Heterogeneous multireference alignment for images with application to 2-D classification in single particle reconstruction

Chao Ma[1], **Tamir Bendory**[1], **Nicolas Boumal**[2], **Fred Sigworth**[3], **Amit Singer**[1,2]

[1]The Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ, USA

[2]Department of Mathematics, Princeton University, Princeton, NJ, USA

[3]Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, CT, USA

Abstract

Motivated by the task of 2-D classification in single particle reconstruction by cryo-electron microscopy (cryo-EM), we consider the problem of heterogeneous multireference alignment of images. In this problem, the goal is to estimate a (typically small) set of target images from a (typically large) collection of observations. Each observation is a rotated, noisy version of one of the target images. For each individual observation, neither the rotation nor which target image has been rotated are known. As the noise level in cryo-EM data is high, clustering the observations and estimating individual rotations is challenging. We propose a framework to estimate the target images directly from the observations, completely bypassing the need to cluster or register the images. The framework consists of two steps. First, we estimate rotation-invariant features of the images, such as the bispectrum. These features can be estimated to any desired accuracy, at any noise level, provided sufficiently many observations are collected. Then, we estimate the images from the invariant features. Numerical experiments on synthetic cryo-EM datasets demonstrate the effectiveness of the method. Ultimately, we outline future developments required to apply this method to experimental data.

Keywords

multireference alignment; cryo-EM; bispectrum; steerable PCA; single particle reconstruction

---

I. Introduction

Single particle reconstruction using cryo-EM is a high-resolution imaging technique used in structural biology to image 3-D structures of macromolecules [1], [2]. In a cryo-EM experiment, multiple samples of a particle are frozen in a thin layer of vitreous ice. Within the ice, the samples are randomly oriented and positioned. The electron microscope produces a tomographic image of the ice and the embedded samples, called a micrograph. The goal is then to estimate the 3-D structure of the particle from the micrograph.

Importantly, the signal to noise ratio (SNR) of the micrograph is usually low because of the limited electron dose that can be applied without causing excessive radiation damage.

The first stage of existing cryo-EM algorithmic pipelines is called *particle picking*. In this stage, one aims to detect the projections of the samples within the micrograph and extract them. We refer to these extracted images as *projection images*. Throughout this paper, we assume perfect particle picking, that is, we obtain a large number of projection images, each containing a centered particle projection corresponding to an unknown viewing direction.

An important intermediate stage in the 3-D reconstruction procedure of cryo-EM is called *2-D classification*. The goal of this stage is to produce K 2-D images—called *class averages*—with higher SNR. Each one of the K images should represent a subset of the projections taken from a similar viewing direction. The 2-D class averages can be used as templates for particle picking, to construct *ab initio* 3-D structures [3]–[6], to provide a quick assessment of the particles, to remove picked particles which are associated with non-informative classes, and for symmetry detection [7]–[9].

Different solutions were proposed for the 2-D classification problem. One approach is the reference-free alignment (RFA) technique [10]. RFA tries to align all projection images globally by estimating all individual rotation parameters. However, when the images arise from many different viewing directions, RFA tends to produce large errors as no assignment of in-plane rotational angles can align all images simultaneously. Methods based on expectation-maximization (EM)—an iterative algorithm that aims to find the marginalized maximum likelihood—are also popular. In the context of cryo-EM, the method is usually referred to as Maximum Likelihood 2-D classification (ML2D). The method was first proposed in [11], and is implemented in the popular software package RELION [12], [13]. Nevertheless, the EM framework lacks theoretical analysis and may be computationally expensive. In addition, EM suffers from an intrinsic resolution-computational load trade-off, since the sampling of the in-plane rotation angles is Nyquist sampling. In Section IV we present some numerical results of EM and discuss more of its properties.

A different 2-D classification technique is based on multireference alignment (MRA) [14]–[16]. In MRA, the images are clustered into K classes and the images within each class are averaged to suppress the noise. The averaged images are the class averages. As projection images can be similar up to rotation, the clustering is based on either rotationally aligning the images within each class, or on features of the images which are invariant under rotations, such as autocorrelation [17] or bispectrum [18]. MRA and invariant features play a key role in this paper and are discussed in detail later. Notably, our proposed method avoids the clustering stage, which may be inaccurate at low SNR. Instead, we aim to estimate the K class averages directly from the projection images, with no intermediate clustering stage.

In this paper, we propose to model the 2-D classification problem as an instance of the *heterogeneous multireference alignment* (hMRA) problem, for the case of 2-D images [19], [20]. In the hMRA problem, we observe N projection images. Each observed image is an in-plane rotated, noisy version of one of the K underlying images—the class averages, corresponding to K viewing directions. For each observation, the specific underlying image

and the in-plane rotation are unknown. Since in cryo-EM all in-plane rotations are equally likely to appear, we assume uniform distribution of rotations. The goal is to estimate the K class averages, as well as the distribution among the class averages. Crucially, for each observed particle, which class average it came from (its label) and which in-plane rotation was applied to it are treated as *nuisance variables*; that is: they are unknown, but we do not seek to estimate them. A detailed mathematical model of the hMRA problem is provided in Section II.

In the high SNR regime, the nuisance variables could be estimated accurately, at least in principle. Given an accurate estimate of these variables, the problem becomes trivial: one can cluster the observations into the K class averages, undo the rotations, and average within each class to suppress the noise. However, in the low SNR regime, estimating the labels and rotations becomes challenging, and indeed impossible as the SNR drops to zero; see for instance [21] for analysis in a related model. Notwithstanding, it was shown in a series of papers that in many MRA setups the underlying signal (or signals in our case) can be estimated at any noise level, provided sufficiently many observations are recorded [19], [22]–[27]. Remarkably, it was shown that in the low SNR regime, the method of moments achieves the optimal sample complexity under rather moderate conditions [24], [28].

Consequently, targeting the low SNR regime, we propose to apply the method of moments of MRA to the 2-D classification problem in cryo-EM. Our work builds upon the notion of *bispectrum*, first proposed by Tukey [29], and currently used in signal processing [30], [31]. The bispectrum is invariant under rotations; that is, the bispectrum of an image remains unchanged after an arbitrary in-plane rotation. This property enables us to bypass estimation of individual rotations associated with each one of the observations. Under the assumption of uniform distribution of rotations, the bispectrum is equivalent to the third-order moment of the image. Inspired by the seminal work of Kam [32], previous works [20], [23] studied the MRA and hMRA problems for 1-D signals using the bispectrum as a simplified model for the 3-D reconstruction problem in cryo-EM. In this paper, we study the more involved hMRA model for 2-D images as a model for 2-D classification.

In a nutshell, our proposed approach for 2-D classification consists of the following stages. First, we expand each image in a *steerable basis*. In this paper, we use the Fourier-Bessel basis, but alternative bases, such as the prolate spheroidal wave functions, can be used alternatively [33]. As explained in Section II, in such a basis all in-plane rotations of an image admit the same expansion coefficients, up to complex phase modulations. This property is called *steerability*. As a result, specific monomials in these coefficients are invariant under in-plane rotations. We refer to these monomials as *invariant features*. Specifically, we make use of monomials of the first-, second- and third-order called the mean, power spectrum and bispectrum, respectively.

In practice, instead of working directly on the expansion coefficients, we employ a dimensionality-reduction and denoising technique called *steerable principal component analysis* (sPCA). This technique is similar to the standard PCA, while boosting the SNR by accounting for all in-plane rotations of the data in an efficient way [34]. In addition, the

sPCA coefficients preserve the steerability property. Therefore, the invariant monomials can be computed in the sPCA space.

After computing the invariant features of each image, we average over all images. These averages are consistent estimators (up to bias terms that can be removed easily) of the mixed invariant features of the K class averages. Ultimately, a nonconvex least squares (LS) optimization problem is designed to recover the sPCA coefficients of the individual class averages from these mixed invariant features. All the ingredients of this algorithmic pipeline are provided in Sections II and III. Figure 1 illustrates the flowchart of the procedure. Numerical results and comparison with EM are provided in Section IV. Section V concludes this work, discusses its limitations and potential future extensions.

II. STATISTICAL MODEL AND INVARIANT FEATURES

In this section, we first describe in detail the hMRA model. Then, we introduce our framework based on computing features that are invariant under rotations in a steerable basis.

A. Statistical model

Let $\{I_1, I_2, ..., I_K\}$ be a set of K images of size $L \times L$, with L odd[1] and pixel values in [0, 1]: these are the class averages, our target parameters. The pixels in an image are indexed by a pair of integers (x, y) with $-(L-1)/2 \le x, y \le (L-1)/2$. The support of the images is assumed to lie in the disk $x^2 + y^2 \le (L-1)^2/4$; as a result, any of their rotations have the same property. Let R_θ be a rotation operator which rotates an image counter-clockwise by angle θ, and let ξ be a random variable following a uniform distribution on $[0, 2\pi]$. In addition, let π be a random variable on the set $\{1, 2, ..., K\}$ with distribution $(\pi_1, ..., \pi_K)$:

$$\pi_k := \mathbb{P}(\pi = k) > 0, \quad k \in \{1, 2, ..., K\}.$$

Then, our observations are i.i.d. random samples from the model

$$Y = R_\xi I_\pi + \varepsilon,$$

(II.1)

where $\varepsilon = (\varepsilon_{ij}) \in \mathbb{R}^{L \times L}$ is a noise matrix of i.i.d. Gaussian variables with zero mean and variance σ^2; the random variables ξ, π, ε are independent. Indeed, it was observed that the background noise in cryo-EM experiments can be treated as Gaussian [35].

Suppose we collect N independent observations from the generative model works when L is even. (II.1),

$$\mathcal{Y} = \{Y_1, Y_2, ..., Y_N\},$$

---

[1]We consider an odd L for convenience of implementation and presentation. Our algorithm also works when L is even.

so that

$$Y_i = R_{\xi_i} I_{\pi_i} + \varepsilon_i, \quad i = 1, \ldots, N.$$

From the observed data $\mathscr{Y}$, we seek to estimate the target images $\{I_1, I_2, \ldots, I_K\}$ (class averages) and, possibly, the distribution π, without estimating the in-plane rotations of individual observations ξ_i or the labels π_i. The model (II.1) suffers from unavoidable ambiguities of rotations (of each class average) and permutation (across the K images). Therefore, naturally, a solution is defined up to these symmetries. In Section IV we define a suitable error metric.

