Bayesian sparse image reconstruction~~:~~ Application to MRFM

Nicolas Dobigeon, Alfred O. Hero and Jean-Yves Tourneret

Abstract—**This paper presents a Bayesian model to reconstruct sparse images when the observations are obtained from linear transformations and corrupted by an additive white Gaussian noise. ~~We propose an appropriate prior distribution for the image to be estimated that takes into account the sparsity and the positivity of the measurements.~~ ~~This prior is~~ based on a weighted mixture of a positive exponential distribution and a mass at zero. The hyperparameters that ~~are inherent of the model~~ are tuned automatically ~~in an unsupervised way. They are estimated in the fully Bayesian scheme, yielding a~~ hierarchical Bayesian model. To overcome the complexity of the ~~resulting~~ posterior distribution, a Gibbs sampling strategy is ~~derived to generate samples asymptotically distributed according to the posterior distribution of interest.~~ ~~These~~ samples can ~~then~~ be used to estimate the image to be recovered. ~~As the~~ posteriors of the parameters are available~~, this~~ algorithm provides more information than other previously proposed sparse reconstruction methods that only give a point estimate. The performance of ~~the proposed~~ sparse reconstruction method is illustrated on synthetic and real data ~~provided by a new nanoscale magnetic resonance imaging technique called MRFM.~~**

Index Terms—**Deconvolution, MRFM imagery, sparse representation, Bayesian inference, MCMC methods.**

I. INTRODUCTION

For several decades, image deconvolution has received increasing interest in the literature [1], [2]. 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 [3], medical imagery [4], remote sensing [5] and photography [6]. More recently, a new imaging technology, ~~so-~~called Magnetic Resonance Force Microscopy (MRFM), has been developed (see [7] and [8] for recent reviews). This non-destructive method allows one to improve the detection sensitivity of standard magnetic resonance imaging [9]. Because of ~~their~~ potential atomic-level resolution[1], the 2-dimensional or 3-dimensional images resulting from this technology are ~~characterized by their sparsit~~y. 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.

~~Deconvolution of sparse signals or images~~ has motivated research in spectral analysis in astronomy [10]~~, for~~ seismic signal analysis in geophysics [11], [12]~~, for~~ deconvolution of ultrasonic B-scans [13]~~, among other examples.~~ We propose here a ~~fully~~ Bayesian model that is based on an appropriate prior distribution for the unknown image~~. This~~ 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 [14]–[18] or more recently [19] and [20]). 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 LAZE prior. This distribution has been used in [21] to solve a denoising problem in a non-Bayesian quasi-maximum likelihood estimation framework. In [22], [23], this prior has also been used for sparse reconstruction of noisy images. ~~In this paper, a~~ new prior composed of a mass at zero and a single-sided exponential distribution is introduced~~. The main motivation of choosing this prior is to take into account the~~ positivity and ~~the~~ sparsity of the pixels in the image. The full Bayesian posterior can then be derived from samples generated by Markov chain Monte Carlo (MCMC) methods [24].

~~With the prior modeling introduced above, the results of the sparse reconstruction critically depend on the parameters chosen to define the mixture. Unfortunately, estimating~~ the ~~"~~hyperparameters~~"~~ involved in the prior distribution described above is ~~a~~ difficult task. Empirical solutions ~~have been~~ proposed in [22], [23] to deal with this issue~~. When compared with other standard methods, the results in [22], [23] are satisfactory at low signal-to-noise ratios (SNR). At high SNRs, these methods display increasingly biased estimation of the hyperparameters that can lead to unstable results.~~ 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 [25], [26] 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 fully Bayesian approach, adopted in this paper, has been suc-

Part of this work has been supported by a DGA fellowship from French Ministry of Defence and by ARO MURI grant No. W911NF-05-1-0403.

Nicolas Dobigeon was with University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA. He is now with University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France. (e-mail: nicolas.dobigeon@enseeiht.fr).

Alfred O. Hero is with University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA. (e-mail: hero@umich.edu).

Jean-Yves Tourneret is with University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France. (e-mail: jean-yves.tourneret@enseeiht.fr).

[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.

Summary of Comments on bayesian_SIR_16.pdf

Page: 1

Number: 1 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:01:17 PM

Number: 2 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:01:13 PM
Hierarchical

Number: 3 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:01:22 PM
with application to

Number: 4 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:29:46 PM
naturally sparse in the standard pixel basis

Number: 5 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:01:27 PM
hierarchical

Number: 6 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:03:27 PM

Number: 7 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:03:26 PM
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.

Number: 8 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:02:19 PM
The motivating application is magnetic resonance force microscopy (MRFM), an emerging atomic-level resolution molecular imaging modality.

Number: 9 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:30:04 PM
Sparse signal and image deconvolution

Number: 10 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:14:23 PM
We propose a prior that is

Number: 11 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:30:22 PM

Number: 12 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:30:16 PM
in many scientific applications including:

Number: 13 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:30:18 PM
;

Number: 14 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:14:39 PM

Number: 15 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:14:37 PM
prior has

Number: 16 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:30:31 PM
;

Number: 17 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:30:36 PM
and

Number: 18 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:30:48 PM

Number: 19 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:14:50 PM
by marginalization over the

Number: 20 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:30:59 PM
hierarchical

Number: 21 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:31:16 PM
selecting

Number: 22 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:15:01 PM
hierarchical Bayesian

Number: 23 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:15:24 PM

Number: 24 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:32:03 PM
and other unknown parameters

Number: 25 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:32:21 PM
The image

Number: 26 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:15:28 PM
proposed.

Number: 27 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:15:35 PM

Number: 28 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:15:33 PM
The Gibbs

Comments from page 1 continued on next page

Bayesian sparse image reconstruction. Application to MRFM

Nicolas Dobigeon, Alfred O. Hero and Jean-Yves Tourneret

***Abstract*—This paper presents a Bayesian model to reconstruct sparse images when the observations are obtained from linear transformations and corrupted by an additive white Gaussian noise. ~~We propose an appropriate prior distribution for the image to be estimated that takes into account the sparsity and the positivity of the measurements.~~ ~~This prior is~~ based on a weighted mixture of a positive exponential distribution and a mass at zero. The hyperparameters that ~~are inherent of the model~~ are tuned automatically ~~in an unsupervised way. They are estimated in the fully Bayesian scheme, yielding a~~ hierarchical Bayesian model. To overcome the complexity of the ~~resulting~~ posterior distribution, a Gibbs sampling strategy is ~~derived to generate samples asymptotically distributed according to the posterior distribution of interest~~. ~~These~~ samples can ~~then~~ be used to estimate the image to be recovered. ~~the~~ posteriors of the parameters are available, this algorithm provides more information than other previously proposed sparse reconstruction methods that only give a point estimate. The performance of the ~~proposed~~ sparse reconstruction method is illustrated on synthetic and real data ~~provided by a new nanoscale magnetic resonance imaging technique called MRFM.~~**

***Index Terms*—Deconvolution, MRFM imagery, sparse representation, Bayesian inference, MCMC methods.**

I. INTRODUCTION

For several decades, image deconvolution has received increasing interest in the literature [1], [2]. 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 [3], medical imagery [4], remote sensing [5] and photography [6]. More recently, a new imaging technology, ~~so-~~called Magnetic Resonance Force Microscopy (MRFM), has been developed (see [7] and [8] for recent reviews). This non-destructive method allows one to improve the detection sensitivity of standard magnetic resonance imaging [9]. Because of their potential atomic-level resolution[1], the 2-dimensional or 3-dimensional images resulting from this technology are ~~characterized by their sparsity~~. 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.

