

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

Emmanuel Moebel, Charles Kervrann

► **To cite this version:**

Emmanuel Moebel, Charles Kervrann. A Monte Carlo framework for noise removal and missing wedge restoration in cryo-electron tomography. 2018. hal-01966821

HAL Id: hal-01966821

<https://hal.inria.fr/hal-01966821>

Preprint submitted on 29 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Emmanuel Moebel, Charles Kervrann

Serpico Project-Team, Inria-Rennes / CNRS-UMR 144
Inria, CNRS, Institut Curie, PSL Research University
Campus Universitaire de Beaulieu
35 042 Rennes Cedex France

Abstract

In this paper, we describe 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 computational method 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 near to atomic resolution three-dimensional views of vitrified samples, 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 (180°) 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”.

This blind spot is observable in the 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 on page 1*). This elongation erases boundaries and makes it difficult to differentiate neighboring features.

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 technique is routinely used in cryo-ET and is known as sub-tomogram averaging [Förster and Hegerl, 2007]. To improve sub-tomogram averaging and compensate the remaining MW artifacts, dedicated tomographic regularization reconstruction algorithms 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 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 then 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. 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.

The remainder of the paper is organized as follows. In the next section, several existing methods for spectrum restoration and non-local and patch-based image denoising are reviewed. In Section 3, we formulate the reconstruction problem as an inverse problem. In Section 4, we recall the general Bayesian approach to derive a MMSE estimator. A Monte-Carlo Markov Chains method based on

the Metropolis-Hastings algorithm are described to compute the underlying high-dimensional integral. In Section 5, we adapt the general Bayesian framework to solve our inverse problem. An original Metropolis-Hastings procedure is presented to explore the large space of admissible solutions and to select relevant samples. Section 6 presents the experimental results obtained on simulated and real data. We illustrate the potential of our method with experiments on real cryo-tomogram images and we compare to other common algorithms.

2 Related work

In this section, 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. A special case consists in extrapolating the band-limited spectrum of an image up to higher frequencies. These two problems are generally formulated as denoising problems with specific reconstruction constraints. Typically, Moisan [Moisan, 2001] and Guichard and Malgouyres [Guichard and Malgouyres, 1998] investigated the Total Variation (TV) minimization of the image to extend the band-limited spectrum of an image. The objective was 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 [?] [?]) 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. In [Maggioni et al., 2013], the authors adapt the concept of BM3D for recovering the missing spectrum applied to MRI imaging with very promising results on real data. In our approach, we also focus on patch-based methods to restore the input image corrupted by noise and non-linear transform. It is worth noting that recent non-local and patch-based methods presented below, have shown to be able to solve a large range of inverse problems with good results [Kindermann et al., 2005, Lou et al., 2010], especially image restoration problems [Kervrann and Boulanger, 2008, Katkovnik et al., 2010, Pizarro et al., 2010, Milanfar, 2013, Sutour et al., 2014].

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], LSSC [Mairal et al., 2009], EPLL [Zoran and Weiss, 2011], NL-Bayes [Lebrun et al., 2013], PEWA [Kervrann, 2014] inspired from the popular N(on)L(ocal)-means [Buades et al., 2005]. The NL-means filter is probably the first non-local approach proposed in the literature. Since, several filters with data-dependent weights have been studied in the continuity of [Buades et al., 2005] for better spatial adaptation: [Kervrann and Boulanger, 2006, Deledalle et al., 2009, Deledalle et al., 2012, Kervrann et al., 2014].

For instance, in [Kervrann and Boulanger, 2007, Van De Ville and Kocher, 2009, Louchet and Moisan, 2011, Duval et al., 2011], the authors focused on the selection of the global bandwidth parameter; in [Jin et al., 2017], the weights and smoothing parameters are locally optimized to minimise the mean square error of the estimator. Surprisingly, several sophisticated NL-means filters and patch-based methods were able to produce state-of-the-art denoising results when the input noisy image constitutes the unique source of information. Typically, the popular BM3D algorithm combines clustering of noisy patches, DCT-based transform and shrinkage operation to achieve the state-of-the-art results for several years [Dabov et al., 2007]. PLOW [Chatterjee and Milanfar, 2012], S-PLE [Wang and Morel, 2013] and NL-Bayes [Lebrun et al., 2013], were able to produce very comparable results. In [Lebrun et al., 2013], the covariances matrices of clustered noisy patches are empirically estimated to compute a Maximum A Posteriori (MAP) or a Minimum-Mean-Squared-Error (MMSE) estimate. The PEWA algorithm, which can be viewed as a generalized NL-means filter, aggregate multiple denoised image computed with basic methods and achieves also the state-of-the-art results.

To complete the brief overview of non-local methods, we mention that a noisy image can be also 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]. For instance, EPLL computes a prior from a mixture of Gaussians trained with a database of clean image patches [Zoran and Weiss, 2011]; denoising is then performed by maximizing the so-called Expected Patch Log Likelihood (EPLL) criteria using an optimization algorithm. In [Burger et al., 2012], a first multi-layer perceptron (MLP) procedure exploiting a training set of noisy and noise-free patches was able to achieve the state-of-the-art performance. While all these methods are attractive and powerful, computation is not always feasible since a very large collection of 3D “clean” patches are required to denoise all patches in the input image. In our study, no library or dictionary of clean patches are available. Accordingly, 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.

Before closing this section, we mention that our approach is inspired from [Maggioni et al., 2013], but exploits any competitive denoising methods for restoring the Fourier coefficients. In summary, the method [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 in in [Maggioni et al., 2013]. It is worth noting the data in [Maggioni et al., 2013] is exclusively synthetic and the input images are corrupted with white Gaussian noise. It remains unclear how the concept

performs on experimental data and non Gaussian noise. To generalise this idea, we propose an 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.

The experimental results are comparable to the method in [Maggioni et al., 2013]. The algorithm is implemented efficiently since all patches can be processed independently.

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

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

where \overline{W} is the sampled spectral region (SR) where the Fourier coefficients are positive and non-zero. The so-called missing wedge W is assumed to be symmetric with respect to the origin as illustrated in Figure 1, and $S = W \cup \overline{W}$. In what follows, we assume that the clean image $\mathcal{F}\mathbf{x}$ is known over the region \overline{W} . Our objective is then to estimate $\mathbf{x} : S \rightarrow \mathbb{R}$ in the whole domain S such that

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

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 φ_W to the spectrum of \mathbf{y} to get an estimator $\hat{\mathbf{x}}$ of \mathbf{x} :

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

In the sequel, we describe a Bayesian approach as conventionally proposed for solving such an inverse problem.

4 Bayesian estimator and Monte Carlo Markov Chain sampling

Solving inverse problems in image processing consists in estimating an unknown image \mathbf{x} given an image \mathbf{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 as follows via 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 probability density function (pdf) and $p(\mathbf{y})$ is the marginal distribution of \mathbf{y} which is in general unknown and not computable.

4.1 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 $C : \Omega \times \Omega \rightarrow \mathbb{R}^+$, 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 .

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

$$\begin{aligned} \hat{\mathbf{x}}_{MAP} &= \arg \max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) \\ &= \arg \min_{\mathbf{x}} \{-\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x})\}, \end{aligned} \quad (6)$$

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

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

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 β 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:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} D(\mathbf{x}, \mathbf{y}) + \phi(\mathbf{x}). \quad (8)$$

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.

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:

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

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

$$\hat{\mathbf{x}}_{MMSE} = \frac{\int_{\mathbb{R}^n} \exp\left(-\frac{U(\mathbf{x}, \mathbf{y})}{\beta}\right) \mathbf{x} d\mathbf{x}}{\int_{\mathbb{R}^n} \exp\left(-\frac{U(\mathbf{x}, \mathbf{y})}{\beta}\right) d\mathbf{x}}. \quad (10)$$

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 [25, 47, 33, 50, 57, 70, 85] [Robert and Casella, 2004] as explained in the next section.

Nevertheless, 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:

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

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 an explicit and dedicated prior discussed in [Protter et al., 2010, Gribonval, 2011, Louchet and Moisan, 2013, Kazerooni 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.

4.2 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 follows

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

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 also 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 $\bar{\mathbf{x}}_T$ decreases with the number of samples:

$$\text{Var}[\bar{\mathbf{x}}_T] = \frac{v^2}{T}, \quad (14)$$

where $v^2 = \text{Var}[\mathbf{X}]$. Finally, $\bar{\mathbf{x}}_T$ is Gaussian distributed due to the central limit theorem: $\bar{\mathbf{x}}_T \sim \mathcal{N}(\hat{\mathbf{x}}_{MMSE}, \frac{v^2}{T})$ 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 discuss the convergence properties of the Metropolis-Hastings algorithm used in our approach to generate a Markov chain with a target stationary distribution π .

4.2.1 Simulating independent samples

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 $\pi(\mathbf{x})$ [53] 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 $T \rightarrow \infty$. Nevertheless, 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 given in (14).

4.2.2 Burn-in phase

Another consequence of the correlation is the burn-in period that the chain requires before converging to the invariant target pdf π . In general, the initial T_b samples are discarded and not included in the computation of the estimator [Robert and Casella, 2004]:

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

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].

4.2.3 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 $q(\mathbf{z}|\mathbf{x})$, $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$ 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

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

- (c) Draw α from a uniform distribution: $\alpha \sim \mathcal{U}(0, 1)$.
- (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 π when t increases [Robert and Casella, 2004]. In general, the MH algorithm satisfies the so-called detailed balance condition:

$$(\mathbf{x}, \mathbf{z}) \in \mathbb{R}^n \times \mathbb{R}^n, p(\mathbf{x}|\mathbf{z})\pi(\mathbf{z}) = p(\mathbf{z}|\mathbf{x})\pi(\mathbf{x}),$$

(the chain is reversible) which is a sufficient condition to guarantee that the chain is ergodic and has π 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 π . In what follows, we mainly focus on the specification of the proposal densities to improve convergence.