B. Steerable basis

As mentioned above, we aim to bypass estimating the nuisance variables by computing features that are invariant under rotation. To this end, we first expand the images with respect to a steerable basis. In polar coordinates, steerable basis functions take the form

$$u^{k,q}(r, \theta) = f^{k,q}(r) e^{\imath k \theta},$$

(II.2)

where $\imath := \sqrt{-1}$. Notice the separation of variables: If we expand an image in $u^{k,q}$,

$$I(r, \theta) = \sum_{k,a} a_{k,q} u^{k,q}(r, \theta),$$

(II.3)

then the expansion of the rotated image follows from:

$$(R_\alpha I)(r, \theta) = I(r, \theta - \alpha) = \sum_{k,a} a_{k,q} u^{k,q}(r, \theta - \alpha)$$
$$= \sum_{k,q} a_{k,q} e^{-\imath k \alpha} u^{k,q}(r, \theta).$$

(II.4)

Since our images are real, the coefficients satisfy a conjugacy symmetry: $a_{k,q} = \overline{a_{-k,q}}$. Therefore, coefficients with $k \geq 0$ suffice to represent the images.

Examples of steerable bases include the Fourier-Bessel basis and prolate spheroidal wavefunctions. See [34] [36] [33] [37] for efficient expansion algorithms. In this paper, we work with the Fourier-Bessel basis on a disk with radius c defined as:

$$u^{k,q}(r, \theta) = \begin{cases} N_{k,q} J_k\left(R_k \dfrac{r}{c}\right) e^{\imath k \theta}, & r \leq c, \\ 0, & r > c, \end{cases}$$

(II.5)

where J_k is the Bessel function of the first kind, $R_{k,q}$ is the qth root of J_k and $N_{k,q} = \left(c\sqrt{\pi}|J_{k+1}(R_{k,q})|\right)^{-1}$ is a normalization factor. We take c to be $(L-1)/2$, in accordance with the assumed support of the images.

To reduce the dimensionality of the representation and denoise the image, we perform sPCA after expanding the images in a steerable basis [34]. The sPCA results in a new, data driven basis to represent the images. Importantly, this new basis preserves the steerability property and consequently the rotation property (II.4) holds true. Section III-A introduces the sPCA technique in more details. With some abuse of notation, in what follows the coefficients in a sPCA basis are also denoted by $\{a_{k,q}\}$.

C.   Invariant features

The steerability property (II.4) enables us to determine features of images which are invariant under rotation. Specifically, features that are invariant to an action of SO (2): the special orthogonal group in 2-D. We assume that the images are "band-limited" in the sense that their expansion in a steerable basis is finite.

From (II.4), it is clear that coefficients $a_{k,q}$ corresponding to $k = 0$ are not affected by rotation. Hence, the first-order invariants are just the mean values, or the "DC components":

$$m_q = a_{0,q},$$

(II.6)

for all q. The second-order invariants, which form the power spectrum, are given by

$$p_{k,q_1,q_2} = a_{k,q_1}\overline{a_{k,q_2}},$$

(II.7)

for all k, q_1, q_2. The power spectrum coefficients are invariant to rotation since for all α:

$$\left(a_{k,q_1}e^{-\imath k\alpha}\right)\overline{\left(a_{k,q_2}e^{-\imath k\alpha}\right)} = a_{k,q_1}\overline{a_{k,q_2}}.$$

Unfortunately, the power spectrum does not determine the image uniquely: a multiplication of the expansion coefficients by $e^{\imath h[k]}$ for an arbitrary function h does not change the power spectrum, yet it does change the image.

The third-order invariant, the bispectrum, is defined by

$$b_{k_1,k_2,q_1,q_2,q_3} = a_{k_1,q_1}a_{k_2,q_2}\overline{a_{k_1+k_2,q_3}},$$

(II.8)

for all k_1, k_2, q_1, q_2, q_3. Using equation (II.4), one can verify that indeed:

$$\left(a_{k_1, q_1}e^{-\imath k_1 \alpha}\right)\left(a_{k_2, q_2}e^{-\imath k_2 \alpha}\right)\overline{\left(a_{k_1 + k_2, q_3}e^{-\imath(k_1 + k_2)\alpha}\right)}$$
$$= \left(a_{k_1, q_1}a_{k_2, q_2}\overline{a_{k_1 + k_2, q_3}}\right)\left(e^{-\imath k_1 \alpha}e^{-\imath k_2 \alpha}\overline{e^{-\imath(k_1 + k_2)\alpha}}\right)$$
$$= a_{k_1, q_1}a_{k_2, q_2}\overline{a_{k_1 + k_2, q_3}}.$$

The combined power spectrum and bispectrum do determine an image uniquely, up to global rotation:

Theorem II.1.—*Consider two images with steerable basis coefficients $a_{k,q}$ and $a'_{k,q}$, respectively, in the range $-k_{max} \le k \le k_{max}$. Assume that for all $-k_{max} \le k \le k_{max}$ there exists at least one q such that $a_{k,q} \ne 0$. If for all indices*

$$a_{k, q_1}\overline{a_{k, q_2}} = a'_{k, q_1}\overline{a'_{k, q_2}},$$

(II.9)

$$a_{k_1, q_1}a_{k_2, q_2}\overline{a_{k_1 + k_2, q_3}} = a'_{k_1, q_1}a'_{k_2, q_2}\overline{a'_{k_1 + k_2, q_3}},$$

(II.10)

then there exists $\theta \in [0, 2\pi)$ such that

$$a'_{k, q} = a_{k, q}e^{-\imath k \theta}.$$

(II.11)

for all k, q. That is, the two images only differ by a rotation.

Proof: Set $q_1 = q_2$ in (II.9), we have $|a_{k,q}| = |a'_{k,q}|$ for any k and q. Hence, $a'_{k,q} \ne 0$ if and only if $a_{k,q} \ne 0$, and there exists $\theta_{k,q} \in [0, 2\pi)$ such that $a'_{k,q} = a_{k,q}e^{-\imath\theta_{k,q}}$. Then, still by (II.9), we have

$$a'_{k, q_1}\overline{a'_{k, q_2}} = a_{k, q_1}\overline{a_{k, q_2}}e^{-\imath\left(\theta_{k, q_1} - \theta_{k, q_2}\right)}.$$

This means that, for fixed k, $\theta_{k,q}$ take a same value (in $[0, 2\pi)$) for all q satisfying $a_{k,q} \ne 0$. Hence, for each k, there exists a single $\theta_k \in [0, 2\pi)$ such that

$$a'_{k, q} = a_{k, q}e^{-\imath\theta_k}.$$

Next, by (II.10), we have for all k_1, k_2, q_1, q_2, and q_3,

$$a'_{k_1, q_1}a'_{k_2, q_2}\overline{a'_{k_1 + k_2, q_3}} = a_{k_1, q_1}a_{k_2, q_2}\overline{a_{k_1 + k_2, q_3}}$$
$$\times e^{-\imath\left(\theta_{k_1} + \theta_{k_2} - \theta_{k_1 + k_2}\right)}.$$

By assumption, we can always choose q_1, q_2 and q_3 such that a_{k_1,q_1}, a_{k_2,q_2} and $a_{k_1+k_2,q_3}$ are all nonzero. Then, we have

$$\theta_{k_1} + \theta_{k_2} = \theta_{k_1+k_2},$$

for all $-k_{max} \leq k_1, k_2 \leq k_{max}$. This in turn implies

$$\theta_k = k\theta,$$

for some constant θ. Thus we conclude that

$$a'_{k,q} = a_{k,q}e^{-\imath k\theta}.$$

∎

III. 2-D CLASSIFICATION USING INVARIANT FEATURES

In this section, we introduce the whole pipeline of our algorithm in detail. We start by expanding the observed images in Fourier-Bessel basis. We then perform sPCA on the resulting coefficients. Next, we estimate the mixed invariants of the K true images (class averages) from the observations' sPCA coefficients. Ultimately, we estimate the sPCA coefficients of the true images, and thus the images themselves, from the mixed invariants via a nonconvex LS optimization problem. Figure 1 shows a schematic flow chart of our algorithm and Algorithm 1 describes our algorithm step by step. Next, we elaborate on each of the steps.

Algorithm 1:

2-D classification by invariant features

---

1 **Input:** Observations $Y_1, \ldots, Y_N$; noise variance σ^2.

2 Expand the observations in the Fourier-Bessel basis, and perform sPCA on the expansion coefficients using the method described in [34].

3 Estimate the mixed invariants of the true images using the sPCA coefficients of the data by (III.5), (III.6) and (III.7).

4 Estimate the sPCA coefficients of the true images and the distribution π by solving the optimization problem (III.13).

5 Recover the images $\hat{I}_1, \ldots, \hat{I}_K$ from the sPCA coefficients by (III.16).

6 **Output:** images $\hat{I}_1, \ldots, \hat{I}_K$ (up to permutation and rotations); distribution $\hat{\pi}$.

A. Fourier-Bessel sPCA

We perform sPCA on the images after they were expanded in a Fourier-Bessel steerable basis, introduced in Section II-B. Like in a standard PCA, the first step is to subtract the mean observed image from each observation to center the data. The mean image is added

back at the last step of the PCA. To ease exposition, we assume (only in this subsection) that the images have zero mean.

For a regular PCA, one would construct the data matrix X such that each column holds the expansion coefficients of one image. Then, PCA would extract the dominant eigenvectors of XX^*, where X^* is the conjugate transpose of X. Note that since XX^* is Hermitian, it has real eigenvalues. Since in our model (and in cryo-EM) any observed images could have been observed after arbitrary in-plane rotation with the same probability, we wish to include all such rotated versions of all images in the PCA procedure. This can be done efficiently owing to steerability, as described in [38]. The dominant eigenimages obtained through sPCA form an orthonormal basis for a lower dimensional subspace, where we now project all observations. Crucially, this eigenbasis is also steerable (since each eigenimage is a linear combination of steerable basis functions.) We still get the two usual benefits of PCA—dimensionality reduction and denoising—with the added benefit that we exploited all of the available information. The sPCA has been proven to be an effective denoising tool for cryo-EM reconstruction [18], [33], [39], [40]. The procedure is actually faster than standard PCA because the data covariance XX^* is block-diagonal upon factoring in all rotated images. In the next subsection, we use the expansion coefficients in the sPCA basis to compute invariant features.

B. Estimating the invariant features of the class average images

After performing sPCA, we get a steerable eigenbasis (a collection of eigenimages) and the expansion coefficients of the observed images in that basis (after projection to the corresponding subspace). Then, we can compute the invariant features using these coefficients. The invariants can be computed according to equations (II.6), (II.7) and (II.8). Next, we estimate the mixed invariants of the underlying class average images using the invariants of the noisy data, which we now explain.

