Robust w-Estimators for Cryo-EM Class Means

Chenxi Huang, *Student Member, IEEE*, and Hemant D. Tagare, *Senior Member, IEEE*

Abstract—A critical step in cryogenic electron microscopy (cryo-EM) image analysis is to calculate the average of all images aligned to a projection direction. This average, called the class mean, improves the signal-to-noise ratio in single-particle reconstruction. The averaging step is often compromised because of the outlier images of ice, contaminants, and particle fragments. Outlier detection and rejection in the majority of current cryo-EM methods are done using cross-correlation with a manually determined threshold. Empirical assessment shows that the performance of these methods is very sensitive to the threshold. This paper proposes an alternative: a w-estimator of the average image, which is robust to outliers and which does not use a threshold. Various properties of the estimator, such as consistency and influence function are investigated. An extension of the estimator to images with different contrast transfer functions is also provided. Experiments with simulated and real cryo-EM images show that the proposed estimator performs quite well in the presence of outliers.

Index Terms—Electron microscopy, single particle reconstruction, robust estimation, class averaging, three-dimensional reconstruction.

I. INTRODUCTION

C RYOGENIC electron microscopy (cryo-EM) is a relatively new imaging technique that aims to reconstruct the three-dimensional (3D) structure of biological macromolecules (e.g. protein molecules), called *particles*, from their two-dimensional (2D) projection images [2]. Cryo-EM does not require crystallization and the particles are maintained in their native hydrated state. These advantages, however, come at a cost: cryo-EM images are extremely noisy with signal-to-noise ratios (SNR) commonly below 0dB. The SNR is improved by averaging images that are believed to be from the same projection direction. The resulting average images are called *class means*.

In practice, many of the images that participate in the averaging turn out to be outliers. They arise from structures that are unrelated to the molecule (e.g., ice, contaminants) [3]. The presence of outliers at the averaging step compromises the fidelity of the reconstruction. Detecting such outlier images is a key problem in cryo-EM reconstruction [4].

C. Huang is with the Department of Biomedical Engineering, Yale University, New Haven, CT 06520 USA (e-mail: chenxi.huang@yale.edu).

H. D. Tagare is with the Department of Diagnostic Radiology, Electrical Engineering and Biomedical Engineering, Yale University, New Haven, CT 06520 USA (e-mail: hemant.tagare@yale.edu).

Current outlier detection strategies in cryo-EM are mostly based on cross-correlating the images with one or more templates [5], [6]. Images with cross-correlation below a threshold are regarded as outliers. This approach is found in popular cryo-EM packages such as EMAN, SPIDER, Bsoft and SIGNATURE [7]–[10]. This classical approach has a serious drawback: its performance is sensitive to the threshold. The useful range of the threshold is narrow, difficult to find, and dependent on the image SNR. The false negative rate and the false positive rate increase rapidly as the threshold deviates from the optimal range [11]. Quite often the threshold has to be manually adjusted to the SNR of the image [7], [12], [13]. Manual adjustment is also problematic because outliers are often not visible to naked eye due to the low SNR.

In this paper, we propose an alternative approach that does not rely on thresholding. Drawing on classical robust estimation theory, we present a "w-estimator" for class means that is robust to outliers. The estimator is based on an outlier model specific to cryo-EM. Theoretical analysis shows that the influence function of the estimator is bounded for typical cryo-EM outliers. The reader should be aware that the analysis of the influence function is rather complex. One part of the analysis is in closed form, while one part is numerical. The numerical part is generic in the sense that it can be applied to any class mean, and is not restricted to a specific example. We also show that the estimator is Fisher-consistent.

A few methods in cryo-EM literature use w-estimator-like techniques, e.g., FREALIGN calculates the 3D reconstruction by a weighted sum of all images [14]. A similarity measure in the form of weighted frequency components is proposed to align cryo-EM images [15]. To our knowledge, these estimators are not explicitly derived in a w-estimator robust estimation theory framework. Their outlier model is not explicit, and theoretical properties such as consistency and robustness are not addressed. The parameters used in these algorithms are also set in an ad-hoc manner.

Common-lines based outlier detection methods correlate images with particle images from all projection directions on the common-lines [16]. Although such a method incorporates more images than the classical approach, it has similar drawbacks resulting from using thresholds for outlier rejection.

Approaches not based on thresholds are also possible. By classifying cryo-EM images into heterogeneous classes (e.g. using RELION [17]) and analyzing the reconstructed classes, classes that are visually inconsistent with the known structures of the molecule components are rejected as outliers [18], [19].

Extensive simulations are provided to compare the performances of the w-estimator and conventional outlier detection.

Additionally, experiments with real cryo-EM data are also reported. They show that the w-estimator can suppress outliers in real-world cases.

Cryo-EM images are affected by the microscope contrast transfer function (CTF). To provide some insensitivity to the CTF, the particle is often imaged at different CTFs. We extend the w-estimator to this case, so that the class mean can be calculated from images acquired at different CTFs. Fisher-consistency and boundedness of the influence function for outliers also hold for this estimator.

The rest of the paper is organized as follows. Section II contains background information on cryo-EM. This section is included for readers who may be unfamiliar with cryo-EM and is also meant for fixing terminology and for stating the two problems we address. Section III contains the proposed estimator for the single CTF case and provides theoretical analysis of the estimator. Section IV extends the estimator to multiple CTFs. Section V shows the results of using the proposed w-estimators with simulated and real cryo-EM data. Section VI concludes the paper. Proofs of all claims are available in the Appendix.

II. THE CRYO-EM CLASS MEAN PROBLEM

A. Cryo-EM and Reconstruction

In cryo-EM, projection images, called *micrographs* containing copies of the same particle embedded in vitreous ice are obtained. Sub-images of individual particles, referred to as *particle images*, are extracted from the micrograph using a template. Imperfections in the micrograph are caused by incompletely formed or disassembled particles due to sample preparation and contaminants. These "non-particles" are often mistakenly selected as particle images.

The formation of a cryo-EM image can be modeled as a tomographic projection of the particle from a random direction, convolved with a filter, followed by additive noise [2]. The kernel of filter is called the contrast transfer function (CTF). CTF arises from the interaction of the electron beam, the molecule, and the ice. Theoretical analysis suggests that CTF has a Fourier transform that is real, circularly symmetric, and taking positive and negative values (Fig.9b). The CTF (the zeros of the CTF) can be changed using the defocus setting of the microscope.

Particle images are used in an iterative procedure to reconstruct the 3D structure of the particle. Fig.1 illustrates the reconstruction process using class means. There are reconstruction methods such as FREALIGN [14], that do not use class means for reconstruction (nevertheless, even for these methods the calculation of a class mean can be used for outlier detection, as this paper shows). Starting from an initial estimate of the structure, projections are obtained from many fixed directions and CTFs are applied to each projection. After allowing for in-plane rotation and translation, every particle image is associated with the most similar CTF-filtered projection, and the aligned images are averaged to give a *class mean*. Correlation or its variants are often used as similarity measure for alignment [20], [21]. The class means are backprojected to obtain a new estimate of the structure
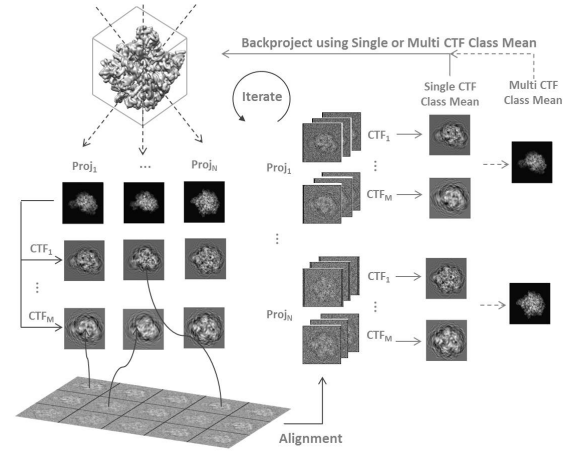

Fig. 1.   Outline of reconstruction procedure.

for the next iteration. An alternative reconstruction strategy is also possible: For every projection direction, the class means at different CTFs can be combined into a CTF-corrected class mean which is then backprojected to give the reconstruction.

Ideally, only particle images which come from a projection direction are aligned to that direction. In reality, some non-particle images plus some particle images from neighboring projection directions are also aligned (misaligned images). We refer to these images as *outliers* and the proposed method deals with these outlier images.

B. The Two Class Mean Estimation Problems

We can now state the two problems that we address in this paper:

1) Single CTF Class Mean: Given a set of images aligned to a single CTF at a projection direction, estimate the class mean robustly in the presence of outliers.

2) Multi CTF Class Mean: Given a set of images aligned to more than one CTF at a projection direction, estimate the CTF-corrected class mean robustly in the presence of outliers.

C. The Outlier Model

To proceed, we need a mathematical model for outliers. Recall that outliers are images of non-particles and misaligned particles and that outliers are detected according to their correlation coefficients with a reference during both particle extraction and alignment. Under this condition, the signal in the outlier has a low, but non-zero correlation coefficient with the matching template. The outlier model used below is based on this observation.

We also assume that the image noise is white. Although cryo-EM images do not follow this assumption exactly, pre-whitening filters are often designed and used [22].

III. CLASS MEAN ESTIMATION FOR SINGLE CTF

A. Image Model

Suppose that $\mathbf{x} \in \mathbb{R}^p$ is a noisy image of p pixels obtained at a CTF, and aligned to a projection direction:

$$\mathbf{x} = \boldsymbol{\theta}_x + \mathbf{n}, \tag{1}$$

where $\mathbf{n}$ is zero-mean Gaussian white noise with covariance matrix $\sigma^2 I$, and $\boldsymbol{\theta}_x \in \mathbb{R}^p$ is the deterministic but unknown signal in the image. The expression for $\boldsymbol{\theta}_x$ depends on whether the image is an inlier or an outlier:

Inlier: When the image $\mathbf{x}$ is an inlier

$$\boldsymbol{\theta}_x = s\boldsymbol{\theta}, \tag{2}$$

where $\boldsymbol{\theta}$ ($\|\boldsymbol{\theta}\| > 0$) is the non-noisy projection of the particle. $\boldsymbol{\theta}$ incorporates the effect of CTF (and possibly the pre-whitening filter). The amplitude factor s is assumed to have a uniform probability density $\pi(s|a, b)$ in an interval (a, b) which is symmetric about 1. This models real-world change in signal amplitude [23]. The probability density function (pdf) of $\mathbf{x}$ is obtained by marginalizing s:

$$f(\mathbf{x}|\boldsymbol{\theta}) = \int_a^b g(\mathbf{x}|s\boldsymbol{\theta})\pi(s|a, b)\mathrm{d}s \tag{3}$$

where $g(\mathbf{x}|s\boldsymbol{\theta})$ is the pdf of a multivariate normal distribution with mean $s\boldsymbol{\theta}$ and covariance matrix $\sigma^2 I$. The corresponding probability distribution of $\mathbf{x}$ is $F(\mathbf{x}|\boldsymbol{\theta})$, which we simply denote as $F(\mathbf{x})$ or F.

Outlier: Following the discussion in section II-C, the model for the signal $\boldsymbol{\theta}_x$ in an outlier image is

1) $\boldsymbol{\theta}_x$ has a low correlation coefficient with $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the projection of the particle as defined in (2).
2) $\boldsymbol{\theta}_x$ has a finite component along $\boldsymbol{\theta}$.

B. The Robust Estimator of Class-Mean

The problem is to estimate $\boldsymbol{\theta}$ from a collection of images $\mathbf{x}_i \in \mathbb{R}^p$, $i = 1, \cdots, N$, some of which may be outliers. This estimate is the *class mean*. We propose a w-estimator [24] of $\boldsymbol{\theta}$ as the fixed point $\mathbf{T}$ of the weighted average of images:

$$\mathbf{T} = \frac{\sum_{i=1}^N \mathbf{x}_i w(\mathbf{x}_i, \mathbf{T})}{\sum_{i=1}^N w(\mathbf{x}_i, \mathbf{T})}, \tag{4}$$

where $w(\mathbf{x}, \mathbf{T}) \geq 0$ depends on the similarity between image $\mathbf{x}_i$ and $\mathbf{T}$. The weight function, defined below, is chosen such that outliers are given lower weights than the inliers, limiting the outlier's influence on the estimate $\mathbf{T}$. $\mathbf{T}$ is usually determined by starting from an initial estimate $\mathbf{T}^{(0)}$ (e.g., the median of all images), and iterating

$$\mathbf{T}^{(j+1)} = \frac{\sum_{i=1}^N \mathbf{x}_i w(\mathbf{x}_i, \mathbf{T}^{(j)})}{\sum_{i=1}^N w(\mathbf{x}_i, \mathbf{T}^{(j)})} \tag{5}$$

until $\mathbf{T}^{(j)}$ converges.

