Variational semi-blind sparse image reconstruction with application to MRFM

Se Un Park, Nicolas Dobigeon, Alfred O. Hero

Variational Semi-blind Sparse Image Reconstruction with Application to MRFM

Se Un Park[a], Nicolas Dobigeon[b] and Alfred O. Hero[a]

[a]University of Michigan, EECS Department, Ann Arbor, USA;
[b]University of Toulouse, IRIT/INP-ENSEEIHT, Toulouse, France

ABSTRACT

This paper addresses the problem of joint image reconstruction and point spread function (PSF) estimation when the PSF of the imaging device is only partially known. To solve this semi-blind deconvolution problem, prior distributions are specified for the PSF and the 3D image. Joint image reconstruction and PSF estimation is then performed within a Bayesian framework, using a variational algorithm to estimate the posterior distribution. The image prior distribution imposes an explicit atomic measure that corresponds to image sparsity. Simulation results demonstrate that the semi-blind deconvolution algorithm compares favorably with previous Markov chain Monte Carlo (MCMC) version of myopic sparse reconstruction. It also outperforms non-myopic algorithms that rely on perfect knowledge of the PSF. The algorithm is illustrated on real data from magnetic resonance force microscopy (MRFM).

Keywords: Variational Bayesian inference, posterior image distribution, image reconstruction, hyperparameter estimation, MRFM experiment.

1. INTRODUCTION

Standard image reconstruction techniques generally assume that the point spread function (PSF) is perfectly known. However, in many situations, the true PSF is either unknown or partially known. In such circumstances, the PSF required in the reconstruction process is mismatched with the true PSF. The quality of standard image reconstruction techniques may suffer from this disparity. To deal with this mismatch, deconvolution methods have been proposed to estimate the unknown image and the PSF jointly. When prior knowledge of the PSF is available, these methods are usually referred to as semi-blind deconvolution[1,2] or myopic deconvolution.[3–5]

We propose to formulate the semi-blind deconvolution task as an estimation problem in a Bayesian setting. Bayesian approaches provide a flexible framework to solve complex model-based problems. Prior information available on the parameters to be estimated can be efficiently included within the model, leading to an implicit regularization of our ill-posed problem. In addition, the Bayes framework produces posterior estimates of uncertainty. Extending our work in Park *et al.*,[5] in this paper we propose a variational estimator for the parameters as contrasted to the Monte Carlo approach in Dobigeon *et al.*[6] and Park *et al.*[5] The extension of the solution of the semi-blind problem[5] to a variational setting is non-trivial. We introduce a novel variational Bayes algorithm that iterates on a hidden variable domain associated with the mixture coefficients. This results in a faster and more scalable algorithm than in Park *et al.*[5]

This paper accounts for uncertainty in the PSF using a Bayesian model, as in Park *et al.*[5] This uncertainty is modeled as the deviation of the a priori known PSF from the true PSF. Applying an eigendecomposition of the PSF covariance, the deviation is represented as a linear combination of orthogonal PSF bases with unknown coefficients that need to be estimated. Furthermore, we assume the desired image is sparse, corresponding to the natural sparsity of molecular or astronomical images. The image prior is a weighted sum of a sparsity inducing part and a continuous distribution, introduced as a positive truncated *Laplacian and atom at zero* (LAZE) prior.[7]

This work was partially supported by ARO, grant number W911NF-05-1-0403.
Further author information: (Send correspondence to A.O.H.)
S.U.P.: E-mail: seunpark@umich.edu
N.D.: E-mail: nicolas.dobigeon@enseeiht.fr
A.O.H.: E-mail: hero@umich.edu

Similar priors have been applied to Gaussian mixtures[8,9] and to model search for mixture distributions.[10] The novelty of our image prior modeling lies in the introduction of a hidden variable that factors the approximate distributions, which consist of discrete mass and a continuous density function.