Let $\{a_{k,q}^i\}$ be the sPCA coefficients of the ith target image I_i. As per our model (II.1), in the absence of noise, the coefficients of an observation Y are given by $a_{k,q}^\pi e^{-\iota k \xi}$. Let $b_{k_1, k_2, q_1, q_2, q_3}^{\pi, \xi}$ be the bispectrum computed from the latter. By construction, this is independent of ξ: this is simply the bispectrum of the target image I_π. Marginalizing over the remaining nuisance variable π, we find

$$\mathbb{E}_\pi b_{k_1, k_2, q_1, q_2, q_3}^{\pi, \xi} = \mathbb{E}_\pi a_{k_1, q_1}^\pi a_{k_2, q_2}^\pi \overline{a_{k_1 + k_2, q_3}^\pi},$$

(III.1)

where $\mathbb{E}_\pi$ represents expectation taken against π: a sum over $\pi = 1, \ldots, K$ weighted by $(\pi_1, \ldots, \pi_K)$.

This relation implies that by averaging over all the bispectra of the observations we can estimate the mixed bispectra of the K class averages. Estimation of mixed mean and power spectra can be similarly obtained. Crucially, we approximate the mixed invariants of the true images without estimating the rotations ξ or the labels π of individual observations.

The same method can be applied in the presence of noise. Now, the coefficients of the observations are given by

$$a_{k,q}^{\pi} e^{-\imath k\xi} + \varepsilon_{k,q}^{c} e^{-\imath k\xi},$$

(III.2)

where ε^c denotes the complex Gaussian noise in the coefficients, satisfying $\mathbb{E}\varepsilon^c = 0$ and $\mathbb{E}\varepsilon^c(\varepsilon^c)^* = \sigma^2 I$, where I is the identity matrix. The noise terms induce bias in the power spectrum and bispectrum estimation. Particularly, the noisy power spectrum of Y satisfies

$$\mathbb{E}_{\pi,\varepsilon} p_{k,q_1,q_2}^{\pi,\xi,\varepsilon} = \mathbb{E}_{\pi} a_{k,q_1}^{\pi} \overline{a_{k,q_2}^{\pi}} + \sigma^2 \delta_{q_1,q_2},$$

(III.3)

where δ_{q_1,q_2} is the Kronecker delta function. Hence, we get a bias term which depends solely on σ, which is usually estimated in the cryo-EM algorithmic pipeline. Similarly, expectation over the noisy bispectrum results in

$$\mathbb{E}_{\pi,\varepsilon} b_{k_1,k_2,q_1,q_2,q_3}^{\pi,\xi,\varepsilon} = \mathbb{E}_{\pi} a_{k_1,q_1}^{\pi} a_{k_2,q_2}^{\pi} \overline{a_{k_1+k_2,q_3}^{\pi}} + \sigma^2 \mathbb{E}_{\pi} A^{\pi},$$

(III.4)

where

$$A^{\pi} := \delta_{q_2,q_3} \delta_{k_1,0} a_{0,q_1}^{\pi} + \delta_{q_1,q_3} \delta_{k_2,0} a_{0,q_2}^{\pi} + \delta_{q_1,q_2} \delta_{k_1+k_2,0} a_{0,q_3}^{\pi}.$$

Here, the bias term depends on both σ^2 and the coefficients $a_{0,q}^{\pi}$. Noise does not introduce bias in estimates of the mean.

Equipped with (III.3) and (III.4), estimating the mixed invariants can be executed by averaging over the invariants of the observations and removing the bias terms. Specifically, let $m_q^{Y_i}, p_{k,q_1,q_2}^{Y_i}, b_{k_1,k_2,q_1,q_2,q_3}^{Y_i}$ be, respectively, the mean, power spectrum and bispectrum of the noisy observation Y_i. Then, our estimators of the mixed invariants are easily computed as:

$$\widehat{m}_q = \frac{1}{N} \sum_{i=1}^{N} m_q^{Y_i},$$

(III.5)

$$\widehat{p}_{k,q_1,q_2} = \frac{1}{N} \sum_{i=1}^{N} p_{k,q_1,q_2}^{Y_i} - \sigma^2 \delta_{q_1,q_2},$$

(III.6)

$$\hat{b}_{k_1, k_2, q_1, q_2, q_3} = \frac{1}{N} \sum_{i=1}^{N} b_{k_1, k_2, q_1, q_2, q_3}^{Y_i} - \sigma^2 E_\pi A^\pi.$$

(III.7)

In (III.7), the bias term $E_\pi A^\pi$ can be estimated by $\hat{m}_q$.

C. Estimating the coefficients of the class averages

In the last section, we showed how the mixed invariants of the class averages can be estimated from the data. Now, we turn our attention to estimating the sPCA coefficients of the class average images from their mixed invariants by a LS optimization problem.

Our optimization problem consists of two types of variables. The first type represents the sPCA coefficients of the target images:

$$\hat{a}^i = \left\{ \hat{a}_{k,q}^i \right\}, \quad i = 1, 2, ..., K.$$

(III.8)

The second type represents the distribution from which the observations are sampled:

$$\hat{\pi} = (\hat{\pi}_1, ... \hat{\pi}_K).$$

(III.9)

In practice, as long as we have sufficiently many observations so that the empirical estimates of the invariant features are accurate, the following identities hold approximately:

$$\sum_{i=1}^{K} \pi_i a_{0,q}^i \approx \hat{m}_{0,q},$$

(III.10)

$$\sum_{i=1}^{K} \pi_i a_{k,q_1}^i \overline{a_{k,q_2}^i} \approx \hat{p}_{k,q_1,q_2},$$

(III.11)

$$\sum_{i=1}^{K} \pi_i a_{k_1,q_2}^i a_{k_2,q_2}^i \overline{a_{k_1+k_2,q_3}^i} \approx \hat{b}_{k_1,k_2,q_1,q_2,q_3}.$$

(III.12)

In an ideal case, we want $\hat{a}^i$ to be coefficients of I_i (or an in-plane rotated version thereof), and $\hat{\pi}_i = \pi_i$. Hence, we design an LS problem to minimize the difference

between left- and right-hand sides of equations (III.10), (III.11) and (III.12). Let $M_q(\hat{a}, \hat{\pi})$, $P_{k, q_1, q_2}(\hat{a}, \hat{\pi})$, $B_{k_1, k_2, q_1, q_2, q_3}(\hat{a}, \hat{\pi})$ capture the left-hand sides of the three equations above, respectively, with the estimators $(\hat{a}, \hat{\pi})$ rather the unknown, underlying parameters (a, π). Our objective function reads

$$
\begin{aligned}
F(\hat{a}, \hat{\pi}) = {} & \sum_q \left| M_q(\hat{a}, \hat{\pi}) - \hat{m}_{0, q} \right|^2 \\
& + \frac{1}{1 + \sigma^2} \sum_{k, q_1, q_2} \left| P_{k, q_1, q_2}(\hat{a}, \hat{\pi}) - \hat{p}_{k, q_1, q_2} \right|^2 \\
& + \frac{1}{1 + \sigma^2 + \sigma^4} \sum_{k_1, k_2, q_1, q_2, q_3} \left| B_{k_1, k_2, q_1, q_2, q_3}(\hat{a}, \hat{\pi}) - \hat{b}_{k_1, k_2, q_1, q_2, q_3} \right|^2.
\end{aligned}
$$

(III.13)

Here we use 1, $1 + \sigma^2$ and $1 + \sigma^2 + \sigma^4$ as rough estimates for the variances of corresponding terms; see also [20].

Some constraints need to be imposed. First, as variables $\hat{\pi}$ are used to represent a distribution, they should lie on the simplex, that is, $\hat{\pi}_i \geq 0$ and $\hat{\pi}_1 + \cdots + \hat{\pi}_K = 1$. Second, as our images are real, and Fourier-Bessel sPCA basis functions with $k = 0$ are real, we have $\Im(a_{0, q}^i) = 0$ (where $\Im$ extracts imaginary part). Consequently, we can force $\hat{a}_{0, q}^i$ to be real. Similarly, coefficients with the same q but opposite k are conjugate,

$$
\hat{a}_{k, q}^i = \overline{\hat{a}_{-k, q}^i}.
$$

(III.14)

In the optimization problem, we can just consider those coefficients with nonnegative k.

To conclude, we aim to solve the following constrained LS optimization problem:

$$
\begin{aligned}
& \min_{\hat{a}_{k, q}^i \in \mathbb{C}, \ \hat{\pi}_i \in \mathbb{R}} \quad F(\hat{a}, \hat{\pi}), \\
& \text{subject to} \quad \sum_{i = 1}^{K} \hat{\pi}_i = 1, \quad \hat{\pi}_i \geq 0, \\
& \qquad\qquad\quad \Im(\hat{a}_{0, q}^i) = 0.
\end{aligned}
$$

(III.15)

While the LS is nonconvex, we find that we can solve it satisfactorily in practice—see Section IV. This is in line with recent related work [20], [23], [26], [41].

We attempt to solve the optimization problem (III.15) by a trust-regions algorithm or conjugate gradient method using Manopt [42]: a toolbox for optimization on manifolds[2]. In our problem, the variable $\hat{a}$ lies in Euclidean space, while $\hat{\pi}$ lies on the simplex, whose relative interior is endowed with a Riemannian geometry in the toolbox [43]. When using

---

[2] www.manopt.org

Manopt, we only provide the gradient on Euclidean space. The gradient on the manifold is computed from the gradient on Euclidean space together with the representation of the manifold. The Hessian is approximated automatically by finite differences of the gradient.

D. Recovery of the images

After solving the optimization problem, we obtain a collection of coefficients $\hat{a}^i_{k,q}$, which are believed to approximate the sPCA coefficients of the target images, $a^i_{k,q}$. To recover the images themselves, up to rotation, we simply compute a linear combination of the basis images given by the sPCA, using coefficients $\hat{a}^i$. We then add the mean image that was subtracted from all observations during the sPCA preprocessing, denoted by I^i_m—see Section III-A. Specifically, letting Φ be the sPCA basis, we recover the images by

$$\hat{I}_i = \Phi\hat{a}^i + I^i_m.$$

(III.16)

E. Computational complexity

In this section we discuss the computational complexity of each step of Algorithm 1. By [34], the computational cost of the sPCA step is $\mathcal{O}\left(NL^3 + L^4\right)$, where N is the number of observations and L is the side length of the images. For each image, assume the sPCA provides M components and the maximum angular frequency is k_{max}. Then, we obtain $\mathcal{O}\left(\frac{M^3}{k_{max}}\right)$ invariants in total [18]. Hence, $\mathcal{O}\left(\frac{NM^3}{k_{max}}\right)$ computations are required to compute the invariants of all observations and estimate the mixed invariants for groundtruth images. Next, in the optimization part, computing the gradient requires going through all terms in the objective function, and each term contributes $\mathcal{O}(K)$ elements of the gradient. Hence, if T iterations are performed, the computational complexity of the optimization step is $\mathcal{O}\left(\frac{TKM^3}{k_{max}}\right)$.

