2021 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-4115-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIP42928.2021.9506435

CRYO-ELECTRON MICROSCOPY IMAGE DENOISING USING MULTI-FREQUENCY VECTOR DIFFUSION MAPS

*Yifeng Fan, Zhizhen Zhao**

Department of ECE and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign

ABSTRACT

Cryo-electron microscopy (EM) single particle reconstruction is a general technique for 3D structure determination of macromolecules. However, because the images are taken at low electron dose, it is extremely hard to visualize the individual particle with low contrast and high noise level. In this paper, we propose a novel framework for cryo-EM single particle image denoising, which incorporates the recently developed multi-frequency vector diffusion maps [1] for improving the identification and alignment of images with similar viewing directions. In addition, we propose a novel filtering scheme combining graph signal processing and truncated Fourier-Bessel expansion of the projection images. Through both simulated and publicly available real data, we demonstrate that our proposed method is efficient and robust to noise compared with the state-of-the-art cryo-EM 2D class averaging algorithms.

Index Terms— Cryo-electron microscopy, multi-frequency vector diffusion maps, image denoising.

1. INTRODUCTION

Cryo-electron microscopy (cryo-EM) is a ground-breaking technology that enables us to image biomolecules in their native functional states at high resolution. In cryo-EM experiments, randomly oriented protein particles are distributed evenly within the special sample grid holes, an electron beam projects such particles and generates a projection image, which is modulated by the contrast transfer function (CTF) [2] and contaminated by noise with an extremely low signal-to-noise ratio (SNR). Single particle images are then boxed out by a particle picking step [3, 4] and the goal is to recover the 3D structure of the particle (see Fig. 1).

Since the SNR of the raw images is typically too low for reconstruction, in current frameworks a crucial step called *2D class averaging* aims to denoise and improve the image quality. It groups, aligns, and averages images with similar viewing directions. The denoised images are then taken as the input to 3D *ab initio* reconstruction [5, 6] and refinement

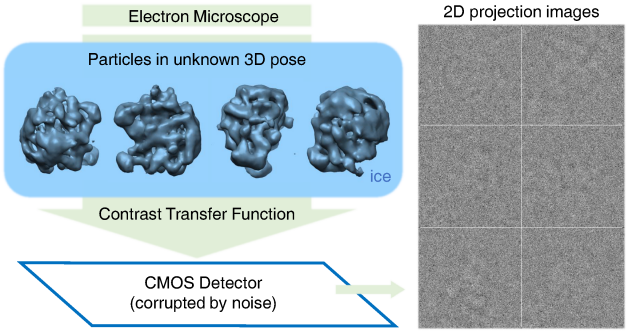


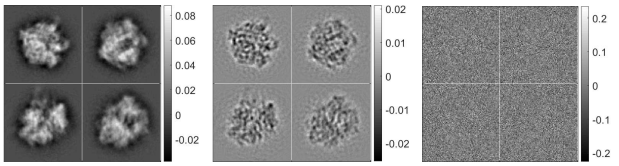
Fig. 1: Image formation in cryo-EM: The electron beam projects the randomly oriented particles and generates projection images, which is corrupted by noise with extremely low SNR.

algorithms. A better denoising result is beneficial in a variety of aspects, not only because it leads to a more accurate 3D reconstruction [7], but also the denoised image itself can be used to directly observe heterogeneity, discover symmetry, and separate particles into subgroups for additional analysis. Therefore, it is important to have fast and accurate algorithms for denoising cryo-EM images.

Many existing software packages [8, 9, 10] for single particle reconstruction contain algorithms for 2D class averaging. Most common methods are basically variants of k-means with iterative rotational alignment and clustering. Another kind of method implemented in the ASPIRE class averaging pipeline [7] is a two-stage algorithm, which first performs an initial nearest neighbor search and alignment estimation, by using rotationally invariant features to identify images of similar views, then it applies *vector diffusion maps* [11, 12] to further improve the initial result by considering the consistency of in-plane rotations among neighboring images. However, when SNR is extremely low, all the methods above might fail to generate satisfying 2D class averaging results, which leads to inaccurate 3D ab initio reconstructions.