The specific weight function we use is:

$$w(\mathbf{x}, \mathbf{T}) = \frac{|<\mathbf{x}, \mathbf{T}>|}{\|\mathbf{x}\|\|\mathbf{T}\|}\exp\left\{-\beta\left\|\mathbf{x} - \frac{<\mathbf{x}, \mathbf{T}>}{\|\mathbf{T}\|^2}\mathbf{T}\right\|^2\right\} \tag{6}$$

where $<, >$ is the inner product and β is a constant whose value is discussed later. Some comments on (6):

1) The weight function has two terms: the first term $|<\mathbf{x}, \mathbf{T}>|/(\|\mathbf{x}\|\|\mathbf{T}\|)$ is the absolute value of the correlation coefficient of image $\mathbf{x}$ and $\mathbf{T}$. The second, exponential term is a function of the component of $\mathbf{x}$ orthogonal to $\mathbf{T}$. The first term is responsible for

limiting the effects of outliers. The second term bounds the influence function of the estimator (Section III-D).
2) (6) is undefined when $\|\mathbf{x}\| = 0$, but this happens with probability zero and does not affect the rest of the argument. In practice, cryo-EM images are so noisy that $\|\mathbf{x}\| \gg 0$.
3) The convergence of (5) is difficult to establish. However, in all real cryo-EM cases that we have investigated, (5) converges reliably to a non-zero $\mathbf{T}$.

The weight function (6) is different from classical weight functions of w-estimators, which have the form $w(\mathbf{x}_i - \mathbf{T})$. Consequently, we cannot borrow classical results about consistency, influence function, etc. for our estimator.

C. Fisher-Consistency of the Estimator

A w-estimator is Fisher-consistent if the fixed point $\mathbf{T}$ of (4) equals $\boldsymbol{\theta}$ asymptotically. With a slight abuse of notation we explicitly denote the dependence of $\mathbf{T}$ on F by $\mathbf{T}(F)$, so that

$$\mathbf{T}(F) = \frac{\int \mathbf{x}w(\mathbf{x}, \mathbf{T}(F))\mathrm{d}F(\mathbf{x})}{\int w(\mathbf{x}, \mathbf{T}(F))\mathrm{d}F(\mathbf{x})}. \tag{7}$$

$\mathbf{T}(F)$ is Fisher-consistent if $\mathbf{T}(F) = \boldsymbol{\theta}$ when $\mathbf{x}$ are inliers. Because $\boldsymbol{\theta}$ is a vector in $\mathbb{R}^p$, it has a direction and a norm. We establish the consistency of $\mathbf{T}(F)$ by considering its direction and norm separately. We first investigate the consistency of the estimator in direction and then the consistency of the norm through numerical evaluation, where the parameters such as the dimension, the SNRs of the images are assigned values typical in cryo-EM.

Because (6) only uses inner products and norms, the weight function is independent of the coordinate system. Using any orthonormal coordinate system in $\mathbb{R}^p$ gives the same weight, and hence the same estimate $\mathbf{T}(F)$ in (7). We use a coordinate system in which the direction of $\mathbf{T}(F)$ is the first coordinate axis, i.e., $\mathbf{T}(F) = [T, 0, \cdots, 0]^T$ ($T \neq 0$).

The consistency of the estimator in direction requires showing $\mathbf{T}(F) = \alpha\boldsymbol{\theta}$ where $\alpha > 0$. This is established by using Claim 1 below, which is proved in the Appendix:

Claim 1: For $\mathbf{x} \sim N(\boldsymbol{\theta}, \sigma^2 I)$ where $\mathbf{x} \in \mathbb{R}^p$, if $h(\mathbf{x})$: $\mathbb{R}^p \to \mathbb{R}$ is spherically symmetric, $h(\mathbf{x}) > 0$ almost everywhere and $h(\mathbf{x}) \leq M < \infty$, then $E_\mathbf{x}[\mathbf{x}h(\mathbf{x})] = \alpha\boldsymbol{\theta}$ with $\alpha > 0$.

Using Claim 1, it is straightforward to establish:

Claim 2: Let $\mathbf{T}(F)$ be defined as (7) with the weight function in (6) and F is the distribution of the pdf in (3). If $\mathbf{T}(F) \neq 0$, then $\mathbf{T}(F) = \alpha\boldsymbol{\theta}$ with $\alpha > 0$.

Proof: Consider two cases:

1) $p = 1$. When both $\mathbf{x}$ and $\mathbf{T}$ are scalars, $w(x, T) = 1$. Thus $T = \int x\mathrm{d}F(x) = E_x[x] = \theta$.
2) $p > 1$. Under the aforementioned coordinate system, $\mathbf{T} = [T, \mathbf{0}]^T$ and $\mathbf{x} = [x_1, x_{\tilde{1}}]^T$ where $\mathbf{0}$ is the zero vector of dimension $p - 1$ and $\tilde{1}$ indexes the second to the pth components. Equation (7) can then be written as

$$\mathbf{T}(F) = \int \mathbf{x}\frac{w(\mathbf{x}, \mathbf{T}(F))}{\int w(\mathbf{x}, \mathbf{T}(F))\mathrm{d}F(\mathbf{x})}\mathrm{d}F(\mathbf{x})$$

$$= \int \mathbf{x}\frac{(|x_1|/\|\mathbf{x}\|)e^{-\beta\|x_{\tilde{1}}\|^2}}{\int(|x_1|/\|\mathbf{x}\|)e^{-\beta\|x_{\tilde{1}}\|^2}\mathrm{d}F(\mathbf{x})}f(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{8}$$

Let $h(\mathbf{x}) \triangleq \frac{(|x_1|/\|\mathbf{x}\|)e^{-\beta\|x_{\bar{1}}\|^2}}{\int(|x_1|/\|\mathbf{x}\|)e^{-\beta\|x_{\bar{1}}\|^2}dF(\mathbf{x})}$,

Then equation (8) simplifies to

$$\int \mathbf{x}h(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \int \mathbf{x}h(x)\int_a^b \frac{1}{b-a}g(\mathbf{x}|s\boldsymbol{\theta})dsd\mathbf{x}$$

$$= \int_a^b \frac{1}{b-a}\int \mathbf{x}h(x)g(\mathbf{x}|s\boldsymbol{\theta})d\mathbf{x}ds$$

$$= E_s\big[E_{\mathbf{x}|s}[\mathbf{x}h(\mathbf{x})]\big]. \qquad (9)$$

Note that for a given x_1, $h(\mathbf{x})$ is a spherically symmetric function of $x_{\bar{1}}$ and $h(\mathbf{x}) > 0$ except $x_1 = 0$. Also, because its numerator $(|x_1|/\|\mathbf{x}\|)e^{-\beta\|x_{\bar{1}}\|^2} \leq 1$ and its denominator $\int(|x_1|/\|\mathbf{x}\|)e^{-\beta\|x_{\bar{1}}\|^2}dF(\mathbf{x}) = Q \neq 0$, $h(\mathbf{x}) \leq 1/Q < \infty$. Applying claim 1, $E_{\mathbf{x}|s}[\mathbf{x}h(\mathbf{x})] = \alpha[s\boldsymbol{\theta}]$ with $\alpha > 0$. Thus, (9) can be written as

$$E_s\big[E_{\mathbf{x}|s}[\mathbf{x}h(\mathbf{x})]\big] = E_s[\alpha[s\boldsymbol{\theta}]] = \alpha E_s[s]\boldsymbol{\theta} \triangleq \alpha^*\boldsymbol{\theta},$$

where $\alpha^* > 0$ because $\alpha > 0$ and $s > 0$. ∎

$\mathbf{T}(F)$ is consistent in norm if $\|\mathbf{T}(F)\| = \|\boldsymbol{\theta}\|$. From Claim 2, $\boldsymbol{\theta}$ and $\mathbf{T}(F)$ have the same direction. Using the same coordinate system as Claim 2, we have $\boldsymbol{\theta} = [\theta, 0, \cdots, 0]^T$ (assume $\theta > 0$), $\mathbf{T}(F) = [T, 0, \cdots, 0]^T$ and $\mathbf{x} = [x_1, x_2, \cdots, x_p]^T$. Evaluating (7) gives $\int x_1 w(\mathbf{x}, \mathbf{T}(F))dF(\mathbf{x}) = T\int \mathbf{x}w(\mathbf{x}, \mathbf{T}(F))dF(\mathbf{x})$ and $\int x_i w(\mathbf{x}, \mathbf{T}(F))dF(\mathbf{x}) = 0, i > 1$. Thus

$$\|\mathbf{T}(F)\| = |T| = \frac{\int x_1 w(\mathbf{x}, \mathbf{T}(F))dF(\mathbf{x})}{\int w(\mathbf{x}, \mathbf{T}(F))dF(\mathbf{x})}. \qquad (10)$$

The numerator of (10) is

$$\int x_1 w(\mathbf{x}, \mathbf{T}(F))dF(\mathbf{x}) = \int x_1 w(\mathbf{x}, \mathbf{T}(F))f(\mathbf{x})d\mathbf{x}$$

$$= \int[\int x_1 \frac{|x_1|}{\|\mathbf{x}\|}f(x_1)dx_1]e^{-\beta\sum_2^p x_i^2}\prod_2^p[f(x_i)dx_i]. \qquad (11)$$

After using $\bar{x}_i = \frac{x_i}{\sigma}$, $\bar{\theta} = \frac{\theta}{\sigma}$, $\bar{\beta} = \beta\sigma^2$, where σ is image noise standard deviation, (11) can be expressed as

$$\sigma \int[\int \bar{x}_1 \frac{|\bar{x}_1|}{\|\bar{\mathbf{x}}\|}f(\bar{x}_1)d\bar{x}_1]e^{-(\bar{\beta}+\frac{1}{2})\sum_2^p \bar{x}_i^2}\prod_2^p \frac{1}{\sqrt{2\pi}}d\bar{x}_i. \qquad (12)$$

(12) can be simplified by writing $\bar{x}_2, \cdots, \bar{x}_p$ in spherical coordinates so that $\bar{x}_2 = r\cos\phi_1, \cdots, \bar{x}_p = r\sin\phi_1\cdots\sin\phi_{p-2}$, $d\bar{x}_2\cdots d\bar{x}_p = r^{p-2}\sin^{p-3}\phi_1\cdots\sin\phi_{p-3}drd\phi_1\cdots d\phi_{p-2}$ and setting $\sum_2^p \bar{x}_i^2 = r^2$. Then (12) simplifies to $\sigma Q_1(p)C(p)$, where

$$C(p) \triangleq (\frac{1}{\sqrt{2\pi}})^{p-1}\int \sin^{p-3}\phi_1\cdots\sin\phi_{p-3}\prod_1^{p-2}d\phi_i, \text{ and}$$

$$Q_1(p) \triangleq \int_0^\infty \int \bar{x}_1 \frac{|\bar{x}_1|}{\sqrt{\bar{x}_1^2 + r^2}}f(\bar{x}_1)e^{-(\bar{\beta}+\frac{1}{2})r^2}r^{p-2}d\bar{x}_1dr.$$

Similarly, the denominator of (10) can be written as $Q_2(p)C(p)$, where

$$Q_2(p) \triangleq \int_0^\infty \int \frac{|\bar{x}_1|}{\sqrt{\bar{x}_1^2 + r^2}}f(\bar{x}_1)e^{-(\bar{\beta}+\frac{1}{2})r^2}r^{p-2}d\bar{x}_1dr.$$
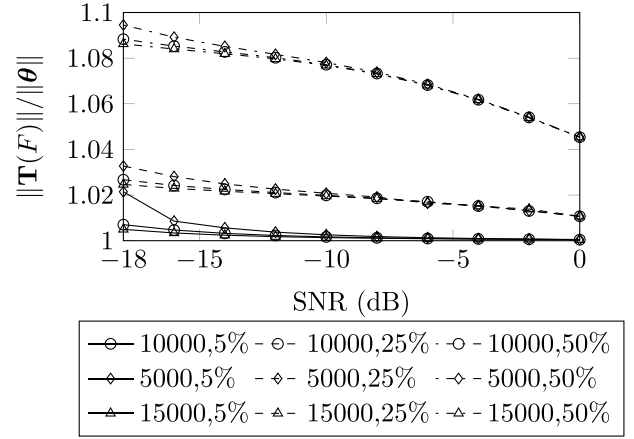

Fig. 2. Consistency of $\|\mathbf{T}(F)\|$. Image size $p = 5000, 10000, 15000$ and contrast variation of 5%, 25% and 50% are considered.

