Denoising and Covariance Estimation of Single Particle Cryo-EM Images

Tejal Bhamre

Department of Physics, Princeton University, Jadwin Hall, Washington Road, Princeton, NJ 08544-0708, USA

Teng Zhang

Department of Mathematics, University of Central Florida, 4393 Andromeda Loop N, Orlando, FL 32816-8007, USA

Amit Singer

Department of Mathematics and PACM, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA

Abstract

The problem of image restoration in cryo-EM entails correcting for the effects of the Contrast Transfer Function (CTF) and noise. Popular methods for image restoration include 'phase flipping', which corrects only for the Fourier phases but not amplitudes, and Wiener filtering, which requires the spectral signal to noise ratio. We propose a new image restoration method which we call 'Covariance Wiener Filtering' (CWF). In CWF, the covariance matrix of the projection images is used within the classical Wiener filtering framework for solving the image restoration deconvolution problem. Our estimation procedure for the covariance matrix is new and successfully corrects for the CTF. We demonstrate the efficacy of CWF by applying it to restore both simulated and experimental cryo-EM images. Results with experimental datasets demonstrate that CWF provides a good way to evaluate the particle images and to see what the dataset contains even without 2D classification and averaging.

Keywords: CTF Correction, Steerable PCA, Wiener Filtering

1. Introduction

Single particle reconstruction (SPR) using cryo-electron microscopy (cryo-EM) is a rapidly advancing technique for determining the structure of biological macromolecules at near-atomic resolution directly in their native state, without any need for crystallization [1, 2, 3, 4, 5]. In SPR, 3D reconstructions are estimated by combining multiple noisy 2D tomographic projections of macro-molecules in different unknown orientations.

The acquired data consists of multiple micrographs from which particle images are extracted in the first step of the computational pipeline. Next, the images are grouped together by similarity in the 2D classification and averaging step [6, 7]. Class averages can be used to inspect the underlying particles, and to estimate viewing angles and form a low resolution ab-initio 3D model. Subsequently, this 3D model is refined to high resolution, and 3D classification might be performed as well.

In this paper we propose an image restoration method that provides a way for visualizing the particle images without performing any 2D classification. While noise reduction is achieved in 2D classification by averaging together different particle images, our method operates on each image separately, and performs contrast transfer function (CTF) correction and denoising in a single step.

Existing image restoration techniques (for denoising and CTF correction) can be broadly categorized into two kinds of approaches [8]. The first is an approach known as 'phase flipping', which involves flipping the sign of the Fourier coefficients at frequencies for which the CTF is negative. Consequently, phase flipping restores the correct phases of the Fourier coefficients, but ignores the effect of the CTF on the amplitudes. Phase flipping preserves the noise statistics and is easy to implement, leading to its widespread usage in several cryo-EM software packages. However, it is suboptimal because it does not restore the correct Fourier amplitudes of the images. The second commonly used approach is Wiener filter based restoration, to which we refer here as traditional Wiener filtering (TWF). Wiener filtering takes into account both the phases and am-

plitudes of the Fourier coefficients, unlike phase flipping. However, calculation of the Wiener filter coefficients requires prior estimation of the spectral signal to noise ratio (SSNR) of the signal, which by itself is a challenging problem. It is therefore customary to either treat the SSNR as a precomputed constant as in the software package SPIDER [9], or to apply Wiener filtering only at later stages of the 3D reconstruction pipeline when the noise level is sufficiently low, such as in EMAN2 [10]. It is also possible to use a combination of the two approaches, by first phase flipping the 2D images, and later correct only for the amplitudes in the 3D reconstruction step, as in IMAGIC [11, 12]. Despite its simplicity, there are several drawbacks to TWF. First, it cannot restore information at the zero crossings of the CTF. Second, it requires estimation of the SSNR. Third, it is restrictive to the Fourier basis which is a fixed basis not adaptive to the image dataset.

We refer to our proposed method as Covariance Wiener Filtering (CWF). CWF consists of first estimating the CTF-corrected covariance matrix of the underlying clean 2D projection images, followed by application of the Wiener filter to denoise the images. Unlike phase flipping, CWF takes into account both the phases and magnitudes of the images. Moreover, unlike TWF that always operates in the data-independent Fourier domain, CWF is performed in the data-dependent basis of principal components (i.e., eigenimages). Crucially, CWF can be applied at preliminary stages of data processing on raw 2D particle images. The resulting denoised images can be used for an early inspection of the dataset, to identify the associated symmetry, and to eliminate 'bad' particle images prior to 2D classification and 3D reconstruction. Additionally, the estimation of the 2D covariance matrix is itself of interest, for example, in Kam's approach for 3D reconstruction [13, 14].

The paper is organized as follows: sections 2.1 and 2.2 detail the estimation of the covariance matrix for two different noise models, first for the simpler model of white noise, and second for the more realistic model of colored noise. In section 2.3 we discuss the steerability property of the covariance matrix [15]. The associated deconvolution problem is solved to obtain denoised

images using the estimated covariance matrix in section 2.4. Finally in section 3, we demonstrate CWF in a number of numerical experiments, with both simulated and experimental datasets. We obtain encouraging results for experimental datasets, in particular, those acquired with the modern direct electron detectors. Image features are clearly observed after CWF denoising. For reproducibility, the MATLAB code for CWF and its dependencies are available in the open source cryo-EM toolbox ASPIRE at `www.spr.math.princeton.edu`. The script *cwf_script.m*, calls the main function *cwf.m*.

2. Methods

The first step of CWF is estimation of the covariance matrix of the underlying clean images, to which we refer as the population covariance. The second step of CWF is solving a deconvolution problem to recover the underlying clean images using the estimated covariance. In the rest of this section, we describe these steps in detail.

2.1. The Model

The image formation model in cryo-EM under the linear, weak phase approximation [16] is given by

$$y_i = a_i * x_i + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where n is the number of images, $*$ denotes the convolution operation, y_i is the noisy, CTF filtered i'th image in real space, x_i is the underlying clean projection image in real space, a_i is the point spread function of the microscope that convolves with the clean image in real space, and ϵ_i is additive Gaussian noise that corrupts the image, for each i. Taking the Fourier transform of eqn. 1 gives

$$Y_i = A_i X_i + \xi_i, \quad i = 1, 2, \ldots, n \tag{2}$$

4

where Y_i, X_i and ξ_i are now in Fourier space. A_i is a diagonal operator, whose diagonal consists of the Fourier transform of the point spread function, and is also commonly known as the CTF. The CTF modulates the phases and the amplitudes of the Fourier coefficients of the image, and contains numerous zero crossings that correspond to frequencies at which no information is obtained. Any image restoration technique that aims to completely correct for the CTF must therefore correctly restore both the phases and the amplitudes. The zero crossings make CTF correction challenging since it cannot be trivially inverted. In experiments, different groups of images are acquired at different defocus values, in the hope that information that is lost from one group could be recovered from another group that has different zero crossings. In the experimental datasets used in this paper, the number of images per defocus group typically ranges from 50 to 1000.