In this paper, we propose a novel framework for denoising cryo-EM images, by applying the recently developed algorithm called *multi-frequency vector diffusion maps* (MFVDM) [1] for improving the efficiency and accuracy of nearest neighbor search and alignment estimation. Furthermore, we propose to denoise the truncated Fourier-Bessel expansion coefficients of the images using the top eigenvalues and eigenvectors of the averaging operators in MFVDM,

*Thanks to NSF DMS-1854791, NSF OAC-1934757, and Alfred P. Sloan Foundation for funding.



(a) Clean (b) w. CTFs (c) SNR $= 0.01$

Fig. 2: Simulated 70S ribosome projection images: (a) clean images; (b) images altered by CTFs; (c) images contaminated by additive white Gaussian noise at SNR $= 0.01$.

and reconstruct the images from the denoised expansion coefficients. Unlike traditional 2D class averaging methods, this approach does not require clustering and rotation of the original Cartesian grid images. We demonstrate the performance of our framework by comparing it with the state-of-the-art algorithms on both simulated and experimental datasets.

2. METHODS

2.1. Image formation model and geometry

In cryo-EM experiments, each particle is embedded in a thin layer at an unknown orientation characterized by a 3×3 orthogonal matrix $R = [R^1, R^2, R^3] \in$ SO(3). Assuming particles are well centered, the projection image I of a single particle can be expressed as a line integral $I(x, y) = \int_{-\infty}^{\infty} \phi(xR^1 + yR^2 + zR^3)dz$, where (x, y) is the pixel location and ϕ is the 3D volume of the particle. In particular, the matrix R can be divided into two parts: (1) the last column R^3 represents the viewing (projection) direction $v = R^3$ of the particle; (2) the first two columns namely R^1 and R^2 form a basis of the tangent plane whose normal vector is v. As a result, projection images with similar viewing directions are almost identical up to some in-plane rotations.

The projection images are corrupted by two phenomena: contrast transfer functions (CTFs) [2] and noise. The CTF modulates the phase amplitudes and flips the signs of the phases in the Fourier domain, which leads to a systematic alteration of the image data (see Figs. 2(a) and 2(b)). Then noise is assumed to be added on the image modulated by CTF with zero mean (see Fig. 2(c)). In the Fourier domain, each noisy image I_i is related to the underlying clean image $\bar{I}_i$ via

$$\mathcal{F}(I_i) = C_i \cdot \mathcal{F}(\bar{I}_i) + \varepsilon_i, \text{ for } i = 1, \ldots, n, \tag{1}$$

where $\mathcal{F}$ denotes the Fourier transform and ϵ_i represents the additive noise. The CTF C_i is assumed to be given for each particle image.

To denoise the images, 2D class averaging [2] is a crucial step that first identifies images with similar viewing directions and registers those images by in-plane rotations (and shifts if images are not well centered), then it denoises by aligning and averaging neighboring images. To identify images of similar views, for a pair of images I_i and I_j, it is natural to define the rotational invariant distance (RID) $d_{\text{RID}}(i, j)$ as

$$d_{\text{RID}}(i, j) = \min_{\alpha \in [0, 2\pi)} \|I_i - \mathcal{R}_\alpha I_j\|, \tag{2}$$

where $\mathcal{R}_\alpha$ is an operator that rotates image I_j by an angle α, and the associated optimal alignment is denoted by α_{ij}. In the noiseless case, since small d_{RID} indicates close viewing directions (assuming no symmetry), class averaging can be easily done by clustering based on d_{RID}. While in the case of extremely low SNR, noise dominates and could align well instead of the underlying signals, then images with distant views could also have a small d_{RID}. As a result, directly using d_{RID} would lead a poor class averaging result.

In [13], we introduced a fast steerable principal component analysis (sPCA) and a Wiener filtering approach to denoise sPCA expansion coefficients of the images. The denoised coefficients can be used to construct rotational invariant features according to [7]. We identify the initial κ-nearest neighbors for each image using a randomized algorithm for fast nearest neighbor search [14] on image features. The computational complexity for the initial nearest neighbor search scales as $O(nd \log n)$ instead of $O(n^2 d)$ for brute force search, where d is the dimension of the invariant feature.