Thus $\|\mathbf{T}(F)\| = \sigma Q_1(p)/Q_2(p)$, so that

$$\frac{\|\mathbf{T}(F)\|}{\|\boldsymbol{\theta}\|} = \frac{Q_1(p)}{Q_2(p)} \times \frac{1}{\|\bar{\boldsymbol{\theta}}\|} \qquad (13)$$

where, the term $\|\bar{\boldsymbol{\theta}}\| = \|\boldsymbol{\theta}\|/\sigma$ is related to the SNRs of the images by SNR $= 10\log_{10}(\|\bar{\boldsymbol{\theta}}\|^2/p)$.

We can investigate $\|\mathbf{T}(F)\|/\|\boldsymbol{\theta}\|$ for different SNRs using (13). Q_1 and Q_2 can be evaluated numerically for typical values of p and of contrast variation a and b in cryo-EM (a and b determine F, see (3)). We consider $p = 5000, 10000, 15000$ which loosely correspond to images of size 64×64, 100×100 and 128×128. We also consider three ranges of contrast variation: 5% ($a = 0.95$ and $b = 1.05$), 25% ($a = 0.75$ and $b = 1.25$), and 50% ($a = 0.5$ and $b = 1.5$). For each combination of image size p and contrast variation, we numerically evaluated $Q_1(p)$ and $Q_2(p)$, and hence $\|\mathbf{T}(F)\|/\|\boldsymbol{\theta}\|$ using (13) for SNRs between -18dB to 0dB. The results are plotted in Fig.2. The figure shows that the difference between $\|\mathbf{T}(F)\|$ and $\|\boldsymbol{\theta}\|$ is less than 10% for all values of p and is negligible when amplitude variation is 25% or smaller. Thus, as a close approximation we may regard $\|\mathbf{T}(F)\|$ and hence $\mathbf{T}(F)$ as consistent.

D. The Influence Function

We investigate the influence function of the estimator for outliers. The influence function of $\mathbf{T}(F)$ is defined as [25]:

$$IF(\mathbf{x}; \mathbf{T}, F) = \lim_{\epsilon \to 0} \frac{\mathbf{T}(F_\epsilon) - \mathbf{T}(F)}{\epsilon} = \frac{\partial}{\partial\epsilon}[\mathbf{T}(F_\epsilon)]_{\epsilon=0} \quad (14)$$

where $F_\epsilon = (1-\epsilon)F + \epsilon\Delta_\mathbf{x}$ is a contaminated distribution by a point mass at $\mathbf{x}$. Replacing F in (7) by F_ϵ, we have

$$\big[\int w(\mathbf{y}, \mathbf{T}(F_\epsilon))dF_\epsilon\big]\mathbf{T}(F_\epsilon) = \int \mathbf{y}w(\mathbf{y}, \mathbf{T}(F_\epsilon))dF_\epsilon. \quad (15)$$
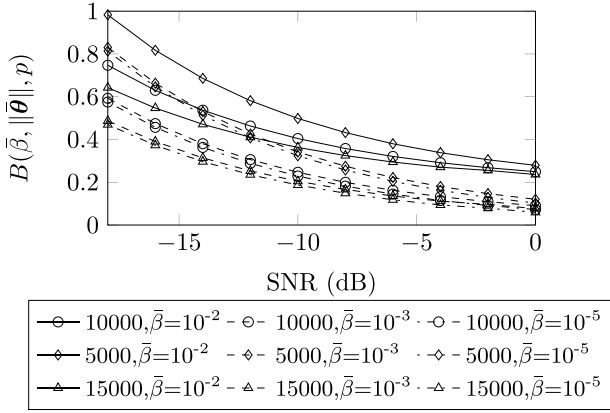
Fig. 3. Numerical evaluation of $B(\bar{\beta}, \|\bar{\theta}\|, p)$. B is evaluated for SNR = −18dB to 0dB (converted from $\|\theta\|$) and $p = 5000, 10000, 15000$ and for $\bar{\beta} = 10^{-2}, 10^{-3}, 10^{-5}$ and 50% contrast variation ($a = 0.5$ and $b = 1.5$).

Taking the derivative of both sides of (15) with respect to ϵ and evaluating at 0 gives

$$w(\mathbf{x}, \mathbf{T}(F))\mathbf{T}(F) - \int w(\mathbf{y}, \mathbf{T}(F))\mathrm{d}F$$
$$+ \left[\int \left[w(\mathbf{y}, \mathbf{T}(F)) + \mathbf{T}(F)\frac{\partial}{\partial \mathbf{t}}[w(\mathbf{y}, \mathbf{t})]_{\mathbf{T}(F)} \right]\mathrm{d}F \right] \frac{\partial}{\partial \epsilon}[\mathbf{T}(F_\epsilon)]_{\epsilon=0}$$
$$= \mathbf{x}w(\mathbf{x}, \mathbf{T}(F)) - \int \mathbf{y}w(\mathbf{y}, \mathbf{T}(F))\mathrm{d}F$$
$$+ \left[\int \mathbf{y}\frac{\partial}{\partial \mathbf{t}}[w(\mathbf{y}, \mathbf{t})]_{\mathbf{T}(F)}\mathrm{d}F \right] \frac{\partial}{\partial \epsilon}[\mathbf{T}(F_\epsilon)]_{\epsilon=0}.$$

Dividing both sides by $\int w(\mathbf{y}, \mathbf{T}(F))\mathrm{d}F$ and applying consistency $\mathbf{T}(F) = \theta$ gives

$$[I - H(\theta)]\frac{\partial}{\partial \epsilon}[\mathbf{T}(F_\epsilon)]_{\epsilon=0} = \frac{w(\mathbf{y}, \theta)(\mathbf{x} - \theta)}{\int w(\mathbf{y}, \theta)\mathrm{d}F},$$

where

$$H(\theta) \triangleq \frac{\int (\mathbf{y} - \theta)\frac{\partial}{\partial \mathbf{t}}[w(\mathbf{y}, \mathbf{t})]_\theta}{\int w(\mathbf{y}, \theta)\mathrm{d}F}$$

and $H(\theta)$ is a matrix of size $p \times p$ (p is the image size). If $I - H(\theta)$ is invertible, the influence function is

$$IF(\mathbf{x}; \mathbf{T}, F) = \frac{\partial}{\partial \epsilon}[\mathbf{T}(F_\epsilon)]_{\epsilon=0} = [\mathbf{I} - H(\theta)]^{-1}\frac{w(\mathbf{x}, \theta)(\mathbf{x} - \theta)}{\int w(\mathbf{y}, \theta)\mathrm{d}F}.$$

The condition that $I - H(\theta)$ is invertible is critical for the existence of the influence function. A simple sufficient condition for $I - H(\theta)$ to be invertible is $\|H(\theta)\|_{op} < 1$ where $\|\cdot\|_{op}$ denotes the operator norm. In the Appendix, we show that $\|H(\theta)\|_{op} \leq B(\bar{\beta}, \|\bar{\theta}\|, p)$, where $\bar{\beta} = \beta\sigma^2$ with β being the parameter in (6) and σ the standard deviation of image noise. Fig.3 is a plot of $B(\bar{\beta}, \|\bar{\theta}\|, p)$ for the same values of p and SNRs as used in Fig.2. We consider three values of $\bar{\beta}$: $10^{-2}, 10^{-3}, 10^{-5}$ and 50% contrast variation ($a = 0.5$ and $b = 1.5$). B is shown as a monotonically decreasing function of SNR. Further, $B < 1$ for all values of p when $\bar{\beta}$ is less than 10^{-3}. We use $\bar{\beta} = 10^{-5}$ for all our experiments. Note that this value of $\bar{\beta}$ does not need tuning; it may be considered as a fixed constant.

Having established that $I - H(\theta)$ is invertible, we can evaluate the influence function for cryo-EM outliers. Recall that an outlier $\mathbf{x}$ in cryo-EM has a low correlation coefficient with θ and a finite component along θ. Let the outlier $\mathbf{x}$ have a component $x_1 \neq 0$ along θ and a component x_2 orthogonal to θ, and the absolute value of the correlation coefficient $\phi = \|x_1\|/\|\mathbf{x}\|$. We consider the value of the influence function as $\phi \to 0$:

$$\|IF(\mathbf{x}; \mathbf{T}, F)\| = \left\| [I - H(\theta)]^{-1}\frac{w(\mathbf{x}, \theta)(\mathbf{x} - \theta)}{\int w(\mathbf{y}, \theta)\mathrm{d}F(\mathbf{y})} \right\|$$
$$\leq \frac{\|[I - H(\theta)]^{-1}\|_{op}}{\int w(\mathbf{y}, \theta)\mathrm{d}F(\mathbf{y})}\|w(\mathbf{x}, \theta)(\mathbf{x} - \theta)\|$$
$$= K(\theta)\|w(\mathbf{x}, \theta)(\mathbf{x} - \theta)\|, \quad (16)$$

where $K(\theta) \triangleq \|[I - H(\theta)]^{-1}\|_{op}/\int w(\mathbf{y}, \theta)\mathrm{d}F(\mathbf{y})$. Then

$$\lim_{\phi \to 0} \|w(\mathbf{x}, \theta)(\mathbf{x} - \theta)\|$$
$$= \lim_{\phi \to 0} \phi e^{-\beta\frac{(1-\phi^2)\|x_1\|^2}{\phi^2}}\sqrt{\|x_1 - \theta\|^2 + \|x_1\|^2(1-\phi^2)/(\phi^2)}$$
$$= \lim_{\phi \to 0} e^{-\beta\frac{\|x_1\|^2}{\phi^2}}\sqrt{\phi^2\|x_1 - \theta\|^2 + \|x_1\|^2(1-\phi^2)} = 0.$$

Further, $K(\theta)$ is finite since both $\int w(\mathbf{y}, \theta)\mathrm{d}F(\mathbf{y})$ and $\|[I - H(\theta)]^{-1}\|_{op}$ are finite. Thus, from (16),

$$\lim_{\phi \to 0}\|IF(\mathbf{x}; \mathbf{T}, F)\| \leq \lim_{\phi \to 0} K(\theta)\|w(\mathbf{x}, \theta)(\mathbf{x} - \theta)\| = 0.$$

Since the influence function goes to zero when the outliers have zero correlation coefficients with the correct projection and the influence function is continuous, it is also bounded for outliers whose correlation coefficients are small.

IV. CLASS MEAN ESTIMATION FOR MULTIPLE CTFs

We now turn to the problem of estimating a CTF-corrected class mean from aligned images having different CTFs.

A. Image Model

We assume that each of the N images $\mathbf{x}_i$, $i = 1, \cdots, N$ has one of L CTFs, c_j, $j = 1, \cdots, L$. Let N_j denote the number of images in the jth CTF group. It will be convenient to use a double index for images, where $\mathbf{x}_{j,k}$, $k = 1, \cdots, N_j$ refers to the kth image in the jth CTF group.

Assuming Gaussian white noise, the image $\mathbf{x}_{j,k} \in \mathbb{R}^p$ is

$$\mathbf{x}_{j,k} = c_j * \boldsymbol{\mu}_{j,k} + \mathbf{n}, \quad (17)$$

where $\boldsymbol{\mu}_{j,k}$ is the CTF-free non-noisy signal and $*$ denotes the convolution operation.

