Steerable *e*PCA: Rotationally Invariant Exponential Family PCA

Zhizhen Zhao, *Member, IEEE*, Lydia T. Liu, and Amit Singer

***Abstract*— In photon-limited imaging, the pixel intensities are affected by photon count noise. Many applications require an accurate estimation of the covariance of the underlying 2-D clean images. For example, in X-ray free electron laser (XFEL) single molecule imaging, the covariance matrix of 2-D diffraction images is used to reconstruct the 3-D molecular structure. Accurate estimation of the covariance from low-photon-count images must take into account that pixel intensities are Poisson distributed, hence the classical sample covariance estimator is highly biased. Moreover, in single molecule imaging, including in-plane rotated copies of all images could further improve the accuracy of covariance estimation. In this paper we introduce an efficient and accurate algorithm for covariance matrix estimation of count noise 2-D images, including their uniform planar rotations and possibly reflections. Our procedure, *steerable ePCA*, combines in a novel way two recently introduced innovations. The first is a methodology for principal component analysis (PCA) for Poisson distributions, and more generally, exponential family distributions, called *e*PCA. The second is steerable PCA, a fast and accurate procedure for including all planar rotations when performing PCA. The resulting principal components are invariant to the rotation and reflection of the input images. We demonstrate the efficiency and accuracy of steerable *e*PCA in numerical experiments involving simulated XFEL datasets and rotated face images from Yale Face Database B.**

***Index Terms*— Poisson noise, X-ray free electron laser, steerable PCA, eigenvalue shrinkage, autocorrelation analysis, image denoising.**

I. Introduction

X-RAY free electron laser (XFEL) is an emerging imaging technique for elucidating the three-dimensional structure of molecules [1], [2]. Single molecule XFEL imaging collects two-dimensional diffraction patterns of single particles at random orientations. The images are very noisy due to the low photon count. The detector count-noise follows an approximately Poisson distribution. Since only one diffraction pattern is captured per particle and the particle orientations are unknown, it is challenging to reconstruct the 3-D structure at low signal-to-noise ratio (SNR). One approach is to use expectation-maximization (EM) [3], [4], but it has a high computational cost at low signal-to-noise ratio (SNR). Alternatively, assuming that the orientations of the particles are uniformly distributed over the special orthogonal group $SO(3)$, Kam's correlation analysis [5]–[12] bypasses orientation estimation and requires just one pass over the data, thus alleviating the computational cost. Kam's method requires a robust estimation of the covariance matrix of the noiseless 2-D images. This serves as our main motivation for developing efficient and accurate covariance estimation and denoising methods for Poisson data. Nevertheless, the methods presented here are quite general and can be applied to other imaging modalities involving Poisson distributions.

Principal Component Analysis (PCA) is widely used for dimension reduction and denoising of large datasets [13], [14]. However, it is most naturally designed for Gaussian data, and there is no consensus on the extension to non-Gaussian settings such as exponential family distributions [13, Sec. 14.4]. For denoising with non-Gaussian noise, popular approaches reduce it to the Gaussian case by a wavelet transform such as the Haar transform [15]; by adaptive wavelet shrinkage [15], [16]; or by approximate variance stabilization such as the Anscombe transform [17]. The latter is known to work well for Poisson signals with large parameters, due to approximate normality. However, the normal approximation breaks down for Poisson distributions with a small parameter, in the case of photon-limited XFEL [18, Sec. 6.6]. Other methods are based on alternating minimization [19]–[21], singular value thresholding (SVT) [22], [23] and Bayesian techniques [24]. Many of aforementioned methods such as [19] are computationally intractable for large datasets and do not have statistical guarantees for covariance estimation. In particular, [21] applied the methodology of alternating minimization [19] to denoise a single image by performing PCA on clusters of patches extracted from the image. However, this approach is not suitable for our problem setting, where the goal is to simultaneously denoise a large number of images and estimate their covariance. Recently, [25] introduced exponential family PCA (*e*PCA), which extends PCA to a wider class of distributions. It involves the eigendecomposition of a new covariance matrix

Manuscript received December 12, 2018; revised July 13, 2019, December 4, 2019, and February 25, 2020; accepted March 29, 2020. Date of publication April 27, 2020; date of current version May 1, 2020. The work of Zhizhen Zhao was supported in part by the National Center for Supercomputing Applications Faculty Fellowship, University of Illinois at Urbana–Champaign College of Engineering Strategic Research Initiative and in part by NSF under Grant DMS-1854791. The work of Amit Singer was supported in part by the NIGMS, under Award R01GM090200, in part by AFOSR under Grant FA9550-17-1-0291, in part by the Simons Investigator Award, in part by the Moore Foundation Data-Driven Discovery Investigator Award, and in part by the NSF BIGDATA Award under Grarnt IIS-1837992. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Giacomo Boracchi. *(Corresponding author: Zhizhen Zhao.)*

Zhizhen Zhao is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Urbana, IL 61820 USA (e-mail: zhizhenz@illinois.edu).

Lydia T. Liu is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: lydiatliu@berkeley.edu).

Amit Singer is with the Department of Mathematics, and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544 USA (e-mail: amits@math.princeton.edu).

Digital Object Identifier 10.1109/TIP.2020.2988139

estimator, constructed in a deterministic and non-iterative way using moment calculations and shrinkage. *e*PCA was shown to be more accurate than PCA and its alternatives for exponential families. It has computational cost similar to that of PCA, substantial theoretical justification building on random matrix theory, and interpretable output. We refer readers to [25] for experiments that benchmark *e*PCA against previous methods.

In XFEL imaging, the orientations of the particles are uniformly distributed over $SO(3)$, so it is equally likely to observe any planar rotation of the given 2-D diffraction pattern. Therefore, it makes sense to include all possible in-plane rotations of the images when performing *e*PCA. To this end, we incorporate steerability in *e*PCA, by adapting the steerable PCA algorithm, which avoids duplicating rotated images [26]. The concept of a steerable filter was introduced in [27] and various methods were proposed for computing data adaptive steerable filters [28]–[31]. We take into account the action of the group $O(2)$ on diffraction patterns by in-plane rotation (and reflection if needed). The resulting principal components are invariant to any $O(2)$ transformation of the input images.

The new algorithm, to which we refer as steerable *e*PCA, combines *e*PCA and steerable PCA in a natural way. Steerable *e*PCA is not an iterative optimization procedure. All steps involve only basic linear algebra operations. The various steps include expansion in a steerable basis, eigen-decomposition, eigenvalue shrinkage, and different normalization steps. The mathematical and statistical rationale for all steps is provided in Section II.

We illustrate the improvement in covariance matrix estimation by applying our method to image denoising in Section III. Specifically, we introduce a Wiener-type filtering using the principal components. Rotation invariance enhances the effectiveness of *e*PCA in covariance estimation and thus achieves better denoising. In addition, the denoised expansion coefficients are useful in building rotationally invariant image features (i.e. bispectrum-like features [32]). Numerical experiments are performed on simulated XFEL diffraction patterns and a natural image dataset—Yale Face Database B [33], [34]. As is the case for standard PCA, the computational complexity of the steerable *e*PCA is lower than that of *e*PCA.

An implementation of steerable *e*PCA in MATLAB is publicly available at github.com/zhizhenz/sepca/.

Algorithm 1 Steerable *e*PCA (s*e*PCA) and Denoising

Input: Image data Y that contains n images of size $L \times L$

Output: Rotationally invariant covariance estimator of noiseless images and denoised images

Compute the sample mean $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ (*e*PCA)

Estimate the support size R and band limit c for the mean image (sPCA)

Compute the Fourier-Bessel expansion coefficients of $F(\bar{Y})$ and estimate the rotationally invariant sample mean $\bar{f}$ as in Eq. (14) (s*e*PCA)

Compute the variance estimate $D_n = \mathrm{diag}[\bar{f}]$ (s*e*PCA)

Prewhiten the image data $Z = D_n^{-1/2} Y$ (*e*PCA)

Estimate the band limit c for whitened images (sPCA)

Compute the truncated Fourier-Bessel expansion coefficients of $F(Z)$ and form the coefficients matrices $A^{(k)}$, for $k = 0, \ldots, k_{\max}$ (sPCA)

for $k = 0, 1, \ldots, k_{\max}$ **do**

  Compute the prewhitened sample covariance matrix $S_h^{(k)}$ as in Eq. (16) and its eigendecomposition $S_h^{(k)} = \hat{U}\Lambda\hat{U}^*$ (*e*PCA)

  Shrink the eigenvalues $S_{h,\eta_{\gamma_k}}^{(k)} = \hat{U}\eta_{\gamma_k}(\Lambda)\hat{U}^*$ of top r_k eigenvalues according to Eq. (19) (*e*PCA, sPCA)

  Compute the recoloring matrix $B^{(k)}$ in eq. (23) and $D^{(k)}$ in eq. (26) (s*e*PCA)

  Recolor the covariance matrix $S_{he}^{(k)} = \left(B^{(k)}\right)^* \cdot S_{h,\gamma_k}^{(k)} \cdot B^{(k)}$ (s*e*PCA)

  Compute the scaling coefficients $\hat{\alpha}$ in Eq. (27) and keep components with $\hat{\alpha} > 0$ (s*e*PCA)

  Scale the covariance matrix $S_s^{(k)} = \sum \hat{\alpha}_i^{(k)} \hat{v}_i^{(k)} \left(\hat{v}_i^{(k)}\right)^*$, where the eigendecomposition of $S_{he}^{(k)}$ is $\sum \hat{v}_i^{(k)} \left(\hat{v}_i^{(k)}\right)^*$ (*e*PCA)

  Denoise $\{A^{(k)}\}_{k=0}^{k_{\max}}$ as in Eqs. (29) and (30) (s*e*PCA)

end

The rotationally invariant covariance matrix estimator $\hat{\mathcal{S}}((x,y),(x',y')) = G^{(0)}(x,y) S_s^{(0)} G^{(0)}(x',y')^* + 2\sum_{k=1}^{k_{\max}} G^{(k)}(x,y) S_s^{(k)} G^{(k)}(x',y')^*$. (sPCA)

Reconstruct the denoised image using Eq. (31) (sPCA)

II. Methods

The goal of steerable *e*PCA is to estimate the rotationally invariant covariance matrix from the images whose pixel intensities are affected by photon count noise. To develop this estimator, we combine our previous works on steerable PCA with *e*PCA in a novel way. The main challenge in combining *e*PCA with steerable PCA is that steerable PCA involves a (nearly orthogonal) change of basis transformation from a Cartesian grid to a steerable basis. While multidimensional Gaussian distributions are invariant to orthogonal transformations, the multidimensional Poisson distribution is not invariant to such transformations. Therefore, steerable PCA and *e*PCA cannot be naively combined in a straightforward manner. Algorithm 1 details the steps of the combined procedure that achieves the goal of estimating a rotationally invariant covariance matrix from Cartesian grid images with pixel intensities following a Poisson distribution. The following subsections explain the main concepts underlying *e*PCA and steerable PCA, and how they are weaved together in a manner that overcomes the aforementioned challenge. In the following subsections, we detail the associated concepts and steps for steerable *e*PCA.