In our statistical model, the Fourier transformed clean images $X_1, \ldots, X_n$ (viewed, for mathematical convenience, as vectors in $\mathbb{C}^p$, where p is the number of pixels) are assumed to be independent, identically distributed (i.i.d.) samples from a distribution with mean $\mathbb{E}[\mathbf{X}] = \mu$ and covariance $\mathbb{E}[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T] = \Sigma$. Since the clean images are two-dimensional projections of the three-dimensional molecule in different orientations, the distribution of $\mathbf{X}$ in our model is determined by the three-dimensional structure, the distribution of orientations, the varying contrast due to changes in ice thickness, and structural variability, all of course unknown at this stage. The covariance matrix Σ therefore represents the overall image variability due to these determinants. While these model assumptions do not necessarily hold in reality [17, 18], they simplify the analysis and, as will be shown later lead to excellent denoising. Quoting George Box, "All models are wrong but some are useful" [19].

Our denoising scheme requires μ and Σ. Since these quantities are not readily given, we estimate them from the noisy images themselves as follows. For simplicity, we first assume that the noise in our model is additive white Gaussian noise such that $\xi_i \sim \mathcal{N}(0, \sigma^2 I_{p \times p})$ in eqn. 2 are i.i.d. The white noise assumption is later replaced by that of the more realistic colored noise. First,

5

notice from eqn. 2 it follows that

$$\mathbb{E}[\mathbf{Y}_i] = A_i \mathbb{E}[\mathbf{X}_i], \quad i = 1, 2, \ldots, n. \tag{3}$$

So,

$$\begin{aligned}
\mathbb{E}[(\mathbf{Y}_i - \mathbb{E}[\mathbf{Y}_i])(\mathbf{Y}_i - \mathbb{E}[\mathbf{Y}_i])^T] &= \mathbb{E}[A_i(\mathbf{X}_i - \mu)(\mathbf{X}_i - \mu)^T A_i^T] + \sigma^2 I \\
&= A_i \Sigma A_i^T + \sigma^2 I.
\end{aligned} \tag{4}$$

Eqn. 4 relates the second order statistics of the noisy images with the population covariance Σ of the clean images, based on which we can estimate Σ.

Next, we construct estimators for the mean μ and population covariance Σ using eqn. 3 and 4. The mean μ of the dataset can be estimated as the solution to a least squares problem

$$\hat{\mu} = \arg\min_{\mu} \sum_{i=1}^{n} ||(Y_i - A_i \mu)||_2^2 + \lambda ||\mu||_2^2 \tag{5}$$

where $\lambda \geq 0$ is a regularization parameter. The solution to 5 is explicitly

$$\hat{\mu} = (\sum_{i=1}^{n} A_i^T A_i + \lambda I)^{-1} (\sum_{i=1}^{n} A_i^T Y_i). \tag{6}$$

The population covariance Σ can be estimated as

$$\begin{aligned}
\hat{\Sigma} &= \arg\min_{\Sigma} \sum_{i=1}^{n} ||(Y_i - \mathbb{E}[\mathbf{Y}_i])(Y_i - \mathbb{E}[\mathbf{Y}_i])^T - (A_i \Sigma A_i^T + \sigma^2 I)||_F^2 \\
&= \arg\min_{\Sigma} \sum_{i=1}^{n} ||A_i \Sigma A_i^T + \sigma^2 I - C_i||_F^2
\end{aligned} \tag{7}$$

where $C_i = (Y_i - A_i \mu)(Y_i - A_i \mu)^T$ and $||.||_F$ is the Frobenius matrix norm. The estimators $\hat{\mu}$ and $\hat{\Sigma}$ can be shown to be consistent in the large sample limit $n \to \infty$, similar to the result in Appendix B of [20].

To ensure that the estimated covariance is positive semidefinite (PSD), we project it onto the space of PSD matrices by computing its spectral decomposition and retaining only the non negative eigenvalues (and their corresponding

eigenvectors). To solve eqn. 7, we differentiate the objective function with respect to Σ and set the derivative to zero. This yields

$$\sum_{i=1}^{n} A_i^T A_i \hat{\Sigma} A_i^T A_i = \sum_{i=1}^{n} A_i^T C_i A_i - \sum_{i=1}^{n} \sigma^2 A_i^T A_i \tag{8}$$

Eqn. 8 defines a system of linear equations for the elements of the matrix $\hat{\Sigma}$. However, direct inversion of this linear system is slow and computationally impractical for large image sizes. Notice that eqn. 8 can be written as

$$L(\hat{\Sigma}) = B \tag{9}$$

where $L : \mathbb{R}^{p \times p} \to \mathbb{R}^{p \times p}$ is the linear operator acting on $\hat{\Sigma}$ defined by the left hand side of eqn. 8, and B is the right hand side. Since applying L only involves matrix multiplications, it can be computed fast, and the conjugate gradient method is employed to efficiently compute $\hat{\Sigma}$ instead of direct inversion, similar to how it is used in [21].

Notice that $L(\hat{\Sigma})$ is a PSD matrix whenever $\hat{\Sigma}$ is PSD (as a sum of PSD matrices), while B may not necessarily be PSD due to finite sample fluctuations (i.e., n is finite). It is therefore natural to project B onto the cone of PSD matrices. This amounts to computing the spectral decomposition of B and setting all negative eigenvalues to 0, which is an instance of eigenvalue thresholding.

We now describe an alternate eigenvalue thresholding procedure, better suited to cases in which the number of images n is not exceedingly large. To that end, we first analyze the matrix B when $X_i = 0$ for all i, i.e., the input images are white noise images containing no signal. Let

$$M = \sum_{i=1}^{n} A_i^T C_i A_i = \sum_{i=1}^{n} A_i^T Y_i Y_i^T A_i. \tag{10}$$

Then, $\mathbb{E}[M] = \sigma^2 \sum_{i=1}^{n} A_i^T A_i$ and $B = M - \mathbb{E}[M]$. Let $S = (\mathbb{E}[M])^{1/2}$, i.e. S is PSD and $\mathbb{E}[M] = S^2$. Then multiplying both sides of eqn. 9 with S^{-1} we get

$$S^{-1} L(\hat{\Sigma}) S^{-1} = S^{-1} (M - \mathbb{E}[M]) S^{-1} = S^{-1} M S^{-1} - I. \tag{11}$$

