



# A Monte Carlo framework for missing wedge restoration and noise removal in cryo-electron tomography

Emmanuel Moebel, Charles Kervrann\*

Inria - Centre de Rennes Bretagne Atlantique, Campus Universitaire de Beaulieu, 35042 Rennes, France  
 Institut Curie, PSL Research University, CNRS UMR 144, UPMC, 75005 Paris, France



## ARTICLE INFO

### Keywords:

Cryo electron tomography  
 Missing wedge  
 Restoration  
 Inverse problems  
 Monte-Carlo simulation

## ABSTRACT

We propose a statistical method to address an important issue in cryo-electron tomography image analysis: reduction of a high amount of noise and artifacts due to the presence of a missing wedge (MW) in the spectral domain. The method takes as an input a 3D tomogram derived from limited-angle tomography, and gives as an output a 3D denoised and artifact compensated volume. The artifact compensation is achieved by filling up the MW with meaningful information. To address this inverse problem, we compute a Minimum Mean Square Error (MMSE) estimator of the uncorrupted image. The underlying high-dimensional integral is computed by applying a dedicated Markov Chain Monte-Carlo (MCMC) sampling procedure based on the Metropolis-Hasting (MH) algorithm. The proposed MWR (Missing Wedge Restoration) algorithm can be used to enhance visualization or as a pre-processing step for image analysis, including segmentation and classification of macromolecules. Results are presented for both synthetic data and real 3D cryo-electron images.

## 1. Introduction

Cryo-electron tomography (cryo-ET) is generally used to explore the structure of an entire cell and constitutes a rapidly growing field in biology. The particularity of cryo-ET is that it is able to produce three-dimensional views of vitrified samples at sub-nanometer resolution, which allows observing the structure of molecular complexes (e.g. ribosomes) in their physiological environment. Nevertheless, observation of highly resolved cellular mechanisms is challenging: i/ due to the low dose of electrons used to preserve specimen integrity during image acquisition, the amount of noise is very high; ii/ due to technical limitations of the microscope, complete tilting of the sample (90°) is impossible, resulting into a blind spot. As a consequence, projections are not available for a determined angle range, hence the term “limited angle tomography”.

The blind spot is observable in Fourier domain, where the missing projections appear as a missing wedge (MW). This separates the Fourier spectrum into two regions: the sampled region (SR) and the unsampled regions (MW). The sharp transition between these two regions is responsible for a Gibbs-like phenomenon: ray- and side-artifacts emanate from high contrast objects (see Fig. 1), which can hide important structural features in the image. Another type of artifact arises from the incomplete angular sampling: objects appear elongated in the direction of the blind spot (see Fig. 1), in other words the data has an anisotropic

resolution (e.g. linear features perpendicular to the tilt axis disappear). This elongation erases boundaries and makes it difficult to differentiate neighboring features. The quality of tomograms can be improved if sophisticated algorithms such as MBIR (Yan et al., 2019) are applied instead of conventional methods (e.g. WPB (Radermacher, 1992), SIRT (Gilbert, 1972)).

Filling up the MW with relevant data enables to potentially reduce or completely suppress these artifacts. Experimentally this can partially be achieved during data acquisition by using dual-axis tomography (Guesdon et al., 2013), where the sample is tilted with respect to the second axis. Consequently the blind spot is smaller and the MW becomes a missing pyramid, which results into a smaller missing spectrum. However dual-axis tomography is technically challenging and requires intensive post-processing in order to correct tilt and movement bias in the microscope. Another reconstruction approach consists in exploiting the symmetry of the underlying structure (Förster and Hegerl, 2007), but this can only be applicable to a limited number of biological objects (e.g. virus with either helical or icosahedral structure). Another common computational approach amounts to combining several hundred or thousands views of the same object, but with distinct blind spots. This so-called sub-tomogram averaging technique (Förster and Hegerl, 2007) is routinely used in cryo-ET, and is continuously improved for structure determination (Xu et al., 2018). To improve sub-tomogram averaging and compensate the remaining MW

\* Corresponding author.

E-mail address: [charles.kervrann@inria.fr](mailto:charles.kervrann@inria.fr) (C. Kervrann).

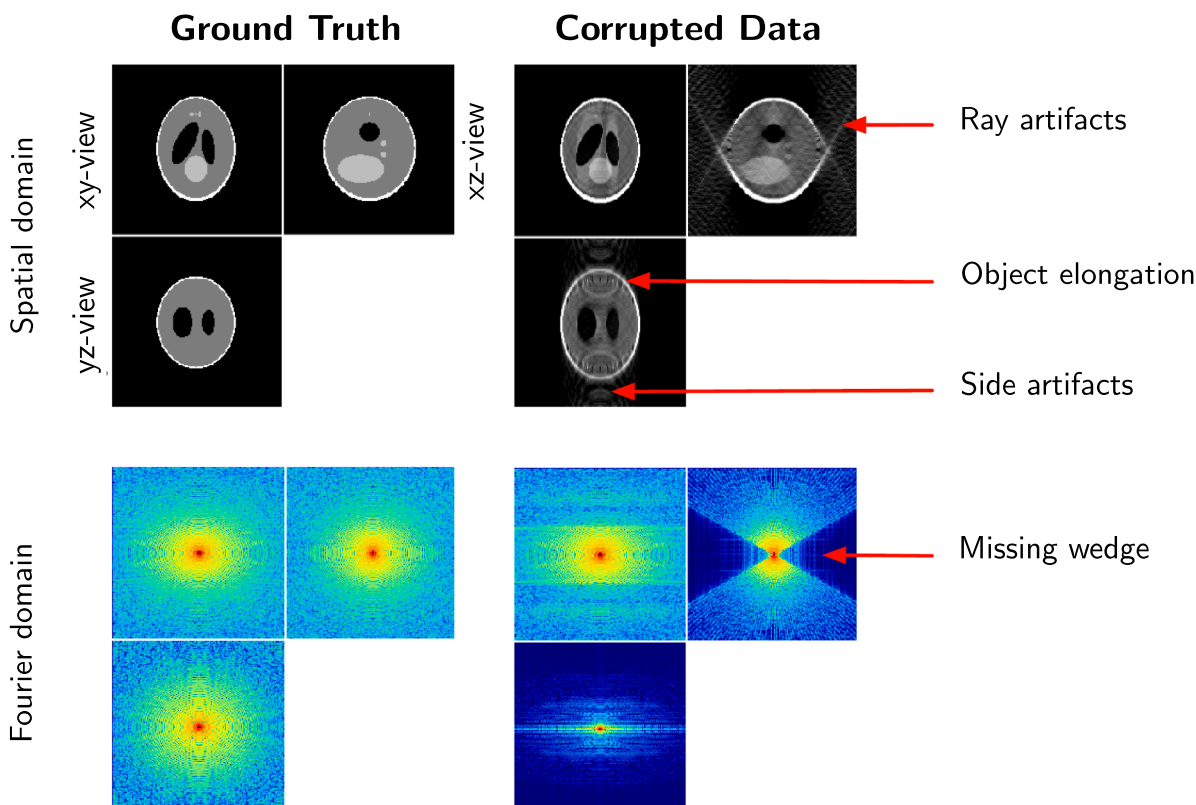


Fig. 1. The effects of the missing wedge on the Shepp-Logan phantom. Left: uncorrupted 3D data displayed as ortho-slices (one central slice along each dimension). Right: data corrupted by the missing wedge. We display the data in spatial domain (top row) and Fourier domain (bottom row).

artifacts, tomographic reconstruction algorithms with dedicated regularization have also been proposed in (Paavolainen et al., 2014; Leary et al., 2013).

The objective of our work is to design a statistical approach for the problem of recovering missing Fourier coefficients from a single volume in the situation where low and high frequency coefficients are missing in a specific and large region of the 3D spectrum. A simple way of handling MW artifacts is described in (Kováčik et al., 2014), where a dedicated spectral filter is used to smooth out the transition between SR and MW; ray- and side-artifacts are reduced with this filter, but the object elongation remains in the resulting image. Inspired from Maggioni et al. (2013), we have rather investigated MCMC methods to compute a MMSE estimator based on any non-local image denoiser to recover the missing information. We show that our Monte-Carlo sampling algorithm performs as well as the iterative method (Maggioni et al., 2013) but converges faster. Nevertheless, our concept is more general since any denoising method can be applied, included denoising algorithms dedicated to cryo-ET images (Frangakis and Hegerl, 2001; Fernandez and Li, 2003; Darbon et al., 2008; Wei and Yin, 2010; Kervrann et al., 2008; Moreno et al., 2018). In this paper, we focus on the cryo-ET restoration problem but the proposed algorithm could be potentially used to address a large range of applications including medical and seismic imaging, and other inverse scattering problems.

**Related work** We first focus on computational methods designed for spectrum restoration and Fourier coefficients recovering. Most of methods have been designed for 2D images and very few of them for 3D imaging. In general, the corruption process is supposed to be known and the artifacts observed in the input image, are due to a set of missing Fourier coefficients, well localized in the spectrum. First, several methods have been investigated to retrieve partially-missing phases of complex coefficients from modulus of coefficients in electron

microscopy (Fienup, 1982) and time-frequency signal analysis (Krémeš et al., 2018). Here, we focus on another special case which consists in extrapolating the band-limited spectrum of an image up to higher frequencies. Nevertheless, these problems are generally formulated as denoising problems with specific reconstruction constraints. For instance, Moisan (2001) and Guichard and Malgouyres (1998) investigated the Total Variation (TV) minimization to extend the band-limited spectrum of an image. In (Lauze et al., 2017), the authors combine TV minimization and positivity constraints to reduce noise and artifacts, providing an inpainting-like mechanism for the sinogram missing data in limited-angle tomography (see also (Frikel and Quinto, 2013; Sentosum et al., 2017)). The common objective is to create new frequencies while preserving discontinuities and details in the restored image. Instead of explicitly imposing some regularity (e.g. Total variation, or robust regularization (Geman and Reynolds, 1992; Charbonnier et al., 1997)) on the solution, another successful restoration approach consists in exploiting the spatial redundancy of the input image. In (Chambolle and Jalalzai, 2014), a non-local method was suggested in the framework of variational methods for image reconstruction. In this approach, a patchwise similarity measure based on atoms corresponding to pseudo Gabor filters is designed to compare corrupted regions. Meanwhile, Maggioni et al. (2013) adapted the concept of BM3D for recovering the missing spectrum applied to MRI imaging with very promising results on synthetic data. BM3D (Dabov et al., 2007) is a popular denoising algorithm which combines clustering of noisy patches, DCT-based transform and shrinkage operation to achieve the state-of-the-art results for several years. In our approach, we also focus on patch-based methods (Kervrann and Boulanger, 2008; Kindermann et al., 2005; Lou et al., 2010; Katkovnik et al., 2010; Pizarro et al., 2010; Milanfar, 2013; Sutour et al., 2014) to restore the input image corrupted by noise and non-linear transform. Indeed, it has been experimentally confirmed that the most competitive denoising

methods are non-local and exploit self-similarities occurring at large distances in images, such as BM3D (Dabov et al., 2007), NL-Bayes (Lebrun et al., 2013), PLOW (Chatterjee and Milanfar, 2012), S-PLC (Wang and Morel, 2013), PEWA (Kervrann, 2014) and many other adaptative filters (Kervrann and Boulanger, 2006; Kervrann and Boulanger, 2007; Van De Ville and Kocher, 2009; Deledalle et al., 2009; Louchet and Moisan, 2011; Duval et al., 2011; Deledalle et al., 2012; Kervrann et al., 2014; Jin et al., 2017), inspired from the seminal N(on) L(ocal)-means algorithm (Buades et al., 2005).

To complete the brief overview of non-local methods, we mention that a noisy image can also be restored from a set of noisy or “clean” patches or a learned dictionary. The statistics of a training set of image patches serve then as priors for denoising (Elad and Aharon, 2006; Mairal et al., 2009; Zoran and Weiss, 2011). Another approach based multi-layer perceptron (MLP) exploiting a training set of noisy and noise-free patches was also able to achieve the state-of-the-art performance (Burger et al., 2012). Very recently, Buchholz et al. (2018) proposed to train content-aware restoration networks for denoising cryo-transmission electron microscopy data. While all these machine learning methods are attractive and powerful, computation is not always feasible in 3D because very large collection of 3D “clean” patches are required. In our study, we focus on unsupervised denoising methods since they are more flexible for real applications. They are less computationally demanding and are still competitive when compared to recent machine learning methods.

Our approach is mainly inspired from Maggioni et al. (2013), but can use any competitive denoising methods for restoring the Fourier coefficients. The method proposed by Maggioni et al. (2013) works by alternatively adding noise into the missing region and applying the BM4D algorithm which is the extension of BM3D (Dabov et al., 2007) to volumes. The authors interpret this iterative restoration method in the framework of compressed sensing with two information theory concepts in mind: *sparsity* of the signal in the transformed domain, and *incoherence* between the transform and the sampling matrix. Actually, BM4D does rely on a transform where the signal is sparse. Moreover, it is not clearly established that this transform is incoherent with the sampling matrix, defined by the support of the sampling region. Therefore, the proof of convergence is not clearly established, even though the authors presented convincing experimental results on synthetic images corrupted with white Gaussian noise. It remains unclear how the concept performs on experimental data and non Gaussian noise. To generalize this idea, we propose a statistical approach well-grounded in the Bayesian and MCMC framework and applied to challenging real data in cryo-ET. Our contributions are the following ones:

1. We present a MMSE estimator dedicated to the problem of MW restoration.
2. We propose an original Monte Carlo Markov Chain (MCMC) sampling procedure to efficiently compute the MMSE estimator.

**Paper organization** The remainder of the paper is organized as follows. In the next section, we present an overview of our computational method. In Section 3, we give the theoretical background and formulate the reconstruction problem as an inverse problem. We shortly describe the usual Bayesian approach to derive a MMSE estimator. Furthermore, a Monte-Carlo Markov Chains method based on the popular Metropolis-Hastings algorithm is proposed to compute the underlying high-dimensional integral. In Section 4, we adapt this general Bayesian framework for MW restoration. An original Metropolis-Hastings procedure is presented to explore the large space of admissible solutions and to select relevant samples. Section 5 presents the experimental results obtained on simulated and real data. We illustrate the potential of our MWR (Missing Wedge Restoration) algorithm with experiments on real cryo-tomogram images and we compare our iterative approach to several competitive algorithms.

## 2. Overview of the MW restoration algorithm

In this section, we present our concept for MW restoration. First, we formulate the problem and introduce notation. Second, we present the two-step algorithm which is embedded in a Monte-Carlo simulation framework.

