Hierarchical Bayesian sparse image reconstruction with application to MRFM

Nicolas Dobigeon[1,2], Alfred O. Hero[2] and Jean-Yves Tourneret[1]

[1] University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, 31071 Toulouse, France.

[2] University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA

{Nicolas.Dobigeon, Jean-Yves.Tourneret}@enseeiht.fr, hero@umich.edu

Abstract

This paper presents a hierarchical Bayesian model to reconstruct sparse images when the observations are obtained from linear transformations and corrupted by an additive white Gaussian noise. Our hierarchical Bayes model is well suited to such naturally sparse image applications as it seamlessly accounts for properties such as sparsity and positivity of the image via appropriate Bayes priors. We propose a prior that is based on a weighted mixture of a positive exponential distribution and a mass at zero. The prior has hyperparameters that are tuned automatically by marginalization over the hierarchical Bayesian model. To overcome the complexity of the posterior distribution, a Gibbs sampling strategy is proposed. The Gibbs samples can be used to estimate the image to be recovered, e.g. by maximizing the estimated posterior distribution. In our fully Bayesian approach the posteriors of all the parameters are available. Thus our algorithm provides more information than other previously proposed sparse reconstruction methods that only give a point estimate. The performance of our hierarchical Bayesian sparse reconstruction method is illustrated on synthetic and real data collected from a tobacco virus sample using a prototype MRFM instrument.

Index Terms

Deconvolution, MRFM imagery, sparse representation, Bayesian inference, MCMC methods.

Part of this work has been supported by ARO MURI grant No. W911NF-05-1-0403.

I. INTRODUCTION

For several decades, image deconvolution has received increasing interest in the literature [2], [44]. Deconvolution mainly consists of reconstructing images from observations provided by optical devices and may include denoising, deblurring or restoration. The applications are numerous including astronomy [46], medical imagery [45], remote sensing [38] and photography [52]. More recently, a new imaging technology, called Magnetic Resonance Force Microscopy (MRFM), has been developed (see [35] and [27] for recent reviews). This non-destructive method allows one to improve the detection sensitivity of standard magnetic resonance imaging [43]. Because of its potential atomic-level resolution[1], the 2-dimensional or 3-dimensional images resulting from this technology are naturally sparse in the standard pixel basis. Indeed, as the observed objects are molecules, most of the image is empty space. In this paper, a hierarchical Bayesian model is proposed to perform reconstruction of such images.

Sparse signal and image deconvolution has motivated research in many scientific applications including: spectral analysis in astronomy [4]; seismic signal analysis in geophysics [7], [42]; and deconvolution of ultrasonic B-scans [36]. We propose here a hierarchical Bayesian model that is based on selecting an appropriate prior distribution for the unknown image and other unknown parameters. The image prior is composed of a weighted mixture of a standard exponential distribution and a mass at zero. When the non-zero part of this prior is chosen to be a centered normal distribution, this prior reduces to a Bernoulli-Gaussian process. This distribution has been widely used in the literature to build Bayesian estimators for sparse deconvolution problems (see [5], [15], [22], [26], [31] or more recently [3] and [16]). However, choosing a distribution with heavier tail may improve the sparsity inducement of the prior. Combining a Laplacian distribution with an atom at zero results in the so-called LAZE prior. This distribution has been used in [25] to solve a general denoising problem in a non-Bayesian quasi-maximum likelihood estimation framework. In [49], [51], this prior has also been used for sparse reconstruction of noisy images, including MRFM. The principal weakness of these previous approaches is the sensitivity to hyperparameters that determine the prior distribution, e.g. the LAZE mixture coefficient and the

[1]Note that the current state of art of the MRFM technology allows one to acquire images with nanoscale resolution. Indeed, several hundreds of nuclei are necessary to get a detectable signal. However, atomic-level resolution might be obtained in the future.

weighting of the prior vs the likelihood function. The hierarchical Bayesian approach proposed in this paper circumvents these difficulties. Specifically, a new prior composed of a mass at zero and a single-sided exponential distribution is introduced, which accounts for positivity and sparsity of the pixels in the image. Conjugate priors on the hyperparameters of the image prior are introduced. It is this step that makes our approach hierarchical Bayesian. The full Bayesian posterior can then be derived from samples generated by Markov chain Monte Carlo (MCMC) methods [41].

The estimation of hyperparameters involved in the prior distribution described above is the most difficult task and poor estimation leads to instability. Empirical Bayes (EB) and Stein unbiased risk (SURE) solutions were proposed in [49], [51] to deal with this issue. However, instability was observed especially at higher signal-to-noise ratios (SNR). In the Bayesian estimation framework, two approaches are available to estimate these hyperparameters. One approach couples MCMC methods to an expectation-maximization (EM) algorithm or to a stochastic EM algorithm [28], [30] to maximize a penalized likelihood function. The second approach defines non-informative prior distributions for the hyperparameters; introducing a second level of hierarchy in the Bayesian formulation. This latter fully Bayesian approach, adopted in this paper, has been successfully applied to signal segmentation [11], [13], [14] and semi-supervised unmixing of hyperspectral imagery [12].

Only a few works in the literature have been dedicated to reconstruction of MRFM image data [6], [8], [53], [54]. In [20], several techniques based on linear filtering and maximum-likelihood principles have been proposed that do not exploit image sparsity. More recently, Ting *et al.* has introduced sparsity penalized reconstruction methods for MRFM (see [51] or [50]). The reconstruction problem has been formulated as a decomposition into a deconvolution step and a denoising step, yielding an iterative thresholding framework. In [51] the hyperparameters are estimated using penalized log-likelihood criteria including the SURE approach [47]. Despite promising results, especially at low SNR, penalized likelihood approaches require iterative algorithms that are often slow to converge and can get stuck on local maxima [10]. In contrast to [51], the fully Bayesian approach presented in this paper converges quickly and produces estimates of the entire posterior and not just local maxima. Indeed, the hierarchical Bayesian formulation proposed here asymptotically generates Bayes-optimal estimates of all image parameters, including the hyperparameters.

In this paper, the response of the MRFM imaging device is assumed to be known. While it may be possible to extend our methods to unknown point spread functions, e.g., along the lines of [21], the case of sparse blind deconvolution is outside of the scope of this paper.

This paper is organized as follows. The deconvolution problem is formulated in Section II. The hierarchical Bayesian model is described in Section III. Section IV presents a Gibbs sampler that allows one to generate samples distributed according to the posterior of interest. Simulation results, including extensive performance comparison, are presented in Section V. In Section VI we apply our hierarchical Bayesian method to reconstruction of a tobacco virus from real MRFM data. Our main conclusions are reported in Section VII.

II. Problem formulation

Let $\mathbf{X}$ denote a $l_1 \times \ldots \times l_n$ unknown n-dimensional pixelated image to be recovered (e.g. $n = 2$ or $n = 3$). Observed are a collection of P projections $\mathbf{y} = [y_1, \ldots, y_P]^T$ which are assumed to follow the model:

$$\mathbf{y} = T(\boldsymbol{\kappa}, \mathbf{X}) + \mathbf{n}, \tag{1}$$

where $T(\cdot, \cdot)$ stands for a bilinear function, $\mathbf{n}$ is a $P \times 1$ dimension noise vector and $\boldsymbol{\kappa}$ is the kernel that characterizes the response of the imaging device. In the right-hand side of (1), $\mathbf{n}$ is an additive Gaussian noise sequence distributed according to $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_P)$, where the σ^2 is assumed to be unknown.

Note that in standard deblurring problems, the function $T(\cdot, \cdot)$ represents the standard n-dimensional convolution operator $\otimes$. In this case, the image $\mathbf{X}$ can be vectorized yielding the unknown image $\mathbf{x} \in \mathbb{R}^M$ with $M = P = l_1 l_2 \ldots l_n$. With this notation, Eq. (1) can be rewritten:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \qquad \text{or} \qquad \mathbf{Y} = \boldsymbol{\kappa} \otimes \mathbf{X} + \mathbf{N} \tag{2}$$

where $\mathbf{y}$ (resp. $\mathbf{n}$) stands for the vectorized version of $\mathbf{Y}$ (resp. $\mathbf{N}$) and $\mathbf{H}$ is an $P \times M$ matrix that describes convolution by the psf $\boldsymbol{\kappa}$.

The problem addressed in the following sections consists of estimating $\mathbf{x}$ and σ^2 under sparsity and positivity constraints on $\mathbf{x}$ given the observations $\mathbf{y}$, the psf $\boldsymbol{\kappa}$ and the bilinear function[2] $T(\cdot, \cdot)$.

[2]In the following, for sake of conciseness, the same notation $T(\cdot, \cdot)$ will be adopted for the bilinear operations used on n-dimensional images $\mathbf{X}$ and used on $M \times 1$ vectorized images $\mathbf{x}$.

III. Hierarchical Bayesian model

A. Likelihood function

The observation model defined in (1) and the Gaussian properties of the noise sequence $\mathbf{n}$ yield:

$$f\left(\mathbf{y}|\mathbf{x},\sigma^2\right) = \left(\frac{1}{2\pi\sigma^2}\right)^P \exp\left(-\frac{\|\mathbf{y}-T\left(\boldsymbol{\kappa},\mathbf{x}\right)\|^2}{2\sigma^2}\right), \tag{3}$$

where $\|\cdot\|$ denotes the standard ℓ_2 norm: $\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x}$.

B. Parameter prior distributions

The unknown parameter vector associated with the observation model defined in (1) is $\boldsymbol{\theta} = \{\mathbf{x},\sigma^2\}$. In this section, we introduce prior distributions for these two parameters; which are assumed to be independent.

1) Image prior: First let consider the exponential distribution with shape parameter $a > 0$:

$$g_a\left(x_i\right) = \frac{1}{a}\exp\left(-\frac{x_i}{a}\right)\mathbf{1}_{\mathbb{R}_+^*}\left(x_i\right), \tag{4}$$

where $\mathbf{1}_{\mathbb{E}}\left(x\right)$ is the indicator function defined on $\mathbb{E}$:

$$\mathbf{1}_{\mathbb{E}}\left(x\right) = \begin{cases} 1, & \text{if } x \in \mathbb{E}, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Choosing $g_a\left(\cdot\right)$ as prior distributions for x_i ($i = 1,\ldots,M$) leads to a MAP estimator of $\mathbf{x}$ that corresponds to a maximum ℓ_1-penalized likelihood estimate with a positivity constraint[3]. Indeed, assuming the component x_i ($i = 1,\ldots,P$) a priori independent allows one to write the full prior distribution for $\mathbf{x} = \left[x_1,\ldots,x_M\right]^T$:

$$g_a\left(\mathbf{x}\right) = \left(\frac{1}{a}\right)^M \exp\left(-\frac{\|\mathbf{x}\|_1}{a}\right)\mathbf{1}_{\{\mathbf{x}\succ 0\}}\left(\mathbf{x}\right), \tag{6}$$

where $\{\mathbf{x}\succ 0\} = \left\{\mathbf{x}\in\mathbb{R}^M; x_i > 0, \forall i = 1,\ldots,M\right\}$ and $\|\cdot\|_1$ is the standard ℓ_1 norm $\|\mathbf{x}\|_1 = \sum_i |x_i|$. This estimator has shown interesting sparse properties for Bayesian estimation [1] and signal representation [19].