$S^{-1}MS^{-1}$ can be viewed as a sample covariance matrix of n vectors in $\mathbb{R}^p$ whose population covariance is the identity matrix. When p is fixed and n goes to infinity, all eigenvalues of $S^{-1}MS^{-1}$ converge to 1. In practice, however, n and p are often comparable. In the limit $p, n \to \infty$ and $p/n \to \gamma$ with $0 < \gamma < \infty$, the limiting spectral density of the eigenvalues converges to the Marčenko Pastur (MP) distribution [22], given by

$$MP(x) = \frac{1}{2\pi} \frac{\sqrt{(\gamma_+ - x)(x - \gamma_-)}}{\gamma x} 1_{[\gamma_-, \gamma_+]}, \quad \gamma_\pm = (1 \pm \sqrt{\gamma})^2 \qquad (12)$$

for $\gamma \leq 1$. It is therefore expected that $S^{-1}MS^{-1}$ would have eigenvalues (considerably) larger than 1, even in the pure white noise case. These large eigenvalues should not be mistakingly attributed to signal. In the case of images containing signal (plus noise), eigenvalues corresponding to the signal can only be detected if they reside outside of the support of the MP distribution. We use the method of [23] to determine the number of eigenvalues corresponding to the signal. We then apply the operator norm eigenvalue shrinkage procedure (see [24]) to those eigenvalues, while setting all other eigenvalues to 0. We then use the conjugate gradient method [1] to solve eqn. 11 for $\hat{\Sigma}$, with the right hand side replaced with its shrinkage version. We observed in numerical simulations (see Fig. 3) that this procedure typically outperforms other shrinkage methods in terms of the accuracy of the estimated covariance matrix.

2.2. Covariance Estimation with Colored noise

So far, we assumed additive white Gaussian noise in the image formation process. In reality, the noise in experimental images is colored. That is, in the image formation model in eqn. 2, ξ_i is additive colored Gaussian noise. We preprocess the images in order to "whiten" the noise. The noise power

---

[1] While L in eqn. 9 is PSD, the new effective operator in the LHS of eqn. 11 is not necessarily PSD in general. In order to use conjugate gradient, we solve the system $S^{-1}L(S^{-1}\Sigma_S S^{-1})S^{-1} = S^{-1}MS^{-1} - I$, where $\Sigma_S = S\Sigma S$, in which the operator acting on Σ_S in the LHS is PSD. Σ is then obtained from the estimated Σ_S.

spectrum can be estimated, for example, using the pixels in the corners of the noisy projection images. To do this, we first estimate using correlograms the 2D autocorrelation of the corner pixels of the images which contain mostly noise and no signal. These corner pixels are used to estimate the 1D autocorrelation, which is then extended to populate the 2D isotropic autocorrelation. We then calculate the Fourier transform of the 2D autocorrelation, which is the 2D power spectrum of noise. The noisy projection images in Fourier space are multiplied element-wise by the inverse of the estimated power spectral density, also called the whitening filter, so that the noise in the resulting images is approximately white. Let W be the "whitening" filter, such that

$$WY_i = WA_iX_i + W\xi_i, \quad i = 1, 2, \ldots, n \tag{13}$$

and $W\xi_i \sim \mathcal{N}(0, \sigma^2 I)$.

Eqn. 13 is reminiscent of eqn. 2. It is tempting to define a new "effective" CTF as WA_i and estimate Σ following the same procedure as in the case of white noise. However, the linear system akin to eqn. 8 for this case is ill-conditioned due to the product of W with the CTF, and it takes a large number of iterations for conjugate gradient to converge to the desired solution. Instead, we seek an approach in which the linear system to solve is well conditioned as that in the case of white noise. Since the CTF's A_i, $i = 1, 2, \ldots n$ and the whitening filter W are diagonal operators in the Fourier basis, they commute, and eqn. 13 becomes

$$WY_i = A_iWX_i + W\xi_i, \quad i = 1, 2, \ldots, n. \tag{14}$$

We therefore absorb W into X_i, and estimate the matrix $\Sigma_W = W\Sigma W^T$ (the population covariance of $W\mathbf{X}$) using the same procedure as before. The population covariance Σ is then estimated as

$$\hat{\Sigma} = W^{-1}\hat{\Sigma}_W(W^T)^{-1}. \tag{15}$$

2.3. Fourier-Bessel Steerable PCA

The population covariance matrix Σ must be invariant under in-plane rotation of the projection images, therefore it is block diagonal in any steerable basis in which the basis elements are outer products of radial functions and angular Fourier modes. Following [15], we choose to represent the images in a Fourier-Bessel basis and it suffices to estimate each diagonal block $\Sigma^{(k)}$, corresponding to the angular frequency k, separately. The Fourier-Bessel basis [15] consists of p_k basis functions (that satisfy the sampling criterion) for each angular frequency k, where p_k decreases with increasing k. The matrix $\Sigma^{(k)}$ is thus of size $p_k \times p_k$.

An important property of the CTF's A_i and the whitening filter W is that they are radially isotropic [2]. Therefore, the CTF's and the whitening filter are also block diagonal in the Fourier Bessel basis. Eqn. 8 (and its analog in the case of colored noise) is hence solved separately for each k to estimate $\Sigma^{(k)}$.

2.4. Wiener Filtering

The estimated covariance is further used to solve the associated deconvolution problem in eqn. 2 using Wiener filtering. The result is a denoised, CTF corrected image for each noisy, CTF affected measurement Y_i for $i = 1, 2, \ldots n$. We estimate X_i in the white noise model using the Wiener filtering procedure as

$$\hat{X}_i = (I - H_i A_i)\hat{\mu} + H_i Y_i \tag{16}$$

where $H_i = \hat{\Sigma} A_i^T (A_i \hat{\Sigma} A_i^T + \sigma^2 I)^{-1}$ is the linear Wiener filter [25]. In the case of colored noise,

$$\hat{X}_i = (I - H_i W A_i)\hat{\mu} + H_i Y_i \tag{17}$$

with $H_i = \hat{\Sigma} A_i^T W^T (W A_i \hat{\Sigma} A_i^T W^T + \sigma^2 I)^{-1}$. Since the estimated covariance is block-diagonal in the Fourier Bessel basis, the Wiener filtering procedure is

---

[2]In the case of astigmatism, where the CTF deviates slightly from radial isotropy, this is a good approximation to obtain low resolution denoised images.

applied to the Fourier Bessel coefficients of the noisy images Y_i for each angular frequency k separately. The denoised Fourier Bessel expansion coefficients are used to reconstruct denoised images in Fourier space that are inverse Fourier transformed to acquire images in real space on a Cartesian grid.