### 2.1. Problem formulation and notation

Let us define a  $n$ -dimensional image  $\mathbf{x}: S \subset \mathbb{Z}^3 \rightarrow \mathbb{R}$  assumed to be periodic and defined over a cubic domain  $\Omega = [0, 1]^3$  and  $n = |\Omega|$ . The discrete Fourier transform of  $\mathbf{x} = \{x(s), s \in S\}$  is then as follows:

$$\mathcal{F}\mathbf{x}: k \rightarrow \sum_{s \in S} \exp(-2i\pi k \cdot s)x(s), \quad (1)$$

where  $s$  is the coordinate of point in spatial domain  $S$ . In our problem, one considers a corrupted image denoted  $\mathbf{y} = \{y(s), s \in S\}$  defined as

$$y(s) = \sum_{k \in \bar{W}} \exp(2i\pi k \cdot s) \mathcal{F}\mathbf{x}[k] \quad (2)$$

where  $\bar{W}$  is the sampled spectral region (SR) where the Fourier coefficients  $\mathcal{F}\mathbf{x}[k]$  are positive and non-zero. The region  $\bar{W}$  is equivalent to the support of a binary mask  $\mathbf{m} \in \{0, 1\}^S$  such as  $\mathbf{m}[k] = 1$  if  $k \in \bar{W}$  and 0 otherwise:  $\bar{W} = \text{supp}(\mathbf{m})$ . The so-called missing wedge  $W$  is assumed to be symmetric with respect to the origin as illustrated in Fig. 1 (bottom right), and  $S = W \cup \bar{W}$ . In what follows, we assume that the clean image  $\mathcal{F}\mathbf{x}$  is known over the region  $\bar{W}$ . Our objective is then to estimate  $\mathbf{x}: S \rightarrow \mathbb{R}$  in the whole domain  $S$  such that

$$\forall k \in \bar{W}, \quad \mathcal{F}\mathbf{x}[k] = \mathcal{F}\mathbf{y}[k] \quad (3)$$

and  $\forall s \in S, x(s) > 0$ . In other words, the set of known Fourier coefficients will be preserved by the restoration procedure. The challenge is to recover the unknown set of low, middle and high frequencies in a large region in the spectrum. This amounts to applying an interpolation operator  $\varphi_W$  to the spectrum of  $\mathbf{y}$  to get an estimator  $\hat{\mathbf{x}}$  of  $\mathbf{x}$ :

$$\hat{\mathbf{x}} = \mathcal{F}^{-1} \circ \varphi_W \circ \mathcal{F}\mathbf{y}. \quad (4)$$

### 2.2. Principles of the iterative two step-algorithm

Our computational method takes as input a noisy image  $\mathbf{y}$  and iterates two steps as illustrated in Fig. 2. At the initialization ( $t = 0$ ), we set  $\mathbf{x}^{(0)} = \mathbf{y}$ :

- In the *proposal* step and at iteration  $t$ , white Gaussian noise  $\varepsilon$  is added to the current image  $\mathbf{x}^{(t-1)}$ . The Fourier coefficients are then non-zero in the MW region  $W$ . We substitute the original Fourier coefficients in the SR region  $\bar{W}$  to the noisy Fourier coefficients to preserve information. Formally, this amounts to applying a projection operator  $\mathcal{P}_W(\cdot)$  to the artificially noisy image. The objective is to randomly initialize the Fourier coefficients in the MW. A denoising algorithm is then used to efficiently remove the majority of the noise while preserving image structure. Due to its random nature, a small amount of the added noise will be close to the signal and is thus “saved” after denoising. The denoising algorithm  $\mathcal{D}(\cdot)$ , which promotes smoothness while preserving contours, acts as a spatial regularizer. In summary, a candidate image  $\mathbf{z} = \mathcal{D}(\mathcal{P}_W(\mathbf{x}^{(t-1)} + \varepsilon))$  is obtained by applying a projection operator  $\mathcal{P}$  in the Fourier domain and a denoising operator  $\mathcal{D}$  in the spatial domain.
- In the second *evaluation* step, the candidate image  $\mathbf{z}$  is compared and is potentially substituted to the current image  $\mathbf{x}^{(t-1)}$  depending on its ability to satisfies several constraints, including fidelity to the original input image  $\mathbf{z}$ . A cost function (or energy) is used to accept/reject this candidate image with a certain probability established in the Metropolis-Hastings simulation framework as presented in the next section.

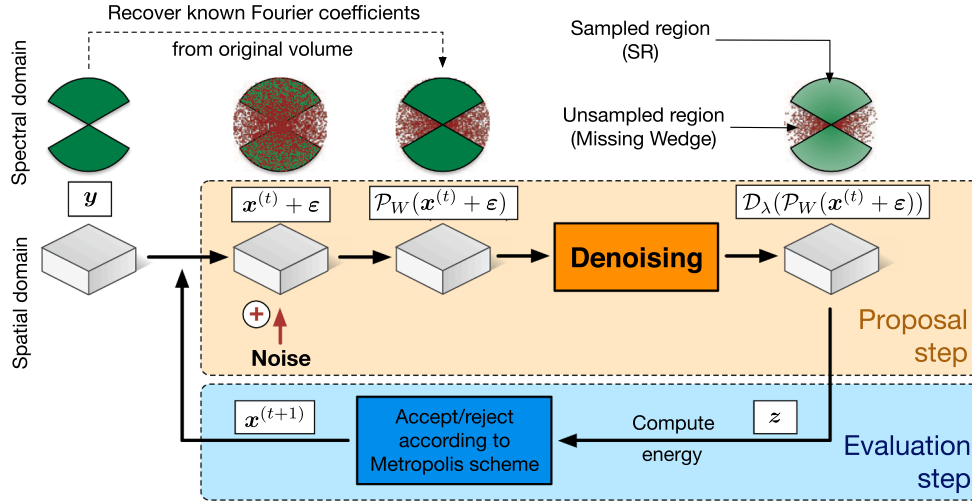


Fig. 2. The MCMC method flowchart. The 1st icon row represents the data in the spectral domain, the 2nd in spatial domain.

By repeating these two steps iteratively, we collect several hundreds of images  $\{\mathbf{x}^{(t)}\}$  to compute an average corresponding to the final restored image. The corresponding MWR (Missing Wedge Restoration) algorithm is actually able to diffuse information from SR into MW. It cannot retrieve unobserved data, but it merely makes the best statistical guess of what the missing data could be, based on what has been observed. The underlying algorithm is controlled by several parameters, especially the variance  $\sigma_\epsilon^2$  of noise added to the current image and constant at each iteration. More importantly, this iterative procedure will be successful provided the denoising algorithm is able to remove the artificial additive Gaussian noise. Any state-of-the-art algorithms can be then applied, especially the BM3D (Dabov et al., 2007) and BM4D (Maggioni et al., 2013) algorithms. This stochastic procedure can be interpreted as a data-driven random search in a large space of possible images. In the next section, we present the theoretical background required to justify the rationale behind this concept, illustrated in Fig. 2.

### 3. Theoretical background

In this section, we describe the theoretical background in the Bayesian framework to justify our original computational method.

#### 3.1. Bayesian estimator and Monte Carlo Markov Chain sampling

Solving inverse problems in image processing consists in estimating an unknown image  $\mathbf{x} \in \mathcal{X}$  given an image  $\mathbf{y} \in \mathcal{Y}$ . Different sources of distortion may cause damages on the ideal image, including noise, blur, and projections. In the Bayesian framework, the whole information once the data have been collected, is represented by the posterior probability density function (pdf) defined via the Bayes' Theorem:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \quad (5)$$

where  $p(\mathbf{y}|\mathbf{x})$  denotes the likelihood function,  $p(\mathbf{x})$  is the prior pdf and  $p(\mathbf{y})$  is the marginal distribution of  $\mathbf{y}$  which is in general unknown and not computable.

##### 3.1.1. Bayesian estimators

In this section,  $\mathbf{x}$  and  $\mathbf{y}$  are realizations of a random variable  $\mathbf{X}$  (with a pdf  $p(\mathbf{x})$ ) and a random realization of  $\mathbf{Y}$  respectively. Given a cost function  $C: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+$ , a Bayesian estimator is defined as the minimizer of expected risk  $\mathbb{E}_p[C(\mathbf{X}, \hat{\mathbf{x}}(\mathbf{Y}))]$  wrt the joint distribution  $p(\mathbf{x}, \mathbf{y})$  of the pair  $(\mathbf{X}, \mathbf{Y})$ . Several Bayesian estimators can be derived based on the choice of the cost function  $C$ .

**MAP estimator** The most conventional choice is  $C(\mathbf{x}, \hat{\mathbf{x}}) = 1 - \delta(\mathbf{x}, \hat{\mathbf{x}})$  where  $\delta$  is the Kronecker symbol. The corresponding Maximum A Posteriori estimator, defined as

$$\hat{\mathbf{x}}_{MAP} = \operatorname{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \operatorname{argmin}_{\mathbf{x}} \left\{ -\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}) \right\}, \quad (6)$$

selects the most likely image  $\mathbf{x}$ , that is the solution corresponding to the mode of the posterior distribution  $p(\mathbf{x}|\mathbf{y})$ .

Furthermore, if we assume that the prior and likelihood functions are represented by Gibbs functions, the posterior distribution has the following form

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} \exp\left(-\frac{U(\mathbf{x}, \mathbf{y})}{\tau}\right) \quad (7)$$

where  $Z$  is normalizing factor,  $U(\mathbf{x}, \mathbf{y}) = D(\mathbf{x}, \mathbf{y}) + \phi(\mathbf{x})$  is an energy functional composed of a data-fidelity term  $D(\mathbf{x}, \mathbf{y})$  and a prior term  $\phi(\mathbf{x})$ , and  $\tau$  can be interpreted as a "temperature" or scale parameter. The prior generally encourages piecewise smoothness (TV) or sparsity of  $\mathbf{x}$ . Hence, the MAP formulation is equivalent to the popular variational problem which amounts to computing the unique image  $\mathbf{x}$  that minimizes the following criterion:

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} D(\mathbf{x}, \mathbf{y}) + \phi(\mathbf{x}). \quad (8)$$

Typically  $\phi(\mathbf{x}) = \|\nabla \mathbf{x}\|_1$  ( $\nabla \cdot$  is the gradient operator) is the total variation regularizer and serves to smooth the image  $\mathbf{x}$  while preserving image discontinuities.

**MMSE estimator** Another well-known Bayesian estimator can be derived if we consider the quadratic cost function  $C(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2$ . The Minimum Mean Square Error (MMSE) estimator is the posterior expectation (or conditional mean) defined as:

$$\hat{\mathbf{x}}_{MMSE} = \mathbb{E} \left[ \mathbf{X} | \mathbf{Y} = \mathbf{y} \right] = \int_{\mathcal{X}} \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}. \quad (9)$$

If the posterior is modeled as Gibbs distribution, we get:

$$\hat{\mathbf{x}}_{MMSE} = \frac{\int_{\mathcal{X}} \exp\left(-\frac{U(\mathbf{x}, \mathbf{y})}{\tau}\right) \mathbf{x} d\mathbf{x}}{\int_{\mathcal{X}} \exp\left(-\frac{U(\mathbf{x}, \mathbf{y})}{\tau}\right) d\mathbf{x}}. \quad (10)$$

In our case, the MMSE estimator is typically intractable since the underlying integral involves several thousands of variables (typically  $n$



is the number of pixels in the image). The MMSE estimator cannot be computed in a closed form, and numerical approximations are typically required. In high-dimensional space, a common approach consists in approximating the integral by using Monte Carlo (MC) simulation techniques (Robert and Casella, 2004) as explained in the next section.

However, we draw the readers' attention to the fact that it has been shown that the MMSE estimator has connections with the variational optimization problem in the case of an image corrupted by white Gaussian noise:

$$\hat{\mathbf{x}}_{MMSE} = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|^2 + \psi(\mathbf{x}), \quad (11)$$

where the function  $\psi(\mathbf{x})$  can be seen as a pseudo-prior which differs from the prior distribution  $p(\mathbf{x}) \propto \exp(-\phi(\mathbf{x}))$ . Nevertheless, except in the case of a explicit and dedicated prior discussed in (Protter et al., 2010; Grignonval, 2011; Louchet and Moisan, 2013; Kazerouni et al., 2013), it is not possible to derive an analytical form of  $\psi(\mathbf{x})$  from  $\phi(\mathbf{x})$ , especially if the data-fidelity term is not quadratic. Accordingly, the most practical way to compute a MMSE estimator in the case of complex data-fidelity terms and prior terms is to applying the MCMC approach.

### 3.1.2. Monte-Carlo integration

Let us consider  $T$  independent and identically distributed (i.i.d.) samples  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(T)}$  drawn from a target pdf  $\pi(\mathbf{x}) := p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ . A consistent estimator can be computed as

$$\bar{\mathbf{x}}_T := \frac{1}{T} \sum_{t=1}^T \mathbf{x}^{(t)} \xrightarrow{p} \hat{\mathbf{x}}_{MMSE}, \quad (12)$$

i.e. the empirical mean of samples converges in probability to  $\hat{\mathbf{x}}_{MMSE}$  due to the weak law of large numbers. Formally, for any positive number  $\epsilon > 0$ , we have

$$\lim_{T \rightarrow +\infty} \Pr(|\bar{\mathbf{x}}_T - \hat{\mathbf{x}}_{MMSE}| > \epsilon) = 0. \quad (13)$$

The Monte-Carlo estimator is unbiased  $\mathbb{E}_{\pi}[\bar{\mathbf{x}}_T] = \hat{\mathbf{x}}_{MMSE}$  and converges provided that the samples  $\mathbf{x}^{(t)}$  are i.i.d.. In that case, the variance of  $\bar{\mathbf{x}}_T$  defined as  $\operatorname{Var}[\bar{\mathbf{x}}_T] = v^2/T$  (where  $v^2 = \operatorname{Var}[\mathbf{X}]$ ) decreases with the number of samples, and  $\bar{\mathbf{x}}_T$  is Gaussian distributed due to the central limit theorem:  $\bar{\mathbf{x}}_T \sim \mathcal{N}(\hat{\mathbf{x}}_{MMSE}, v^2/T)$  when  $T \rightarrow +\infty$ .

A central question is then to draw a series of i.i.d. samples. The most conventional approach in high-dimensional space is to consider Markov chain Monte Carlo (MCMC) algorithms (Gilks et al., 1995; Robert and Casella, 2004; Liang et al., 2010). In the sequel, we focus on a few important components of the MCMC machinery and we ss the convergence properties of the Metropolis-Hastings algorithm used in our approach to generate a Markov chain with a target stationary distribution  $\pi$ .

#### Simulating independent samples and fusion of multiple chains

Drawing independent samples from the target pdf  $\pi(\mathbf{x})$  cannot be directly applied. A MCMC procedure is able to simulate an ergodic and stationary Markov chain given a target pdf  $\pi(\mathbf{x})$  and a starting state  $\mathbf{x}^{(0)}$ . The set of samples  $\mathbf{x}^{(0)} \rightarrow \dots \rightarrow \mathbf{x}^{(T)}$  are generally correlated samples, but it has been established that Monte-Carlo estimator is consistent as  $T \rightarrow +\infty$ . In (Louchet et al., 2008; Louchet and Moisan, 2013), the authors also studied the behavior of the expected approximation error

$$\mathbb{E}[\|\bar{\mathbf{x}}_T - \hat{\mathbf{x}}_{MMSE}\|^2] = \mathbb{E}[\|\bar{\mathbf{x}}_T - \mathbb{E}[\bar{\mathbf{x}}_T]\|^2] + \|\mathbb{E}[\bar{\mathbf{x}}_T] - \hat{\mathbf{x}}_{MMSE}\|^2 \quad (14)$$

decomposed into the sum of the trace of the covariance matrix (or span) and the squared bias which entails the loss of efficiency of the sampling procedure.