~~Deconvolution of sparse signals or images~~ has motivated research ~~for~~ spectral analysis in astronomy [10]~~, for~~ seismic signal analysis in geophysics [11], [12] ~~or for~~ deconvolution of ultrasonic B-scans [13]~~, among other examples.~~ We propose here a ~~fully~~ Bayesian model that is based on an appropriate prior distribution for the unknown image. ~~This~~ 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 [14]–[18] or more recently [19] and [20]). 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 LAZE prior. This distribution has been used in [21] to solve a denoising problem in a non-Bayesian quasi-maximum likelihood estimation framework. In [22], [23], this prior has also been used for sparse reconstruction of noisy images. In ~~this paper, a~~ new prior composed of a mass at zero and a single-sided exponential distribution is introduced. ~~The main motivation of choosing this prior is to take into account the~~ positivity and ~~the~~ sparsity of the pixels in the image. The full Bayesian posterior can then be derived from samples generated by Markov chain Monte Carlo (MCMC) methods [24].

~~With the prior modeling introduced above, the results of the sparse reconstruction critically depend on the parameters chosen to define the mixture. Unfortunately, estimating~~ the hyperparameters involved in the prior distribution described above is a difficult task. Empirical solutions ~~have been~~ proposed in [22], [23] to deal with this issue. ~~When compared with other standard methods, the results in [22], [23] are satisfactory at low signal-to-noise ratios (SNR). At high SNRs, these methods display increasingly biased estimation of the hyperparameters that can lead to unstable results.~~ 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 [25], [26] 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 fully Bayesian approach, adopted in this paper, has been suc-

Part of this work has been supported by a DGA fellowship from French Ministry of Defence and by ARO MURI grant No. W911NF-05-1-0403.

Nicolas Dobigeon was with University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA. He is now with University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France. (e-mail: nicolas.dobigeon@enseeiht.fr).

Alfred O. Hero is with University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA. (e-mail: hero@umich.edu).

Jean-Yves Tourneret is with University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France. (e-mail: jean-yves.tourneret@enseeiht.fr).

[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.

Number: 29 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:16:13 PM

Number: 30 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:16:09 PM
, e.g. by maximizing the estimated posterior distribution.

Number: 31 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:27:32 PM
In our fully Bayesian approach the

Number: 32 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:27:59 PM

Number: 33 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:27:52 PM
all

Number: 34 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:28:04 PM
Thus our

Number: 35 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:28:14 PM
our hierarchical Bayesian

Number: 36 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:57:39 PM
collected from a tobacco virus sample using a prototype MRFM instrument.

Number: 37 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:32:54 PM
so-called

Number: 38 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:33:06 PM
general

Number: 39 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:36:03 PM
, 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 weighting of the prior vs the likelihood function. The hierarchical Bayesian approach proposed in this paper circumvents these difficulties.

Number: 40 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:36:16 PM
Specifically, a

Number: 41 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:36:28 PM
, which accounts for

Number: 42 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:36:29 PM

Number: 43 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:36:32 PM

Number: 44 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:37:24 PM
Conjugate priors on the hyperparameters of the image prior are introduced. It is this step that makes our approach hierarchical Bayesian.

Number: 45 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:28:50 PM

Number: 46 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:37:43 PM
The estimation of

Number: 47 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:37:47 PM

Number: 48 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:37:48 PM

Number: 49 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:29:33 PM
its

Number: 50 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:37:55 PM
the most

Number: 51 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:38:04 PM
and poor estimation leads to instability

Number: 52 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:38:59 PM
Bayes (EB) and Stein unbiased risk (SURE)

Number: 53 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:38:08 PM
were

Number: 54 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:38:49 PM
However, instability was observed especially at higher signal-to-noise ratios (SNR).

Number: 55 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:39:10 PM
hierarchical

Number: 56 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:39:23 PM
latter

cessfully applied to signal segmentation [27]–[29] and semi-supervised unmixing of hyperspectral imagery [30].

A In this paper, the response of the MRFM imaging device is assumed to be known. ~~This standard assumption makes the sparse image reconstruction a non-blind deconvolution problem that is a standard linear inverse problem [31].~~ The hierarchical Bayesian formulation proposed here ~~naturally introduces an appropriate regularization for the ill-posed problem where the hyperparameters are estimated in an unsupervised scheme.~~ B Only a few works in the literature are dedicated to reconstruction of MRFM image data [32]–[35]. In [36], several techniques based on linear filtering or maximum-likelihood principle have been proposed. ~~Nevertheless, none of these models and algorithms takes advantage of the sparse nature of the image to be analyzed.~~ More recently, Ting *et al.* has introduced sparsity penalized reconstruction methods ~~motivated by~~ MRFM ~~applications~~ [23]. The reconstruction problem is ~~decomposed~~ into a deconvolution step and a denoising step, yielding an iterative thresholding framework. ~~However, in~~ [23], the hyperparameters are estimated ~~via a heuristic manner by applying~~ the Stein's unbiased risk estimator [37], ~~contrary to our fully Bayesian approach that allows them to be marginalized. As it has been pointed out above, this *ad hoc* hyperparameter choice can lead to unreliable results. Moreover, a posterior analysis is not possible with the strategy proposed in [23]. As a consequence, the parameter estimation is only based on the peak of the penalized likelihood function, whose research thanks to the EM algorithm can be subjected to slow convergence and local maxima [38].~~ B'

A' 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. ~~The simulation~~ results, including ~~comparison of performances,~~ are presented in Section V for MRFM. 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 $n = 3$). ~~This image is available as~~ a collection of P projections $\mathbf{y} = [y_1, \ldots, y_P]^T$ which follow~~s~~ 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. ~~Typical point spread responses $\boldsymbol{\kappa}$ of MRFM tip can be found in [39] for horizontal and vertical configurations.~~ In (1), $\mathbf{n}$ is an additive Gaussian noise sequence distributed according to $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_P)$.

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 $\mathbf{x}$ under sparsity and positivity constraints on $\mathbf{x}$ given the observations $\mathbf{y}$, the psf $\boldsymbol{\kappa}$ and the bilinear function[2] $T(\cdot,\cdot)$.

III. Hierarchical Bayesian model

A. Likelihood function

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

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

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 $a > 0$:

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

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

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

Choosing $g_a(\cdot)$ 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 $\mathbf{x} = [x_1, \ldots, x_M]^T$:

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

where $\{\mathbf{x} \succ 0\} = \{\mathbf{x} \in \mathbb{R}^M; x_i > 0, \forall i = 1, \ldots, M\}$ 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 [41] and signal representation [42].

Coupling a standard probability density function (pdf) with an atom at zero is another ~~classical~~ alternative to ensure sparsity. This strategy has for instance been used for located event detection [14] such as spike train deconvolution [11], [17]. In order to increase the sparsity of the prior, we propose

[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 $\mathbf{x}$.

[3] Note that a similar estimator using a Laplacian prior for x_i ($i = 1, \ldots, M$) was proposed in [40] for regression problems and is usually referred to as the LASSO estimator but without positivity constraint.

Page: 2

Number: 1 Author: vmuser Subject: Pencil Date: 9/17/2008 12:50:10 PM

Number: 2 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:40:38 PM

Number: 3 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:49:27 PM
While it may be possible to extend our methods to unknown point spread functions, e.g., along the lines of [54], the case of sparse blind deconvolution is outside of the scope of this paper.

Number: 4 Author: vmuser Subject: Pencil Date: 9/17/2008 12:50:41 PM

Number: 5 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:49:58 PM
asymptotically generates Bayes-optimal estimates of all image parameters, including the hyperparameters.

Number: 6 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:51:06 PM

Number: 7 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:51:11 PM
have been

Number: 8 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:51:14 PM
and

Number: 9 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:51:43 PM
s

Number: 10 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:51:54 PM
that do not exploit image sparsity.

Number: 11 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:51:56 PM

Number: 12 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:52:03 PM
for

Number: 13 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:52:05 PM

Number: 14 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:52:23 PM

Number: 15 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:52:30 PM
formulated as a decomposition

Number: 16 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:52:36 PM

Number: 17 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:52:35 PM
In

Number: 18 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:54:31 PM
using penalized log-likelihood approaches including

Number: 19 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:53:36 PM
(SURE)

Number: 20 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:52:59 PM
approach.