4.2.4 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 discuss 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 a 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.
- The proposal is generated in accordance with an approximation of the Langevin diffusion process: $z \sim \mathcal{N}(x + \frac{\delta}{2} \nabla \log \pi(x), \delta)$ for a given small value δ .
- 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. More generally, let us consider the whole history $(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t-1)})$. The new state \mathbf{z} is obtained from the proposal distribution $q(\mathbf{z}|\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(t-1)})$ 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 π . 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.

In summary, the convergence of the chain depends on the proposal density but it is also established that the ideal proposal pdf must as close to 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.

5 A MH algorithm for missing wedge restoration

In this section, we propose an original and statistical approach for MW restoration. First, we consider the following image model

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

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.

The idea is to compute a MMSE estimator defined as

$$\hat{\mathbf{x}}_{MMSE} = \frac{\int_{\mathbb{R}^n} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathbf{x}d\mathbf{x}}{\int_{\mathbb{R}^n} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}}$$

by using a dedicated MH algorithm, where $p(\mathbf{x})$ is the prior used to encourage the solution to be positive and piecewise smooth. More specifically, we assume here that the prior is uniform over the set Γ defined as:

$$\Gamma = \{\mathbf{x} \in \mathbb{R}^n : \forall k \in \overline{W} \quad |\mathcal{F}\mathbf{x}(k) - \mathcal{F}\mathbf{y}(k)| < \delta, \quad (15)$$

$$\forall s \in S, x(s) \geq 0, \text{ and } \|\nabla\mathbf{x}\|_1 < \lambda\}$$

where $\|\nabla\mathbf{x}\|_1 = \sum_{s \in S} |\nabla x(s)|$ is the discrete Total Variation of the image [Rudin et al., 1992]. Here we impose that the Fourier coefficients of the reconstructed image are very similar to the known coefficients of the corrupted image \mathbf{y} . Consequently, the performance of the proposed estimator depends on the pre-specified thresholds λ and δ which controls the size of the space Γ . In practice, these values do not need to be accurately adjusted as discussed below. Hence the MMSE estimator can be reformulated as

$$\hat{\mathbf{x}}_{MMSE} = \frac{\int_{\Gamma} \exp\left(-\frac{D(\mathbf{x}, \mathbf{y})}{\beta}\right) \mathbf{x} d\mathbf{x}}{\int_{\Gamma} \exp\left(-\frac{D(\mathbf{x}, \mathbf{y})}{\beta}\right) d\mathbf{x}}, \quad (16)$$

where the likelihood function is a Gibbs model and $D(\mathbf{x}, \mathbf{y})$ is the data fidelity term defined as the L_2 norm between the corrupted image and the restored image degraded by the operator \mathcal{D}_W , i.e.:

$$D(\mathbf{x}, \mathbf{y}) = \sum_{s \in S} \left(y(s) - \sum_{k \in S} e^{2i\pi k \cdot s} \mathcal{F}\mathbf{x}(k) \right)^2. \quad (17)$$

In our experiments, we have considered alternative data fidelity terms yielding similar performance. Because of the high dimensionality of the problem, we need efficient computational algorithms, and therefore we propose an original sampling approach to compute the MMSE estimator.

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. 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 $[0.25 - 0.6]$. In theoretical studies, it has been shown that $a^* = 0.234$ [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

$$\mathbf{x}' = \mathcal{D}_\lambda(\mathcal{P}_W(\mathbf{x} + \boldsymbol{\varepsilon})), \quad (18)$$

where $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, I_{n \times n} \sigma_\varepsilon^2)$ 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 $\forall k \in \overline{W}$, $\mathcal{F}(\mathbf{x} + \boldsymbol{\varepsilon})(k) = \mathcal{F}\mathbf{y}(k)$ and \mathcal{D}_λ is denoising operator that ensures that the total variation of the denoised image is lower than λ . Consequently, the distribution of increments $\mathbf{x}' - \mathbf{x}$ is not parametric due to the nonlinear operators involved in (18). In the sequel, we only assume that this non-parametric distribution q is approximately symmetric. Even though visualization of the empirical high-dimensional proposal density is not possible, we suggest that (18) tends to produces similar samples (denoised images) concentrated around some empirical mean belonging to Γ , with a few moves quite far away from this mean.

In other words, 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 and Moisan, 2008, Louchet and Moisan, 2013], only one pixel is modified at each iteration to compute the TV-LSE estimator. Our approach is 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 correlation between samples is small. To our knowledge, this is the first time that such MH approach given below, is used in the context of image restoration and inverse problems.

Our dedicated MH algorithm for MW restoration is then as follows:

The Metropolis-Hastings algorithm for MW restoration

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

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 α from a uniform distribution: $\alpha \sim \mathcal{U}(0, 1)$
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

In [Louchet and Moisan, 2008, Louchet and Moisan, 2013], the authors also studied the behavior of the expected approximation error

$$\mathbb{E}[\|\tilde{\mathbf{x}}_T - \hat{\mathbf{x}}_{MMSE}\|^2] = \mathbb{E}[\|\tilde{\mathbf{x}}_T - \mathbb{E}[\tilde{\mathbf{x}}_T]\|^2] + \|\mathbb{E}[\tilde{\mathbf{x}}_T] - \hat{\mathbf{x}}_{MMSE}\|^2$$

decomposed into the sum of two terms: the trace of the covariance matrix (or span) and the squared bias which entails the loss of efficiency of the sampling procedure. If we consider another Markov chain x' defined as x , it is established that [Louchet and Moisan, 2008, Louchet and Moisan, 2013]

$$\mathbb{E}[\|\tilde{x}_T - \tilde{x}'_T\|^2] = 2\mathbb{E}[\|\tilde{x}_T - \mathbb{E}[\tilde{x}_T]\|^2]$$

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.:

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

This suggests that averaging multiple independent Markov chains should provide better estimators. In this case, several denoisers (e.g. BM3D, NL-Bayes) associated to independent chains can be potentially combined to improve performance of MW restoration.

In the end, the method is governed by three parameters: the number of iterations T , the noise variance σ_ϵ^2 and the scaling parameter β . At each iteration t , the denoising algorithm removes the perturbation noise. The parameter β affects the acceptance rate of the evaluation step. The higher the value of β , the higher the acceptance rate. For a high enough β value, all proposed samples are accepted and we fall back on the original method [Maggioni et al., 2013]. This method 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.

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 the standard deviation σ_ϵ decreases through iterations, and the final estimate is obtained by weighting the samples $\mathbf{x}^{(t)}$ with weights equal to $1/\sigma_\epsilon$ updated at each iteration. Hence, this gives more importance to the last samples. In our case, σ_ϵ 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

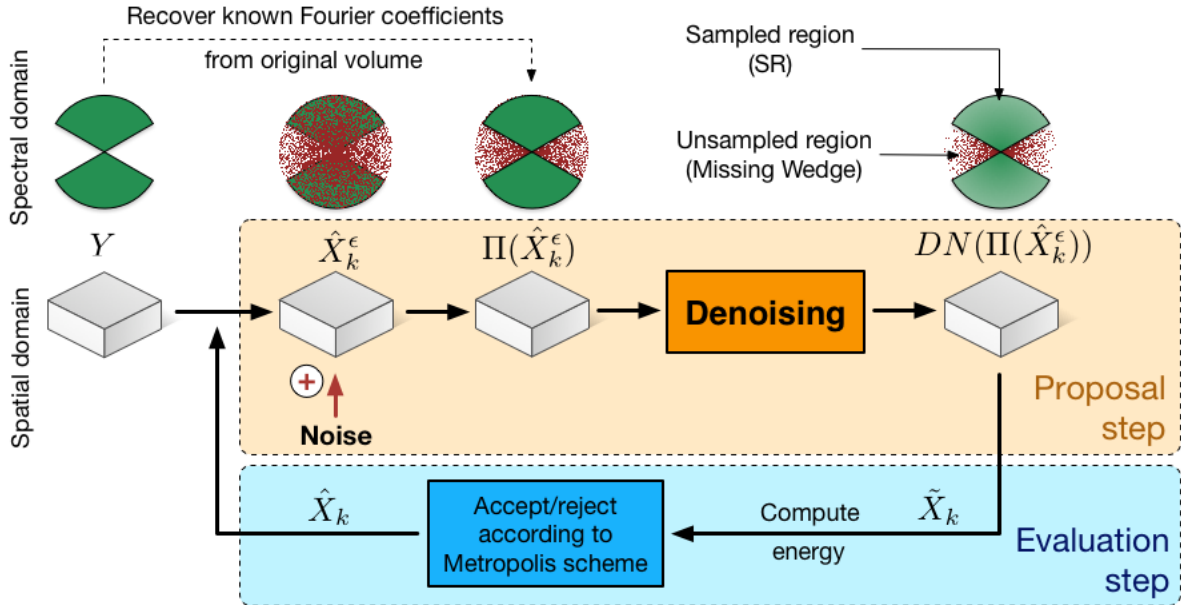


Figure 1: The method flowchart. The 1st icon row represents the data in the spectral domain, the 2nd in the spatial domain.

higher weights in the computation of the Monte-Carlo estimator (12). Finally, we add a “periodic plus smooth image decomposition” [Moisan, 2011] operation at the end of the proposal step. 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.