In practice, a MCMC method will provide better performance than another MCMC method if the samples present less correlation. On the contrary, it is required to generate many samples to reduce the variance of the estimator.

Finally, if we consider another Markov chain  $\mathbf{x}'$  defined as  $\mathbf{x}$ , it is

established that (Louchet et al., 2008; Louchet and Moisan, 2013):  $\mathbb{E}[\|\bar{\mathbf{x}}_T - \bar{\mathbf{x}}_T'\|^2] = 2 \mathbb{E}[\|\bar{\mathbf{x}}_T - \mathbb{E}[\bar{\mathbf{x}}_T]\|^2]$ . It follows that the average of the two chains has a smaller span than the span of each independent chain while keeping the same bias, i.e..

$$\mathbb{E} \left\| \left\| \frac{\bar{\mathbf{x}}_T + \bar{\mathbf{x}}_T'}{2} - \hat{\mathbf{x}}_{MMSE} \right\|^2 \right\| \leq \frac{1}{2} \mathbb{E}[\|\bar{\mathbf{x}}_T - \mathbb{E}[\bar{\mathbf{x}}_T]\|^2] + \left\| \mathbb{E}[\bar{\mathbf{x}}_T] - \hat{\mathbf{x}}_{MMSE} \right\|^2.$$

This suggests that averaging multiple independent Markov chains should provide better estimators.

**Burn-in phase** Another consequence of the correlation is the burn-in period that the chain requires before converging to the invariant target pdf  $\pi$ . In general, the initial  $T_b$  samples are discarded and not included in the computation of the estimator (Robert and Casella, 2004):

$$\tilde{\mathbf{x}}_T := \frac{1}{T - T_b} \sum_{t=T_b}^T \mathbf{x}^{(t)}. \quad (15)$$

However, the length  $T_b$  of the burn-in period cannot be easily predicted even if a few studies in the literature focused on that problem (Gelman and Rubin, 1992; Brooks and Gelman, 1998).

**Metropolis-Hastings algorithm** The most popular and widely applied MCMC algorithm is based on the Metropolis-Hastings procedure (Metropolis et al., 1953; Hastings, 1970) described below. The MH algorithm involves the definition of the proposal density  $q(\mathbf{z}|\mathbf{x})$ ,  $\mathbf{x}, \mathbf{z} \in \mathcal{X}$  to move from the state  $\mathbf{x}$  to state  $\mathbf{z}$ , and the acceptance probability  $0 \leq a(\mathbf{x}, \mathbf{z}) \leq 1$ . The transition probability is then defined as:  $p(\mathbf{z}|\mathbf{x}) = q(\mathbf{z}|\mathbf{x})a(\mathbf{z}, \mathbf{x})$ . In the MH procedure, a sample  $\mathbf{z}$  is drawn from the proposal distribution and then a test is applied to accept the transition from the state  $\mathbf{x}$  to the state  $\mathbf{z}$  or not. If the transition is not accepted, the chain remains in the same state as before:

#### The Metropolis-Hastings algorithm

1. Set an initial state  $\mathbf{x}^{(0)}$ .
2. **For**  $t = 1, \dots, T$  **do**
  - (a) Draw a sample  $\mathbf{z} \sim q(\mathbf{x}|\mathbf{x}^{(t-1)})$ .
  - (b) Compute the acceptance probability

$$a\left(\mathbf{x}^{(t-1)}, \mathbf{z}\right) = \min \left[ 1, \frac{\pi(\mathbf{z})q(\mathbf{x}^{(t-1)}|\mathbf{z})}{\pi(\mathbf{x}^{(t-1)})q(\mathbf{z}|\mathbf{x}^{(t-1)})} \right].$$

- (c) Draw  $\alpha$  from a uniform distribution:  $\alpha \sim U(0, 1)$ .
- (d) **If**  $\alpha \leq a(\mathbf{x}^{(t-1)}, \mathbf{z})$  **then**  $\mathbf{x}^{(t)} = \mathbf{z}$ ,  
**else**  $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)}$ .  
**end if**  
**end for**

The MH algorithm returns a set of  $T_b - T$  correlated samples if we discard the  $T_b$  first samples. Under some mild regularity conditions, it has been established that the pdf of  $\mathbf{x}^{(t)}$  converges to the target pdf  $\pi$  when  $t$  increases (Robert and Casella, 2004). In general, the MH algorithm satisfies the so-called detailed balance condition:

$$p(\mathbf{x}, \mathbf{z}) \in \mathcal{X} \times \mathcal{X}, p(\mathbf{x}|\mathbf{z})\pi(\mathbf{z}) = p(\mathbf{z}|\mathbf{x})\pi(\mathbf{x}),$$

(the chain is reversible) which is a sufficient condition to guarantee that the chain is ergodic and has  $\pi$  as stationary distribution (Robert and Casella, 2004). Note that reversibility of the chain is not a necessary condition; recent studies experimentally show that non-reversible Markov chains may provide better convergence, i.e. the number  $T$  of samples can be lower than the reversible chains (Neal, 2004; Bierkens, 2015).

In practice, the proposal density and the acceptance probability can be modified in order to improve the performance of the algorithm, and always ensuring the ergodicity of the chain. Actually, the proposal pdf  $q$

should be chosen as close as possible to the target pdf  $\pi$ . In what follows, we mainly focus on the specification of the proposal densities to improve convergence.

**Choices of the proposal density** There is a large flexibility in the choice of proposal function and it is a challenge to find a proposal function that is able to use the data efficiently in order to obtain satisfactory convergence. Below, we ss four possible proposals.

- Assume that the proposal satisfies the equality  $q(\mathbf{z}|\mathbf{x}) = q(\mathbf{x}|\mathbf{z})$  (e.g. uniform distribution), then the acceptance probability is simplified since a sample  $\mathbf{z}$  having a higher value  $\pi(\mathbf{z})$  is always accepted, whereas the samples with smaller values  $\pi(\mathbf{z}) < \pi(\mathbf{x})$  are accepted with a probability lower than 1.
- The proposal pdf has the following form:  $q(\mathbf{z}|\mathbf{x}) = q(\mathbf{z} - \mathbf{x})$ . This means that the new state is explicitly randomly found in the neighborhood of the current state  $\mathbf{x}$ . This proposal is then viewed as random walk and enables to progressively explore the large space of possible states. Nevertheless, the random walk MH algorithm (1953) tends to stay in the same state for a long period but the chain has not converged.
- When the target density is differentiable, the proposal can be generated in accordance with an approximation of the Langevin diffusion process (Girolami and Calderhead, 2011):  $\mathbf{z} \sim \mathcal{N}\left(\mathbf{x} + \frac{\delta}{2} \nabla \log \pi(\mathbf{x}), \delta\right)$  for a given small value  $\delta$ .
- The idea of adaptive MH algorithm consists in updating the proposal distribution by using all the information collected so far about the target distribution (Holden et al., 2009). First, it has been suggested to model the proposal density as a Gaussian distribution centered on the current state with a covariance computed from a fixed finite number of previous states. Given the whole history  $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t-1)})$ , the new state  $\mathbf{z}$  is obtained from  $q(\mathbf{z}|\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t-1)})$  assumed to be symmetric.
- If the proposal pdf  $q(\mathbf{z}|\mathbf{x}) = q(\mathbf{x})$  does not depend on the state  $\mathbf{x}$ , the acceptance probability is defined as

$$a(\mathbf{x}, \mathbf{z}) = \min \left[ 1, \frac{\pi(\mathbf{z})q(\mathbf{x})}{\pi(\mathbf{x})q(\mathbf{z})} \right].$$

The independent Metropolis-Hastings algorithm is an efficient sampling algorithm only if  $q$  is reasonably close to  $\pi$ . An attractive property of independent proposals is their ability to make large jumps while keeping the acceptance rate high. Consequently, the autocorrelation of the chain decreases rapidly.

Note that more sophisticated proposal rules are generally recommended to address high-dimensional problems (Girolami and Calderhead, 2011). For instance, a two-step optimization approach is typically appropriate for sampling Gaussian distributions (Papandreou and Yuille, 2010; Gilavert et al., 2015). Another sophisticated approach is based on data augmentation and the adding of auxiliary variables to improve convergence speed if the samples are correlated (Marnissi et al., 2018).

In summary, the convergence of the chain depends on the specification of the proposal density but it is also established that the ideal proposal pdf must as close as possible to the target pdf. In that case, the MH procedure generates a sequence of states with low correlations and converges faster. In our approach described in the next section, we investigated a stochastic scheme to generate samples with low correlations in the context of MW restoration.

## 4. Our MH algorithm for missing wedge restoration

### 4.1. Gibbs energy modeling

Let us consider the following image model

$$\mathbf{y} = \mathcal{D}_W(\mathbf{x}), \quad (16)$$

where  $\mathbf{y}, \mathbf{x} \in \mathbb{R}^n$ , and  $\mathcal{D}_W(\cdot)$  is a degradation operator setting to zero the Fourier coefficients belonging to the MW support  $W$  assumed to be known. Our objective is to compute a MMSE estimator defined as

$$\hat{\mathbf{x}}_{MMSE} = \frac{\int_{\mathcal{X}} p(\mathbf{x}|\mathbf{y})\mathbf{x}d\mathbf{x}}{\int_{\mathcal{X}} p(\mathbf{x}|\mathbf{y})d\mathbf{x}} = \frac{\int_{\mathcal{X}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathbf{x}d\mathbf{x}}{\int_{\mathcal{X}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}} \quad (17)$$

by using a dedicated MH algorithm, where  $p(\mathbf{x})$  is the prior used to encourage the solution to be positive and piecewise smooth. In our modeling, the likelihood function is composed of two terms: i/ we impose that the Fourier coefficients of the reconstructed image are very similar to the known coefficients of the corrupted image  $\mathbf{y}$ , i.e.  $|\mathcal{F}\mathbf{x}[k] - \mathcal{F}\mathbf{y}[k]| < \delta$ ; ii/we consider a data fidelity term defined as the  $L_2$  norm (alternative data fidelity terms will be considered in our experiments) between the corrupted image and the restored image degraded by the operator  $\mathcal{D}_W$ , i.e.

$$D(\mathbf{x}, \mathbf{y}) = \sum_{s \in S} \left( y(s) - \sum_{k \in S} e^{2i\pi k \cdot s} \mathcal{F}\mathbf{x}[k] \right)^2. \quad (18)$$

It follows that the posterior is defined as ( $\mathbb{1}_{\mathcal{A}}(\mathbf{z}) = 1$  if  $\mathbf{z}$  belongs to the set  $\mathcal{A}$  and zero otherwise)

$$p\left(\mathbf{x} \middle| \mathbf{y}\right) \propto \underbrace{\exp\left(-\frac{D(\mathbf{x}, \mathbf{y})}{\beta}\right)}_{\text{likelihood}} \underbrace{\mathbb{1}_{\mathcal{A}_{\delta, \mathbf{y}}}(\mathbf{x})}_{\text{prior}} \mathbb{1}_{\mathcal{A}_{\lambda}^+}(\nabla \mathbf{x}) \mathbb{1}_{\mathcal{A}_+}(\mathbf{x}), \quad (19)$$

such that  $\mathcal{A}_{\delta, \mathbf{y}}(\mathbf{x}) = \{\mathbf{x} \in \mathcal{X} : \forall k \in \bar{W} |\mathcal{F}\mathbf{x}[k] - \mathcal{F}\mathbf{y}[k]| < \delta\}$  and  $\mathcal{A}_{\lambda}^+(\nabla \mathbf{x}) = \{\mathbf{x} \in \mathcal{X} : \|\nabla \mathbf{x}\|_1 \leq \lambda\}$  where  $\|\nabla \mathbf{x}\|_1 = \sum_{s \in S} |\nabla x(s)|$  is the discrete Total Variation (Rudin et al., 1992). Here,  $\mathcal{A}_+$  denotes the set of positive solutions, that is the set of images for which  $x(s) \geq 0, \forall s \in S$ . As mentioned in Section 3.1,  $\beta$  can be interpreted in (19) as a ‘‘temperature’’ parameter. In Eq. (19), the prior term imposes positivity and regularity of the solution. The TV norm of the restored image is then assumed to be lower than a prescribed threshold  $\lambda$ . The likelihood is based on reconstruction error (18) and adequacy of frequency in the SR region  $\bar{W}$ .

Consequently, the MMSE estimator can reformulated as

$$\hat{\mathbf{x}}_{MMSE} = \frac{\int_{\Gamma} \exp\left(-\frac{D(\mathbf{x}, \mathbf{y})}{\beta}\right) \mathbf{x} d\mathbf{x}}{\int_{\Gamma} \exp\left(-\frac{D(\mathbf{x}, \mathbf{y})}{\beta}\right) d\mathbf{x}}, \quad (20)$$

where the set  $\Gamma$  of admissible solutions is defined as

$$\Gamma = \{\mathbf{x} \in \mathcal{X} : \forall k \in \bar{W} |\mathcal{F}\mathbf{x}[k] - \mathcal{F}\mathbf{y}[k]| < \delta, \quad \forall s \in S, x(s) \geq 0, \text{ and } \|\nabla \mathbf{x}\|_1 < \lambda\}.$$

The performance of MMSE (20) depends on the pre-specified thresholds  $\lambda$  and  $\delta$ . In practice, these values do not need to be accurately adjusted in practice as ssed below. Meanwhile, because of the high dimensionality of the problem, we need an efficient MH algorithm to compute MMSE.

### 4.2. Implementation of the MH algorithm

The efficiency of a MH algorithm depends on the choice of the proposal distribution  $q(\cdot|\mathbf{x}^{(t-1)})$ . In practice, the proposal generator leads to correlated samples  $\mathbf{x}^{(t-1)}$  induced by the two following factors: i/ by construction, the newly proposed state  $\mathbf{x}^{(t)} \sim q(\cdot|\mathbf{x}^{(t-1)})$  is generated from the current state; ii/ the new state  $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)}$  when the proposed move has been rejected. Note that this correlation is not known in advance but can be empirically estimated and updated from the previous samples (Holden et al., 2009). To achieve good performance, a well-chosen proposal distribution both allows significant changes between the subsequent states with a high probability of acceptance (see

Section 4.2.4). Unfortunately these requirements cannot be satisfied easily in practice. If we choose a proposal distribution with small moves, the probability of acceptance will be high, however the resulting chain will be highly correlated, as  $\mathbf{x}^{(t)}$  changes only very slowly. In return, if we choose a proposal distribution with large moves, the probability of acceptance will be rather low. Accordingly, we investigated an original strategy to generate a sequence of moves with a probability of acceptance in the range of  $a \in [0.25 - 0.6]$ . In theoretical studies, it has been shown that the optimal probability of acceptance  $a^* = 0.234$  (Roberts et al., 1997) whereas  $a^* = 0.574$  in Roberts and Rosenthal (2001).