Number: 21 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:53:09 PM

Number: 22 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:56:18 PM

Number: 23 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:56:30 PM
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 [38]. In contrast to [23], the fully Bayesian approach presented in this paper converges quickly and produces estimates of the entire posterior and not just local maxima.

Number: 24 Author: vmuser Subject: Pencil Date: 9/17/2008 12:50:30 PM

Number: 25 Author: vmuser Subject: Pencil Date: 9/17/2008 12:50:16 PM

Number: 26 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:56:36 PM
is

Number: 27 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:56:42 PM

Number: 28 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:56:43 PM
S

Comments from page 2 continued on next page

cessfully applied to signal segmentation [27]–[29] and semi-supervised unmixing of hyperspectral imagery [30].

A

In this paper, the response of the MRFM imaging device is assumed to be known~~. This standard assumption makes the sparse image reconstruction a non-blind deconvolution problem that is a standard linear inverse problem [31]~~. The hierarchical Bayesian formulation proposed here ~~naturally introduces an appropriate regularization for the ill-posed problem where the hyperparameters are estimated in an unsupervised scheme.~~ Only a few works in the literature ~~are~~ dedicated to reconstruction of MRFM image data [32]–[35]. In [36], several techniques based on linear filtering ~~or~~ maximum-likelihood principle have been proposed. ~~Nevertheless, none of these models and algorithms takes advantage of the sparse nature of the image to be analyzed.~~ More recently, Ting *et al.* has introduced sparsity penalized reconstruction methods ~~motivated by~~ MRFM ~~applications~~ [23]. The reconstruction problem is ~~decomposed~~ into a deconvolution step and a denoising step, yielding an iterative thresholding framework. ~~However, in~~ [23]~~,~~ the hyperparameters are estimated ~~via a heuristic manner by applying~~ the Stein's unbiased risk estimator [37]~~, contrary to our fully Bayesian approach that allows them to be marginalized. As it has been pointed out above, this *ad hoc* hyperparameter choice can lead to unreliable results. Moreover, a full posterior analysis is not possible with the strategy proposed in [23]. As a consequence, the parameter estimation is only based on the peak of the penalized likelihood function, whose research thanks to the EM algorithm can be subjected to slow convergence and local maxima [38]~~. B'

B

A'

This paper is organized as follows. The deconvolution problem is formulated in Section II. The hierarchical Bayesian model ~~are~~ described in Section III. Section IV presents a Gibbs sampler that allows one to generate samples distributed according to the posterior of interest. ~~Some~~ simulation results, including ~~comparison of performances,~~ are presented in Section V for MRFM. 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 $n = 3$). ~~This image is available as~~ collection of P projections $\mathbf{y} = [y_1, \ldots, y_P]^T$ which follows 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. ~~Typical point spread responses $\boldsymbol{\kappa}$ of MRFM tip can be found in [39] for horizontal and vertical configurations.~~ In (1), $\mathbf{n}$ is an additive Gaussian noise sequence distributed according to $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_P)$.

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:

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

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 $\mathbf{x}$ under sparsity and positivity constraints on $\mathbf{x}$ given the observations $\mathbf{y}$, the psf $\boldsymbol{\kappa}$ and the bilinear function[2] $T(\cdot, \cdot)$.

III. Hierarchical Bayesian Model

A. Likelihood function

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

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

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 $a > 0$:

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

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

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

Choosing $g_a(\cdot)$ 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 $\mathbf{x} = [x_1, \ldots, x_M]^T$:

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

where $\{\mathbf{x} \succ 0\} = \{\mathbf{x} \in \mathbb{R}^M; x_i > 0, \forall i = 1, \ldots, M\}$ 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 [41] and signal representation [42].

Coupling a standard probability density function (pdf) with an atom at zero is another ~~classical~~ alternative to ensure sparsity. This strategy has for instance been used for located event detection [14] such as spike train deconvolution [11], [17]. In order to increase the sparsity of the prior, we propose

[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 $\mathbf{x}$.

[3]Note that a similar estimator using a Laplacian prior for x_i $(i = 1, \ldots, M)$ was proposed in [40] for regression problems and is usually referred to as the LASSO estimator but without positivity constraint.

Number: 29 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:56:54 PM

Number: 30 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:56:51 PM
extensive performance comparison

Number: 31 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:57:19 PM
In Section VI we apply our hierarchical Bayesian method to reconstruction of a tobacco virus from real MRFM data.

Number: 32 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:58:00 PM
Observed are

Number: 33 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:57:52 PM

Number: 34 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:58:06 PM
are asumed to

Number: 35 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:58:07 PM

Number: 36 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:58:13 PM

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 [21] and used, for example, in [22], [23] 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 [43, 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 [28], [30], [44]).

C. Hyperparameter priors

The hyperparameter vector associated with the previous prior distributions is $\boldsymbol{\Phi} = \{a, \gamma, w\}$. 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:

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

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 [45] for a similar choice).

2) Hyperparameter γ: ~~No~~ informative Jeffreys' prior [46], [47] is assumed for hyperparameter γ:

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

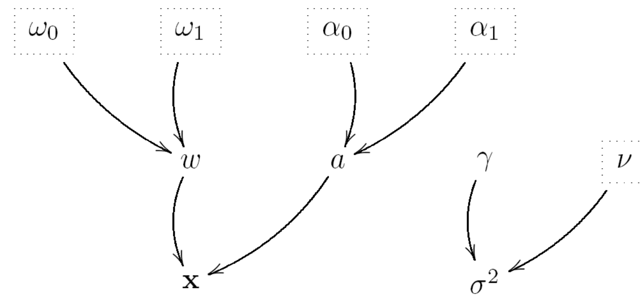

Fig. 1. DAG for the parameter priors and hyperpriors (the fixed hyperparameters appear in dashed boxes).

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

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

with $\boldsymbol{\omega} = [\omega_0, \omega_1]^T$ and where $\mathcal{B}(a,b)$ 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:

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

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

D. Posterior distribution

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

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

with

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

where $f(\mathbf{y}|\boldsymbol{\theta})$ and $f(\boldsymbol{\Phi}|\boldsymbol{\alpha},\boldsymbol{\omega})$ 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:

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

where, as defined in paragraph III-B1, $n_1 = \|\mathbf{x}\|_0$ and $n_0 = M - \|\mathbf{x}\|_0$.

The next section presents an appropriate Gibbs sampling strategy [24] that allows ones to generate samples distributed according to the posterior distribution $f(\mathbf{x}|\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\omega})$.

Number: 1 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:58:20 PM
A non

Number: 2 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:58:24 PM

IV. A GIBBS SAMPLING STRATEGY FOR SPARSE IMAGE RECONSTRUCTION

~~We propose in~~ this section a Gibbs sampling strategy ~~that allows one to generate~~ 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-B1. 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 $f\left(\mathbf{x}\left|w,a,\sigma^2,\mathbf{y}\right.\right)$: 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 A yield:

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

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 A. 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 $\left(w_i,\mu_i,\eta_i^2\right)$. Appendix C presents an algorithm that can be used to generate samples ~~distributed according to~~ this distribution.

To summarize, generating samples distributed according to $f\left(\mathbf{x}\left|w,\sigma^2,a,,\mathbf{y}\right.\right)$ 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 $f\left(\sigma^2\left|\mathbf{x},\mathbf{y}\right.\right)$

Samples are generated as the following way:

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

V. SIMULATION ON SYNTHETIC IMAGES

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

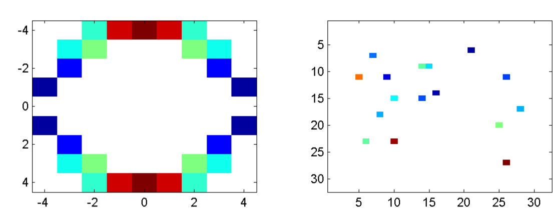

Fig. 2. Left: Psf of the MRFM tip. Right: unknown sparse image to be estimated.