Finally, building the recovered images just involves K linear combinations of the principal components with computational cost $\mathcal{O}\left(KML^2\right)$.

IV. NUMERICAL EXPERIMENTS

In this section, we show results of numerical experiments using our algorithm. First, we use random projections of the E. coli 70S ribosome volume [44] as the groundtruth images to explore the performance of our algorithm under different noise levels and distributions. The volume is available in the software ASPIRE package[3]. The size of each image is 129^2 (i.e., $L = 129$). Figure 2 shows some examples of the class averages and noisy input images. Later, we apply our algorithm on two other molecules with projections of larger size. Code for our algorithm and all experiments is available at https://github.com/chaom1026/2DhMRA. The experiments presented below are conducted by MATLAB on a machine with 4× E7–8880 v3 CPUs, and 750GB of RAM.

---

[3] www.spr.math.princeton.edu

Following [20], we define error metrics suitable for the inherent symmetries of our problem. For two images I and $\hat{I}$, we define a rotationally invariant distance as

$$\mathrm{dist}(I, \hat{I}) = \min_{\theta \in [0, 2\pi]} \left\| R_\theta I - \hat{I} \right\|_{\mathrm{F}}.$$

(IV.1)

This distance measures the Frobenius norm between all rotational alignments of I and $\hat{I}$. Let $\mathcal{I} = \{I_1, ..., I_K\}$ be a set of underlying images, and let $\widehat{\mathcal{I}} = \left\{ \hat{I}_1, ..., \hat{I}_K \right\}$ be the estimates. To be invariant under both rotations and permutations, we use the following definition:

$$\mathrm{dist}(\mathcal{I}, \widehat{\mathcal{I}})^2 = \min_{p \in S_K} \sum_{i=1}^{K} \mathrm{dist}\left(I_i, \hat{I}_{p(i)}\right)^2,$$

(IV.2)

where S_K is the set of all permutations of $\{1, 2, ..., K\}$. The relative error between $\mathcal{I}$ and $\widehat{\mathcal{I}}$ is defined by

$$\mathrm{dist}_r(\mathcal{I}, \widehat{\mathcal{I}}) = \frac{\mathrm{dist}(\mathcal{I}, \widehat{\mathcal{I}})}{\sqrt{\sum_{i=1}^{K} \|I_i\|_{\mathrm{F}}^2}}.$$

(IV.3)

Note that when computing errors we only consider the disk area with diameter L and ignore the corners. We measure the error between $\hat{\pi}$ and the true distribution π by the total variation (TV) distance which takes values in [0,1]:

$$\mathrm{dist}_{\mathrm{TV}}(\hat{\pi}, \pi) = \frac{1}{2} \sum_{i=1}^{K} |\hat{\pi}_i - \pi_i|.$$

(IV.4)

Here we assume that a permutation given by equation (IV.2) has been applied to $\hat{\pi}$. In what follows, we define the SNR as

$$\mathrm{SNR} = \frac{\mathbb{E}\left(\mathrm{Signal}^2\right)}{\mathbb{E}\left(\mathrm{Noise}^2\right)} := \frac{\sum_{i=1}^{K} \|I_i\|_{\mathrm{F}}^2}{K L^2 \sigma^2}.$$

(IV.5)

During sPCA, we use the method introduced in [34] to choose the eigenimages automatically, based on properties of the Marchenko-Pastur distribution. Specifically, for each frequency k, we take those eigenimages with eigenvalues satisfying

$$\lambda^{(k)} > 1.005\sigma^2\left(1 + \sqrt{\gamma_k}\right),$$

<div align="right">(IV.6)</div>

where σ^2 is the variance of the noise, $\gamma_0 = \frac{p_0}{N}$, and $\gamma_k = \frac{p_k}{2N}$ for $k \neq 0$. Here, p_k is the number of eigenimages for frequency k and N is the total number of observations. The factor 1.005 in (IV.6) is chosen heuristically to control the number of sPCA coefficients.

In our first experiment we consider a uniform distribution π, and assume that the algorithm knows that π is uniform. Hence, in the optimization problem, variables $\hat{\pi}_i$ are fixed to be $1/K$. We choose $K = 10$, SNR = 1/50, and take $N = 10^4$ observations in total. Examples of noisy observations are shown in the third column of Figure 3. With this much noise, it would be challenging to rotationally align and cluster the observations. Nevertheless, the proposed algorithm gets estimates of the images without (even implicitly) doing either alignment or clustering. During sPCA, 83 coefficients are automatically chosen to represent each image, according to (IV.6). Hence, in total we have 830 variables. Figure 3 shows some examples of original images (before and after sPCA), noisy observations and the recovered images by our algorithm. We can see that our algorithm produces accurate recovery of the groundtruth images from noisy samples. We split the measured error into two terms: the error caused by sPCA (the error between groundtruth images before and after sPCA) and estimation error in the sPCA space caused by the optimization problem (the error between recovered images and groundtruth images after sPCA). We refer to these errors as sPCA error and estimation error, respectively. For this experiment, the relative sPCA error compared to the groundtruth images is about 19.6%, while the relative estimation error compared to groundtruth images after sPCA is about 5.2%. Table I shows the CPU time of each step of the algorithm. In this experiment, and all the experiments on uniform distribution in the following, conjugate gradient method is used to solve the optimization problem.

We conducted experiments to study how the recovery error increases with the noise level. As before, we set $K = 10$, π is the (known) uniform distribution, and the number of observations is 10^4. Figure 4 shows the recovery results and relative errors for different SNRs. As can be seen, both the relative sPCA error and estimation error are, more or less, inversely proportional to the SNR. When the noise is larger, the sPCA gives less coefficients and results in larger sPCA errors. In addition, the estimation of the invariants, and thus the coefficients of groundtruth images, are less accurate under larger noise. Of course, when the noise is larger, we need more observations to average out the noise.

The next experiment aims to examine our algorithm when optimizing over $\hat{\pi}$ and the images simultaneously. As before, we fix $K = 10$ and SNR = 1/50. We take 500 observations for each class for the first 5 classes, and 1500 observations for the other 5 classes, so that

$$\pi_i = 0.05, \quad i = 1, \ldots, 5; \quad \pi_i = 0.15, \quad i = 6, \ldots, 10.$$

<div align="right">(IV.7)</div>

After applying our algorithm, the TV distance between $\hat{\pi}$ and π turns out to be 0.0086. The relative estimation error of all 10 images compared to groundtruth images after sPCA is 6.05%. The relative estimation error of the 5 images with $\pi_i = 0.05$ is 7.39%, while the

relative estimation error of the other 5 images with $\pi_i = 0.15$ is 3.94%. Empirically, a non-uniform distribution does not influence much the overall quality of the recovery, though it seems underrepresented images suffer more. Figure 5 shows some results of this experiment with nonuniform distribution. CPU time cost is shown in table I. For the experiments with non-uniform distribution, trust-regions algorithm is used to solve the optimization problem. Usually, trust-regions algorithm provide more accurate estimates than conjugate gradient at the cost of running times.

Next, we study the influence of π on the quality of the recovery. In this experiment, we take $K = 2$, SNR = 1/50, and $N = 10^4$. Table II shows the relative estimation errors on recovered images and the TV error on the recovered distribution as the distribution (π_1, π_2) shifts away from uniform. From the table, we can see that in all cases our estimated distributions are close to the true distribution with TV distance less than 0.01. When the number of observations for different images are not equal, the image with more observations tends to have lower estimation error.

Usually the projection images are not perfectly centered because the particle picking is not ideal. We conducted experiments to examine the robustness of our algorithm to small shifts of the input images, albeit our model does not take these shifts into account. This time, we consider a problem with only one class ($K = 1$) of size 129×129, 5×10^3 noisy observations and SNR = 1/50. A random shift is applied to each observation. The shift is generated by a 2-D uniform distribution on all the shifts within a circle of radius s. When s ranges from 0 to 5, the relative estimation errors of the recovered images are 3.30%, 4.86%, 6.58%, 10.28%, 11.64% and 18.30%, respectively. While the error increases with the size of the shifts, it does so at a reasonable pace: when the shifts are small, the recovery errors are too.

Comparison of our algorithm with the EM method was made. We implement a vanilla version of the EM algorithm, which is applied on sPCA coefficients rather than the images themselves, and considers only a finite set of in-plane rotations. Our EM is different from the EM-based algorithms implemented in cryo-EM software packages (such as RELION), which are more sophisticated and include many heuristics to improve running time and accuracy. Here we aim to underscore the resolution-computational load trade-off of EM In the first experiment, we take $K = 5$ classes and 10^3 observations per class with SNR = 1/50. Figure 6 compares the relative estimation error compared to groundtruth images after sPCA for each class and the running time with different number of in-plane rotations considered by EM. From the figure we can see that EM performs better when the number of rotations is large (32 in this experiments). However, at the same time EM becomes time-costing, taking nearly 10 times more CPU-time than our algorithm. In the right plot of Figure 6, the time costs of computing invariant features are shown, which are already included in the time cost of our algorithm (the red line). Figure 7 shows the results of another experiment for nonuniform distribution and larger noise. We take the distribution (IV.7) and SNR = 1/100, with $K = 10$ and $N = 10^4$. We can observe similar phenomena as the last experiment.

To demonstrate that our algorithm applies to other data sets as well, we considered two additional molecules: the transient receptor potential cation channel subfamily V member 1

(TrpV1) [45] and the yeast mitochondrial ribosome [46]. The volumes are downloaded from The Electronic Microscopy Data Bank[4]. Similarly to the experiments before, groundtruth images are projections randomly generated from the volumes. The size of the projections is 181×181 and we set $K = 10$, SNR = 1/50, and $N = 10^4$. Figure 8 shows some results of TrpV1 data. In this experiment, the distribution is taken to be uniform and the variable $\hat{\pi}$ is fixed. The relative estimation error compared to groundtruth images after sPCA is 4.29%. Figure 9 shows part of the results from the yeast mitochondrial ribosome data. In this experiment, we use the distribution given by (IV.7). The total variation error of $\hat{\pi}$ is 0.0141. The relative estimation error of all the images compared to groundtruth images after sPCA is 6.03%, and 3.83% and 7.96% for images with $\pi_i = 0.15$ and $\pi_i = 0.05$, respectively.