Given an initial state  $\mathbf{x}^{(0)} \in \Gamma$ , we explore the neighborhood of the current value set  $\mathbf{x}^{(t)}$  of the chain. Our proposal distribution  $q$ , which enables to potentially move from  $\mathbf{x} \in \Gamma$  to  $\mathbf{x}' \in \Gamma$  is chosen as (also see Fig. 2):

$$\mathbf{x}' = \mathcal{D}_\lambda(\mathcal{P}_W(\mathbf{x} + \varepsilon)), \quad (21)$$

where  $\varepsilon \sim \mathcal{N}(0, I_{n \times n} \sigma_\varepsilon^2)$  is a white Gaussian noise,  $I_{n \times n}$  is the  $n$ -dimensional identity matrix,  $\mathcal{P}_W$  is a projection operator that impose that the  $\forall k \in \overline{W}$ ,  $\mathcal{F}(\mathbf{x} + \varepsilon)[k] = \mathcal{F}\mathbf{y}[k]$  and  $\mathcal{D}_\lambda$  is denoising operator that ensures that the total variation of the denoised image is lower than  $\lambda$ . Consequently, the distribution of increments  $\mathbf{x}' - \mathbf{x}$  is not parametric due to the nonlinear operators involved in (21). In the sequel, we only assume that this non-parametric distribution  $q$  is approximately symmetric. Even though visualization of the empirical proposal density is not possible, we suggest that (21) tends to produces similar samples (denoised images) concentrated around some empirical mean belonging to  $\Gamma$ , with a few moves quite far away from this mean.

Our simulator can be viewed as a random walk in a high-dimensional space, where all the pixels of the images are modified at once. Note that in (Louchet et al., 2008; Louchet and Moisan, 2013), only one pixel is modified at each iteration to compute the TV-LSE estimator. Our approach can be seen also a blockwise MH sampling procedure but only one block corresponding to the whole image, is considered in procedure. In our experiments, we observed a high acceptance rate, suggesting that the new sample is not far from the previous one (i.e.  $\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\|$  is small). To our knowledge, this is the first time that such a proposal, is used in the context of image restoration and inverse problems. Our MH algorithm for MW restoration is then as follows (see Fig. 2):

#### The Missing Wedge Restoration (MWR) algorithm

Set an initial state  $\mathbf{x}^{(0)} \in \Gamma$ .

For  $t = 1, \dots, T$  do

1. Generate a new state  $\mathbf{z}$  with the a three-step approach:
  - **Perturbation:**  $\mathbf{x}^{(t-1)} + \varepsilon$ ,  $\varepsilon \sim \mathcal{N}(0, I_{n \times n} \sigma_\varepsilon^2)$ .
  - **Projection:** of  $\mathcal{P}_W(\mathbf{x}^{(t-1)} + \varepsilon)$  onto the subspace of images having the same observed frequencies as  $\mathbf{y}$  if  $k \in \overline{W}$ .
  - **Denoising** of  $\mathcal{P}_W(\mathbf{x}^{(t-1)} + \varepsilon)$  to get an image with a small  $\|\nabla \mathbf{x}\|$  and set  $\mathbf{z} = \mathcal{D}_\lambda(\mathcal{P}_W(\mathbf{x}^{(t-1)} + \varepsilon))$ .
2. Compute the acceptance probability

$$a(\mathbf{x}^{(t-1)}, \mathbf{z}) = \min \left[ 1, \exp \left( \frac{D(\mathbf{x}^{(t-1)}, \mathbf{y}) - D(\mathbf{z}, \mathbf{y})}{\beta} \right) \right].$$

3. Draw  $\alpha$  from a uniform distribution:  $\alpha \sim U(0, 1)$
4. If  $\alpha \leq a(\mathbf{x}^{(t-1)}, \mathbf{z})$  then  $\mathbf{x}^{(t)} = \mathbf{z}$ 
  - else  $\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)}$ .
  - end if

end for

#### 4.3. Setting of parameters

In the end, the computational method is governed actually by three

parameters: the number of iterations  $T$ , the noise variance  $\sigma_\varepsilon^2$  and the scaling parameter  $\beta$ . At each iteration of the procedure, we consider that any state-of-the-art algorithm produces a satisfyingly smooth image, that is an image for which the total variation is lower than  $\lambda$ . In practice, it is not required to accurately set this parameter  $\lambda$ . At each iteration  $t$ , the denoising algorithm removes the perturbation noise. The parameter  $\beta$  affects the acceptance rate of the evaluation step. The higher the value of  $\beta$ , the higher the acceptance rate. For a high enough  $\beta$  value, all proposed samples are accepted and we fall back on the original method (Maggioni et al., 2013).

Unlike Maggioni et al. (2013), we propose a statistical physics and energy minimization framework for MW restoration. In (Maggioni et al., 2013), all candidate samples are accepted at each iteration and used to compute an aggregated estimator. It is worth noting that in (Maggioni et al., 2013), the standard deviation  $\sigma_\varepsilon$  decreases through iterations, and the final estimate is obtained by weighting the samples  $\mathbf{x}^{(t)}$  with weights equal to  $1/\sigma_\varepsilon$  updated at each iteration. Hence, this gives more importance to the last samples. In our case,  $\sigma_\varepsilon$  is constant through iterations, and the weights results naturally from the MCMC sampling which selects the most appropriate generated samples. The most frequent accepted samples have higher weights in the computation of the Monte-Carlo estimator (Eq. 12).

Similar to (Maggioni et al., 2013), we used BM4D as denoising operator in MWR. Actually BM4D uses two denoising steps: Step #1 is a hard thresholding performed in the discrete cosine transform domain; Step #2 is Wiener filtering. To save computing time (about factor 2), we focused on Step #1 in our iterative MWR algorithm. Additional experiments on real cryo-ET data (see Fig. 13) also confirmed that the denoising results were worse after Step #2.

Finally, we add a ‘‘periodic plus smooth image decomposition’’ (Moisan, 2011) operation prior to Fourier transforms, in order to reduce artifacts originating from image borders. Indeed, we noticed that a cross-structure tends to emerge in the restored MW and gets amplified through iterations. This structure is a well-known spectral artifact of the Fourier transform, resulting from the false assumption that the images are periodic signals, when in reality the images rarely have similar opposite borders. Applying this decomposition allows to reduce the cross structure and solves the problem.

## 5. Experimental results

The MW restoration method has been evaluated experimentally on synthetic noisy data by varying the parameters and the components of the MH algorithm. Furthermore, we demonstrate the potential of the method on real cryo-ET data.

### 5.1. Results on synthetic data

In this section, we justify the choice of algorithm parameters, then evaluate robustness to noise, and finally compare our approach to a few other competitive methods. We consider an artificial dataset (Dataset A) which consists of a density map of the 20S proteasome corrupted by additive white Gaussian noise and by applying an artificial MW process, which amounts to setting to zero the Fourier coefficients within an artificial wedge shaped mask. Given the ground-truth  $\mathbf{x}$ , we use two similarity measures for quantitatively evaluating the restoration results  $\hat{\mathbf{x}}$ : the peak signal-to-noise ratio (PSNR) and the constrained correlation coefficient (CCC), defined as

$$\text{PSNR} \left( \mathbf{x}, \hat{\mathbf{x}} \right) = \frac{\max_{s \in S} x(s)}{\frac{1}{n} \sum_{s \in S} (x(s) - \hat{x}(s))^2}, \quad (22)$$

$$\text{CCC} \left( \mathbf{x}, \hat{\mathbf{x}} \right) = \frac{\sum_{k \in W} \mathcal{F}\mathbf{x}[k] \mathcal{F}\hat{\mathbf{x}}[k]^*}{\sqrt{\sum_{k \in W} \mathcal{F}\mathbf{x}[k]^2 \sum_{k \in W} \mathcal{F}\hat{\mathbf{x}}[k]^2}} \quad (23)$$

The PSNR is a common score in image denoising, and is well adapted to estimate the global quality of processed images. CCC is a score used in cryo-ET for sub-tomogram averaging (Förster and Hegerl, 2007). This score is very similar to the Pearson's correlation coefficient measured in Fourier domain. Note that only a constrained region of the Fourier domain is considered for CCC computation. In our work, we use CCC to quantify the quality of retrieved Fourier coefficients located in the region  $W$ .

### 5.1.1. Analysis of performance of denoising algorithms

First, we study the influence of several denoising algorithms embedded in our MCMC framework. We focus on three 2D methods: BM3D (Dabov et al., 2007), non-local Bayes (NL-Bayes) (Lebrun et al., 2013), total variation (TV) (Rudin et al., 1992). Since it was not possible to extend all denoisers in 3D, we applied the method on a 2D slice of the 3D volume for assessment. As shown in Fig. 3, the best results have been obtained by using BM3D, both in terms of PSNR and visual quality. Nevertheless, NL-Bayes produced very similar results. TV denoising produced noticeable worse results. It turns out that the performance strongly depends on the ability of the denoising algorithm to remove the Gaussian noise. This experiment confirms that our MCMC procedure achieves better results than traditional denoising algorithms. In addition, the best results are obtained with the BM3D and NL-Bayes algorithms embedded in our MH algorithm. This is consistent with the literature in image denoising. Finally, we confirm that any image denoising algorithm (we also tested NL-means (Buades et al., 2005) PEWA (Kervrann, 2014), OWE (Jin et al., 2017)) allow us to produce a restored image with less MW artifacts. Because faster and very performant in terms of PSNR values, we used BM4D (Maggioni et al., 2013) which a 3D extension of BM3D, for processing 3D cryo-ET data.

### 5.1.2. Acceptance rate of MH algorithm

In this section, we evaluate the sensitivity of the parameter  $\beta$  controlling the acceptance rate of the MH procedure (combined with BM4D). In Fig. 4(left), it is confirmed that the acceptance rate affects the convergence speed. Nevertheless, whatever the parameter  $\beta$  chosen in the range  $[1.5 - 4.0] \times 10^{-5}$ , the algorithm provides solutions with a similar PSNR value close to 34 dB. Note that we got similar reconstructed images by uniformly aggregating all the samples or by aggregating samples with weights equal to the exponential form of the data fidelity term.

In theory, the recommended acceptance rate is about 0.234 (Breyer and Roberts, 2000) in the MH algorithm if we consider a Gaussian proposal distribution. In our case, we get faster convergence since the maximum acceptance rate is about 70%, suggesting that the set of proposed samples are relevant.

In what follows, we set  $\beta = 4.0 \times 10^{-5}$  since it provides faster convergence as shown in Fig. 4(left).

### 5.1.3. Data-fidelity terms

We have tested several data-fidelity terms  $D(\mathbf{x}, \mathbf{y})$ , corresponding to PSNR (see (22)),  $L_1$  and  $L_2$  norms defined as

$$D_{L_1}(\mathbf{x}, \mathbf{y}) = \sum_{s \in S} |x(s) - y(s)|, \quad (24)$$

$$D_{L_2}(\mathbf{x}, \mathbf{y}) = \sum_{s \in S} (x(s) - y(s))^2, \quad (25)$$

and the Pearson's correlation coefficient (CC):

$$D_{\text{CC}}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{s \in S} (x(s) - \mu_x)(y(s) - \mu_y)}{\sqrt{\sum_{s \in S} (x(s) - \mu_x)^2} \sqrt{\sum_{s \in S} (y(s) - \mu_y)^2}}, \quad (26)$$

where  $\mu_x$  and  $\mu_y$  are the means of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively. We have also evaluated a data-fidelity term based on the mutual information (MI):

$$D_{\text{MI}}(\mathbf{x}, \mathbf{y}) = \sum_{i,j} p_{xy}(i,j) \log \frac{p_{xy}(i,j)}{p_x(i)p_y(j)}, \quad (27)$$

where  $p_{xy}(i,j)$  is the joint probability function of  $\mathbf{x}$  and  $\mathbf{y}$  with intensity bins  $i$  and  $j$ , and  $p_x(i)$  and  $p_y(j)$  denote the marginal probability distribution functions of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively. In our implementation, we approximate the probability functions by histograms of pixel values.

It turns out that the resulting images are very similar for all these data-fidelity terms (see Fig. 4(right)). We observed a maximum error of 0.1 db between the final restored images. In the sequel, we decided to focus on the  $D_{L_2}$  data-fidelity term to evaluate the components of the MH algorithm.

### 5.1.4. Spectral analysis of MW

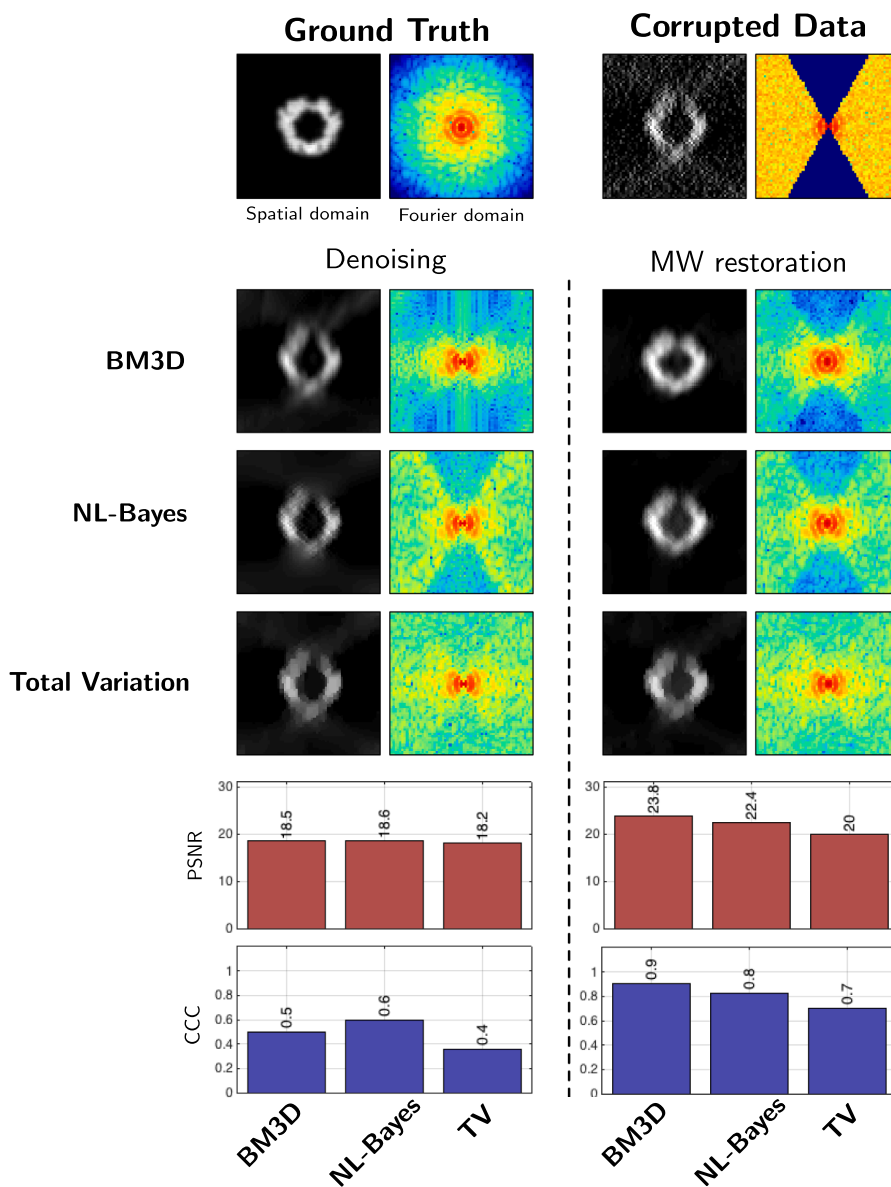
The MW has different shapes if we consider different reciprocal spaces as illustrated in Fig. 5. In our framework, any spectral transform, provided that the transform allows us to decompose the spectral domain into connected components, that is two regions corresponding to non-zero and zero coefficients. The wavelet transform is typically not appropriate as shown in Fig. 5. The MW region should be as small as possible to make restoration successful. Accordingly, we investigated several spectral transforms to improve image restoration with our approach (see Fig. 6). The best result is achieved with the discrete fast Fourier transform (FFT), followed very closely by the discrete cosine transform (DCT). The pseudo-polar fast Fourier transform (PP-FFT), already considered in cryo-ET (Miao et al., 2005), achieves a worse result, both visually and in terms of PSNR values. Finally, it turns out that the performance of our method is not impacted by the transform type, but is actually influenced by the potential precision of the inverse transform. When evaluating the implementations of the considered three transforms, the resulting mean squared errors are in the range of  $10^{-34}$  for DFT,  $10^{-32}$  for DCT and  $10^{-12}$  for PP-DFT. These errors perfectly correlate with the performance given in Fig. 6. Therefore, similar results could be achieved with another transform, provided that the inverse transform is precise enough.