A. The Observation Model and Homogenization

We adopt the same observation model introduced in [25]. We observe n noisy images $Y_i \in \mathbb{R}^p$ (i.e., p is the number of pixels), for $i = 1, \ldots, n$. These are random vectors sampled from a hierarchical model defined as follows.

First, a latent vector—or hyperparameter—$\omega \in \mathbb{R}^p$ is drawn from a probability distribution P with mean μ_ω and covariance matrix Σ_ω. Conditioning on ω, each coordinate of $Y = (Y(1), \ldots, Y(p))^\top$ is drawn independently from a canonical one-parameter exponential family,

$$Y(j)|\omega(j) \sim p_{\omega(j)}(y), \quad Y = (Y(1), \ldots, Y(p))^\top, \tag{1}$$

with density

$$p_{\omega(j)}(y) = \exp[\omega(j)y - A(\omega(j))] \tag{2}$$

with respect to a σ-finite measure ν on $\mathbb{R}$, where the jth entry of the latent vector, $\omega(j) \in \mathbb{R}$, is the natural parameter of the family and $A(\omega(j)) = \log \int \exp(\omega(j)y)d\nu(y)$ is the corresponding log-partition function. The mean and variance of the random variable $Y(j)$ can be expressed as $A'(\omega(j))$ and $A''(\omega(j))$, where we denote $A'(\omega) = dA(\omega)/d\omega$. Therefore, the mean of Y conditioning on ω is

$$X := \mathbb{E}(Y|\omega) = (A'(\omega(1)), \ldots, A'(\omega(p)))^\top = A'(\omega),$$

so the noisy data vector Y can be expressed as $Y = A'(\omega) + \tilde{\varepsilon}$, with $\mathbb{E}(\tilde{\varepsilon}|\omega) = 0$ and the marginal mean of Y is $\mathbb{E}Y = \mathbb{E}A'(\omega)$. Thus, one can think of Y as a noisy realization of the clean vector $X = A'(\omega)$. However, the latent vector ω is also random and varies from sample to sample. In the XFEL application, $X = A'(\omega)$ are the unobserved noiseless images, and their randomness stems from the random (and unobserved) orientation of the molecule. We may write[1] $Y = A'(\omega) + \text{diag}[A''(\omega)]^{1/2}\varepsilon$, where the coordinates of ε are conditionally independent and standardized given ω. The covariance of Y is given by the law of total covariance:

$$\begin{aligned}\text{Cov}[Y] &= \text{Cov}[\mathbb{E}(Y|\omega)] + \mathbb{E}[\text{Cov}[Y|\omega]] \\ &= \text{Cov}[A'(\omega)] + \mathbb{E}\,\text{diag}[A''(\omega)].\end{aligned} \tag{3}$$

The off-diagonal entries of the covariance matrix of the noisy images are therefore the same as those of the clean images. However, the diagonal of the covariance matrix (i.e., the variance) of the noisy images is further inflated by the noise variance. Unlike white noise, the noise variance here often changes from one coordinate to another (i.e., there is heteroscedasticity). In *e*PCA, the homogenization step is a particular weighting method that improves the signal-to-noise ratio [25, Section 4.2.2]. Specifically, the homogenized vector is defined as

$$Z = \text{diag}[\mathbb{E}A''(\omega)]^{-1/2}Y = \text{diag}[\mathbb{E}A''(\omega)]^{-1/2}A'(\omega) + \varepsilon. \tag{4}$$

Then the corresponding homogenized covariance matrix is

$$\begin{aligned}\text{Cov}[Z] = {} & \text{diag}[\mathbb{E}A''(\omega)]^{-1/2}\text{Cov}[A'(\omega)] \\ & \times \text{diag}[\mathbb{E}A''(\omega)]^{-1/2} + I.\end{aligned} \tag{5}$$

This step is commonly called "prewhitening" in signal and image processing, while the term "homogenization" is more commonly used in statistics. The terms "prewhitened" and "homogenized" are synonyms in the context of this paper. In Section II-C, we discuss how to estimate the homogenized rotationally invariant covariance matrix.

For the special case of Poisson observations, where $Y \sim \text{Poisson}_p(X)$ and $X \in \mathbb{R}^p$ is random, we can write $Y = X + \text{diag}(X)^{1/2}\varepsilon$. The natural parameter is the vector ω with $\omega(j) = \log X(j)$ and $A'(\omega(j)) = A''(\omega(j)) = \exp(\omega(j)) = X(j)$. Therefore, we have $\mathbb{E}Y = \mathbb{E}X$, and

$$\text{Cov}[Y] = \text{Cov}[X] + \text{diag}[\mathbb{E}X]. \tag{6}$$

In other words, while the mean of the noisy images agrees with the mean of the clean images, their covariance matrices differ by a diagonal matrix that depends solely on the mean image. If we homogenize the observations by $Z = \text{diag}(\mathbb{E}[X])^{-1/2}Y$, then Eq. (6) becomes,

$$\begin{aligned}\text{Cov}[Z] &= \text{diag}(\mathbb{E}[X])^{-1/2}\text{Cov}[Y]\,\text{diag}(\mathbb{E}[X])^{-1/2} \\ &= \text{diag}(\mathbb{E}[X])^{-1/2}\text{Cov}[X]\,\text{diag}(\mathbb{E}[X])^{-1/2} + I.\end{aligned} \tag{7}$$

We can estimate the covariance of X from the homogenized covariance Z according to Eq. (7), i.e. $\text{Cov}[X] = \text{diag}(\mathbb{E}[X])^{1/2}(\text{Cov}[Z] - I)\,\text{diag}(\mathbb{E}[X])^{1/2}$. In Sec. II-E, we detail the corresponding recoloring step to estimate $\text{Cov}[X]$.

In Alg. 1, Steps 1 and 5, with $D_n = \text{diag}[\bar{Y}]$, correspond to the homogenization procedure in *e*PCA. We provide more details on how to estimate the homogenized rotationally invariant covariance matrix (Steps 1–5) in Sec. II-B and Sec. II-C.

B. Steerable Basis Expansion

Under the observation model in Sec. II-A, we develop a method that estimates the rotationally invariant $\text{Cov}[X]$ efficiently and accurately from the image dataset Y. We assume that a digital image I is composed of discrete samples from a continuous function f with band limit c. The Fourier transform of f, denoted $\mathcal{F}(f)$, can be expanded in any orthogonal basis for the class of square-integrable functions in a disk of radius c. For the purpose of steerable PCA, it is beneficial to choose a basis whose elements are products of radial functions with Fourier angular modes, such as the Fourier-Bessel functions, or 2-D prolate functions [35]. For concreteness, in the following we use the Fourier-Bessel functions given by

$$\psi_c^{k,q}(\xi, \theta) = \begin{cases} N_{k,q}\, J_k\left(R_{k,q}\dfrac{\xi}{c}\right)e^{\imath k\theta}, & \xi \le c, \\ 0, & \xi > c, \end{cases} \tag{8}$$

where (ξ, θ) are polar coordinates in the Fourier domain (i.e., $\xi_1 = \xi\cos\theta$, $\xi_2 = \xi\sin\theta$, $\xi \ge 0$, and $\theta \in [0, 2\pi)$; $N_{k,q} = (c\sqrt{\pi}|J_{k+1}(R_{k,q})|)^{-1}$ is the normalization factor; J_k is the Bessel function of the first kind of integer order k; and $R_{k,q}$ is the qth root of the Bessel function J_k. We also assume that the functions of interest are concentrated in a disk of radius R in real domain. In order to avoid aliasing, we only use Fourier-Bessel functions that satisfy the following criterion [26], [36]

$$R_{k,q+1} \le 2\pi c R. \tag{9}$$

For each angular frequency k, we denote by p_k the number of components satisfying Eq. (9). The total number of components is $p = \sum_{k=-k_{\max}}^{k_{\max}} p_k$, where $k_{\max}$ is the maximal possible

[1] $\text{diag}[x]$ for $x \in \mathbb{R}^p$ denotes a $p \times p$ diagonal matrix whose diagonal entries are $x(j)$ for $j = 1, \cdots, p$.

value of k satisfying Eq. (9). We also denote $\gamma_k = \frac{p_k}{2n}$ for $k > 0$ and $\gamma_0 = \frac{p_0}{n}$.

The inverse Fourier transform (IFT) of $\psi_c^{k,q}$ is

$$\mathcal{F}^{-1}(\psi_c^{k,q})(r,\phi) = \frac{2\, c\sqrt{\pi}(-1)^q R_{k,q} J_k(2\pi cr)}{\imath^k (2\pi cr)^2 - R_{k,q}^2} e^{\imath k\phi} \equiv g_c^{k,q}(r) e^{\imath k\phi}, \tag{10}$$

where $g_c^{k,q}(r)$ is the radial part of the inverse Fourier transform of the Fourier-Bessel function. Therefore, we can approximate f using the truncated expansion

$$f(r,\phi) \approx \sum_{k=-k_{\max}}^{k_{\max}} \sum_{q=1}^{p_k} a_{k,q} g_c^{k,q}(r) e^{\imath k\phi}. \tag{11}$$

The approximation error in Eq. (11) is due to the finite truncation of the Fourier-Bessel expansion. For essentially band-limited functions, the approximation error is controlled using the asymptotic behavior of the Bessel functions, see [37, Section 2] for more details. We evaluate the Fourier-Bessel expansion coefficients numerically as in [26] using a quadrature rule that consists of equally spaced points in the angular direction $\theta_l = \frac{2\pi l}{n_\theta}$, with $l = 0, \ldots, n_\theta - 1$ and a Gaussian quadrature rule in the radial direction ξ_j for $j = 1, \ldots, n_\xi$ with the associated weights $w(\xi_j)$. Using the sampling criterion introduced in [26], the values of n_ξ and n_θ depend on the compact support radius R and the band limit c. Our previous work found that using $n_\xi = \lceil 4cR \rceil$ and $n_\theta = \lceil 16cR \rceil$ results in highly-accurate numerical evaluation of the integral to evaluate the expansion coefficients. To evaluate $a_{k,q}$, we need to sample the discrete Fourier transform of the image I, denoted $F(I)$, at the quadrature nodes,

$$F(I)(\xi_j,\theta_l) = \frac{1}{2R} \sum_{i_1=-R}^{R-1} \sum_{i_2=-R}^{R-1} I(i_1,i_2) \times \exp\left(-\imath 2\pi(\xi_j \cos\theta_l i_1 + \xi_j \sin\theta_l i_2)\right), \tag{12}$$

which can be evaluated efficiently using the the nonuniform discrete Fourier transform [38], and we get

$$a_{k,q} \approx \sum_{j=1}^{n_\xi} N_{k,q} J_{k,q}\left(\frac{\xi_j}{c}\right) \widehat{F(I)}(\xi_j,k)\xi_j w(\xi_j), \tag{13}$$

where $\widehat{F(I)}(\xi_j,k)$ is the 1D FFT of $F(I)$ on each concentric circle of radius ξ_j. For real-valued images, it is sufficient to evaluate the coefficients with $k \geq 0$, since $a_{-k,q} = a_{k,q}^*$. In addition, the coefficients have the following properties: under counter-clockwise rotation by an angle α, $a_{k,q}$ changes to $a_{k,q}e^{-\imath k\alpha}$; and under reflection, $a_{k,q}$ changes to $a_{-k,q}$. The numerical integration error in Eq. (13) was analyzed in [26] and drops below 10^{-17} for the chosen values of n_ξ and n_θ.