6 Experimental results

First, our MW restoration method is evaluated on synthetic noisy data by varying the parameters and the components of the MH algorithm. Second, we demonstrate the potential of the method on real cryo-ET data.

6.1 Results on synthetic data

In this section, we first optimize the parameters, then evaluate its 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 x , we use two similarity measures for quantitatively evaluating the restoration results \hat{x} : the peak signal-to-noise ratio (PSNR) and the constrained correlation coefficient (CCC), defined as follows:

$$\text{PSNR}(x, \hat{x}) = \frac{\max(x(s))}{\frac{1}{n} \sum_{s \in S} (x(s) - \hat{x}(s))^2} \quad (19)$$

$$\text{CCC}(x, \hat{x}) = \frac{\sum_{k \in W} \mathcal{F}x(k) \mathcal{F}\hat{x}(k)^*}{\sqrt{\sum_{k \in W} \mathcal{F}x(k)^2 \sum_{k \in W} \mathcal{F}\hat{x}(k)^2}} \quad (20)$$

The PSNR is a common score in image denoising, and is well adapted to estimate the global quality of processed images. The CCC is a score used in cryo-ET for sub-tomogram averaging [Förster and Hegerl, 2007], and it is actually Pearson’s correlation coefficient measured in Fourier domain. In addition, only a constrained region of the Fourier domain is considered for computing the CCC, hence its name. In our work, we use the CCC to estimate the quality of recovered Fourier coefficients, which is why we compute the CCC over W only.

Some of the algorithms we have compared are implemented in 2D only, such as NL-Bayes and sMAPEM. In these cases, we apply 2D processing to ensure a fair comparison (see Figures 2 and 8). In all other cases, we apply 3D processing.

6.1.1 Analysis of performance of denoising algorithms

Here we examine the influence of several denoising algorithms embedded in our MCMC framework: BM3D [Dabov et al., 2007], non-local Bayes (NL-Bayes) [Lebrun et al., 2013], total variation (TV) [Rudin et al., 1992]. Note that all these algorithms are implemented in 2D. Since it was not possible to extend all of them in 3D, we applied the method on a 2D slice of the 3D volume for assessment. As shown in Fig. 2, 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. In this experiment, the best results are obtained with the BM3D and NL-Bayes algorithms embedded in the MH algorithm. This is consistent with the literature in image denoising [Kervrann, 2014]. From this experiment, we can confirm that any image denoising algorithm allow us to produce a restored image with less MW artifacts. Because faster and very performant in terms of PSNR values, we validated BM4D for processing 3D cryo-ET data.

In Figure 2, we also illustrate the results of basic applications of considered denoising algorithms. In all cases, our method achieves better performance. This better performance is attributable to the restoration of the MW, as shown by the clear increase in the CCC score (Eq. 20). This result confirms that our method achieves more than simple denoising, and therefore justifies its use for data corrupted by a MW.

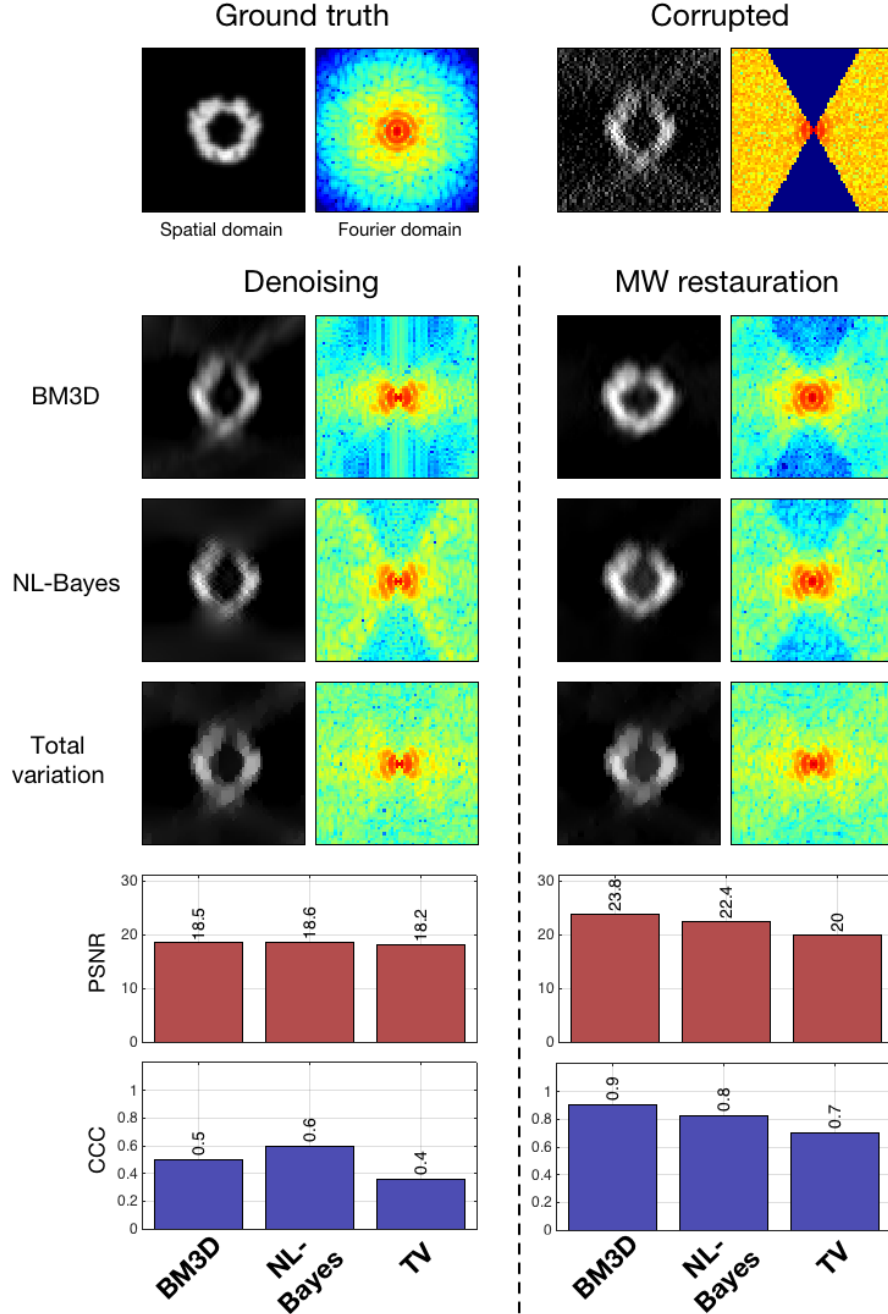


Figure 2: Dataset A (2D): comparing denoising algorithms. In this figure, we make two comparisons. First, we compare simple denoising (left column) to processing by our iterative method (right column). Second, we compare the influence of using different denoising algorithms: BM3D, NL-Bayes, NL-means and ROF denoising. In this figure, all processing is performed in 2D. On the bottom we evaluate the processed images, using two scores: the PSNR and the CCC. As can be observed, our method always performs better than simple denoising, regardless of the denoising algorithm used.

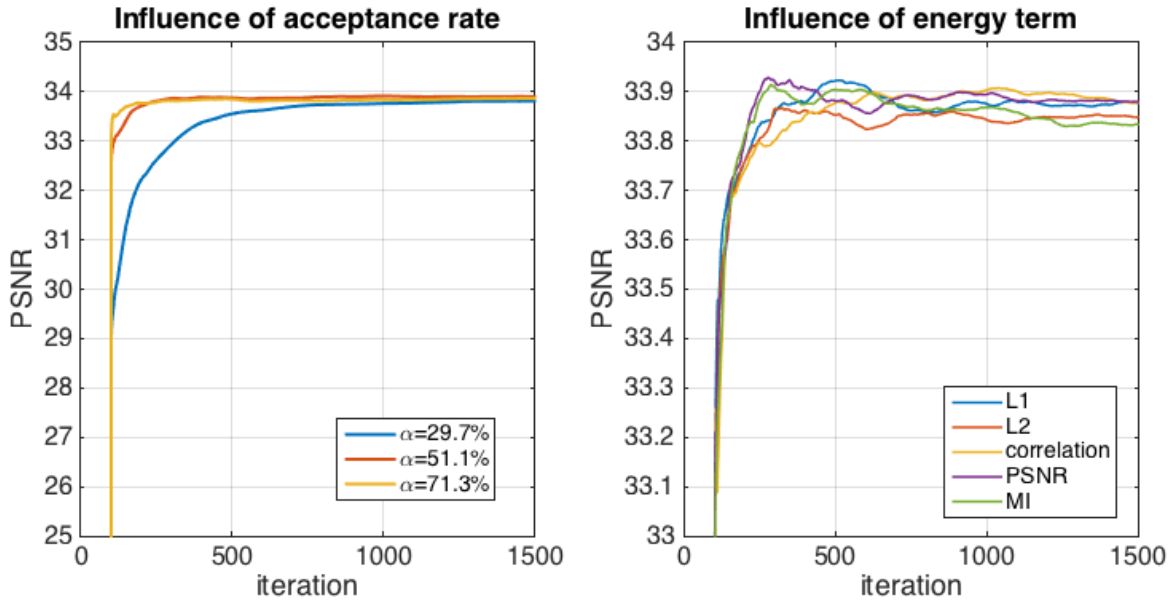


Figure 3: Dataset A: This figure shows the effect that parameters of our method have. We illustrate this effect in terms of PSNR through iterations. We do not show images because obtained results are visually identical. On the left, the influence of parameter β on the performance of our method. β affects the acceptance rate of the Metropolis sampling. As can be observed, the choice of β 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. We have tested the L_1 norm (Eq. 21), L_2 norm (Eq. 22), correlation coefficient (Eq. 23), PSNR (Eq. 19) and mutual information (MI, Eq. 24). Obtained results are very similar (maximum error of 0.1 dB between restored images).