### 5.1.5. Comparing MAP and MMSE estimators

Both the MAP and MMSE have been used considered for solving image restoration problems. For instance, in (Louchet et al., 2008; Louchet and Moisan, 2013), the authors compared TV-MAP and TV-LSE and consider the TV norm as a priori to encourage piece-wise constant images. TV-MAP is more appropriate for restoring piecewise constant images (e.g. cartoons, shepp-Logan phantom), but washes out textures present in natural images and in microscopy images. To reduce stair-casing artifacts produced by TV-MAP, Louchet and Moisan (2013) proposed the TV-LSE estimator to favor smooth transitions between contrasted regions instead of sharp ones.

In our modeling framework, we do not directly use TV as in (Louchet et al., 2008; Louchet and Moisan, 2013). The TV constraint is actually used to a priori discard irrelevant samples during the proposal step in the MCMC sampling procedure. Nevertheless, we can compare the performance MMSE estimator to the MAP estimator corresponding to the most frequent sample in the MCMC sequence. At first glance, both estimates look visually similar, even though the MAP estimate produces more background noise (see Fig. 7). The difference is more noticeable in the spectral domain: the MW of the MAP estimate contains higher amplitudes in the high frequencies. However, this does not mean





**Fig. 3.** Dataset A (2D): comparison of denoising algorithms. First, we compare conventional denoisers (left column) to our MH method (right column). Second, we compare the results obtained by applying three different 2D denoising algorithms: BM3D, NL-Bayes, and ROF denoising. On the bottom, we evaluate the performance in terms of PSNR and CCC values. It turns out that our MH method performs better than conventional denoisers in all situations.

that the MW restoration is of better quality. Indeed, according to the PSNR values, MMSE is closer to the ground-truth than any generated sample. Therefore the higher MW amplitudes in the MAP estimate most likely carry noise rather than meaningful information.

#### 5.1.6. Robustness to noise, comparison to BM4D (Maggioni et al., 2013)

From the results on dataset A (see Fig. 8(a)), it can be seen how well our method works in the absence of noise ( $\sigma_n = 0$ ): a quasi perfect image recovery has been achieved, despite the complexity of the object. Increasing the amount of noise deteriorates the performance, but as it can be observed for  $\sigma_n = 0.2$ , the result is still satisfying. For high amounts of noise ( $\sigma_n = 0.4$ ), the object contrast is still greatly enhanced but the MW artifacts could not be completely removed. In Fourier domain (Fig. 8(b)), the MW has been filled up completely when  $\sigma_n = 0$ , whereas for an increasing amount of noise the MW reconstruction is increasingly restrained to the low frequencies. This is because high frequency components of a signal are more affected by noise, which makes them more difficult to recover. In Fig. 8(c), the evolution of the PSNR over time shows that the method converges to a stable solution.

In Fig. 8(d), we compare our MH method to the method proposed by Maggioni et al. (2013). Both methods produce visually identical results in spatial domain, as well as in the spectral domain, as confirmed by the final PSNR values. However, the difference lies in the convergence speed: our method takes about half as long as the original method (Maggioni et al., 2013). Even though the synthetic dataset A is a simplified case of data corruption in cryo-ET, it gives a good intuition of the performance of our method.

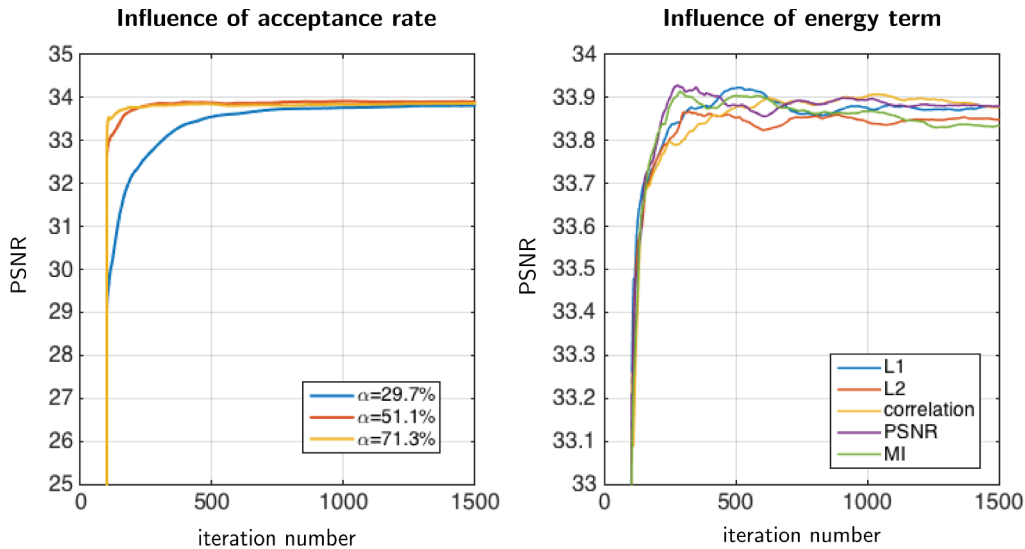
#### 5.1.7. Comparison to other MW restoration methods

We compare our results to those produced by three competing methods (see Fig. 9): sMAPEM (Paavola et al., 2014), BFLY (Kováčik et al., 2014), and a TV method with spectral constraints (Moisan, 2001), each adopting a different strategy to reduce MW artifacts. We implemented the Moisan's method as follows:

**Our implementation of the Moisan (2001)'s method:**

Set an initial state  $\mathbf{x}^{(0)} \in \Gamma$ .

**For**  $t = 1, \dots, T$  **do**



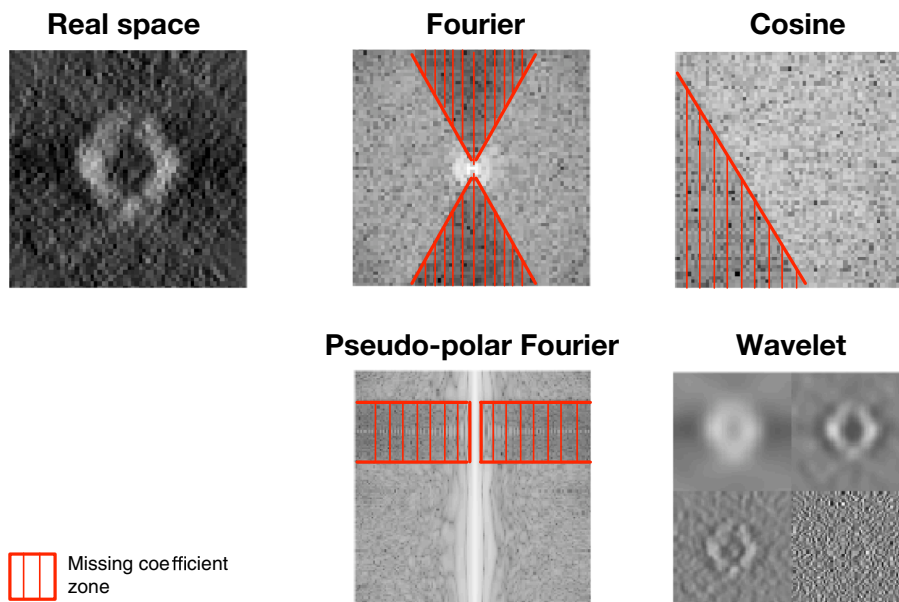
**Fig. 4.** Dataset A: This figure shows the impact of algorithm parameters. We illustrate these effects in terms of PSNR through iterations. We do not show images because obtained results are visually similar. On the left, the influence of parameter  $\beta$  controlling the acceptance rate of the MH sampling, is shown:  $\beta = 1.5 \times 10^{-5}$  (blue),  $\beta = 2.0 \times 10^{-5}$  (red),  $\beta = 4.0 \times 10^{-5}$  (yellow). Clearly, the choice of  $\beta$  affects the convergence speed, however in all cases our method converges to the same result. On the right, the influence of the data fidelity term is presented. We compare the  $L_1$  (24) and  $L_2$  norms (25), the correlation coefficient (26), the PSNR (22), and the mutual information (27). The results are very similar (maximum error of 0.1 dB between restored images). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1. **Projection:**  $\mathbf{z}^{(t)} = \mathcal{P}_W(\mathbf{x}^{(t)})$
2. **Denoising:**  $\min_{\mathbf{x}} \|\mathbf{z}^{(t)} - \mathbf{x}\|^2 + \lambda \|\nabla \mathbf{x}\|_1$

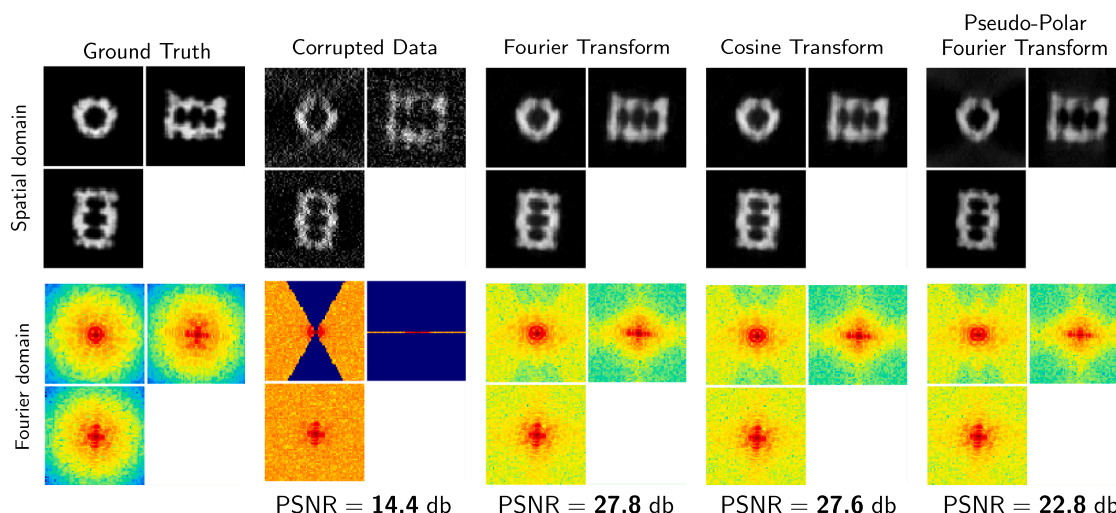
end for which amounts to alternatively minimizing TV (Rudin et al., 1992) and satisfying the spectral constraints. The sMAPEM algorithm (Paavolainen et al., 2014) is an iterative tomogram reconstruction procedure, designed to reduce MW artifacts and achieve isotropic resolution. In our experiments, we performed the comparison in 2D mainly because sMAPEM is not available in 3D. Our method, BFLY and the TV method operate on tomograms (2D in our case), while sMAPEM operates on projections (1D). These projections were obtained from the dataset A ground-truth, with a tilt-range of  $-60^\circ$ – $60^\circ$  and a tilt increment of 2 degrees. We then added Gaussian noise ( $\sigma_n = 0.17$ ) to the

projections. In order to make a fair comparison, the inputs should originate from these same projections. Therefore we use the weighted back-projection (WBP (Radermacher, 1992)) algorithm to produce the 2D data needed for the other methods.

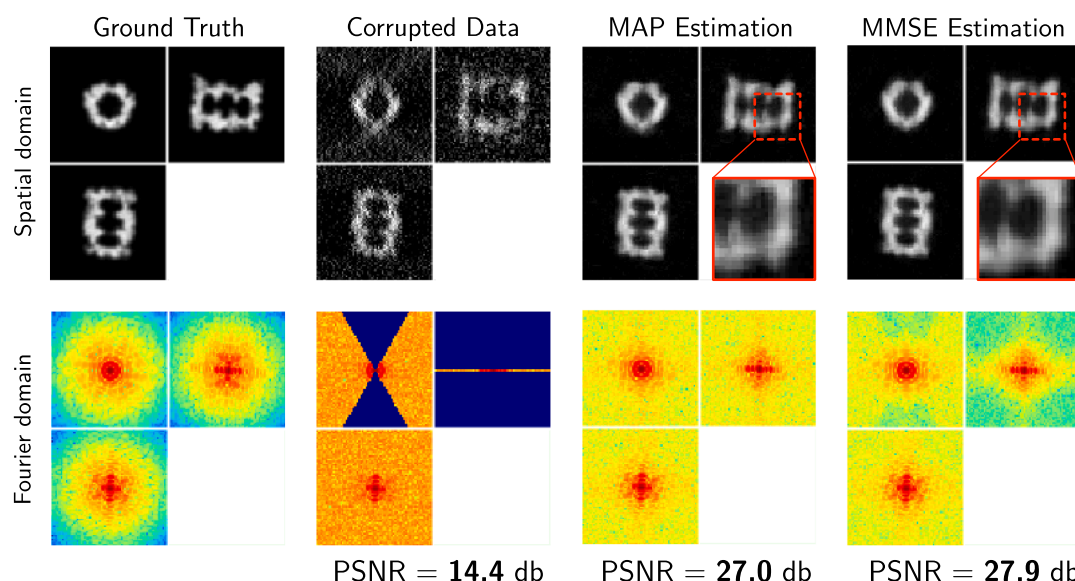
As explained above, the strategy of sMAPEM differs from ours, in the sense that it takes as an input projections and gives as an output a reconstructed tomogram. One may argue that the best strategy would be the one adopted by sMAPEM, because it directly uses the projections, instead of a tomogram which is already contaminated by MW artifacts. However, as shown in Fig. 9, our method achieves a better PSNR value than sMAPEM. Visually both methods seem to approach isotropic resolution, however the result of our method is smoother, while the result of sMAPEM contains pixelated artifacts. The result of the Moisan’s



**Fig. 5.** The missing wedge shape (in red) for different transforms: Fourier transform, cosine transform, and pseudo-polar Fourier transform. Note that the missing wedge is not apparent in all transforms, as is illustrated with the wavelet transform.