The steerable basis expansion are applied in two parts of the steerable *e*PCA: (1) the rotationally invariant sample mean estimation in Step 3 of Alg. 1 and (2) the expansion of the whitened images in Step 7 of Alg. 1.

C. The Sample Rotationally-Invariant Homogenized Covariance Matrix

Suppose $I_1, \ldots, I_n$ are n discretized input images sampled from $f_1, \ldots, f_n$. Here, the observation vectors Y_i for *e*PCA are simply $Y_i = I_i$. In *e*PCA, the first step is to prewhiten the data using the sample mean as suggested by Eqs. (4) and (6). However, the sample mean $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ is not necessarily rotationally invariant. With the estimated band limit and support size in Step 2, we compute the truncated Fourier-Bessel expansion coefficients of $F(\bar{Y})$, denoted by $\bar{a}_{k,q}$. The rotationally invariant sample mean can be evaluated from $\bar{a}_{k,q}$,

$$\bar{f}(r,\phi) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{n}\sum_{i=1}^n f_i(r,\phi-\alpha)d\alpha \approx \sum_{q=1}^{p_0} \bar{a}_{0,q} g_c^{0,q}(r). \tag{14}$$

The approximation error in Eq. (14) follows directly from that of Eq. (11). The rotationally invariant sample mean is circularly symmetric. We denote by $\bar{A}$ a vector that contains all the coefficients $\bar{a}_{0,q}$ ordered by the radial index q. Although the input images are non-negative, the finite truncation may result in small negative values in the estimated mean, so we threshold any negative entries to zero. As mentioned in Sec. II-A, the expected covariance matrix of the Poisson observations differs from the covariance matrix of clean data by a diagonal matrix, where the diagonal entries are equal to the mean image. Therefore, in Step 4 of Alg. 1, we have $D_n = \mathrm{diag}[\bar{f}]$ and it is used in Step 5 to compute the homogenized vectors, similar to Eq. (4).

We prewhiten the images by the estimated mean image to create new images $Z_1, \ldots, Z_n$ as $Z_i(x,y) = \bar{f}(x,y)^{-1/2} Y_i(x,y)$, when $\bar{f}(x,y) > 0$, and $Z_i(x,y)$ is 0 otherwise. The whitening step might change the band limit. Therefore, we estimate the band limit of the whitened images in Step 6. Combining Eqs. (9) and (13), we compute the truncated Fourier-Bessel expansion coefficients $a_{k,q}^i$ of $F(Z_i)$. Let us denote by $A^{(k)}$ the matrix of expansion coefficients with angular frequency k, obtained by putting $a_{k,q}^i$ into a matrix, where the columns are indexed by the image number i and the rows are ordered by the radial index q. The coefficient matrix $A^{(k)}$ is of size $p_k \times n$.

We use $\bar{Z}$ to represent the rotationally invariant sample mean of the whitened images. Under the action of the group $O(2)$, i.e. counter-clock wise rotation by an angle $\alpha \in [0, 2\pi)$ and reflection $\beta \in \{+,-\}$, where '+' indicates no reflection and '−' indicates with reflection, the image Z_i is transformed to $Z_i^{\alpha,\beta}$. Since the truncated Fourier-Bessel transform is almost unitary [26], the rotationally invariant covariance kernel $\mathcal{S}\left((x,y),(x',y')\right)$ built from the whitened image data with all possible in-plane rotations and reflections, defined as,

$$\mathcal{S}\left((x,y),(x',y')\right) = \frac{1}{4\pi n}\sum_{i=1}^n \sum_{\beta\in\{+,-\}} \int_0^{2\pi} \left(Z_i^{\alpha,\beta}(x,y) - \bar{Z}(x,y)\right)\left(Z_i^{\alpha,\beta}(x',y') - \bar{Z}(x',y')\right) d\alpha, \tag{15}$$

can be computed in terms of the IFT of the Fourier-Bessel basis and the associated expansion coefficients. Subtracting the

sample mean is equivalent to subtracting $\frac{1}{n}\sum_{j=1}^{n} a_{0,q}^{j}$ from the coefficients $a_{0,q}^{i}$, while keeping other coefficients unchanged. Therefore, we first update the zero angular frequency coefficients by $a_{0,q}^{i} \leftarrow a_{0,q}^{i} - \frac{1}{n}\sum_{j=1}^{n} a_{0,q}^{j}$. In terms of the expansion coefficients, the rotationally invariant homogenized sample covariance matrix is $S_h = \bigoplus_{k=-k_{\max}}^{k_{\max}} S_h^{(k)}$, with

$$S_h^{(k)} = \frac{1}{n}\mathrm{Re}\left\{A^{(k)}\left(A^{(k)}\right)^*\right\}. \tag{16}$$

We further denote the eigenvalues and eigenvectors of $S_h^{(k)}$ by $\lambda_i^{(k)}$ and $\hat{u}_i^{(k)}$, that is,

$$S_h^{(k)} = \sum_{i=1}^{p_k} \lambda_i^{(k)} \hat{u}_i^{(k)} (\hat{u}_i^{(k)})^*. \tag{17}$$

The procedure of homogenization for steerable *e*PCA is detailed in Steps 1–5 in Alg. 1. Step 9 of Alg. 1 computes the rotationally invariant homogenized sample covariance matrix.

D. Eigenvalue Shrinkage

For data corrupted by additive white noise (with noise variance 1 in each coordinate), previous work [39]–[42] showed that if the population eigenvalue ℓ is above the Baik, Ben Arous, Péché (BBP) phase transition, then the corresponding sample eigenvalue lies to the right of the Marčenko-Pastur distribution of the "noise" eigenvalues. The sample eigenvalue will converge to the value given by *the spike forward map* as $p_k, n \to \infty$, and $p_k/n = \gamma_k$:

$$\lambda(\ell;\gamma_k) = \begin{cases} (1+\ell)\left(1+\dfrac{\gamma_k}{\ell}\right) & \text{if} \quad \ell > \sqrt{\gamma_k}, \\ \left(1+\sqrt{\gamma_k}\right)^2 & \text{otherwise.} \end{cases} \tag{18}$$

The underlying clean population covariance eigenvalues can be estimated by solving the quadratic equation in Eq. (18),

$$\hat{\ell} = \eta_{\gamma_k}(\lambda) = \begin{cases} \dfrac{\left(\lambda - 1 - \gamma_k + \sqrt{(\lambda-1-\gamma_k)^2 - 4\gamma_k}\right)}{2}, & \lambda > (1+\sqrt{\gamma_k})^2, \\ 0 & \lambda \le (1+\sqrt{\gamma_k})^2. \end{cases} \tag{19}$$

Shrinking the eigenvalues improves the estimation of the sample covariance matrix [43]. Since the homogenized$_i$ sample covariance matrix S_h is decoupled into small sub-blocks $S_h^{(k)}$, the shrinkers are defined for each frequency k separately. The shrinkers $\eta_{\gamma_k}(\lambda)$ set all noise eigenvalues to zero for λ within the support of the Marčenko-Pastur distribution and reduce other eigenvalues according to Eq. (19). Then the denoised covariance matrices are

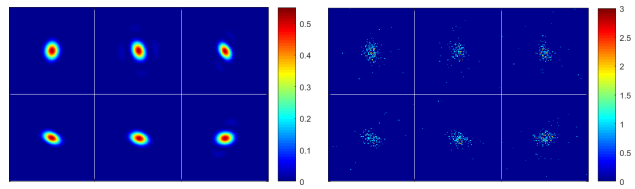
$$S_{h,\eta}^{(k)} = \sum_{i=1}^{r_k} \eta_{\gamma_k}\left(\lambda_i^{(k)}\right) \hat{u}_i^{(k)} (\hat{u}_i^{(k)})^*, \tag{20}$$

where r_k is the number of components with $\eta_{\gamma_k}(\lambda) > 0$. The empirical eigenvector $\hat{u}^{(k)}$ of $S_h^{(k)}$ is an inconsistent estimator of the true eigenvector. We can heuristically quantify the inconsistency based on results from the Gaussian standard spiked model, even though the noise is non-Gaussian. Under this model, the empirical and true eigenvectors have an asymptotically deterministic angle: $\left(\left(u^{(k)}\right)^* \hat{u}^{(k)}\right)^2 \to c^2(\ell;\gamma_k)$ almost surely, where $c(\ell;\gamma_k)$ is the *cosine forward map* given by [41], [42]:

$$c^2(\ell;\gamma_k) = \begin{cases} \dfrac{1-\gamma_k/\ell^2}{1+\gamma_k/\ell} & \text{if} \quad \ell > \sqrt{\gamma_k}, \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

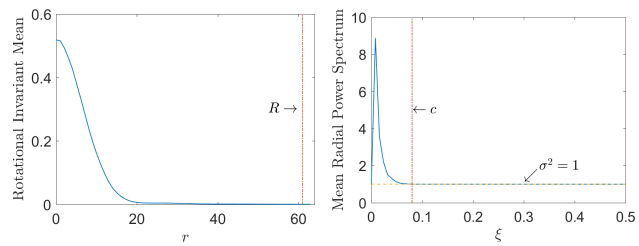
Therefore, asymptotically for the population eigenvectors beyond the BBP phase transition, the sample eigenvectors have positive correlation with the population eigenvectors, but this correlation is less than 1 [44]–[46]. We denote by $\hat{c}$ an estimate of c using the estimated clean covariance eigenvalues $\hat{\ell}$ in Eq. (19) and $\hat{s}^2 = 1 - \hat{c}^2$.

In short, Step 10 of Alg. 1 improves the estimation of the rotationally invariant homogenized covariance matrix through eigenvalue shrinkage.



(a) clean (b) noisy

Fig. 1. Sample clean and noisy images of the XFEL dataset. Image size is 128×128 with mean per pixel photon count $= 0.01$.