2.2. Improving nearest neighbor identification and rotational alignment

To improve the accuracy of nearest neighbor identification from the noisy image data, we use the similarity measure based on *multi-frequency vector diffusion maps* (MFVDM) [1]. The first step is to build a graph $G = (V, E)$ based on the initial κ-nearest neighbor list and pairwise alignment angles. Each node in the graph represents an image and two nodes i and j are connected if the (i, j) pair is in the κ-nearest neighbor list. Each edge contains the optimal alignment α_{ij}.

We notice that for clean images of the same viewing directions, I_i, I_j and I_l, the pairwise alignment angles α_{ij}, α_{jk} and α_{kl} must satisfy

$$k(\alpha_{ij} + \alpha_{jk} + \alpha_{kl}) = 0 \mod 2\pi, \text{ for } k = 1, 2, \ldots, K. \tag{3}$$

This consistency in Eq. (3) can be extended to cycles of four or more images. The *cycle consistency* allows us to detect images with similar viewing directions in the presence of wrongly identified nearest neighbors. To systematically incorporate the alignment information and impose the consistency of alignments, given the initial graph $G = (V, E)$, we construct a set of $n \times n$ weight matrices W_k for $k = 1, 2, \ldots, K$ such that

$$W_k(i, j) = \begin{cases} e^{\imath k \alpha_{ij}} & (i, j) \in E, \\ 0 & (i, j) \notin E, \end{cases} \tag{4}$$

where $\alpha_{ij} = -\alpha_{ji}$. In [15], we detail the spectral properties of the matrices W_k. We define the diagonal degree matrix D as $D(i, i) = \deg(i) = \sum_{j:(i,j)\in E} 1$. Then the matrix $S_k = D^{-1} W_k$ can be viewed as an averaging operator of the vector fields, because for $v \in \mathbb{C}^n$, $(S_k v)(i) = \frac{1}{\deg(i)} \sum_{j:(i,j)\in E} e^{\imath k \alpha_{ij}} v(j)$ and it averages the vectors at nodes j connected to i after being rotated by α_{ij}.

$S_k^{2t}(i,j)$ is a weighted sum of the transformations along all length-$2t$ paths connecting i and j. A large value of $|S_k^{2t}(i,j)|$ means the transformations along all the paths are consistent, which indicates that I_i and I_j are true neighbors with high probability. This affinity can be efficiently approximated using the top eigenvalues $\lambda_1^{(k)} \geq \lambda_2^{(k)} \geq \cdots \geq \lambda_{m_k}^{(k)}$ and eigenvectors $\{u_\ell^{(k)}\}_{\ell=1}^{m_k}$ of a similar matrix $\widetilde{S}_k = D^{1/2} S_k D^{-1/2} = D^{-1/2} W_k D^{-1/2}$. Please refer to [1] for more details. The refined nearest neighbor list is determined by $\sum_{k=1}^{K} |S_k^{2t}(i,j)|$.

For nodes i and j of the same viewing directions, the eigenvectors satisfy $u_\ell^{(k)}(i) = e^{\imath k \alpha_{ij}} u_\ell^{(k)}(j)$. Therefore, we can estimate the in-plane rotational alignment angles from the top eigenvalues and eigenvectors of $\widetilde{S}_k$,

$$\hat{\alpha}_{ij} = \underset{\alpha \in [0,2\pi)}{\operatorname{argmax}} \sum_{k=1}^{k_{\max}} \left(\sum_{\ell=1}^{m_k} \left(\lambda_\ell^{(k)}\right)^{2t} u_\ell^{(k)}(i) \overline{u_\ell^{(k)}(j)} \right) e^{-\imath k \alpha}. \quad (5)$$

Based on the new nearest neighbor list and alignment angles, we can build a set of new matrices $\hat{W}_k$, $\hat{D}$, and $\hat{S}_k$ following the procedure of constructing W_k, D, and S_k. The averaging operator $\hat{S}_k$ is used for image denoising.