6.1.2 Acceptance rate of MH algorithm

In Fig. 3 (left), it is confirmed that the acceptance rate controlled by the parameter β affects the convergence speed. Nevertheless, whatever the parameter β , the algorithm provides solutions with a similar PSNR value close to 34 dB. We can notice that we get a similar reconstructed image 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%.

6.1.3 Data-fidelity terms

We have tested several data-fidelity terms $D(\mathbf{x}, \mathbf{y})$, including L_1 and L_2 norms and PSNR (see (19)):

$$L_1(\mathbf{x}, \mathbf{y}) = \sum_{s \in S} |x(s) - y(s)|, \quad (21)$$

$$L_2(\mathbf{x}, \mathbf{y}) = \sum_{s \in S} (x(s) - y(s))^2. \quad (22)$$

We have also considered the Pearson's correlation coefficient (CC):

$$\text{CC}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{s \in S} (x(s) - \mu_x)(y(s) - \mu_y)}{\sqrt{\sum_{s \in S} (x(s) - \mu_x)^2} \sqrt{\sum_{s \in S} (y(s) - \mu_y)^2}}, \quad (23)$$

where μ_x and μ_y are the means of \mathbf{x} and \mathbf{y} , respectively, and the mutual information (MI):

$$\text{MI}(\mathbf{x}, \mathbf{y}) = \sum_y \sum_x p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right), \quad (24)$$

where $p(x, y)$ is the joint probability function of \mathbf{x} and \mathbf{y} , and $p(x)$ and $p(y)$ are 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 as shown in Fig. 3 (right). We observed a maximum error of 0.1 db between the final restored images. In the sequel, we decided to focus on the L_2 norm (see 17)) to evaluate the components of the MH algorithm.

6.1.4 Spectral analysis of MW

The MW has different shapes if we consider different reciprocal spaces as illustrated in Fig. 4. 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. 4. 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. 5). 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. 5. Therefore, similar or better results could be achieved with another transform, provided that the inverse transform is precise enough.

6.1.5 Comparing MAP and LSE estimates

Both the MAP and MMSE have been used considered for solving image restoration problems. In [Louchet and Moisan, 2008, Louchet and Moisan, 2013], the authors compared TV-MAP and TV-LSE and consider TV as a prior to encourage piece-wise constant images. TV-MAP is more appropriate for

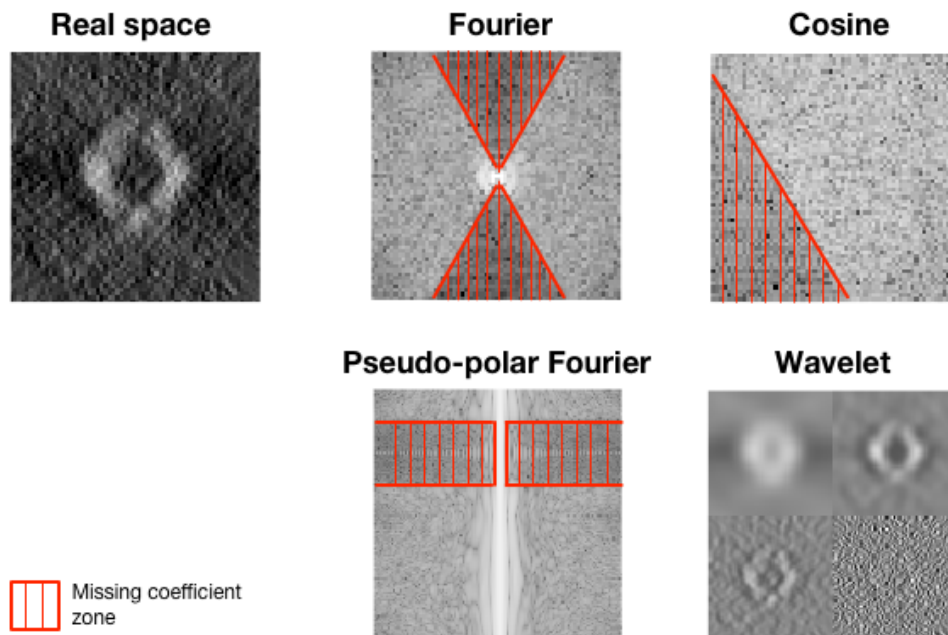


Figure 4: The missing wedge shape (in red) for different transforms: Fourier transform, cosine transform, and pseudo-polar Fourier transform. Note that the missing wedge is not apparent in all transforms, as is illustrated with the wavelet transform.

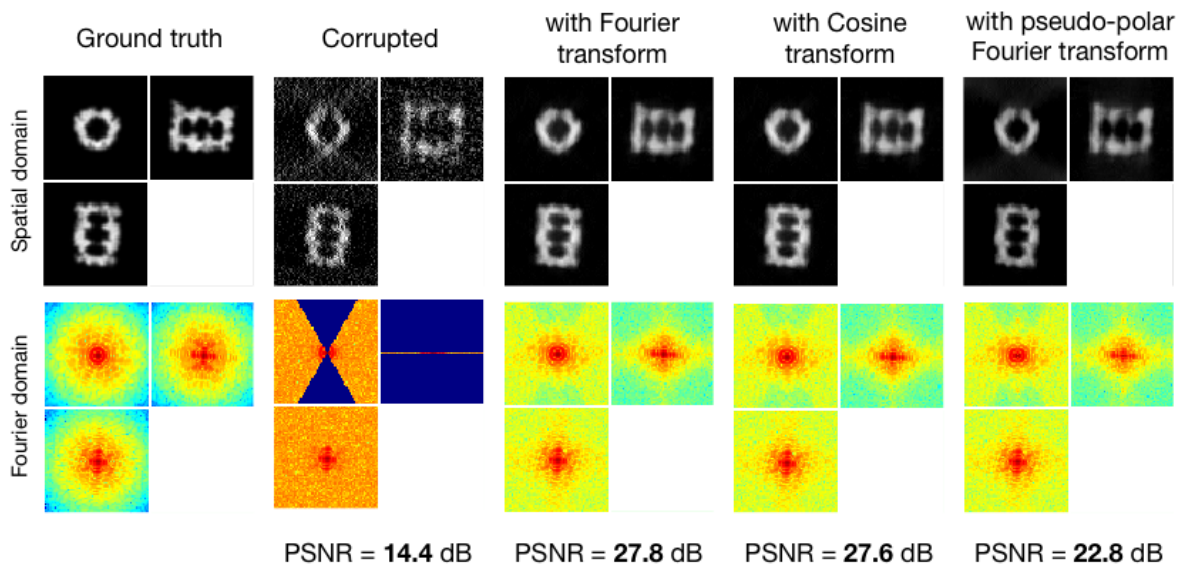


Figure 5: Dataset A (3D): influence of the transform type for the spectral constraint. From left to right: the ground truth (used as reference for measuring the PSNR), the corrupted image (used as input for the method), and the processed images using the Fourier transform, cosine transform and pseudo-polar Fourier transform. The best result is obtained by using the Fourier transform.

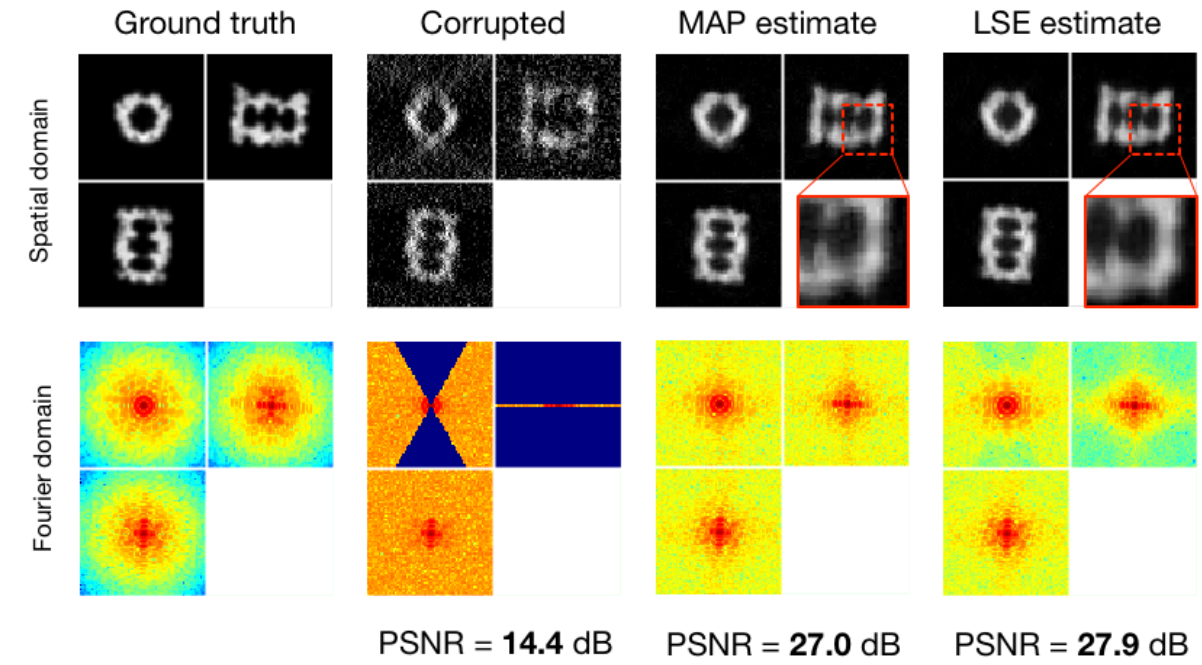


Figure 6: Dataset A (3D): comparing MAP and MMSE estimators. From left to right: the ground truth (reference for PSNR), the corrupted image (method input), the MAP estimate and the MMSE estimate from our MCMC sampling. As can be observed in the zoomed in regions (the red frames), the MMSE estimate is less noisy than the MAP estimate, which is confirmed by its higher PSNR.

restoring piecewise constant images (e.g. cartoons, shepp-Logan phantom), but washes out textures present in natural images and in microscopy images. To overcome stair-casing artifacts of TV-MAP, Louchet and Moisan investigated the TV-LSE estimator. The resulting restored image has smooth transitions instead of sharp ones, and the stair-casing effects are reduced.

