Poisson image denoising by piecewise principal component analysis and its application in single-particle X-ray diffraction imaging

Qiyu Jin, Osamu Miyashita, Florence Tama, Jie Yang, Slavica Jonic

IET Research Journals

This paper is a postprint of a paper submitted to and accepted for publication in IET Image Processing and is subject to Institution of Engineering and Technology Copyright. The copy of record is available at IET Digital Library. doi: 10.1049/iet-ipr.2018.5145

IET Journals
The Institution of Engineering and Technology

Poisson image denoising by piecewise principal component analysis and its application in single-particle X-ray diffraction imaging

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

Qiyu Jin[1*], *Osamu Miyashita*[2], *Florence Tama*[3,6,2], *Jie Yang*[4], *Slavica Jonic*[5*]

[1] *School of Mathematical Science Inner Mongolia University, Hohhot, 010021, China*
[2] *RIKEN Center for Computational Science, 6-7-1, Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan*
[3] *Department of Physics, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan*
[4] *Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, No. 800 Dongchuan Road, Minhang District, Shanghai 200240, China*
[5] *Sorbonne Université, UMR CNRS 7590, Muséum National d'Histoire Naturelle, IRD, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, IMPMC, 75005 Paris, France*
[6] *Institute of Transformative Bio-Molecules, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aich, 464-8602, Japan*
** E-mail: qyjin2015@aliyun.com*
** E-mail: slavica.jonic@upmc.fr*

Abstract: This article describes an improved method for Poisson image denoising that is based on a state-of-the-art Poisson denoising approach known as Non-Local Principal Component Analysis (NLPCA). The new method is referred to as PieceWise Principal Component Analysis (PWPCA). In PWPCA, the given image is first split into pieces, then NLPCA is run on each image piece, and finally, the entire image is reconstituted by a weighted combination of the NLPCA-processed image pieces. Using standard test images with Poisson noise, we show that PWPCA restores images more effectively than state-of-the-art Poisson denoising approaches. In addition, and to the best of our knowledge, we show a first application of such approaches to single-particle X-ray free-electron laser (XFEL) data. We show that the resolution of 3D reconstruction from XFEL diffraction images is improved when the data are preprocessed with PWPCA. XFELs are currently under rapid development to allow high-resolution biomolecular structure determination at near-physiological conditions. Data analysis methods developments follow these technological advances and are expected to have high impact in structural biology and drug design. This article contributes to these developments. As little experimental single-particle XFEL data is available still, the XFEL experiments shown here were performed with simulated data.

1 Introduction

Poisson noise (also called photon noise or shot noise [1]) is due to a discrete ("quantized") nature of the signal measured by a detector. For instance, light can be considered as a discrete flow of individual "packets" (photons) that arrive to the detector randomly and independently from each other. As a result, photon counting obeys a classical Poisson statistics, and the number of photons y measured by a given detector element over a time interval t is described by the following probability distribution:

$$\mathcal{P}(y|f) = \frac{(f)^y}{y!}e^{-f}, \tag{1}$$

where f is the expected number of photons in time interval t. This is the standard Poisson distribution with rate f.

For large numbers, the Poisson distribution approaches a normal distribution, meaning that shot noise is practically indistinguishable from a true Gaussian noise, except in the case of very few elementary events (photons, electrons, etc.) that are individually observed. Since the standard deviation of the shot noise is equal to the square root of the expected number of incident photons f, the signal-to-noise ratio (SNR) is given by:

$$\mathrm{SNR} = \frac{f}{\sqrt{f}} = \sqrt{f}. \tag{2}$$

This equation indicates that the SNR is large when f is large. However, in many applications (e.g., night vision, medical imaging, weather forecasting, astronomy, and spectral imaging), very few photons are collected by the detector, due to various constraints (weak light source, short exposure time, phototoxicity). This is referred to as photon-limited imaging. Thus, the limited number of available photons limits the SNR, due to the signal-dependent Poisson statistics.

An approach to improve the SNR is to first reduce the dependence of the noise on the signal by an image transformation (e.g., Variable-Stabilizing Transformation (VST), multiscale Variable-Stabilizing Transformation (MVST), conditional variance stabilization (CVS), or Fisz transform [2–19]) and, then, to use methods that assume signal-independent noise [20–23]. Other approaches to the SNR improvement have also been proposed, such as maximum likelihood estimation [24–32],Plug-and-Play scheme [33], Deep convolutional denoising [34] Kullback-Leibler divergence [35–37], optimal weights filtering [38] based on Non-Local Means approach [20], or Non-Local Principal Component Analysis (NLPCA) [39].

NLPCA combines Principal Component Analysis (PCA) and sparse Poisson intensity estimation methods in a nonlocal estimation framework and takes into account the heteroscedastic nature of observations. It clusters (K-Means clustering) a set of patches extracted from the entire image into 14 groups, each group containing similar patches. Then, it separates noise and signal in each group by reducing the dimension of the data in the group using PCA. More precisely, most of the low-frequency information will be contained

in the first few principal components and the other principal components, corresponding to the high-frequency information, will not be taken into account in the signal reconstruction. Therefore, NLPCA has lower efficiency with images rich in details [39].

Theoretical analyses and simulation results of Jin et al. [40, 41] suggest that larger sizes of the window from which the patches are extracted do not necessarily produce better denoising results. More precisely, there is an optimal value of the window size for the best denoising results [40, 41]. The probability of similarity of patches is higher and the similarity estimation is more accurate for the patches extracted from the same area than for the patches extracted from different areas of the same image. Building on the results of that work [40, 41], we here propose a denoising method that is based on combining the NLPCA results obtained on different areas of the same image. The method first splits the given image into large pieces. Then, the pieces are individually denoised using NLPCA. Finally, the denoised pieces are combined into a full-size image. The information dimension (complexity) in a piece (local area) of the image is much lower than the complexity of information in the entire image. Therefore, it is easier to separate noise from signal in the local areas than in the entire image.

The approach that we have developed for Poisson denoising using the principle of piecewise PCA is referred to as PieceWise PCA (PWPCA) and it is described in this article. We show that it provides better results than several state-of-the-art approaches, including NLPCA. PWPCA was evaluated using standard real-space test images affected by Poisson noise as well as using single-particle X-ray diffraction images, such as those that can be collected with an X-ray free-electron laser (XFEL). XFEL experiments are currently under rapid development to allow high-resolution determination of biomolecular structures under near-physiological hydration conditions [42–45]. The diffraction image collected by XFEL is, in a first approximation, the intensity (the square of the amplitude) of the 2D Fourier transform of the object's 2D real-space projection (or a central slice of the object's 3D Fourier-transform intensity, according to the central slice theorem), whereas the phase information cannot be collected experimentally. One diffraction image contains the diffraction data collected from one biomolecular particle at a random and unknown orientation in 3D space. Therefore, to compute a 3D reconstruction from the 2D diffraction data collected from many copies of the same particle, their orientations (the orientation of different particles hit by the XFEL beam) and phases should be determined computationally. The amplitude volume can be obtained from the square root of the volume assembled from the oriented diffraction images. The phase images can be retrieved from the amplitude images and assembled in 3D at the determined orientations or the 3D phase volume can be retrieved from the 3D amplitude volume, and the phase retrieval is usually done with the standard Fienup method or its variants [43, 46, 47]. The real-space 3D object can then be computed by inverse Fourier transform of the combined amplitude and phase volumes. Both orientation and phase determination are difficult tasks because the majority of non-zero diffraction-image pixels are concentrated in a very small central region of the diffraction image and the pixel intensities obey Poisson statistics. As very few XFEL experimental data sets had been obtained so far, we have performed experiments using simulated data, as commonly done when validating new methods for XFEL. To the best of our knowledge, diffraction data restoration using Poisson denoising methods has not been previously addressed.

The remaining part of the article is organized as follows. Basic concepts of NLPCA are provided in Section 2. PWPCA is described in Section 3. Results obtained with PWPCA using simulated noisy real-space images and using simulated XFEL data are presented and discussed in Sections 4 and 5, respectively. Conclusions are provided in Section 6.

2 Background: Non-Local Principal Component Analysis

In this section, we briefly remind the NLPCA approach [39] that we here propose to be used in a piecewise fashion over an image (Section 3). The set of image pixels is denoted by $\mathbf{I} = \{x = (x^{(1)}, x^{(2)}) | x^{(1)}, x^{(2)} = 1, 2, \cdots, \sqrt{M}\}$, where both $M > 0$ and $\sqrt{M}$ are integers. Let $\mathbf{y} = \{y(x) | x = (x^{(1)}, x^{(2)}) \in \mathbf{I}\}$ be the observed image obtained by an image acquisition device and each $y(x)$ is an independent random Poisson variable given by (Eq. 1), whose mean $f(x) > 0$ is the underlying intensity value to be estimated. Let $\mathbf{Y}$ be the $M \times N$ matrix of all denoising patches of size $\sqrt{N} \times \sqrt{N}$ pixels reordered in N-long vectors referred to as denoising vectors (one patch per image pixel for the image size of $\sqrt{M} \times \sqrt{M}$ pixels). $Y_{i,j}$ is the jth pixel in the ith patch of the noisy image. Let $\mathbf{F}$ be defined similarly to $\mathbf{Y}$ for the underlying intensity $\mathbf{f} = \{f(x) | x = (x^{(1)}, x^{(2)}) \in \mathbf{I}\}$.