[3]Note that a similar estimator using a Laplacian prior for x_i ($i = 1,\ldots,M$) was proposed in [48] for regression problems and is usually referred to as the LASSO estimator but without positivity constraint.

Coupling a standard probability density function (pdf) with an atom at zero is another alternative to ensure sparsity. This strategy has for instance been used for located event detection [26] such as spike train deconvolution [5], [7]. In order to increase the sparsity of the prior, we propose to use the following distribution derived from $g_a(\cdot)$ as prior distribution for x_i:

$$f(x_i|w,a) = (1-w)\delta(x_i) + w g_a(x_i), \tag{7}$$

where $\delta(\cdot)$ is the Dirac function. This prior is similar to the LAZE distribution (Laplacian pdf and an atom at zero) introduced in [25] and used, for example, in [49], [51] for MRFM. However, the proposed prior in (7) allows one to take into account the positivity of the pixel value to be estimated. By assuming the components x_i to be a priori independent ($i = 1, \ldots, M$), the following prior distribution is obtained for $\mathbf{x}$:

$$f(\mathbf{x}|w,a) = \prod_{i=1}^{M} \left[(1-w)\delta(x_i) + w g_a(x_i)\right]. \tag{8}$$

Introducing the index subsets $\mathcal{I}_0 = \{i; x_i = 0\}$ and $\mathcal{I}_1 = \overline{\mathcal{I}}_0 = \{i; x_i \neq 0\}$ allows one to rewrite the previous equation as follows:

$$f(\mathbf{x}|w,a) = \left[(1-w)^{n_0} \prod_{i\in\mathcal{I}_0} \delta(x_i)\right] \left[w^{n_1} \prod_{i\in\mathcal{I}_1} g_a(x_i)\right], \tag{9}$$

with $n_\epsilon = \operatorname{card}\{\mathcal{I}_\epsilon\}$, $\epsilon \in \{0,1\}$. Note that $n_0 = M - n_1$ and $n_1 = \|\mathbf{x}\|_0$ where $\|\cdot\|_0$ is the standard ℓ_0 norm $\|\mathbf{x}\|_0 = \#\{i; x_i \neq 0\}$.

2) Noise variance prior: A conjugate inverse-Gamma distribution with parameters $\frac{\nu}{2}$ and $\frac{\gamma}{2}$ is chosen as prior distribution for the noise variance [40, Appendix A]:

$$\sigma^2|\nu,\gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right). \tag{10}$$

In the following, ν will be fixed to $\nu = 2$ and γ will be an hyperparameter to be estimated (see [12], [13], [37]).

C. Hyperparameter priors

The hyperparameter vector associated with the previous prior distributions is $\mathbf{\Phi} = \{a, \gamma, w\}$. Obviously, the accuracy of the proposed Bayesian model depends on the values of these hyperparameters. If prior knowledge, e.g. mean number of the non-zero pixels, is available, these parameters can be tuned manually to their actual values. However, in practical situations, such

prior information is not available. In this case, as outlined in Section I, these hyperparameters can be estimated directly from the data. Priors for these hyperparameters, sometimes referred to as "hyperpriors" are detailed below.

1) Hyperparameter a: A conjugate inverse-Gamma distribution is assumed for hyperparameter a:

$$a|\boldsymbol{\alpha} \sim \mathcal{IG}\left(\alpha_0, \alpha_1\right), \tag{11}$$

with $\boldsymbol{\alpha} = [\alpha_0, \alpha_1]^T$. The fixed hyperparameters α_0 and α_1 have been chosen to obtain a vague prior: $\alpha_0 = \alpha_1 = 10^{-10}$ (see for example [18] for a similar choice).

2) Hyperparameter γ: A non informative Jeffreys' prior [23], [24] is assumed for hyperparameter γ:

$$f\left(\gamma\right) \propto \frac{1}{\gamma}\mathbf{1}_{\mathbb{R}_+}\left(\gamma\right). \tag{12}$$

3) Hyperparameter w: A conjugate beta distribution with fixed hyperparameters ω_1 and ω_0 is chosen as prior distribution for w:

$$w|\boldsymbol{\omega} \sim \mathcal{B}\left(\omega_1, \omega_0\right), \tag{13}$$

with $\boldsymbol{\omega} = [\omega_0, \omega_1]^T$ and where $\mathcal{B}\left(a, b\right)$ denotes the Beta distribution with parameters (a, b). Note that by choosing $\omega_0 = \omega_1 = 1$, the Beta distribution reduces to the uniform distribution on $[0, 1]$, which gives the least informative prior.

Assuming that the individual hyperparameters are independent the full hyperparameter prior distribution for $\boldsymbol{\Phi}$ can be expressed as:

$$\begin{aligned} f\left(\boldsymbol{\Phi}|\boldsymbol{\alpha}, \boldsymbol{\omega}\right) &= f\left(w\right) f\left(\gamma\right) f\left(a\right) \\ &= \frac{w^{\omega_1 - 1}\left(1 - w\right)^{\omega_0 - 1}}{aw\gamma B\left(\omega_1, \omega_0\right)}\mathbf{1}_{[0,1]}\left(w\right)\mathbf{1}_{\mathbb{R}^+}\left(a\right)\mathbf{1}_{\mathbb{R}^+}\left(\gamma\right), \end{aligned} \tag{14}$$

with $B\left(\omega_1, \omega_0\right) = \frac{\Gamma(\omega_1)\Gamma(\omega_0)}{\Gamma(\omega_1 + \omega_0)}$, where $\Gamma(\cdot)$ denotes the Gamma function.

D. Posterior distribution

The posterior distribution of $\{\boldsymbol{\theta}, \boldsymbol{\Phi}\}$ can be computed as follows:

$$f\left(\boldsymbol{\theta}, \boldsymbol{\Phi}|\mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\omega}\right) \propto f\left(\mathbf{y}|\boldsymbol{\theta}\right) f\left(\boldsymbol{\theta}|\boldsymbol{\Phi}\right) f\left(\boldsymbol{\Phi}|\boldsymbol{\alpha}, \boldsymbol{\omega}\right), \tag{15}$$

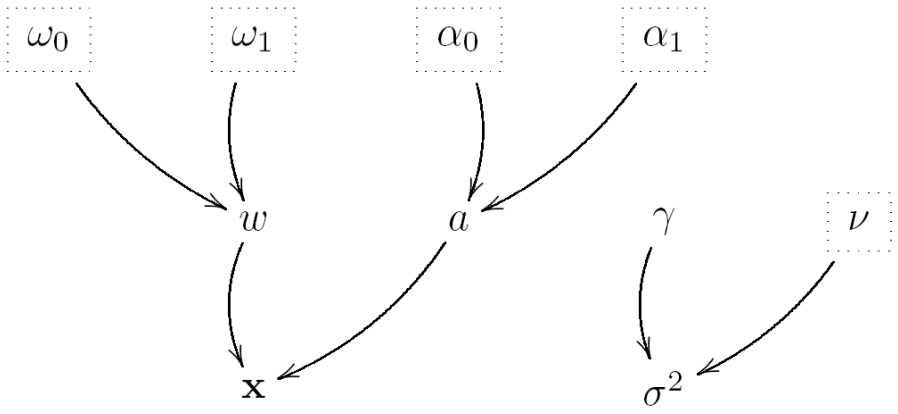

Fig. 1. DAG for the parameter priors and hyperpriors (the fixed hyperparameters appear in dashed boxes).

with

$$f\left(\boldsymbol{\theta}|\boldsymbol{\Phi}\right) = f\left(\mathbf{x}|a,w\right) f\left(\sigma^2|\gamma\right), \tag{16}$$

where $f\left(\mathbf{y}|\boldsymbol{\theta}\right)$ and $f\left(\boldsymbol{\Phi}|\boldsymbol{\alpha},\boldsymbol{\omega}\right)$ have been defined in (3) and (14). This hierarchical structure, represented on the directed acyclic graph (DAG) in Fig. 1, allows one to integrate out the parameter σ^2 and the hyperparameter vector $\boldsymbol{\Phi}$ in the full posterior distribution (15), yielding:

$$f\left(\mathbf{x}|\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\omega}\right) \propto \frac{B\left(\omega_1+n_1,\omega_0+n_0\right)}{\left\|\mathbf{y}-T\left(\boldsymbol{\kappa},\mathbf{x}\right)\right\|^P} \frac{\Gamma\left(n_1+\alpha_0\right)}{\left[\left\|\mathbf{x}\right\|_1+\alpha_1\right]^{n_1+\alpha_0}}. \tag{17}$$

where, as defined in paragraph III-B.1, $n_1 = \left\|\mathbf{x}\right\|_0$ and $n_0 = M - \left\|\mathbf{x}\right\|_0$.

The next section presents an appropriate Gibbs sampling strategy [41] that allows one to generate samples distributed according to the posterior distribution $f\left(\mathbf{x}|\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\omega}\right)$.

IV. A Gibbs sampling strategy for sparse image reconstruction

In this section we describe the Gibbs sampling strategy for generating samples $\left\{\mathbf{x}^{(t)}\right\}_{t=1,\ldots}$ distributed according to the posterior distribution in (17). As simulating directly according to (17) is a difficult task, it is much more convenient to generate samples distributed according to the joint posterior $f\left(\mathbf{x},\sigma^2|\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\omega}\right)$. The main steps of this algorithm are detailed in subsections IV-A and IV-B (see also Algorithm 1 below).

ALGORITHM 1:

Gibbs sampling algorithm for sparse image reconstruction

- Initialization:
  - Sample parameter $\mathbf{x}^{(0)}$ from pdf in (9),
  - Sample parameters $\widetilde{\sigma}^{2(0)}$ from the pdf in (10),
  - Set $t \leftarrow 1$,
- Iterations: for $t = 1, 2, \ldots,$ do
  1. Sample hyperparameter $w^{(t)}$ from the pdf in (19),
  2. Sample hyperparameter $a^{(t)}$ from the pdf in (20),
  3. For $i = 1, \ldots, M$, sample parameter $x_i^{(t)}$ from pdf in (21),
  4. Sample parameter $\widetilde{\sigma}^{2(t)}$ from the pdf in (24),
  5. Set $t \leftarrow t + 1$.

A. Generation of samples according to $f\left(\mathbf{x}\left|\sigma^2, \mathbf{y}, \boldsymbol{\alpha}, \boldsymbol{\omega}\right.\right)$

To generate samples distributed according to $f\left(\mathbf{x}\left|\sigma^2, \mathbf{y}, \boldsymbol{\omega}\right.\right)$, it is very convenient to sample according to $f\left(\mathbf{x}, w, a\left|\sigma^2, \mathbf{y}, \boldsymbol{\omega}\right.\right)$ in the following 3-step procedure.