In our modeling framework, we do not directly use TV as in [Louchet and Moisan, 2008, Louchet and Moisan, 2013]. The TV constraint is rather used to a priori discard irrelevant samples produced by the MCMC sampling procedure. In what follows, we investigate the performance MMSE and MAP estimators. The MAP estimator corresponds 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. 6). The difference is more noticeable in the spectral domain: the MW of the MAP estimate contains higher amplitudes in the high frequencies. Actually, this does not mean that the MW restoration is more effective; indeed, according to the PSNR values the MMSE estimator 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.

6.1.6 Robustness to noise and comparison to BM4D [Maggioni et al., 2013]

From the results on dataset A (Fig. 7 (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. Let us examine the Fourier domain (Fig. 7 (b)): in the absence of noise, the MW has been filled up completely, 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. 7 (c), the evolution of the PSNR over time shows that the method converges to a stable solution.

In Fig. 7 (d), we compare our MH method to the method [Maggioni et al., 2013]. Both methods produce visually identical results in the spatial domain, as well as in the spectral domain, as confirmed by the final PSNR values. However, the difference lies in 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 method performance.

6.1.7 Comparison to other MW restoration methods

Here we compare our results to those produced by three competing methods (see Figure 8), sMAPEM [Paavolaianen 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 perform the comparison in 2D, as sMAPEM and the TV method are implemented for 2D data. These methods do not all have the same type of input: 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° to 60° and a tilt increment of 2° . We then added Gaussian noise ($\sigma_n = 17\%$) to the projections. In order to make a fair comparison, the other method inputs should originate from these same projections. Therefore we use the weighted back-projection (WBP) algorithm to produce the 2D data needed for the other methods. The WBP algorithm is widely used, fast and straightforward. In the remainder of this Section, we will point out how competing methods differ from ours, and comment obtained results.

sMAPEM [Paavolaianen et al., 2014] is an iterative tomogram reconstruction procedure, designed to reduce MW artifacts and achieve isotropic resolution. 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. Our method on the other hand takes an already reconstructed tomogram as an input. The common points between the methods are that they both rely on Bayesian statistics, and none of them uses prior morphological knowledge about the specimen. 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, according to Fig. 8, our method achieves a better

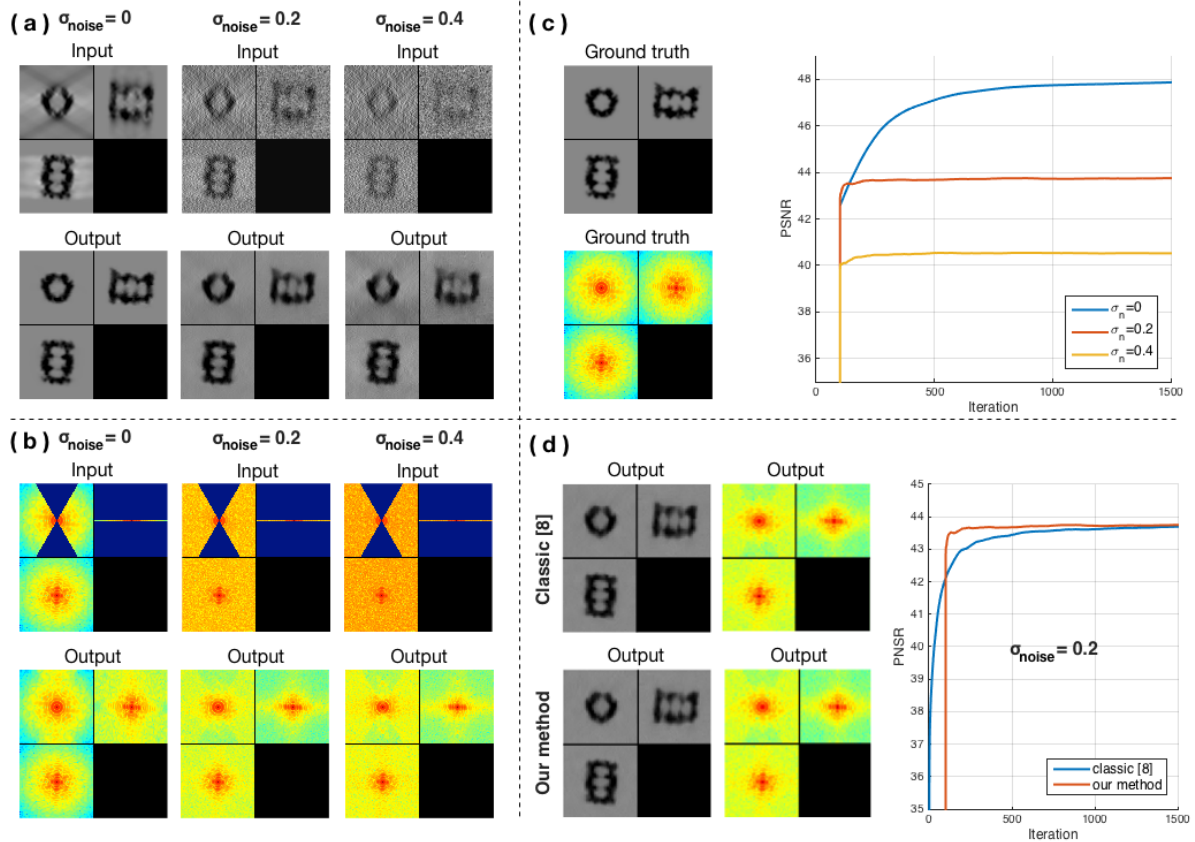


Figure 7: Simulated data of the 20S proteasome, for varying amounts of noise (dataset A). All images depict ortho-slices of 3D volumes. The volume size is $64 \times 64 \times 64$ voxels. For (a) and (b), top row: method inputs, bottom row: method outputs. Results are shown in the spatial domain (a) and in the spectral domain (b). In (c) can be observed the ground-truth and the evolution of the PSNR over iterations. Finally, in (d) we compare our method to the original method described in [Maggioni et al., 2013]

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.

Moisan’s TV method has similarities with [Maggioni et al., 2013]: it uses the same spectrum projection, and applies an iterative TV minimization, which is close to a denoising operation [Rudin et al., 1992].

We implemented Moisan’s method as follows:

Our implementation of Moisan’s method:

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

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

1. **Projection:** $\mathcal{P}_W(\mathbf{x}^{(t-1)})$
2. **Denoising:** apply TV denoising $\min_x \|y - x\|^2 + \lambda|\nabla x|$

end for

The result of Moisan’s 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 the CCC score, 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 artefacts 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 scores.

In the end, our approach outperforms competing methods both in terms of PSNR and CCC scores. Visually, our approach produces a smoother image, while sMAPEM and 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.

6.2 Results on experimental data

In this section, we evaluate the performance of our method on real images to confirm the results obtained on artificial images .

6.2.1 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).