The estimate of $\mathbf{F}$ is given by

$$\widehat{F}_{i,j} = \exp\left([\mathbf{U}^*\mathbf{V}^*]_{i,j}\right), i = 1, 2, \cdots, M, j = 1, 2, \cdots, N, \quad (3)$$

where

$$(\mathbf{U}^*, \mathbf{V}^*) = \arg \min_{(\mathbf{U},\mathbf{V})} L(\mathbf{U}, \mathbf{V}). \quad (4)$$

In Eq. 4, the matrices $\mathbf{U}/\mathbf{U}^*$ and $\mathbf{V}/\mathbf{V}^*$ have dimensions $M \times l$ and $l \times N$ and are referred to as matrix of coefficients and matrix of dictionary components (axes), respectively. L is the loss function that accounts for the Bregman divergence D_Φ to measure the proximity between two matrices X and Y of size $M \times N$:

$$D_\Phi(X||Y) = \Phi(Y) - \Phi(X) - Tr((\nabla\Phi(X))^\top (X - Y)), \quad (5)$$

where $\Phi(\theta) = < \exp(\theta) | \mathbf{1}_n > = \sum_{i=1}^{n} e^{\theta_i}$, $< \mathbf{a} | \mathbf{b} >$ is inner product of $\mathbf{a}$ and $\mathbf{b}$, $\theta = (\theta_1, \cdots, \theta_n)^\top$, exp is the component-wise exponential function: $\exp : \theta \mapsto (e^{\theta_1}, \cdots, e^{\theta_n})^\top$, $\mathbf{1}_n$ is the vector $(1, \cdots, 1)^\top \in \mathbf{R}^n$ and $\nabla\Phi(\theta)$ is the gradient of $\Phi(\theta)$. Moreover $\nabla\Phi(\theta) = \exp(\theta)$.

The use of Eq. 5 amounts to searching for $\mathbf{U}$ and $\mathbf{V}$ that minimize the following loss function:

$$L(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^{M} \sum_{j=1}^{N} \exp\left([\mathbf{UV}]_{i,j}\right) - Y_{i,j}[\mathbf{UV}]_{i,j}, \quad (6)$$

the rows of $\mathbf{V}$ are dictionary components and l ($l \ll M$) is the number of components. The purpose of the rank l is to regularize the solution.

More details on the NLPCA algorithm are provided in Appendices A and B, in the context of denoising image pieces (Section 3.2).

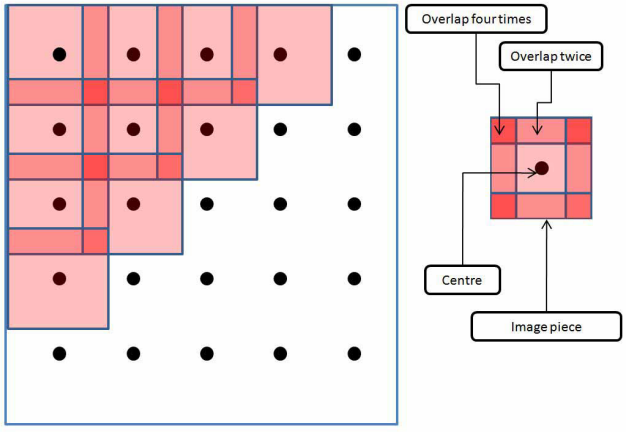

Fig. 1: Piecewise image denoising in PWPCA. Image pieces (right) overlap (left) to eliminate block type of artifacts.

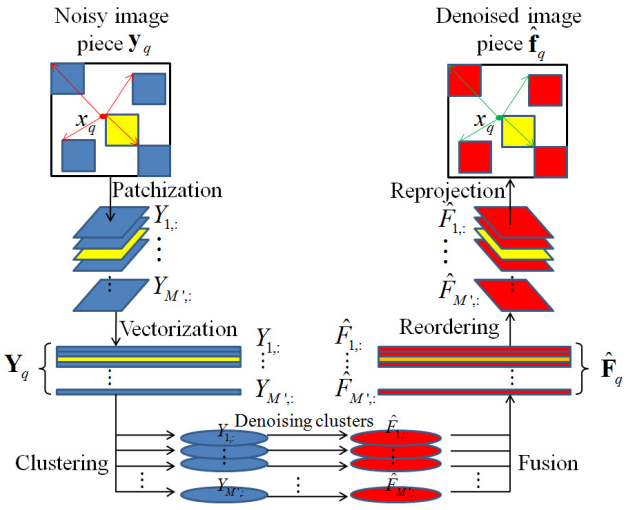

Fig. 2: Denoising of each image piece in PWPCA.

3 New Approach: Poisson Piecewise Principal Component Analysis

The PWPCA approach proposed here consists of three steps. The given image is first divided into a set of overlapping image pieces. Then, Poisson denoising of each individual piece is performed. Finally, all denoised pieces are combined, with some weights, into a full (original size) image. In the remaining part of this section, we describe in detail each of the three steps. The pseudocode of the algorithm is provided in Algorithm 1.

3.1 Extraction of image pieces

The first step is to split the given image $\mathbf{y}$ i.e. its pixels $\mathbf{I}$ into a set of pieces of size $\sqrt{M'} \times \sqrt{M'}$ pixels. The set of pixels in each image piece $\mathbf{y}_q$, $q = 1, 2, \cdots, Q$ is denoted by $\mathbf{B}_q = \{x = (x^{(1)}, x^{(2)}) \in \mathbf{I} | x^{(1)} \in [x_q^{(1)} - m, x_q^{(1)} + m], x^{(2)} \in [x_q^{(2)} - m, x_q^{(2)} + m]\}$, where m is the number of pixels covered by the image piece from its central pixel and $x_q = (x_q^{(1)}, x_q^{(2)})$ are the coordinates of the central pixel, meaning that $\sqrt{M'} = 2 \times m + 1$. This produces the total number of $\sqrt{Q} \times \sqrt{Q}$ pieces, where $\sqrt{Q} = \lceil (\sqrt{M} - \sqrt{M'})/(\frac{2}{3}\sqrt{M'}) \rceil$ and $\lceil t \rceil = \min\{n \in \mathbb{Z} | t \leq n\}$ is the top integral function whose value is the smallest integer greater than the independent variable or equal to it. Let us define T as $T = \sqrt{M} - \sqrt{M'}$. Then, the central pixel of each piece, x_q, is located in the set $\{(x^{(1)}, x^{(2)}) | x^{(1)}, x^{(1)} \in \{1 + m, \lceil T/\sqrt{Q} \rceil + 1 + m, \lceil 2T/\sqrt{Q} \rceil + 1 + m, T + 1 + m\}\}$ and the sets of pixels $\mathbf{B}_q$, $q = 1, 2, \cdots, Q$ cover $\mathbf{I}$ completely. As shown in the Fig. 1, the pieces overlap to avoid block type of artifacts and each pixel is covered at least once but not more than 4 times. Note here that each piece of the image (Fig. 1) will be further split into smaller pieces (as it will explained in Section 3.2) and that these sub-pieces will here be referred to as "denoising patches" (Fig. 2), as was the case in the original publication ([39]).

3.2 Denoising of each image piece

In this step, Poisson denoising of each image piece $\mathbf{y}_q = \{y(x) | x = (x^{(1)}, x^{(2)}) \in \mathbf{B}_q\}$ is preformed using the NLPCA approach (Section 2) that was adapted to process pieces of an image.

Let $\mathbf{Y}_q$ be the $M' \times N$ matrix of all (M') patches of size $\sqrt{N} \times \sqrt{N}$ pixels extracted from $\mathbf{y}_q$ (one patch per pixel of $\mathbf{y}_q$) and reordered in N-long vectors (referred to as denoising vectors).

Algorithm 1 Pseudocode of the Poisson Piecewise PCA (PWPCA) method proposed here

Require: Noisy image $\mathbf{y}$
Parameters: Number of pixels in square-size image pieces, M' (piece size: $\sqrt{M'} \times \sqrt{M'}$ pixels); length of denoising vectors containing vectorized denoising patches of square size, N (denoising patch size: $\sqrt{N} \times \sqrt{N}$ pixels); number of patch clusters for each image piece, K; number of dictionary components, l; maximum number of iterations, N_{iter}; stopping tolerance, ε_{stop}.
Ensure: Estimated image $\widehat{\mathbf{f}}$
Partitioning: divide image $\mathbf{y}$ into square-size pieces $\mathbf{y}_q$, $q = 1, 2, \cdots, Q$, each containing M' pixels.
for each $q = 1, 2, \cdots, Q$ **do**
    Patchization: create $M' \times N$ matrix $\mathbf{Y}_q$ of all denoising vectors (one patch per pixel of piece $\mathbf{y}_q$).
    Clustering: create K clusters of patches using K-means algorithm (see Algorithm 2 in Appendix B). The kth cluster (represented by a matrix $\mathbf{Y}_q^k$) has M_q^k elements.
    for each cluster k **do**
        Initialize $\mathbf{U}_q^0 = \text{randn}(M_q^k, l)$ and $\mathbf{V}_q^0 = \text{randn}(l, N)$
        while $t \leq N_{iter}$ and test $> \varepsilon_{stop}$ **do**
            for all $i \leq M_q^k$ **do**
                Update the ith row of U_q^t using Eq. 20 (Appendix A)
            end for
            for all $j \leq l$ **do**
                Update the jth column of V_q^t using Eq. 21 (Appendix A)
            end for
            t:=t+1
        end while
        $\widehat{\mathbf{F}}_q^k = \exp(\mathbf{U}_q^t \mathbf{V}_q^t)$
    end for
    Concatenation: fuse the denoising clusters in patches $\widehat{\mathbf{F}}_q^k$
    Reprojection: To get the estimate of the qth image piece ($\widehat{\mathbf{f}}_q$), average the pixel estimates in overlapping areas
end for
for each $x \in \mathbf{I}$ **do**
    Calculate $\widehat{f}(x)$ using Eq. (10)
end for

Let $\mathbf{F}_q$ be defined similarly to $\mathbf{Y}_q$ for the underlying intensity to be estimated $\mathbf{f}_q = \{f(x) | x = (x^{(1)}, x^{(2)}) \in \mathbf{B}_q\}$. We approximate each $\mathbf{F}_q$, $q = 1, 2, \cdots, Q$ as follows:

$$\mathbf{F}_q \approx \exp\left([\mathbf{U}_q \mathbf{V}_q]_{i,j}\right) \quad i \in \{1, 2, \cdots, M'\}, j \in \{1, 2, \cdots, N\}, \tag{7}$$

where $\mathbf{U}_q$ is the $M' \times l$ matrix of coefficients, $\mathbf{V}_q$ is the $l \times N$ matrix of dictionary components or axes, and $l \ll M'$.