2.3. Denoising and CTF correction

We assume that each image I_i takes discrete samples on a Cartesian grid from an underlying continuous function with bandlimit s and essentially concentrated within a radius of R. Then the image in the Fourier domain can be well approximated by the truncated expansion on Fourier-Bessel basis $\psi_s^{k,q}(\xi,\theta)$ [13] supported on the disk of radius s,

$$P_{s,R}\mathcal{F}(I_i)(\xi,\theta) = \sum_{k=-k_{\max}}^{k_{\max}} \sum_{q=1}^{p_k} a_{k,q}^i \psi_s^{k,q}(\xi,\theta), \quad (6)$$

where $P_{s,R}$ is the orthogonal projection from the space of L_2 functions supported on a disk of radius s to the space of functions spanned by a finite number of Fourier-Bessel functions. The truncation $k_{\max}$ and $\{p_k\}_{k=1}^{k_{\max}}$ are determined through a sampling criterion detailed in [16, 13]. When I_i is rotated clockwise by α, the Fourier-Bessel expansion coefficient $a_{k,q}^i$ is simply transformed to $a_{k,q}^i e^{\imath k \alpha}$, and 2D class averaging can instead be performed on $a_{k,q}^i$.

Let $A^{(k)}$ be the $n \times p_k$ matrix of the coefficients with angular frequency k for all the n images, such that the i^{th} row corresponds to $\{a_{k,q}^i\}_{q=1}^{p_k}$ for the i^{th} image. Then applying $\hat{S}_k$ to $A^{(k)}$ for all k achieves denoising by phase aligning and averaging the coefficients among the nearest neighbors.

However, simply averaging nearest neighbors could result in over-smoothing, since each image is not exactly identical to its aligned neighbors and the estimation error of alignment exists. To resolve this issue, we use $2\hat{S}_k - \hat{S}_k^2$ instead of $\hat{S}_k$ as the denoising operator, which is related to forward and backward diffusion on the graph that first smooths and then sharpens the signal defined on the graph [17, 18]. As a result, this operator reduces the blurring caused by simply using $\hat{S}_k$. It is equivalent to applying a spectral filter $h(\cdot)$ to the eigenvalues of $\hat{S}_k$ such that $h(\lambda) = 2\lambda - \lambda^2$. In addition, since eigenvectors with smaller eigenvalues are more oscillatory and more easily perturbed by noise, it is necessary to set $h(\lambda) = 0$ for $\lambda < \delta$, which is equivalent to keeping the top $m_k \ll n$ eigenvalues and eigenvectors of $\hat{S}_k$. Then the matrices of the denoised coefficients at frequency k is given as

$$\begin{aligned} \hat{A}^{(k)} &= h(\hat{S}_k) A^{(k)} = \hat{D}^{-1/2} h(\underbrace{\hat{D}^{-1/2} \hat{W}_k \hat{D}^{-1/2}}_{=U_k \Lambda_k U_k^*}) \hat{D}^{1/2} A^{(k)} \\ &= \hat{D}^{-1/2} U_k h(\Lambda_k) U_k^* \hat{D}^{1/2} A^{(k)}, \end{aligned} \quad (7)$$

where Λ_k is a diagonal matrix that contains the eigenvalues of $\hat{S}_k$. The Fourier coefficients of the denoised image without CTF correction can be recovered from $\hat{A}^{(k)}$ as

$$\mathcal{F}\left(\widetilde{I}_i\right)(\xi_1,\xi_2) = \sum_{k=-k_{\max}}^{k_{\max}} \sum_{q=1}^{p_k} \hat{a}_{k,q}^i \psi_s^{k,q}(\xi_1,\xi_2), \quad (8)$$

where (ξ_1,ξ_2) are located on the Cartesian grid points.

CTF correction: It remains to correct the effect of CTFs. From Eq. (1), each C_i is a mask on the Fourier domain, and we can also compute its Fourier-Bessel expansion coefficients $c_{k,q}^i$ similar to Eq. (6). Then the effect of CTF on $\widetilde{I}_i$ can be estimated by applying the same operator as in Eq. (7) and replace $a_{k,q}^i$ by $c_{k,q}^i$, which yields the Fourier coefficients $\widetilde{C}_i(\xi_1,\xi_2)$ of the effective CTF according to Eq. (8). Based on these, we perform the following CTF correction to estimate the underlying clean projection image:

$$\hat{I}_i(x,y) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\widetilde{I}_i)(\xi_1,\xi_2)}{\widetilde{C}_i(\xi_1,\xi_2)} \right)(x,y). \quad (9)$$

The main advantage of our approach compared to other existing class averaging methods (e.g., [9, 8]) is that it does not require clustering and rotation of the Cartesian grid images. Also, the denoised coefficients in Eq. (7) for different k can be computed in parallel.