V.   Conclusion

In this paper, we studied the problem of heterogeneous MRA for 2-D images and proposed a new algorithmic framework for 2-D classification for SPR. Experimental results show that our algorithm can provide high-quality recovery of the groundtruth images (class averages), even when the noise level is high. The algorithm requires only one pass over the data and thus suits for large experimental data sets.

In practice, the projection images in cryo-EM suffer from small random shifts. Hence, a more accurate generative model reads

$$Y = T_s R_\xi I_\pi + \varepsilon,$$

where T_s is a small random shift by s; compare with (II.1). In future work we intend to extend our framework to take shifts into account. A recent study [26] shows that non-uniform distributions of translations makes MRA easier. We may take this issue into account in the future. Meanwhile, we have shown that our algorithm is robust against small shifts.

More importantly, our algorithm considers a discrete set of viewing directions. Yet, more realistically, cryo-EM micrographs contain projections sampled from a continuous distribution of viewing directions. We hope to extend our algorithm to the continuous case in the future. To apply the proposed techniques to experimental data, it is necessary to handle effects of the contrast transfer functions (CTF) and of colored noise as well.

In [20], [47], it was shown that the number of classes that can be demixed in 1-D hMRA is, approximately, $\sqrt{L}$, where L is the length of the signals. Our experiments indicate that we can demix $40 - 50$ classes. How this number depends on the size of the image or number of sPCA coefficients is left for future study.

Acknowledgment

---

[4] www.emdatabank.org

APPENDIX A GRADIENT OF THE OBJECTIVE FUNCTION

In this section we give the gradient of our objective function (III.13) in Euclidean space. For complex variables $\hat{a}$, we treat them as a matrix, and define the gradient $\partial F / \partial \hat{a}$ as the only matrix $g = g(\hat{a}, \hat{\pi})$ satisfying

$$\Re\{\mathrm{Tr}(g^* y)\} = DF(\hat{a}, \hat{\pi})[y],$$

(A.1)

where y is a matrix with the same size as $\hat{a}$, $\Re\{\mathrm{Tr}(g^* y)\}$ is an inner product, and $DF(\hat{a}, \hat{\pi})[y]$ is the directional derivative of F at $\hat{a}$ along y. As the objective function is a summation of least squares, the gradient of the objective function is the summation of the gradient of least squares terms. Hence, we only need to compute gradients for the following 3 groups of terms:

$$\left| \sum_{i=1}^{K} \hat{\pi}_i \hat{a}_{0,q}^i - \hat{m}_{0,q} \right|^2,$$

(A.2)

$$\left| \sum_{i=1}^{K} \hat{\pi}_i \hat{a}_{k,q_1}^i \overline{\hat{a}_{k,q_2}^i} - \hat{p}_{k,q_1,q_2} \right|^2,$$

(A.3)

$$\left| \sum_{i=1}^{K} \hat{\pi}_i \hat{a}_{k_1,q_2}^i \hat{a}_{k_2,q_2}^i \overline{\hat{a}_{k_1+k_2,q_3}^i} - \hat{b}_{k_1,k_2,q_1,q_2,q_3} \right|^2.$$

(A.4)

We call (A.2), (A.3) and (A.4) the first-, second- and third-order terms according to the order of moments they contain.

A.   First-order terms

Let

$$\mathcal{M}_q(\hat{a}, \hat{\pi}) = \sum_{i=1}^{K} \hat{\pi}_i \hat{a}_{0,q}^i - \hat{m}_{0,q}.$$

Then, for real variables $\hat{\pi}_i$ we can easily get

$$\frac{\partial \left|\mathcal{M}_q(\hat{a}, \hat{\pi})\right|^2}{\partial \hat{\pi}_i} = 2\Re\left\{\mathcal{M}_q(\hat{a}, \hat{\pi})\hat{a}_{0,q}^i\right\}.$$

(A.5)

For complex variables $\hat{a}_{0,q}^i$, we have

$$\frac{\partial \left|\mathcal{M}_q(\hat{a}, \hat{\pi})\right|^2}{\partial \hat{a}_{0,q}^i} = 2\mathcal{M}_q(\hat{a}, \hat{\pi})\hat{\pi}_i.$$

(A.6)

For $\hat{a}_{k',q'}^i$ with $k' \neq 0$ or $q' \neq q$ we always have

$$\frac{\partial \left|\mathcal{M}_q(\hat{a}, \hat{\pi})\right|^2}{\partial \hat{a}_{k',q'}^i} = 0,$$

(A.7)

as they do not appear in the term. In the following subsections, we ignore the gradient with respect to such variables.

B. Second-order terms

Let

$$\mathcal{P}(\hat{a}, \hat{\pi}) := \mathcal{P}_{k,q_1,q_2}(\hat{a}, \hat{\pi}) = \sum_{i=1}^{K} \hat{\pi}_i \hat{a}_{k,q_1}^i \overline{\hat{a}_{k,q_2}^i} - \hat{p}_{k,q_1,q_2}.$$

Then, for $\hat{\pi}_i$, similar to the last subsection, we have

$$\frac{\partial \left|\mathcal{P}(\hat{a}, \hat{\pi})\right|^2}{\partial \hat{\pi}_i} = 2\Re\left\{\mathcal{P}(\hat{a}, \hat{\pi})\hat{a}_{k,q_1}^i \overline{\hat{a}_{k,q_2}^i}\right\}.$$

(A.8)

For $\hat{a}_{k,q}^i$, there are two cases. If $q_1 = q_2$, then

$$\frac{\partial \left|\mathcal{P}(\hat{a}, \hat{\pi})\right|^2}{\partial \hat{a}_{k,q_1}^i} = 4\Re\left\{\mathcal{P}(\hat{a}, \hat{\pi})\right\}\hat{a}_{k,q_1}^i \hat{\pi}_i.$$

(A.9)

If $q_1 \neq q_2$ then we have

$$\frac{\partial \left|\mathcal{P}(\hat{a}, \hat{\pi})\right|^2}{\partial \hat{a}_{k,q_1}^i} = 2\mathcal{P}(\hat{a}, \hat{\pi})\hat{a}_{k,q_2}^i \hat{\pi}_i,$$

<div align="right">(A.10)</div>

and

$$\frac{\partial \left| \mathscr{P}(\hat{a}, \hat{\pi}) \right|^2}{\partial \hat{a}_{k, q_2}^i} = 2 \overline{\mathscr{P}(\hat{a}, \hat{\pi})} \hat{a}_{k, q_1}^i \hat{\pi}_i,$$

<div align="right">(A.11)</div>

C. Third-order terms

Let

$$\begin{aligned} \mathscr{B}(\hat{a}, \hat{\pi}) \quad &:= \mathscr{B}_{k_1, k_2, q_1, q_2, q_3}(\hat{a}, \hat{\pi}) \\ &= \sum_{i=1}^{K} \hat{\pi}_i \hat{a}_{k_1, q_2}^i \hat{a}_{k_2, q_2}^i \overline{\hat{a}_{k_1+k_2, q_3}^i} - \hat{b}_{k_1, k_2, q_1, q_2, q_3}. \end{aligned}$$

Then, firstly we have

$$\frac{\partial \left| \mathscr{B}(\hat{a}, \hat{\pi}) \right|^2}{\partial \hat{\pi}_i} = 2 \Re \left\{ \mathscr{B}(\hat{a}, \hat{\pi}) \hat{a}_{k_1, q_2}^i \hat{a}_{k_2, q_2}^i \overline{\hat{a}_{k_1+k_2, q_3}^i} \right\}.$$

<div align="right">(A.12)</div>

Next, again we consider two cases. If $k_1 = k_2$ and $q_1 = q_2$, then

$$\frac{\partial \left| \mathscr{B}(\hat{a}, \hat{\pi}) \right|^2}{\partial \hat{a}_{k_1, q_1}^i} = 4 \mathscr{B}(\hat{a}, \hat{\pi}) \overline{\hat{a}_{k_1, q_1}^i} \hat{a}_{k_1+k_2, q_3}^i \hat{\pi}_i,$$

<div align="right">(A.13)</div>

$$\frac{\partial \left| \mathscr{B}(\hat{a}, \hat{\pi}) \right|^2}{\partial \hat{a}_{k_1+k_2, q_3}^i} = 2 \overline{\mathscr{B}(\hat{a}, \hat{\pi})} \left(\hat{a}_{k_1, q_1}^i \right)^2 \hat{\pi}_i.$$

<div align="right">(A.14)</div>

Otherwise, we have

$$\frac{\partial \left| \mathscr{B}(\hat{a}, \hat{\pi}) \right|^2}{\partial \hat{a}_{k_1, q_1}^i} = 2 \mathscr{B}(\hat{a}, \hat{\pi}) \overline{\hat{a}_{k_2, q_2}^i} \hat{a}_{k_1+k_2, q_3}^i \hat{\pi}_i,$$

<div align="right">(A.15)</div>

$$\frac{\partial \left| \mathscr{B}(\hat{a}, \hat{\pi}) \right|^2}{\partial \hat{a}_{k_2, q_2}^i} = 2 \mathscr{B}(\hat{a}, \hat{\pi}) \overline{\hat{a}_{k_1, q_1}^i} \hat{a}_{k_1+k_2, q_3}^i \hat{\pi}_i,$$

<div align="right">(A.16)</div>

$$\frac{\partial \left| \mathscr{B}(\hat{a}, \hat{\pi}) \right|^2}{\partial \hat{a}^i_{k_1 + k_2, q_3}} = 2\overline{\mathscr{B}(\hat{a}, \hat{\pi})} \hat{a}^i_{k_1, q_1} \hat{a}^i_{k_2, q_2} \hat{\pi}_i.$$

(A.17)

References

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

[2]. Kühlbrandt W, "The resolution revolution," Science, vol. 343, no. 6178, pp. 1443–1444, 2014. [PubMed: 24675944]

[3]. Singer A. and Shkolnisky Y, "Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming," SIAM journal on imaging sciences, vol. 4, no. 2, pp. 543–572, 2011. [PubMed: 22536457]

[4]. Sieben C, Douglass KM, Gonczy P, and Manley S, "Multicolor single-particle reconstruction of protein complexes," Nature Methods, vol. 15, pp. 777–780, 2018. [PubMed: 30275574]

[5]. Van Heel M, "Angular reconstitution: a posteriori assignment of projection directions for 3d reconstruction," Ultramicroscopy, vol. 21, no. 2, pp. 111–123, 1987. [PubMed: 12425301]

[6]. Goncharov A, Vainshtein B, Ryskin A, and Vagin A, "Three-dimensional reconstruction of arbitrarily oriented identical particles from their electron photomicrographs," Sov. Phys. Crystallogr, vol. 32, pp. 504–509, 1987.