Bayesian inference of parameters from the posterior distribution typically requires challenging computations, such as maximization and numerical integration. One widely advocated strategy relies on approximations to the minimum mean square error (MMSE) or maximum a posteriori (MAP) estimators using samples drawn from the posterior distribution. Generation of these samples is typically accomplished using Markov chain Monte Carlo methods (MCMC).[11] MCMC has been successfully adopted in numerous imaging problems such as image segmentation, denoising, and deblurring.[11,12] Recently, to solve blind deconvolution, two promising semi-blind MCMC methods have been suggested by Orieux *et al.*[13] and Park *et al.*[5] However, these sampling methods have the disadvantage that estimation is done by storing sufficiently many samples.

An alternative to Monte Carlo integration are variational approximations to the posterior distribution, and this approach is adopted in this paper. These approximations have been extensively exploited to conduct inference with graphical models.[14] They can produce an analytical posterior distribution from which Bayesian estimators can be efficiently computed. Compared to MCMC, variational methods have the advantage of stability and lower computational complexity, since they avoid stochastic simulation. Additionally, variational approaches do not require as much storage space as MCMC needs. However, variational Bayes (VB) approaches are difficult to apply when latent variables have mixture distributions, as they do in our proposed sparse image model. A principal contribution of this paper is the implementation of a VB algorithm for this case. In order to perform variational Bayesian inference, prior distributions are first assigned to the image variables, noise level, and PSF coefficients. Then, the posterior distributions are estimated by minimizing the Kullback-Leibler (KL) distance between the model and the empirical distances. Simulations conducted on synthetic images show that the resulting myopic deconvolution algorithm outperforms previous non-myopic algorithms and competes with the previous MCMC-based semi-blind method.[5]

We illustrate the proposed method on real data from magnetic resonance force microscopy (MRFM) experiments. MRFM is an emerging molecular imaging modality that has the potential for achieving 3D atomic scale resolution.[15–17] Recently, MRFM has successfully demonstrated imaging[18,19] of a tobacco mosaic virus.[20] The 3D image reconstruction problem for MRFM experiments was investigated with Wiener filters,[19,21,22] iterative least square reconstruction approaches,[20,23,24] and, recently, the Bayesian estimation framework.[5–7] The drawback of these approaches is that they require prior knowledge of the PSF. However, in many practical situations of MRFM imaging, the exact PSF, i.e., the response of the MRFM tip, is only partially known. The proposed semi-blind reconstruction method accounts for this partial knowledge.

2. FORMULATION

2.1 Image Model

As in Park *et al.*,[5] the image model is defined as:

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

where $\mathbf{y}$ is a $P \times 1$ vectorized measurement, $\mathbf{x} = [x_1, \ldots, x_N]^T \succeq 0$ is an $N \times 1$ vectorized sparse image to be recovered, $T(\boldsymbol{\kappa}, \cdot)$ is a convolution operator with the PSF $\boldsymbol{\kappa}$, $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]$ is an equivalent system matrix, and $\mathbf{n}$ is the measurement noise vector. In this work, the noise vector $\mathbf{n}$ is assumed to be Gaussian*, $\mathbf{n} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_P\right)$. The PSF $\boldsymbol{\kappa}$ is assumed to be unknown but a nominal PSF estimate $\boldsymbol{\kappa}_0$ is available. The semi-blind deconvolution problem addressed in this paper consists of the joint estimation of $\mathbf{x}$ and $\boldsymbol{\kappa}$ from the noisy measurements $\mathbf{y}$ and nominal PSF $\boldsymbol{\kappa}_0$.