1) Generation of samples according to $f\left(w\left|\mathbf{x}, \boldsymbol{\omega}\right.\right)$*:* Using (9), the following result can be obtained:

$$f\left(w\left|\mathbf{x}, \boldsymbol{\omega}\right.\right) \propto (1-w)^{n_0+\omega_0-1} w^{n_1+\omega_1-1}, \tag{18}$$

where n_0 and n_1 have been defined in paragraph III-B.1. Therefore, generation of samples according to $f\left(w\left|\mathbf{x}, \boldsymbol{\omega}\right.\right)$ is achieved as follows:

$$w\left|\mathbf{x}, \boldsymbol{\omega}\right. \sim \mathcal{B}e\left(\omega_1 + n_1, \omega_0 + n_0\right). \tag{19}$$

2) Generation of samples according to $f\left(a\left|\mathbf{x}, \boldsymbol{\alpha}\right.\right)$*:* Looking at the joint posterior distribution (15), it yields:

$$a\left|\mathbf{x}, \boldsymbol{\alpha}\right. \sim \mathcal{IG}\left(\left\|\mathbf{x}\right\|_0 + \alpha_0, \left\|\mathbf{x}\right\|_1 + \alpha_1\right). \tag{20}$$

3) Generation of samples according to $f(\mathbf{x}|w,a,\sigma^2,\mathbf{y})$: The prior chosen for x_i ($i=1,\ldots,M$) yields a posterior distribution of $\mathbf{x}$ that is not closed form. However, the posterior distribution of each component x_i ($i=1,\ldots,M$) conditionally upon the others can be easily derived. Indeed straightforward computations detailed in Appendix I yield:

$$\begin{aligned} f\left(x_i|w,a,\sigma^2,\mathbf{x}_{-i},\mathbf{y}\right) \propto (1-w_i)\delta\left(x_i\right) \\ + w_i\phi_+\left(x_i|\mu_i,\eta_i^2\right), \end{aligned} \tag{21}$$

where $\mathbf{x}_{-i}$ stands for the vector $\mathbf{x}$ whose ith component has been removed and μ_i and η_i^2 are given in Appendix I. In (21), $\phi_+\left(\cdot,m,s^2\right)$ stands for the pdf of the truncated Gaussian distribution defined on $\mathbb{R}_+^*$ with hidden parameters equal to mean m and variance s^2:

$$\phi_+\left(x,m,s^2\right) = \frac{1}{C\left(m,s^2\right)}\exp\left[-\frac{(x-m)^2}{2s^2}\right]\mathbf{1}_{\mathbb{R}_+^*}(x), \tag{22}$$

with

$$C\left(m,s^2\right) = \sqrt{\frac{\pi s^2}{2}}\left[1+\operatorname{erf}\left(\frac{m}{\sqrt{2s^2}}\right)\right]. \tag{23}$$

The form in (21) specifies $x_i|w,a,\sigma^2,\mathbf{x}_{-i},\mathbf{y}$ as a Bernoulli-truncated Gaussian variable with parameter (w_i,μ_i,η_i^2). Appendix III presents an algorithm that can be used to generate samples from this distribution.

To summarize, generating samples distributed according to $f(\mathbf{x}|w,\sigma^2,a,,\mathbf{y})$ can be performed by updating the coordinates of $\mathbf{x}$ successively using M Gibbs moves (requiring generation of Bernoulli-truncated Gaussian variables).

B. Generation of samples according to $f(\sigma^2|\mathbf{x},\mathbf{y})$

Samples are generated as the following way:

$$\sigma^2|\mathbf{x},\mathbf{y} \sim \mathcal{IG}\left(\frac{P}{2},\frac{\|\mathbf{y}-T(\boldsymbol{\kappa},\mathbf{x})\|^2}{2}\right). \tag{24}$$

V. Simulation on synthetic images

A. Reconstruction of 2-dimensional image

In this subsection, a 32×32 synthetic image, depicted in Fig. 2 (right), is simulated using the prior in (9) with parameter $a=1$ and $w=0.02$. In this figure and in the following ones, white pixels stands for identically null values. A general analytical derivation of the psf of the

TABLE I

PARAMETERS USED TO COMPUTE THE MRFM PSF.

Parameter		Value
Description	Name	
Amplitude of external magnetic field	B_{ext}	9.4×10^3 G
Value of B_{mag} in the resonant slice	B_{res}	1.0×10^4 G
Radius of tip	R_0	4.0 nm
Distance from tip to sample	d	6.0 nm
Cantilever tip moment	m	4.6×10^5 emu
Peak cantilever oscillation oscillation	x_{pk}	0.8 nm
Maximum magnetic field gradient	G_{max}	125

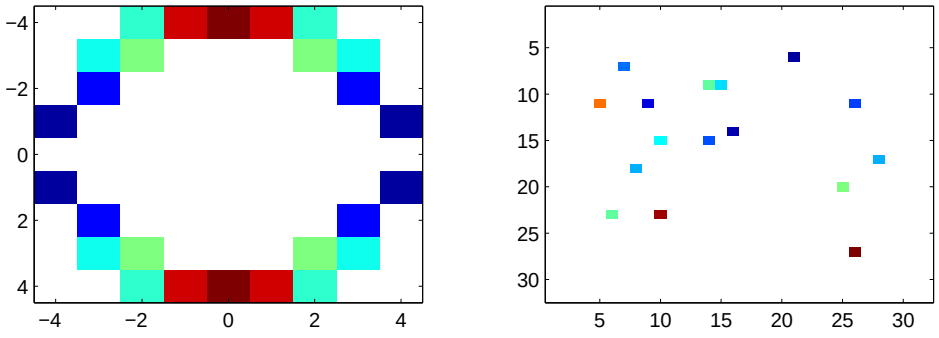

Fig. 2. Left: Psf of the MRFM tip. Right: unknown sparse image to be estimated.

MRFM tip has been given in [32] and is explained in [51]. Following this model, a 10×10 2-dimensional convolution kernel, represented in Fig. 2 (left), has been generated when the physical parameters are tuned to the values gathered in Table I. The corresponding matrix $\mathbf{H}$ introduced in (2) is of size 1024×1024. The observed measurements $\mathbf{y}$, depicted in Fig. 2 (right) are of size $P = 1024$. These observations are corrupted by an additive Gaussian noise with two different variances $\sigma^2 = 1.2 \times 10^{-1}$ and $\sigma^2 = 1.6 \times 10^{-3}$, corresponding to signal-to-noise ratios $\text{SNR} = 2\text{dB}$ and $\text{SNR} = 20\text{dB}$ respectively.

1) Simulation results: The observations are processed by the proposed algorithm that consists of $N_{\text{MC}} = 2000$ iterations of the Gibbs sampler with $N_{\text{bi}} = 300$ burn-in iterations. Then the MAP estimator of the unknown image $\mathbf{x}$ is computed by retaining among $\mathcal{X} = \left\{\mathbf{x}^{(t)}\right\}_{t=1,\ldots,N_{\text{MC}}}$ the generated sample that maximizes the posterior distribution in (17):

$$\begin{aligned}\hat{\mathbf{x}}_{\text{MAP}} &= \underset{\mathbf{x}\in\mathbb{R}_+^M}{\operatorname{argmax}}\, f\left(\mathbf{x}|\mathbf{y}\right) \\ &\approx \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmax}}\, f\left(\mathbf{x}|\mathbf{y}\right).\end{aligned} \tag{25}$$

These estimates are depicted in Fig. 3 for the two levels of noise considered. Observe that the estimated image is very similar to the actual image, even at low SNR.

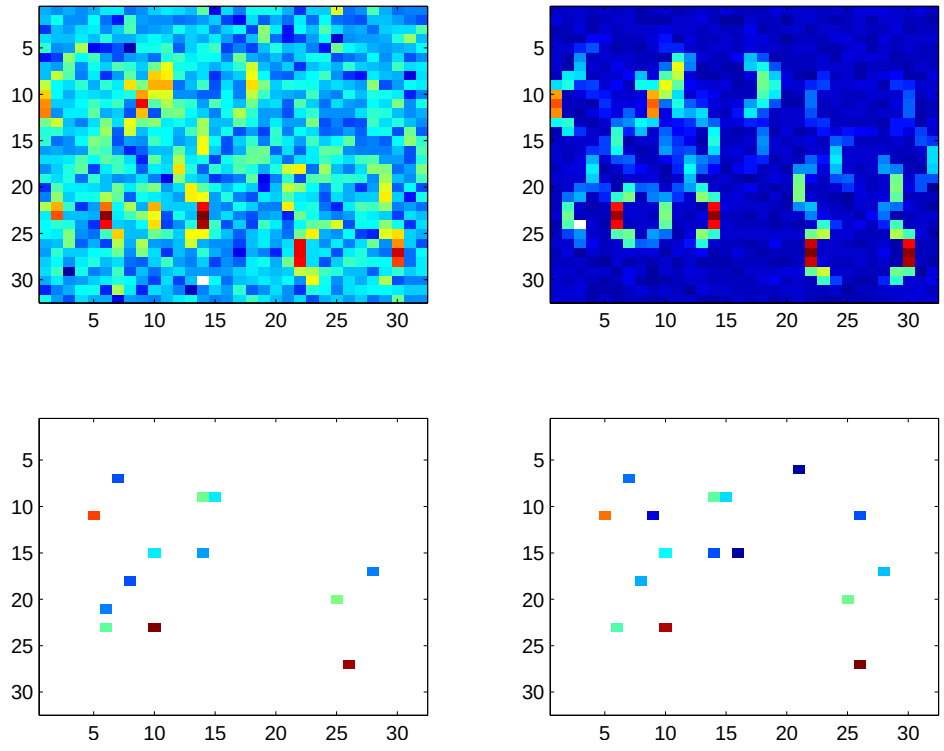

Fig. 3. Top, left (resp. right): noisy observations for SNR $=$ 2dB (resp. 20dB). Bottom, left (resp. right): reconstructed image for SNR $=$ 2dB (resp. 20dB).

Moreover, as the proposed algorithm generates samples distributed according to the posterior distribution in (17), these samples can be used to compute the posterior distributions of each parameter. As examples, the posterior distributions of the hyperparameters a and w, as well as the noise variance σ^2, are shown in Fig. 4, 5 and 6. These estimated distributions are in good

agreement with the ground truth values of these parameters, randomly drawn from the prior distribution.