Inlier: When the image $\mathbf{x}_{j,k}$ is an inlier,

$$\boldsymbol{\mu}_{j,k} = s\boldsymbol{\mu} \quad (18)$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ is the CTF-free non-noisy projection. As before, s models contrast variation and has a uniform density $\pi(s|a, b)$ where a and b are symmetric about 1. Marginalizing s results in the pdf of inliers $\mathbf{x}_{j,k}$:

$$f(\mathbf{x}_{j,k}|\boldsymbol{\mu}) = \int_a^b g(\mathbf{x}_{j,k}|c_j * s\boldsymbol{\mu})\pi(s|a, b)\mathrm{d}s \quad (19)$$

Outlier: Similar to the single-CTF case, for an outlier image:

1) $\boldsymbol{\mu}_{j,k}$ has a low correlation coefficient with $\boldsymbol{\mu}$, and
2) $\boldsymbol{\mu}_{j,k}$ has a finite component along $\boldsymbol{\mu}$.

The goal is to robustly estimate $\boldsymbol{\mu}$ from the images.

B. The Robust CTF-Corrected Class Mean Estimator

We formulate this estimator in the discrete Fourier transform space where the convolution $c_j * \boldsymbol{\mu}$ becomes the point-wise multiplication $C_j(\mathbf{s})M(\mathbf{s})$ where the capital letters denote the 2D discrete Fourier transforms of c_j and $\boldsymbol{\mu}$, and $\mathbf{s}$ denotes a point in the 2D discrete Fourier space. With a slight abuse of notation, $C_j(\mathbf{s})M(\mathbf{s})$ can be written as a matrix operation $C_j M$, by 1) constructing a diagonal matrix called C_j whose diagonal elements are the components of $C_j(\mathbf{s})$, and 2) scanning M into a column vector.

Our robust CTF-corrected class mean estimate is a weighted version of the estimate in [26] defined as

$$\mathbf{T} = \Big[\sum_{j=1}^{L} C_j^T C_j N_j\Big]^{-1}\Big[\sum_{j=1}^{L} C_j^T \sum_{k=1}^{N_j} \mathbf{X}_{j,k}\Big], \qquad (20)$$

where N_j is the number of images in the jth CTF group and $\mathbf{X}_{j,k}$ is the discrete Fourier transform of image $x_{j,k}$. Note that N_j could take value 1, which simply states that each image has its own CTF. We extend (20) to a robust estimator of the CTF-corrected class mean:

$$\mathbf{T} = \Big[\sum_{j=1}^{L} C_j^T C_j \sum_{k=1}^{N_j} w(\mathbf{X}_{j,k}, C_j\mathbf{T})\Big]^{-1}$$
$$\cdot \Big[\sum_{j=1}^{L} C_j^T \sum_{k=1}^{N_j} w(\mathbf{X}_{j,k}, C_j\mathbf{T})\mathbf{X}_{j,k}\Big] \qquad (21)$$

by incorporating the weight function $w(\mathbf{X}, \mathbf{T})$ defined in (6).

Similarly, the estimate can be determined by an iterative algorithm.

C. Fisher-Consistency of the Estimator

The Fisher-consistency of $\mathbf{T}$ in (21) can be easily shown by applying consistency of the single CTF case to each CTF group. We write $\mathbf{T}$ as a statistical functional of F:

$$\mathbf{T}(F) = \Big[\sum_{j=1}^{L} C_j^T C_j \int w(\mathbf{X}_j, C_j\mathbf{T}(F))\mathrm{d}F_j(\mathbf{X}_j)\Big]^{-1}$$
$$\cdot \Big[\sum_{j=1}^{L} C_j^T \int w(\mathbf{X}_j, C_j\mathbf{T}(F))\mathbf{X}_j\mathrm{d}F_j(\mathbf{X}_j)\Big], \quad (22)$$

where $\mathbf{X}_j$ denotes the image in the jth CTF group and F_j is the distribution of $\mathbf{X}_j$. $\mathbf{T}(F)$ is Fisher-consistent if $\mathbf{T}(F) = M$ where M is the Fourier transform of $\boldsymbol{\mu}$.

From the consistency of the single CTF case,

$$C_j M = \frac{\int w(\mathbf{X}_j, C_j M)\mathbf{X}_j\mathrm{d}F_j(\mathbf{X}_j)}{\int w(\mathbf{X}_j, C_j M)\mathrm{d}F_j(\mathbf{X}_j)},$$

which gives

$$C_j M \int w(\mathbf{X}_j, C_j M)\mathrm{d}F_j = \int w(\mathbf{X}_j, C_j M)\mathbf{X}_j\mathrm{d}F_j.$$

Multiplying both sides by C_j^T and summing over all js,

$$\sum_{j=1}^{L} \Big[C_j^T C_j \int w(\mathbf{X}_j, C_j M)\mathrm{d}F_j\Big]$$
$$\cdot M = \sum_{j=1}^{L} C_j^T \int w(\mathbf{X}_j, C_j M)\mathbf{X}_j\mathrm{d}F_j.$$

Rearranging the terms,

$$M = \Big[\sum_{j=1}^{L} C_j^T C_j \int w(\mathbf{X}_j, C_j M)\mathrm{d}F_j\Big]^{-1}$$
$$\cdot \Big[\sum_{j=1}^{L} C_j^T \int w(\mathbf{X}_j, C_j M)\mathbf{X}_j\mathrm{d}F_j\Big]. \qquad (23)$$

Comparing (23) and (22) shows that $\mathbf{T}(F) = M$.

D. The Influence Function

In this section we derive the influence function of the estimator and show that it is bounded for outliers. We evaluate the influence function for an outlier $\mathbf{X}$ as its correlation coefficient with the CTF-affected projection $\boldsymbol{\Theta}$ goes to zero while the component of $\mathbf{X}$ along $\boldsymbol{\Theta}$ remains finite ($\boldsymbol{\Theta} = CM$ where C is the CTF of the outlier image $\mathbf{X}$).

The influence function of $\mathbf{T}$ in (22) at a distribution F is calculated by $IF(\mathbf{X}; \mathbf{T}, F) = \frac{\partial}{\partial\epsilon}[\mathbf{T}(F_\epsilon)]_{\epsilon=0}$, where F_ϵ is a contaminated distribution by putting a ϵ mass at one of the distributions $F_j, j = 1, \cdots, L$ of CTF group. Let the kth CTF group be this contaminated CTF group and denote the contaminated distribution $F_{k_\epsilon} = (1 - \epsilon)F_k + \epsilon \Delta_{\mathbf{X}}$. The influence function can then be derived by replacing F by F_ϵ in (22), taking derivative of (22) with respect to ϵ and evaluating at 0 and applying the consistency $\mathbf{T}(F) = M$, which gives

$$[I - Z(M)]\frac{\partial}{\partial\epsilon}[\mathbf{T}(F_\epsilon)]_{\epsilon=0} = \Big[\sum_{j} C_j^T C_j \int w(\mathbf{Y}_j, C_j M)\mathrm{d}F_j\Big]^{-1}$$
$$\cdot C_k^T w(\mathbf{X}, C_k M)(\mathbf{X} - C_k M),$$

where

$$Z(M) \triangleq \Big[\sum_{j} C_j^T C_j \int w(\mathbf{Y}_j, C_j M)\mathrm{d}F_j\Big]^{-1}$$
$$\cdot \Big[\sum_{j} C_j^T \int (\mathbf{Y}_j - C_j M)\frac{\partial}{\partial\mathbf{t}}[w(\mathbf{Y}_j, \mathbf{t})]_{C_j M}\mathrm{d}F_j C_j\Big].$$

If $I - Z(M)$ is invertible, then the influence function is

$$IF(\mathbf{x}; \mathbf{T}, F) = \Big[(\sum_{j} C_j^T C_j \int w(\mathbf{Y}_i, C_j M)\mathrm{d}F_j)(I - Z(M))\Big]^{-1}$$
$$\cdot C_k^T w(\mathbf{X}, C_k M)(\mathbf{X} - C_k M).$$

$I - Z(M)$ has to be invertible for the existence of the influence function. Using the results of the single CTF case, we prove in the appendix the following sufficient condition: $B(\bar{\beta}, \|\overline{\boldsymbol{\Theta}}\|, p) < 1, \forall 1 \le j \le L$, where $B(\bar{\beta}, \|\overline{\boldsymbol{\Theta}}\|, p)$ is derived in the single CTF case and $\bar{\beta} = \beta\sigma^2$, $\overline{\boldsymbol{\Theta}} = \boldsymbol{\Theta}/\sigma$. Such a condition can be satisfied by choosing proper $\bar{\beta}$.

We have shown in Section III-D that for relevant $\|\overline{\mathbf{\Theta}}\|$ and p, $\bar{\beta} = 10^{-5}$ satisfies the condition.

We next evaluate the influence function for outliers $\mathbf{X}$. Let $\mathbf{X} = [X_1, X_2]^T$ where X_1 and X_2 are along and orthogonal to $\mathbf{\Theta}_k = C_k M$. Similarly, we consider the influence function as $\phi \triangleq \|X_1\|/\|\mathbf{X}\|$ goes to zero while $\|X_1\|$ is finite. First,

$$\|IF(\mathbf{X}; \mathbf{T}, F)\| \leq K(M)\|C_k^T w(\mathbf{X}, C_k M)(\mathbf{X} - C_k M)\|,$$

$$K(M) \triangleq \left\| \left[\left(\sum_j C_j^T C_j \int w(\mathbf{Y}_i, C_j M) dF_j \right)(I - Z(M)) \right]^{-1} \right\|_{op}.$$

Since $K(M)$ is also finite, $\quad \lim_{\phi \to 0} \|IF(\mathbf{X}; \mathbf{T}, F)\|$

$$\leq \lim_{\phi \to 0} K(M)\|w(\mathbf{X}, C_k M)(\mathbf{X} - C_k M)\| = 0.$$

Thus the estimator for the CTF-corrected class-mean also has a zero influence function for outliers.

V. EXPERIMENTAL RESULTS

We present the results of our estimator applied to simulated and real cryo-EM images.

A. Simulated Data

Simulated cryo-EM images were generated from the atomic structure of the 50S ribosomal subunit from the Protein Data Bank (PDB ID: 1JJ2) using a hydration model [27], and was sampled at a pixel size of 3Å. The 2D particle images (inliers) were obtained from the 3D structure according to the image model in (17) and (18) by: 1) projecting the structure and applying a CTF; 2) applying the amplitude factor from a uniform distribution between 0.5 and 1.5; and 3) adding Gaussian white noise.

To model the outliers in cryo-EM, we followed a strategy reported in [28] and generated a uniform mixture of images from five classes: misclassified images, projection of a sphere, a plane and a cylinder, and pure noise image. Misclassified images are projections from other directions than the inliers with the difference of the projection directions larger than 65 degrees. Fig.4 shows typical images from these five classes of outliers. CTF, amplitude variation and Gaussian white noise were also applied to these projections in the same way as for the inliers.

For the single CTF case, a CTF with defocus $1.3\mu m$ was used. For multiple CTF case, five CTFs with defocus values $1.0\mu m$, $1.8\mu m$, $2.0\mu m$, $2.8\mu m$ and $3.5\mu m$ were used.

B. Cryo-EM Data

Real cryo-EM images were the 50S ribosomal subunit images from the National Resource for Automated Molecular Microscopy. Images were classified into five CTF groups (defocus of $1.2\mu m$, $1.6\mu m$, $1.9\mu m$, $2.2\mu m$ and $2.5\mu m$) and aligned by the software package SPIDER [8].

C. Performance Measures

We measured the ability of the proposed w-estimator (referred to as wme) to reject outliers and compared it with the classical cross-correlation with thresholds (referred to as cc).
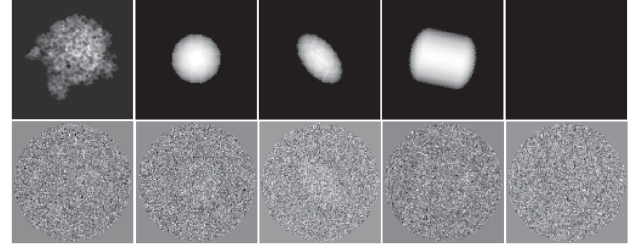

Fig. 4. Simulated outliers. Examples of non-noisy outliers (top row) and their CTF-affected and noisy version (bottom row), including (from left to right) a misaligned particle, projection of a sphere, a plane, a cylinder and image of pure noise.

