fourier and real space. The fact that the absolute value of $S(\vec{K})$ is known is used in every step and only the latest calculated phase is kept. The real space projection is a bit different from the standard procedure where the calculated density is set to zero outside of the support. Instead it is seen as a driving force for the next projection on the Fourier space. Those are standard steps and the reconstructed electron density is shown on top of the molecule in Fig. 4. The agreement with the ball-and-stick-model is really good.

5. Conclusions

Analyzing the metric which links rotated X-ray images, we have identified too diffuse form factors as a source for the failure of the diffusion map algorithm to be able to orient scattering images. Enhancing these images for orientation by rescaling the width of the form factor to make it more compact has enabled us to orient the images and reconstruct from the successfully oriented original images the molecule, explicitly demonstrated for Chignolin.

We expect this enhancement of the X-ray images to be applicable in orientation recovery methods which use image differences for manifold reduction, e.g., also in the isomap algorithm recently used to reconstruct a time sequence of nano-images in a soft X-ray FEL [28]. Since essentially all kernel-driven manifold reduction algorithms are equivalent [29], this should hold for a large class of algorithms. One caveat, however, refers to the quality of the single shot images: If the wings of the form factors have extremely low photon-count they are very noisy and it may not be possible to amplify them by rescaling such they become a meaningful input for the orientation recovery.

In the long run it would be interesting to also compare methods based on very different ideas [16–19] regarding their potential for structure determination from randomly oriented samples.

A. Scattering metric

To allow for an analytical expression for the metric given in Eq. (9) Gaussian form-factors $f_j(k) = n_j e^{-k^2/[4\kappa_j]^2}$ are used corresponding to an electron density

$$\rho_j(r) = [n_j/\sigma_j^3] e^{-4\ln 2 r^2/\sigma_j^2} \quad (14)$$

containing n_j electrons with a radial extension σ_j (it is $\kappa_j = \sqrt{\ln 2}/\sigma_j$). After some lengthy but straightforward calculation one arrives at an explicit expression for the metric

$$g_{\mu\nu}(\vec{\gamma}) = 8\pi \sum_{jl, mn}^N n_{jlmn} \kappa_{jlmn}^4 e^{-\kappa_{jlmn}^2 [(d_{1,jlmn})^2 + (d_{2,jlmn})^2]} \times \left[[\vec{r}_{jt} \underline{A}^{\mu\nu} \vec{r}_{mn}] - 2\kappa_{jlmn}^2 \sum_{s,t=1}^2 d_{s,jlmn} d_{t,jlmn} [\vec{r}_{jt} \underline{A}_{st}^{\mu\nu} \vec{r}_{mn}] \right], \quad (15a)$$

whereby the abbreviations

$$n_{jlmn} \equiv n_j n_l n_m n_n \quad (15b)$$

$$\kappa_{jlmn}^2 \equiv 4 \ln 2 / (\sigma_j^2 + \sigma_l^2 + \sigma_m^2 + \sigma_n^2) \quad (15c)$$

$$d_{s,jlmn} \equiv \vec{e}_s \underline{R}_{\vec{\gamma}} [\vec{r}_{jl} - \vec{r}_{mn}] \quad (15d)$$

$$\underline{A}_{st}^{\mu\nu} \equiv [\partial_\mu \underline{R}_{-\vec{\gamma}}] \vec{e}_s \otimes \vec{e}_t [\partial_\nu \underline{R}_{\vec{\gamma}}] \quad (15e)$$

$$\underline{A}^{\mu\nu} \equiv [\partial_\mu \underline{R}_{-\vec{\gamma}}] \underline{P} [\partial_\nu \underline{R}_{\vec{\gamma}}] \quad (15f)$$

have been used. Recall that \vec{e}_s with $s = 1, 2$ are two orthogonal unit vectors spanning the detector plane.

B. Forming the 3D scattering image

According to the Fourier-projection-slice theorem, the projection of the molecular density on the detector plane is just a 2D cut through the 3D Fourier-space. When the normal of each cut is known ($\underline{R}_{\vec{\gamma}}^{-1} \vec{e}_z$), they can be placed at the correct angle in the 3D Fourier-space. Since the resulting grid is irregularly spaced, an interpolation on a regular one is necessary. At first the two planes closest to the regular grid-point \vec{K}_j are determined by minimizing the absolute value of the scalar product

$$d_{ji} = \frac{1}{|\vec{K}_j|} \vec{K}_j \underline{R}_{\vec{\gamma}_i}^{-1} \vec{e}_z. \quad (16)$$

The Euler angles of these two planes are denoted by $\vec{\gamma}_{j_1}$ and $\vec{\gamma}_{j_2}$.

On each plane the point closest to \vec{K}_j is chosen (denoted by \vec{k}_{j_1} and \vec{k}_{j_2}) for the linear interpolation of the three dimensional scattering image

$$S(\vec{K}_j) = \begin{cases} S_{\vec{\gamma}_{j_1}}(\vec{k}_{j_1}) + [S_{\vec{\gamma}_{j_2}}(\vec{k}_{j_2}) - S_{\vec{\gamma}_{j_1}}(\vec{k}_{j_1})] \frac{|\vec{k}_{j_1} - \vec{K}_j|}{|\vec{k}_{j_1} - \vec{K}_j| + |\vec{k}_{j_2} - \vec{K}_j|}, & \text{if } (\vec{k}_{j_1} - \vec{K}_j) \cdot (\vec{k}_{j_2} - \vec{K}_j) \leq 0 \\ S_{\vec{\gamma}_{j_1}}(\vec{k}_{j_1}) \pm [S_{\vec{\gamma}_{j_2}}(\vec{k}_{j_2}) - S_{\vec{\gamma}_{j_1}}(\vec{k}_{j_1})] \frac{|\vec{k}_{j_1} - \vec{K}_j|}{|\vec{k}_{j_1} - \vec{k}_{j_2}|}, & \text{otherwise.} \end{cases} \quad (17)$$

In the first case the point \vec{K}_j is between the points \vec{k}_{j_1} and \vec{k}_{j_2} and a standard interpolation is used. For the second case the opposite is true and the value at \vec{K}_j is extrapolated and with the right direction is chosen by the \pm sign.