**Fig. 6.** Dataset A (3D): influence of the transform type for MW restoration. From left to right: the ground truth (reference for measuring the PSNR), the corrupted image (used as input for the method), and the processed images using the Fourier transform, cosine transform and pseudo-polar Fourier transform. The best result is obtained by using the Fourier transform.



**Fig. 7.** Dataset A (3D): comparison of MAP and MMSE estimators. From left to right: the ground truth (reference for PSNR), the corrupted image, the MAP and MMSE estimators computed with our MCMC procedure. As can be observed in the zoomed in regions (the red frames), the MMSE estimator is less noisy than the MAP estimate, with a higher PSNR value.

method is visually satisfying. In real space, noise appears to be removed, however the result suffers from staircasing artifacts, as well known with TV denoising. In Fourier space, the entire MW appears to be restored, as opposed to our method and sMAPEM where the restoration is constrained to lower frequencies. However according to CCC values, these restored Fourier coefficient do not correlate as well with the ground truth as our method and sMAPEM. The BFLY filter aims at reducing the MW ray-artifacts by smoothing the sharp transition between the MW and the SR. The object elongation and side artifacts however remain. Unlike other competing methods, BFLY does not recover missing Fourier coefficients, it is therefore not surprising that it has the worst performance, both visually and in terms of PSNR and CCC values.

In summary, our approach outperforms competing methods both in terms of PSNR and CCC values. Visually, our approach produces a smoother image, while sMAPEM and the Moisan's method introduce artifacts in the result. The weakest performance is achieved by BFLY, however this was expected as the other methods have much higher complexity. That being said, BFLY is the fastest method.

## 5.2. Results on experimental data

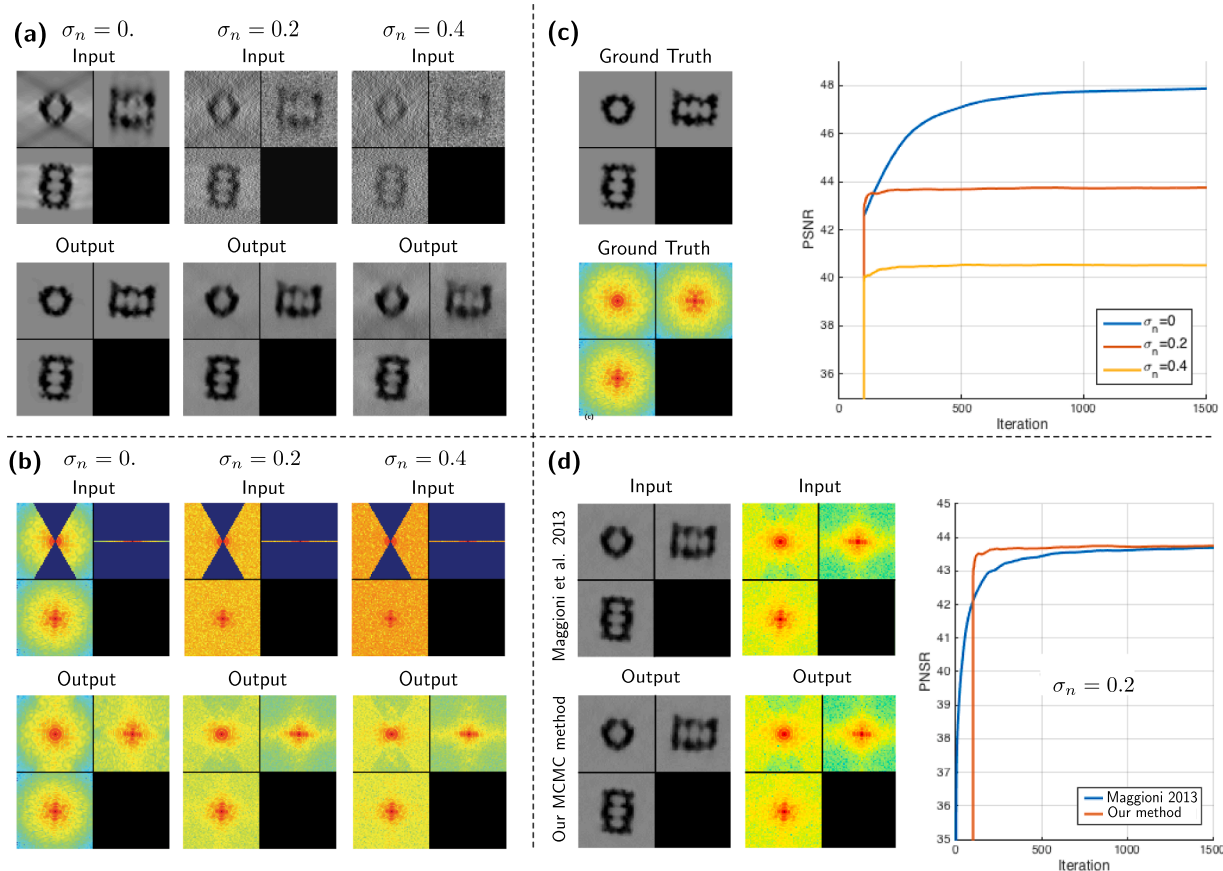
In this section, we evaluate the performance of our MH method (combined with BM4D) on real images to confirm the results obtained on artificial images.

### 5.2.1. Experimental datasets

Three datasets (B, C and D) have been used to evaluate the performance of the method on experimental data. Dataset B is an experimental sub-tomogram containing a gold particle. Dataset C is an experimental sub-tomogram containing 80S ribosomes attached to a membrane. Dataset D is an experimental sub-tomogram depicting a region of an E. Coli bacteria, and contains unidentified macromolecules next to a membrane. Unlike data-sets B and C, the dataset D is available as single-axis and double-axis data (see Section 1).

### 5.2.2. Criteria and method for evaluation

The evaluation differs depending on the dataset. In dataset B, the



**Fig. 8.** Simulated data of the 20S proteasome, for varying amounts of noise (dataset A). All images depict ortho-slices of 3D volumes. The volume size is  $64 \times 64 \times 64$  voxels. For (a) and (b), top row: method inputs, bottom row: method outputs. Results are shown in spatial domain (a) and spectral domain (b). In (c) can be observed the ground-truth and the increase of the PSNR values over iterations. In (d) we compare our method to the original method (Maggioni et al., 2013).

gold particle is deformed and elongated (ellipse) due to the MW artifacts. Improving the sphericity of the object is thus a good evaluation criterion. For dataset C, we measure the similarity between the central ribosome and a reference (obtained via sub-tomogram averaging). The evaluation criterion is the Fourier shell correlation (FSC), commonly used in cryo-ET (Van Heel and Schatz, 2005). In order to measure the quality of the recovered MW only, we also compute the FSC over the MW support (“constrained” FSC or cFSC). For dataset D, we have both single-axis and double-axis versions of the data. A double-axis volume has less missing Fourier coefficients than a single-axis volume. Therefore, when processing the single-axis volume, the additional Fourier coefficients of the double-axis volume can act as an experimental ground truth. We evaluate the result with the CCC score, as illustrated in Fig. 13.

### 5.2.3. Analysis of restoration results

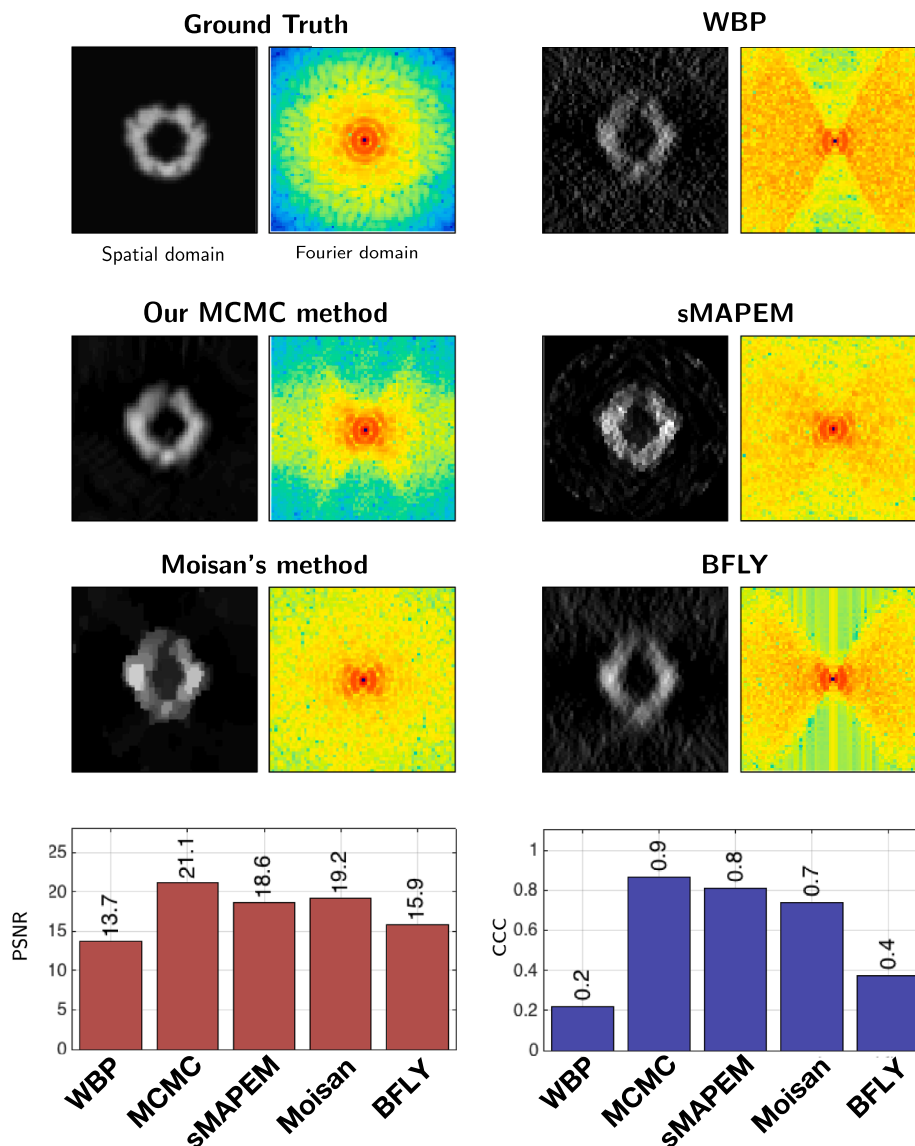
The result on dataset B shows that noise is reduced and a significant part of the MW was recovered (see Fig. 10). Even though the recovery is not complete, it is enough to reduce the MW artifacts while preserving and enhancing image details. The ray and side artifacts induced by the high contrast of the gold particle are reduced and its sphericity has been improved, bringing the image closer to the expected object shape. The result on this dataset shows that the method is able to handle experimental noise in cryo-ET.

The dataset C (ribosomes, Fig. 11) is more challenging because the objects have a more complex shape and less contrast, i.e. the SNR is lower. Nonetheless, the method significantly enhanced the contrast and, according to the FSC criteria, the signal has indeed been improved. Although visually it is more difficult to conclude that the MW artifacts

have been affected, the Fourier spectrum shows that Fourier coefficients were recovered. As expected, the amount of recovered high frequencies is less than for dataset B, because of the lower SNR. It is now necessary to provide a proof that the recovered coefficients carry a coherent signal, therefore the cFSC has been measured. The black curve in Fig. 11 depicts the cFSC between the unprocessed image and the reference: given that the MW contains no information, the curve represents noise correlation. Consequently, everything above the black curve is signal, which is indeed the case for the processed data (red curve in Fig. 11). To illustrate how the method can improve visualization, a simple thresholding has been performed on the data (3D views in Fig. 11). While it is difficult to distinguish objects in the unprocessed data, the shape of ribosomes become clearly visible and it can be observed how they are attached to the membrane. In addition, in order to demonstrate that our procedure is not limited to ribosomes (considered as easy targets because of their good contrast), we perform the same evaluation procedure (i.e. dataset C) on subtomograms containing proteasomes (see Fig. 12). In both processed proteasome sub-tomograms, the contrast has been enhanced and the FSC and cFSC curves provide proof of an improved signal.

Dataset D has a broader field of view than precedent datasets. As can be observed in Fig. 13, the double-axis volume has a better contrast than the single-axis volume, the reason being that it has more sampled Fourier coefficients, and therefore contains more signal and less noise. For this dataset, we processed the single-axis volume and investigated how close to the double-axis volume the restoration results are. We compare the results obtained with the MWR algorithm to those obtained with BM4D (Step #1 and Step #2). From Fig. 13, MWR tends to enhance contrast, which facilitates the identification of membrane and





**Fig. 9.** Dataset A (2D): comparing our approach to competitive methods: i/sMAPEM, a regularized tomographic reconstruction method designed to achieve isotropic resolution; ii/ the Moisan's method designed to extrapolate missing regions in Fourier space; iii/BFLY, a filter designed to reduce MW artifacts. The sMAPEM method takes projections as input, therefore we used the same projections to produce the 2D input (via WBP) for the other methods. On the bottom we display the PSNR and CCC scores obtained for all tested methods.

macromolecules. Also, the global contrast is higher in the images restored with MWR than those produced by BM4D. Moreover, we can notice that the output of Step #1 ( $D_1$  in Fig. 13) has a better contrast than the output of Step #2 ( $D_2$  in Fig. 13), which is surprising at first glance because Step #2 is generally used to "boost" the result of Step #1. This suggests that the assumptions made in Step #2 (i.e. additive noise and known stationary signal and noise spectra), do not hold after Step #1 in our specific case in cryo-ET. When examining the data in Fourier domain, it is clear that several Fourier coefficients have been partially recovered, and correlate better with the double-axis volume as confirmed by the CCC values (0.29 before and 0.71 after applying MWR). The volumes obtained with BM4D show little (or no) improvement in terms of CCC values (0.39 for  $D_1$  and 0.29 for  $D_2$ ). All these results confirm that our MWR algorithm improves the restoration of MW when compared to conventional denoising algorithms.

## 6. Conclusion

In this paper, we addressed the problem of restoring an image corrupted by missing Fourier coefficients in a well-localized spectral region (missing wedge). We proposed an original Monte-Carlo method to denoise 3D cryo-ET images and compensate MW artifacts. Our algorithm cannot recover unobserved data, but it merely makes the best statistical guess of what the missing data could be, based on what has been observed. Any non-local or patch-based denoiser can be used in our Bayesian framework, and the procedure converges faster than (Maggioni et al., 2013). Our experiments on both synthetic and experimental cryo-ET data demonstrate that even for high amounts of noise, the method is able to enhance the signal. The method performs better if the contrast of the object of interest is high, which is not always the case in cryo-ET.