We use two figures of merit: the precision rate (P) and the recall rate (R). For an outlier detection method that makes hard decision about whether an observation is an outlier or not, P and R are calculated by:

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN}, \tag{24}$$

where TP, FP, TN and FN are the number of true positives (correctly classified inliers), false positives (outliers misclassified as inliers), true negatives (correctly classified outliers) and false negatives (inliers misclassified as outliers) [29]. The precision rate measures the fraction of images which participate in the averaging step that are inliers. The recall rate measures the fraction of inliers included in averaging with respect to all available inliers. Since the w-estimator does not make hard decision about outliers but rather assigns weights to the outliers, we calculate two equivalent metrics to P and R:

$$\widehat{P} = \frac{\sum_{j \in S} w_j}{\sum_j w_j} \quad \widehat{R} = \frac{\sum_{j \in S} \min(1, \frac{w_j}{\sum_i w_i} N_{in})}{N_{in}} \tag{25}$$

where w_j are the image weights, S contains indices of inliers and N_{in} is the cardinality of S. $\widehat{P}$ measures the contribution of inliers to the averaging and is thus equivalent to P in (24). The numerator of $\widehat{R}$ calculates the equivalent number of TP by examining the contribution of each inlier that participates in the averaging. Since $TP + FN = N_{in}$, $\widehat{R}$ is equivalent to R in (24).

A second criterion that we used for performance comparison is the standard deviations of the precision and recall rates, which characterize the consistency of the performance.

Finally, to assess the accuracy of the resulting class means, we calculated the mean square error between the class means and the non-noisy projection in different frequency rings.

D. Results for Single CTF Class Means

1) Simulations: Before presenting the results for the precision and recall rates, we first demonstrate that the w-estimator works as designed, i.e. at convergence it down-weighs the contribution of the outliers to the class mean. We took a total of 30 images, 12 of which are outliers (40% outliers). In cryo-EM, the number of images per direction goes up to the order of 100. We chose to report results with vastly fewer images because this is the more challenging case (a significant
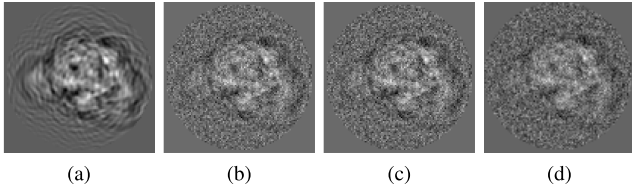
Fig. 5. Results for simulated data in the single CTF case. (a) shows the non-noisy projection θ. (b), (c) and (d) are the w-estimate **T**, the mean of inliers and the mean of all images.
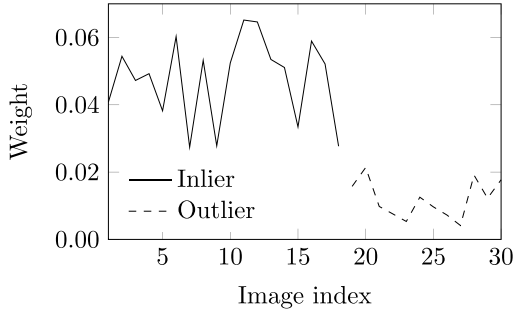

Fig. 6. Weights of the images at convergence for simulated data in the single-CTF case. 30 images with 40% of outliers with SNR of −16dB.

amount of noise remains in the estimated class mean). Noise was added to create an SNR of −16dB.

Fig.5 shows the non-noisy projection, the w-estimate, the mean of inliers and the mean of all images. The w-estimate is close to the mean of inliers. Fig.6 shows the converged weights of images. Note that the weights of the outliers are on average lower than 20% of the weights of the inliers, showing that the effect of outliers on the estimate is greatly diminished.

To continue, we compared the performance of wme with that of cc using precision and recall rates. We evaluated the performances for different SNRs and different outlier percentages.

First, we evaluated wme and cc on images that represent the typical highest and lowest SNRs observed in cryo-EM: −8dB and −16dB. A total of 30 images of each SNR were generated, both with 45% of outliers. This was repeated 100 times. Fig.7 a-b shows the mean values of the precision (P) and recall (R) rates for wme and cc for SNR −8dB and −16dB respectively. P and R for cc depend on threshold, which is indicated on the x-axis with varying thresholds. The wme does not require a threshold and its rates are shown as horizontal lines. Two observations can be made from Fig.7: 1) for both noise levels, there is a very small range of thresholds (denoted as shaded area) in which the cc performance is comparable to the wme performance. The performance of cc degrades rapidly when the threshold falls out of this range, with either P or R falling off from its high value. 2) The aforementioned ranges for the two SNRs are very different from each other. This clearly shows that the threshold has to be adjusted for images of different SNRs. In contrast, wme naturally adapts to different SNRs without requiring any manual adjustment. The standard deviations of P and R for wme and cc are shown in Table I where cc was evaluated with the optimal
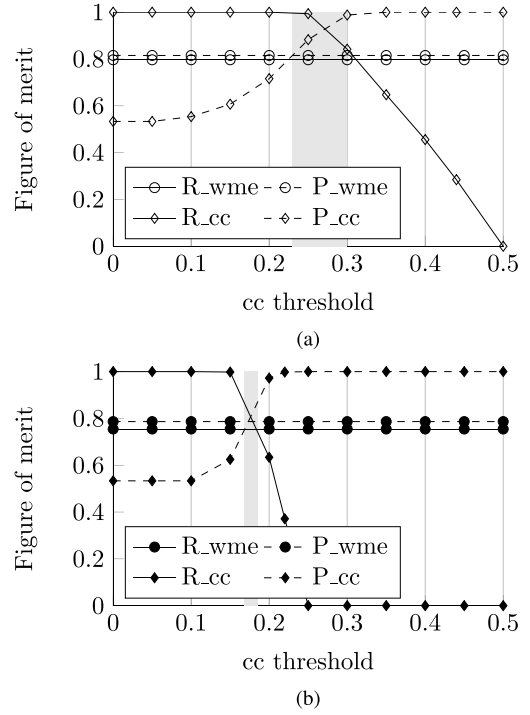

Fig. 7. Performance of cross-correlation with thresholds (cc) and the proposed w-estimator (wme) at two signal-to-noise ratios (SNR): (a) −8dB and (b) −16dB. 45% outliers are present in both cases. P and R are the precision and recall rates.

TABLE I

STANDARD DEVIATION OF PRECISION AND RECALL RATES OF
CROSS-CORRELATION WITH THRESHOLDS (CC) AND THE
W-ESTIMATOR (wme) IN THE SINGLE CTF CASE

	−8dB		−16dB	
	cc[1]	wme	cc[2]	wme
Precision	0.07	0.03	0.09	0.04
Recall	0.04	0.02	0.08	0.03

[1] at threshold of 0.27
[2] at threshold of 0.18

threshold (e.g., 0.18 is used for images of −16dB since Fig.7b shows that it generates the best performance). The wme demonstrates consistently smaller standard deviations than cc, further supporting the claim that it reliably reduces the influence of outliers even when the SNR changes.

Second, we compared the performances of wme and cc for data with different outlier percentages. The precision and recall rates are calculated for data containing 10% to 45% outliers with SNR −16dB. The thresholds used in evaluating the performance of cc are the values that generate the best performance. Fig.8 shows the mean and standard deviation of the rates for wme and cc, calculated from 100 experiments for each outlier percentage. The wme is shown to have comparable rates to those of cc while the standard deviations of the rates for wme are consistently lower for all outlier percentages, demonstrating the reliability of wme performance.

To further compare the quality of the class means and to understand the effects of outliers, we calculated the mean squared error (MSE) between the class means and the non-noisy projection in frequency rings. Similar to calculating
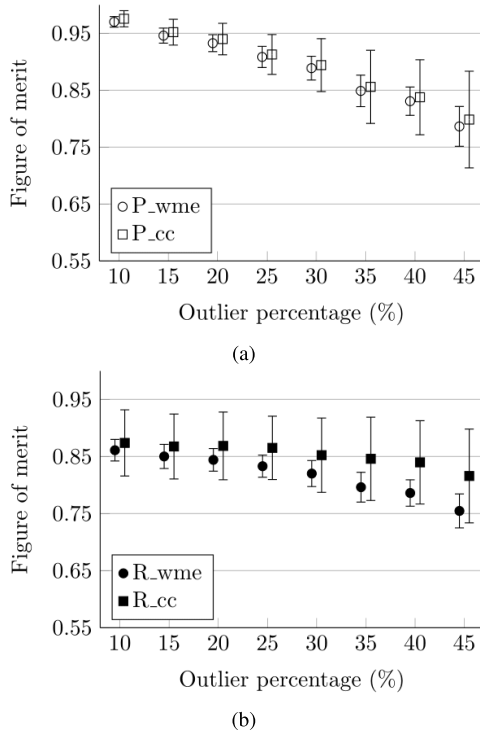
Fig. 8. Performance of cross-correlation with thresholds (cc) and the proposed w-estimator (wme) to varying outlier percentages. Optimal thresholds are used for cc. (a) and (b) show the result of precision rate (P) and recall rate (R) respectively. 10% to 45% outliers and SNR of −16dB.
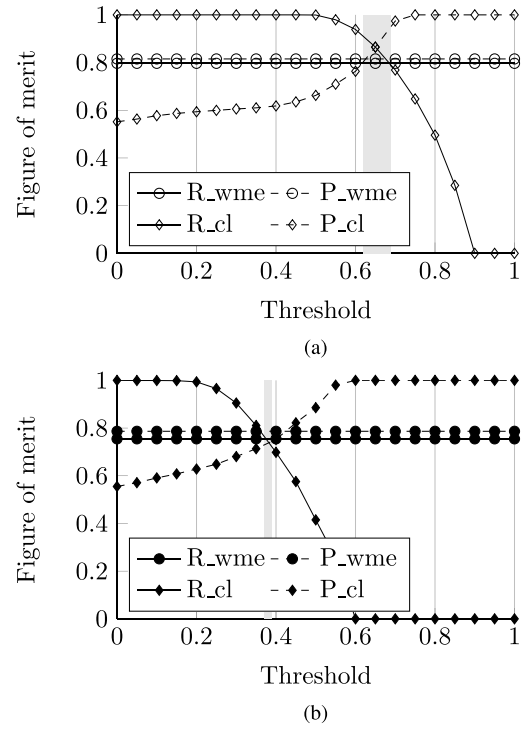

Fig. 10. Performance of common-line based approach (cl) and the proposed w-estimator (wme) at two signal-to-noise ratios (SNR): (a) −8dB and (b) −16dB. A total of 900 common-lines and 45% outliers. P and R are the precision and recall rates.
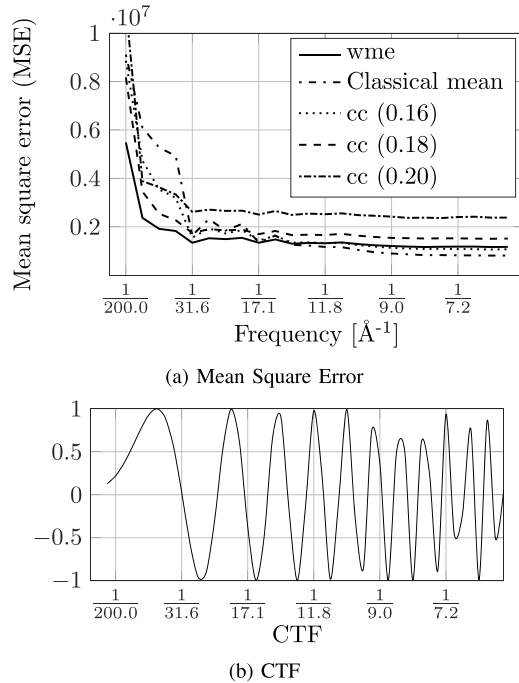

Fig. 9. Mean square error (MSE) of the w-estimator (wme), classical mean and cross-correlation (cc) with different thresholds for outlier rejection. 0.18 is the optimal threshold for cc. 45% outliers and SNR of −16dB. (a) shows the MSEs at different frequencies. (b) is the CTF of the images.