Given $\mathbf{Y}_q \in \mathbb{R}^{M' \times N}$, with $\mathbf{Y}_{q,i,j}$ being the jth pixel in the ith patch, the underlying intensity $\mathbf{F}_q$ is estimated by minimizing the following loss function:

$$L(\mathbf{U}_q, \mathbf{V}_q) = \sum_{i=1}^{M'} \sum_{j=1}^{N} \exp\left([\mathbf{U}_q \mathbf{V}_q]_{i,j}\right) - Y_{i,j} \left[\mathbf{U}_q \mathbf{V}_q\right]_{i,j} \tag{8}$$

with respect to the matrices $\mathbf{U}_q$ and $\mathbf{V}_q$. The minimization of L is done using Newton's method (for more details, see Appendix A).

For each image piece $\mathbf{y}_q$, we separate its patches into different groups (clusters) according to their similarity and apply matrix factorization on each cluster, which allows representing the data within each cluster with a lower dimensional dictionary and results in more efficient denoising and higher SNR in the denoised image. The clustering of patches of each image piece is here done using K-means clustering algorithm (for more details, see Appendix B). The K-means algorithm is used in the original NLPCA approach as well, where it serves to cluster patches extracted from the entire

image [39]. However, denoising results may be poor when clustering patches of the entire noisy image because noise lowers the clustering accuracy and large cluster dimensions make the matrix factorization difficult. As a remedy to this problem, we here propose to perform a separate patch clustering for each piece extracted from the image.

The denoising of each image piece is vizualized in Fig. 2. Note here that the reprojection of patches in Fig. 2 means a reconstitution of the original, full-size piece of the image from the denoised patches. During the reprojection, the overlapping pixels are averaged.

3.3 Full-size image reconstitution

The final entire estimate $\widehat{\mathbf{f}}$ is computed by a weighted average of the piecewise estimates $\widehat{\mathbf{f}}_q$ obtained as explained in Section 3.2. Let pixel x_q from the set $\mathbf{B}_q$ be the center pixel of the denoised image piece $\widehat{\mathbf{f}}_q$ (whose size is $\sqrt{M'} \times \sqrt{M'}$ pixels, $\sqrt{M'} = 2 \times m + 1$). In the merging procedure, each pixel $x \in \mathbf{I}$ is first assigned the following weight [38]:

$$\kappa_{q,m}(x) = \begin{cases} \sum\limits_{s=\|x-x_q\|_\infty}^{m} \frac{1}{m(2s+1)^2} & \text{if } x \neq x_q \text{ and } x \in \mathbf{B}_q, \\ \sum\limits_{s=1}^{m} \frac{1}{m(2s+1)^2} & \text{if } x = x_q, \\ 0 & \text{if } x \in \mathbf{I} \setminus \mathbf{B}_q, \end{cases} \tag{9}$$

where $\| \bullet \|_\infty$ denotes supremum norm. Then, the full-size estimate $\widehat{\mathbf{f}}$ is obtained as follows ([48]):

$$\widehat{f}(x) = \frac{\sum\limits_{q=1}^{Q} \kappa_{q,m}(x) \widehat{f}_q(x)}{\sum\limits_{q=1}^{Q} \chi_{q,m}(x) \kappa_{q,m}(x)}, \quad \forall x \in \mathbf{I}, \tag{10}$$

where

$$\chi_{q,m}(x) = \begin{cases} 1, & x \in \mathbf{B}_q, \\ 0, & x \in \mathbf{I} \setminus \mathbf{B}_q. \end{cases}$$

For the purpose of illustration, we here give an example of the weighting kernel if the entire image is processed as a single piece (i.e. $Q = 1$) and the image size is 5×5 pixels ($m = 2$). The weighting kernel in this case ($\kappa_{1,2}$) is as follows:

$$\underbrace{\begin{pmatrix} \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} \\ \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} \\ \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} \\ \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} \\ \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} \end{pmatrix}}_{\frac{1}{2}} + \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{18} & \frac{1}{18} & \frac{1}{18} & 0 \\ 0 & \frac{1}{18} & \frac{1}{18} & \frac{1}{18} & 0 \\ 0 & \frac{1}{18} & \frac{1}{18} & \frac{1}{18} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}}_{\frac{1}{2}}$$

$$= \underbrace{\begin{pmatrix} \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} \\ \frac{1}{50} & \frac{34}{450} & \frac{34}{450} & \frac{34}{450} & \frac{1}{50} \\ \frac{1}{50} & \frac{34}{450} & \frac{34}{450} & \frac{34}{450} & \frac{1}{50} \\ \frac{1}{50} & \frac{34}{450} & \frac{34}{450} & \frac{34}{450} & \frac{1}{50} \\ \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} & \frac{1}{50} \end{pmatrix}}_{1}.$$

However, note that the number of image pieces is usually larger than 1 and m is usually much larger than 2 (see Eq 11).

Table 1 PWPCA parameters and their values in the experiments shown in this article

Parameter	Meaning	Value
M'	Image piece size	See Eq. (11)
N	Denoising patch size	400 (=20 × 20)
l	Number of dictionary components (axes)	4
K	Number of clusters	14
N_{iter}	Iteration limit	20
ε_{stop}	Stoping tolerance	10^{-1}

3.4 Parameters

Parameters of the PWPCA method are listed in Table 1, together with their values in the experiments shown here (Sections 4 and 5). The image piece size is defined as follows:

$$\sqrt{M'} = \begin{cases} 257 & Peak \leq 0.3, \\ 161 & 0.3 < Peak \leq 0.8, \\ 121 & 0.8 < Peak \leq 3, \\ 101 & Peak > 3, \end{cases} \tag{11}$$

which takes into account the size of the images used in our experiments (256×256 or 257×257 pixels). All other parameters are the same as in [39] (Table 1).

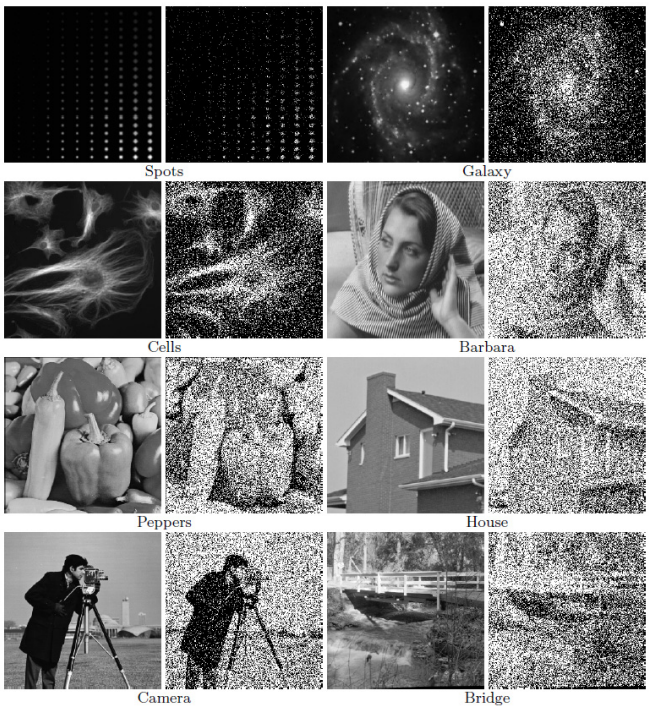

Fig. 3: No-noise images used in this article and synthetic Poisson-noise images with Peak = 2 ("Spots", "Galaxy", "Cells", "Barbara", "Peppers", "House", "Camera", "Bridge").

Table 2 Comparison between PWPCA and NLPCA. ΔPSNR is The difference PSNR value between them. The "Average" means the average of PSNR values for all images denoised with the same method.

Method	Peak	Spots	Galaxy	Cells	Barbara	Peppers	House	Camera	Bridge		Average
NLPCA	0.5	28.02	24.59	22.81	18.50	18.77	20.89	19.36	18.49		21.43
PWPCA	0.5	28.99	24.74	23.08	18.58	18.87	21.04	19.51	18.59		21.68
ΔPSNR	0.5	0.97	0.15	0.27	0.08	0.10	0.15	0.15	0.10		0.25
NLPCA	1	30.95	25.37	24.13	19.19	19.64	22.19	20.34	18.98		22.60
PWPCA	1	30.81	25.52	24.57	19.62	19.92	22.78	20.59	19.44		22.90
ΔPSNR	1	-0.14	0.15	0.44	0.43	0.28	0.59	0.25	0.46		0.30
NLPCA	2	34.07	25.73	25.57	19.67	20.63	23.51	20.76	19.52		23.68
PWPCA	2	34.82	26.16	25.92	20.63	21.05	24.23	21.47	20.03		24.29
ΔPSNR	2	0.75	0.43	0.35	0.96	0.42	0.72	0.71	0.51		0.61
NLPCA	4	35.01	26.75	26.34	20.36	20.97	24.81	21.15	20.25		24.46
PWPCA	4	36.24	27.60	27.01	21.78	22.41	26.21	22.09	20.95		25.54
ΔPSNR	4	1.23	0.85	0.67	1.42	1.44	1.40	0.94	0.70		1.08

Table 3 Comparison of PWPCA against several other methods using synthetic Poisson-noise images with different Peak values and the PSNR measure (in decibels). The best PSNR results are marked in bold. The "Average" means the average of PSNR values for all images denoised with the same method.