Page: 4

Number: 1 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:58:34 PM
In

Number: 2 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:58:36 PM
we describe the

Number: 3 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:58:41 PM
for

Number: 4 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:58:44 PM
generating

Number: 5 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:58:54 PM
from

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 MRFM tip has been given in [39] and is explained in [23]. 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 $\mathrm{SNR} = 2\mathrm{dB}$ and $\mathrm{SNR} = 20\mathrm{dB}$ respectively.

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

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

These estimates are depicted in Fig. 3 for the two levels of noise considered. ~~It can be noticed~~ that the estimated image is very similar to the actual image, even ~~with a~~ low SNR.

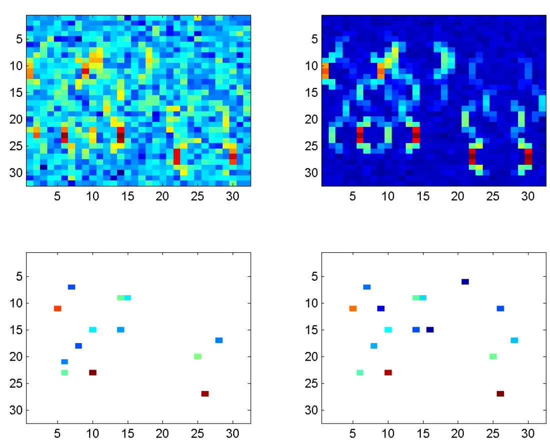

Fig. 3. Top, left (resp. right): noisy observations for SNR = 2dB (resp. 20dB). Bottom, left (resp. right): reconstructed image for SNR = 2dB (resp. 20dB).

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 ~~actual~~ values of these parameters ~~for the two SNR levels.~~

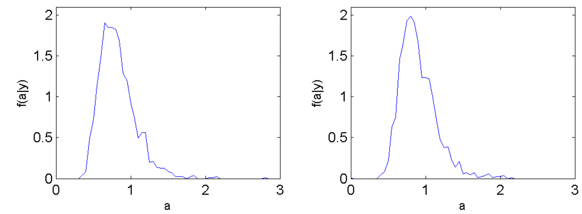

Fig. 4. Posterior distribution of hyperparameter a (left: SNR = 2dB, right: SNR = 20dB).

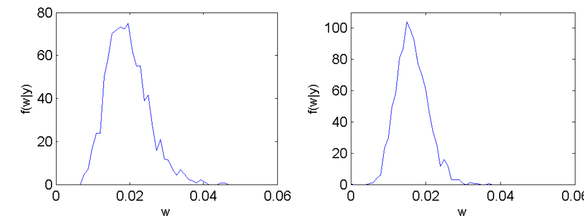

Fig. 5. Posterior distribution of hyperparameter w (left: SNR = 2dB, right: SNR = 20dB).

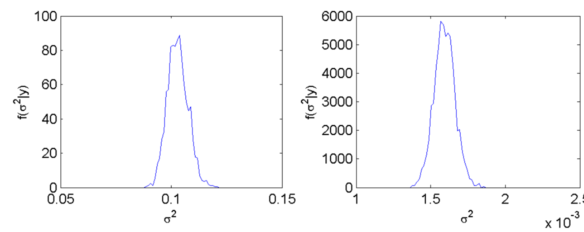

Fig. 6. Posterior distribution of hyperparameter σ^2 (left: SNR = 2dB, right: SNR = 20dB).

The posterior distributions of four different pixels are depicted in Fig. 7. These posteriors are ~~also in agreement~~ with the actual values of these pixels that are represented ~~in~~ dotted red line in these figures.

2) Comparison of reconstruction performances: ~~The results provided by the proposed method have been compared with those provided by methods that also estimate the hyperparameters automatically.~~ ~~Firstly, the~~ techniques proposed in [22], [23] are based on 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, ~~are studied. Here we compare the estimators of [22], [23] to the MMSE estimator and the MAP estimator under the model and the algorithm presented in Sections III and IV. The~~ MMSE estimator of the unknown parameter $\mathbf{x}$ is obtained by an empirical averaging over the last $N_r = 1700$ outputs of the sampler according to:

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

Page: 5

Number: 1 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:59:00 PM
retaining

Number: 2 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:59:11 PM
Observe

Number: 3 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:59:14 PM
at

Number: 4 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:59:44 PM
consistent

Number: 5 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:59:47 PM
as

Number: 6 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:59:48 PM
s

Number: 7 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:00:42 PM
Here we compare our proposed hierarchical Bayesian method to the methods of [22],[23].

Number: 8 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:00:17 PM
The

Number: 9 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:00:24 PM
penalized likelihood

Number: 10 Author: vmuser Subject: Pencil Date: 9/17/2008 1:03:10 PM

Number: 11 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:01:10 PM
were proposed.

Number: 12 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:02:23 PM
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

Number: 13 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:02:25 PM

Number: 14 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:02:28 PM
Gibbs

Number: 15 Author: vmuser Subject: Replacement Text Date: 9/17/2008 12:59:21 PM
ground truth

Number: 16 Author: vmuser Subject: Cross-Out Date: 9/17/2008 12:59:31 PM

Number: 17 Author: vmuser Subject: Inserted Text Date: 9/17/2008 12:59:30 PM
; randomly drawn from the prior distribution

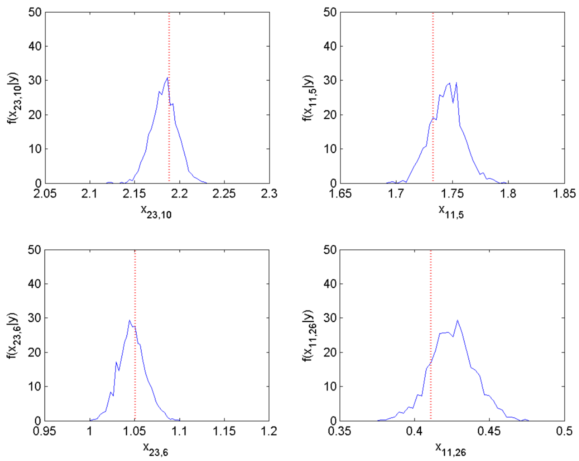

Fig. 7. Posteriors distributions of the non-zero values of $\mathbf{x}$ for SNR $= 20$dB, (actual values are depicted with dotted red lines).

~~Finally, the results are compared~~ with the estimator provided by a standard Landweber algorithm [48]. ~~The proposed comparison is conducted~~ with respect to several ~~measures of~~ performance. ~~First 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. ~~To measure the performance of the sparse reconstruction, criteria inspired by [23] have been used:~~ the ℓ_0, ℓ_1 and ℓ_2-norms of $\mathbf{e}$ to measure the accuracy of the reconstruction and the ℓ_0-norm of the estimator $\hat{\mathbf{x}}$ to measure its sparsity. ~~Moreover, as noticed~~ in [23], small non-zero values of the pixel are usually not distinguishable from exactly zero values by a human being. ~~Following this remark~~, a less strict measure of sparsity than the ℓ_0-norm ~~has been introduced. This measure~~[4], denoted $\|\cdot\|_\delta$, ~~is~~ the number of ~~components~~ that are less than a given threshold δ:

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

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-A1 for two levels of SNR: $\|e\|_0$, $\|e\|_\delta$, $\|e\|_1$, $\|e\|_2$, $\|\hat{\mathbf{x}}\|_0$ and $\|\hat{\mathbf{x}}\|_\delta$.