*$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a Gaussian random variable with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

2.2 PSF Basis Expansion

The PSF $\boldsymbol{\kappa}$ (resp. $\{\mathbf{H}\}$) is modeled as a perturbation about a nominal PSF $\boldsymbol{\kappa}_0$ (resp. $\{\mathbf{H}^0\}$). We use a basis expansion as in Park *et al.*[5] constructed by applying principal component analysis to the space of perturbed $\boldsymbol{\kappa}$ about $\boldsymbol{\kappa}_0$. This produces a set of basis $\{\boldsymbol{\kappa}_i\}, i = 1, \ldots, K$ (resp. $\{\mathbf{H}^i\}$) that yield the following linear approximation to $\boldsymbol{\kappa}$,

$$\boldsymbol{\kappa}(\mathbf{c}) = \boldsymbol{\kappa}_0 + \sum_{i=1}^{K} c_i \boldsymbol{\kappa}_i, \tag{2}$$

where the $\{c_i\}$ determine the PSF relative to this bases. With this parameterization, the objective of semi-blind deconvolution is to estimate the unknown image, $\mathbf{x}$, and the linear expansion coefficients $\mathbf{c} = [c_1, \ldots, c_K]^T$.

2.3 Determination of Priors

The priors on the PSF, the image, and the noise are constructed as latent variables in a hierarchical Bayesian model.

2.3.1 Likelihood function

Under the hypothesis that the noise in (1) is white Gaussian, the likelihood function takes the form

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

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

2.3.2 Image and label priors

To induce sparsity and positivity of the image, we use an image prior consisting of "a mixture of a point mass at zero and a single-sided exponential distribution".[5–7] This prior is a convex combination of an atom at zero and an exponential distribution:

$$p(x_i|a, w) = (1 - w)\delta(x_i) + w g(x_i|a). \tag{4}$$

In (4), $\delta(\cdot)$ is the Dirac delta function, $w = \mathrm{P}\left(x_i \neq 0\right)$ is the prior probability of a non-zero pixel and $g(x_i|a) = \frac{1}{a}\exp\left(-\frac{x_i}{a}\right)\mathbf{1}_{\mathbb{R}_+^*}(x_i)$ is a single-sided exponential distribution where $\mathbb{R}_+^*$ is a set of positive real numbers and $\mathbf{1}_{\mathbb{E}}(\cdot)$ denotes the indicator function on the set $\mathbb{E}$. A distinctive property of the image prior (4) is that it can be expressed as a latent variable model

$$p(x_i|a, z_i) = (1 - z_i)\delta(x_i) + z_i g(x_i|a),$$

where the binary variables $\{z_i\}_1^N$ are independent and identically distributed and indicate if the pixel x_i is active

$$z_i = \begin{cases} 1, & \text{if } x_i \neq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

and have the Bernoulli probabilities: $z_i \sim Ber(w)$.

The prior distribution of pixel value x_i in (4) can be rewritten conditionally upon latent variable z_i such that

$$p(x_i|a, z_i) = \delta(x_i)^{1-z_i} g(x_i|a)^{z_i}. \tag{6}$$

By assuming each component x_i to be conditionally independent given z_i and a, the following conditional prior distribution is obtained for $\mathbf{x}$:

$$p(\mathbf{x}|a, \mathbf{z}) = \prod_{i=1}^{N} \left[\delta(x_i)^{1-z_i} g(x_i|a)^{z_i}\right] \tag{7}$$

where $\mathbf{z} = [z_1, \ldots, z_N]$.

This factorized form will turn out to be crucial for simplifying the variational Bayes reconstruction algorithm in Section 3.

2.3.3 PSF parameter prior

We assume that the PSF parameters $c_1, \ldots, c_K$ are independent and c_k is uniformly distributed over intervals

$$\mathcal{S}_k = [-\Delta c_k, \Delta c_k]\,. \tag{8}$$

These intervals are specified a priori and are associated with error tolerances of the imaging instrument. The joint prior distribution of $\mathbf{c} = [c_1, \ldots, c_K]^T$ is therefore:

$$p(\mathbf{c}) = \prod_{k=1}^{K} \frac{1}{2\Delta c_k} \mathbf{1}_{\mathcal{S}_k}(c_k)\,. \tag{9}$$

2.3.4 Noise variance prior

A conjugate inverse-Gamma distribution with parameters ς_0 and ς_1 is assumed as the prior distribution for the noise variance:

$$\sigma^2 | \varsigma_0, \varsigma_1 \sim \mathcal{IG}(\varsigma_0, \varsigma_1)\,. \tag{10}$$

The parameters ς_0 and ς_1 will be fixed to a number small enough to obtain a vague hyperprior, unless we have good prior knowledge.