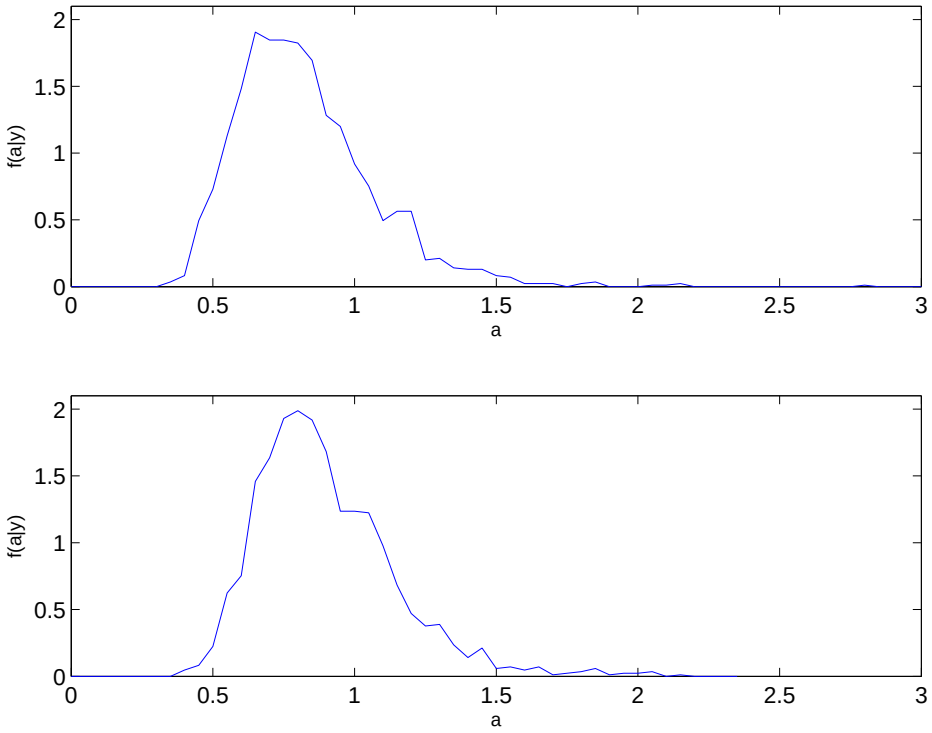

Fig. 4. Posterior distribution of hyperparameter a (left: $\mathrm{SNR} = 2$dB, right: $\mathrm{SNR} = 20$dB).

The posterior distributions of four different pixels are depicted in Fig. 7. These posteriors are consistent with the actual values of these pixels that are represented as dotted red lines in these figures.

2) Comparison of reconstruction performances: Here we compare our proposed hierarchical Bayesian method to the methods of [49], [51]. The techniques proposed in [49], [51] are based on penalized likelihood EM algorithms that perform empirical estimation of the unknown hyperparameters. Therein, two empirical Bayesian estimators, denoted Emp-MAP-Lap and Emp-MAP-LAZE, based on a Laplacian or a LAZE prior respectively, were proposed. We also compare with the estimator provided by a standard Landweber algorithm [29]. These will be compared to our hierarchical Bayesian MAP estimator, given in (25), and also to a minimum mean square error (MMSE) estimator extracted from the estimated full Bayes posterior (17). The MMSE estimator of the unknown parameter $\mathbf{x}$ is obtained by an empirical averaging over

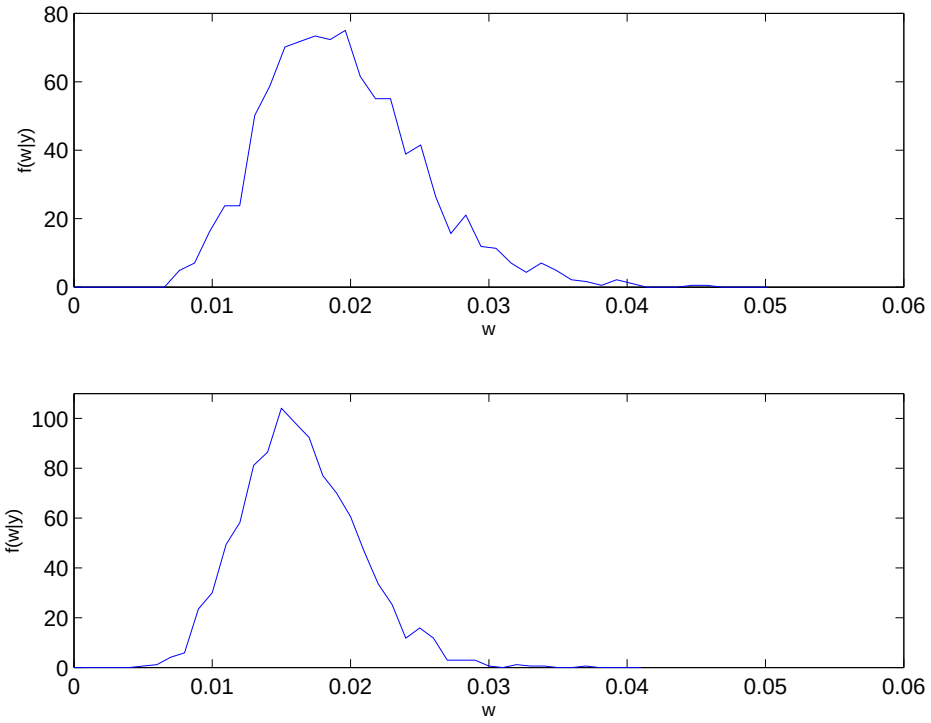

Fig. 5. Posterior distribution of hyperparameter w (left: $\mathrm{SNR} = 2$dB, right: $\mathrm{SNR} = 20$dB).

the last $N_r = 1700$ outputs of the Gibbs sampler according to:

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MMSE}} &= \mathrm{E}\left[\mathbf{x}|\mathbf{y}\right] \\ &\approx \frac{1}{N_r}\sum_{t=1}^{N_r}\mathbf{x}^{(N_{\text{bi}}+t)}. \end{aligned} \tag{26}$$

As in [51] we compare estimators with respect to several performance criteria. Let $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ denote the reconstruction error when $\hat{\mathbf{x}}$ is the estimator of the image $\mathbf{x}$ to be recovered. These criteria are: the ℓ_0, ℓ_1 and ℓ_2-norms of $\mathbf{e}$, which measures the accuracy of the reconstruction, and the ℓ_0-norm of the estimator $\hat{\mathbf{x}}$, which measures its sparsity. As pointed out in [51], small non-zero values of the pixel are usually not distinguishable from exactly zero values by a human being. Thus, a less strict measure[4] of sparsity than the ℓ_0-norm, which is denoted $\|\cdot\|_\delta$, is the

[4]The introduced measure of sparsity is denoted $\|\cdot\|_\delta$. However, it has to be mentioned that is not a norm.

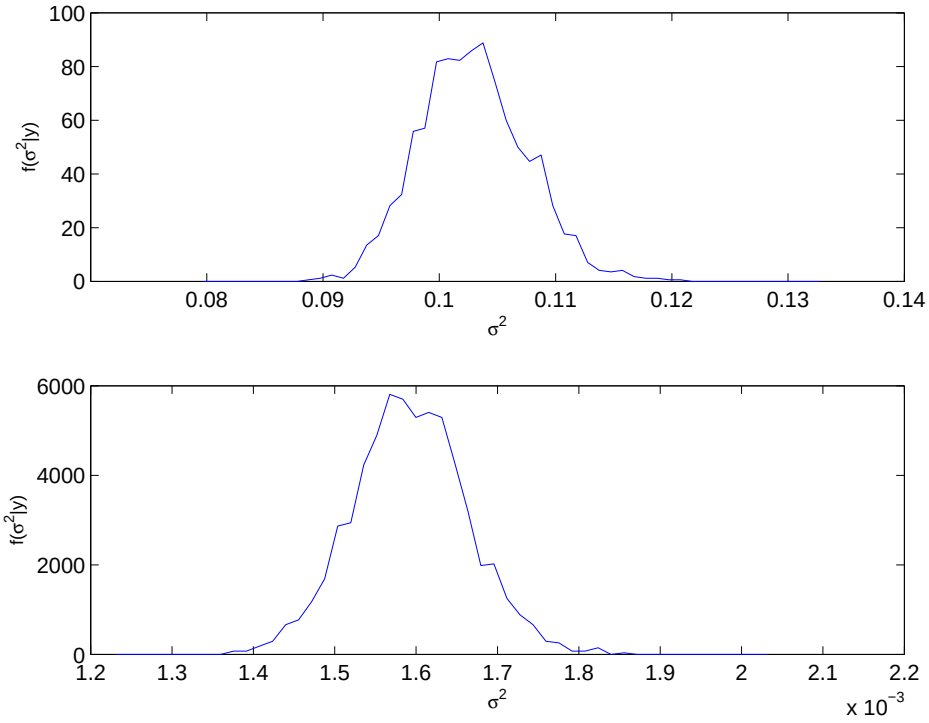

Fig. 6. Posterior distribution of hyperparameter σ^2 (left: $\mathrm{SNR} = 2$dB, right: $\mathrm{SNR} = 20$dB).

number of reconstructed image pixels that are less than a given threshold δ:

$$\begin{aligned} \|\hat{\mathbf{x}}\|_\delta &= \sum_{i=1}^{M} \mathbf{1}_{\hat{x}_i<\delta}\left(\hat{x}_i\right), \\ \|\mathbf{e}\|_\delta &= \sum_{i=1}^{M} \mathbf{1}_{e_i<\delta}\left(e_i\right). \end{aligned} \tag{27}$$

It what follows, δ has been chosen as $\delta = 10^{-2}\|\mathbf{x}\|_\infty$. To summarize, the following criteria have been computed for the image in paragraph V-A.1 for two levels of SNR: $\|e\|_0$, $\|e\|_\delta$, $\|e\|_1$, $\|e\|_2$, $\|\hat{\mathbf{x}}\|_0$ and $\|\hat{\mathbf{x}}\|_\delta$.

Table II shows the six performance measures for the five different algorithms studied. The proposed Bayesian methods (labeled "proposed MMSE" and "proposed MAP" in the table) outperform the other estimators in terms of ℓ_1 or ℓ_2-norms. Note that the MMSE estimation of the unknown image is a non sparse estimator in the ℓ_0-norm sense. This is due to the very small but non-zero posterior probability of non-zero value at many pixels. The sparsity measure $\|\cdot\|_\delta$ indicates that most of the pixels are in fact very close to zero. The MAP estimator seems to be a

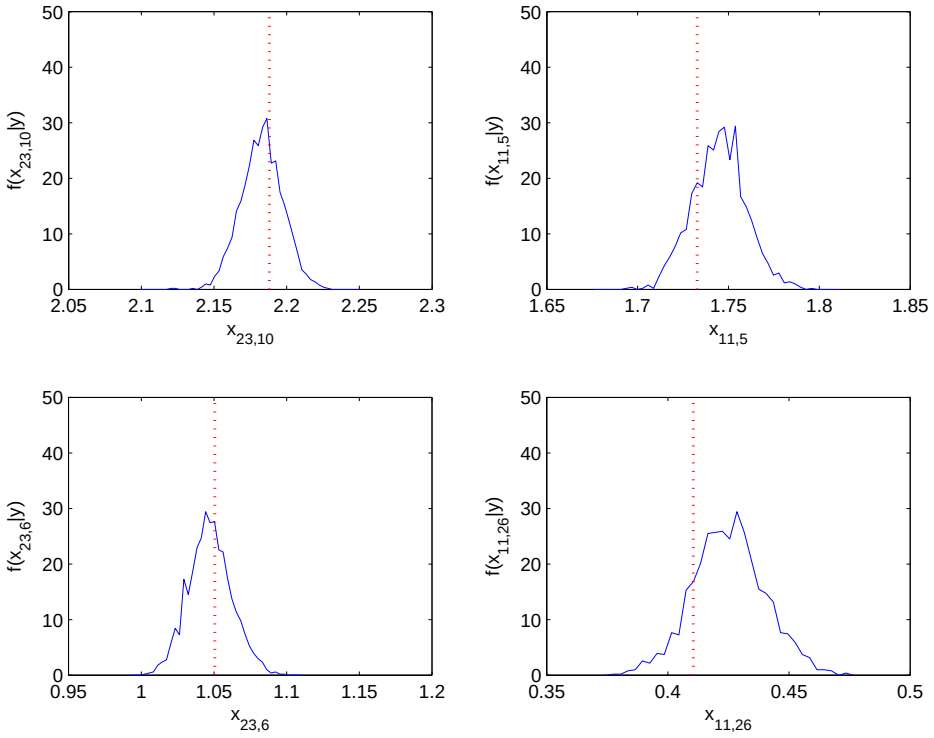

Fig. 7. Posteriors distributions of the non-zero values of $\mathbf{x}$ for $\mathrm{SNR} = 20$dB, (actual values are depicted with dotted red lines).

very powerful estimator for the sparse reconstruction problem as it seems to balance the sparsity of the solution and the minimization of the reconstruction error. However, by construction, the MMSE estimation will always have lower mean square error.