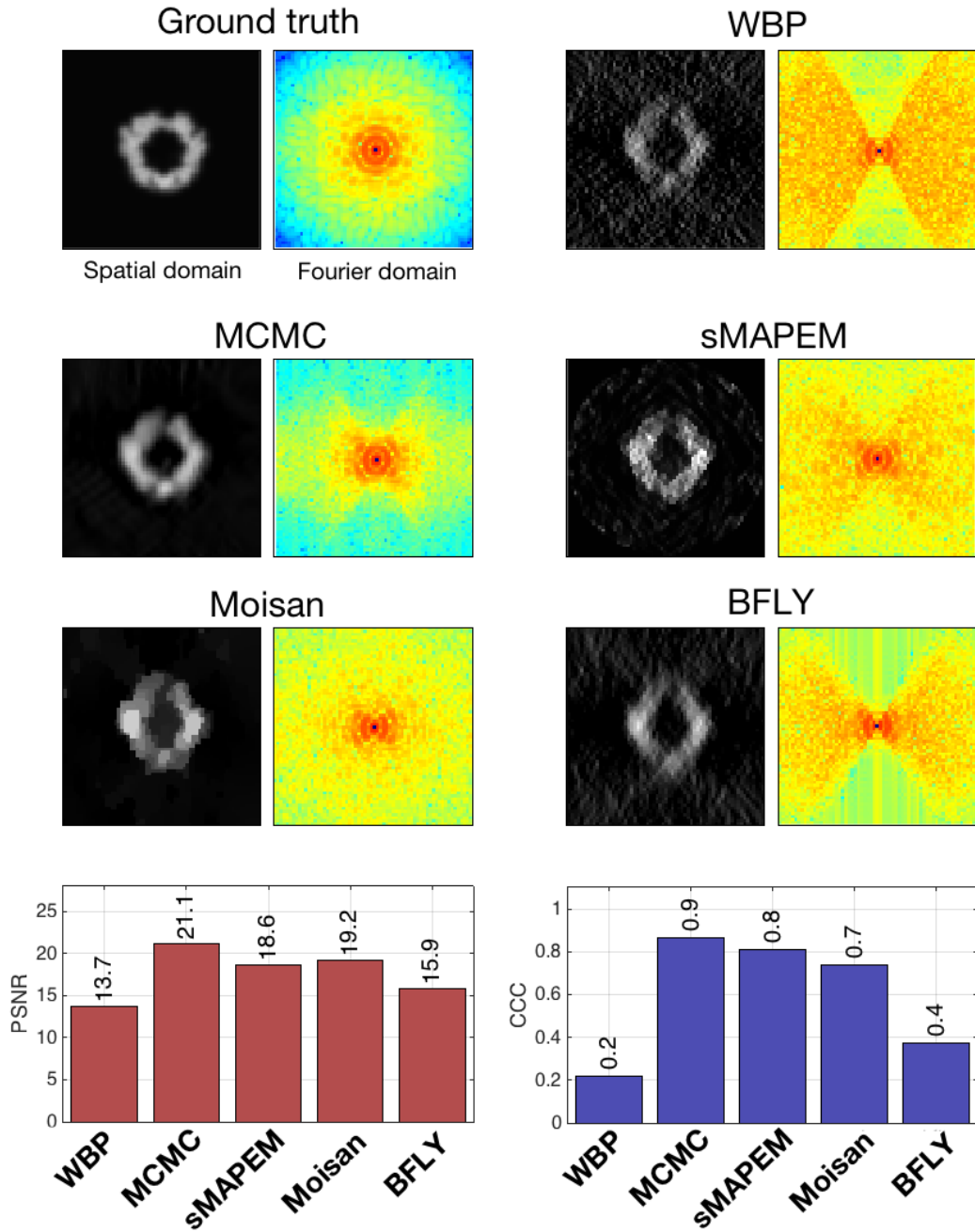


Figure 8: Dataset A (2D): comparing our approach to competitive methods. sMAPEM, a regularized tomographic reconstruction method, designed to achieve isotropic resolution. Moisan, a method designed to extrapolate missing regions in Fourier space. BFLY, a filter designed to reduce MW artifacts. sMAPEM 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 achieved PSNR and CCC scores.

6.2.2 Evaluation method

The evaluation differs depending on the dataset. In dataset B, the 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. As explained in Section 1, 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 Figure 12.

6.2.3 Results

The result on dataset B shows that noise is reduced and a significant part of the MW could be recovered (see Fig. 9). 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 contains molecules (ribosomes) that have more interest for biologists (see Fig. 10). This case is more challenging, because the objects have a more complex shape and less contrast, i.e. the SNR is lower. Nonetheless, the method could significantly enhance 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 could be 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. 10 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. 10). To illustrate how the method can improve visualization, a simple thresholding has been performed on the data (3D views in Fig. 10). 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 Figure 11). In both processed proteasome sub-tomograms, the contrast has been enhanced and the FSC and cFSC curves provide proof of an improved signal.

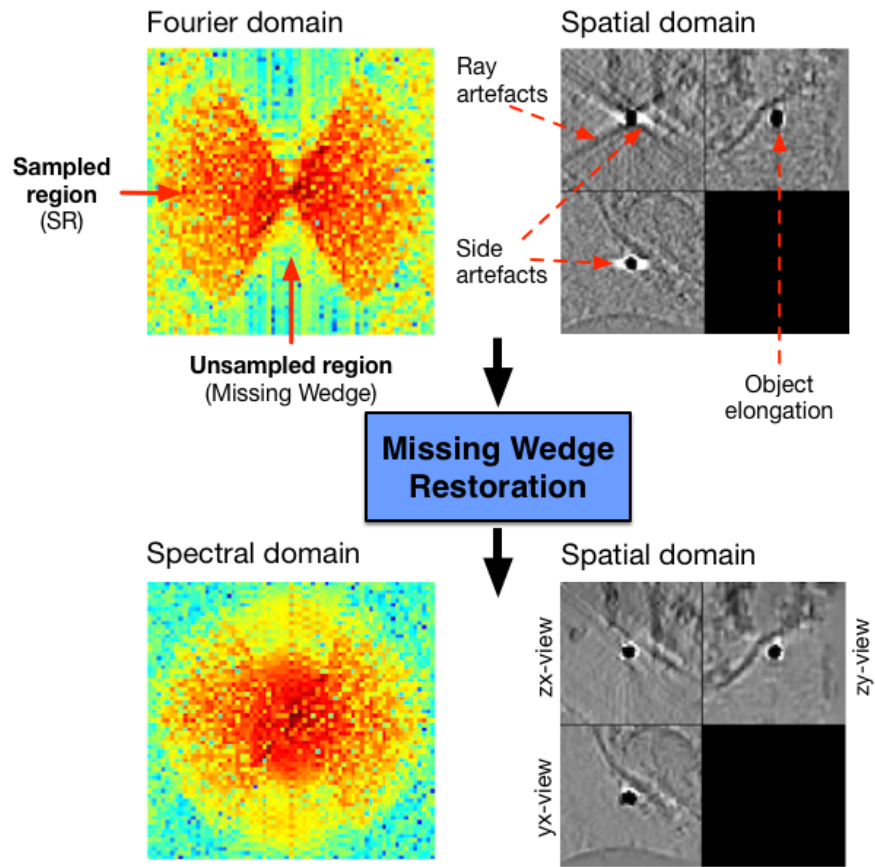


Figure 9: Experimental sub-tomogram ($61 \times 61 \times 61$ voxels) containing a gold particle (dataset B). The top row shows the input in the spectral and spatial domains, the bottom row shows the output.

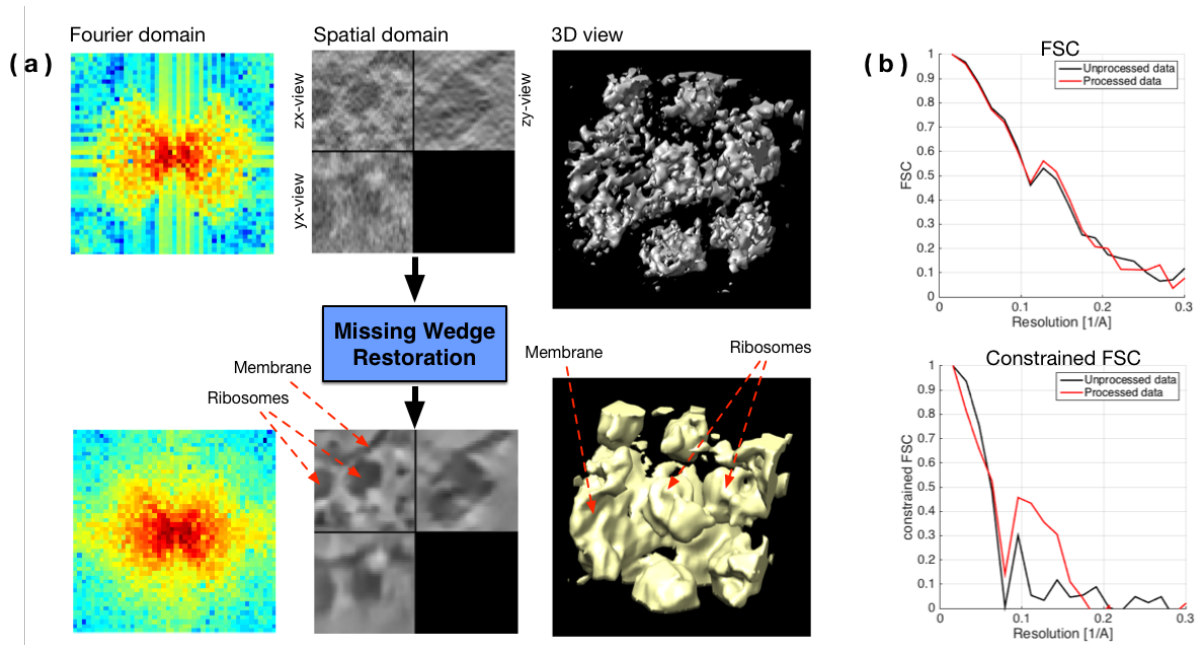


Figure 10: Experimental sub-tomogram ($46 \times 46 \times 46$ voxels) containing ribosomes attached to a membrane (dataset C). (a) Top row: input image in spectral domain, spatial domain and 3D view of the thresholded data. Bottom row: the same representations for the output. (b) FSC and cFSC measures 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. the missing wedge). All measures are wrt the same reference.