Method	Peak	Spots	Galaxy	Cells	Barbara	Peppers	House	Camera	Bridge		Average
PURE-LET	0.5	25.93	24.13	22.19	18.11	18.18	20.11	18.67	18.30		20.70
BM3D	0.5	23.82	**24.75**	22.66	18.04	18.48	20.36	18.66	18.36		20.64
SPDA	0.5	28.34	25.30	**23.48**	18.53	18.66	20.59	18.80	18.54		21.53
NLPCA	0.5	28.02	24.59	22.81	18.50	18.77	20.89	19.36	18.49		21.43
PWPCA	0.5	**28.99**	24.74	23.08	**18.58**	**18.87**	**21.04**	**19.51**	**18.59**		**21.68**
PURE-LET	1	27.34	24.98	23.45	18.87	19.33	21.37	19.83	19.06		21.78
BM3D	1	24.84	25.13	24.36	18.92	**20.01**	22.34	20.49	19.18		21.91
SPDA	1	**31.14**	**25.62**	**24.76**	19.39	19.99	22.63	20.17	19.21		22.86
NLPCA	1	30.95	25.37	24.13	19.19	19.64	22.19	20.34	18.98		22.60
PWPCA	1	30.81	25.52	24.57	**19.62**	19.92	**22.78**	**20.59**	**19.44**		**22.90**
PURE-LET	2	29.59	26.20	24.88	19.71	20.59	22.37	20.86	19.87		23.00
BM3D	2	27.59	26.24	25.68	20.49	**22.10**	23.95	**22.23**	**20.26**		23.56
SPDA	2	32.60	**26.47**	25.98	19.98	21.23	**25.05**	21.49	20.14		24.11
NLPCA	2	34.07	25.73	25.57	19.67	20.63	23.51	20.76	19.52		23.68
PWPCA	2	**34.82**	26.16	**25.92**	**20.63**	21.05	24.23	21.47	20.03		**24.29**
PURE-LET	4	31.60	27.39	26.00	20.46	21.86	23.90	22.12	20.96		24.29
BM3D	4	30.82	27.59	**27.06**	**22.24**	**24.19**	26.13	**24.06**	**21.62**		25.46
SPDA	4	34.25	27.22	26.71	20.52	22.09	26.02	21.87	20.55		24.90
NLPCA	4	35.01	26.75	26.34	20.36	20.97	24.81	21.15	20.25		24.46
PWPCA	4	**36.24**	**27.60**	27.01	21.78	22.41	**26.21**	22.09	20.95		**25.54**

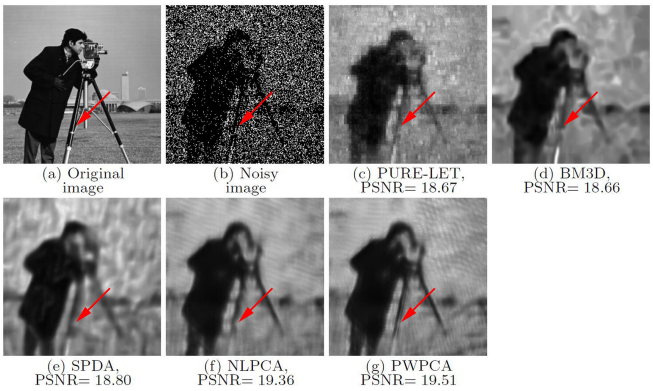

Fig. 4: Denoising of Camera (Peak=0.5) using different methods, with the resulting PSNR indicated in decibels.

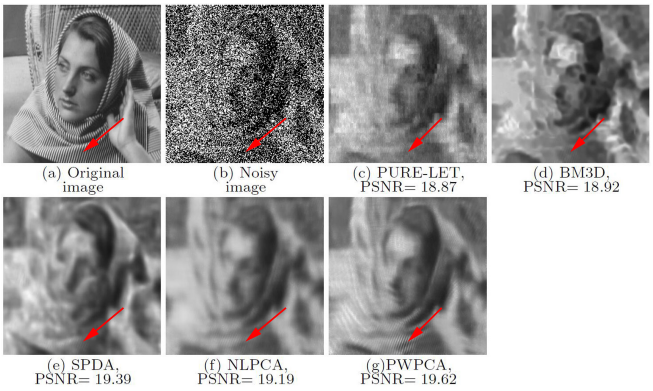

Fig. 5: Denoising of Barbara (Peak=1) using different methods, with the resulting PSNR indicated in decibels.

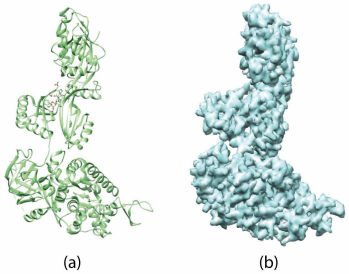

Fig. 6: Structure used to simulate XFEL diffraction data and the corresponding density map. (a) Atomic-resolution structure of yeast elongation factor 2 obtained by X-ray crystallography (from Protein Data Bank; PDB: 1N0U). (b) Density map computed from the structure shown in (a), used for computing the Fourier shell correlation curves shown in Fig. 9 and Fig. 12.

3.5 Computation complexity

The complexity of NLPCA is $O(MlN^{\frac{3}{2}})$ (cf. [32]). As explained in section 3.1, the method proposed here splits the given image into pieces so that each pixel is covered not more than 4 times. Thus, the complexity of the proposed method is not more than $O(4MlN^{\frac{3}{2}})$, meaning that the proposed method has the same order of computational cost as NLPCA.

4 Experiments with standard test images

In this section, we show results of experiments conducted with the new method (PWPCA) using standard test images with Poisson noise. Also, we compare the results of PWPCA with those of several state-of-the-art denoising methods, namely PURE-LET [10], BM3D [6, 49], NLPCA [39, 50], and SPDA [32, 51].

The noise-free standard test images (image size: 256×256 pixels) were graciously obtained from Salmon et al. [39] and Giryes and Elad [32]. A noise-free image f with the Peak value $Peak$ was obtained as follows:

$$f(x) = Peak \times S(x)/\max\{S(x)|x \in \mathbf{I}\},$$

where S is a standard test image. The image with Poisson noise (y) corresponding to image f was generated using MATLAB. The standard test images with $Peak = 2$ and Poisson noise are presented in Fig. 3. When the image intensity is high, the Poisson distribution approaches a normal distribution. In this section, we show the results obtained with images of low intensities, characterized by Peak values of 0.5, 1, 2 and 4.

The denoising results obtained with different methods were evaluated using the so-called Peak Signal-to-Noise Ratio (PSNR) [6], expressed in decibels (dB) as follows:

$$\text{PSNR} = 10\log_{10}\frac{Peak^2}{\text{MSE}}, \tag{12}$$

$$\text{MSE} = \frac{1}{\text{card}(\mathbf{I})}\sum_{x\in\mathbf{I}}(f(x) - \widehat{f}(x))^2, \tag{13}$$

where $\text{card}(\mathbf{I})$ is the number of elements of the set $\mathbf{I}$, f is the noise-free image and $\widehat{f}$ is the denoised image.

The PWPCA and NLPCA denoising results for images with these Peak values, in terms of PSNR, and the difference between the PSNR values obtained with these two methods are shown in Table 2. The PSNR results of all tested methods are listed in Table 3. The proposed PWPCA method achieves $0.25db - 1.08db$ improvement over the NLPCA method, on average, for all four noise levels and the PSNR difference increases significantly as the Peak value increases (Table 2). Also, the proposed PWPCA method achieves the highest average PSNR value in all four noise cases (Table 3). In Fig. 4 and Fig. 5, we compare the visual quality of images denoised by different methods. Fig. 4 shows that PWPCA outperforms the other methods regarding the restoration of image details from the noisy observation such as in the tripod leg area of ąřCameraąś image, as SPDA and NLPCA result in oversmoothing and PURE-LET and BM3D produce more artifacts (the area indicated by arrows in Fig. 4). Furthermore, Fig. 5 shows that PWPCA produces fewer artifacts and preserves the texture better than other methods (e.g., in the area indicated by arrows in Fig. 5). In summary, PWPCA has strong denoising capability and produces better quantitative (PSNR) and visual results than other tested methods.

5 Experiments with simulated XFEL diffraction data

In this section, we show an application of Poisson denoising methods in 3D reconstruction of biomolecular structures from single-particle XFEL diffraction data. As very few single-particle XFEL experimental data were obtained so far, we here used simulated data, which is common in evaluating new methods for XFEL.