2.5. Computational Complexity

In practice, instead of each image being affected by a distinct CTF, all images within a given defocus group have the same CTF. So, given D defocus groups with d_i images in group i, one can equivalently minimize the objective function $\sum_{i=1}^{D} d_i ||(A_i \Sigma A_i^T + \sigma^2 I) - \sum_{j=1}^{d_i} \frac{1}{d_i}(Y_{i_j} - \mathbb{E}[\mathbf{Y}_{i_j}])(Y_{i_j} - \mathbb{E}[\mathbf{Y}_{i_j}])^T||_F^2$ in eqn. 7 (here A_i denotes the CTF of the i'th defocus group, and i_j index images in that group). As a result, the sums in eqn. 8 range from 1 to D instead of from 1 to n, thereby reducing the computational cost of some operations. For images of size $L \times L$, estimating the mean using eqn. 6 takes $O(nL^2)$ (since A_i is diagonal in the Fourier basis for each i). Computing the Fourier Bessel expansion coefficients takes $O(nL^3)$, as detailed in [15]. When solving the linear system in eqn. 8 to estimate each $\Sigma^{(k)}$ separately, the matrices in eqn. 8 are of size $p_k \times p_k$. It is shown in [15] that $\sum_k p_k = O(L^2)$, $\sum_k p_k^2 = O(L^3)$, and $\sum_k p_k^3 = O(L^4)$. While solving eqn. 9 using conjugate gradient for a given angular frequency, computing the action of the linear operation L on $\Sigma^{(k)}$ takes $O(Dp_k^3)$ per iteration, while computing B takes $O(Dp_k^3 + np_k^2)$. Thus, each iteration of conjugate gradient takes $O(D\sum_k p_k^3)$, that is, $O(DL^4)$ and there is also a one time computation of $O(nL^3)$. Wiener filtering the Fourier Bessel coefficients of an image for a given angular frequency k takes $O(p_k^2)$. So the overall complexity for Wiener filtering the coefficients of all images is $O(nL^3)$. In summary, the overall complexity for CWF is $O(TDL^4 + nL^3)$, where T is the number of conjugate gradient iterations.

3. Results

In this section, we apply our algorithm to synthetic and experimental datasets to obtain denoised images. All algorithms are implemented in the UNIX envi-

ronment, on a machine with 60 cores, running at 2.3 GHz, with total RAM of 1.5TB. We perform numerical experiments with (i) a synthetic dataset with additive white and colored Gaussian noise and (ii) four experimental datasets, two of which were acquired with older detectors, and the other two with state-of-the-art direct electron detectors. For all the experimental datasets, the corresponding estimated CTF parameters were provided with the dataset. For all simulations, we use centered projection images. The algorithm does not require centered images. However, having non-centered images would result in an additional 'blurring' effect in the denoised images.

3.1. Simulated Noisy Dataset with White Noise

For the first experiment with simulated data, we construct a synthetic dataset by modeling the image formation process in cryo-EM. The synthetic dataset is prepared from the 3D structure of the P. falciparum 80S ribosome bound to E-tRNA, available on the Electron Microscopy Data Bank (EMDB) as EMDB-6454. We first generate clean 2D projection images starting from a 3D volume, at directions sampled uniformly over the sphere, and then corrupt the generated clean projection images with different CTF's and additive white Gaussian noise. The projection images are divided into 10 defocus groups, with the defocus value ranging from $1\mu m$ to $4\mu m$. The B-factor of the decay envelope was chosen as $10\mathring{A}^2$, the amplitude contrast as 7%, the voltage as 300kV, and the spherical aberration as 2mm. To ensure that the denoising quality of CWF is robust to the mean estimation of the dataset, the regularization parameter λ in the least squares mean estimation in eqn. 6 was fixed at 1 for all the experiments described here.

Figure 1 shows the results of denoising raw, CTF-affected noisy images with CWF and TWF at various levels of the SNR. We have used the EMAN2 [10] implementation of TWF (note that we perform phase flipping followed by TWF only on the raw images in EMAN2, and not on averages). The SNR used here is defined relative to the CTF affected images that constitute the clean signal, and is calculated as an average value for the entire dataset. Using 20
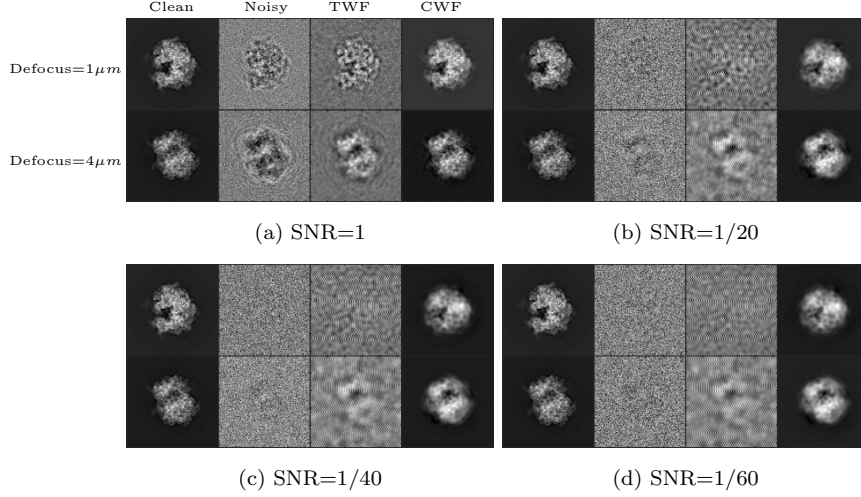
12

Figure 1: **Synthetic white noise**: A comparison of the denoising results of traditional Wiener filtering (TWF) and CWF for the synthetic dataset prepared from EMDB-6454, the P. falciparum 80S ribosome bound to E-tRNA. The dataset consists of 10000 images of size 105×105, which are divided into 10 defocus groups, with the defocus value ranging from $1\mu m$ to $4\mu m$. The two rows in each subfigure correspond to two clean images belonging to different defocus groups; the first one belongs to the group with the smallest defocus value of $1\mu m$, while the second image belongs to the group with the largest defocus value of $4\mu m$.

cores, calculating the Fourier Bessel coefficients took 79 seconds while covariance estimation and Wiener filtering together took 6 seconds in the experiment with SNR= 1/60.

It is seen that TWF works very well at high SNR (≥ 1), but deteriorates at lower SNR's as expected. Note that the denoising results of TWF depend strongly on the defocus value. The location of the zeros in the CTF is such that images corresponding to high defocus values preserve low frequency information, while images corresponding to low defocus values retain more high frequency information. With CWF, there is no such strong dependence on the defocus value, since the covariance matrix is estimated using information from all defocus groups.