2.4 Hyperparameter Priors

As reported in Ting *et al.*[7] and Dogigeon *et al.*,[6] the values of the hyperparameters $\{a, w\}$ greatly impact the quality of the deconvolution. Following the approach in Dobigeon *et al.*,[6] we propose to include them within the Bayesian model, leading to a second level of hierarchy in the Bayesian paradigm. This hierarchical Bayesian model requires the definition of prior distributions for these hyperparameters, also referred to as *hyperpriors* which are defined below.

2.4.1 Hyperparameter a

A conjugate inverse-Gamma distribution is assumed for the Laplacian scale parameter a:

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

with $\boldsymbol{\alpha} = [\alpha_0, \alpha_1]^T$. The parameters α_0 and α_1 will be fixed to a number small enough to obtain a vague hyperprior, unless we have good prior knowledge.

2.4.2 Hyperparameter w

We assume a Beta random variable with parameters (β_0, β_1). We set $\beta_0 = \beta_1 = 1$ to obtain a non-informative prior:

$$w \sim \mathcal{B}(\beta_0, \beta_1). \tag{12}$$

2.5 Posterior Distribution

The conditional relationships between variables is illustrated in Fig. 1. The resulting posterior of hidden variables given the observation is

$$p(\mathbf{x}, \sigma^2, a, \mathbf{z}, w, \mathbf{c} | \mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}, \sigma^2, \mathbf{c}) \times p(\mathbf{x}|a, \mathbf{z}) p(\mathbf{z}|w) p(w) p(a) p(\sigma^2) p(\mathbf{c}). \tag{13}$$

Since it is too complex to derive exact Bayesian estimators from this posterior, a variational approximation of this distribution is proposed in the next section.

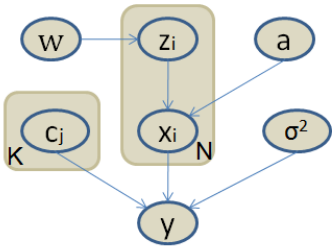


Figure 1. Conditional relationships between variables. A node at an arrow tail conditions the node at the arrow head.

3. VARIATIONAL APPROXIMATION

3.1 Basics of Variational Inference

In this section, we show how to approximate the posterior densities within a variational Bayes framework. Denote by $\mathbf{U}$ the set of all hidden parameter variables including the image variable $\mathbf{x}$ in the model, denoted by $\mathcal{M}$. The hierarchical model implies the Markov representation $p(\mathbf{y}, \mathbf{U}|\mathcal{M}) = p(\mathbf{y}|\mathbf{U}, \mathcal{M})p(\mathbf{U}|\mathcal{M})$. Our objective is to compute the posterior $p(\mathbf{x}|\mathbf{y}, \mathcal{M}) = \int p(\mathbf{y}|\mathbf{U}, \mathcal{M})p(\mathbf{U}|\mathcal{M})d\mathbf{U}_{\setminus \mathbf{x}}/p(\mathbf{y}|\mathcal{M})$, where $\mathbf{U}_{\setminus \mathbf{x}}$ is a set of variables in $\mathbf{U}$ except $\mathbf{x}$. Let q be any arbitrary distribution of $\mathbf{U}$. Then

$$\ln p(\mathbf{y}|\mathcal{M}) = \mathcal{L}(q) + \mathrm{KL}(q\|p) \tag{14}$$

with

$$\mathcal{L}(q) = \int q(\mathbf{U}|\mathcal{M}) \ln \left(\frac{p(\mathbf{y}, \mathbf{U}|\mathcal{M})}{q(\mathbf{U}|\mathcal{M})} \right) d\mathbf{U}$$
$$\mathrm{KL}(q\|p) = -\int q(\mathbf{U}|\mathcal{M}) \ln \left(\frac{p(\mathbf{U}|\mathbf{y}, \mathcal{M})}{q(\mathbf{U}|\mathcal{M})} \right) d\mathbf{U}.$$