[7]. Dube P, Tavares P, Lurz R, and Van Heel M, "The portal protein of bacteriophage spp1: a dna pump with 13-fold symmetry.," The EMBO journal, vol. 12, no. 4, pp. 1303–1309, 1993. [PubMed: 8467790]

[8]. Singer A, "Mathematics for cryo-electron microscopy," Proceedings of the International Congress of Mathematicians, 2018.

[9]. Bendory T, Bartesaghi A, and Singer A, "Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities," arXiv preprint arXiv:1908.00574, 2019.

[10]. Penczek P, Radermacher M, and Frank J, "Three-dimensional reconstruction of single particles embedded in ice," Ultramicroscopy, vol. 40, no. 1, pp. 33–53, 1992. [PubMed: 1580010]

[11]. Sigworth F, "A maximum-likelihood approach to single-particle image refinement," Journal of structural biology, vol. 122, no. 3, pp. 328–339, 1998. [PubMed: 9774537]

[12]. Scheres SH, "RELION: implementation of a bayesian approach to cryo-EM structure determination," Journal of structural biology, vol. 180, no. 3, pp. 519–530, 2012. [PubMed: 23000701]

[13]. Scheres SH, Valle M, Nuñez R, Sorzano CO, Marabini R, Herman GT, and Carazo J-M, "Maximum-likelihood multi-reference refinement for electron microscopy images," Journal of molecular biology, vol. 348, no. 1, pp. 139–149, 2005. [PubMed: 15808859]

[14]. van Heel M, Harauz G, Orlova EV, Schmidt R, and Schatz M, "A new generation of the IMAGIC image processing system," Journal of structural biology, vol. 116, no. 1, pp. 17–24, 1996. [PubMed: 8742718]

[15]. Sorzano C, Bilbao-Castro J, Shkolnisky Y, Alcorlo M, Melero R, Caffarena-Fernandez G, Li M, Xu G, Marabini R, and Carazo J, "Á clustering approach to multireference alignment of single-particle projections in electron microscopy," Journal of structural biology, vol. 171, no. 2, pp. 197–206, 2010. [PubMed: 20362059]

[16]. Yang Z, Fang J, Chittuluru J, Asturias FJ, and Penczek PA, "Iterative stable alignment and clustering of 2D transmission electron microscope images," Structure, vol. 20, no. 2, pp. 237–247, 2012. [PubMed: 22325773]

[17]. Schatz M. and Van Heel M, "Invariant classification of molecular views in electron micrographs," Ultramicroscopy, vol. 32, no. 3, pp. 255–264, 1990. [PubMed: 2382314]

[18]. Zhao Z. and Singer A, "Rotationally invariant image representation for viewing direction classification in cryo-EM," Journal of structural biology, vol. 186, no. 1, pp. 153–166, 2014. [PubMed: 24631969]

[19]. Perry A, Weed J, Bandeira AS, Rigollet P, and Singer A, "The sample complexity of multireference alignment," SIAM Journal on Mathematics of Data Science, vol. 1, no. 3, pp. 497–517, 2019.

[20]. Boumal N, Bendory T, Lederman RR, and Singer A, "Heterogeneous multireference alignment: A single pass approach," in Information Sciences and Systems (CISS), 2018 52nd Annual Conference on, pp. 1–6, IEEE, 2018.

[21]. Aguerrebere C, Delbracio M, Bartesaghi A, and Sapiro G, "Fundamental limits in multi-image alignment," IEEE Transactions on Signal Processing, vol. 64, no. 21, pp. 5707–5722, 2016.

[22]. Bandeira AS, Charikar M, Singer A, and Zhu A, "Multireference alignment using semidefinite programming," in Proceedings of the 5th conference on Innovations in theoretical computer science, pp. 459–470, ACM, 2014.

[23]. Bendory T, Boumal N, Ma C, Zhao Z, and Singer A, "Bispectrum inversion with application to multireference alignment," IEEE Transactions on Signal Processing, vol. 66, no. 4, pp. 1037–1050, 2017. [PubMed: 29805244]

[24]. Bandeira A, Rigollet P, and Weed J, "Optimal rates of estimation for multi-reference alignment," arXiv preprint arXiv:1702.08546, 2017.

[25]. Abbe E, Pereira JM, and Singer A, "Sample complexity of the Boolean multireference alignment problem," in Proceedings. IEEE International Symposium on Information Theory, vol. 2017, p. 1316, NIH Public Access, 2017.

[26]. Abbe E, Bendory T, Leeb W, Pereira JM, Sharon N, and Singer A, "Multireference alignment is easier with an aperiodic translation distribution," IEEE Transactions on Information Theory, vol. 65, no. 6, pp. 3565–3584, 2018.

[27]. Bandeira AS, Blum-Smith B, Kileel J, Perry A, Weed J, and Wein AS, "Estimation under group actions: recovering orbits from invariants," arXiv preprint arXiv:1712.10163, 2017.

[28]. Abbe E, Pereira JM, and Singer A, "Estimation in the group action channel," in 2018 IEEE International Symposium on Information Theory (ISIT), pp. 561–565, IEEE, 2018.

[29]. Tukey J, "The spectral representation and transformation properties of the higher moments of stationary time series," Reprinted in The Collected Works of John W. Tukey, vol. 1, pp. 165–184, 1953.

[30]. Giannakis GB, "Signal reconstruction from multiple correlations: frequency-and time-domain approaches," JOSA A, vol. 6, no. 5, pp. 682–697, 1989.

[31]. Sadler BM and Giannakis GB, "Shift-and rotation-invariant object reconstruction using the bispectrum," JOSA A, vol. 9, no. 1, pp. 57–69, 1992.

[32]. Kam Z, "The reconstruction of structure from electron micrographs of randomly oriented particles," in Electron Microscopy at Molecular Dimensions, pp. 270–277, Springer, 1980.

[33]. Landa B. and Shkolnisky Y, "Steerable principal components for space-frequency localized images," SIAM journal on imaging sciences, vol. 10, no. 2, pp. 508–534, 2017. [PubMed: 29081879]

[34]. Zhao Z, Shkolnisky Y, and Singer A, "Fast steerable principal component analysis," IEEE transactions on computational imaging, vol. 2, no. 1, pp. 1–12, 2016. [PubMed: 27570801]

[35]. Park W, Midgett CR, Madden DR, and Chirikjian GS, "A stochastic kinematic model of class averaging in single-particle electron microscopy," The International journal of robotics research, vol. 30, no. 6, pp. 730–754, 2011. [PubMed: 21660125]

[36]. Landa B. and Shkolnisky Y, "Approximation scheme for essentially bandlimited and space-concentrated functions on a disk," Applied and Computational Harmonic Analysis, vol. 43, no. 3, pp. 381–403, 2017.

[37]. Lederman RR, "Numerical algorithms for the computation of generalized prolate spheroidal functions," arXiv preprint arXiv:1710.02874, 2017.

[38]. Zhao Z. and Singer A, "Fourier–Bessel rotational invariant eigenimages," JOSA A, vol. 30, no. 5, pp. 871–877, 2013. [PubMed: 23695317]

[39]. Bhamre T, Zhang T, and Singer A, "Denoising and covariance estimation of single particle cryo-em images," Journal of structural biology, vol. 195, no. 1, pp. 72–81, 2016. [PubMed: 27129418]

[40]. Levin E, Bendory T, Boumal N, Kileel J, and Singer A, "3D ab initio modeling in cryo-em by autocorrelation analysis," in Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, pp. 1569–1573, IEEE, 2018.

[41]. Bendory T, Boumal N, Leeb W, Levin E, and Singer A, "Multitarget detection with application to cryo-electron microscopy," Inverse Problems, 2019.

[42]. Boumal N, Mishra B, Absil P-A, and Sepulchre R, "Manopt, a Matlab toolbox for optimization on manifolds," The Journal of Machine Learning Research, vol. 15, no. 1, pp. 1455–1459, 2014.

[43]. Sun Y, Gao J, Hong X, Mishra B, and Yin B, "Heterogeneous tensor decomposition for clustering via manifold optimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 3, pp. 476–489, 2016. [PubMed: 27046492]

[44]. Shaikh TR, Gao H, Baxter WT, Asturias FJ, Boisset N, Leith A, and Frank J, "SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs," Nature protocols, vol. 3, no. 12, p. 1941, 2008. [PubMed: 19180078]

[45]. Gao Y, Cao E, Julius D, and Cheng Y, "Trpv1 structures in nanodiscs reveal mechanisms of ligand and lipid action," Nature, vol. 534, no. 7607, p. 347, 2016. [PubMed: 27281200]

[46]. Desai N, Brown A, Amunts A, and Ramakrishnan V, "The structure of the yeast mitochondrial ribosome," Science, vol. 355, no. 6324, pp. 528–531, 2017. [PubMed: 28154081]