Figure 2a shows the relative MSE of denoised images as a function of the SNR of the dataset. The MSE (norm of the difference between the denoised
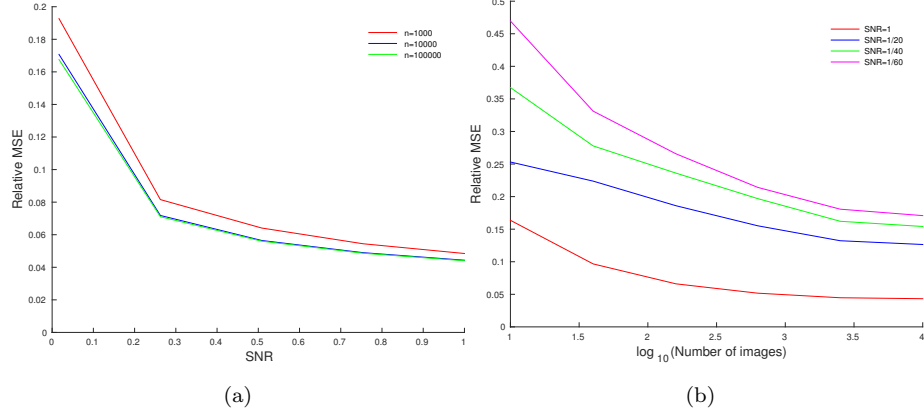
Figure 2: (a) **Relative MSE versus the SNR, for a fixed number of images**: The relative MSE of the denoised images as a function of the SNR, for synthetic data generated using EMDB-6454. The MSE reported here is averaged over all images. n denotes the number of images used in the experiment.(b) **Relative MSE versus the number of images, for a fixed SNR**: The relative MSE of the denoised images as a function of the number of images, for synthetic data generated using EMDB-6454. The MSE reported here is averaged over all images.

image and the original, clean image) shown here corresponds to the same range of SNR's (from 1/60 to 1) as in Figure 1. Figure 2b shows the relative MSE of the denoised images as a function of the number of images used to estimate the covariance in the experiment. The covariance estimation improves as the number of images in the dataset increases, and so the denoising is also expected to improve, as seen from Figure 2b.

The importance of the eigenvalue shrinkage procedure is elucidated in Figure 3. Here, we compare the error in the estimated covariance with and without eigenvalue shrinkage, for varying number of images used in the experiment. The relative MSE of the estimated covariance $\hat{\Sigma}$ is defined as

$$MSE_{rel} = \frac{||\Sigma - \hat{\Sigma}||_F^2}{||\Sigma||_F^2} \tag{18}$$
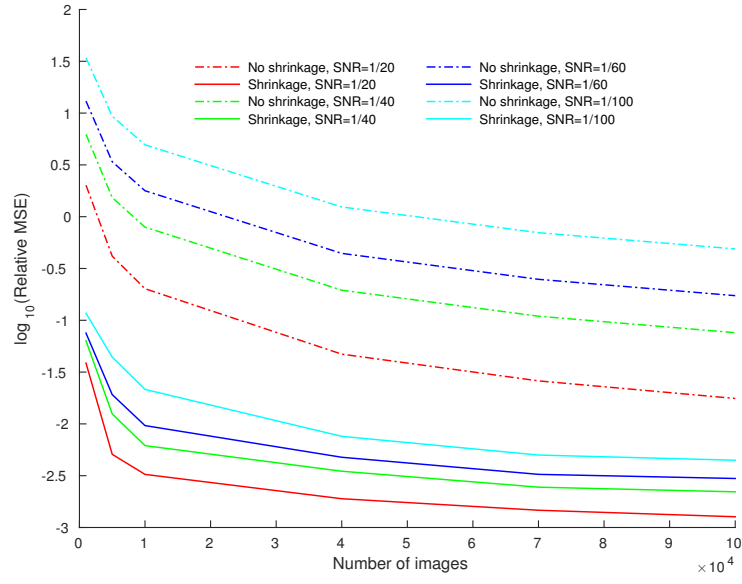
14

Figure 3: **Relative MSE of the estimated covariance versus the number of images**: The relative MSE of the estimated covariance $\hat{\Sigma}$, with and without using eigenvalue shrinkage, as a function of number of images, for synthetic data generated using EMDB-6454.
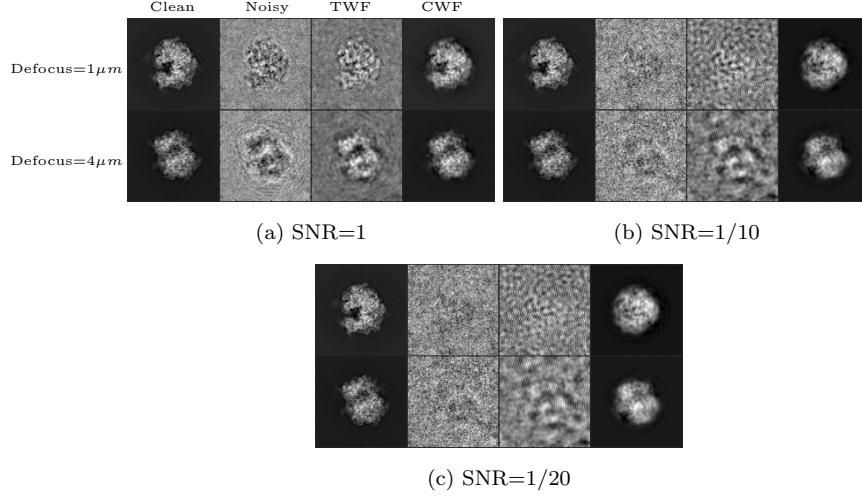
(a) SNR=1

(b) SNR=1/10

(c) SNR=1/20

Figure 4: **Synthetic colored noise**: Denoising results of CWF for the synthetic dataset with additive colored Gaussian noise, prepared from EMDB-6454, the P. falciparum 80S ribosome bound to E-tRNA, as detailed in the caption of Figure 1.

3.2. Simulated Noisy Dataset with Colored Noise

The noise that corrupts images in cryo-EM is not perfectly white, but often colored. To simulate this, we perform experiments with synthetic data generated from EMDB-6454 as described in 3.1, this time adding colored Gaussian noise with the noise response $f(\mathrm{k}) = \frac{1}{\sqrt{(1+\mathrm{k}^2)}}$ (k is the radial frequency) to each clean, CTF-affected projection image. Figure 4 shows the denoised images for this case.