B. Reconstruction of undersampled 3-dimensional images

As discussed in Section VI, the prototype MRFM instrument collects data projections as irregularly spaced, or undersampled, spatial samples. In this subsection, we indicate how the image reconstruction algorithm can be adapted to this undersampled scenario in 3D. For concreteness, we illustrate by a concrete example. First, a $24 \times 24 \times 6$ image is generated such as 4 pixels have non-zero values in each z slice. The resulting data is depicted in Fig. 9 (right) and Fig. 8 (top). This image to be recovered is assumed to be convolved with a $5 \times 5 \times 3$ kernel that is represented in Fig. 9 (right). The resulting convolved image is depicted in Fig. 10 (left). However, the actually observed image is an undersampled version of this image. More precisely, the sampling rates are assumed to be $d_x = 2$, $d_y = 3$ $d_z = 1$ respectively in the 3 dimensions. Consequently the

TABLE II

RECONSTRUCTION PERFORMANCES FOR DIFFERENT SPARSE DECONVOLUTION ALGORITHMS.

Method	Error criterion					
	$\|e\|_0$	$\|e\|_\delta$	$\|e\|_1$	$\|e\|_2$	$\|\hat{\mathbf{x}}\|_0$	$\|\hat{\mathbf{x}}\|_\delta$
SNR = 2dB						
Landweber	1024	990	339.76	13.32	1024	990
Emp-MAP-Lap	18	17	14.13	4.40	0	0
Emp-MAP-LAZE	60	58	9.49	1.44	55	55
Proposed MMSE	1001	30	3.84	0.72	1001	27
Proposed MAP	19	16	2.38	0.81	13	13
SNR = 20dB						
Landweber	1024	931	168.85	6.67	1024	931
Emp-MAP-Lap	33	18	1.27	0.31	28	23
Emp-MAP-LAZE	144	19	1.68	0.22	144	27
Proposed MMSE	541	5	0.36	0.11	541	16
Proposed MAP	19	7	0.39	0.13	16	16

observed 3D image, shown in Fig. 10, is of size $12 \times 8 \times 6$. Finally, an i.i.d. Gaussian noise with $\sigma = 0.02$ is added following the model in (1). Note that under these assumptions, the application $T(\cdot, \cdot)$ can be split into two standard operations following the composition:

$$T(\boldsymbol{\kappa}, \mathbf{X}) = g_{d_x,d_y,d_z}(\boldsymbol{\kappa} \otimes \mathbf{X}), \tag{28}$$

where $g_{d_x,d_y,d_z}(\cdot)$ stands for the undersampling function.

For illustration the proposed hierarchical Bayesian algorithm is used to perform the sparse reconstruction with undersampled data. The number of Monte Carlo runs was fixed to $N_{\text{MC}} = 2000$ with $N_{\text{bi}} = 300$ burn-in iterations. Figure 8 shows the result of applying the proposed MAP estimator to the estimated posterior.

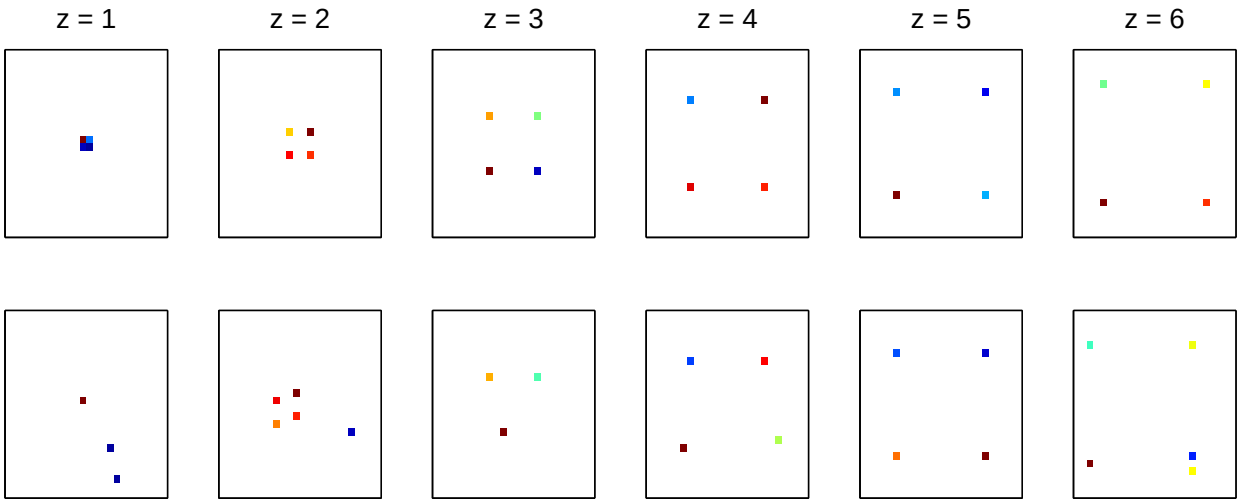

Fig. 8. Top: slices of the sparse image to be recovered. Bottom: slices of the estimated sparse image.

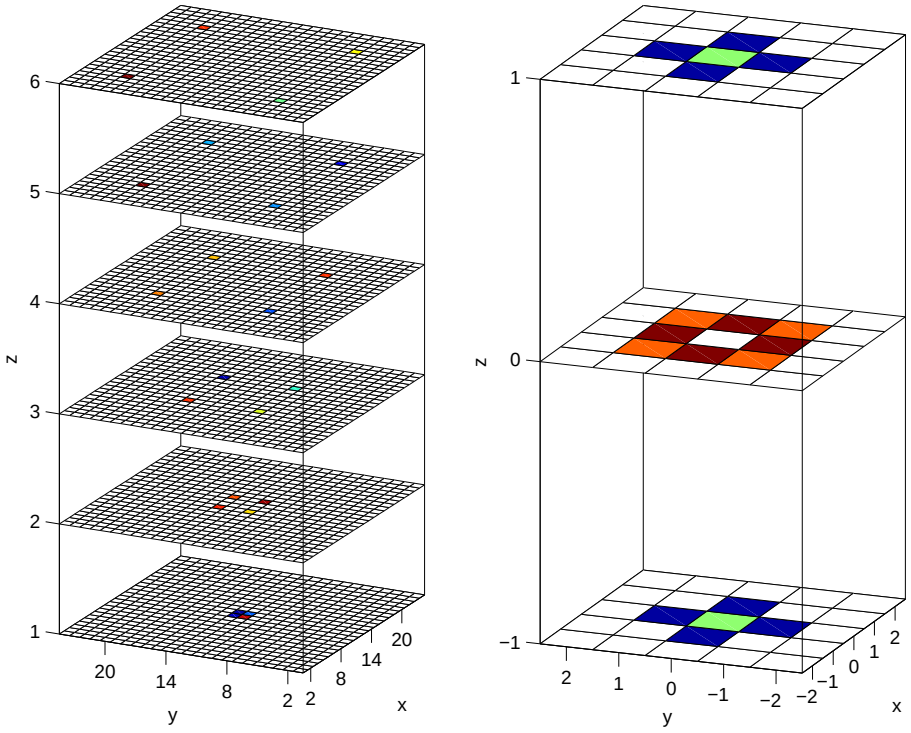

Fig. 9. Left: $24 \times 24 \times 6$ unknown image to be recovered. Right: $5 \times 5 \times 3$ kernel modeling the psf.

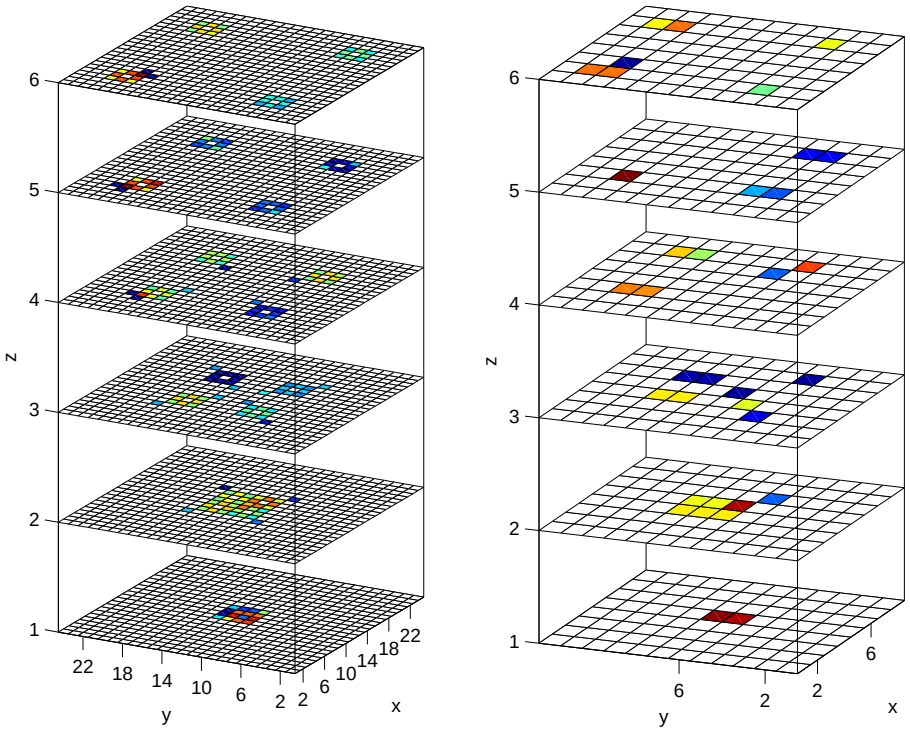

Fig. 10. Left: $24 \times 24 \times 6$ regularly sampled convolved image. Left: $12 \times 8 \times 6$ undersampled observed image.