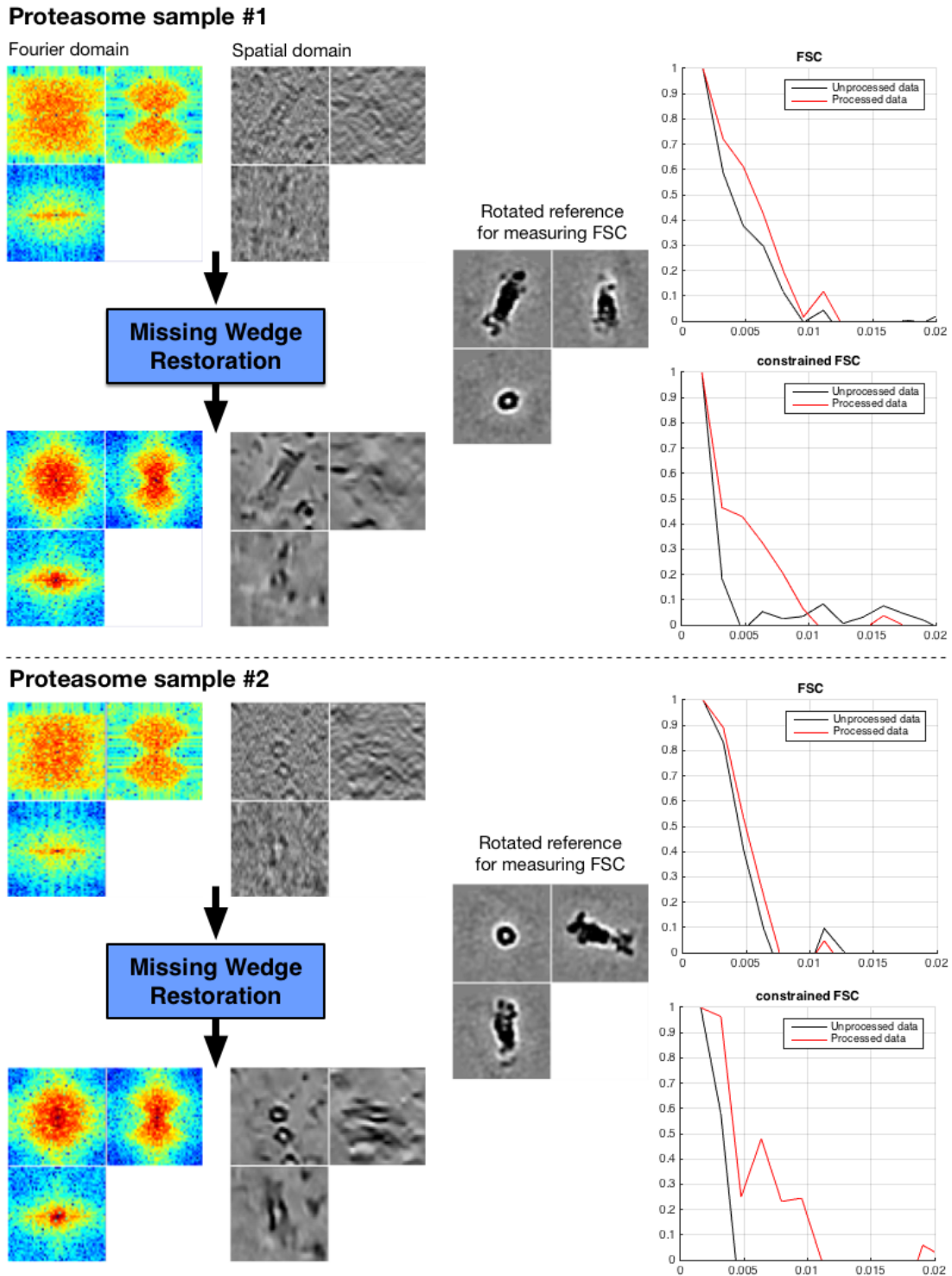


Figure 11: Two experimental sub-tomograms ($46 \times 46 \times 46$ voxels) containing proteasomes. Data is displayed in both Fourier domain and spatial domain. We evaluate the result with FSC and cFSC measures of the method input (in black) and output (in red). The FSC measures overall quality, while the cFSC measures quality of recovered Fourier coefficients only (i.e. the missing wedge). The reference has been obtained via subtomogram averaging of 2949 proteasomes. 28

Dataset D has a broader field of view than precedent datasets. As can be observed in Figure 12, 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 process the single-axis volume and investigate how close to the double-axis volume the result is. According to Figure 12, our method could indeed enhance the contrast, membrane and unknown macromolecules become easier to identify. When observing the data in Fourier domain, we can confirm that Fourier coefficients have been recovered. More importantly, according the CCC score, these recovered Fourier coefficient correlate with the double-axis volume. Indeed, the CCC score rised from 0.29 to 0.72 after applying our method.

7 Conclusion

In this paper, we have considered the problem of restoring an image corrupted by missing Fourier coefficients in a well-localized spectral region (missing wedge). We have proposed a Monte-Carlo simulation method to restore and compensate MW artifacts in cryo-ET images. 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 data show that even for high amounts of noise, the method is able to enhance the signal. However, the method needs a reasonable contrast of the object of interest to perform well, which is not always the case in cryo-ET. Nevertheless, with improving electron microscopy techniques like direct electron detection sensors and phase contrast methods, the method will be able to produce even more impressive results. The effectiveness of the method being demonstrated for the challenging case of cryo-ET, the method can be applied to other imaging modalities, especially on images with high SNR values.

Acknowledgment

This work was jointly supported by Fourmentin-Guilbert Scientific Foundation, and Région Bretagne (Brittany Council). Experiments on real data (courtesy of MPI Biochemistry, Martinsried, Germany) and on synthetic data 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

- [Bierkens, 2015] Bierkens, J. (2015). Non-reversible Metropolis-Hastings. *Stat. Comput.*, 26(6):1213–1228.
- [Breyer and Roberts, 2000] Breyer, L. A. and Roberts, G. O. (2000). From metropolis to diffusions: Gibbs states and optimal scaling. *Stoch. Process. their Appl.*, 90(2):181–206.

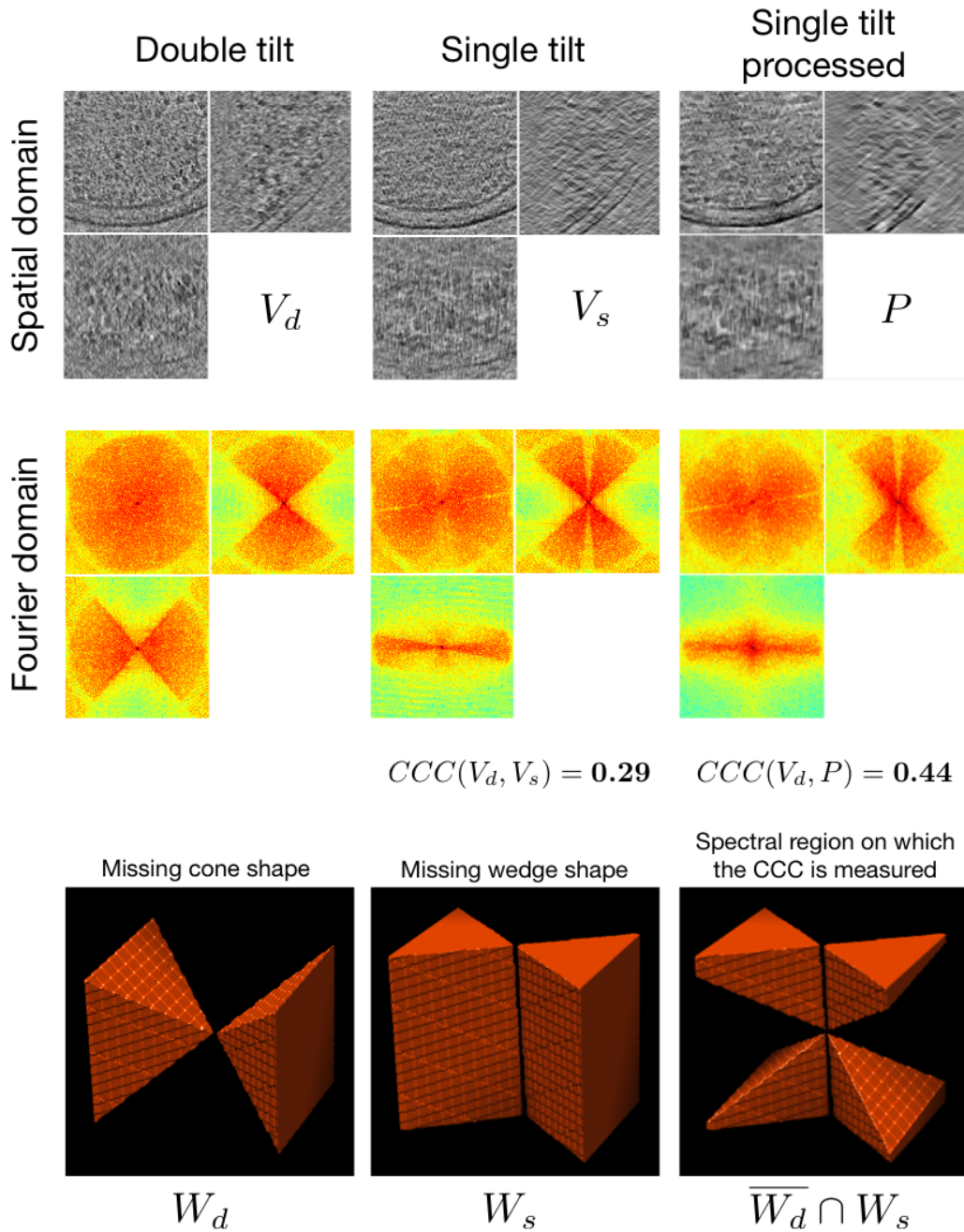


Figure 12: Dataset D: Experimental double-axis sub-tomogram ($128 \times 128 \times 128$ voxels) containing multiple macromolecules (see black dots). In this example, we process the single-axis version of the volume (top middle) and compare to the double-axis volume (top left), which acts as a ground truth. We evaluate the processed volume (top right) using the CCC score, as illustrated in the bottom right image. All volumes are displayed in spatial domain (top row) and Fourier domain (second row). W_d and W_s (bottom row) illustrate the shape of the missing Fourier region for double-axis and single-axis data.