(a) Estimate R (b) Estimate c

Fig. 2. Estimating R and c from $n = 7 \times 10^4$ simulated noisy XFEL diffraction intensity maps of lysozyme. Each image is of size 128×128 pixels. (a) The radial profile of the rotationally invariant sample mean image. The radius of compact support is chosen at $R = 61$. (b) Mean radial power spectrum of the whitened noisy images. The curve levels off at $\sigma^2 = 1$.

E. Recoloring

Homogenization changes the direction of the clean eigenvectors. Therefore, after eigenvalue shrinkage, we recolor (heterogenize) the covariance matrix by conjugating the recoloring matrix B with $S_{h,\eta}$: $S_{he} = B^* \cdot S_{h,\eta} \cdot B$. The recoloring matrix is derived as,

$$\begin{aligned} B_{k_1,q_1;k_2,q_2} &= \int_0^R \int_0^{2\pi} \sqrt{\bar{f}(r)}\, \mathcal{F}^{-1}\left(\psi_c^{k_1,q_1}\right)(r,\theta) \\ &\quad \times \left(\overline{\mathcal{F}^{-1}\left(\psi_c^{k_2,q_2}\right)(r,\theta)}\right) r\, dr d\theta \\ &= \int_0^R \sqrt{\bar{f}(r)}\, g_c^{k_1,q_1}(r)\, \overline{g_c^{k_2,q_2}(r)}\, r\, dr \\ &\quad \times \int_0^{2\pi} e^{\imath(-k_1+k_2)\theta} d\theta \\ &= \delta_{k_1,k_2} \int_0^R \sqrt{\bar{f}(r)}\, g_c^{k_1,q_1}(r)\, \overline{g_c^{k_2,q_2}(r)}\, r dr, \end{aligned} \tag{22}$$

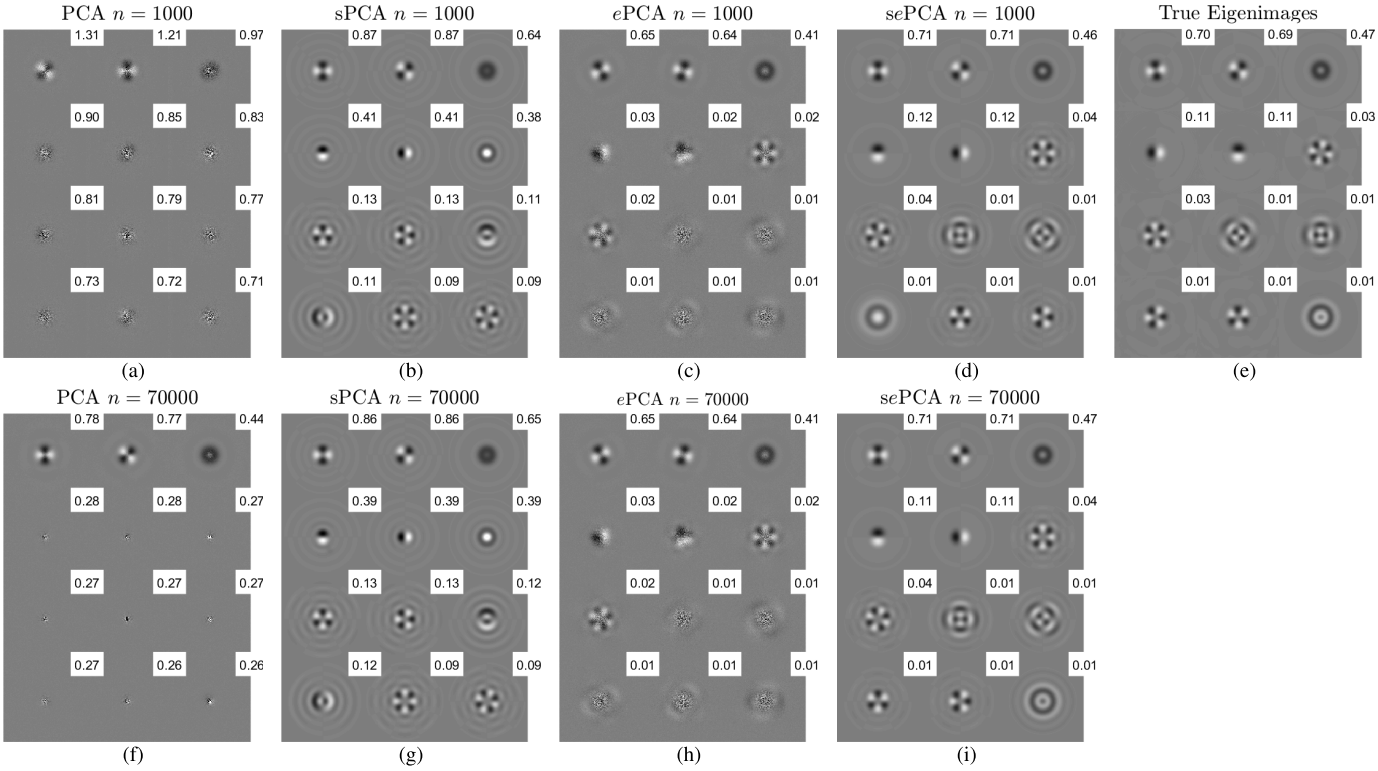


Fig. 3. Eigenimages estimated from noisy XFEL data using PCA, steerable PCA (sPCA), ePCA, and steerable ePCA (se$PCA), ordered by eigenvalues. Input images are corrupted by poisson noise with mean photon count 0.01 (shown in Figure 1b).

which has a block diagonal structure and is decoupled for each angular frequency, $B = \bigoplus_{k=-k_{\max}}^{k_{\max}} B^{(k)}$, with

$$B_{q_1,q_2}^{(k)} = \int_0^R \sqrt{\bar{f}(r)}\, g_c^{k,q_1}(r)\, \overline{g_c^{k,q_2}(r)}\, r\, dr. \tag{23}$$

The radial integral in Eq. (23) is numerically evaluated using the Gauss-Legendre quadrature rule [47, Chap. 4], which determines the locations of $n_r = \lceil 4cR \rceil$ points $\{r_j\}_{j=1}^{n_r}$ on the interval $[0, R]$ and the associated weights $w(r_j)$. The integral in Eq. (23) is thus approximated by

$$B_{q_1,q_2}^{(k)} \approx \sum_{j=1}^{n_r} \sqrt{\bar{f}(r_j)}\, g_c^{k,q_1}(r_j)\, \overline{g_c^{k,q_2}(r_j)} r_j w(r_j). \tag{24}$$

The recoloring step is also decoupled for each angular frequency sub-block. The heterogenized covariance estimators are

$$S_{he}^{(k)} = \left(B^{(k)}\right)^* \cdot S_{h,\eta_{\gamma_k}}^{(k)} \cdot B^{(k)}. \tag{25}$$

Similar to Eq. (23), we define $D^{(k)}$ which will be used to scale the heterogenized covariance matrix estimator (see Eq. (27)) and denoise the expansion coefficients (see Eqs. (29) and (30)),

$$\begin{aligned} D_{q_1,q_2}^{(k)} &= \int_0^R \bar{f}(r)\, g_c^{k,q_1}(r)\, \overline{g_c^{k,q_2}(r)}\, r\, dr \\ &\approx \sum_{j=1}^{n_r} \bar{f}(r_j)\, g_c^{k,q_1}(r_j)\, \overline{g_c^{k,q_2}(r_j)} r_j w(r_j). \end{aligned} \tag{26}$$

In Alg. 1, Steps 11 and 12 summarize the recoloring procedure.

F. Scaling

The eigendecomposition of $S_{he}^{(k)}$ gives $S_{he}^{(k)} = \sum_{i=1}^{r_k} \hat{v}_i^{(k)} \left(\hat{v}_i^{(k)}\right)^*$. The empirical eigenvalues are $\hat{t} = \|\hat{v}^{(k)}\|^2$ which is a biased estimate of the true eigenvalues of the clean covariance matrix Σ_X. In [25, Sec. 4.2.3], a scaling rule was proposed to correct the bias. We extend it in Steps 13 and 14 in Alg. 1 to the steerable case and scale each eigenvalue of $S_{he}^{(k)}$ by a parameter $\hat{\alpha}^{(k)}$,

$$\hat{\alpha}^{(k)} = \begin{cases} \dfrac{1 - \hat{s}^2 \tau^{(k)}}{\hat{c}^2}, & \text{for } \; 1 - \hat{s}^2 \tau^{(k)} > 0 \;\text{ and }\; \hat{c}^2 > 0 \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

where the parameter $\tau^{(k)} = \frac{\mathrm{tr} D^{(k)}}{p_k} \cdot \frac{\hat{\ell}}{\|\hat{v}^{(k)}\|^2}$. The scaled covariance matrices are

$$S_s^{(k)} = \sum_{i=1}^{r_k} \hat{\alpha}_i^{(k)} \hat{v}_i^{(k)} \left(\hat{v}_i^{(k)}\right)^*. \tag{28}$$

The rotationally invariant covariance kernel $\mathcal{S}((x, y), (x', y'))$ is well approximated by $\sum_{k=-k_{\max}}^{k_{\max}} G^{(k)}(x, y) S_s^{(k)} \left(G^{(k)}(x', y')\right)^*$, where $G^{(k)}$ contains IFT of all $\psi_c^{k,q}$ with angular frequency k (see Step 17) in Alg. 1). The computational complexity of steerable ePCA is $O(nL^3 + L^4)$, same as steerable PCA, and it is lower than the complexity of ePCA which is $O(\min(nL^4 + L^6, n^2L^2 + n^3))$.

In summary, Step 14 scales the heterogenized covariance matrix S_{he}. The covariance matrix in the original pixel domain is efficiently computed from $S_s^{(k)}$ in Step 17.

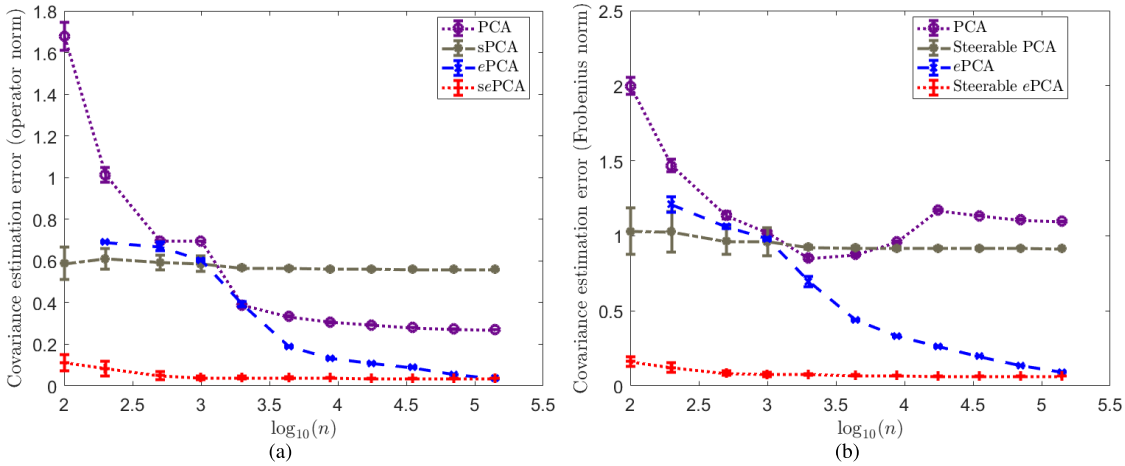


Fig. 4. Error of covariance matrix estimation, measured as the (a) operator norm and (b) Frobenius norm of the difference between each covariance estimate and the true covariance matrix. The sample size n ranges from 100 to 140,000.