Table II ~~gathers~~ the six performance measures for the five different ~~studied~~ algorithms. ~~It clearly appears that the~~ proposed Bayesian method outperform~~s~~ the other~~s~~ in ~~the~~ ℓ_1 or ℓ_2-norm ~~evaluations of the error reconstruction, whatever the estimator chosen (MAP or MMSE). This can be easily explained by the accurate estimation of the hyperparameters thanks to the introduced hierarchical model. It also appears that an~~ MMSE estimation of the unknown image ~~yields to a~~ non sparse estimator in ~~a~~ ℓ_0-norm sense. This ~~can be explained by a~~ very ~~weak~~ posterior probability of ~~having~~ non-zero value ~~for each~~ pixel. ~~A less draconian decision, by using the~~ sparsity measure $\|\cdot\|_\delta$ ~~for instance, can overcome this drawback, as shown in the Table. Finally, the~~ MAP estimator ~~seems to be a 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, ~~it has to be noticed that~~ MMSE estimation ~~contains more information than a point estimation and can be useful to derive confidence intervals~~.

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

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 $= 2$dB						
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 $= 20$dB						
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

B. Reconstruction of undersampled 3-dimensional images

In this subsection, ~~some simulation results are presented to illustrated the performance of the algorithm when applied on undersampled 3D images.~~ 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. 8 (right) and Fig. 10 (top). This image to be recovered is assumed to be convolved with a $5\times5\times3$ kernel that is represented in Fig. 8 (right). The resulting convolved image is depicted in Fig. 9 (left). However, the actually observed image is ~~assumed to be 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 observed 3D image, shown in Fig. 9, 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.

~~The proposed~~ Bayesian algorithm is used to perform the sparse reconstruction. The number of Monte Carlo runs ~~has been~~ fixed to $N_{\mathrm{MC}} = 2000$ with $N_{\mathrm{bi}} = 200$ burn-in iterations. ~~The MAP estimator has been chosen as the reconstructed image estimate since it outperforms the MMSE ones (as explained in paragraph V-A2). This MAP estimator, depicted in Fig. 10 (bottom), is quite satisfactory given the problem difficulty introduced by the undersampling.~~

Page: 6

Number: 1 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:06:34 PM

Number: 2 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:06:25 PM
at many

Number: 3 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:06:25 PM
s

Number: 4 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:06:38 PM
The

Number: 5 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:06:42 PM

Number: 6 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:06:53 PM
indicates that most of the pixels are in fact very close to zero.

Number: 7 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:07:06 PM

Number: 8 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:07:02 PM
The

Number: 9 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:07:28 PM
by construction the

Number: 10 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:07:42 PM
will always have lower mean square error

Number: 11 Author: vmuser Subject: Pencil Date: 9/17/2008 1:02:56 PM

Number: 12 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:02:50 PM
We also compare

Number: 13 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:03:24 PM

Number: 14 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:03:33 PM
As in [23] we compare estimators

Number: 15 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:03:24 PM

Number: 16 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:03:34 PM

Number: 17 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:03:40 PM

Number: 18 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:03:38 PM
criteria

Number: 19 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:03:42 PM
Let

Number: 20 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:03:53 PM
These criteria are:

Number: 21 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:03:56 PM

Number: 22 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:03:55 PM
, which

Number: 23 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:04:06 PM
s

Number: 24 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:04:01 PM

Number: 25 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:04:00 PM

Number: 26 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:04:04 PM
, which

Number: 27 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:03:57 PM
,

Number: 28 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:04:04 PM
s

Comments from page 6 continued on next page

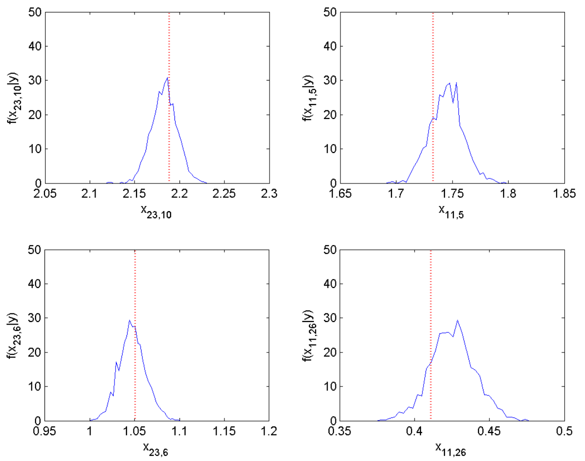

Fig. 7. Posteriors distributions of the non-zero values of $\mathbf{x}$ for SNR $= 20$dB, (actual values are depicted with dotted red lines).

~~Finally, the results are compared~~ with the estimator provided by a standard Landweber algorithm [48].

~~The proposed comparison is conducted~~ with respect to several ~~measures of~~ performance. ~~First 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. ~~To measure the performance of the sparse reconstruction, criteria inspired by [23] have been used:~~ the ℓ_0, ℓ_1 and ℓ_2-norms of $\mathbf{e}$ ~~to~~ measure the accuracy of the reconstruction and the ℓ_0-norm of the estimator $\hat{\mathbf{x}}$ ~~to~~ measure its sparsity. ~~Moreover, as noticed~~ in [23], small non-zero values of the pixel are usually not distinguishable from exactly zero values by a human being. ~~Following this remark~~, a less strict measure of sparsity than the ℓ_0-norm ~~has been introduced. This measure~~[4], denoted $\|\cdot\|_\delta$, is the number of ~~components~~ 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-A1 for two levels of SNR: $\|e\|_0$, $\|e\|_\delta$, $\|e\|_1$, $\|e\|_2$, $\|\hat{\mathbf{x}}\|_0$ and $\|\hat{\mathbf{x}}\|_\delta$.

Table II ~~gathers~~ the six performance measures for the five different ~~studied~~ algorithms. ~~It clearly appears that the~~ proposed Bayesian method outperforms the others ~~in~~ terms of ℓ_1 or ℓ_2-norm ~~evaluations of the error reconstruction, whatever the estimator chosen (MAP or MMSE). This can be easily explained by the accurate estimation of the hyperparameters thanks to the introduced hierarchical model. It also appears that an~~ MMSE estimation of the unknown image ~~yields to a~~ non sparse estimator in ~~a~~ ℓ_0-norm sense. This ~~can be explained by a~~ very ~~weak~~ posterior probability of ~~having~~ non-zero value ~~for each~~ pixel. ~~A less draconian decision, by using the~~ sparsity measure $\|\cdot\|_\delta$ ~~for instance, can overcome this drawback, as shown in the Table~~. ~~Finally, the~~ MAP estimator ~~seems to be a 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, ~~it has to be noticed that~~ MMSE estimation ~~contains more information than a point estimation and can be useful to derive confidence intervals~~.

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

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

B. Reconstruction of undersampled 3-dimensional images

~~In~~ this subsection, ~~the simulation results are presented to illustrated the performance of the algorithm when applied on undersampled 3D images.~~ 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. 8 (right) and Fig. 10 (top). This image to be recovered is assumed to be convolved with a $5 \times 5 \times 3$ kernel that is represented in Fig. 8 (right). The resulting convolved image is depicted in Fig. 9 (left). However, the actually observed image is ~~assumed to be 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 observed 3D image, shown in Fig. 9, 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. ~~The proposed~~ Bayesian algorithm is used to perform the sparse reconstruction. ~~The~~ number of Monte Carlo runs ~~has been~~ fixed to $N_{\mathrm{MC}} = 2000$ with $N_{\mathrm{bi}} = 200$ burn-in iterations. ~~The MAP estimator has been chosen as the reconstructed image estimate since it outperforms the MMSE ones (as explained in paragraph V-A2). This MAP estimator, depicted in Fig. 10 (bottom), is quite satisfactory given the problem difficulty introduced by the undersampling.~~

Number: 29 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:04:14 PM
As pointed out

Number: 30 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:04:10 PM
Thus

Number: 31 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:08:43 PM
As iscussed in Sec. VI, the prototype MRFM instrument collects data projections as irregularly spaced, or undersampled, spatial samples.