VI. Application on real MRFM images

Here we illustrate the hierarchical Bayesian MAP reconstruction algorithm for real three dimensional MRFM data. The data is a set of MRFM projections of a sample of tobacco virus. Comprehensive details of both the experiment and the MRFM data acquisition protocol are given in [8]. The observed sample consists of a collection of Tobacco mosaic virus particles that are divided into a whole viral segment in addition to viral fragments. The projections are computed from the measured proton distribution and the 3-dimensional psf following the protocol described in [8] and [9]. The resulting scan data are depicted in Figure 11 (top) for four different distances between the MRFM tip and the sample: $d = 24$nm, $d = 37$nm, $d = 50$nm and $d = 62$nm. Each of these x-y slices is of size 60×32 pixels.

These experimental data are undersampled, i.e. the spatial resolution of the MRFM tip, and therefore the psf function, is finer than the resolution of the observed slices. Consequently, these data have been deconvolved taking into account the oversampling rates defined by $d_x = 3$, $d_y = 2$

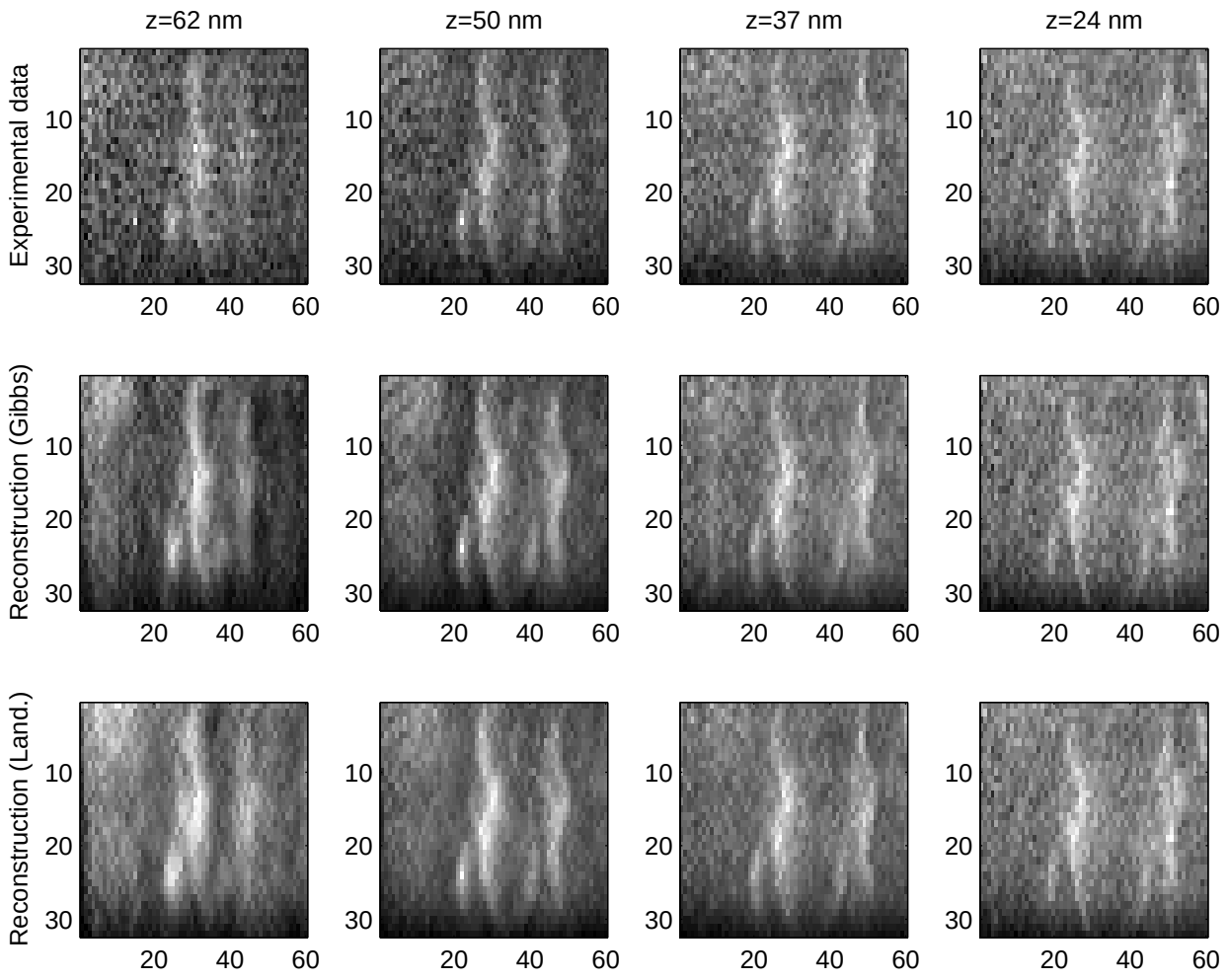

Fig. 11. Top: experimental scan data. Middle: scan data reconstructed from the proposed hierarchical Bayesian algorithm. Bottom: scan data reconstructed from the Landweber algorithm.

and $d_z = 3$ in the three directions. The MAP estimate of the unknown image is computed from $N_{\mathrm{MC}} = 1000$ Gibbs samples of the proposed Bayesian algorithm initialized with the output of a single Landweber iteration. Several more iterations of the Landweber algorithm would produce the reconstructions reported in [35]. The image reconstructions produced by the Landweber and Bayesian MAP algorithm are shown in Figs. 11-13. Three horizontal slices of the estimated image[5] are depicted in Figure 12. A 3-dimensional view of the estimated profile of the virus

[5]Note that most part of the estimated 3 dimensional image is empty space due to the very localizated position of the imaged data.

fragments is shown in Figure 13. The MMSE estimates of the parameters introduced in Section III are $\hat{\sigma}^2_{\text{MMSE}} = 0.10$, $\hat{a}_{\text{MMSE}} = 1.9 \times 10^{-12}$ and $\hat{w}_{\text{MMSE}} = 1.4 \times 10^{-2}$.

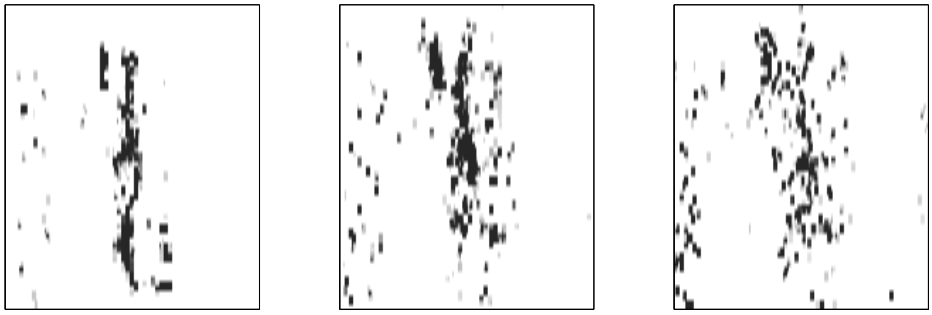

Fig. 12. Three horizontal slices of the estimated image.

By forward projecting the estimated virus image through the point spread function one can visually evaluate the goodness of fit of the reconstruction to the raw measured data. This is depicted in Fig. 11. These figures are clearly in good agreement with the observed data (top). To evaluate the convergence speed, the reconstruction error is represented in Figure 14 as a function of the iterations for the proposed Bayesian and the Landweber algorithms. This shows that the convergence rate of our algorithm is significantly better than the Landweber algorithm.

VII. Conclusions

This paper presented a hierarchical Bayesian algorithm for deconvolving sparse positive images corrupted by additive Gaussian noise. A Bernoulli-truncated exponential distribution was proposed as prior distribution for the sparse image to be recovered. The unknown hyperparameters of the model were integrated out from the posterior distribution of the image producing a full posterior distribution that can be used for estimation of the pixel values by maximization (MAP) or integration (MMSE). An efficient Gibbs sampler was used to generate samples according to this posterior distribution. The derived Bayesian estimators significantly outperformed several previously proposed sparse reconstruction algorithms. Our approach was implemented on real MRFM data to form a 3D image of a tobacco virus. Future work will include extension of the proposed method to other sparse bases, inclusion of uncertain point spread functions, and investigation of molecular priors.

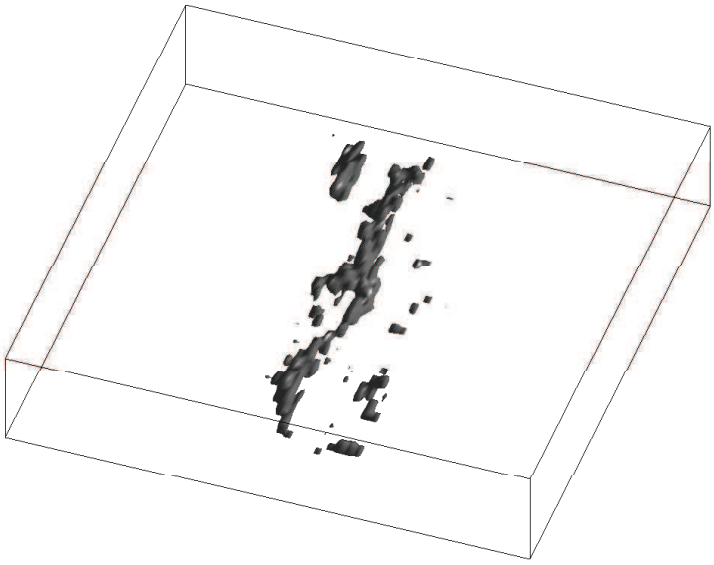

Fig. 13. 3-dimensional view of the estimated profile of the Tobacco virus fragments.

APPENDIX I
DERIVATION OF THE CONDITIONAL

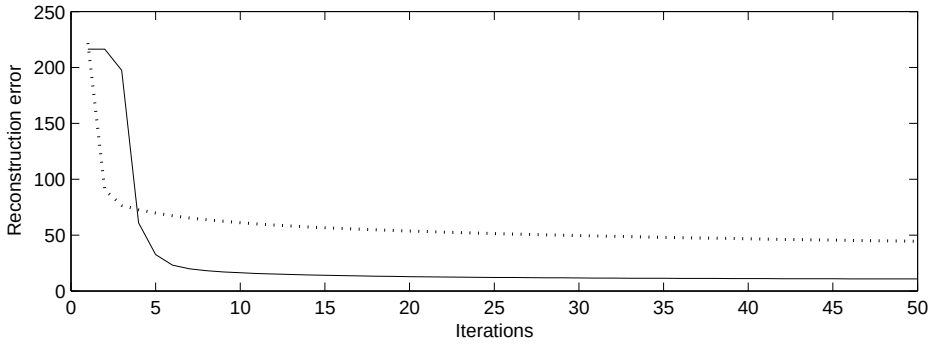

Fig. 14. Error reconstructions as functions of the iteration number for the proposed algorithm (continuous blue line) and Landweber algorithm (dotted red line).