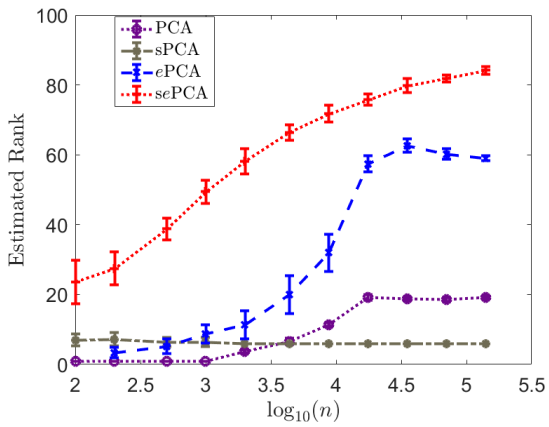


Fig. 5. The estimated number of signal principal components using PCA, sPCA, *e*PCA and *se*PCA. Images are corrupted by poisson noise with mean per pixel photon count 0.01.

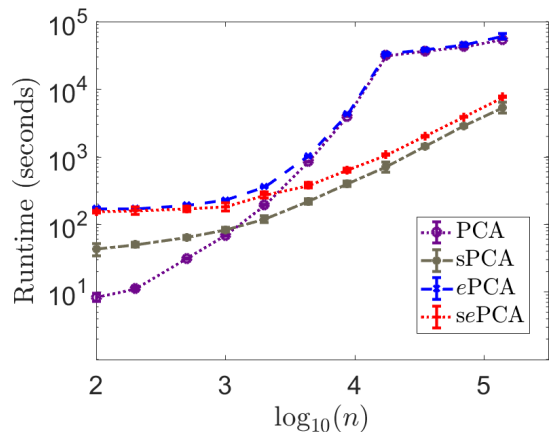
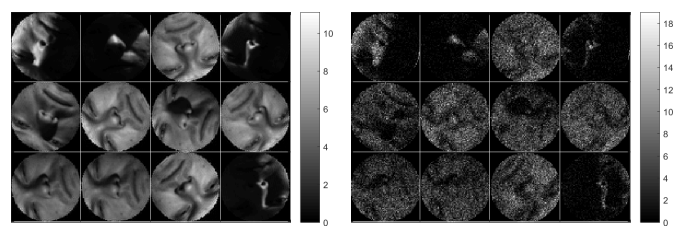


Fig. 6. The runtime for computing the principal components using PCA, sPCA, *e*PCA and *se*PCA. Images are corrupted by poisson noise with mean per pixel photon count = 0.01.



(a) Clean (b) Noisy

Fig. 7. Sample clean and noisy images of randomly rotated yale face database B. Image size is 64×64 with mean intensity 2.3 photons per pixel.

G. Denoising

As an application of steerable *e*PCA, we develop a method to denoise the photon-limited images. Since the Fourier-Bessel expansion coefficients are computed from the prewhitened images, we first recolor the coefficients by multiplying $B^{(k)}$ with $A^{(k)}$ and then we apply Wiener-type filtering to denoise the coefficients. For the steerable basis expansion coefficients $A^{(k)}$ with angular frequency $k \neq 0$,

$$\hat{A}^{(k)} = S_s^{(k)} \left(D^{(k)} + S_s^{(k)}\right)^{-1} B^{(k)} A^{(k)}. \tag{29}$$

For $k = 0$, we need to take into account the rotationally invariant mean expansion coefficients,

$$\hat{A}^{(0)} = S_s^{(0)} \left(D^{(0)} + S_s^{(0)}\right)^{-1} B^{(0)} A^{(0)} + D^{(0)} \left(D^{(0)} + S_s^{(0)}\right)^{-1} \bar{A} \mathbf{1}_n^\top. \tag{30}$$

The denoised image sampled on the Cartesian grid (x, y) in real domain are computed from the filtered expansion coefficients $\hat{a}_{k,q}^i$,

$$\hat{X}_i(x, y) = \sum_{q=1}^{p_0} \hat{a}_{0,q}^i \, g_c^{0,q}(r_{x,y}) + 2\mathrm{Re}\left[\sum_{k=1}^{k_{\max}} \sum_{q=1}^{p_k} \hat{a}_{k,q}^i \, g_c^{k,q}(r_{x,y}) e^{-\imath k \theta_{x,y}}\right], \tag{31}$$

where $r_{x,y} = \sqrt{x^2 + y^2}$ and $\theta_{x,y} = \tan^{-1}\left(\frac{y}{x}\right)$.

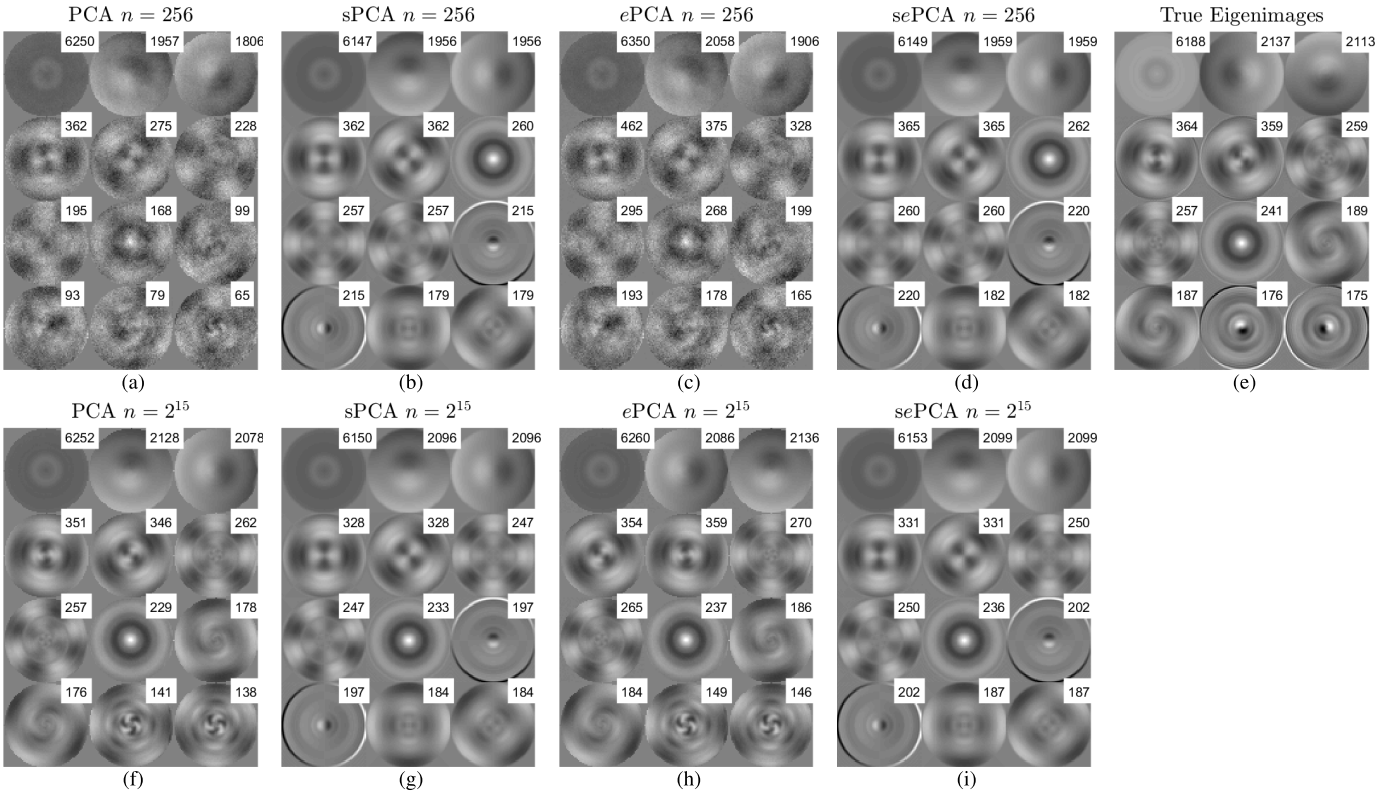


Fig. 8. Eigenimages estimated from noisy rotated yale face database B using PCA, sPCA, *e*PCA, and s*e*PCA, ordered by eigenvalues. Input images are corrupted by poisson noise with mean intensity 2.3 photons per pixel (shown in Fig. 7b).

In essence, we first denoise the recolored steerable expansion coefficients in Step 15 of Alg. 1 according to Eqs. (29) and (30) and then reconstruct the image using Eq. (31) in Step 18 of Alg. 1.