Number: 32 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:09:11 PM
we indicate how the image reconstruction algorithm can be adapted to this undersampled scenario in 3D.

Number: 33 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:04:25 PM

Number: 34 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:09:13 PM

Number: 35 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:04:26 PM
,

Number: 36 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:09:24 PM
For concreteness, we illustrate by a concrete example.

Number: 37 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:09:13 PM

Number: 38 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:04:33 PM
which is

Number: 39 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:04:43 PM
reconstructed image pixels

Number: 40 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:09:28 PM

Number: 41 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:09:31 PM
an

Number: 42 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:04:49 PM
shows

Number: 43 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:04:51 PM

Number: 44 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:04:59 PM
studied

Number: 45 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:05:03 PM
The

Number: 46 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:05:15 PM

Number: 47 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:05:14 PM
s (labeled "proposed MMSE" and "proposed MAP" in the table)

Number: 48 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:05:19 PM
estimators

Number: 49 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:05:28 PM
terms of

Number: 50 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:05:31 PM

Number: 51 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:05:35 PM
s

Number: 52 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:09:50 PM

Number: 53 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:05:50 PM
Note that the

Number: 54 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:10:24 PM

Number: 55 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:10:24 PM

Number: 56 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:10:30 PM
For illustration the proposed

Number: 57 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:09:43 PM
hierarchical

Number: 58 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:10:37 PM
with undersampled data

Comments from page 6 continued on next page

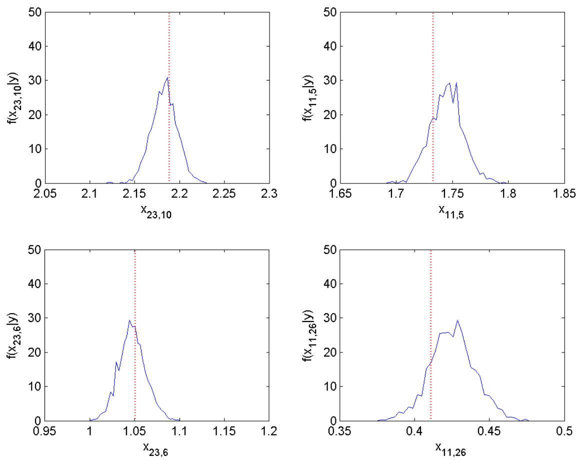

Fig. 7. Posteriors distributions of the non-zero values of $\mathbf{x}$ for SNR = 20dB, (actual values are depicted with dotted red lines).

~~Finally, the results are compared~~ with the estimator provided by a standard Landweber algorithm [48].

~~The proposed comparison is conducted~~ with respect to several ~~measures of~~ performance. ~~First 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. ~~To measure the performance of the sparse reconstruction, criteria inspired by [23] have been used:~~ the ℓ_0, ℓ_1 and ℓ_2-norms of $\mathbf{e}$ ~~to~~ measure the accuracy of the reconstruction and the ℓ_0-norm of the estimator $\hat{\mathbf{x}}$ ~~to~~ measure its sparsity. ~~Moreover, as noticed~~ in [23], small non-zero values of the pixel are usually not distinguishable from exactly zero values by a human being. ~~Following this remark~~, a less strict measure of sparsity than the ℓ_0-norm ~~has been introduced. This measure~~[4], denoted $\|\cdot\|_\delta$, ~~is~~ the number of ~~components~~ 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-A1 for two levels of SNR: $\|e\|_0$, $\|e\|_\delta$, $\|e\|_1$, $\|e\|_2$, $\|\hat{\mathbf{x}}\|_0$ and $\|\hat{\mathbf{x}}\|_\delta$.

Table II ~~gathers~~ the six performance measures for the five different ~~studied~~ algorithms. ~~It clearly appears that the~~ proposed Bayesian method outperforms the others in ~~the~~ ℓ_1 or ℓ_2-norm ~~evaluations of the error reconstruction, whatever the estimator chosen (MAP or MMSE). This can be easily explained by the accurate estimation of the hyperparameters thanks to the introduced hierarchical model. It also appears that an~~ MMSE estimation of the unknown image ~~yields to a~~ non sparse estimator in the ℓ_0-norm sense. This ~~can be explained by a~~ very weak posterior probability of ~~having~~ non-zero value ~~for each~~ pixel. ~~A less draconian decision, by using the~~ sparsity measure $\|\cdot\|_\delta$ ~~for instance, can overcome this drawback, as shown in the Table~~. ~~Finally, the~~ MAP estimator ~~seems to be a 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, ~~it has to be noticed that~~ MMSE estimation ~~contains more information than a point estimation and can be useful to derive confidence intervals~~.

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

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

B. Reconstruction of undersampled 3-dimensional images

In this subsection, ~~some simulation results are presented to illustrated the performance of the algorithm when applied on undersampled 3D images.~~ 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. 8 (right) and Fig. 10 (top). This image to be recovered is assumed to be convolved with a $5 \times 5 \times 3$ kernel that is represented in Fig. 8 (right). The resulting convolved image is depicted in Fig. 9 (left). However, the actually observed image is ~~assumed to be 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 observed 3D image, shown in Fig. 9, 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.

~~The proposed~~ Bayesian algorithm is used to perform the sparse reconstruction. The number of Monte Carlo runs ~~has been~~ fixed to $N_{\rm MC} = 2000$ with $N_{\rm bi} = 200$ burn-in iterations. ~~The MAP estimator has been chosen as the reconstructed image estimate since it outperforms the MMSE ones (as explained in paragraph V-A2). This MAP estimator, depicted in Fig. 10 (bottom), is quite satisfactory given the problem difficulty introduced by the undersampling.~~

Number: 59 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:10:40 PM
was

Number: 60 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:05:44 PM

Number: 61 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:05:55 PM
is a

Number: 62 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:05:55 PM

Number: 63 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:05:55 PM

Number: 64 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:11:18 PM
Figure 10 shows the result of applying the proposed MAP estimator to the estimated posterior.

Number: 65 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:05:57 PM
the

Number: 66 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:06:05 PM
is due to the

Number: 67 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:06:14 PM

Number: 68 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:06:11 PM
small

Number: 69 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:06:08 PM
but non-zero

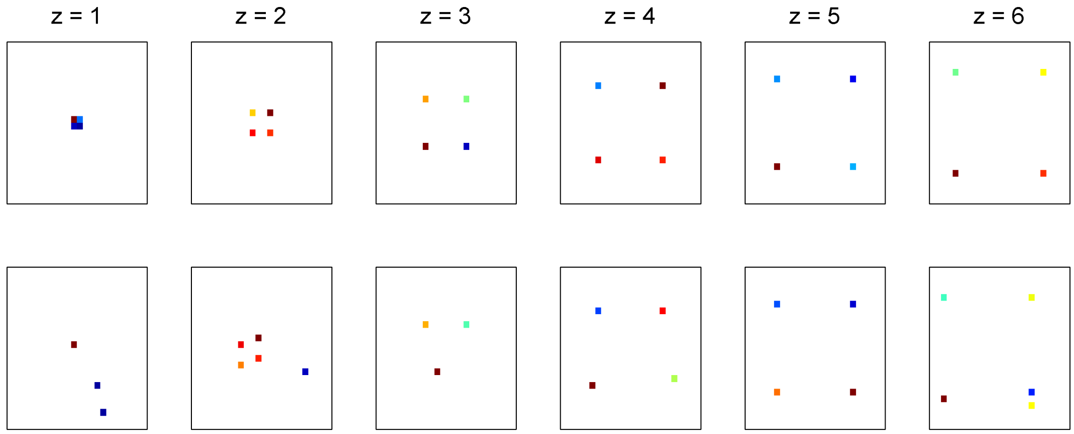

Fig. 10. Top: slices of the sparse image to be recovered. Bottom: slices of the estimated sparse image.