[47]. Wein A, Statistical Estimation in the Presence of Group Actions. PhD thesis, Massachusetts Institute of Technology, 2018.
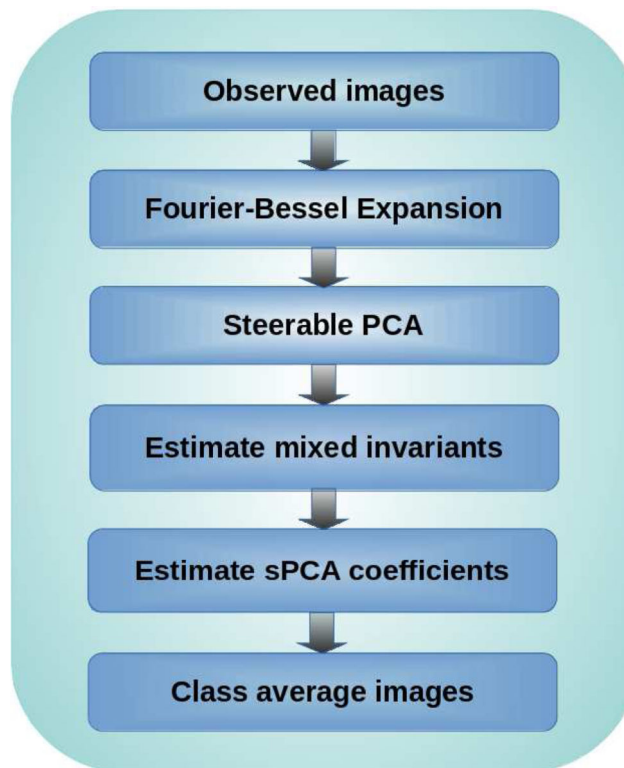
Fig. 1.
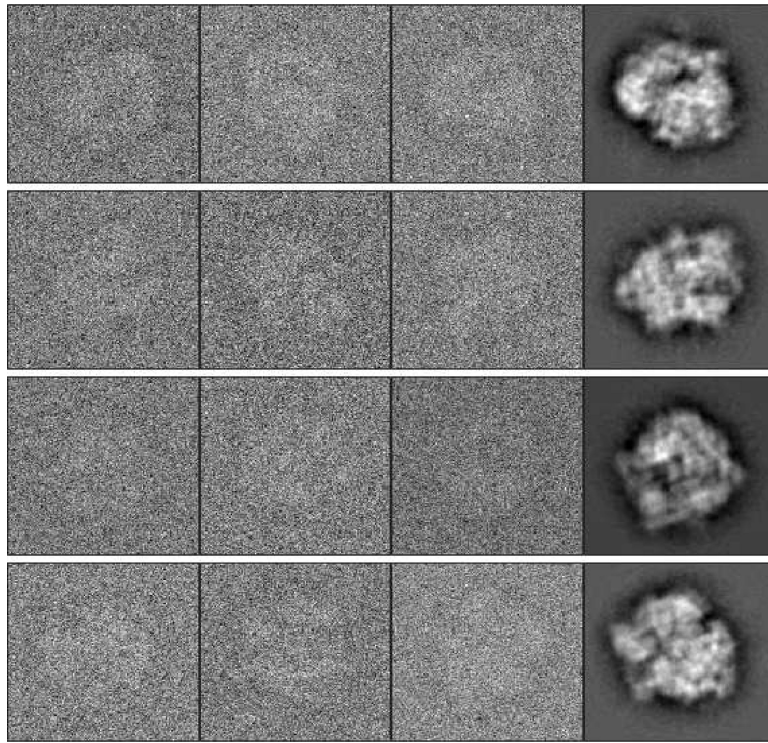Flow chart of Algorithm 1 for 2-D classification using rotationally invariant features.

Fig. 2.
Examples of class averages (the right column) and rotated noisy images (the first three columns) for the E. coli 70S ribosome volume. SNR = 1/50.
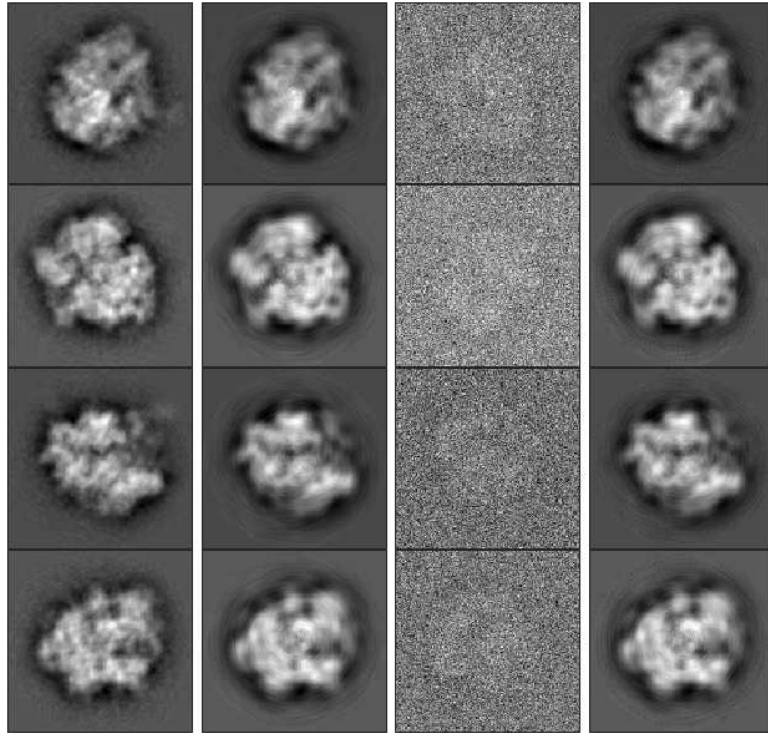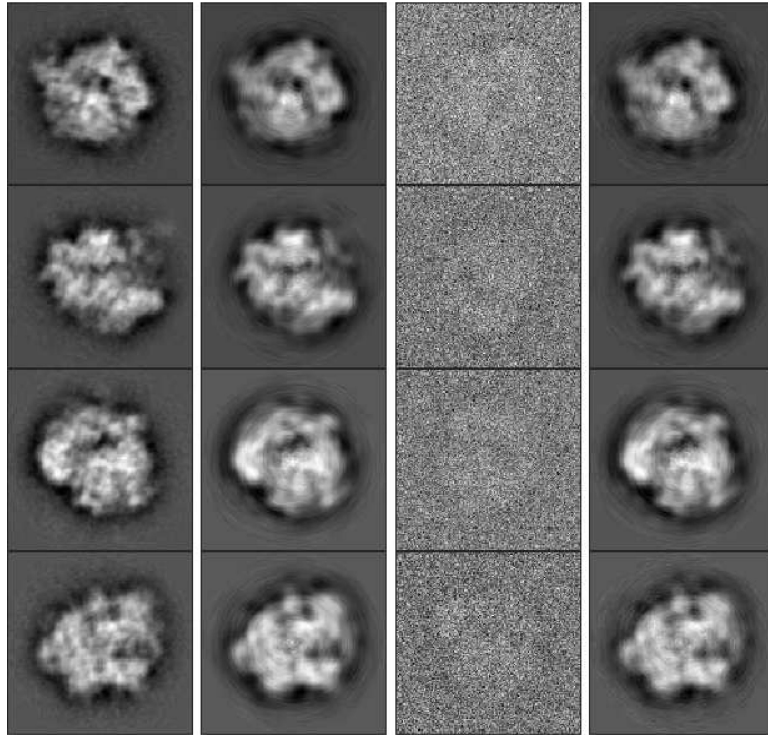
Fig. 3.
First column (left to right): Groundtruth images before sPCA. **Second column:** Groundtruth images after sPCA. **Third column:** examples of noisy observations. **Fourth column:** recovered images by our algorithm, rotated and permuted to align with groundtruth images.
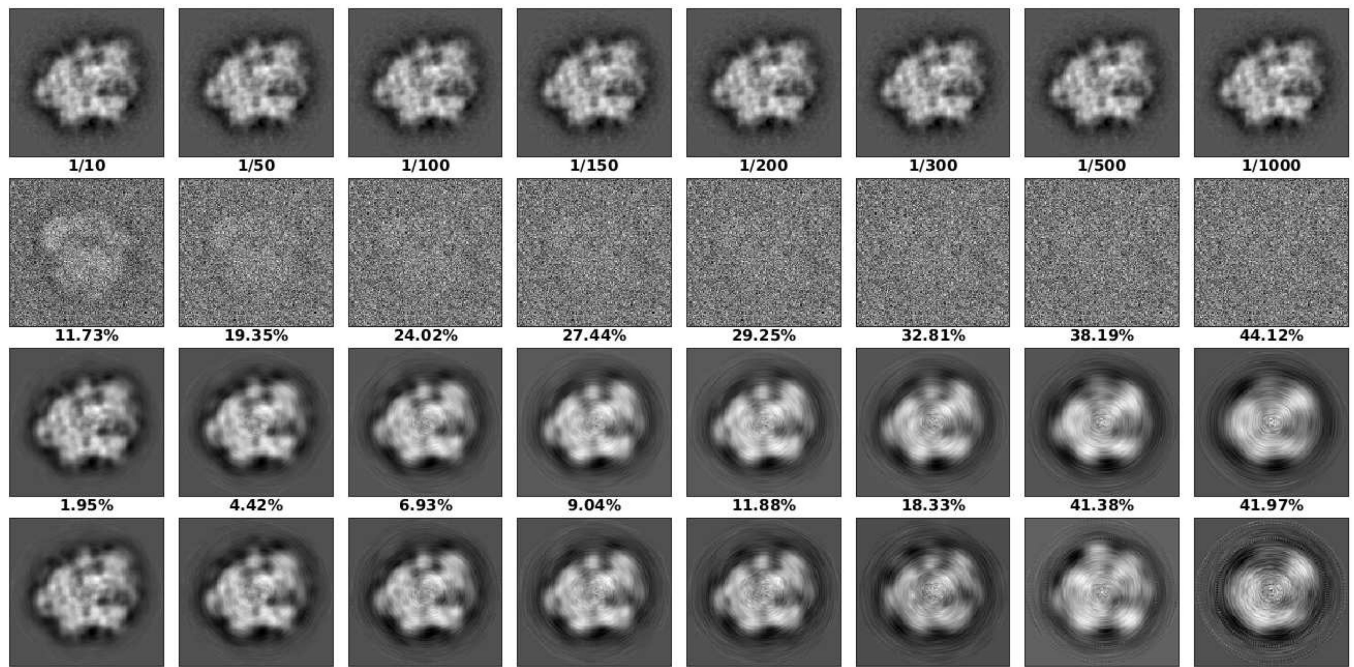
Fig. 4.
Recovery results for different SNR. The first to fourth rows of images are groundtruth images, noisy images, groundtruth images after sPCA, and recovered images, respectively. The first to third rows of numbers are SNR, relative sPCA error compared to groundtruth images (the first row), and relative estimation error compared to groundtruth images after sPCA (the third row), respectively. From left to right, the numbers of coefficients chosen by sPCA are 172, 88, 60, 45, 41, 33, 20 and 13.