3. EXPERIMENTAL RESULTS

We apply our method on both simulated and experimental datasets. We compare with several existing approaches: (1) ASPIRE 2D class averaging [7], (2) Covariance Wiener Filter (CWF) [19], (3) RELION [8], (4) Steerable Graph Laplacian (SGL) [20]. All experiments were performed on a Linux system with a 60 cores Intel Xeon CPUs (12 cores were used), running at 2.3GHz with 512GB RAM in total.

Simulated dataset: We generate $n = 10^4$ projection images from the centered volume of 70S ribosome, where viewing directions are uniformly distributed over SO(3). The effects of CTF and additive white Gaussian noise are added with SNR $= 0.05$ and 0.01, which are typical in real problems. In Fig. 3, we display samples of denoised images by different

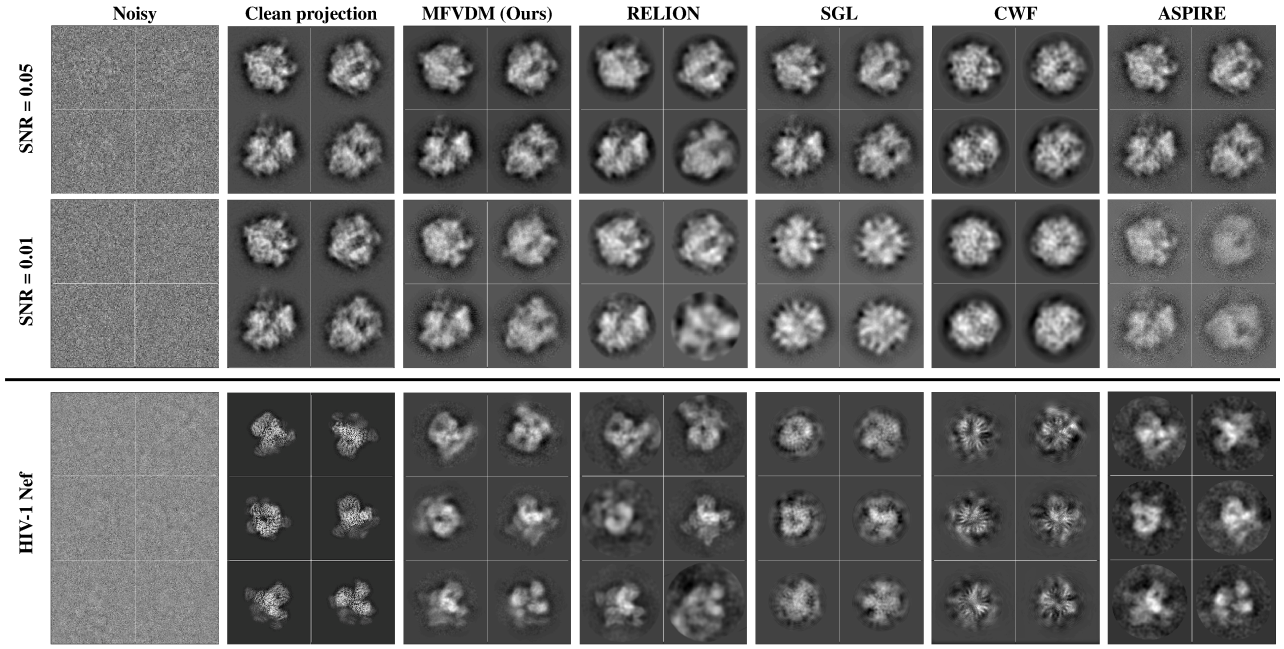


Fig. 3: *Top:* Image samples of simulated 70S ribosome dataset with SNR $= 0.05$ and 0.01. *Bottom:* Image samples of HIV-1 Nef dataset. Different columns display noisy images, clean projections and denoising results of different methods.

Table 1: MSE of the denoised images on 70S ribosome.