POSTERIOR DISTRIBUTION $f\left(x_i\,|w, a, \sigma^2, \mathbf{x}_{-i}, \mathbf{y}\right)$

The posterior distribution of each component x_i ($i = 1, \ldots, M$) conditionally upon the others is linked to the likelihood function (3) and the prior distribution (7) via the Bayes' formula:

$$f\left(x_i|w, a, \sigma^2, \mathbf{x}_{-i}, \mathbf{y}\right) \propto f\left(\mathbf{y}|\mathbf{x}, \sigma^2\right) f\left(x_i|w, a\right). \tag{29}$$

This distribution can be easily derived by decomposing $\mathbf{x}$ on the standard orthonormal basis

$$\mathbb{B} = \left\{\mathbf{u}_1, \ldots, \mathbf{u}_M\right\}, \tag{30}$$

where $\mathbf{u}_i$ is the ith column of the $M \times M$ identity matrix. Indeed, let decompose

$$\mathbf{x} = \tilde{\mathbf{x}}_i + x_i\mathbf{u}_i, \tag{31}$$

where $\tilde{\mathbf{x}}_i$ is the vector $\mathbf{x}$ whose ith element has been replaced by 0. Then the linear property of the operator $T\left(\boldsymbol{\kappa}, \cdot\right)$ allows one to state:

$$T\left(\boldsymbol{\kappa}, \mathbf{x}\right) = T\left(\boldsymbol{\kappa}, \tilde{\mathbf{x}}_i\right) + x_i T\left(\boldsymbol{\kappa}, \mathbf{u}_i\right). \tag{32}$$

Consequently, (29) can be rewritten

$$\begin{aligned} f\left(x_i|w, a, \sigma^2, \mathbf{x}_{-i}, \mathbf{y}\right) &\propto \exp\left(-\frac{\left\|\mathbf{e}_i - x_i\mathbf{h}_i\right\|^2}{2\sigma^2}\right) \\ &\times \left[(1-w)\delta\left(x_i\right) + \frac{w}{a}\exp\left(-\frac{x_i}{a}\right)\mathbf{1}_{\mathbb{R}_+^*}\left(x_i\right)\right], \end{aligned} \tag{33}$$

where[6]

$$\begin{cases} \mathbf{e}_i = \mathbf{y} - T\left(\boldsymbol{\kappa}, \tilde{\mathbf{x}}_i\right), \\ \mathbf{h}_i = T\left(\boldsymbol{\kappa}, \mathbf{u}_i\right). \end{cases} \tag{34}$$

An efficient way to compute $\mathbf{e}_i$ within the Gibbs sampler scheme is reported in Appendix II. Then, straightforward computations similar to those in [7] and [34, Annex B] yield to the following distribution:

$$\begin{aligned} f\left(x_i | w, a, \sigma^2, \mathbf{x}_{-i}, \mathbf{y}\right) \propto (1 - w_i)\delta\left(x_i\right) \\ + w_i \phi_+\left(x_i | \mu_i, \eta_i^2\right), \end{aligned} \tag{35}$$

with

$$\begin{cases} \eta_i^2 = \dfrac{\sigma^2}{\|\mathbf{h}_i\|^2}, \\ \mu_i = \eta_i^2 \left(\dfrac{\mathbf{h}_i^T \mathbf{e}_i}{\sigma^2} - \dfrac{1}{a} \right), \end{cases} \tag{36}$$

and

$$\begin{cases} u_i = \dfrac{w}{a} C\left(\mu_i, \eta_i^2\right) \exp\left(\dfrac{\mu_i^2}{2\eta_i^2}\right), \\ w_i = \dfrac{u_i}{u_i + (1 - w)}. \end{cases} \tag{37}$$

The distribution in (35) is a Bernoulli-truncated Gaussian distribution with hidden mean μ_i and hidden variance η_i^2.

APPENDIX II
FAST RECURSIVE COMPUTATIONS FOR SIMULATING ACCORDING TO $f\left(\mathbf{x} | w, a, \sigma^2, \mathbf{y}\right)$

In the Gibbs sampling strategy presented in Section IV, the main computationally expensive task is the generation of samples distributed according to $f\left(x_i | w, a, \sigma^2, \mathbf{x}_{-i}, \mathbf{y}\right)$. Indeed, the evaluation of the hidden mean and hidden variance in (36) of the Bernoulli-truncated Gaussian distribution may be costly, especially when the bilinear application $T\left(\cdot, \cdot\right)$ is not easily computable. In this appendix, an appropriate recursive strategy is proposed to accelerate the Gibbs sampling by efficiently updating the coordinate i of the vector $\mathbf{x}$ at iteration t of the Gibbs sampler.

[6]It can be noticed that, for deblurring applications, $\mathbf{h}_i$ is also the ith column of the matrix $\mathbf{H}$ introduced in (2).

Let $\mathbf{x}^{(t,i-1)}$ denote the current Monte Carlo state of the unknown vectorized image $\mathbf{x}$ ($i = 1,\ldots,M$):

$$\mathbf{x}^{(t,i-1)} = \left[x_1^{(t)},\ldots,x_{i-1}^{(t)},x_i^{(t-1)},x_{i+1}^{(t-1)},\ldots,x_M^{(t-1)}\right]^T. \tag{38}$$

with, by definition, $\mathbf{x}^{(t,0)} = \mathbf{x}^{(t-1,M)}$. Updating $\mathbf{x}^{(t,i-1)}$ consists of drawing $x_i^{(t)}$ according to the Bernoulli-truncated Gaussian distribution $f\left(x_i \left| w,a,\sigma^2,\mathbf{x}_{-i}^{(t,i-1)},\mathbf{y}\right.\right)$ in (21) with:

$$\mathbf{x}_{-i}^{(t,i-1)} = \left[x_1^{(t)},\ldots,x_{i-1}^{(t)},x_{i+1}^{(t-1)},\ldots,x_M^{(t-1)}\right]^T. \tag{39}$$

The proposed strategy to simulate efficiently according to (21) is based on the following property.

Property: Given the quantity $T\left(\boldsymbol{\kappa},\mathbf{x}^{(0)}\right)$ and the vectors $\{\mathbf{h}_i\}_{i=1,\ldots,M}$, simulating according to $f\left(x_i \left| w,a,\sigma^2,\mathbf{x}_{-i}^{(t,i)},\mathbf{y}\right.\right)$ can be performed without evaluating the bilinear function $T(\cdot,\cdot)$.

Proof: Simulating according to (21) mainly requires to compute the vector $\mathbf{e}_i$ introduced by (34):

$$\mathbf{e}_i = \mathbf{y} - T\left(\boldsymbol{\kappa},\tilde{\mathbf{x}}_i^{(t,i-1)}\right), \tag{40}$$

with

$$\tilde{\mathbf{x}}_i^{(t,i-1)} = \left[x_1^{(t)},\ldots,x_{i-1}^{(t)},0,x_{i+1}^{(t-1)},\ldots,x_M^{(t-1)}\right]^T. \tag{41}$$

Moreover, by using the decomposition in (31) and by exploiting the linear property of $T(\boldsymbol{\kappa},\cdot)$, the vector $T\left(\boldsymbol{\kappa},\tilde{\mathbf{x}}_i^{(t,i-1)}\right)$ in the right-hand side of (40) can be rewritten as:

$$T\left(\boldsymbol{\kappa},\tilde{\mathbf{x}}_i^{(t,i-1)}\right) = T\left(\boldsymbol{\kappa},\mathbf{x}^{(t,i-1)}\right) - x_i^{(t-1)}\mathbf{h}_i, \tag{42}$$

where $\mathbf{h}_i$ has been introduced in (34). Consequently, to prove the property, we have to demonstrate that the vector series $\left\{T\left(\boldsymbol{\kappa},\mathbf{x}^{(t,k)}\right)\right\}_{k=1,\ldots,M}$ can be computed recursively without using $T(\cdot,\cdot)$. Assume that $T\left(\boldsymbol{\kappa},\mathbf{x}^{(t,i-1)}\right)$ is available at this stage of the Gibbs sampling and that $x_i^{(t)}$ has been drawn. The new Monte Carlo state is then:

$$\mathbf{x}^{(t,i)} = \left[x_1^{(t)},\ldots,x_{i-1}^{(t)},x_i^{(t)},x_{i+1}^{(t-1)},\ldots,x_M^{(t-1)}\right]^T. \tag{43}$$

Similarly to (42), the vector $T\left(\boldsymbol{\kappa},\mathbf{x}^{(t,i)}\right)$ can be decomposed as follows:

$$T\left(\boldsymbol{\kappa},\mathbf{x}^{(t,i)}\right) = T\left(\boldsymbol{\kappa},\tilde{\mathbf{x}}_i^{(t,i-1)}\right) + x_i^{(t)}\mathbf{h}_i. \tag{44}$$

Therefore, combining (42) and (44) allow one to state:

$$T\left(\boldsymbol{\kappa}, \mathbf{x}^{(t,i)}\right) = T\left(\boldsymbol{\kappa}, \mathbf{x}^{(t,i-1)}\right) + \left(x_i^{(t)} - x_i^{(t-1)}\right)\mathbf{h}_i.$$

■

The bilinear function $T(\cdot,\cdot)$ only needs to be used at the very beginning of the Gibbs sampling algorithm to evaluate $T\left(\boldsymbol{\kappa}, \mathbf{x}^{(0)}\right)$ and the vectors $\{\mathbf{h}_i\}_{i=1,\ldots,M}$. The resulting simulation scheme corresponding to step 3 of Algorithm 1 is shown in Algorithm 2.

ALGORITHM 2:

Efficient simulation according to $f\left(\mathbf{x}\left|w, a, \sigma^2, \mathbf{y}\right.\right)$

For $i = 1, \ldots, M$, update the ith coordinate of the vector

$$\mathbf{x}^{(t,i-1)} = \left[x_1^{(t)}, \ldots, x_{i-1}^{(t)}, x_i^{(t-1)}, x_{i+1}^{(t-1)}, \ldots, x_M^{(t-1)}\right]^T$$

via the following steps:

1. compute $\|\mathbf{h}_i\|^2$,
2. set $T\left(\boldsymbol{\kappa}, \tilde{\mathbf{x}}_i^{(t,i-1)}\right) = T\left(\boldsymbol{\kappa}, \mathbf{x}^{(t,i-1)}\right) - x_i^{(t-1)}\mathbf{h}_i$,
3. set $\mathbf{e}_i = \mathbf{x} - T\left(\boldsymbol{\kappa}, \tilde{\mathbf{x}}_i^{(t,i-1)}\right)$,
4. compute μ_i, η_i^2 and w_i as defined in (36) and (37),
5. draw $x_i^{(t)}$ according to (21),
6. set $\mathbf{x}^{(t,i)} = \left[x_1^{(t)}, \ldots, x_{i-1}^{(t)}, x_i^{(t)}, x_{i+1}^{(t-1)}, \ldots, x_M^{(t-1)}\right]^T$,
7. set $T\left(\boldsymbol{\kappa}, \mathbf{x}^{(t,i)}\right) = T\left(\boldsymbol{\kappa}, \tilde{\mathbf{x}}_i^{(t,i-1)}\right) + x_i^{(t)}\mathbf{h}_i$.