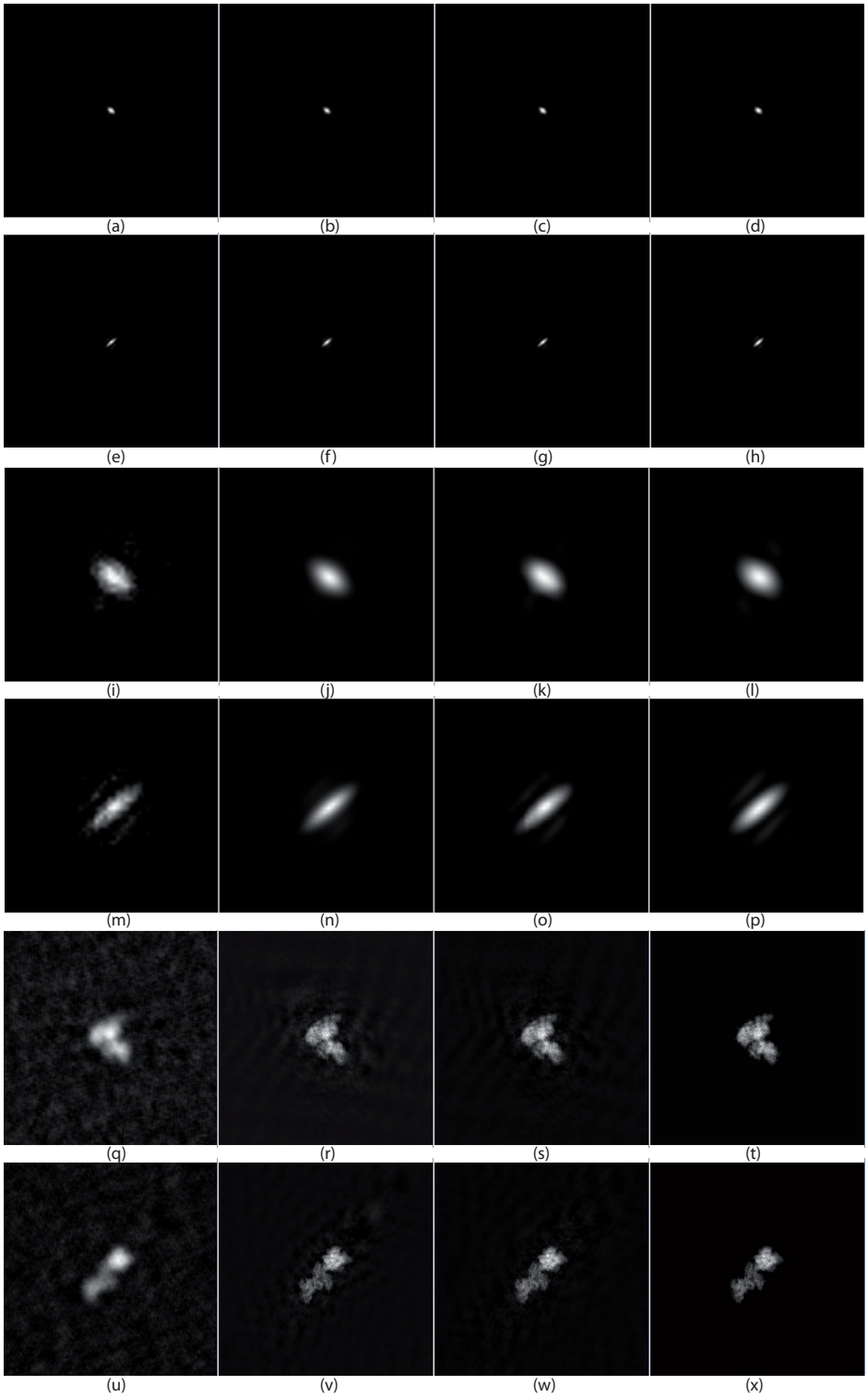

Fig. 7: Filtering of diffraction images simulated for the beam intensity of 10^{14} photons/pulse/μm^2 and the influence of filtering of diffraction images on real-space projection images. (a,e) Two examples of simulated diffraction images corresponding to two different beam orientations (257×257 pixels). (b,f) Filtering of images shown in (a,e) using PWPCA, respectively. (c,g) Filtering of images shown in (a,e) using BM3D, respectively. (d,h) Ground-truth diffraction images related to the simulated diffraction images shown in (a,e), respectively (Fourier intensity data representing the diffraction data as they would be obtained in ideal conditions that are unachievable in practice). (i-p) Central region (50×50 pixels) of the diffraction images shown in (a-h), respectively. (q-x) Real-space images obtained by combining the square root of the data shown in (a-h) and the ground-truth phases (not shown), respectively.

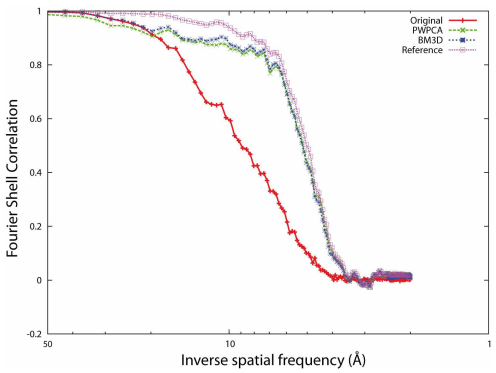

Fig. 8: Fourier shell correlation between the density map of the structure used to simulate XFEL diffraction data (Fig. 7b) and 3D reconstructions from four data sets for the beam intensity of 10^{14} photons/pulse/μm^2 (original, processed by PWPCA, processed by BM3D, and ground-truth sets, with ground-truth phases and orientations) (Fig. 10a-d).

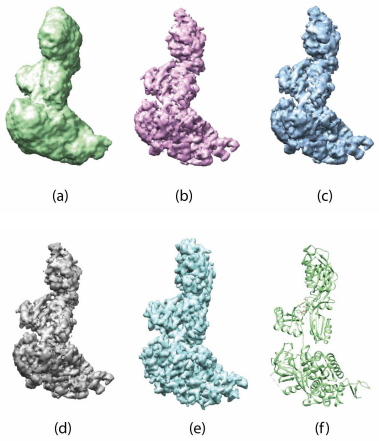

Fig. 9: 3D reconstructions in the case of beam intensity of 10^{14} photons/pulse/μm^2. (a-d) 3D reconstructions from the following four data sets: the original data set (a), the set processed by PWPCA (b), the set processed by BM3D (c), and the ground-truth data set (d) (ground-truth phases and orientations were used in each set). (e-f) Ground-truth atomic structure (f) and the corresponding density map (e), which are also shown in Fig. 7b,a, respectively.

XFEL diffraction images were simulated using an atomic-resolution structure of yeast elongation factor 2 obtained by X-ray crystallography, which is available in the Protein Data Bank (PDB entry: 1N0U [52]) (Fig. 6a). The simulation was done using SPSIM [53], for two different XFEL beam intensities i.e. 10^{13} photons/pulse/μm^2 and 10^{14} photons/pulse/μm^2. We found that 10^{13} photons/pulse/μm^2 was amongst the lowest beam intensities that could be used with this structure and that the Peak value of the simulated diffraction images was larger than 3 in both cases of beam intensity. Though Table 3 was obtained with a different type of data, it shows that, for $Peak > 3$, BM3D is the next best performing method after PWPCA. Thus, the results of processing these data using PWPCA are here compared with those obtained using BM3D.

The simulated data contain 642 XFEL diffraction images (257×257 pixels) for each beam intensity, corresponding to 642 different XFEL beam directions (i.e. to 642 projection directions) uniformly distributed over a sphere with an angular sampling rate of 8°. In addition, we computed the amplitudes and the phases of the 2D Fourier transforms of the object's 2D projections (or, according to the central slice theorem, central slices of the object's 3D Fourier transform at the same orientations as those of the simulated diffraction images). Two examples of diffraction images (corresponding to two different beam orientations) for the beam intensity of 10^{14} photons/pulse/μm^2 are provided in Fig. 7a,e and the same two examples for the beam intensity of 10^{13} photons/pulse/μm^2 are provided in Fig. 10a,e. One can notice a concentration of non-zero pixels in a small central region of the image (50×50 pixels) and a smaller number of non-zero pixels in the case of lower beam intensity (Figs. 7i,m and 10i,m). These diffraction images can be compared with the Fourier intensity data (Figs. 7d,h and 10d,h), whose central regions are shown in Figs. 7l,p and 10l,p. The Fourier intensity data represent expected mean value (the diffraction data as they would be obtained in ideal conditions that are unachievable in practice) and will be here referred to as ground-truth diffraction data. Instead of showing the phase that corresponds to a simulated diffraction image (here referred to as ground-truth phase), we here show the corresponding real-space image. The real-space images obtained by combining the square root of diffraction images and the phase images for the two example beam orientations and intensities are shown in Figs. 7q,u and 10q,u while those obtained by combining the Fourier amplitude and phase images (here referred to as ground-truth real-space images) are shown in Figs. 7t,x and 10t,x.

The PWPCA data processing was performed using the parameters provided in Table 1 (with $\sqrt{M'} = 101$, which corresponds to $Peak > 3$ according to Eq. 11) while the BM3D data processing was preformed using the same parameters as in [6]. The results of PWPCA for the two example beam orientations and intensities are shown in Figs. 7b,f and 10b,f (the central regions are shown in Figs. 7j,n and 10j,n) while those of BM3D are shown in Figs. 7c,g and 10c,g (the central regions are shown in Figs. 7k,o and 10k,o). One can note that both methods enhance the given diffraction patterns and that PWPCA results in more non-zero pixels recovered than BM3D, which is particularly visible in the case of lower beam intensity (Fig. 10). The real-space images obtained by combining the results of PWPCA (or BM3D) and the ground-truth phases are more similar to the ground-truth real-space images in the case of higher beam intensity (Figs. 7r,v and 7t,x) than in the case of lower beam intensity (Figs. 10r,v and 10t,x). In both intensity cases, the 3D structure is globally well recovered, though its resolution is higher in the case of higher beam intensity (Figs. 8-9) than in the case of lower beam intensity (Figs. 11-12). Indeed, the Fourier shell correlation (FSC) between the recovered density map of the ground-truth structure (Fig. 6b; the density map was obtained using the method introduced in [54]) and the 3D reconstructions (reconstruction in Fourier space using Kaiser-Bessel interpolation) from four data sets for each of the two beam intensities (original, processed by PWPCA, processed by BM3D, and ground-truth sets, with ground-truth phases and ground-truth orientations; Figs. 9 and 12) shows that data processing using any of the two methods (PWPCA or BM3D) improves the resolution of the 3D reconstruction (the spatial frequencies at which the FSC falls below 0.5 are higher for the data processed by PWPCA and BM3D than for the original data) and that this resolution improvement is slightly worse in the case of lower beam intensity (Fig. 11) than in the case of higher beam intensity (Fig. 8). Interestingly enough, PWPCA performs slightly better than BM3D regarding the resolution improvement in the case of lower beam intensity (the FSC falls below 0.5 at a higher spatial frequency when using PWPCA than when using BM3D; Fig. 11).