III. Numerical Results

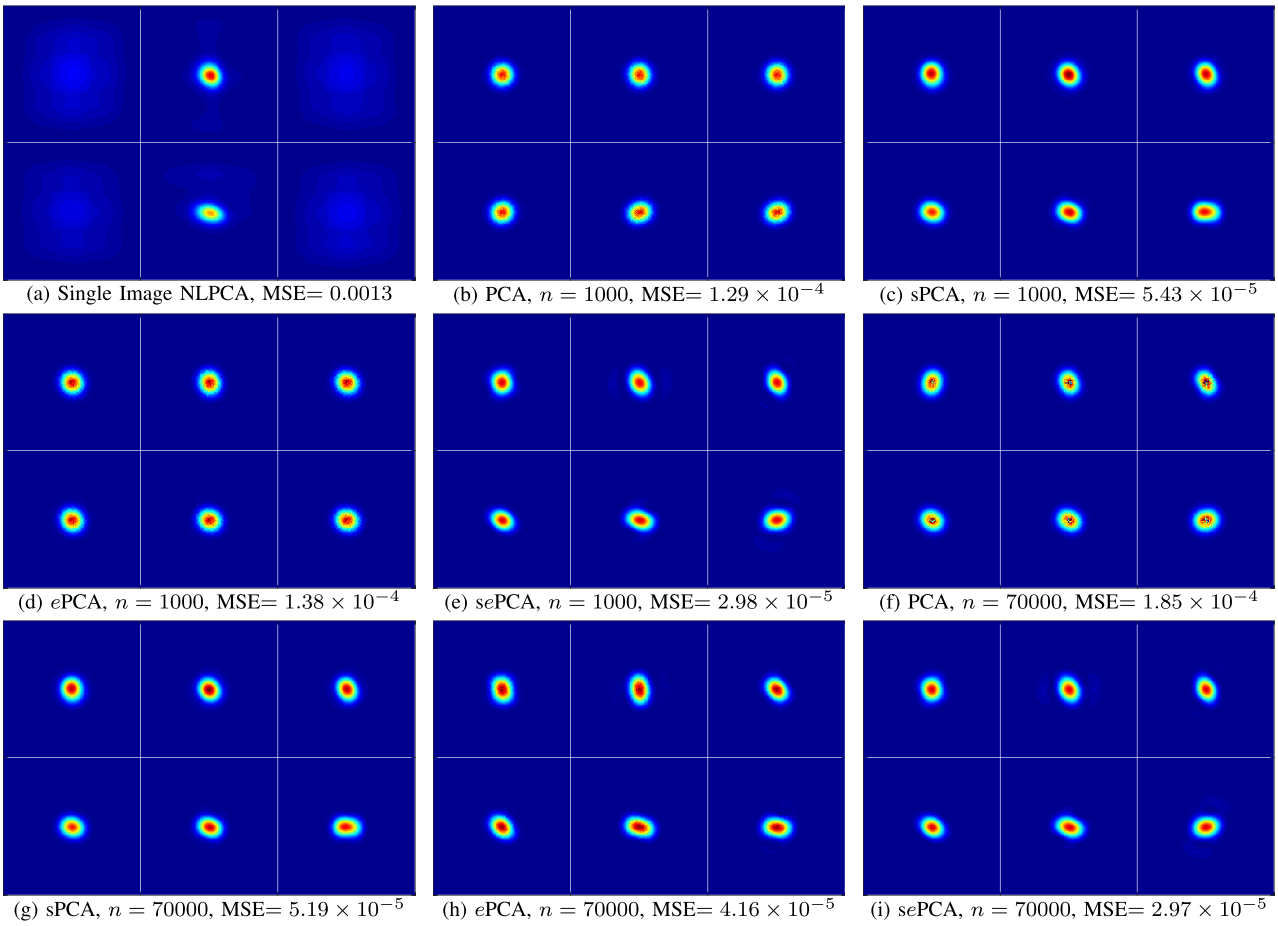
We apply PCA, *e*PCA, steerable PCA, and steerable *e*PCA to a simulated XFEL dataset and compare the results for covariance estimation and denoising. The algorithms are implemented in MATLAB on a machine with 60 cores, running at 2.3 GHz, with total RAM of 1.5TB. Only 8 cores were used in our experiments.

We simulate $n = 140,000$ noiseless XFEL diffraction intensity maps of a lysozyme (Protein Data Bank 1AKI) with Condor [48]. The average pixel intensity is rescaled to be 0.01 for image size 128×128 pixels such that shot noise dominates [49]. To sample an arbitrary number n of noisy diffraction patterns, we sample an intensity map at random, and then sample the photon count of each detector pixel from a Poisson distribution whose mean is the pixel intensity. Figs. 1a and 1b illustrate the clean intensity maps and the resulting noisy diffraction patterns.

We estimate the radius of the concentration (R) of the diffraction intensities in real domain and the band limit (c) in Fourier domain from the noisy images in the following way. The data variance is proportional to the sample mean. The region with non-zero rotationally invariant sample mean contains the object of interest. We show the radial profile of the estimated sample mean in Fig. 2a, which is also the variance map of the dataset averaged in the angular direction. At large r, the radial part of the rotational invaraint sample mean levels off at 0. We compute the cumulative variance by integrating the radial sample mean over r with a Jacobian weight rdr. For the XFEL data, the fraction of the cumulative mean exceeds $\eta = 99.9\%$ at $r = 61$, and therefore R was chosen to be 61 (see Fig. 2a). We compute the whitened projection images using the rotationally invariant sample mean. To estimate the parameter c for the whitened images, we compute the angular average of the mean 2D power spectrum. With a large number of images, a region in the Fourier domain contains signal information, if the corresponding radial mean power spectrum values are larger than the noise variance $\sigma^2 = 1$. The curve in Fig. 2b levels off at the noise variance $\sigma^2 = 1$ when ξ is large. We use the same method as before to compute the cumulative radial power spectrum. The fraction reaches $\eta = 99.9\%$ at $\xi = 0.08$, therefore the band limit is chosen to be $c = 0.08$. The band limit c and support radius R are used in both steerable PCA and steerable *e*PCA. Both parameters can be chosen by the user through either setting the values of R and c directly, or controlling the threshold value η. The proposed parameter selection procedure leads to good empirical results for covariance estimation. The impact of different choices of c and R on denoising is illustrated in Fig. 13.

A. Covariance Estimation and Principal Components

Fig. 3 shows the top 12 eigenimages for clean XFEL diffraction patterns (Fig. 3e), and noisy diffraction patterns



(a) Single Image NLPCA, MSE= 0.0013 (b) PCA, $n = 1000$, MSE= 1.29×10^{-4} (c) sPCA, $n = 1000$, MSE= 5.43×10^{-5}

(d) *e*PCA, $n = 1000$, MSE= 1.38×10^{-4} (e) s*e*PCA, $n = 1000$, MSE= 2.98×10^{-5} (f) PCA, $n = 70000$, MSE= 1.85×10^{-4}

(g) sPCA, $n = 70000$, MSE= 5.19×10^{-5} (h) *e*PCA, $n = 70000$, MSE= 4.16×10^{-5} (i) s*e*PCA, $n = 70000$, MSE= 2.97×10^{-5}

Fig. 9. Sample denoised images of the XFEL dataset illustrated in Fig. 1 using NLPCA, PCA, sPCA, *e*PCA, s*e*PCA. Image size is 128×128.

with mean photon count per pixel 0.01 (Figs. 3c–fig:eigVeci) using PCA, steerable PCA, *e*PCA, and steerable *e*PCA. The true eigenimages in Fig. 3e are computed from 70000 clean diffraction patterns whose orientations are uniformly distributed over $SO(3)$. Figs. 3a–3d are computed from 1000 noisy images. Since the number of samples is much smaller than the size of the image and the noise type is non-Gaussian, PCA can only recover the first two or three components. *e*PCA improves the estimation and is able to extract the top 7 eigenimages. Moreover, steerable PCA and steerable *e*PCA achieve much better estimation of the underlying true eigenimages for a given sample size. Steerable *e*PCA achieves the best performance in estimating both the eigenvalues and eigenimages.

Furthermore, we compare the operator norms and Frobenius norms of the difference between the covariance estimates and the true covariance matrix. Fig. 4 shows that steerable *e*PCA significantly improves the covariance estimation, especially when the sample size is small.

For experiments using *e*PCA, we use a permutation bootstrap-based method to estimate the rank of the covariance matrix, following e.g. [50]. By randomly permuting each column of the mean-subtracted data matrix, we completely destroy structural information including linear structure, while the noise statistics remain the same (see [51], [52] for an analysis). Singular values of the randomly permuted

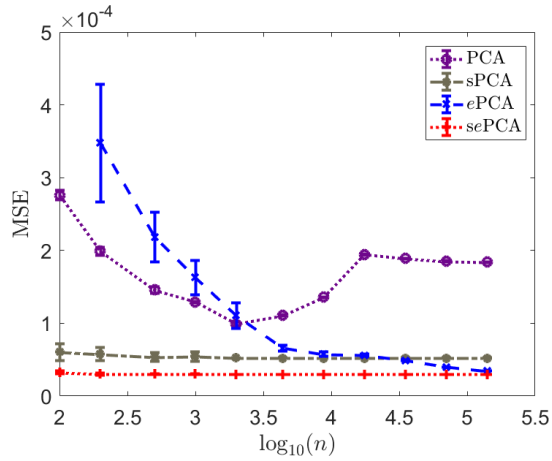
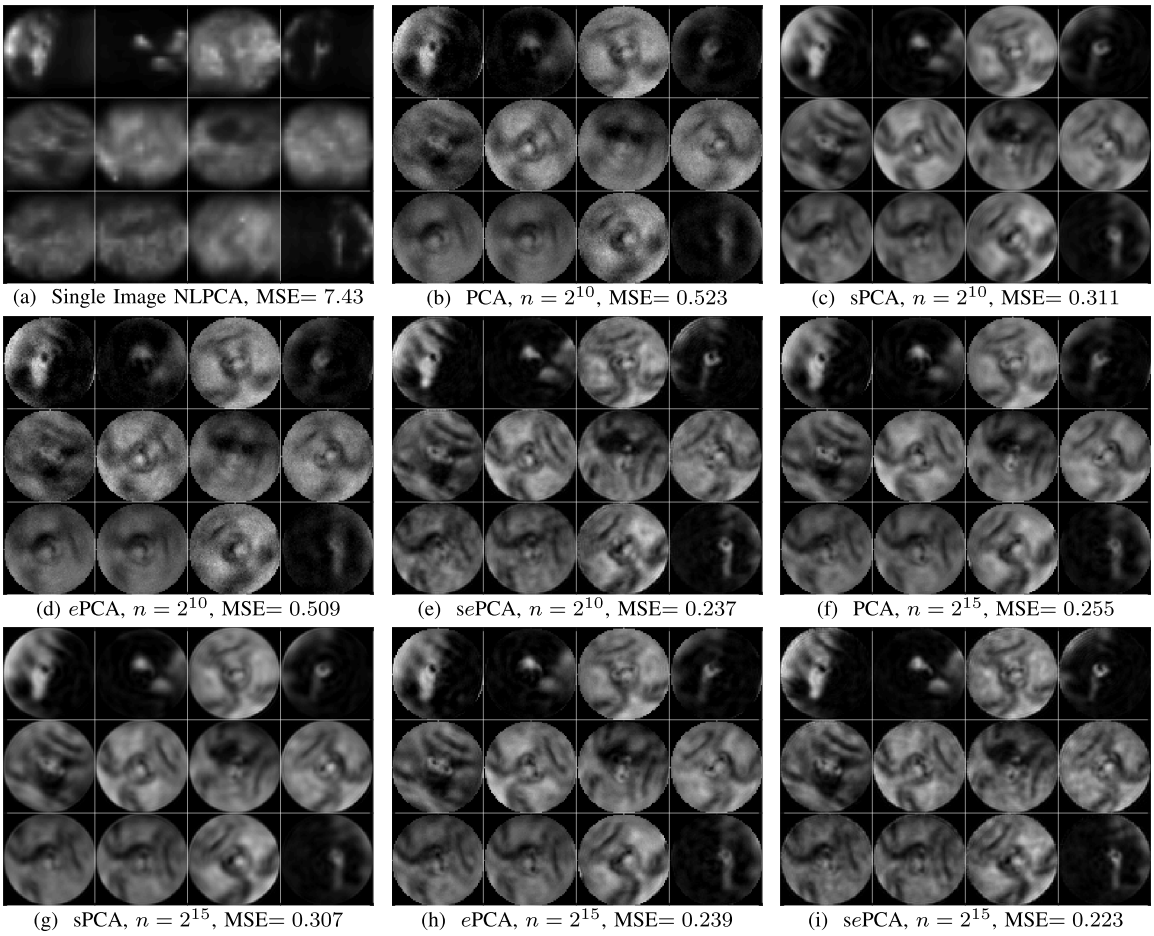