SNR	MFVDM	SGL	CWF	ASPIRE
0.05	**0.0177**	0.0193	0.0313	0.0376
0.01	**0.0383**	0.0527	0.0637	0.273

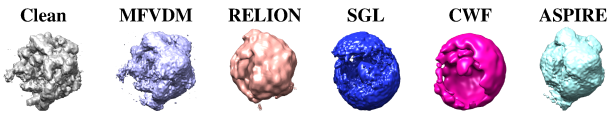


Fig. 4: 3D *ab initio* models of 70S ribosome at SNR $= 0.01$.

methods. The averaged mean squared errors (MSEs) between the denoised images and the clean projections are shown in Tab. 1. Our method achieves the lowest MSEs at all SNR levels. To further demonstrate the importance of denoising, we use a common-lines based algorithm [21] to generate the 3D *ab initio* models from 10^3 denoised images of each method for SNR $= 0.01$ (see Fig. 4). Fig. 4 shows that only our approach obtains a comparatively accurate model.

Experimental dataset: We also test our method on the experimental dataset HIV-1 Nef [22] which is available in the Electron Microscope Pilot Image Archive (EMPIAR) [23]. Samples of noisy and clean projections are shown in the first column of Fig. 3. Again, we take $n = 10^4$ images as input to perform denoising by different methods. As shown in Fig. 3, visually the result of our approach is cleaner than the results of other methods. In Fig. 5, we show the 3D *ab initio* models generated from the denoised images and compare them to the reference volume. In addition to visual inspection, we use the Fourier shell correlation (FSC) [2] between the reconstruction and the original volume to quantify the resolution, which is determined by the 0.143 cutoff of FSC (see Tab. 2). Our models achieve the highest resolutions.

Runtime: In Tab. 3, we report the runtime of different approaches on each dataset for denoising 10^4 images.

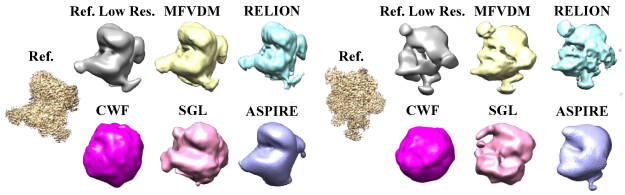


Fig. 5: 3D *ab initio* models of HIV-1 Nef. Two views are shown here. 'Ref.' and 'Ref. Low Res.' represent the original reference volume the Gaussian filtered volume,

Table 2: The volume resolutions (lower is better).

Dataset	MFVDM	RELION	CWF	SGL	ASPIRE
70S rib.	**14.8Å**	25.4Å	44.4Å	39.5Å	17.7Å
HIV-1 Nef	**8.69Å**	9.78Å	39.12Å	39.12Å	16.76Å

Table 3: The runtime (in minutes) of different methods on 70S ribosome (SNR $= 0.01$) and HIV-1 Nef datasets.

dataset	MFVDM	RELION	SGL	CWF	ASPIRE
70S rib.	12	117	12	**11**	15
HIV-1 Nef	**32**	127	71	36	49

4. CONCLUSION

In this paper, we present a novel framework for cryo-EM image denoising by using MFVDM for improving the identification and alignment of images with similar views. In addition, denoising and CTF correction are efficiently performed on the truncated Fourier-Bessel expansion coefficients of images which avoids interpolation error occurred in directly rotating and averaging Cartesian grid images. Our method outperforms existing methods in terms of the quality of the denoised images as well as the resulting 3D *ab initio* reconstruction, especially at low SNR.

5. REFERENCES

[1] Y. Fan and Z. Zhao, "Multi-frequency vector diffusion maps," in *International Conference on Machine Learning*, pp. 1843–1852, 2019.

[2] J. Frank, *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. Oxford University Press, 2006.

[3] R. Langlois, J. Pallesen, J. T. Ash, D. N. Ho, J. L. Rubinstein, and J. Frank, "Automated particle picking for low-contrast macromolecules in cryo-electron microscopy," *Journal of structural biology*, vol. 186, no. 1, pp. 1–7, 2014.

[4] A. Heimowitz, J. Andén, and A. Singer, "Apple picker: Automatic particle picking, a low-effort cryo-em framework," *Journal of structural biology*, vol. 204, no. 2, pp. 215–227, 2018.