Fig. 5.
First column: Groundtruth images before sPCA. **Second column:** Groundtruth images after sPCA. **Third column:** examples of noisy observations. **Fourth column:** recovered images by our algorithm, rotated to align with groundtruth images. **The rows:** The first two rows are images with $\pi = 0.15$ and the last two rows are images with $\pi_i = 0.05$.
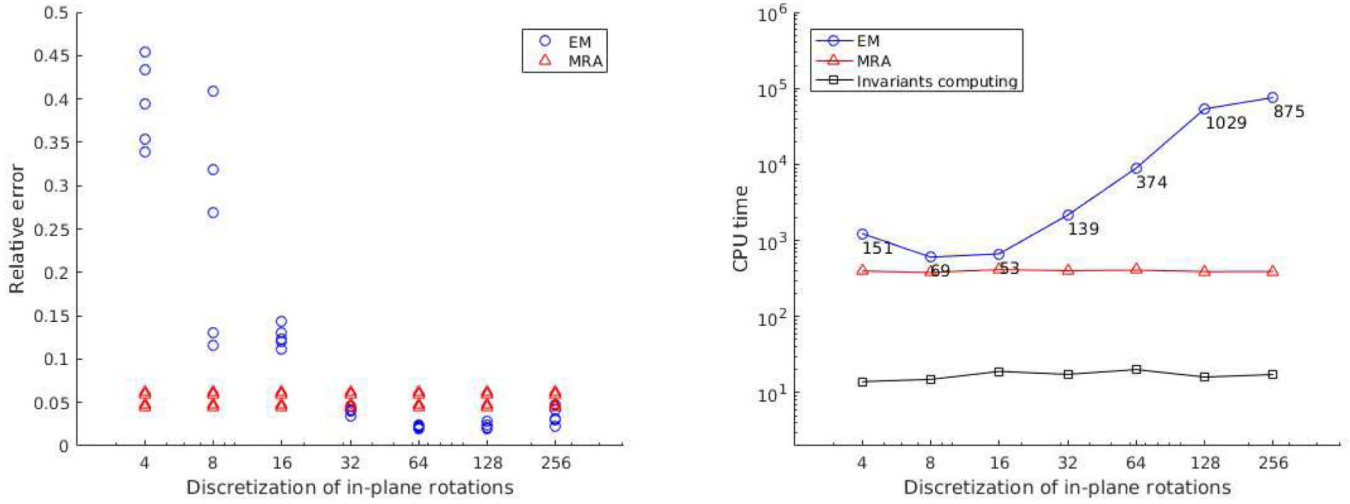
Fig. 6.

Comparison of our algorithm (MRA) with EM in terms of relative estimation error and computation time. Images are generated from 5 classes with uniformly random in-plane rotations. Contrary to MRA, EM needs to assume the rotations are selected from a discrete set. Here, we see the accuracy/computation time trade-off of EM for SNR = 1/50. Left panel: for each discretization value (i.e., number of uniformly sampled angles) and algorithm, each point represents the relative estimation error compared to groundtruth images after sPCA for one of the classes. Right panel: integers indicate the number of EM iterations. The CPU time is measured in seconds.
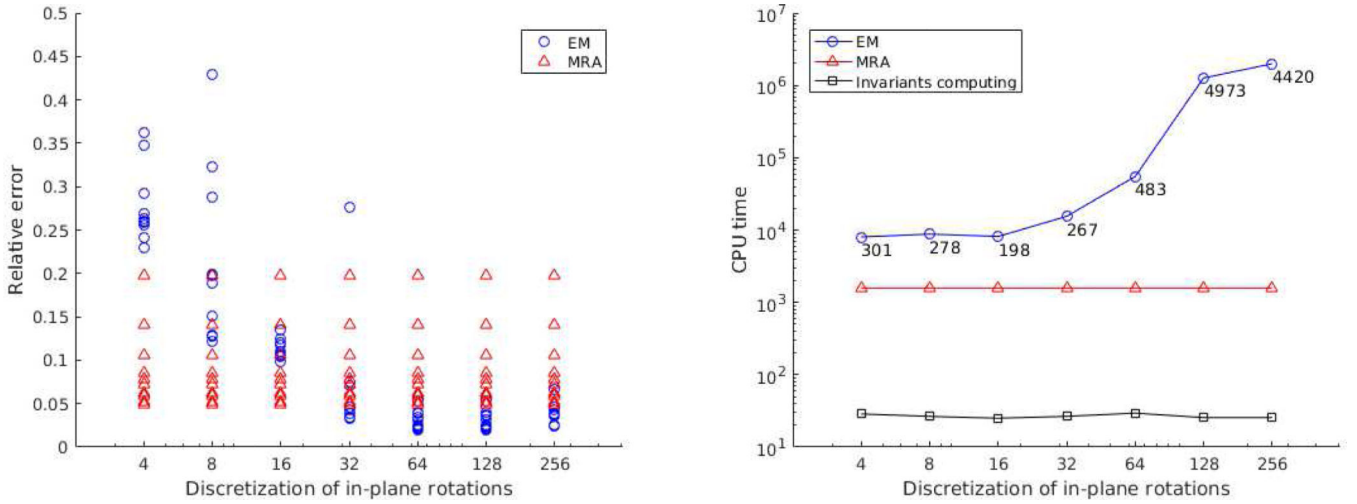
Fig. 7.

Comparison of our algorithm (MRA) with EM in terms of relative estimation error and computation time. The 10 classes are distributed nonuniformly according to (IV.7) and SNR = 1/100. Left panel: for each discretization value and algorithm, each point represents the relative estimation error compared to groundtruth images after sPCA for one of the classes. Right panel: integers indicate the number of EM iterations. The CPU time is measured in seconds.
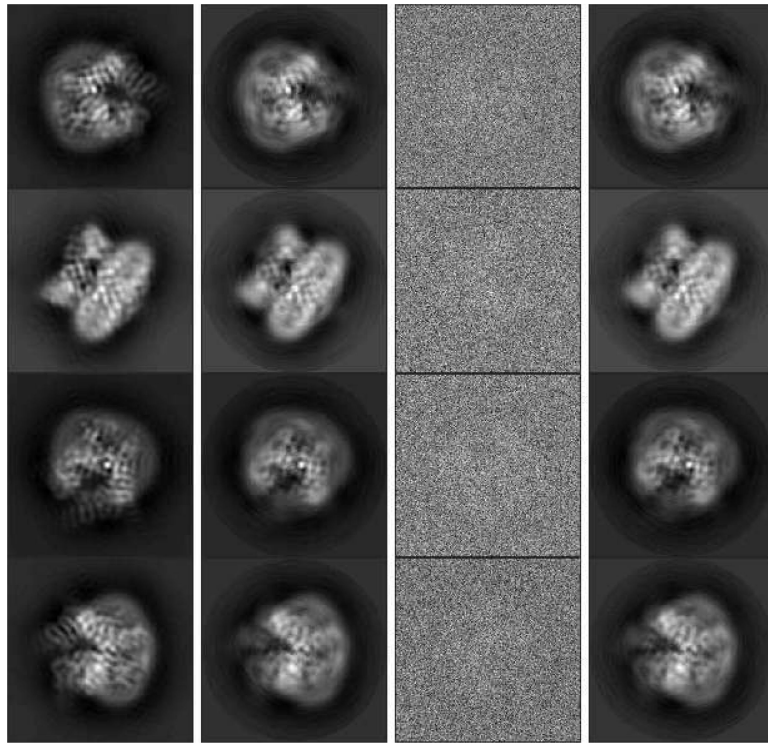
Fig. 8.
Recovery results of the TrpV1 data. **First column:** Groundtruth images before sPCA. **Second column:** Groundtruth images after sPCA. **Third column:** examples of noisy observations. **Fourth column:** recovered images by our algorithm, rotated to align with groundtruth images.
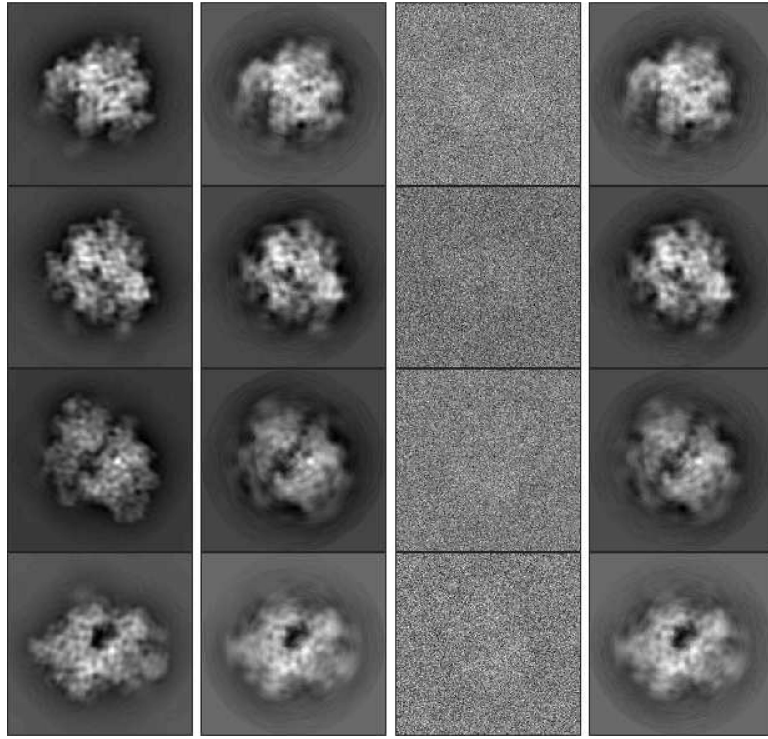
Fig. 9.
Recovery results of the yeast mitochondrial ribosome data. **First column:** Groundtruth images before sPCA. **Second column:** Groundtruth images after sPCA. **Third column:** examples of noisy observations. **Fourth column:** recovered images by our algorithm, rotated to align with groundtruth images. **The rows:** The first two rows are images with $\pi = 0.15$. The last two rows are images with $\pi_i = 0.05$.

TABLE I

CPU TIME COST (IN SECONDS) OF DIFFERENT STEPS FOR EXPERIMENTS ON E. COLI 70S RIBOSOME WITH UNIFORM AND NON-UNIFORM DISTRIBUTION (CORRESPONDING TO Figure 3 AND 5).

Step \ Distribution	Uniform	Non-uniform
Computing sPCA	252.2s	259.7s
Computing mixed invariants	16.3s	17.5s
Optimization	428.2s	7157.3s

TABLE II

RELATIVE ESTIMATION ERROR COMPARED TO GROUNDTRUTH IMAGES AFTER sPCA AND TOTAL VARIATION ERROR OF EXPERIMENTAL RESULTS WITH VARYING DISTRIBUTION $\pi = (\pi_1, \pi_2)$; error_i REPRESENTS THE RELATIVE ERROR ON IMAGE $i = 1, 2$.

π_1	0.1	0.2	0.3	0.4	0.5
dist_{TV}	0.0025	0.0007	0.0022	0.0009	0.0014
error_1	3.59%	5.01%	1.59%	3.60%	3.95%
error_2	2.01%	4.66%	0.91%	4.08%	4.63%