Fourier ring correlation (FRC) [30], the MSE between the Fourier-transformed class means and non-noisy projection was computed for concentric rings of increasing radius centered at $(0, 0)$ frequency. Fig.9a shows the MSE of wme, the

classical mean (no outlier removal) and cc with different thresholds (0.18 is the optimal threshold). We used a total of 30 images with SNR −16dB and 45% outliers and results were calculated from 100 repeated experiments. Several observations can be made: 1) The MSE at around $1/200Å^{-1}$ are comparably high for all methods since CTF causes loss of low frequency information. 2) The classical mean and cc with thresholds 0.16 and 0.18 all have higher MSE between $1/200Å^{-1}$ and $1/31.6Å^{-1}$ which is the range of frequencies within the first peak of the corresponding CTF (Fig.9b). This range of frequencies is where the outliers have the most influence. 3) For cc, small deviation from the optimal threshold causes the MSE to increase, showing that the performance of cc is very sensitive to thresholds. 4) The low MSE of the wme estimate in lower frequency regions shows that it outperforms cc and the classical mean. (The MSE of the classical mean and cc with threshold 0.16 are slightly lower than wme in higher frequency regions, because they average more images and only noise is present in these regions.)

We also compared the performance of wme with the common-lines (cl) based approach in [16]. Two sets of 900 class means from uniform projections of the north hemisphere of the simulated 50S ribosome subunit were generated, with SNRs equivalent to averaging 30 images with SNRs of −8dB and −16dB. Along one projection direction, 30 images (SNRs of −8dB or −16dB) with 45% outliers were created. Their common-line projection correlation coefficient with all of the class means from the remaining 899 directions were calculated and used for outlier rejection. Fig.10 shows the
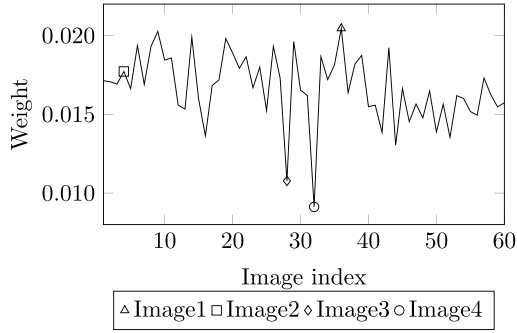
Fig. 11. Weights of images at convergence for cryo-EM images. Four images are labeled: image1 ($\triangle$) with high weight, image2 ($\square$) with median weight, image3 ($\diamond$) and image4 ($\circ$) with low weights.
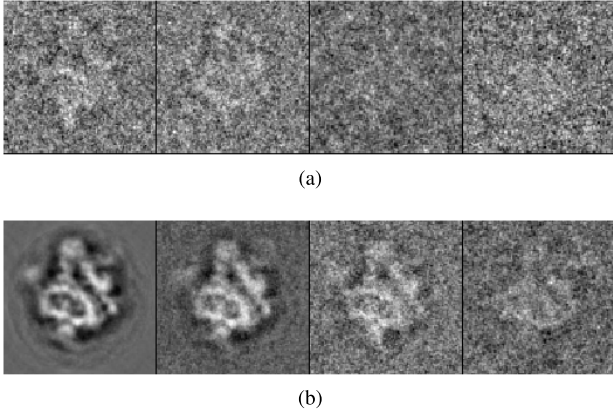

(a)

(b)

Fig. 12. Results for cryo-EM images in the single CTF case. (a) From left to right: image1 ($\triangle$) with high weight, image2 ($\square$) with median weight, image3 ($\diamond$) and image4 ($\circ$) with low weights. (b) From left to right: projection of reconstruction, estimate **T**, mean of 5 images with highest weights and mean of 5 images with lowest weights.

mean precision and recall rates from 100 repeated experiments. The results show that the performance of correlating the common-lines is also sensitive to the choice of thresholds, and wme has performance comparable to the best common-lines performance where the "optimal" threshold (gray area) is used.

2) Real Cryo-EM Images: We next show the results on real cryo-EM data. Sixty images of 50S Ribosome subunit aligned to a projection direction from CTF5 group (highest defocus) were used. Fig.11 shows the converged weights of wme. Notice that two images have significantly lower weights. To further assess the quality of these images, we chose to examine four images (denotes as Image1-4 in Fig.11): Image1 with a high weight, Image2 with median weight, and Image3 and Image4 with lower weights. Fig.12a displays these four images. The projection of the reconstruction obtained by the algorithm in [31] serves as a reference for the non-noisy projection of the particle and is shown in Fig.12b. Comparing the images in Fig.12a with the reference in Fig.12b strongly suggests that the two images with lower weights do not appear to contain the signal of the projected structure. To further confirm this, we took 5 images with the highest and lowest converged weights and calculate the mean images from them. The resulting mean images, displayed in Fig.12b, show that the images with lowest weights contain signal different from
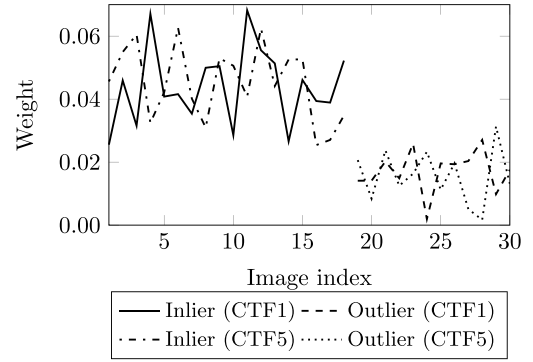
Fig. 13. Weights of images at convergence for simulated images in the multiple CTF case. Results of CTF1 (lowest defocus) and CTF5 (highest defocus) are shown. Each CTF group has 40% outliers and average SNR of -15dB. Weights are normalized within each CTF group.
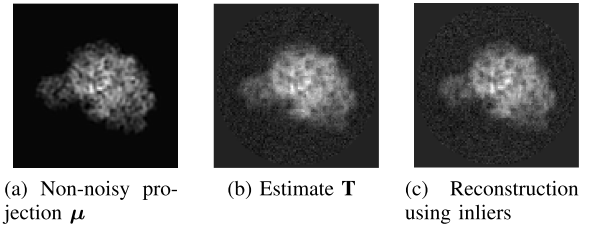

(a) Non-noisy projection μ (b) Estimate **T**   (c) Reconstruction using inliers

Fig. 14. Results for simulated data in the multiple CTF case. (a) shows the non-noisy CTF-free projection μ. (b) is the w-estimate **T** calculated from the images and (c) is the ideal estimate calculated from the inliers.

(a)

(b)
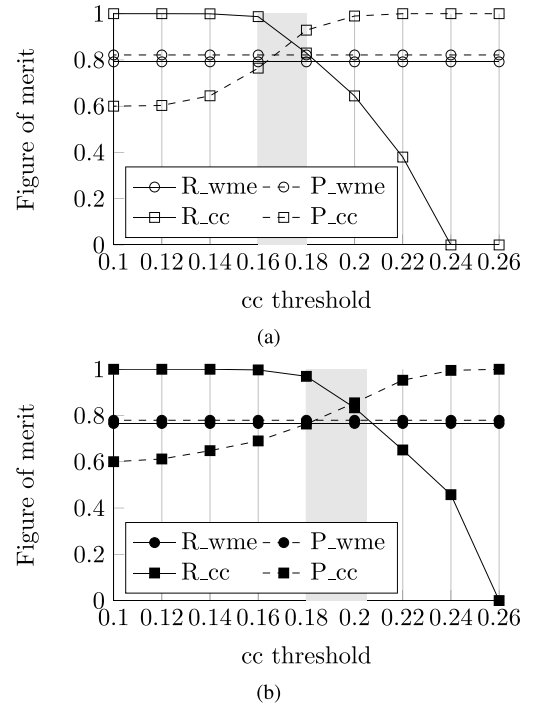
Fig. 15. Performance of cross-correlation with thresholds (cc) and the w-estimator (wme) for multiple-CTF case. cc is performed independently for each CTF groups. Precision (P) and Recall (R) rates of two CTFs with defocus values of: (a) $1.0\mu m$ and (b) $3.5\mu m$ are shown. 40% outliers.

the projected structure and the rest of the images. Fig.12b also shows the estimate of the class-mean from our estimator.

To summarize, experiments with the single CTF case have shown that: 1) the performance of cross-correlation is sensitive

TABLE II

PRECISION RATE AND RECALL RATE AND THEIR STANDARD DEVIATIONS OF THE w-ESTIMATOR FOR SINGLE (wme(s)) AND MULTI CTF (wme(M)) CASES AND CROSS-CORRELATION WITH THRESHOLDS (cc) IN THE MULTIPLE CTF CASE

10% outlier	−7dB			−15dB		
	cc	wme(s)	wme(m)	cc	wme(s)	wme(m)
Precision	0.98(0.02)	0.97(0.01)	0.97(0.01)	0.96(0.03)	0.96(0.01)	0.96(0.01)
Recall	0.91(0.05)	0.89(0.01)	0.89(0.01)	0.88(0.06)	0.86(0.02)	0.87(0.02)
40% outlier	−7dB			−15dB		
	cc	wme(s)	wme(m)	cc	wme(s)	wme(m)
Precision	0.86(0.07)	0.84(0.03)	0.84(0.02)	0.78(0.07)	0.80(0.03)	0.81(0.03)
Recall	0.82(0.06)	0.83(0.02)	0.83(0.02)	0.80(0.08)	0.78(0.03)	0.78(0.03)

to the threshold and has to be adjusted for images of different SNRs. 2) the w-estimator performs as well as the cross-correlation (with optimal threshold) without the need of adjustment for varying SNRs and outlier percentages. 3) w-estimator is able to down-weight outliers in real cryo-EM case.

E. Results for Multiple CTF Class Means

We first show that the proposed w-estimator is capable of handling outliers in more than one CTF groups. 30 images with 40% outliers were generated for each of the five CTF groups. The standard deviation of the image noise is $\sigma = 85$ (Because different CTFs lead to different signal energy, the images do not have the same SNRs. This σ corresponds to an average SNR of −15dB.). Fig.13 shows the converged weights of images for the CTF groups with the smallest (CTF1) and largest (CTF5) defocus values. The weights of other CTF groups have similar behaviors and were omitted in the plot for ease of visualization. For the same reason, the weights shown in Fig.13 are normalized such that the weights of all images within each CTF group sum up to 1. For all CTF groups, the weights of the outliers are lower than 20% of the weights of the inliers. Fig.14 shows the CTF-free non-noisy projection, the estimate from the w-estimator and the estimate from a hypothetical "ideal" method that uses equal weights for inliers and zero weights for outliers. The w-estimate is similar to the class mean produced by the ideal method.

Next, we compare the performance of cross-correlation with thresholds (cc) and the w-estimator (wme) for the multiple CTF case. We have already shown from the single CTF case that the performance of cc is sensitive to threshold for varying SNRs. For the multiple CTF case, cc is performed independently for each CTF group. Fig.15 shows the precision and recall rates of wme and cc. Results of two CTF groups with defocus values of $1.0 \mu m$ and $3.5 \mu m$ are shown. The shaded area in the two plots indicate the range of thresholds within which the performance of cc is comparable to wme. These ranges are narrow for both CTF groups and the two ranges have very limited overlapping region. This is evidence that the same threshold cannot be used for all CTF groups. In contrast, wme achieves satisfactory performance for different CTF groups without adjustments.

We next show that wme for the multiple CTF case also adapts to varying SNRs and outlier percentages. We calculated the precision and recall rates of wme and cc for data with combinations of two SNRs (−7dB and −15dB) and two outlier
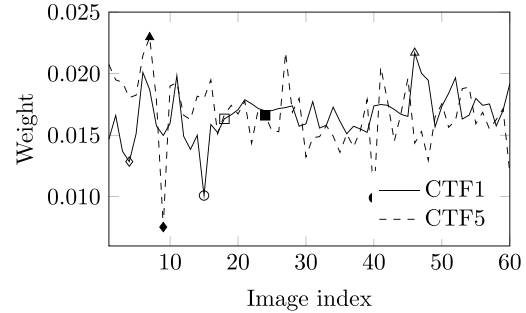

Fig. 16. Weights of images at convergence for cryo-EM images. Results of CTF1 (lowest defocus) and CTF5 (highest defocus) are shown. Weights are normalized within each CTF group. Four images are labeled for each CTF group: image1 (△/▲) with high weight, image2 (□/■) with median weight, image3 (◇/◆) and image4 (○/●) with low weights.