APPENDIX III

SIMULATION ACCORDING TO A

BERNOULLI-TRUNCATED GAUSSIAN DISTRIBUTION

This appendix describes how we generate random variables distributed according to a Bernoulli-truncated Gaussian distribution with parameters (w, m, s^2) whose pdf is:

$$f\left(x|\lambda, m, s^2\right) = (1-\lambda)\,\delta(x) + \frac{\lambda}{C\left(m, s^2\right)} \exp\left[-\frac{(x-m)^2}{2s^2}\right] \mathbf{1}_{\mathbb{R}_+^*}(x)$$

where $C\left(m, s^2\right)$ has been defined in (23). Monte Carlo draws from this density can be obtained by using an auxiliary binary variable ε following the strategy shown in Algorithm 3.

ALGORITHM 3:

Simulation according to

a Bernoulli-truncated Gaussian distribution

1. generate ε according to $\varepsilon \sim \mathcal{B}er\left(\lambda\right)$,
2. set $\begin{cases} x = 0, & \text{if } \varepsilon = 0; \\ x \sim \mathcal{N}^+\left(m, s^2\right), & \text{if } \varepsilon = 1. \end{cases}$

In Algorithm 3, $\mathcal{B}er\left(\cdot\right)$ and $\mathcal{N}^+\left(\cdot,\cdot\right)$ denote the Bernoulli and the positive truncated Gaussian distributions respectively. In step 2, samples distributed according to the truncated Gaussian distribution can be generated by using an appropriate accept-reject procedure with instrumental distributions [17], [33], [39].

ACKNOWLEDGEMENTS

The authors would like to thank M. Ting for providing the code to generate point spread functions of MRFM tip and for interesting suggestions regarding this work. The authors are also very grateful to Dr. Daniel Rugar who provided the real data used in Section VI as well as a valuable feedback about this paper.

REFERENCES

[1] S. Alliney and S. A. Ruzinsky, "An algorithm for the minimization of mixed l1 and l2 norms with application to Bayesian estimation," *IEEE Trans. Signal Processing*, vol. 42, no. 3, pp. 618–627, March 1994.

[2] H. Andrews and B. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

[3] S. Bourguignon and H. Carfantan, "Bernoulli-Gaussian spectral analysis of unevenly spaced astrophysical data," in *Proc. IEEE Workshop on Stat. Signal Processing (SSP)*, Bordeaux, France, July 2005, pp. 811–816.

[4] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Sel. Topics Signal Processing*, vol. 1, no. 4, Dec. 2007.

[5] F. Champagnat, Y. Goussard, and J. Idier, "Unsupervised deconvolution of sparse spike trains using stochastic approximation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 2988–2998, Dec. 1996.

[6] S. Chao, W. M. Dougherty, J. L. Garbini, and J. A. Sidles, "Nanometer-scale magnetic resonance imaging," *Review Sci. Instrum.*, vol. 75, no. 5, pp. 1175–1181, April 2004.

[7] Q. Cheng, R. Chen, and T.-H. Li, "Simultaneous wavelet estimation and deconvolution of reflection seismic signals," *IEEE Trans. Geosci. and Remote Sensing*, vol. 34, no. 2, pp. 377–384, March 1996.

[8] C. L. Degen, M. Poggio, H. J. Mamin, C. T. Rettner, and D. Rugar, "Nanoscale magnetic resonance imaging," *Proc. Nat. Academy of Science*, 2008, submitted.

[9] ——, "Nanoscale magnetic resonance imaging. Supplementary information," *Proc. Nat. Academy of Science*, 2008, submitted.

[10] J. Diebolt and E. H. S. Ip., "Stochastic EM: method and application," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London: Chapman & Hall, 1996, pp. 259–273.

[11] N. Dobigeon and J.-Y. Tourneret, "Joint segmentation of wind speed and direction using a hierarchical model," *Comput. Stat. & Data Analysis*, vol. 51, no. 12, pp. 5603–5621, Aug. 2007.

[12] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 2684–2695, July 2008.

[13] N. Dobigeon, J.-Y. Tourneret, and M. Davy, "Joint segmentation of piecewise constant autoregressive processes processes by using a hierarchical model and a Bayesian sampling approach," *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1251–1263, April 2007.

[14] N. Dobigeon, J.-Y. Tourneret, and J. D. Scargle, "Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model," *IEEE Trans. Signal Processing*, vol. 55, no. 2, pp. 414–423, Feb. 2007.

[15] A. Doucet and P. Duvaut, "Bayesian estimation of state-space models applied to deconvolution of Bernoulli-Gaussian processes," *Signal Processing*, vol. 57, no. 2, pp. 147–161, March 1997.

[16] C. Févotte, B. Torrésani, L. Daudet, , and S. J. Godsill, "Sparse linear regression with structured priors and application to denoising of musical audio," *IEEE Trans. Audio, Speech, Language Processing*, vol. 16, no. 1, pp. 174–185, Jan. 2008.

[17] J. Geweke, "Efficient simulation from the multivariate normal and Student-T distributions subject to linear constraints," in *Computing Science and Statistics, Proc. of the 23th Symposium on the Interface*, E. M. Keramidas, Ed. Fairfax, VA: Interface Foundation of North America, Inc., 1991, pp. 571–578.

[18] S. J. Godsill and P. J. W. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech, Audio Proc.*, vol. 6, no. 4, pp. 352–372, July 1998.

[19] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[20] P. C. Hammel, D. V. Pelekhov, P. E. Wigen, T. R. Gosnell, M. M. Midzor, and M. L. Roukes, "The Magnetic-Resonance Force Microscope: A new tool for high-resolution, 3-D, subsurface scanned probe imaging," *Proc. IEEE*, vol. 91, no. 5, pp. 789–798, May 2003.

[21] K. Herrity, R. Raich, and A. O. Hero, "Blind deconvolution for sparse molecular imaging," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal (ICASSP)*, Las Vegas, USA, April 2008, pp. 545–548.

[22] J. Idier and Y. Goussard, "Stack algorithm for recursive deconvolution of Bernoulli-gaussian processes," *IEEE Trans. Signal Processing*, vol. 28, no. 5, pp. 67–79, Sept. 1990.

[23] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proc. of the Royal Society of London. Series A*, vol. 186, no. 1007, pp. 453–461, 1946.

[24] ——, *Theory of Probability*, 3rd ed. London: Oxford University Press, 1961.

[25] I. M. Johnstone and B. W. Silverman, "Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences," *Ann. Stat.*, vol. 32, no. 4, pp. 1594–1649, 2004.

[26] J. J. Kormylo and J. M. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, pp. 482–488, May 1982.

[27] S. Kuehn, S. A. Hickman, and J. A. Marohn, "Advances in mechanical detection of magnetic resonance," *J. Chemical Physics*, vol. 128, no. 5, Feb. 2008.

[28] E. Kuhn and M. Lavielle, "Coupling a stochastic approximation version of EM with an MCMC procedure," *ESAIM Probab. Statist.*, vol. 8, pp. 115–131, 2004.

[29] L. Landweber, "An iteration formula for Fredholm integral equations of the first kind," *Amer. J. Math.*, vol. 73, no. 3, pp. 615–624, July 1951.

[30] M. Lavielle and E. Lebarbier, "An application of MCMC methods for the multiple change-points problem," *Signal Processing*, vol. 81, no. 1, pp. 39–53, Jan. 2004.

[31] M. Lavielle, "Bayesian deconvolution of Bernoulli-Gaussian processes," *Signal Processing*, vol. 33, no. 1, pp. 67–79, July 1993.

[32] J. Mamin, R. Budakian, and D. Rugar, "Point response function of an MRFM tip," IBM Research Division, Tech. Rep., Oct. 2003.

[33] V. Mazet, D. Brie, and J. Idier, "Simulation of postive normal variables using several proposal distributions," in *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, Bordeaux, France, July 2005, pp. 37–42.

[34] V. Mazet, "Développement de méthodes de traitement de signaux spectroscopiques : estimation de la ligne de base et du spectre de raies," Ph.D. dissertation, Univ. Henri Poincaré, Nancy, France, Dec. 2006, in French.

[35] D. Mounce, "Magnetic resonance force microscopy," *IEEE Instr. Meas. Magazine*, vol. 8, no. 2, pp. 20–26, June 2005.

[36] T. Olofsson and E. Wennerström, "Sparse deconvolution of B-scan images," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 54, no. 8, Aug. 2007.

[37] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 747–758, March 2002.

[38] S. E. Reichenbach, D. E. Koehler, and D. W. Strelow, "Restoration and reconstruction of AVHRR images," *IEEE Trans. Geosci. and Remote Sensing*, vol. 33, no. 4, pp. 997–1007, July 1995.

[39] C. P. Robert, "Simulation of truncated normal variables," *Statistics and Computing*, vol. 5, no. 2, pp. 121–125, June 1995.

[40] ——, *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*, 2nd ed., ser. Springer Texts in Statistics. New York: Springer-Verlag, 2007.

[41] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

[42] O. Rosec, J.-M. Boucher, B. Nsiri, and T. Chonavel, "Blind marine seismic deconvolution using statistical MCMC methods," *IEEE J. Ocean. Eng.*, vol. 28, no. 3, pp. 502–512, July 2003.

[43] D. Rugar, R. Budakian, H. J. Mamin, and B. W. Chui, "Single spin detection by magnetic resonance force microscopy," *Nature*, vol. 430, pp. 329–332, July 2004.

[44] J. C. Russ, *The image processing handbook*, 5th ed. Boca Raton, FL: CRC Press, 2006.

[45] P. Sarder and A. Nehorai, "Deconvolution methods for 3-D fluorescence microscopy images," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 32–45, May 2006.

[46] J.-L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*, 2nd ed. Berlin Heidelberg: Springer-Verlag, 2006.

[47] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.

[48] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[49] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction using sparse priors," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Oct. 2006, pp. 1261–1264.

[50] ——, "Sparse image reconstruction for molecular imaging," *IEEE Trans. Image Processing*, 2008, submitted.

[51] M. Y. Ting, "Signal processing for magnetic resonance force microscopy," Ph.D. dissertation, Univ. of Michigan, Ann Arbor, MI, May 2006.

[52] F. Šroubek and J. Flusser, "Multichannel blind iterative image restoration," *IEEE Trans. Image Processing*, vol. 12, no. 9, pp. 1094–1106, Sept. 2003.

[53] O. Zuger, S. T. Hoen, C. S. Yannoni, and D. Rugar, "Three-dimensional imaging with a nuclear magnetic resonance force microscope," *J. Appl. Phys.*, vol. 79, no. 4, pp. 1881–1884, Feb. 1996.

[54] O. Zuger and D. Rugar, "Magnetic resonance detection and imaging using force microscope techniques," *J. Appl. Phys.*, vol. 75, no. 10, pp. 6211–6216, May 1994.