- [Brooks and Gelman, 1998] Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.*, 7(4):434–455.
- [Buades et al., 2005] Buades, A., Coll, B., and Morel, J.-m. (2005). A non-local algorithm for image denoising. *Comput. Vis. Pattern Recognit.*, 2:60–65.
- [Burger et al., 2012] Burger, H. C., Schuler, C. J., and Harmeling, S. (2012). Image denoising: can plain neural networks compete with BM3D?
- [Chambolle and Jalalzai, 2014] Chambolle, A. and Jalalzai, K. (2014). Adapted basis for nonlocal reconstruction of missing spectrum. *SIAM J. Imaging Sci.*, 7(3):1484–1502.
- [Chatterjee and Milanfar, 2012] Chatterjee, P. and Milanfar, P. (2012). Patch-based near-optimal image denoising. *IEEE Trans. Image Process.*, 21(4):1635–1649.
- [Dabov et al., 2007] Dabov, K., Foi, A., and Egiazarian, K. (2007). Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):145–149.
- [Deledalle et al., 2009] Deledalle, C. A., Denis, L., and Tupin, F. (2009). Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Trans. Image Process.*, 18(12):2661–2672.
- [Deledalle et al., 2012] Deledalle, C. A., Duval, V., and Salmon, J. (2012). Non-local methods with shape-adaptive patches (NLM-SAP). *J. Math. Imaging Vis.*, 43(2):103–120.
- [Duval et al., 2011] Duval, V., Aujol, J.-F., and Gousseau, Y. (2011). A bias-variance approach for the nonlocal means. *SIAM J. Imaging Sci.*, 4(2):760–788.
- [Elad and Aharon, 2006] Elad, M. and Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. In *IEEE Trans. Image Process.*, volume 15, pages 3736–3745.
- [Förster and Hegerl, 2007] Förster, F. and Hegerl, R. (2007). Structure determination In Situ by averaging of tomograms. In *Cell. Electron Microsc.*, volume 79, pages 741–767.
- [Gelman and Rubin, 1992] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.*, 7(4):457–511.
- [Gilks et al., 1995] Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1995). *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*.
- [Gribonval, 2011] Gribonval, R. (2011). Should penalized least squares regression be interpreted as maximum a posteriori estimation? In *IEEE Trans. Signal Process.*, volume 59, pages 2405–2410.
- [Guesdon et al., 2013] Guesdon, A., Blestel, S., Kervrann, C., and Chrétien, D. (2013). Single versus dual-axis cryo-electron tomography of microtubules assembled in vitro: limits and perspectives. *J. Struct. Biol.*, 181(2):169–78.

- [Guichard and Malgouyres, 1998] Guichard, F. and Malgouyres, F. (1998). Total variation based interpolation. In *Eur. Signal Process. Conf.*, volume 3, pages 1741–1744.
- [Hastings, 1970] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109.
- [Holden et al., 2009] Holden, L., Hauge, R., and Holden, M. (2009). Adaptive independent Metropolis-Hastings. *Ann. Appl. Probab.*, 19(1):395–413.
- [Jin et al., 2017] Jin, Q., Grama, I., Kervrann, C., and Liu, Q. (2017). Nonlocal means and optimal weights for noise removal. *SIAM J. Imaging Sci.*, 10(4):1878–1920.
- [Katkovnik et al., 2010] Katkovnik, V., Foi, A., Egiazarian, K., and Astola, J. (2010). From local kernel to nonlocal multiple-model image denoising. *Int. J. Comput. Vis.*, 86(1):1–32.
- [Kazerouni et al., 2013] Kazerouni, A., Kamilov, U. S., Bostan, E., and Unser, M. (2013). Bayesian denoising: from MAP to MMSE using consistent cycle spinning. In *IEEE Signal Process. Lett.*, volume 20, pages 249–252.
- [Kervrann, 2014] Kervrann, C. (2014). PEWA: Patch-based exponentially weighted aggregation for image denoising. In *Adv. Neural Inf. Process. Syst.*, volume 27, pages 1–9.
- [Kervrann and Boulanger, 2006] Kervrann, C. and Boulanger, J. (2006). Optimal spatial adaptation for patch-based image denoising. In *IEEE Trans. Image Process.*, volume 15, pages 2866–2878.
- [Kervrann and Boulanger, 2007] Kervrann, C. and Boulanger, J. (2007). Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In *Scale Sp. Var. Methods Comput. Vis.*, pages 520–532.
- [Kervrann and Boulanger, 2008] Kervrann, C. and Boulanger, J. (2008). Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *Int. J. Comput. Vis.*, 79(1):45–69.
- [Kervrann et al., 2014] Kervrann, C., Roudot, P., and Waharte, F. (2014). Approximate bayesian computation, stochastic algorithms and non-local means for complex noise models. In *IEEE Int. Conf. Image Process.*, pages 2834–2838.
- [Kindermann et al., 2005] Kindermann, S., Osher, S., and Jones, P. W. (2005). Deblurring and denoising of images by nonlocal functionals. *Multiscale Model. Simul.*, 4(4):1091–1115.
- [Kováčik et al., 2014] Kováčik, L., Kereiche, S., Höög, J. L., Jůda, P., Matula, P., and 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(1):141–52.
- [Leary et al., 2013] Leary, R., Saghi, Z., Midgley, P. A., and Holland, D. J. (2013). Compressed sensing electron tomography. *Ultramicroscopy*, 131:70–91.

- [Lebrun et al., 2013] Lebrun, M., Buades, A., and Morel, J. (2013). A nonlocal Bayesian image denoising algorithm. *SIAM J. Imaging Sci.*, 6(3):1665–1688.
- [Liang et al., 2010] Liang, F., Liu, C., and Carroll, R. J. (2010). *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*.
- [Lou et al., 2010] Lou, Y., Zhang, X., Osher, S., and Bertozzi, A. (2010). Image recovery via nonlocal operators. *J. Sci. Comput.*, 42(2):185–197.
- [Louchet and Moisan, 2008] Louchet, C. and Moisan, L. (2008). Total variation denoising using posterior expectation. *16th Eur. Signal Process. Conf.*, 5:1–5.
- [Louchet and Moisan, 2011] Louchet, C. and Moisan, L. (2011). Total variation as a local filter. *SIAM J. Imaging Sci.*, 4(3):651–694.
- [Louchet and Moisan, 2013] Louchet, C. and Moisan, L. (2013). Posterior expectation of the total variation model: properties and experiments. *SIAM J. Imaging Sci.*, 6(4):2640–2684.
- [Maggioni et al., 2013] Maggioni, M., Katkovnic, V., Egiazarian, K., and Foi, A. (2013). Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.*, 22(1):119–133.
- [Mairal et al., 2009] Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2009). Non-local sparse models for image restoration. In *IEEE Int. Conf. Comput. Vis.*, pages 2272–2279.
- [Metropolis et al., 1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations. *J. Chem. Phys.*, 21(6):1087–1092.
- [Miao et al., 2005] Miao, J., Förster, F., and Levi, O. (2005). Equally sloped tomography with oversampling reconstruction. *Phys. Rev. B*, 72(5):3–6.
- [Milanfar, 2013] Milanfar, P. (2013). A tour of modern image filtering. *IEEE Signal Process. Mag.*, 30(1):106–128.
- [Moisan, 2011] Moisan, L. (2011). Periodic plus smooth image decomposition. *J. Math. Imaging Vis.*, 39(2):161–179.
- [Moisan, 2001] Moisan, L. C. (2001). Extrapolation de spectre et variation totale pondérée. In *18e Colloq. sur le Trait. du Signal des Images*, pages 892–895.
- [Neal, 2004] Neal, R. M. (2004). Improving asymptotic variance of MCMC estimators : non-reversible chains are better. Technical Report 0406.
- [Paavolainen et al., 2014] Paavolainen, L., Acar, E., Tuna, U., Peltonen, S., Moriya, T., Pan, S., Marjom, V., Cheng, R. H., and Ruotsalainen, U. (2014). Compensation of missing wedge effects with sequential statistical reconstruction in electron tomography. *PLoS One*, 9(10):1–23.

- [Pizarro et al., 2010] Pizarro, L., Mrázek, P., Didas, S., Grewenig, S., and Weickert, J. (2010). Generalised nonlocal image smoothing. *Int. J. Comput. Vis.*, 90(1):62–87.
- [Protter et al., 2010] Protter, M., Yavneh, I., and Elad, M. (2010). Closed-Form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary. *IEEE Trans. Signal Process.*, 58(7):3471–3484.
- [Robert and Casella, 2004] Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer.
- [Roberts et al., 1997] Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120.
- [Roberts and Rosenthal, 2001] Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Stat. Sci.*, 16(4):351–367.
- [Rudin et al., 1992] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys. D*, 60:259–268.
- [Sutour et al., 2014] Sutour, C., Deledalle, C.-a., and Aujol, J.-f. (2014). Adaptive regularization of the NL-means: application to image and video denoising. *IEEE Trans. Image Process.*, 23(8):3506–3521.
- [Van De Ville and Kocher, 2009] Van De Ville, D. and Kocher, M. (2009). SURE-based non-local means. *IEEE Signal Process. Lett.*, 16(11):973–976.
- [Van Heel and Schatz, 2005] Van Heel, M. and Schatz, M. (2005). Fourier shell correlation threshold criteria. *J. Struct. Biol.*, 151:250–262.
- [Wang and Morel, 2013] Wang, Y.-Q. and Morel, J.-M. (2013). SURE guided Gaussian mixture image denoising. *SIAM J. Imaging Sci.*, 6(2):999–1034.
- [Zoran and Weiss, 2011] Zoran, D. and Weiss, Y. (2011). From learning models of natural image patches to whole image restoration. In *IEEE Int. Conf. Comput. Vis.*, pages 479–486.