(a) Images from CTF1 (top) and CTF5 (bottom). (b) Projection of reconstruction
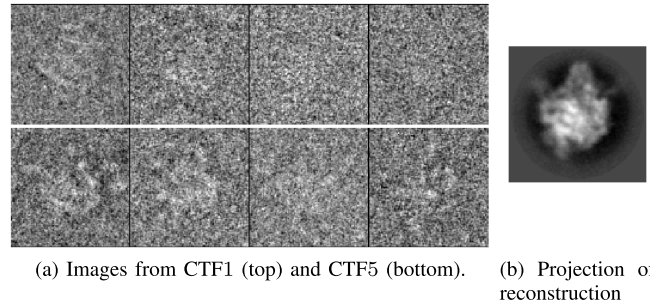
Fig. 17. Results for cryo-EM images in the multiple CTF case. (a) shows the images from two CTF groups that are labeled in Fig. 16. From left to right: image1 (△/▲) with high weight, image2 (□/■) with median weight, image3 (◇/◆) and image4 (○/●) with low weights. (b) is the projection of the reconstruction.

percentages (10% and 40%). The rates and their standard deviations are reported in Table II. We included the results of wme for both single CTF (wme(s), i.e. applying wme independently for each CTF group) and multiple CTF (wme(m)). For both wme(s) and wme(m), their rates are comparable to cc but with much lower standard deviations. The performance of wme(m) is on average better than wme(s), wme(m) uses images from more than one CTF.

Finally, we show results of real cryo-EM images with five CTF groups, each having 60 images. Fig.16 shows the converged weights for CTF1 (smallest defocus) and CTF5 (largest defocus) groups. For both groups, four images were chosen: an image with a high weight, an image with a median weight and two images with lower weights. Top and bottom rows of Fig.17a shows images in CTF1 and CTF5. The projection of the reconstruction is given in Fig.17b. The

images with low weights of CTF5 do not appear to contain the same particle signal as the images with higher weights; the images with low weights of CTF1 appear to contain minimal signals.

VI. CONCLUSION AND DISCUSSION

We presented a new approach to estimate class means in the presence of outliers in cryo-EM. The new estimator is applied to the images after they are aligned according to their projection directions and CTFs. Instead of attempting to reject outlier images with a threshold, this approach aims to calculate the class means by a weighted average of images where the weight function limits the influence of outliers. The estimator is robust against outliers; its influence function is bounded and goes to zero asymptotically.

Classical methods for outlier detection require a manually adjusted threshold, such as a threshold for the correlation coefficient. Simulations show that performance of such methods is very sensitive to the choice of threshold. Optimal thresholds are also difficult to find in practice. The main advantage of our approach is that it eliminates thresholds and automatically adapts to the SNR and the outlier percentage of the data.

We also extended the proposed estimator to multiple CTFs. This estimator is capable of estimating the CTF-corrected class mean while limiting the effects of outliers. Experiments with simulated data demonstrate its ability to deal with outliers in more than one CTF group.

We applied the estimator to experimental cryo-EM data of 50S ribosomal subunit. For both single and multiple CTF cases, the estimator assigns lower weights to possible outlier images and limits their influence of on the class means.

The proposed estimator calculates robust 2D class means when outliers are present, by using the weighting scheme of the w-estimators. Such a strategy can be adopted for robust 3D reconstruction by incorporating weights in the reconstruction calculation. Although weighted reconstruction were proposed in [14] and [17], they either do not address the outliers or require heuristics in choosing the parameter values. By posing reconstruction as a robust estimation problem and designing an estimator with desired properties, images from all projection directions can be used for a robust 3D reconstruction.

The robust estimation framework proposed in this work is also useful for other image processing problems. One possible application is robustifying the non-local means (NLM) algorithm [32] for image denoising. The performance of NLM can be improved by using weight functions similar to one proposed here.

APPENDIX A
PROOF OF CLAIM 1

First, using Cauchy-Schwarz inequality, $E_{\mathbf{x}}[\|\mathbf{x}h(\mathbf{x})\|] = \int \|\mathbf{x}\| h(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \leq \sqrt{\int \|\mathbf{x}\|^2 f(\mathbf{x}) d\mathbf{x}} \sqrt{\int h(\mathbf{x})^2 f(\mathbf{x}) d\mathbf{x}}$. Since $\int h(\mathbf{x})^2 f(\mathbf{x}) d \leq M^2$ ($h(\mathbf{x}) \leq M$ and $\int f(\mathbf{x}) d\mathbf{x} = 1$) and $\int \|\mathbf{x}\|^2 f(\mathbf{x}) d\mathbf{x} = Q^2 < \infty$, $E_{\mathbf{x}}[\|\mathbf{x}h(\mathbf{x})\|] \leq QM < \infty$. Thus $\|E_{\mathbf{x}}[\mathbf{x}h(\mathbf{x})]\| \leq E_{\mathbf{x}}[\|\mathbf{x}h(\mathbf{x})\|] < \infty$. Consider two cases:

1) $p = 1$. $h(x)$ is an even function, i.e., $h(-x) = h(x)$. Then $E_x[xh(x)] = \int_{-\infty}^{\infty} xh(x)f(x)dx = \int_0^{\infty} xh(x)[f(x) - f(-x)]dx$. Also, if $x > 0$, $f(x) > f(-x)$ hence $E_x[xh(x)] > 0$ when $\theta > 0$ and $f(x) < f(-x)$ hence $E_x[xh(x)] < 0$ when $\theta < 0$. Thus, $E_x[xh(x)] = \alpha\theta$ where $\alpha > 0$.

2) $p > 1$. Let $\mathbf{x} = [x_1, x_{\bar{1}}]$, where $x_{\bar{1}}$ denotes the vector containing the second to the pth components. Let $\boldsymbol{\theta} = [\theta, 0]$. Then $x_1 \sim N(\theta, \sigma^2)$ and $x_{\bar{1}} \sim N(\mathbf{0}, \sigma^2 I)$. Thus $E_{x_{\bar{1}}|x_1}[x_{\bar{1}}h(\mathbf{x})] = 0$. Then $E_{\mathbf{x}}[x_{\bar{1}}h(\mathbf{x})] = E_{x_1}[E_{x_{\bar{1}}|x_1}[x_{\bar{1}}h(\mathbf{x})]] = 0$ and $E_{\mathbf{x}}[\mathbf{x}h(\mathbf{x})] = E_{\mathbf{x}}[x_1h(\mathbf{x})] = E_{x_1}[E_{x_{\bar{1}}|x_1}[x_1h(\mathbf{x})]] = E_{x_1}[x_1 E_{x_{\bar{1}}|x_1}[h(\mathbf{x})]] = E_{x_1}[x_1\beta(x_1)]$ where $\beta(x_1) \triangleq E_{x_{\bar{1}}|x_1}[h(\mathbf{x})]$. $\beta(x_1) > 0$ except $x_1 = 0$, and $\beta(x_1)$ is an even function of x_1. From case 1), $E_{x_1}[x_1\beta(x_1)] = \alpha\theta, \alpha > 0$. Therefore, $E_{\mathbf{x}}[\mathbf{x}h(\mathbf{x})] = \alpha\boldsymbol{\theta}$ where $\alpha > 0$.

APPENDIX B
UPPER BOUND OF $\|H(\boldsymbol{\theta})\|_{op}$

Define a coordinate system where $\boldsymbol{\theta} = [\theta, 0, \cdots, 0]^T$ and $\mathbf{y} = [y_1, \cdots, y_p]^T$. We have $f(\mathbf{y}) = f(y_1|\theta) f(y_2) \cdots f(y_p)$ where $f(y_1|\theta) = \int_a^b 1/(b-a)g(y_1|s\theta)ds$ and $f(y_i) = g(y_i|0)$, $i = 2, \cdots, p$. $g(y|\gamma)$ is the pdf of $N(\gamma, \sigma^2)$.

Let $U(\boldsymbol{\theta}) = \int (\mathbf{y} - \boldsymbol{\theta}) \frac{\partial}{\partial \mathbf{t}}[w(\mathbf{y}, \mathbf{t})]|_{\boldsymbol{\theta}} dF$ and $V(\boldsymbol{\theta}) = \int w(\mathbf{y}, \mathbf{T}(F)) dF$. $U(\boldsymbol{\theta})$ is a diagonal matrix with components

$$a_{jj} = \int y_i \left[\text{sgn}(y_1) \frac{y_j e^{-\beta \sum_2^p y_i^2}}{\|\mathbf{y}\|\|\boldsymbol{\theta}\|}(1 + 2\beta y_1^2) \right] dF.$$

Furthermore, a_{jj} are identical except $a_{11} = 0$. An upper bound of a_{jj} can be derived as follows:
Let $\bar{\mathbf{y}} = \frac{\mathbf{y}}{\sigma}, \bar{\boldsymbol{\theta}} = \frac{\boldsymbol{\theta}}{\sigma}, \bar{\beta} = \beta\sigma^2$ and $f^*(\bar{y}_1) = f(\bar{y}_1) - f(-\bar{y}_1)$.

$$
\begin{aligned}
a_{jj} &= \iint_0^{\infty} e^{-\bar{\beta} \sum_2^p \bar{y}_i^2} \frac{\bar{y}_j^2(1 + 2\bar{\beta}\bar{y}_1^2)}{\|\bar{\mathbf{y}}\|\|\bar{\boldsymbol{\theta}}\|} f^*(\bar{y}_1) d\bar{y}_1 \prod_2^p f(\bar{y}_i) d\bar{y}_i \\
&\leq \iint_0^{\infty} e^{-\bar{\beta} \sum_2^p \bar{y}_i^2} \frac{|\bar{y}_j|(1 + 2\bar{\beta}\bar{y}_1|\bar{y}_j|)}{\|\bar{\boldsymbol{\theta}}\|} \\
&\quad \times f^*(\bar{y}_1) d\bar{y}_1 \prod_2^p f(\bar{y}_i) d\bar{y}_i \\
&= Q_1(\bar{\beta}, \|\bar{\boldsymbol{\theta}}\|) \prod_{i \neq 1, j} e^{-\bar{\beta} \sum_{i\neq1,j}^p \bar{y}_i^2} f(\bar{y}_i) d\bar{y}_i, \quad (26)
\end{aligned}
$$

$$
\begin{aligned}
&Q_1(\bar{\beta}, \|\bar{\boldsymbol{\theta}}\|) \\
&\triangleq \iint_0^{\infty} e^{-\bar{\beta}\bar{y}_j^2} \frac{|\bar{y}_j|(1 + 2\bar{\beta}\bar{y}_1|\bar{y}_j|)}{\|\bar{\boldsymbol{\theta}}\|} f^*(\bar{y}_1) f(\bar{y}_j) d\bar{y}_1 d\bar{y}_j.
\end{aligned}
$$

Evaluating (26) in spherical coordinates and let $C(p) \triangleq (\frac{1}{\sqrt{2\pi}})^{p-2} \int \sin\phi_1 \cdots \sin\phi_{p-3} d\phi_1 \cdots d\phi_{p-3}$,

$$
\begin{aligned}
\|U(\boldsymbol{\theta})\|_{op} &\leq C(p) Q_1(\bar{\beta}, \|\bar{\boldsymbol{\theta}}\|) \int_0^{\infty} e^{-(\bar{\beta}+\frac{1}{2})r^2} r^{p-3} dr \\
&= C(p) Q_1(\bar{\beta}, \|\bar{\boldsymbol{\theta}}\|) \frac{\Gamma(\frac{p-2}{2})}{2(\bar{\beta}+\frac{1}{2})^{\frac{p-2}{2}}}.
\end{aligned}
$$