We observe that maximizing the lower bound $\mathcal{L}(q)$ is equivalent to minimizing the Kullback-Leibler (KL) divergence $\mathrm{KL}(q\|p)$. Consequently, instead of directly evaluating $p(\mathbf{y}|\mathcal{M})$ given $\mathcal{M}$, we will specify a distribution $q(\mathbf{U}|\mathcal{M})$ that approximates the posterior $p(\mathbf{U}|\mathbf{y}, \mathcal{M})$. The best approximation maximizes $\mathcal{L}(q)$. We present Algorithm 1 that iteratively increases the value of $\mathcal{L}(q)$ by updating posterior surrogate densities. To obtain a tractable approximating distribution q, we will assume a factorized form as $q(\mathbf{U}) = \prod_j q(\mathbf{U}_j)$ where $\mathbf{U}$ has been partitioned into disjoint groups $\mathbf{U}_j$. Subject to this factorization constraint, the optimal distribution $q^*(\mathbf{U}) = \prod_j q^*(\mathbf{U}_j)$ is given by

$$\ln q_j^*(\mathbf{U}_j) = \mathbb{E}_{\setminus \mathbf{U}_j}[\ln p(\mathbf{U}, \mathbf{y})] + (\text{const}), \quad \forall j \tag{15}$$

where $\mathbb{E}_{\setminus \mathbf{U}_j}$ denotes the expectation† with respect to all factors $\mathbf{U}_i$ except $i = j$. We will call $q^*(\mathbf{U})$ the posterior surrogate for p.

3.2 Suggested Factorization

Based on our assumptions on the image and hidden parameters, the random vector is $\mathbf{U} \triangleq \{\boldsymbol{\theta}, \boldsymbol{\phi}\} = \{\mathbf{x}, \mathbf{z}, \mathbf{c}, w, a, \sigma^2\}$ with $\boldsymbol{\theta} = \{\mathbf{x}, \mathbf{z}, \mathbf{c}\}$ and $\boldsymbol{\phi} = \{w, a, \sigma^2\}$. We propose the following factorized approximating distribution

$$q(\mathbf{U}) = q(\mathbf{x}, \sigma^2, a, \mathbf{z}, w, \mathbf{c}) = q(\sigma^2, a, w)q(\mathbf{x}, \mathbf{z}, \mathbf{c}).$$

†In the sequel, we use both $\mathbb{E}(\cdot)$ and $\langle \cdot \rangle$ to denote the expectation. To make our expressions more compact, we use subscripts to denote expectation with respect to the random variables in the subscripts. These notations with the subscripts of '$\setminus \mathbf{v}$' denote expectation with respect to all random variables except for the variable $\mathbf{v}$. e.g. $\mathbb{E}_{\setminus \mathbf{U}_j}$

Ignoring constants[‡], (15) leads to

$$\ln q(\sigma^2, a, w) = \underbrace{\mathbb{E}_{\setminus\sigma^2} \ln p(\mathbf{y}|\mathbf{x}, \sigma^2)p(\sigma^2)}_{\ln q(\sigma^2)} + \underbrace{\mathbb{E}_{\setminus a} \ln p(\mathbf{x}|a, \mathbf{z})p(a)}_{\ln q(a)} + \underbrace{\mathbb{E}_{\setminus w} \ln p(\mathbf{z}|w)p(w)}_{\ln q(w)} \tag{16}$$

which induces the factorization

$$q(\boldsymbol{\phi}) = q(\sigma^2)q(a)q(w). \tag{17}$$

Similarly, the factorized distribution for $\mathbf{x}$, $\mathbf{z}$ and $\mathbf{c}$ is

$$q(\boldsymbol{\theta}) = \left[\prod_i q(x_i|z_i)\right] q(a)q(\mathbf{z}) \tag{18}$$

leading to the fully factorized distribution

$$q(\boldsymbol{\theta}, \boldsymbol{\phi}) = \left[\prod_i q(x_i|z_i)\right] q(a)q(\mathbf{z})q(w)q(\sigma^2)q(\mathbf{c}) \tag{19}$$