Fig. 10. Comparing the denoising quality of the XFEL dataset ($p = 16384$) with various number of images.

matrices reveal what should be the largest covariance matrix eigenvalues that correspond to noise, up to a user-selected confidence level ρ. This can replace the other rank estimation methods that assume Gaussianity when the noise model is non-Gaussian, such as in our case. Empirically, we observe that $\rho = 0.1$ gives the best performance in covariance estimation. For steerable *e*PCA, we estimate the number of components



(a) Single Image NLPCA, MSE= 7.43 (b) PCA, $n = 2^{10}$, MSE= 0.523 (c) sPCA, $n = 2^{10}$, MSE= 0.311

(d) ePCA, $n = 2^{10}$, MSE= 0.509 (e) sePCA, $n = 2^{10}$, MSE= 0.237 (f) PCA, $n = 2^{15}$, MSE= 0.255

(g) sPCA, $n = 2^{15}$, MSE= 0.307 (h) ePCA, $n = 2^{15}$, MSE= 0.239 (i) sePCA, $n = 2^{15}$, MSE= 0.223

Fig. 11. Sample denoised images from yale face database B illustrated in Fig. 7 using NLPCA, PCA, sPCA, ePCA, and sePCA. Image size is 64×64. The intensity range is [0, 2.6] for (a), and is [0, 11.1] for (b)–(i).

using the right edge of the Marčenko-Pastur distribution for homogenized covariance matrices $S_h^{(k)}$ and include only the components whose scaling factor $\hat{\alpha}^{(k)}$ are above zero.

Steerable ePCA is able to recover more signal principal components from noisy images than PCA, steerable PCA, and ePCA (see Fig. 5). When the sample size $n = 1000$, the mean number of estimated components is 11 and 49 for ePCA and steerable ePCA respectively. For $n = 70000$, the estimated number of components is 59 and 81 for ePCA and steerable ePCA respectively. Fig. 6 shows that steerable ePCA is more efficient than ePCA and PCA. Because steerable ePCA contains extra steps such as prewhitening, recoloring, and scaling, its runtime is slightly longer than steerable PCA. When $n = 140000$, steerable ePCA is 8 times faster than ePCA.

In addition to the XFEL diffraction intensity data, we include a natural image dataset–Yale Face Database B [33], [34]–to illustrate the efficacy of the proposed method. The database contains 5760 single light source images of 10 subjects with 9 poses. For every subject in a particular pose, 64 images were captured under different ambient illuminations. The original images of a single face under different lighting conditions inhabit an approximately 9-dimensional linear space [53]. In the experiments, we take one subject at a particular pose with 64 different illumination conditions. We downsample the original images to the size of 64×64 pixels and scale the intensity such that the mean photon count for the whole dataset is 2.3 photons per pixel (i.e., "faces in the dark"). Since we need to rotate the images, we mask images by a disk of radius $R = 32$. We uniformly sample n clean images from the original images and apply a random in-plane rotation to each sample. Fig. 7 shows 12 samples of the clean data and the corresponding noisy observations. The corresponding principal components are illustrated in Fig. 8. The true eigenimages are evaluated using the clean images rotated at every 0.7 degrees.

B. Denoising

We compare the denoising effects of steerable ePCA with PCA, steerable PCA, ePCA, and patch-based single image non-local PCA (NLPCA) [21], by the mean squared error, MSE:= $(pn)^{-1} \sum_{i=1}^{n} \|\hat{X}_i - X_i\|^2$. We perform "$e$PCA denoising" using the empirical best linear predictor (EBLP) [25], which had been shown to outperform "PCA denoising," i.e., orthogonal projection onto sample or ePCA / heterogenized eigenimages, as well as the exponential family PCA method proposed by [19]. Note that in our implementation of ePCA Wiener-type filtering, to avoid inverting a singular

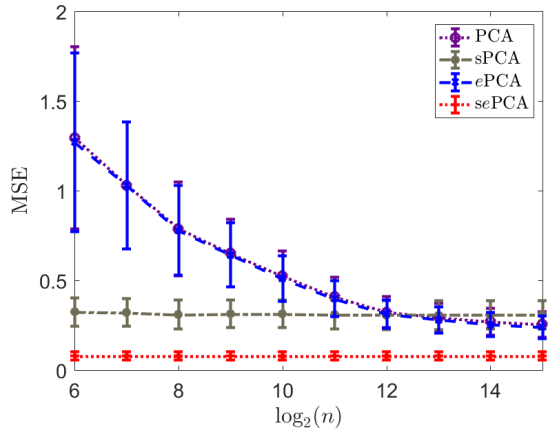
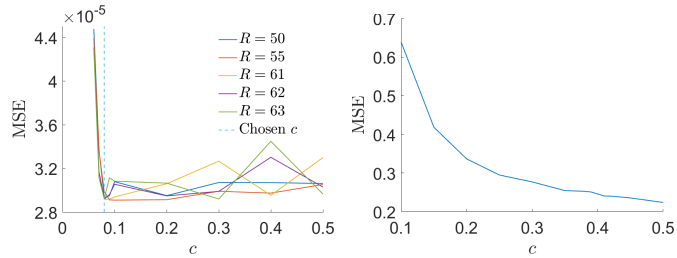


Fig. 12. Comparing the denoising quality of the yale face database B ($p = 4096$) with various number of images.



(a) XFEL images (b) Rotated Yale Face Database B

Fig. 13. Sensitivity of the denoising performance over the parameters c and R. For yale face database B, we cropped the images by the disc of radius $R = 32$.

matrix (when some coordinates have 0 sample mean and sample variance), we compute $\mathrm{diag}[\bar{Y}]+S_s$ with regularization, $\mathrm{diag}[\bar{Y}] + S_s \leftarrow S_s + (1-\epsilon)\,\mathrm{diag}[\bar{Y}] + \epsilon m I$ where $\epsilon = 0.1$, $m = \frac{1}{p}\bar{Y}\cdot\mathbf{1}$, and I is the $p \times p$ identity matrix. The number of components are estimated by the permutation rank estimation as described in the previous section.

Fig. 9 shows some examples of denoised XFEL images for sample size $n = 1000$ and 70000. For robustness, we repeated the numerical experiment for another dataset simulated from the small protein chignolin (Protein Data Bank entry 1UAO) and obtained qualitatively similar results. Steerable *e*PCA is able to recover the images with lower MSEs compared to PCA, steerable PCA, and *e*PCA (see Fig. 10), especially when the sample size n is small.

Fig. 11 shows examples of the denoised face images for sample size $n = 2^{10}$ and 2^{15}. We use the permutation rank estimation described in the previous section with confidence level $\rho = 0.1$ for PCA, sPCA, and *e*PCA. In this example, we see that when the number of images is small, PCA and *e*PCA can only recover rough contours of the faces, whereas sPCA and s*e*PCA are able to recover finer details in the faces. The MSEs of the denoised images with varying number of samples are plotted in Fig. 12. In this experiment, the data only contains 64 unrotated clean images and the mean photon count is 2.3 per pixel. Compared to the diffraction patterns with mean photon count 0.01 per pixel. Therefore, s*e*PCA is able to sufficiently capture the clean data subspace with only a small number of images, resulting in a much flatter MSE curve with respect to the number of images.

We also test how sensitive our method is to the choice of parameters, i.e. the band limit c and the support radius R. For the diffraction intensity data, we use $n = 10000$ diffraction patterns with mean photon count 0.01 per pixel. The underlying clean diffraction patterns are very smooth. We vary the band limit from $c = 0.06$ to 0.5 and the support radius R ranges from 50 to 63. With the chosen parameters indicated in Fig. 2, the MSE for the denoised image is small (see Fig. 13a). Although the MSEs vary when we increase the parameter c above the chosen band limit and change R within the range of 50 to 63, the errors are still relatively small ($2.9-3.5\times 10^{-5}$), compared to above 5×10^{-5} with *e*PCA and sPCA. For the face data, we compare the results with $n = 8192$ noisy images with 2.3 photons per pixel. Since we crop out the region within a disk of radius $R = L/2$, the parameter R is fixed to be 32 and we vary the band limit c from 0.1 to 0.5. We observe that choosing $c = 0.5$ provides the best denoising performance (see Fig. 13b), and this agrees with the fact that the face images contain more high-frequency information than the XFEL data.

IV. Conclusion

We presented steerable *e*PCA, a method for principal component analysis of a set of low-photon-count images and their uniform in-plane rotations. This work has been mostly motivated by its application to XFEL, but is relevant to other low-photon-count imaging applications. The computational complexity of the new algorithm is $O(nL^3 + L^4)$, whereas that of *e*PCA is $O(\min(nL^4 + L^6, n^2\ L^2 + n^3))$. Incorporating rotational invariance allows more robust estimation of the true eigenvalues and eigenvectors. Our numerical experiments showed that steerable *e*PCA more accurately estimates the covariance matrix and achieves better denoising results. Finally, we remark that the Fourier-Bessel basis can be replaced with other suitable bases, for example, the 2-D prolate spheroidal wave functions (PSWF) on a disk [54], [55]. Nevertheless, our method has certain limitations. For example, if the noise distribution is not specified in advance, it is hard to estimate the noise covariance $\mathbb{E}\,\mathrm{diag}[A''(\omega)]$, which will affect the homogenization step of the proposed method. In addition, the optimal shrinkage function for the covariance estimation depends on how we measure the error, for example using Frobenius norm, operator norm, or other criteria. The corresponding theoretical guarantee in the non-Gaussian case is still an open problem. Thus, the statistical optimality is beyond the scope of this paper. For future work, we will study how sensitive the proposed method is to the misspecification of the noise distribution and the statistical optimality of our procedure in the non-Gaussian case.

Acknowledgments

The authors would like to thank associate editor and the anonymous referees for their valuable comments. They would also like to thank E. Dobriban and B. Landa for discussions.

REFERENCES

[1] V. Favre-Nicolin, J. Baruchel, H. Renevier, J. Eymery, and A. Borbély, "XTOP: High-resolution X-ray diffraction and imaging," *J. Appl. Crystallogr.*, vol. 48, no. 3, pp. 620–620, Jun. 2015.

[2] F. R. N. C. Maia and J. Hajdu, "The trickle before the torrent—Diffraction data from X-ray lasers," *Sci. Data*, vol. 3, no. 1, Dec. 2016, Art. no. 160059.