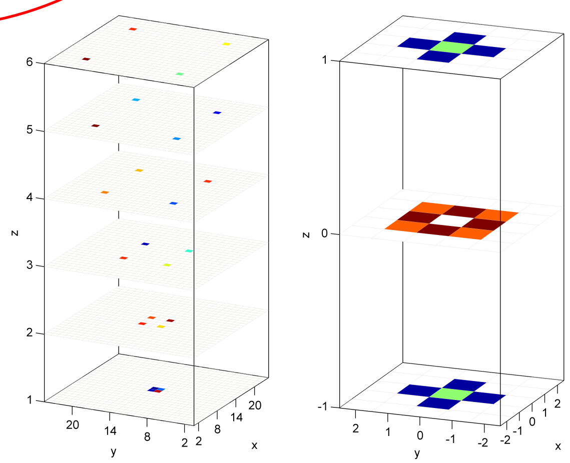

Fig. 8. Left: $24 \times 24 \times 6$ unknown image to be recovered. Right: $5 \times 5 \times 3$ kernel modeling the psf.

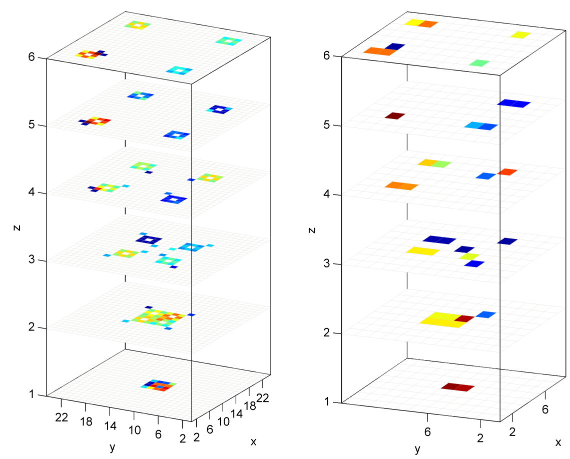

Fig. 9. Left: $24 \times 24 \times 6$ regularly sampled convolved image. Left: $12 \times 8 \times 6$ undersampled observed image.

VI. Application on real MRFM images

~~The 3-dimensional real data used in this section have been initially presented in [35] to illustrate the nanometer spatial resolution of MRFM.~~ The observed sample consists of a collection of Tobacco mosaic virus particles that are divided into a whole segment ~~and others~~ fragments. The ~~map of the signal is~~ computed from the measured proton distribution and the 3-dimensional psf following the protocol described in [35] and [49]. 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.

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 $d_x = 3$, $d_y = 2$ and $d_z = 3$ in the three directions. The MAP estimate of the unknown image is computed ~~after~~ $N_{\mathrm{MC}} = 1000$ (with $N_{\mathrm{bi}} = 200$) of the proposed Bayesian algorithm initialized with the output of the Landweber iteration. Three horizontal slices of the estimated image[5] ~~is~~ depicted in Figure 12. A 3-dimensional view of the estimated profile of the virus fragments is ~~also available~~ in Figure 13. The MMSE estimates of the parameters introduced in Section III are $\hat{\sigma}^2_{\mathrm{MMSE}} = 0.10$, $\hat{a}_{\mathrm{MMSE}} = 1.9 \times 10^{-12}$ and $\hat{w}_{\mathrm{MMSE}} = 1.4 \times 10^{-2}$.

~~To illustrate the performance of the proposed deconvolution algorithm, the data reconstructed from the estimated 3-dimensional image are depicted in Figure 11 (bottom). These figures are clearly in good agreement with the observed data (top). Moreover, 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. ~~It clearly appears~~ that the convergence rate of our

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

Page: 7

Number: 1 Author: vmuser Subject: Pencil Date: 9/17/2008 1:14:57 PM

Number: 2 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:15:12 PM
This figure is out of order!

Number: 3 Author: vmuser Subject: Typewritten Text Date: 9/17/2008 1:15:03 PM

Number: 4 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:13:31 PM
from

Number: 5 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:13:34 PM
Gibbs samples

Number: 6 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:13:38 PM
. The algorithm was

Number: 7 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:13:41 PM
a single

Number: 8 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:13:02 PM
Here we illustrate the heirarchical Bayes MAP reconstruction algorithm for real lthree 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 [35].

Number: 9 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:14:44 PM
Several more iterations of the Landweber algorithm would produce the reconstructions reported in [35]. Theimage reconstructions produced by the Landweber and Bayesian MAP algorithm are shown in Figs. 11-13.

Number: 10 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:15:31 PM
are

Number: 11 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:15:35 PM
shown

Number: 12 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:12:49 PM
viral

Number: 13 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:12:47 PM
in addiiton to viral

Number: 14 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:13:09 PM
projections are

Number: 15 Author: vmuser Subject: Sticky Note Date: 9/17/2008 1:17:29 PM
Nicolas - pourras-tu mettre une troisieme serie d'images pour comparer Bayes contre Landweber dans Fig 11?

Number: 16 Author: vmuser Subject: Pencil Date: 9/17/2008 1:17:09 PM

Number: 17 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:16:38 PM
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.

Number: 18 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:16:58 PM

Number: 19 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:16:57 PM

Number: 20 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:13:21 PM
pixels

Number: 21 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:17:01 PM
To

Number: 22 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:17:48 PM
This shows

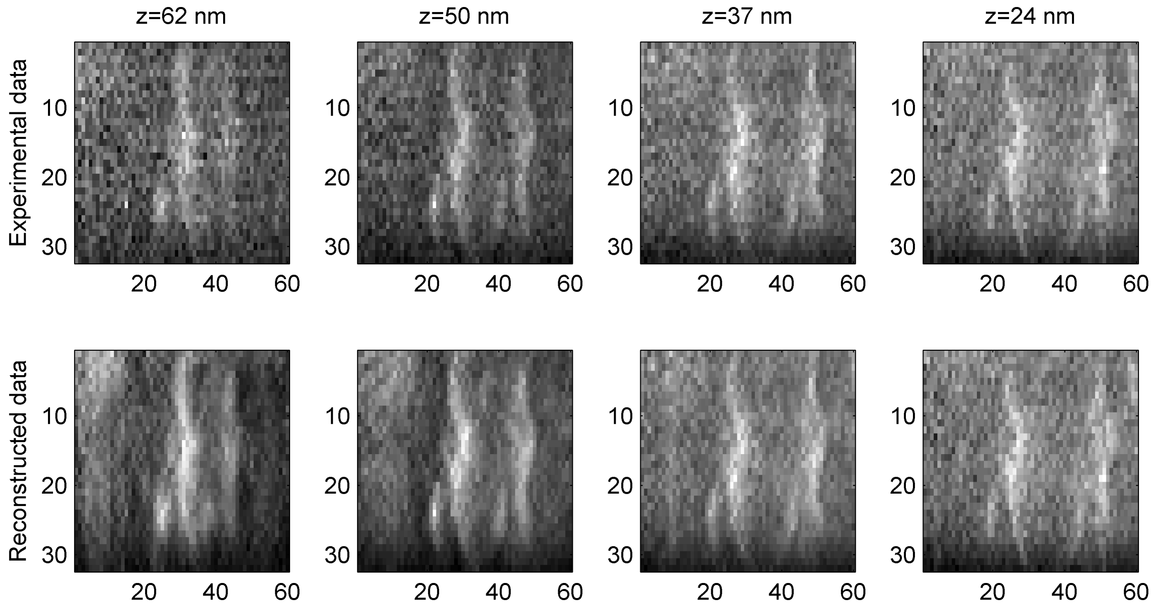

Fig. 11. Top: experimental scan data. Bottom: scan data computed from the proposed Bayesian reconstruction.

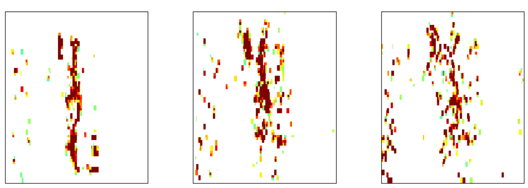

Fig. 12. Three horizontal slices of the estimated image.