3.3 Approximating Distribution q

In this section, we specify the marginal distributions in the approximated posterior distribution required in (19). The parameters for the posterior distributions are evaluated iteratively due to the mutual dependence of the parameters in the distributions for the hidden variables, as illustrated in Algorithm 1.

3.3.1 Posterior surrogate for a

$$q(a) = \mathcal{IG}(\tilde{\alpha}_0, \tilde{\alpha}_1), \tag{20}$$

with $\tilde{\alpha}_0 = \alpha_0 + \sum\langle z_i\rangle$, $\tilde{\alpha}_1 = \alpha_1 + \sum\langle z_i x_i\rangle$.

3.3.2 Posterior surrogate for w

$$q(w) = \mathcal{B}(\tilde{\beta}_0, \tilde{\beta}_1), \tag{21}$$

with $\tilde{\beta}_0 = \beta_0 + N - \sum\langle z_i\rangle$, $\tilde{\beta}_1 = \beta_1 + \sum\langle z_i\rangle$.

3.3.3 Posterior surrogate for σ^2

$$q(\sigma^2) = \mathcal{IG}(\tilde{\varsigma}_0, \tilde{\varsigma}_1), \tag{22}$$

with $\tilde{\varsigma}_0 = P/2 + \varsigma_0$, $\tilde{\varsigma}_1 = \langle\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\rangle/2 + \varsigma_1$, and $\langle\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\rangle = \|\mathbf{y} - \langle\mathbf{H}\rangle\langle\mathbf{x}\rangle\|^2 + \sum \mathrm{var}[x_i]\left[\|\langle\boldsymbol{\kappa}\rangle\|^2 + \sum_l \sigma_{c_l}\|\boldsymbol{\kappa}_l\|^2\right] + \sum_l \sigma_{c_l}\|\mathbf{H}^l\langle\mathbf{x}\rangle\|^2$, where σ_{c_l} is the variance of the Gaussian distribution $q(c_l)$ given in (28) and $\mathrm{var}[x_i]$ is computed under the distribution $q(x_i)$ defined in the next section.

3.3.4 Posterior surrogate for x

We first note that

$$\ln q(\mathbf{x}, \mathbf{z}) = \ln q(\mathbf{x}|\mathbf{z})q(\mathbf{z}) = \mathbb{E}[\ln p(\mathbf{y}|\mathbf{x}, \sigma^2)p(\mathbf{x}|a, \mathbf{z})p(\mathbf{z}|w)]. \tag{23}$$

The conditional density of $\mathbf{x}$ given $\mathbf{z}$ is $p(\mathbf{x}|a, \mathbf{z}) = \prod_i^N g_{z_i}(x_i)$, where $g_0(x_i) \triangleq \delta(x_i), g_1(x_i) \triangleq g(x_i|a)$. Therefore, the conditional posterior surrogate for x_i is

$$q(x_i|z_i = 0) = \delta(x_i), \tag{24}$$

$$q(x_i|z_i = 1) = \phi_+(\mu_i, \eta_i), \tag{25}$$

where $\phi_+(\mu, \sigma^2)$ is a positively truncated-Gaussian density function with the hidden mean μ and variance σ^2, $\eta_i = 1/[\langle\|\mathbf{h}_i\|^2\rangle\langle 1/\sigma^2\rangle]$, $\mu_i = \eta_i[\langle\mathbf{h}_i^T\mathbf{e}_i\rangle\langle 1/\sigma^2\rangle - \langle 1/a\rangle]$, $\mathbf{e}_i = \mathbf{y} - \mathbf{H}\mathbf{x}_{-i}$, $\mathbf{x}_{-i}$ is $\mathbf{x}$ except for the ith entry replaced with 0, and $\mathbf{h}_i$ is the ith column of $\mathbf{H}$. Therefore,