The 3D reconstruction step requires that phases and orientations of all diffraction images are known (determined by some means prior to 3D reconstruction). As the tasks of orientation determination and phase retrieval are not trivial and prone to errors, we used the ground-truth phases and orientations for the 3D reconstructions. Also, we used the same 3D reconstruction algorithm for all 3D reconstructions. Thus, differences between 3D reconstructions obtained from data processed with PWPCA and BM3D can fully be explained by differences in denoising results of the two methods.

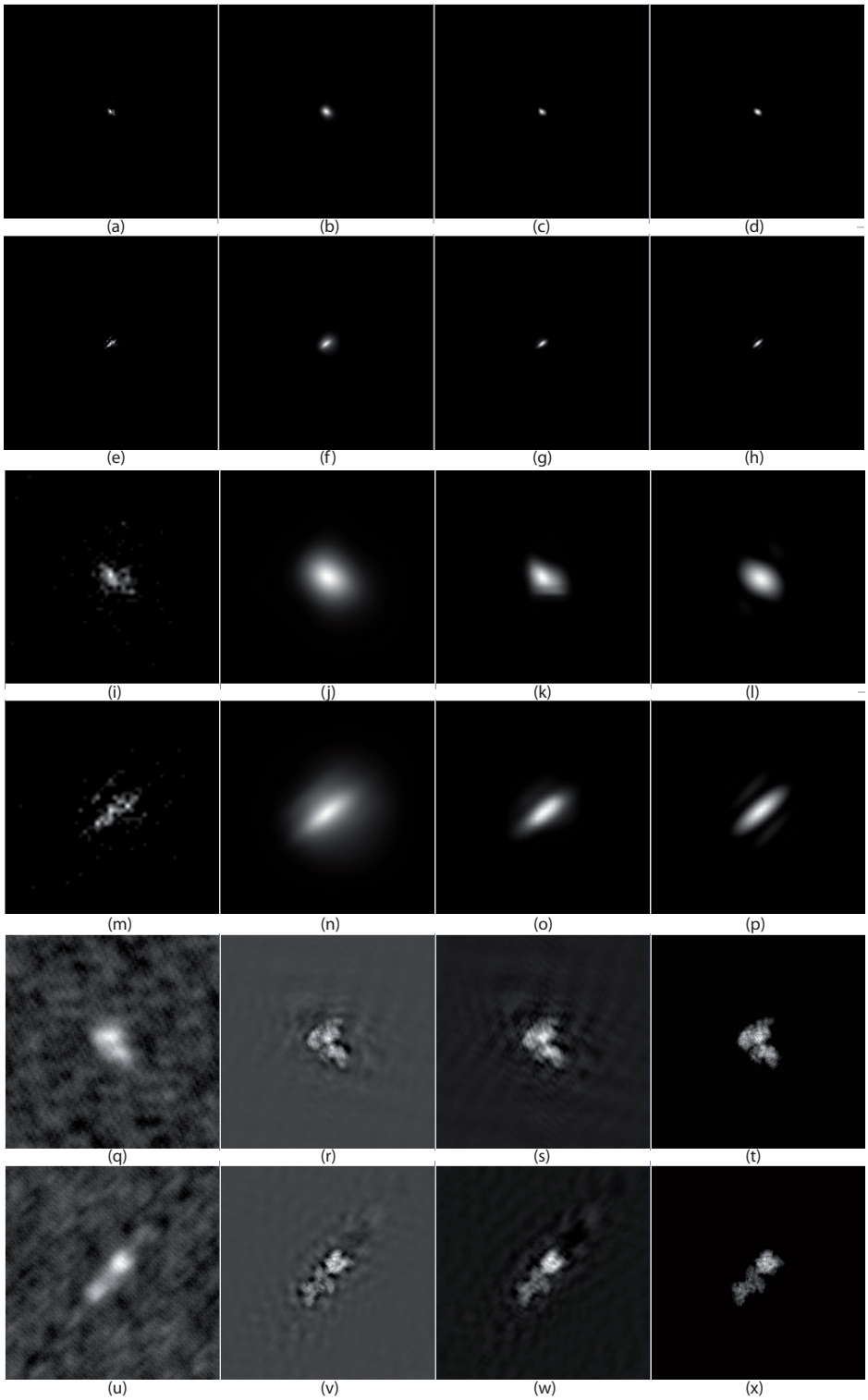

Fig. 10: Filtering of diffraction images simulated for the beam intensity of 10^{13} photons/pulse/μm^2 and the influence of filtering of diffraction images on real-space projection images. (a,e) Two examples of simulated diffraction images corresponding to two different beam orientations (257×257 pixels). (b,f) Filtering of images shown in (a,e) using PWPCA, respectively. (c,g) Filtering of images shown in (a,e) using BM3D, respectively. (d,h) Ground-truth diffraction images related to the simulated diffraction images shown in (a,e), respectively (Fourier intensity data representing the diffraction data as they would be obtained in ideal conditions that are unachievable in practice). (i-p) Central region (50×50 pixels) of the diffraction images shown in (a-h), respectively. (q-x) Real-space images obtained by combining the square root of the data shown in (a-h) and the ground-truth phases (not shown), respectively.

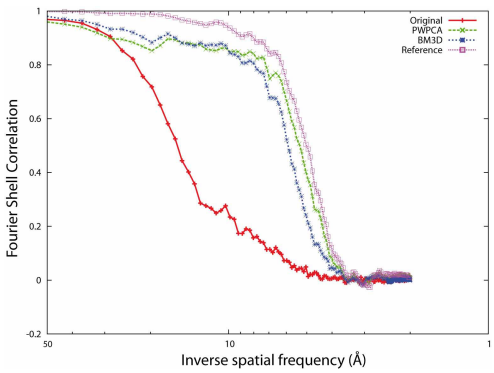

Fig. 11: Fourier shell correlation between the density map of the structure used to simulate XFEL diffraction data (Fig. 7b) and 3D reconstructions from four data sets for the beam intensity of 10^{13} photons/pulse/μm^2 (original, processed by PWPCA, processed by BM3D, and ground-truth sets, with ground-truth phases and orientations) (Fig. 13a-d).

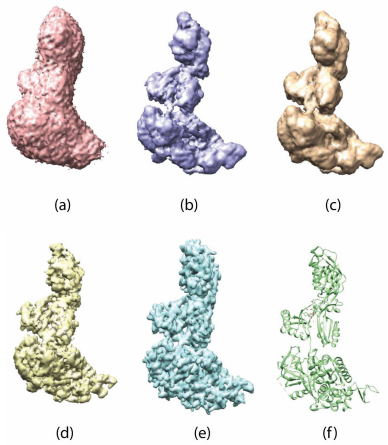

Fig. 12: 3D reconstructions in the case of beam intensity of 10^{13} photons/pulse/μm^2. (a-d) 3D reconstructions from the following four data sets: the original data set (a), the set processed by PWPCA (b), the set processed by BM3D (c), and the ground-truth data set (d) (ground-truth phases and orientations were used in each set). (e-f) Ground-truth atomic structure (f) and the corresponding density map (e), which are also shown in Fig. 7b,a, respectively.

6 Conclusions

In this article, we presented a new method for reducing Poisson noise from images and compared it with several state-of-the-art methods. We showed that the proposed method, referred to as Piecewise Principal Component Analysis (PWPCA), has the best overall performance with respect to other methods when denoising images with Peak $= 0.5$, 1, 2 and 4. The overall performance of a method for a particular Peak value was evaluated by averaging the Peak Signal-to-Noise Ratio (PSNR) values that the given method achieved on different images with the same, given value of Peak. The visual quality of recovered images was consistent with the obtained PSNR values.

Additionally, we showed that PWPCA can be used to preprocess single-particle XFEL diffraction data of biological macromolecules in order to improve the resolution of 3D reconstruction of the macromolecular structure from such data. In this context, PWPCA was compared with BM3D. The two methods provided similar results in the case of high XFEL beam intensities, such as 10^{14} photons/pulse/μm^2. Furthermore, PWPCA provided slightly better results than BM3D in the case of low beam intensities, more difficult to work with, such as 10^{13} photons/pulse/μm^2. These experiments were performed using simulated data. For our future work, we plan to extend the method to applications with experimental XFEL data, which should allow gaining new biological insights.

MATLAB implementation of PWPCA is available from the authors upon request.

7 Acknowledgements

This work was supported by a grant from the National Natural Science Foundation of China (No. 61661039,61661038,61661040),Natural Science Fund of Inner Mongolia Autonomous Region (No. 2016MS0107) and Scientific Research Projection of Higher Schools of Inner Mongolia (No. NJZY16017); and FOCUS for Establishing Supercomputing Center of Excellence (Japan). The authors would also like to thank Joseph Salmon, Zachary Harmany, Charles-Alban Deledalle, and Rebecca Willett for providing their NLPCA algorithm and test images, and Raja Giryes and Michael Elad for providing their SPDA algorithm and test images.

APPENDIX A. Newton Method for Minimizing L

The estimate of $\mathbf{F}_q$ is given by

$$\widehat{\mathbf{F}}_q = \exp\left(\mathbf{U}_q^* \mathbf{V}_q^*\right), \tag{14}$$

where the elements of the matrix $\widehat{\mathbf{F}}_q$ are $\exp\left([\mathbf{U}_q^* \mathbf{V}_q^*]_{i,j}\right), i = 1, 2, \cdots, M', j = 1, 2, \cdots, N$. The matrices $\mathbf{U}_q^* \in \mathbb{R}^{M' \times l}$ and $\mathbf{V}_q^* \in \mathbb{R}^{l \times N}$ are found by minimizing, with respect to $\mathbf{U}_q$ and $\mathbf{V}_q$, the loss function L given by

$$L(\mathbf{U}_q, \mathbf{V}_q) = \sum_{i=1}^{M'} \sum_{j=1}^{N} \exp\left([\mathbf{U}_q \mathbf{V}_q]_{i,j}\right) - Y_{q,i,j} [\mathbf{U}_q \mathbf{V}_q]_{i,j}. \tag{15}$$

We use Newton method [55, 56] to minimize L. As L is not convex jointly in $\mathbf{U}_q$ and $\mathbf{V}_q$, the minimization is done by solving two separate partial optimization problems (the problem is convex for one variable fixed and the other one varying).