16

3.3. Experimental Dataset - TRPV1

We apply CWF to an experimental dataset of the TRPV1 ion channel, taken using a K2 direct electron detector. It is available on the public database Electron Microscope Pilot Image Archive (EMPIAR) as EMPIAR-10005, and the 3D reconstruction is available on EMDB as EMDB-5778, courtesy of Liao et al. [26]. The dataset consists of 35645 motion corrected, picked particle images of size 256×256 pixels with a pixel size of $1.2156\mathring{A}$. Using 20 cores, calculating the Fourier Bessel coefficients took 312 seconds while covariance estimation and Wiener filtering together took 574 seconds. The result is shown in Figure 5. CWF retains 384 eigenvalues of Σ.
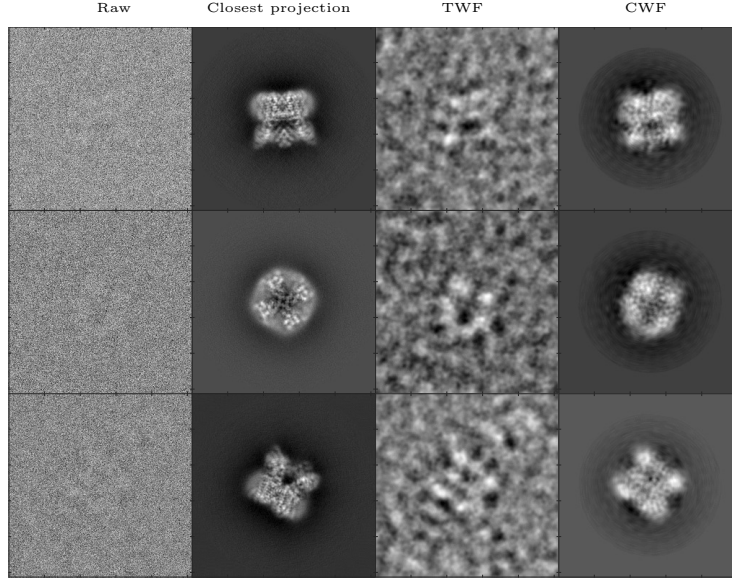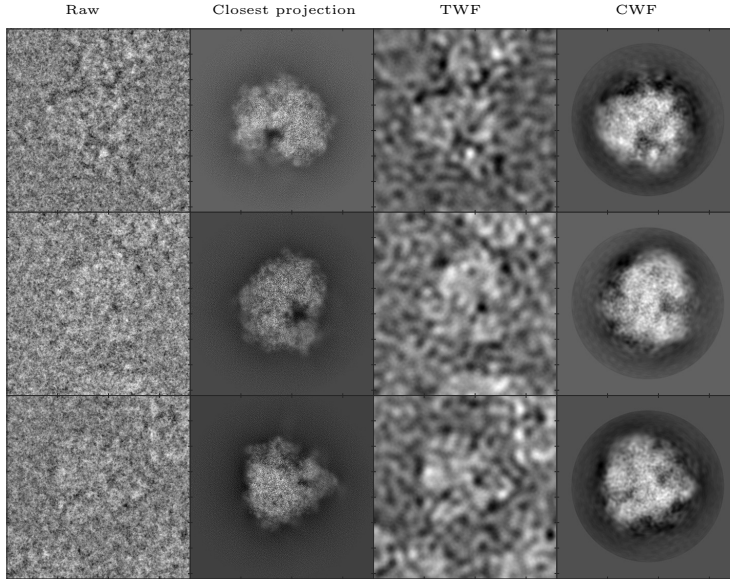

Figure 5: **Denoising an experimental dataset of TRPV1 [26]**: Here we show, for three images in the dataset, the raw image, the closest true projection image generated from the 3D reconstruction of the molecule (EMDB 5778), the denoised image obtained using TWF, and the denoised image obtained using CWF. In this experiment, 35645 images of size 256×256 belonging to 935 defocus groups were used. The amplitude contrast is 10%, the spherical aberration is 2mm, and the voltage is 300kV.

3.4. Experimental Dataset - 80S ribosome

We apply CWF to an experimental dataset of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine, taken using a FEI FALCON II 4k × 4k direct electron detector. The raw micrographs and picked particles are available on the public database EMPIAR as EMPIAR-10028, and the 3D reconstruction is available on EMDB as EMDB-2660, courtesy of Wong et al. [27]. The dataset we used was provided by Dr. Sjors Scheres, and consists of 105247 motion corrected, picked particle images of size 360×360 with a pixel size of 1.34Å. Using 20 cores, calculating the Fourier Bessel coefficients took 731 seconds while covariance estimation and Wiener filtering together took 385 seconds. The result is shown in Figure 6. CWF retains 962 eigenvalues of Σ.
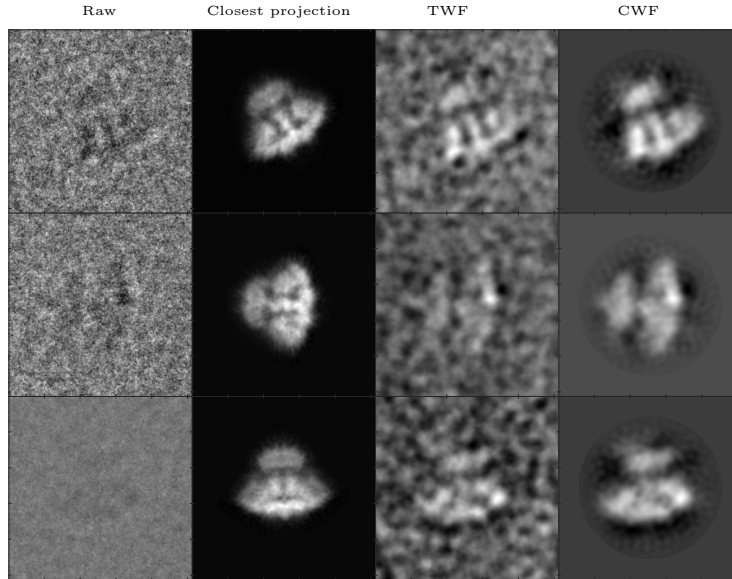

Figure 6: **Denoising an experimental dataset of the 80S ribosome** [**27**]: Here we show, for three images in the dataset, the raw image, the closest true projection image generated from the 3D reconstruction of the molecule (EMDB 2660), the denoised image obtained using TWF, and the denoised image obtained using CWF. In this experiment, the first 30000 images out of the 105247 images in the dataset were used for covariance estimation. The images are of size 360×360 and belong to 290 defocus groups. The amplitude contrast is 10%, the spherical aberration is 2mm, and the voltage is 300kV.

3.5. Experimental Dataset - IP_3R1

We apply CWF to an experimental dataset of the Inositol 1, 4, 5-triphosphate receptor 1 (IP_3R1) provided by Dr. Irina Serysheva, obtained using the older Gatan 4k × 4k CCD camera [28]. The 3D reconstruction obtained from this dataset is available on EMDB as EMDB-5278. The dataset consists of 37382 images of size 256×256 pixels with a pixel size of $1.81\mathring{A}$. Using 20 cores, calculating the Fourier Bessel coefficients took 429 seconds while covariance estimation and Wiener filtering together took 589 seconds. The result is shown in Figure 7. CWF retains 290 eigenvalues of Σ.
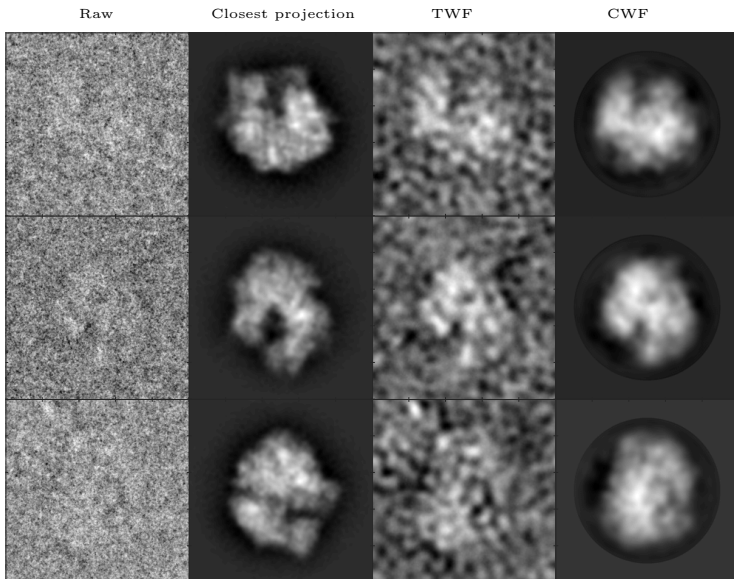