In terms of complexity (linear with respect to the number of pixels), the timings of MWR are actually 5.5 min (with a non-optimized version of the algorithm (Matlab) with no parallel implementation) if the

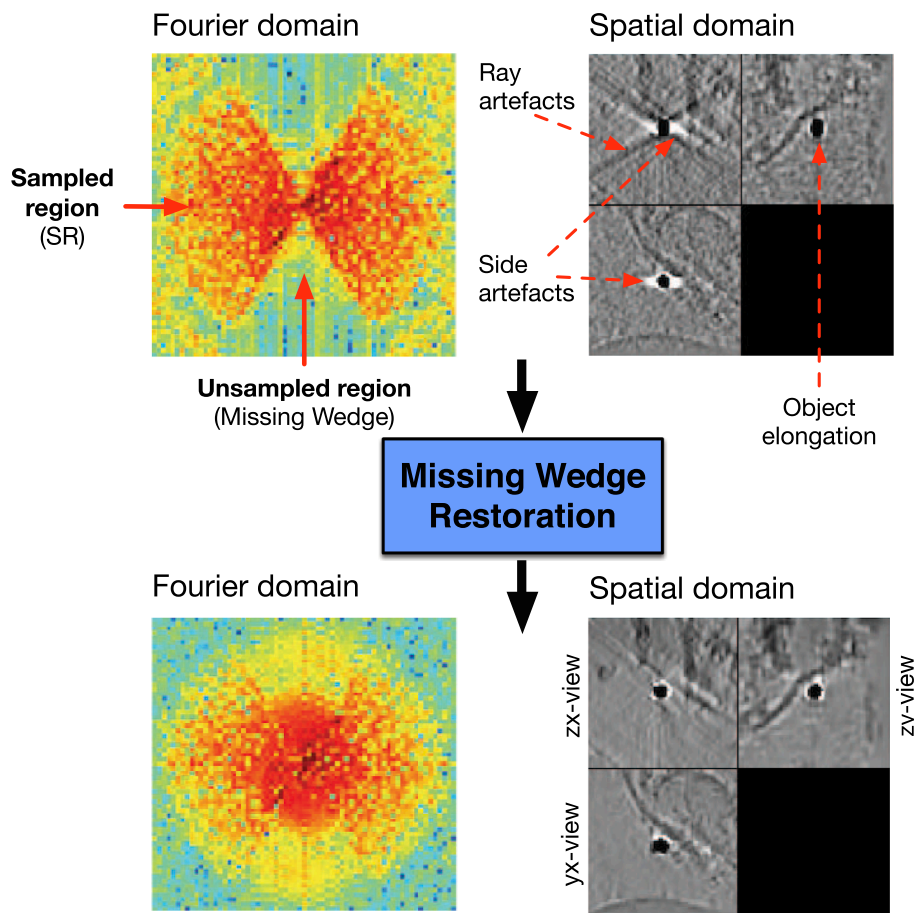


Fig. 10. Experimental sub-tomogram ( $61 \times 61 \times 61$  voxels) containing a gold particle (dataset B). The top row shows the input in the spectral and spatial domains, the bottom row shows the restored image and spectrum.

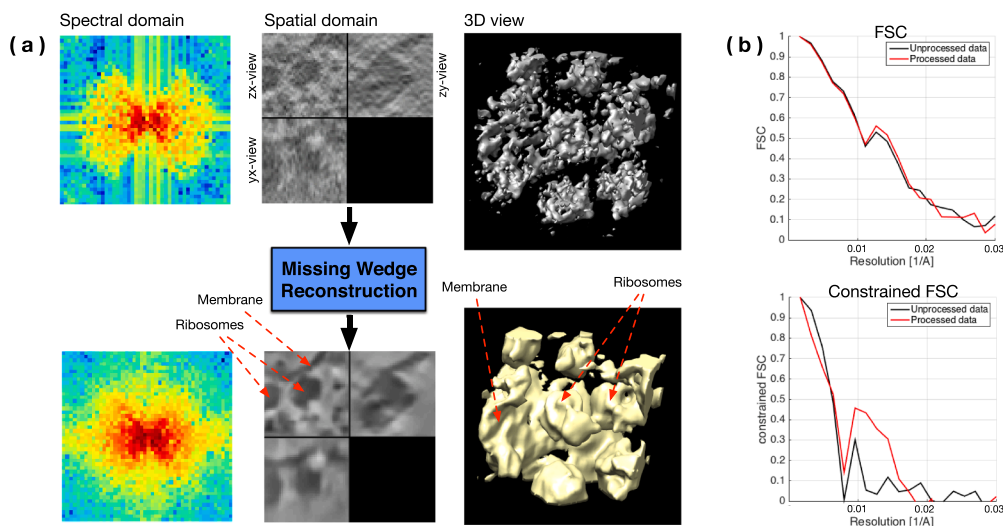
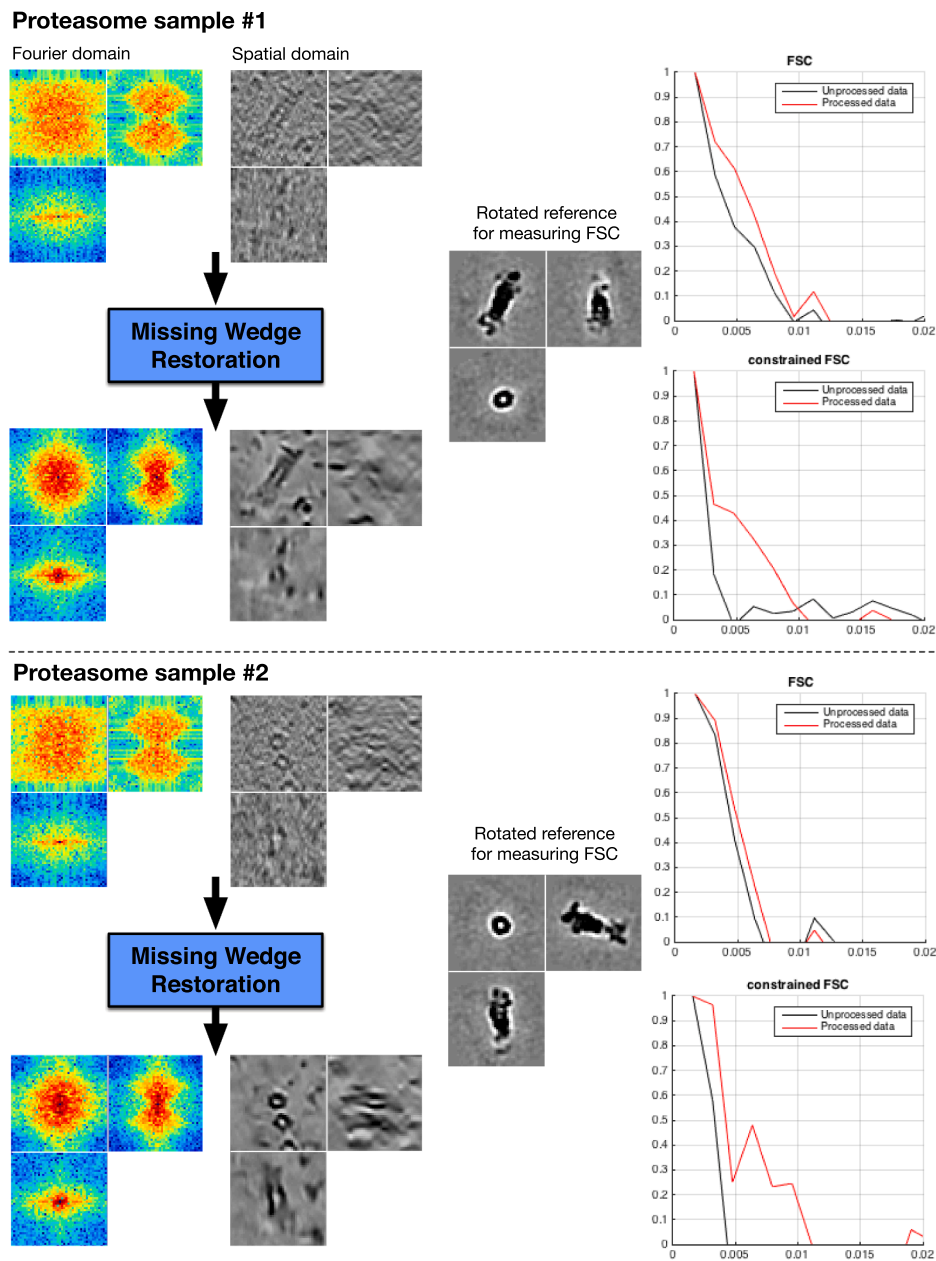


Fig. 11. Experimental sub-tomogram ( $46 \times 46 \times 46$  voxels) containing ribosomes attached to a membrane (dataset C). (a) Top row: input image in spectral domain, spatial domain and 3D view of the thresholded data. Bottom row: the same representations for the output. (b) FSC and cFSC measures of the method input (in black) and output (in red). The FSC measures overall quality, while the cFSC measures quality of recovered Fourier coefficients only (i.e. MW). All measures are wrt the same reference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

number of iterations is set to 500. However, we usually get very similar results after 100 iterations (see Fig. 4), that is 1 min for processing a  $64 \times 64 \times 64$  voxels image. To improve computation, several investigations can be performed. We suggest to handle in parallel and fuse several shorter Markov chains (see Louchet et al., 2008; Louchet and

Moisan, 2013). Hence, we can exploit multithreading since the Markov chains are independent. At the end, we can expect a gain of factor 10 if we consider only 100 iterations and several Markov chains. For future work, a GPU implementation of the algorithm can be also investigated to process larger volumes.



**Fig. 12.** Two experimental sub-tomograms ( $46 \times 46 \times 46$  voxels) containing proteasomes. Data is displayed in both Fourier and spatial domains. We evaluate the result with FSC and cFSC measures of the method input (in black) and output (in red). The FSC measures overall quality, while the cFSC measures quality of recovered Fourier coefficients only (i.e. MW). The reference has been obtained via subtomogram averaging of 2949 proteasomes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Software** In terms of computational performance, MWR takes 5 min and 30 s (0.66 s/iteration,  $T = 500$  samples) on a standard volume of  $64 \times 64 \times 64$  voxels on a Macbook Pro equipped with 2.9 Ghz Intel Core i7, 16 Gb of RAM and the Mac OS X v.10.12.3 operating system. The computing time increases linearly with the number of voxels. The MRW software (Matlab code) can be downloaded from the Git-Hub website: <https://gitlab.inria.fr/serpico/mwr>.

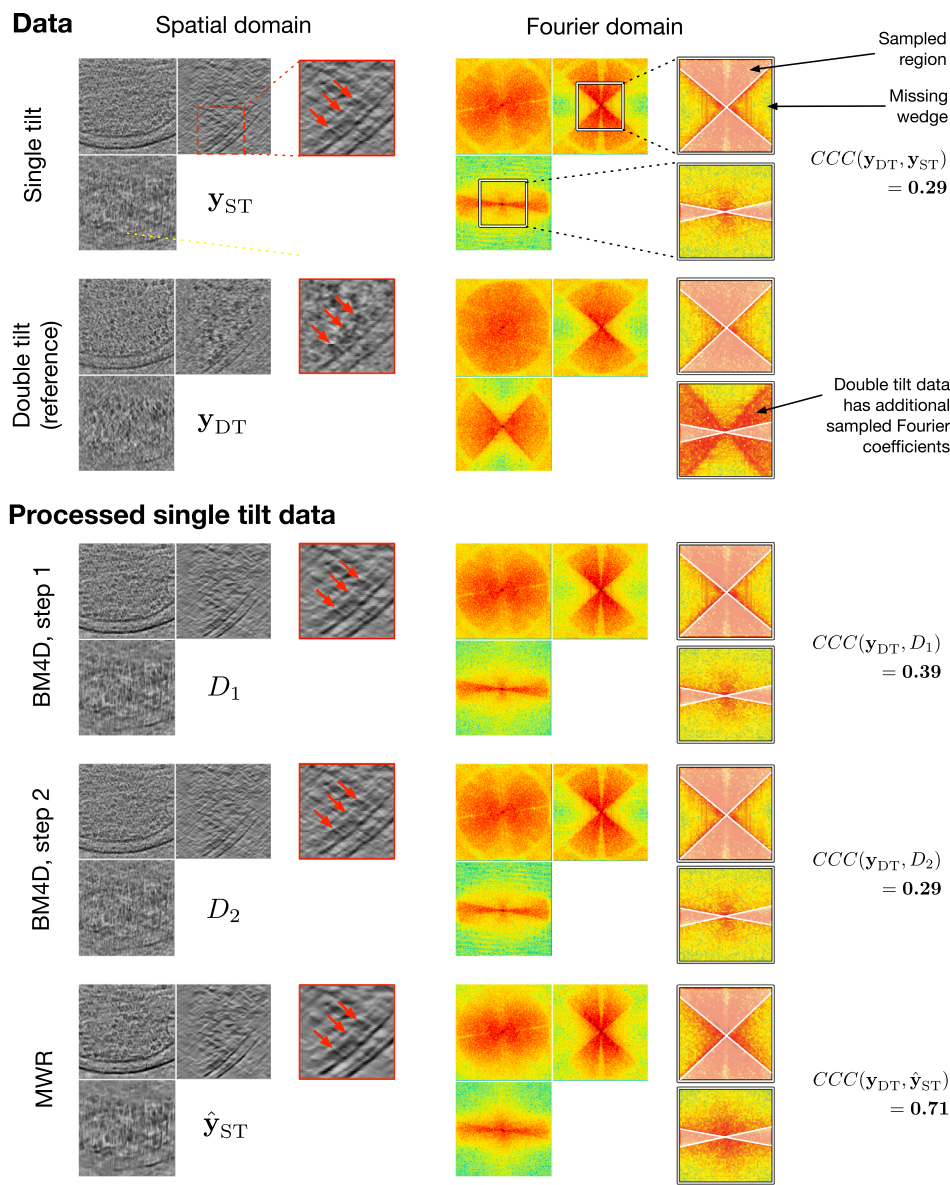
**Data** Datasets B and C have been obtained from tomograms of *Chlamydomonas Reinhardtii* cells (see (Pfeffer et al., 2017; Albert et al., 2017)) for details about data acquisition). Dataset D has been obtained from a tomogram of *E. coli* cell, obtained via dual-axis tilt Cryo-ET on FIB-milled samples, and acquired by J. Ortiz (Max-Planck Institute,

Chemistry Department, Martinsried, Germany).