$$q(x_i) = q(z_i = 0)\delta(x_i) + q(z_i = 1)\phi_+(\mu_i, \eta_i), \tag{26}$$

which is a Bernoulli truncated-Gaussian density.

[‡]In the sequel, constant terms with respect to the variables of interest can be omitted in equations.

3.3.5 Posterior surrogate for z

For $i = 1, \ldots, N$,

$$q(z_i = 1) = 1/[1 + C_i'] \text{ and } q(z_i = 0) = 1 - q(z_i = 1), \tag{27}$$

with $C_i' = \exp(C_i/2 \times \tilde{\varsigma}_0/\tilde{\varsigma}_1 + \mu_i\tilde{\alpha}_0/\tilde{\alpha}_1 + \ln\tilde{\alpha}_1 - \psi(\tilde{\alpha}_0) + \psi(\tilde{\beta}_0) - \psi(\tilde{\beta}_1))$. ψ is the digamma function and $C_i = \langle\|\mathbf{h}_i\|^2\rangle(\mu_i^2 + \eta_i) - 2\langle\mathbf{e}_i^T\mathbf{h}_i\rangle\mu_i$.

3.3.6 Posterior surrogate for c

For $j = 1, \ldots, K$,

$$q(c_j) = \phi(\mu_{c_j}, \sigma_{c_j}), \tag{28}$$

where $\phi(\mu, \sigma)$ is the probability density function for the normal distribution with the mean μ and variance σ, $\mu_{c_j} = \langle\mathbf{x}^T{\mathbf{H}^j}^T\mathbf{y} - \mathbf{x}{\mathbf{H}^j}^T\mathbf{H}^0\mathbf{x} - \sum_{l\neq j}\mathbf{x}^T{\mathbf{H}^j}^T\mathbf{H}^l c_l\mathbf{x}\rangle/\langle\mathbf{x}^T{\mathbf{H}^j}^T\mathbf{H}^j\mathbf{x}\rangle$, and $1/\sigma_{c_j} = \langle 1/\sigma^2\rangle\langle\mathbf{x}^T{\mathbf{H}^j}^T\mathbf{H}^j\mathbf{x}\rangle$.

Algorithm 1 VB semi-blind image reconstruction algorithm

% Initialization:
Initialize estimates $\langle\mathbf{x}^{(0)}\rangle$, $\langle\mathbf{z}^{(0)}\rangle$, and $w^{(0)}$, and set $\mathbf{c} = \mathbf{0}$ to have $\hat{\boldsymbol{\kappa}}^{(0)} = \boldsymbol{\kappa}_0$,
% Iterations:
for $t = 1, 2, \ldots,$ **do**
  Evaluate $\tilde{\alpha}_0^{(t)}, \tilde{\alpha}_1^{(t)}$ in (20) by using $\langle\mathbf{x}^{(t-1)}\rangle, \langle\mathbf{z}^{(t-1)}\rangle$,
  Evaluate $\tilde{\beta}_0^{(t)}, \tilde{\beta}_1^{(t)}$ in (21) by using $\langle\mathbf{z}^{(t-1)}\rangle$,
  Evaluate $\tilde{\varsigma}_0^{(t)}, \tilde{\varsigma}_1^{(t)}$ in (22) from $\langle\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\rangle$,
  for $i = 1, 2, \ldots, N$ **do**
    Evaluate necessary statistics (μ_i, η_i) for $q(x_i|z_i = 1)$ in (24),
    Evaluate $q(z_i = 1)$ in (27),
    Evaluate $\langle x_i\rangle, \mathrm{var}[x_i]$,
    For $l = 1, \ldots, K$, evaluate $\mu_{c_l}, 1/\sigma_{c_l}$ for $q(c_l)$ in (28),
  end for
end for