Figure 7: **Denoising an experimental dataset of IP_3R1 [28]**: Here we show, for three images in the dataset, the raw image, the closest true projection image generated from the 3D reconstruction of the molecule (EMDB 5278), the denoised image obtained using TWF, and the denoised image obtained using CWF. In this experiment, 37382 images of size 256×256 belonging to 851 defocus groups were used. The amplitude contrast is 15%, the spherical aberration is 2mm, and the voltage is 200kV.

3.6. Experimental Dataset - 70S ribosome

We apply CWF to an experimental dataset of the 70S ribosome provided by Dr. Joachim Frank's group [29]. This heterogeneous dataset consists of 216517

images of size 250×250 pixels with a pixel size of $1.5\mathring{A}$, obtained using the older TVIPS TEMCAM-F415 (4k x 4k) CCD detector. The 3D reconstruction obtained from this dataset is available on EMDB as EMDB-5360. Using 20 cores, calculating the Fourier Bessel coefficients took 1174 seconds while covariance estimation and Wiener filtering together took 113 seconds. The result is shown in Figure 8. CWF retains 219 eigenvalues of Σ.

Figure 8: **Denoising an experimental dataset of 70S [29]**: Here we show, for three images in the dataset, the raw image, the closest true projection image generated from the 3D reconstruction of the molecule (EMDB 5360), the denoised image obtained using TWF, and the denoised image obtained using CWF. In this experiment, the first 99979 images out of the 216517 images in the dataset were used for covariance estimation. The images are of size 250×250 and belong to 38 defocus groups. The amplitude contrast is 10%, the spherical aberration is 2.26mm, and the voltage is 300kV.

3.7. Outlier Detection

In the cryo-EM pipeline, a significant amount of time is spent on discarding outliers by visual inspection after the particle picking step. CWF provides an automatic way to classify picked particles into "good" particles and outliers.

The classifier uses the contrast of a denoised image to determine if it is an outlier.

The specimen particles can be at various depths in the ice layer at the time of imaging, so the acquired projection images can have different contrasts. The contrast can be modeled as an additional scalar parameter α for each acquired noisy projection image as in eqn. 19, typically as a uniformly distributed random variable spread about its mean at 1.

$$Y_i = \alpha_i A_i X_i + \xi_i, \quad i = 1, 2, \ldots, n \tag{19}$$

We absorb the contrast α into $\mathbf{X}$ and estimate $\alpha_i X_i$ in this case, using the same procedure as before. We perform an experiment with synthetic data generated using EMDB-6454 with additive colored Gaussian noise at SNR=1/20, and $\alpha \in [0.75, 1.5]$. 10% of the projection images are replaced by "outliers", that is, pure noise images containing no signal. Fig. 9c shows the estimated mean image μ, and Fig. 9d shows the top 6 principal components of the estimated covariance $\hat{\Sigma}$, also known as eigenimages. Fig. 9a and Fig. 9b show a sample of raw and denoised images respectively. High contrast images enjoy a higher SNR and are thus of interest for subsequent steps of the pipeline. On the other hand, outlier images, which typically have low contrast after denoising, can be automatically detected by a linear classifier after CWF and discarded from the dataset. In the experiment shown in Fig. 9a and 9b, a classifier with a threshold of 0.95 for the contrast discards 95% of the outliers, while 3% of the inliers are also discarded in the process.

One can also use a different classifier based on features like the relative energy of the image before and after denoising, etc. However, outliers that look like particles, for example, images belonging to a different class of a heterogeneous dataset which act as "contaminants", are difficult to detect using this method.
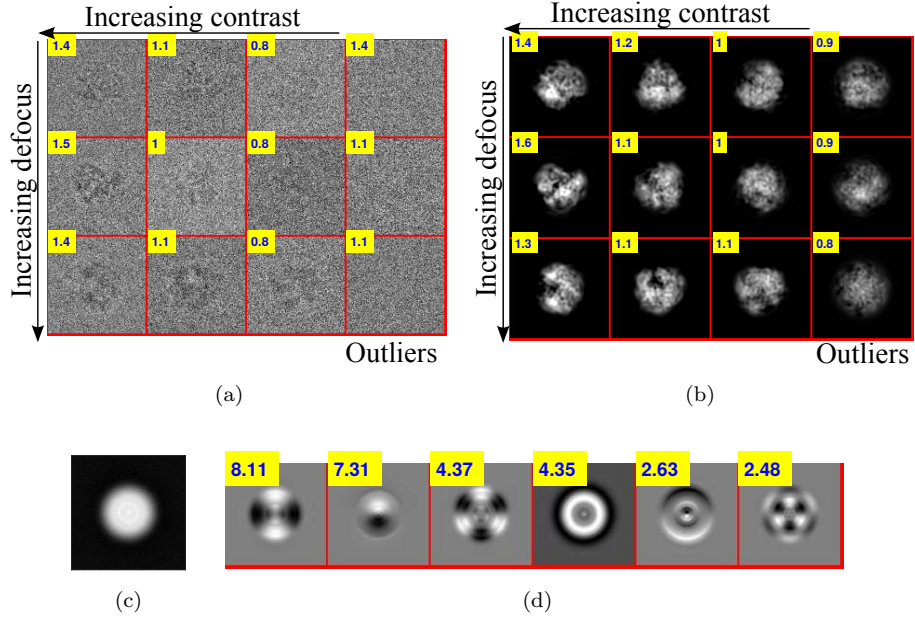
21

Figure 9: (a) **Raw images**: A sample of synthetic data generated using EMDB-6454 with additive colored Gaussian noise at SNR=1/20. 10% of the projection images are replaced by pure noise. The contrast parameter α ranges from 0.75 to 1.5. The outliers are shown in the last column. Inset in a yellow box is the contrast of each image. (b) **Denoised images**: The denoised images using CWF. Notice the low contrast outliers in the last column. (c) **Estimated Mean Image** (d) **Top 6 eigenimages**: Inset in a yellow box is the corresponding eigenvalue.