[3] S. H. W. Scheres *et al.*, "Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization," *Nature Methods*, vol. 4, no. 1, pp. 27–29, Jan. 2007.

[4] N.-T.-D. Loh and V. Elser, "Reconstruction algorithm for single-particle diffraction imaging experiments," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 2, Aug. 2009, Art. no. 026705.

[5] Z. Kam, "Determination of macromolecular structure in solution by spatial correlation of scattering fluctuations," *Macromolecules*, vol. 10, no. 5, pp. 927–934, Sep. 1977.

[6] D. K. Saldin, V. L. Shneerson, R. Fung, and A. Ourmazd, "Structure of isolated biomolecules obtained from ultrashort X-ray pulses: Exploiting the symmetry of random orientations," *J. Phys., Condens. Matter*, vol. 21, no. 13, Apr. 2009, Art. no. 134014.

[7] D. Starodub *et al.*, "Single-particle structure determination by correlations of snapshot X-ray diffraction patterns," *Nature Commun.*, vol. 3, no. 1, p. 1276, Jan. 2012.

[8] K. Pande, P. Schwander, M. Schmidt, and D. K. Saldin, "Deducing fast electron density changes in randomly orientated uncrystallized biomolecules in a pump–probe experiment," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 369, no. 1647, Jul. 2014, Art. no. 20130332.

[9] R. P. Kurta *et al.*, "Correlations in scattered X-ray laser pulses reveal nanoscale structural features of viruses," *Phys. Rev. Lett.*, vol. 119, no. 15, Oct. 2017, Art. no. 158102.

[10] J. J. Donatelli, J. A. Sethian, and P. H. Zwart, "Reconstruction from limited single-particle diffraction data via simultaneous determination of state, orientation, intensity, and phase," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 28, pp. 7222–7227, Jul. 2017.

[11] B. von Ardenne, M. Mechelke, and H. Grubmüller, "Structure determination from single molecule X-ray scattering with three photons per image," *Nature Commun.*, vol. 9, no. 1, p. 2375, Dec. 2018.

[12] K. Pande *et al.*, "Ab initio structure determination from experimental fluctuation X-ray scattering data," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 46, pp. 11772–11777, Nov. 2018. [Online]. Available: http://www.pnas.org/content/115/46/11772

[13] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2002.

[14] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York, NY, USA: Wiley, 2003.

[15] R. D. Nowak and R. G. Baraniuk, "Wavelet-domain filtering for photon imaging systems," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 666–678, May 1999.

[16] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 593–606, Mar. 2007.

[17] F. J. Anscombe, "The transformation of poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, nos. 3–4, pp. 246–254, 1948.

[18] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse Image Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[19] M. Collins, S. Dasgupta, and R. E. Schapire, "A generalization of principal components analysis to the exponential family," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2001, pp. 617–624.

[20] A. P. Singh and G. J. Gordon, "A unified view of matrix factorization models," in *Machine Learning and Knowledge Discovery in Databases*, W. Daelemans, B. Goethals, and K. Morik, Eds. Berlin, Germany: Springer, 2008, pp. 358–373.

[21] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *J. Math. Imag. Vis.*, vol. 48, no. 2, pp. 279–294, Feb. 2014.

[22] T. Furnival, R. K. Leary, and P. A. Midgley, "Denoising time-resolved microscopy image sequences with singular value thresholding," *Ultramicroscopy*, vol. 178, pp. 112–124, Jul. 2017.

[23] Y. Cao and Y. Xie, "Low-rank matrix recovery in Poisson noise," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2014, pp. 384–388.

[24] M. Sonnleitner, J. Jeffers, and S. M. Barnett, "Local retrodiction models for photon-noise-limited images," in *Proc. Opt., Photon. Digit. Technol. for Imag. Appl. IV*, Apr. 2016, p. 98.

[25] L. T. Liu, E. Dobriban, and A. Singer, "*e*PCA: High dimensional exponential family PCA," *Ann. Appl. Statist.*, vol. 12, no. 4, pp. 2121–2150, 2018.

[26] Z. Zhao, Y. Shkolnisky, and A. Singer, "Fast steerable principal component analysis," *IEEE Trans. Comput. Imag.*, vol. 2, no. 1, pp. 1–12, Mar. 2016.

[27] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[28] R. Hilai and J. Rubinstein, "Recognition of rotated images by invariant Karhunen–Loéve expansion," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 11, no. 5, pp. 1610–1618, 1994.

[29] P. Perona, "Deformable kernels for early vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 5, pp. 488–499, May 1995.

[30] M. Uenohara and T. Kanade, "Optimal approximation of uniformly rotated images: Relationship between Karhunen–Loéve expansion and discrete cosine transform," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 116–119, Jan. 1998.

[31] C. Ponce and A. Singer, "Computing steerable principal components of a large set of images and their rotations," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3051–3062, Nov. 2011.

[32] Z. Zhao and A. Singer, "Rotationally invariant image representation for viewing direction classification in cryo-EM," *J. Struct. Biol.*, vol. 186, no. 1, pp. 153–166, Apr. 2014.

[33] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[34] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[35] B. Landa and Y. Shkolnisky, "Approximation scheme for essentially bandlimited and space-concentrated functions on a disk," *Appl. Comput. Harmon. Anal.*, vol. 43, no. 3, pp. 381–403, Nov. 2017.

[36] A. Klug and R. A. Crowther, "Three-dimensional image reconstruction from the viewpoint of information theory," *Nature*, vol. 238, no. 5365, pp. 435–440, Aug. 1972.

[37] Q. Wang, O. Ronneberger, and H. Burkhardt, "Rotational invariance based on Fourier analysis in polar and spherical coordinates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1715–1722, Sep. 2009.

[38] L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast Fourier transform," *SIAM Rev.*, vol. 46, no. 3, pp. 443–454, Jan. 2004.

[39] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Probab.*, vol. 33, no. 5, pp. 1643–1697, 2005.

[40] J. Baik and J. W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *J. Multivariate Anal.*, vol. 97, no. 6, pp. 1382–1408, Jul. 2006.

[41] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Stat. Sinica*, vol. 17, no. 4, pp. 1617–1642, 2007.

[42] F. Benaych-Georges and R. R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Adv. Math.*, vol. 227, no. 1, pp. 494–521, May 2011.

[43] D. Donoho, M. Gavish, and I. Johnstone, "Optimal shrinkage of eigenvalues in the spiked covariance model," *Ann. Statist.*, vol. 46, no. 4, pp. 1742–1778, Aug. 2018.

[44] I. M. Johnstone and A. Y. Lu, "On consistency and sparsity for principal components analysis in high dimensions," *J. Amer. Stat. Assoc.*, vol. 104, no. 486, pp. 682–693, Jun. 2009.

[45] S. Jung and J. S. Marron, "PCA consistency in high dimension, low sample size context," *Ann. Statist.*, vol. 37, no. 6B, pp. 4104–4130, Dec. 2009.

[46] I. M. Johnstone and D. Paul, "PCA in high dimensions: An orientation," *Proc. IEEE*, vol. 106, no. 8, pp. 1277–1292, Aug. 2018.

[47] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 90*, vol. 2. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[48] M. F. Hantke, T. Ekeberg, and F. R. N. C. Maia, "Condor: A simulation tool for flash X-ray imaging," *J. Appl. Crystallogr.*, vol. 49, no. 4, pp. 1356–1362, Aug. 2016.

[49] P. Schwander, D. Giannakis, C. H. Yoon, and A. Ourmazd, "The symmetries of image formation by scattering. II. Applications," *Opt. Express*, vol. 20, no. 12, pp. 12827–12849, Jun. 2012.

[50] J. Landgrebe, W. Wurst, and G. Welzl, "Permutation-validated principal components analysis of microarray data," *Genome Biol.*, vol. 3, no. 4, p. research0019.1, 2002.

[51] E. Dobriban, "Permutation methods for factor analysis and PCA," 2017, *arXiv:1710.00479*. [Online]. Available: http://arxiv.org/abs/1710.00479

[52] E. Dobriban and A. B. Owen, "Deterministic parallel analysis: An improved method for selecting factors and principal components," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 81, no. 1, pp. 163–183, Feb. 2019.

[53] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[54] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—IV: Extensions to many dimensions; generalized prolate spheroidal functions," *Bell Syst. Tech. J.*, vol. 43, no. 6, pp. 3009–3057, 1964.

[55] B. Landa and Y. Shkolnisky, "Steerable principal components for space-frequency localized images," *SIAM J. Imag. Sci.*, vol. 10, no. 2, pp. 508–534, Jan. 2017.

Zhizhen Zhao (Member, IEEE) received the B.A. and M.Sc. degrees in physics from Trinity College, Cambridge University, in 2008, and the Ph.D. degree in physics from Princeton University in 2013. She joined the University of Illinois in 2016. From 2014 to 2016, she was a Courant Instructor with the Courant Institute of Mathematical Sciences, New York University. She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign. Her research interests include applied and computational harmonic analysis, signal processing, data analysis and the applications in structural biology and atmospheric, and oceanic sciences. She has been a recipient of Alfred P. Sloan Research Fellowship for the period 2020–2022.

Lydia T. Liu received the B.S.E. degree from Princeton University in 2017. She is currently pursuing the Ph.D. degree in computer science with the University of California, Berkeley. Her current research investigates the theoretical foundations of machine learning and algorithmic decision-making, with a focus on societal impact and welfare. Other interests include microeconomic theory and high-dimensional statistics. She was a recipient of an International Conference on Machine Learning Best Paper Award in 2018, the Microsoft Ada Lovelace Fellowship in 2019, and the Open Philanthropy AI Fellowship 2019.

Amit Singer received the B.Sc. degree in physics and mathematics and the Ph.D. degree in applied mathematics from Tel Aviv University, Israel, in 1997 and 2005, respectively. He is currently a Professor of mathematics and member of the Executive Committee of the Program in applied and computational mathematics (PACM) and of the Executive Committee for the Center for Statistics and Machine Learning (CSML), Princeton University. He joined Princeton as an Assistant Professor in 2008. From 2005 to 2008, he was a Gibbs Assistant Professor of applied mathematics with the Department of Mathematics, Yale University. He served in the Israeli Defense Forces from 1997 to 2003. His current research in applied mathematics focuses on theoretical and computational aspects of data science, and on developing computational methods for structural biology. His list of awards includes the Haim Nessyahu Prize for Best Ph.D. in Mathematics in Israel in 2007, the Presidential Early Career Award for Scientists and Engineers in 2010, the Alfred P. Sloan Research Fellowship in 2010, the Simons Investigator Award in 2012, the Moore Investigator in Data-Driven Discovery in 2014, the National Finalist for Blavatnik Awards for Young Scientists in 2016, and the Simons Math+X Investigator Award in 2017.