[5] L. Wang, A. Singer, and Z. Wen, "Orientation determination of cryo-EM images using least unsquared deviations," *SIAM journal on imaging sciences*, vol. 6, no. 4, pp. 2450–2483, 2013.

[6] Z. Kam, "The reconstruction of structure from electron micrographs of randomly oriented particles," *Journal of Theoretical Biology*, vol. 82, pp. 15–39, 1980.

[7] Z. Zhao and A. Singer, "Rotationally invariant image representation for viewing direction classification in cryo-EM," *Journal of structural biology*, vol. 186, no. 1, pp. 153–166, 2014.

[8] S. H. Scheres, "Relion: Implementation of a Bayesian approach to cryo-EM structure determination," *Journal of Structural Biology*, vol. 180, pp. 519–530, 2012.

[9] J. De la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldivar, J. Vargas, J. Carazo, and C. Sorzano, "Xmipp 3.0: an improved software suite for image processing in electron microscopy," *Journal of structural biology*, vol. 184, no. 2, pp. 321–328, 2013.

[10] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, "EMAN2: an extensible image processing suite for electron microscopy," *Journal of structural biology*, vol. 157, no. 1, pp. 38–46, 2007.

[11] A. Singer, Z. Zhao, Y. Shkolnisky, and R. Hadani, "Viewing angle classification of cryo-electron microscopy images using eigenvectors," *SIAM Journal on Imaging Sciences*, vol. 4, no. 2, pp. 723–759, 2011.

[12] A. Singer and H.-t. Wu, "Vector diffusion maps and the connection Laplacian," *Communications on pure and applied mathematics*, vol. 65, no. 8, pp. 1067–1144, 2012.

[13] Z. Zhao, Y. Shkolnisky, and A. Singer, "Fast steerable principal component analysis," *IEEE transactions on computational imaging*, vol. 2, no. 1, pp. 1–12, 2016.

[14] P. W. Jones, A. Osipov, and V. Rokhlin, "Randomized approximate nearest neighbors algorithm," *Proceedings of the National Academy of Sciences*, vol. 108, no. 38, pp. 15679–15686, 2011.

[15] Y. Fan, T. Gao, and Z. Zhao, "Representation theoretic patterns in multi-frequency class averaging for three-dimensional cryo-electron microscopy," *Inference and Information: A Journal of IMA*, 2021.

[16] Z. Zhao and A. Singer, "Fourier–bessel rotational invariant eigenimages," *JOSA A*, vol. 30, no. 5, pp. 871–877, 2013.

[17] A. Singer, Y. Shkolnisky, and B. Nadler, "Diffusion interpretation of nonlocal neighborhood filters for signal denoising," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 118–139, 2009.

[18] G. Gilboa, N. Sochen, and Y. Y. Zeevi, "Forward-and-backward diffusion processes for adaptive image enhancement and denoising," *IEEE transactions on image processing*, vol. 11, no. 7, pp. 689–703, 2002.

[19] T. Bhamre, T. Zhang, and A. Singer, "Denoising and covariance estimation of single particle cryo-EM images," *Journal of structural biology*, vol. 195, no. 1, pp. 72–81, 2016.

[20] B. Landa and Y. Shkolnisky, "The steerable graph laplacian and its application to filtering image datasets," *SIAM Journal on Imaging Sciences*, vol. 11, no. 4, pp. 2254–2304, 2018.

[21] L. Wang and A. Singer, "Exact and stable recovery of rotations for robust synchronization," *Information and Inference: A Journal of the IMA*, vol. 2, no. 2, pp. 145–193, 2013.

[22] K. L. Morris, C. Z. Buffalo, C. M. Stürzel, E. Heusinger, F. Kirchhoff, X. Ren, and J. H. Hurley, "HIV-1 Nefs are cargo-sensitive AP-1 trimerization switches in tetherin downregulation," *Cell*, vol. 174, no. 3, pp. 659–671, 2018.

[23] A. Iudin, P. K. Korir, J. Salavert-Torres, G. J. Kleywegt, and A. Patwardhan, "EMPIAR: a public archive for raw electron microscopy image data," *Nature methods*, vol. 13, no. 5, pp. 387–388, 2016.