4. Conclusion

In this paper we presented a new approach for image restoration of cryo-EM images, CWF, whose main algorithmic components are covariance estimation and deconvolution using Wiener filtering. CWF performs both CTF correction, by correcting the Fourier phases and amplitudes of the images, as well as denoising, by eliminating the noise thereby improving the SNR of the resulting images. In particular, since CWF applies Wiener filtering in the data-dependent basis of principal components ("eigenimages"), while TWF applies Wiener filtering in the data-independent Fourier basis, we see in numerical experiments that CWF performs better than TWF, and considerably better at high noise levels. We demonstrated the ability of CWF to restore images for several experimental datasets, acquired with both CCD detectors and the state-of-the-art direct electron detectors.

Due to the high noise level typical in cryo-EM images, 2D classification is performed before estimating a 3D ab-initio model. Class averages enjoy a higher SNR and are used to estimate viewing angles and obtain an initial model. For future work, it remains to be seen whether the resulting denoised images from CWF can be directly used to estimate viewing angles, without performing classification and averaging. Another possible future direction is integration of CWF into existing 2D class averaging procedures in order to improve their performance.

5. Acknowledgements

[1] X. C. Bai, G. McMullan, and S. H. Scheres. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.*, 40(1):49–57, Jan 2015.

[2] J. L. Milne, M. J. Borgnia, A. Bartesaghi, E. E. Tran, L. A. Earl, D. M. Schauder, J. Lengyel, J. Pierson, A. Patwardhan, and S. Subramaniam. Cryo-electron microscopy–a primer for the non-microscopist. *FEBS J.*, 280(1):28–45, Jan 2013.

[3] Eva Nogales. The development of cryo-em into a mainstream structural biology technique. *Nat Meth*, 13(1):24–27, Jan 2016. Historical Commentary.

[4] Fred J. Sigworth. Principles of cryo-em single-particle image processing. *Microscopy*, 2015.

[5] Werner Kühlbrandt. Cryo-em enters a new era. *eLife*, 3:e03678, Aug 2014. 25122623[pmid].

[6] Zhizhen Zhao and Amit Singer. Rotationally invariant image representation for viewing direction classification in cryo-em. *Journal of Structural Biology*, 186(1):153 – 166, 2014.

[7] Wooram Park and Gregory S. Chirikjian. An assembly automation approach to alignment of noncircular projections in electron microscopy. *IEEE Transactions on Automation Science and Engineering*, 11(3):668 – 679, 2014.

[8] Pawel A. Penczek. Chapter two - image restoration in cryo-electron microscopy. In Grant J. Jensen, editor, *Cryo-EM, Part B: 3-D Reconstruction*, volume 482 of *Methods in Enzymology*, pages 35 – 72. Academic Press, 2010.

[9] J. Frank, M. Radermacher, P. Penczek, J. Zhu, Y. Li, M. Ladjadj, and A. Leith. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.*, 116(1):190–199, 1996.

[10] Guang Tang, Liwei Peng, Philip R. Baldwin, Deepinder S. Mann, Wen Jiang, Ian Rees, and Steven J. Ludtke. Eman2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1):38 – 46, 2007. Software tools for macromolecular microscopy.

[11] M. van Heel, G. Harauz, E. V. Orlova, R. Schmidt, and M. Schatz. A new generation of the IMAGIC image processing system. *J. Struct. Biol.*, 116(1):17–24, 1996.

[12] M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, and A. Patwardhan. Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys.*, 33(4):307–369, Nov 2000.

[13] Z. Kam. The reconstruction of structure from electron micrographs of randomly oriented particles. *Journal of Theoretical Biology*, 82(1):15 – 39, 1980.

[14] T. Bhamre, Teng Zhang, and A. Singer. Orthogonal matrix retrieval in cryo-electron microscopy. In *IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 1048–1052, April 2015.

[15] Z. Zhao, Y. Shkolnisky, and A. Singer. Fast steerable principal component analysis. *IEEE Transactions on Computational Imaging*, 2(1):1–12, March 2016.

[16] Joachim Frank. Chapter 3 - two-dimensional averaging techniques. In *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, pages 54 – 125. Academic Press, Burlington, 1996.

[17] C.O.S. Sorzano, S. Jonic, R. Nez-Ramrez, N. Boisset, and J.M. Carazo. Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function. *Journal of Structural Biology*, 160(2):249 – 262, 2007.

[18] Carlos O.S. Sorzano, Javier Vargas, Joaqun Otn, Vahid Abrishami, Jos M. de la Rosa-Trevn, Sandra del Riego, Alejandro Fernndez-Alderete, Carlos Martnez-Rey, Roberto Marabini, and Jos M. Carazo. Fast and accurate conversion of atomic models into electron density maps. *AIMS Biophysics*, 2(20150102):8–20, 2015.

[19] George E. P. Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.

[20] E. Katsevich, A. Katsevich, and A. Singer. Covariance matrix estimation for the cryo-em heterogeneity problem. *SIAM Journal on Imaging Sciences*, 8(1):126–185, 2015.

[21] Joakim Andén, Eugene Katsevich, and Amit Singer. ovariance estimation using conjugate gradient for 3d classification in CRYO-EM. In *IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 200–204, April 2015.

[22] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483, April 1967.

[23] Shira Kritchman and Boaz Nadler. Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems*, 94:19–32, 2008.

[24] David Donoho, Matan Gavish, and I M Johnstone. Optimal shrinkage of eigenvalues in the spiked covariance model. *CoRR*, abs/1311.0851, 2014.

[25] D. J. C. MacKay. Chapter 46 - deconvolution. In *Information Theory, Inference and Learning Algorithms*, pages 550 – 551. Cambridge University Press, Cambridge, UK, 2004.

[26] Maofu Liao, Erhu Cao, David Julius, and Yifan Cheng. Structure of the trpv1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478):107112, December 2013.

[27] Wilson Wong, Xiao-chen Bai, Alan Brown, Israel S. Fernandez, Eric Hanssen, Melanie Condron, Yan Hong Tan, Jake Baum, and Sjors HW Scheres. Cryo-em structure of the plasmodium falciparum 80s ribosome bound to the anti-protozoan drug emetine. *eLife*, 3:e03080, Jun 2014. 24913268[pmid].

[28] StevenJ. Ludtke, ThaoP. Tran, QueT. Ngo, VeraYu. Moiseenkova-Bell, Wah Chiu, and IrinaI. Serysheva. Flexible architecture of IP$_3$R1 by cryo-em. *Structure*, 19(8):1192 – 1199, 2011.

[29] Xabier Agirrezabala, Hstau Y. Liao, Eduard Schreiner, Jie Fu, Rodrigo F. Ortiz-Meoz, Klaus Schulten, Rachel Green, and Joachim Frank. Structural characterization of mrna-trna translocation intermediates. *Proceedings of the National Academy of Sciences*, 109(16):6094–6099, 2012.