**Author contributions** C.K. conceived the project and designed the research. E.M. developed the algorithm, implemented the code, analyzed and interpreted data. C.K. and E.M wrote the paper, agreed to all the contents and agreed the submission.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



**Fig. 13.** Dataset D: Experimental double-axis sub-tomogram ( $128 \times 128 \times 128$  voxels) containing multiple macromolecules (see red arrows). We process the single-axis version of the volume ( $y_{ST}$ ) and compare to the double-axis volume ( $y_{DT}$ ), which acts as a ground truth. We compare the results obtained with BM4D (Step #1 and Step #2) and MWR. All volumes are displayed in spatial domain (left column) and Fourier domain (right column). The thumbnails provide zoomed-in views of the data. For the thumbnails in Fourier domain (right column), the sampled Fourier region has been bleached, in order to focus the attention on the missing wedge. The processed volumes  $D_1$ ,  $D_2$  and  $\hat{y}_{ST}$  are evaluated by computing the CCC score.  $y_{DT}$  has additional Fourier coefficients yielding signal compared to  $y_{ST}$  (i.e.  $y_{DT}$  has less missing data than  $y_{ST}$ ). Therefore the CCC score is computed over the support of additional Fourier coefficients. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## Acknowledgments

The authors thank the members of the Max-Planck Institute, Biochemistry Department, Martinsried (Germany) for providing the tomograms and for fruitful sessions: W. Baumeister, F. Förter, A. Martinez, J. Ortiz, S. Pfeffer, S. Albert. We also thank D. Larivière et E. Fourmentin (Fourmentin-Guilbert Foundation) who participated to the design of the research project.

**Funding** This work was jointly supported by Fourmentin-Guilbert Foundation, and Région Bretagne (Brittany Council). Experiments on real data (courtesy of MPI Biochemistry, Martinsried, Germany) were performed on the Inria Rennes computing grid facilities partly funded by France-BioImaging infrastructure (French National Research Agency - ANR-10-INBS-04-07, Investments for the future).

## References

- Albert, S., Schaffer, M., Beck, F., Mosalaganti, S., Asano, S., Thomas, H., Plitzko, J., Beck, M., Baumeister, W., Engel, B., 2017. Dissecting the molecular organization of the translocon-associated protein complex. *Proc. Natl. Acad. Sci. U.S.A.* 114, 13726–13731.
- Bierkens, J., 2015. Non-reversible Metropolis-Hastings. *Stat. Comput.* 26, 1213–1228.
- Breyer, L.A., Roberts, G.O., 2000. From metropolis to diffusions: Gibbs states and optimal scaling. *Stoch. Process. Appl.* 90, 181–206.
- Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7, 434–455.
- Buades, A., Coll, B., M, M.J., 2005. A review of image denoising algorithms, with a new one. *SIAM J. Multiscale Model. Simul.* 4, 490–530.
- Buchholz, T., Jordan, M., Pigo, G., Jug, F., 2018. Cryo-care: content-aware image restoration for cryo-transmission electron microscopy data. arXiv: 1810.05420.
- Burger, H.C., Schuler, C.J., Harmeling, S., 2012. Image denoising: can plain neural networks compete with BM3D? In: *IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, Rhode Island, pp. 2392–2399.
- Chambolle, A., Jalalzai, K., 2014. Adapted basis for nonlocal reconstruction of missing spectrum. *SIAM J. Imaging Sci.* 7, 1484–1502.



- Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M., 1997. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.* 6, 298–311.
- Chatterjee, P., Milanfar, P., 2012. Patch-based near-optimal image denoising. *IEEE Trans. Image Process.* 21, 1635–1649.
- Dabov, K., Foi, A., Egiazarian, K., 2007. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.* 16, 145–149.
- Darbon, J., Cunha, A., Chan, F., Osher, S., Jensen, G., 2008. Fast nonlocal filtering applied to electron cryo-microscopy. In: *IEEE Int. Symp. Biomedical Imaging (ISBI)*, Paris, France, pp. 1331–1334.
- Deledalle, C.A., Denis, L., Tupin, F., 2009. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Trans. Image Process.* 18, 2661–2672.
- Deledalle, C.A., Duval, V., Salmon, J., 2012. Non-local methods with shape-adaptive patches (NLM-SAP). *J. Math. Imaging Vis.* 43, 103–120.
- Duval, V., Aujol, J.F., Gousseau, Y., 2011. A bias-variance approach for the nonlocal means. *SIAM J. Imaging Sci.* 4, 760–788.
- Elad, M., Aharon, M., 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* 3736–3745.
- Fernandez, J., Li, S., 2003. An improved algorithm for anisotropic diffusion for denoising tomograms. *J. Struct. Biol.* 144, 152–161.
- Fienu, J., 1982. Phase retrieval algorithms: a comparison. *Appl. Opt.* 21.
- Förster, F., Hegerl, R., 2007. Structure determination *In Situ* by averaging of tomograms. In: *Cell. Electron Microsc.*, vol. 79, pp. 741–767.
- Frangakis, A., Hegerl, R., 2001. Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. *J. Struct. Biol.* 135, 239–250.
- Friel, J., Quinto, E.T., 2013. Characterization and reduction of artifacts in limited angle tomography. *Inverse Prob.* 29.
- Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statist. Science* 7, 457–511.
- Geman, D., Reynolds, G., 1992. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 367–383.
- Gilavert, C., Moussaoui, S., Idier, J., 2015. Efficient Gaussian sampling for solving large-scale inverse problems using MCMC. *IEEE Trans. Signal Process.* 63, 70–80.
- Gilbert, P., 1972. Iterative methods for the three-dimensional reconstruction of an object from projections. *J. Theor. Biol.* 36, 145–149.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1995. *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*.
- Girolami, M., Calderhead, B., 2011. Riemannian manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Soc. Ser. B Stat. Meth.* 73, 123–214.
- Gribonval, R., 2011. Should penalized least squares regression be interpreted as maximum a posteriori estimation? *IEEE Trans. Signal Process.* 2405–2410.
- Guesdon, A., Blestel, S., Kervrann, C., Chrétien, D., 2013. Single versus dual-axis cryo-electron tomography of microtubules assembled *in vitro*: limits and perspectives. *J. Struct. Biol.* 181, 169–178.
- Guichard, F., Malgouyres, F., 1998. Total variation based interpolation. In: *Eur. Signal Processing Conf. (EUSIPCO)*, Island of Rhodes, Greece, pp. 1741–1744.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
- Holden, L., Hauge, R., Holden, M., 2009. Adaptive independent Metropolis-Hastings. *Ann. Appl. Probab.* 19, 395–413.
- Jin, Q., Grama, I., Kervrann, C., Liu, Q., 2017. Nonlocal means and optimal weights for noise removal. *SIAM J. Imaging Sci.* 10, 1878–1920.
- Katkovnik, V., Foi, A., Egiazarian, K., Astola, J., 2010. From local kernel to nonlocal multiple-model image denoising. *Int. J. Comput. Vis.* 86, 1–32.
- Kazerouni, A., Kamilov, U.S., Bostan, E., Unser, M., 2013. Bayesian denoising: from MAP to MMSE using consistent cycle spinning. *IEEE Signal Process. Lett.* 249–252.
- Kervrann, C., 2014. PEWA: Patch-based exponentially weighted aggregation for image denoising. In: *Neural Inf. Process. Systems (NIPS)*, Montreal, Canada, pp. 1–9.
- Kervrann, C., Boulanger, J., 2006. Optimal spatial adaptation for patch-based image denoising. *IEEE Trans. Image Process.* 2866–2878.
- Kervrann, C., Boulanger, J., 2007. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In: *Int. Conf. Scale Space (SSVM)*, Ischia, Italy, pp. 520–532.
- Kervrann, C., Boulanger, J., 2008. Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *Int. J. Comput. Vis.* 79, 45–69.
- Kervrann, C., Roudot, P., Waharte, F., 2008. Approximate Bayesian computation, stochastic algorithms and non-local means for complex noise models. In: *IEEE Int. Conf. Image Processing (ICIP)*, Paris, France, pp. 2834–2838.
- Kervrann, C., Roudot, P., Waharte, F., 2014. Approximate Bayesian computation, stochastic algorithms and non-local means for complex noise models. In: *IEEE Int. Conf. Image Processing (ICIP)*, Paris, France, pp. 2834–2838.
- Kindermann, S., Osher, S., Jones, P.W., 2005. Deblurring and denoising of images by nonlocal functionals. *SIAM J. Multiscale Model. Simul.* 4, 1091–1115.
- Kováčik, L., Kerešič, S., Höög, J.L., Juda, P., Matula, P., Raška, I., 2014. A simple Fourier filter for suppression of the missing wedge ray artefacts in single-axis electron tomographic reconstructions. *J. Struct. Biol.* 186, 141–152.
- Kréme, M., Emiva, V., Chau, C., 2018. Phase reconstruction for time-frequency inpainting. In: *Int. Conf. LVA/ICA*, Guildford, UK, pp. 417–426.
- Lauze, F., Quéau, Y., Plenge, E., 2017. Simultaneous reconstruction and segmentation of CT scans with shadowed data. In: *Int. Conf. Scale Space (SSVM)*, Kolding, Denmark, pp. 308–319.
- Leary, R., Saggi, Z., Midgley, P.A., Holland, D.J., 2013. Compressed sensing electron tomography. *Ultramicroscopy* 131, 70–91.
- Lebrun, M., Buades, A., Morel, J., 2013. A nonlocal Bayesian image denoising algorithm. *SIAM J. Imaging Sci.* 6, 1665–1688.
- Liang, F., Liu, C., Carroll, R.J., 2010. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. Wiley Series Computational Statistics, England.
- Lou, Y., Zhang, X., Osher, S., Bertozzi, A., 2010. Image recovery via nonlocal operators. *J. Sci. Comput.* 42, 185–197.
- Louchet, C., Moisan, L., 2008. Total variation denoising using posterior expectation. In: *Eur. Signal Processing Conf. (EUSIPCO)*, Lausanne, Switzerland, pp. 1–5.
- Louchet, C., Moisan, L., 2011. Total variation as a local filter. *SIAM J. Imaging Science* 4, 651–694.
- Louchet, C., Moisan, L., 2013. Posterior expectation of the total variation model: properties and experiments. *SIAM J. Imaging Sci.* 6, 2640–2684.
- Maggioni, M., Katkovnic, V., Egiazarian, K., Foi, A., 2013. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.* 22, 119–133.
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A., 2009. Non-local sparse models for image restoration. In: *IEEE Int. Conf. Comput. Vis. (ICCV)*, Tokyo, Japan, pp. 2272–2279.
- Marnissi, Y., Chouzenoux, E., Benazza-Benyahia, A., Pesquet, J., 2018. An auxiliary variable method for Markov chain Monte Carlo algorithms in high dimension. *Entropy* 20, 1–35.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations. *J. Chem. Phys.* 21, 1087–1092.
- Miao, J., Förster, F., Levi, O., 2005. Equally sloped tomography with oversampling reconstruction. *Physical Rev. B* 72, 3–6.
- Milanfar, P., 2013. A tour of modern image filtering. *IEEE Signal Process. Mag.* 30, 106–128.
- Moisan, L., 2001. Extrapolation de spectre et variation totale pondérée. In: *Colloque Trait. du Signal des Images (GRETSD)*, Juan-les-Pins, France, pp. 892–895.
- Moisan, L., 2011. Periodic plus smooth image decomposition. *J. Math. Imaging Vis.* 39, 161–179.
- Moreno, J., Martinez-Sanchez, A., Martinez, J., Garzon, E., Fernandez, J., 2018. TomoEED: fast edge-enhancing denoising of tomographic volumes. *Bioinformatics* 34, 3776–3778.
- Neal, R.M., 2004. Improving asymptotic variance of MCMC estimators: non-reversible chains are better (Technical Report 0406). arXiv: 0407281v1.
- Paavola, L., Acar, E., Tuna, U., Peltonen, S., Moriya, T., Pan, S., Marjom, V., Cheng, R.H., Ruotsalainen, U., 2014. Compensation of missing wedge effects with sequential statistical reconstruction in electron tomography. *PLoS One* 9, 1–23.
- Papandreou, G., Yuille, A., 2010. Gaussian sampling by local perturbations. In: *Neural Inf. Process. Systems (NIPS)*, Vancouver, Canada, pp. 1858–1866.
- Pfeffer, S., Dudek, J., Schaffer, M., Ng, B.G., Albert, S., Plitzko, J.M., Baumeister, W., Zimmermann, R., Freeze, H.H., Engel, B.D., Förster, F., 2017. Dissecting the molecular organization of the translocon-associated protein complex. *Nat. Commun.* 8, 14516.
- Pizarro, L., Mrázek, P., Didas, S., Grewenig, S., Weickert, J., 2010. Generalised nonlocal image smoothing. *Int. J. Comput. Vis.* 90, 62–87.
- Protter, M., Yavneh, I., Elad, M., 2010. Closed-Form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary. *IEEE Trans. Signal Processing* 58, 3471–3484.
- Radermacher, M., 1992. Weighted back-projection methods. *Electron Tomography* 91–115.
- Robert, C.P., Casella, G., 2004. *Monte Carlo Statistical Methods*. Springer.
- Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* 7, 110–120. <https://doi.org/10.1214/aop/1034625254>.
- Roberts, G.O., Rosenthal, J.S., 2001. Optimal scaling for various Metropolis-Hastings algorithms. *Stat. Sci.* 16, 351–367.
- Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Physica D* 60, 259–268.
- Sentosum, K., Lobato, I., Bladt, E., Zhang, Y., Palenstijn, W., Batenburg, K., Van Dyck, D., Bals, S., 2017. Artifact reduction based on sinogram interpolation for the 3D reconstruction of nanoparticles using electron tomography. *Part. Part. Syst. Charact.*, 1700287.
- Soutour, C., Deledalle, C.A., Aujol, J.F., 2014. Adaptive regularization of the NL-means: application to image and video denoising. *IEEE Trans. Image Process.* 23, 3506–3521.
- Van De Ville, D., Kocher, M., 2009. SURE-based non-local means. *IEEE Signal Process. Lett.* 16, 973–976.
- Van Heel, M., Schatz, M., 2005. Fourier shell correlation threshold criteria. *J. Struct. Biol.* 151, 250–262.
- Wang, Y.Q., Morel, J.M., 2013. SURE guided Gaussian mixture image denoising. *SIAM J. Imaging Sci.* 6, 999–1034.
- Wei, D., Yin, C., 2010. An optimized locally adaptive non-local means denoising filter for cryo-electron microscopy data. *J. Struct. Biol.* 172, 211–218.
- Xu, M., Singla, J., Tocheva, E., Chang, Y., Stevens, R., Jensen, G., Alber, F., 2018. De novo visual proteomics in single cells through pattern mining. under review. arXiv: 1512.09347.
- Yan, R., Venkatakrishnan, S., Liu, J., Bouman, C., Jiang, W., 2019. MBIR: a cryo-ET 3D reconstruction method that effectively minimizes missing wedge artifacts and restores missing information. *J. Struct. Biol.* 206, 183–192.
- Zoran, D., Weiss, Y., 2011. From learning models of natural image patches to whole image restoration. In: *IEEE Int. Conf. Comput. Vis. (ICCV)*, Barcelona, Spain, pp. 479–486.