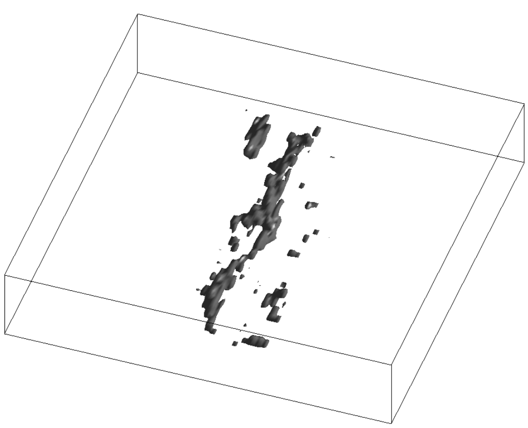

Fig. 13. 3-dimensional view of the estimated profile of the Tobacco virus fragments.

algorithm is significantly better than the Landweber algorithm.

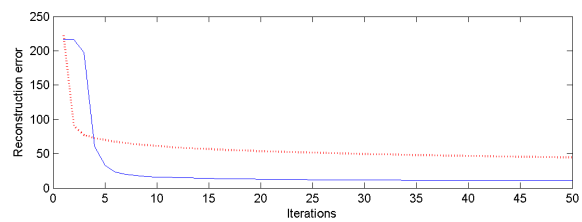

Fig. 14. Error reconstructions as functions of the iteration number for the proposed algorithm (continuous blue line) and Landweber algorithm (dotted red line).

VII. Conclusions

This paper presented a Bayesian ~~sampling~~ algorithm for ~~solving~~ deconv~~olution of~~ sparse images corrupted by additive Gaussian noise. A Bernoulli-truncated exponential distribution was proposed as prior distribution for the sparse image to be recovered. The hyperparameters of the model were ~~estimated in a fully Bayesian scheme by choosing prior distributions for them and by integrating them out from the full posterior distribution~~. An efficient Gibbs sampler ~~allowed one to~~ generate samples ~~distributed~~ according to this posterior distribution. The derived Bayesian estimators ~~outperformed significantly the estimators classically used to solve~~ sparse reconstruction ~~problems. This was mainly due to a performing estimation of the hyperparameters via the proposed hierarchical model.~~

Appendix A
Derivation of the conditional

Page: 8

Number: 1 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:18:00 PM

Number: 2 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:17:58 PM
hierarchical

Number: 3 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:18:02 PM

Number: 4 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:18:04 PM
ing

Number: 5 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:18:10 PM
positive

Number: 6 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:18:26 PM
unknown

Number: 7 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:19:22 PM
integrated out of 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).

Number: 8 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:19:25 PM

Number: 9 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:19:29 PM
was used to

Number: 10 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:19:31 PM

Number: 11 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:19:37 PM
significantly outperformed

Number: 12 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:19:43 PM
several previously proposed

Number: 13 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:19:50 PM
algorithms

Number: 14 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:21:16 PM
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.

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

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' ~~paradigm~~:

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

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 B. Then, straightforward computations similar to those in [11] and [50, 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

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

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.

[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).

Appendix B
Fast recursive computations for simulating according to $f\left(\mathbf{x} \left| w, a, \sigma^2, \mathbf{y}\right.\right)$

In the Gibbs sampling strategy presented in Section IV, the main computationally expensive task is the generation of samples distributed according to $f\left(x_i \left| w, a, \sigma^2, \mathbf{x}_{-i}, \mathbf{y}\right.\right)$. Indeed, the evaluation of the hidden mean and hidden variance in (36) of the Bernoulli-truncated Gaussian distribution may be ~~really~~ 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 ~~make this~~ Gibbs ~~move faster.~~ ~~More precisely, we describe how to update~~ efficiently the coordinate i of the vector $\mathbf{x}$ at iteration t of the Gibbs sampler.

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 $\left\{\mathbf{h}_i\right\}_{i=1,\ldots,M}$, 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 ~~resorting~~ to the bilinear ~~application~~ $T\left(\cdot, \cdot\right)$.

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 $T\left(\boldsymbol{\kappa}, \cdot\right)$, 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 $T\left(\cdot, \cdot\right)$. 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:

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

Number: 1 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:21:23 PM
formula

Number: 2 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:21:25 PM

Number: 3 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:21:32 PM
accelerate the

Number: 4 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:21:45 PM
sampling by

Number: 5 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:21:47 PM

Number: 6 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:21:51 PM
updating

Number: 7 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:22:00 PM
evaluating

Number: 8 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:22:06 PM
function

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

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

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

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

$\blacksquare$

~~As a conclusion,~~ ~~the~~ bilinear ~~application~~ $T(\cdot,\cdot)$ ~~has~~ only to be used at the very beginning of the 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 ~~detailed~~ in Algorithm 2.

ALGORITHM 2:

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

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 C
SIMULATION ACCORDING TO A BERNOULLI-TRUNCATED GAUSSIAN DISTRIBUTION

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

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

where $C(m, s^2)$ has been defined in (23). ~~This generation can be conducted~~ by using an auxiliary binary variable ε following the strategy ~~detailed~~ in Algorithm 3.

In ~~the algorithm presented above~~, $\mathcal{B}er(\cdot)$ and $\mathcal{N}^+(\cdot,\cdot)$ denote the Bernoulli and the positive truncated Gaussian distributions respectively. In step 2, ~~generation of~~ samples distributed according to the truncated Gaussian distribution can be ~~achieved~~ by using an appropriate accept-reject procedure with ~~different~~ instrumental distributions [51]–[53].

ALGORITHM 3:

Simulation according to a Bernoulli-truncated Gaussian distribution

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

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. Dan Rugar who provided the real data used in Section VIas well as a valuable feedback about this paper.

REFERENCES

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

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

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

[4] 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.

[5] 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.

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

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

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

[17] 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.

[18] 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.

Page: 10

Number: 1 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:22:14 PM

Number: 2 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:22:22 PM

Number: 3 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:22:17 PM
The

Number: 4 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:22:19 PM
function

Number: 5 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:22:25 PM
needs

Number: 6 Author: vmuser Subject: Inserted Text Date: 9/17/2008 1:22:28 PM
Gibbs sampling

Number: 7 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:22:30 PM

Number: 8 Author: vmuser Subject: Pencil Date: 9/17/2008 1:23:40 PM

Number: 9 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:22:39 PM
shown

Number: 10 Author: vmuser Subject: Sticky Note Date: 9/17/2008 1:23:55 PM
Consider making Dan Rugar a co-author (I should be last on the list)

Number: 11 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:22:50 PM
describes how we

Number: 12 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:23:09 PM
Monte Carlo draws from this density can be obtained

Number: 13 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:23:12 PM
shown

Number: 14 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:23:22 PM
Algorithm 3

Number: 15 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:23:25 PM

Number: 16 Author: vmuser Subject: Replacement Text Date: 9/17/2008 1:23:29 PM
generated

Number: 17 Author: vmuser Subject: Cross-Out Date: 9/17/2008 1:23:33 PM

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

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

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Bristol, UK: Institute of Physics Publishing, 1998.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

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

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

[41] 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.

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

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

[44] 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.

[45] 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.

[46] 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.

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

[48] 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.

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

[50] 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.

[51] 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.

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

[53] 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.

Number: 1 Author: vmuser Subject: Sticky Note Date: 9/17/2008 1:25:38 PM

Include other references:

M. Ting, R. Raich and A. O. Hero, "Empirical approaches to sparse image reconstruction," submitted to IEEE Trans. on Image Processing, 2008.

Kyle Herrity, Raviv Raich and Alfred O. Hero, "Blind deconvolution for sparse molecular imaging," IEEE Intl Conf. on Acoustics, Speech and Signal Processing , April 2008.