Similarly, a lower bound of $V(\boldsymbol{\theta})$ can be derived as:

$$
\begin{aligned}
V(\boldsymbol{\theta}) &= \int \Big[\iint e^{-\bar{\beta}\bar{y}_j} \frac{|\bar{y}_1|}{\|\bar{\mathbf{y}}\|} f(\bar{y}_1) f(\bar{y}_j) \mathrm{d}\bar{y}_1 \mathrm{d}\bar{y}_j \Big] \\
&\quad \cdot \prod_{i\neq 1,j} e^{-\bar{\beta}\sum_{i\neq 1,j}^{p} \bar{y}_i^2} f(\bar{y}_i) \mathrm{d}\bar{y}_i \\
&= C(p) \int_0^{\infty} \Big[\iint \frac{e^{-\bar{\beta}\bar{y}_j}|\bar{y}_1| f(\bar{y}_1) f(\bar{y}_j)}{\sqrt{\bar{y}_1^2 + \bar{y}_j^2 + r^2}} \mathrm{d}\bar{y}_1 \mathrm{d}\bar{y}_j \Big] \\
&\quad \times e^{-(\bar{\beta}+\frac{1}{2})r^2} r^{p-3} \mathrm{d}r \\
&\geq C(p) \iint \frac{e^{-\bar{\beta}\bar{y}_j}|\bar{y}_1| f(\bar{y}_1) f(\bar{y}_j)}{\sqrt{\bar{y}_1^2 + \bar{y}_j^2 + r_{max}^2}} \mathrm{d}\bar{y}_1 \mathrm{d}\bar{y}_j \\
&\quad \times \int_0^{r_{max}} e^{-(\bar{\beta}+\frac{1}{2})r^2} r^{p-3} \mathrm{d}r \\
&= C(p) Q_2(\bar{\beta}, \|\bar{\boldsymbol{\theta}}\|, r_{max}) \frac{\gamma\left(\frac{p-2}{2}, (\bar{\beta}+\frac{1}{2})r_{max}^2\right)}{2(\bar{\beta}+\frac{1}{2})^{\frac{p-2}{2}}},
\end{aligned}
$$

where

$$
Q_2(\bar{\beta}, \|\bar{\boldsymbol{\theta}}\|, r_{max}) \triangleq \iint \frac{e^{-\bar{\beta}\bar{y}_j}|\bar{y}_1| f(\bar{y}_1) f(\bar{y}_j)}{\sqrt{\bar{y}_1^2 + \bar{y}_j^2 + r_{max}^2}} \mathrm{d}\bar{y}_1 \mathrm{d}\bar{y}_j.
$$

Define $Q_3(\bar{\beta}, \|\bar{\boldsymbol{\theta}}\|, p) = \underset{r_{max}}{\operatorname{argmax}}[Q_2(\bar{\beta}, \|\bar{\boldsymbol{\theta}}\|, r_{max})\gamma(\frac{p-2}{2},$ $(\bar{\beta}+\frac{1}{2})r_{max}^2)]$, $V(\boldsymbol{\theta}) \geq C(p)\frac{Q_3(\bar{\beta},\|\bar{\boldsymbol{\theta}}\|,p)}{2(\bar{\beta}+\frac{1}{2})^{\frac{p-2}{2}}}$. Thus, $\|H(\boldsymbol{\theta})\|_{op} =$ $\frac{\|U(\boldsymbol{\theta})\|_{op}}{V(\boldsymbol{\theta})} \leq \frac{Q_1(\bar{\beta},\|\bar{\boldsymbol{\theta}}\|)\Gamma(\frac{p-2}{2})}{Q_3(\bar{\beta},\|\bar{\boldsymbol{\theta}}\|,p)} \triangleq B(\bar{\beta}, \|\bar{\boldsymbol{\theta}}\|, p).$

APPENDIX C
INVERTIBILITY OF $I - Z(M)$

If $B(\bar{\beta}, \|\overline{\boldsymbol{\Theta}_j}\|, p) < 1, \forall 1 \leq j \leq L$, then $\|H(\boldsymbol{\Theta}_j)\|_{op} \leq B(\bar{\beta}, \|\overline{\boldsymbol{\Theta}_j}\|, p) < 1$, where $\boldsymbol{\Theta}_j = C_j M$, which gives

$$
\Big\| \int (\mathbf{Y}_j - C_j M) \frac{\partial}{\partial \mathbf{t}} [w(\mathbf{Y}_i, \mathbf{t})]_{C_j M} \mathrm{d}F_j \Big\|_{op}
$$
$$
< \Big| \int w(\mathbf{Y}_j, C_j M) \mathrm{d}F_j \Big|.
$$

Further, we have shown $\int (\mathbf{Y}_j - C_j M) \frac{\partial}{\partial \mathbf{t}}[w(\mathbf{Y}_j, \mathbf{t})]_{C_j M} \mathrm{d}F_j$ is a diagonal matrix with identical diagonal elements except the first one being zero. Denote this diagonal matrix as $\operatorname{diag}\{0, a_j, \cdots, a_j\}$. Let $b_j \triangleq \int w(\mathbf{Y}_j, C_j M) \mathrm{d}F_j$. Thus $|a_j| < |b_j|$. C_j are diagonal matrices with elements being the components of the CTF. Let $C_j(i)$ denote its ith component.

$$
\begin{aligned}
A &\triangleq \sum_j C_j^T \int (\mathbf{Y}_j - C_j M) \frac{\partial}{\partial \mathbf{t}}[w(\mathbf{Y}_j, \mathbf{t})]_{C_j M} \mathrm{d}F_j C_j \\
&= \operatorname{diag}\{0, \sum_j C_j^2(2) a_j, \cdots, \sum_j C_j^2(p) a_j\}, \text{ and} \\
B &\triangleq \sum_j C_j^T C_j \int w(\mathbf{Y}_j, C_j M) \mathrm{d}F_j \\
&= \operatorname{diag}\{\sum_j C_j^2(1) b_j, \cdots, \sum_j C_j^2(p) b_j\}.
\end{aligned}
$$

Because $Z(M) = B^{-1}A$ and $|a_j| < |b_j| \; \forall i$, $\|Z(M)\|_{op} = \max_i\{|\sum_j C_j^2(i)a_j / \sum_j C_j^2(i)b_j|\} < 1$.

REFERENCES

[1] C. Huang and H. D. Tagare, "Robust estimation for class averaging in cryo-EM single particle reconstruction," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2014, pp. 3329–3332.

[2] J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. San Diego, CA, USA: Academic, Jan. 1996.

[3] K. R. Lata, P. Penczek, and J. Frank, "Automatic particle picking from electron micrographs," *Ultramicroscopy*, vol. 58, nos. 3–4, pp. 381–391, Jun. 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/030439919500002I

[4] T. R. Shaikh, R. Trujillo, J. S. LeBarron, W. T. Baxter, and J. Frank, "Particle-verification for single-particle, reference-based reconstruction using multivariate data analysis and classification," *J. Struct. Biol.*, vol. 164, no. 1, pp. 41–48, Oct. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847708001457

[5] W. V. Nicholson and R. M. Glaeser, "Review: Automatic particle detection in electron microscopy," *J. Struct. Biol.*, vol. 133, nos. 2–3, pp. 90–101, Feb. 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847701943485

[6] H. C. Wong, J. Chen, F. Mouche, I. Rouiller, and M. Bern, "Model-based particle picking for cryo-electron microscopy," *J. Struct. Biol.*, vol. 145, nos. 1–2, pp. 157–167, Jan. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S104784770300296X

[7] S. J. Ludtke, P. R. Baldwin, and W. Chiu, "EMAN: Semiautomated software for high-resolution single-particle reconstructions," *J. Struct. Biol.*, vol. 128, no. 1, pp. 82–97, Dec. 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847799941746

[8] T. R. Shaikh *et al.*, "SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs," *Nature Protocols*, vol. 3, no. 12, pp. 1941–1974, 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2737740/

[9] J. B. Heymann and D. M. Belnap, "Bsoft: Image processing and molecular modeling for electron microscopy," *J. Struct. Biol.*, vol. 157, no. 1, pp. 3–18, Jan. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847706001997

[10] J. Z. Chen and N. Grigorieff, "SIGNATURE: A single-particle selection system for molecular electron microscopy," *J. Struct. Biol.*, vol. 157, no. 1, pp. 168–173, Jan. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847706001961

[11] J. Vargas *et al.*, "Particle quality assessment and sorting for automatic and semiautomatic particle-picking techniques," *J. Struct. Biol.*, vol. 183, no. 3, pp. 342–353, Sep. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847713001950

[12] H. C. Wong, J. Chen, F. Mouche, I. Rouiller, and M. Bern, "Model-based particle picking for cryo-electron microscopy," *J. Struct. Biol.*, vol. 145, nos. 1–2, pp. 157–167, 2004.

[13] A. M. Roseman, "FindEM—A fast, efficient program for automatic selection of particles from electron micrographs," *J. Struct. Biol.*, vol. 145, nos. 1–2, pp. 91–99, Jan. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S104784770300279X

[14] N. Grigorieff, "FREALIGN: High-resolution refinement of single particle structures," *J. Struct. Biol.*, vol. 157, no. 1, pp. 117–125, Jan. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847706001699

[15] A. Stewart and N. Grigorieff, "Noise bias in the refinement of structures derived from single particles," *Ultramicroscopy*, vol. 102, no. 1, pp. 67–84, Dec. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15556702

[16] M. Shatsky, R. J. Hall, E. Nogales, J. Malik, and S. E. Brenner, "Automated multi-model reconstruction from single-particle electron microscopy data," *J. Struct. Biol.*, vol. 170, no. 1, pp. 98–108, Apr. 2010.

[17] S. H. W. Scheres, "RELION: Implementation of a Bayesian approach to cryo-EM structure determination," *J. Struct. Biol.*, vol. 180, no. 3, pp. 519–530, Dec. 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847712002481

[18] S. H. W. Scheres, "Maximum-likelihood methods in cryo-EM. Part II: Application to experimental data," *Methods Enzymol.*, vol. 482, pp. 295–320, 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/

[19] Y. Hashem *et al.*, "Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40s subunit," *Nature*, vol. 503, no. 7477, pp. 539–543, Nov. 2013. [Online]. Available: http://www.nature.com/nature/journal/v503/n7477/full/nature12658.html

[20] W. O. Saxton and J. Frank, "Motif detection in quantum noise-limited electron micrographs by cross-correlation," *Ultramicroscopy*, vol. 2, pp. 219–227, Apr. 1977. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304399176913851

[21] J. Frank, A. Verschoor, and M. Boublik, "Computer averaging of electron micrographs of 40S ribosomal subunits," *Science*, vol. 214, no. 4527, pp. 1353–1355, Dec. 1981.

[22] X. Zeng, H. Stahlberg, and N. Grigorieff, "A maximum likelihood approach to two-dimensional crystals," *J. Struct. Biol.*, vol. 160, no. 3, pp. 362–374, Dec. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847707002304

[23] F. J. Sigworth, "Classical detection theory and the cryo-EM particle selection problem," *J. Struct. Biol.*, vol. 145, nos. 1–2, pp. 111–122, Jan. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847703002405

[24] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA, USA: Addison-Wesley, 1977.

[25] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. New York, NY, USA: Wiley, 1986.

[26] S. J. Ludtke, J. Jakana, J.-L. Song, D. T. Chuang, and W. Chiu, "A 11.5 Å single particle reconstruction of GroEL using EMAN[1]," *J. Molecular Biol.*, vol. 314, no. 2, pp. 253–262, Nov. 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022283601951332

[27] Z. Shang and F. J. Sigworth, "Hydration-layer models for cryo-EM image simulation," *J. Struct. Biol.*, vol. 180, no. 1, pp. 10–16, Oct. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3577369/

[28] R. Norousi *et al.*, "Automatic post-picking using MAPPOS improves particle image detection from cryo-EM micrographs," *J. Struct. Biol.*, vol. 182, no. 2, pp. 59–66, May 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847713000452

[29] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, New York, NY, USA, 2006, pp. 233–240. [Online]. Available: http://doi.acm.org/10.1145/1143844.1143874

[30] N. Banterle, K. H. Bui, E. A. Lemke, and M. Beck, "Fourier ring correlation as a resolution criterion for super-resolution microscopy," *J. Struct. Biol.*, vol. 183, no. 3, pp. 363–367, Sep. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847713001184

[31] A. Kucukelbir, F. J. Sigworth, and H. D. Tagare, "A Bayesian adaptive basis algorithm for single particle reconstruction," *J. Struct. Biol.*, vol. 179, no. 1, pp. 56–67, Jul. 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847712001153

[32] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Jun. 2005, pp. 60–65.

Chenxi Huang received the B.E. degree in information engineering from Shanghai Jiao Tong University, and the M.S. degree in electrical engineering from Yale University. She is currently pursuing the Ph.D. degree with the Department of Biomedical Engineering, Yale University. Her research interests are mathematical and computational approaches to biomedical data analysis.

Hemant D. Tagare (SM'87) received the Ph.D. degree from Rice University. His research interests are in biomedical signal and image analysis. He is a Professor with the Department of Diagnostic Radiology, the Department of Biomedical Engineering, and the Department of Electrical Engineering, Yale University.