The gradient of $L(\mathbf{U}_q, \mathbf{V}_q)$ with respect to the matrix $\mathbf{U}_q$ is given by

$$\nabla_{\mathbf{U}_q} L\left(\mathbf{U}_q, \mathbf{V}_q\right) = \left(\exp\left(\mathbf{U}_q \mathbf{V}_q\right) - \mathbf{Y}_q\right) \mathbf{V}_q^\top \tag{16}$$

and the gradient of $L(\mathbf{U}_q, \mathbf{V}_q)$ with respect to the matrix $\mathbf{V}_q$ is given by

$$\nabla_{\mathbf{V}_q} L\left(\mathbf{U}_q, \mathbf{V}_q\right) = \mathbf{U}_q^\top \left(\exp\left(\mathbf{U}_q \mathbf{V}_q\right) - \mathbf{Y}_q\right). \tag{17}$$

Using element-wise representation, this is equivalent to

$$\frac{\partial L\left(\mathbf{U}_q, \mathbf{V}_q\right)}{\partial U_{q,a,b}} = \sum_{j=1}^{N} \left(\exp\left([\mathbf{U}_q \mathbf{V}_q]_{a,j}\right) - Y_{q,a,j}\right) V_{q,b,j} \tag{18}$$

and

$$\frac{\partial L\left(\mathbf{U}_q, \mathbf{V}_q\right)}{\partial V_{q,a,b}} = \sum_{i=1}^{M'} U_{q,i,a} \left(\exp\left([\mathbf{U}_q \mathbf{V}_q]_{q,i,b}\right) - Y_{q,i,b}\right). \tag{19}$$

In element-wise form, the Hessian matrices $H_{\mathbf{U}_q} = \nabla^2_{\mathbf{U}_q} L(\mathbf{U}_q, \mathbf{V}_q)$ ($M'l \times M'l$ diagonal matrix) and $H_{\mathbf{V}_q} = \nabla^2_{\mathbf{V}_q} L(\mathbf{U}_q, \mathbf{V}_q)$ ($lN \times$

lN diagonal matrix) are respectively written as follows:

$$\begin{aligned}&\frac{\partial^2 L(\mathbf{U}_q,\mathbf{V}_q)}{\partial U_{q,a,b}\partial U_{q,c,d}}\\ =\;&\begin{cases}\sum_{j=1}^{N}\exp\left([\mathbf{U}_q\mathbf{V}_q]_{a,j}\right)V_{q,b,j}^2, & \text{if } (a,b)=(c,d),\\ 0, & \text{if } (a,b)\neq(c,d);\end{cases}\end{aligned}$$

and

$$\begin{aligned}&\frac{\partial^2 L(\mathbf{U}_q,\mathbf{V}_q)}{\partial V_{q,a,b}\partial V_{q,c,d}}\\ =\;&\begin{cases}\sum_{i=1}^{M'}U_{q,i,a}^2\exp\left([\mathbf{U}_q\mathbf{V}_q]_{i,b}\right), & \text{if } (a,b)=(c,d),\\ 0, & \text{if } (a,b)\neq(c,d).\end{cases}\end{aligned}$$

To facilitate the calculation, let us introduce the function

$$\begin{array}{rcl}\mathrm{Vect_C}: & & \\ \mathbb{R}^{l\times N} & \mapsto & \mathbb{R}^{lN\times 1}\\ \mathbf{V}_q=\left(\mathbf{V}_{q,1,:},\cdots,\mathbf{V}_{q,N,:}\right) & \mapsto & \left(\mathbf{V}_{q,1,:}^\top,\cdots,\mathbf{V}_{q,N,:}^\top\right)^\top\end{array}$$

that transforms a matrix into a single column (concatenates the columns) and the function $\mathrm{Vect_R}$

$$\begin{array}{rcl}\mathrm{Vect_R}: & & \\ \mathbb{R}^{M'\times l} & \mapsto & \mathbb{R}^{M'l\times 1}\\ \mathbf{U}_q=\left(\mathbf{U}_{q,:,1}^\top,\cdots,\mathbf{U}_{q,:,M'}^\top\right) & \mapsto & \left(\mathbf{U}_{q,:,1},\cdots,\mathbf{U}_{q,:,M'}\right)\end{array}$$

that transforms a matrix into a single row (concatenates the rows). The Newton updating steps for $\mathbf{U}_q$ and $\mathbf{V}_q$ can then be written as follows:

$$\begin{aligned}&\mathrm{Vect_R}\left(\mathbf{U}_q^{(t+1)}\right)\\ &=\mathrm{Vect_R}(\mathbf{U}_q^{(t)})-\mathrm{Vest_R}(\nabla_{\mathbf{U}_q}L(\mathbf{U}_q^{(t)},\mathbf{V}_q^{(t)}))\mathbf{H}_{\mathbf{U}_q^{(t)}}^{-1},\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Vect_C}(\mathbf{V}_q^{(t+1)})\\ &=\mathrm{Vect_C}(\mathbf{V}_q^{(t)})-\mathbf{H}_{\mathbf{V}_q^{(t)}}^{-1}\mathrm{Vest_C}(\nabla_{\mathbf{V}_q}L(\mathbf{U}_q^{(t)},\mathbf{V}_q^{(t)})).\end{aligned}$$

Considering (18), the ith row of $\mathbf{U}_q^{(t+1)}$, denoted by $U_{q,i,:}^{(t+1)}$, is updated as follows:

$$\begin{aligned}&U_{q,i,:}^{(t+1)}=\\ &U_{q,i,:}^{(t)}-(\exp([\mathbf{U}_q^t\mathbf{V}_q^{(t)}]_{i,:})-Y_{q,i,:})\mathbf{V}_q^{(t)\top}(\mathbf{V}_q^{(t)}\mathbf{D}_i\mathbf{V}_q^{(t)\top})^{-1},\end{aligned}\tag{20}$$

where $\mathbf{D}_i$ is a diagonal matrix of size $N\times N$ i.e.

$$\mathbf{D}_i=\mathrm{diag}(\exp([\mathbf{U}_q^{(t)}\mathbf{V}_q^{(t)}]_{i,1}),\cdots,\exp([\mathbf{U}_q^{(t)}\mathbf{V}_q^{(t)}]_{i,N})).$$

The term $V_{q,:,j}^{(t+1)}$, the jth column of $\mathbf{V}_q^{(t+1)}$, is updated in the same way as follows:

$$\begin{aligned}&V_{q,:,j}^{(t+1)}=\\ &V_{q,:,j}^{(t)}-(\mathbf{U}_q^{(t+1)\top}\mathbf{E}_j\mathbf{U}_q^{(t+1)})^{-1}\mathbf{U}_q^{(t+1)\top}([\exp(\mathbf{U}_q^{(t+1)}\mathbf{V}_q^{(t)})]_{:,j}-Y_{:,j}),\end{aligned}\tag{21}$$

where $\mathbf{E}_j$ is a diagonal matrix of size $M'\times M'$ i.e.

$$\mathbf{E}_j=\mathrm{diag}(\exp([\mathbf{U}_q^{(t+1)}\mathbf{V}_q^{(t)}]_{1,j}),\cdots,\exp([\mathbf{U}_q^{(t+1)}\mathbf{V}_q^{(t)}]_{M',j})).$$

APPENDIX B. Patch clustering algorithm

The K-means clustering algorithm can be generalized using Bregman divergences (Eq. 5), which was introduced in [57] and refereed to as Bregman clustering. We have adapted this algorithm to our piecewise processing approach as shown in Algorithm 2.

Algorithm 2 Bregman clustering

Require: Data points: $\{Y_{i,:}\}_{i=1}^{M'}\in\mathbf{R}^N$, number of clusters: K, Bregman divergence: D_Φ given by (5)
Ensure: Clusters centers: $\{\mu_k\}_{k=1}^{K}$, partition associated: $\{C_k\}_{k=1}^{K}$

Initialize $\{\mu_k\}_{k=1}^{K}$ by randomly selecting K elements among $\{Y_{i,:}\}_{i=1}^{M'}$
repeat
  (The Assignment step: Cluster updates)
  Set $C_k:=\emptyset, 1\leq k\leq K$
  for each $i=1,2,\cdots,M'$ **do**
    $C_{k^*}:=C_{k^*}\cup\{Y_{i,:}\}$
    $k^*=\arg\min_{k'=1,\cdots,K}D_\Phi(Y_{i,:}||\mu_k')$
  end for
  (The Estimation step: Center updates)
  for each $k=1,2,\cdots,K$ **do**
    $\mu_k:=\frac{1}{\sharp C_k}\sum_{Y_{i,:}\in C_k}Y_{i,:}$
  end for
until convergence

8 References

1 W. Schottky, "Über spontane stromschwankungen in verschiedenen elektrizitätsleitern," *Annalen der Physik*, vol. 362, no. 23, pp. 541–567, 1918.
2 A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Denoising of multispectral images via nonlocal groupwise spectrum-pca," in *Conference on Colour in Graphics, Imaging, and Vision*, vol. 2010, pp. 261–266, Society for Imaging Science and Technology, 2010.
3 J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J. B. Sibarita, and J. Salamero, "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 2, pp. 442–454, 2010.
4 C. A. Deledalle, F. Tupin, and L. Denis, "Poisson nl means: Unsupervised non local means for poisson noise," in *Image processing (ICIP), 2010 17th IEEE international conference on*, pp. 801–804, IEEE, 2010.
5 M. Makitalo and A. Foi, "On the inversion of the Anscombe transformation in low-count Poisson image denoising," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2012.
6 M. Makitalo and A. Foi, "Optimal inversion of the anscombe transformation in low-count poisson image denoising," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 99–109, 2011.
7 J. Boulanger, J. B. Sibarita, C. Kervrann, and P. Bouthemy, "Non-parametric regression for patch-based fluorescence microscopy image sequence denoising," in *in Proc. of IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, ISBIï£¡ï£¡2008*, pp. 748–751, IEEE, 2008.
8 B. Zhang, J. M. Fadili, and J. L. Starck, "Wavelets, ridgelets, and curvelets for poisson noise removal," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1093–1108, 2008.
9 S. Lefkimmiatis, P. Maragos, and G. Papandreou, "Bayesian inference on multiscale models for poisson intensity estimation: Applications to photon-limited image denoising," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1724–1741, 2009.
10 F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of poisson-corrupted images," *Signal Process.*, vol. 90, no. 2, pp. 415–427, 2010.
11 F. J. Anscombe, "The transformation of poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.
12 A. A. Borovkov, "Estimates for the distribution of sums and maxima of sums of random variables without the cramer condition," *Siberian Mathematical Journal*, vol. 41, no. 5, pp. 811–848, 2000.
13 M. Fisz, "The limiting distribution of a function of two independent random variables and its statistical application," in *Colloquium Mathematicae*, vol. 3, pp. 138–146, Institute of Mathematics Polish Academy of Sciences, 1955.
14 P. Fryzlewicz and G. P. Nason, "A haar-fisz algorithm for poisson intensity estimation," *J. Comp. Graph. Stat.*, vol. 13, no. 3, pp. 621–638, 2004.
15 P. Fryzlewicz, V. Delouille, and G. P. Nason, "Goes-8 x-ray sensor variance stabilization using the multiscale data-driven haar–fisz transform," *J. Roy. Statist. Soc. ser. C*, vol. 56, no. 1, pp. 99–116, 2007.
16 M. Jansen, "Multiscale poisson data smoothing," *J. Roy. Statist. Soc. B*, vol. 68, no. 1, pp. 27–48, 2006.
17 P. R. Prucnal and B. E. Saleh, "Transformation of image-signal-dependent noise into image-signal-independent noise," *Optics letters*, vol. 6, no. 7, pp. 316–318,

1981.
18 A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *Image Processing, IEEE Transactions on*, vol. 17, no. 10, pp. 1737–1754, 2008.
19 L. Azzari and A. Foi, "Variance stabilization for noisy+ estimate combination in iterative poisson denoising," *IEEE signal processing letters*, vol. 23, no. 8, pp. 1086–1090, 2016.
20 A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.
21 C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866–2878, 2006.
22 D. K. Hammond and E. P. Simoncelli, "Image modeling and denoising with orientation-adapted gaussian scale mixtures," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2089–2101, 2008.
23 V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *Int. J. Comput. Vis.*, vol. 86, no. 1, pp. 1–32, 2010.
24 R. H. Chan and K. Chen, "Multilevel algorithm for a poisson noise removal model with total-variation regularization," *International Journal of Computer Mathematics*, vol. 84, no. 8, pp. 1183–1198, 2007.
25 R. Srivastava and S. Srivastava, "Restoration of poisson noise corrupted digital images with nonlinear pde based filters along with the choice of regularization parameter estimation," *Pattern Recognition Letters*, vol. 34, no. 10, pp. 1175–1185, 2013.
26 L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *Medical Imaging, IEEE Transactions on*, vol. 1, no. 2, pp. 113–122, 1982.
27 G. Van Kempen, H. Van der Voort, J. Bauman, and K. C. Strasters, "Comparing maximum likelihood estimation and constrained tikhonov-miller restoration," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 15, no. 1, pp. 76–83, 1996.
28 I. Moon and B. Javidi, "Three dimensional imaging and recognition using truncated photon counting model and parametric maximum likelihood estimator," *Optics express*, vol. 17, no. 18, pp. 15709–15715, 2009.
29 J. M. Bardsley and N. Laobeul, "Tikhonov regularized poisson likelihood estimation: theoretical justification and a computational method," *Inverse Problems in Science and Engineering*, vol. 16, no. 2, pp. 199–215, 2008.
30 M. B. Johnathan and G. John, "Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation," *Inverse Problems*, vol. 25, 2009.
31 T. Le, R. Chartrand, and T. J. Asaki, "A variational approach to reconstructing images corrupted by poisson noise," *Journal of Mathematical Imaging and Vision*, vol. 27, no. 3, pp. 257–263, 2007.
32 R. Giryes and M. Elad, "Sparsity-based poisson denoising with dictionary learning," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5057–5069, 2014.
33 A. Rond, R. Giryes, and M. Elad, "Poisson inverse problems by the plug-and-play scheme," *Journal of Visual Communication and Image Representation*, vol. 41, pp. 96–108, 2016.
34 T. Remez, O. Litany, R. Giryes, and A. M. Bronstein, "Deep convolutional denoising of low-light images," *arXiv preprint arXiv:1701.01687*, 2017.
35 C. Brune, A. Sawatzky, F. Wübbeling, T. Kösters, and M. Burger, "Forward-backward em-tv methods for inverse problems with poisson noise," *Preprint*, 2010.
36 Y. Wen, R. Chan, and T. Zeng, "Primal-dual algorithms for total variation based image restoration under poisson noise," *Sci. China Math.*, vol. 59, no. 1, pp. 141–160, 2016.
37 J. Li, Z. Shen, R. Yin, and X. Zhang, "A reweighted l2 method for image restoration with poisson and mixed poisson-gaussian noise," *Inverse Probl. Imaging (Springfield)*, vol. 9, no. 3, pp. 875–894, 2015.
38 Q. Y. Jin, I. Grama, and Q. S. Liu, "A new poisson noise filter based on weights optimization," *Journal of Scientific Computing*, vol. 58, no. 3, pp. 548–573, 2014.
39 J. Salmon, Z. Harmany, C. A. Deledalle, and R. Willett, "Poisson noise reduction with non-local pca," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 279–294, 2014.
40 Q. Jin, I. Grama, C. Kervrann, and Q. Liu, "Nonlocal means and optimal weights for noise removal," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1878–1920, 2017.
41 Q. Jin, I. Grama, and Q. Liu, "Convergence theorems for the non-local means filter," *arXiv:1211.6143*, 2012.
42 A. Hosseinizadeh, G. Mashayekhi, J. Copperman, P. Schwander, A. Dashti, R. Sepehr, R. Fung, M. Schmidt, C. H. Yoon, and B. G. Hogue, "Conformational landscape of a virus by single-particle x-ray scattering," *Nature Methods*, vol. 14, no. 9, p. 877, 2017.
43 J. Miao, K. O. Hodgson, and D. Sayre, "An approach to three-dimensional structures of biomolecules by using single-molecule diffraction images.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 12, p. 6641, 2001.
44 M. M. Seibert, T. Ekeberg, F. R. Maia, M. Svenda, J. Andreasson, O. Jönsson, D. Odić, B. Iwan, A. Rocker, D. Westphal, *et al.*, "Single mimivirus particles intercepted and imaged with an x-ray laser," *Nature*, vol. 470, no. 7332, pp. 78–81, 2011.
45 O. M. Yefanov and I. A. Vartanyants, "Orientation determination in single particle x-ray coherent diffraction imaging experiments," *Journal of Physics B Atomic Molecular & Optical Physics*, vol. 46, no. 46, pp. 8323–8331, 2013.
46 J. R. Fienup, "Reconstruction of an object from the modulus of its fourier transform," *Optics Letters*, vol. 3, no. 1, p. 27, 1978.
47 M. Tegze and G. Bortel, "Atomic structure of a single large biomolecule from diffraction patterns of random orientations.," *Journal of Structural Biology*, vol. 179, no. 1, pp. 41–45, 2012.
48 K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
49 M. Makitalo and A. Foi, "A closed-form approximation of the exact unbiased inverse of the anscombe variance-stabilizing transformation," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2697–2698, 2011.
50 J. Salmon, C. A. Deledalle, R. Willett, and Z. Harmany, " Poisson noise reduction with non-local PCA ," in *International Workshop on Local and Non-Local Approximation in Image Processing*, 2009.
51 R. Giryes and M. Elad, "Sparsity based poisson denoising," in *IEEE 27th Convention of Electrical Electronics Engineers in Israel (IEEEI)*,, pp. 1–5, 2012.
52 R. Jørgensen, P. A. Ortiz, A. Carr-Schmid, P. Nissen, T. G. Kinzy, and G. R. Andersen, " Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase," *Nature Structural & Molecular Biology*, vol. 10, pp. 379–385, 2003.
53 F. Maia, "Single particle diffraction simulator, SPSIM," 2008. http://xray.bmc.uu.se/ filipe/?q=hawk/spsim/.
54 C. Sorzano, J. Vargas, J. Otón, V. Abrishami, J. M. de la Rosa-Trevín, S. del Riego, A. Fernï£¡ï£¡ndez-Alderete, C. Martí£¡ï£¡nez-Rey, R. Marabini, and J. M. Carazo, " Fast and accurate conversion of atomic models into electron density maps," *AIMS Biophysics*, vol. 2, pp. 8–20, 2015.
55 N. Roy, G. J. Gordon, and S. Thrun, "Finding Approximate POMDP solutions Through Belief Compression," *Journal of Artificial Intelligence Research*, vol. 23, pp. 1–40, 2005.
56 G. J. Gordon, "Generalized^2 Linear^2 Models," in *Neural Information Processing Systems*, pp. 577–584, 2002.
57 A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with bregman divergences," *The Journal of Machine Learning Research*, vol. 6, pp. 1705–1749, 2005.