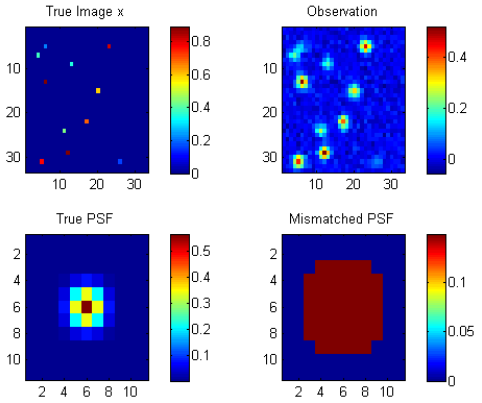
The final iterative algorithm is presented in Algorithm 1, where required *shaping* parameters under distributional assumptions and related statistics are iteratively updated.

4. SIMULATION RESULTS

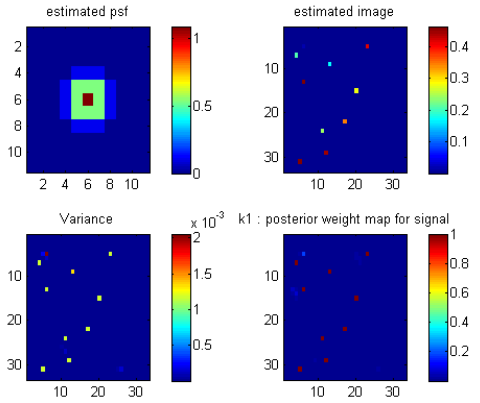
We first present numerical results obtained for Gaussian and typical MRFM PSFs, shown in Fig. 2(a) and Fig. 3(a), respectively. Then the proposed variational algorithm is applied to a tobacco virus MRFM data set. There are many possible approaches to selecting hyperparamters, including the non-informative approach[6] and the expectation-maximization approach.[9] In our experiments, hyper-parameters ς_0, ς_1, α_0, and α_1 for the densities are chosen based on the framework advocated in Dobigeon *et al.*[6] This leads to vague priors corresponding to selecting $\varsigma_0 = \varsigma_1 = \alpha_0 = \alpha_1 = 1$. The initial image used to initialize the algorithm is obtained from one Landweber iteration.[25]

4.1 Simulation with Gaussian PSF

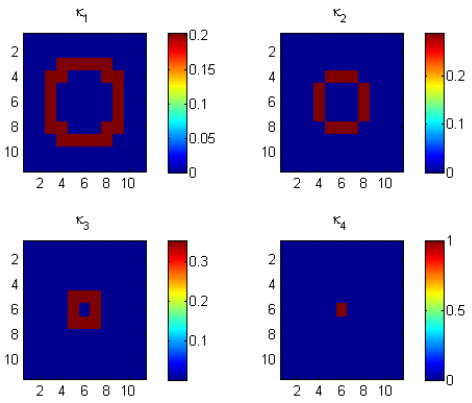
The true image $\mathbf{x}$ used to generate the data, observation $\mathbf{y}$, the true PSF, and the initial, mismatched PSF are shown in Fig. 2(a). The final estimated PSF and reconstructed image are depicted in Fig. 2(b), along with the reconstructed variances and posterior probability of $z_i \neq 0$. We dissected the support region of the true PSF, to produce 4 PSF bases $\{\boldsymbol{\kappa}_i\}_i$ shown in Fig. 2(c). The reconstructed PSF clearly matches the true one, as seen in Fig. 2(b) and Fig. 2(a). Note that the restored image is attenuated while the restored PSF is amplified because of intrinsic scale ambiguity.



(a) True image, observation, true PSF, and mismatched PSF ($\boldsymbol{\kappa}_0$).

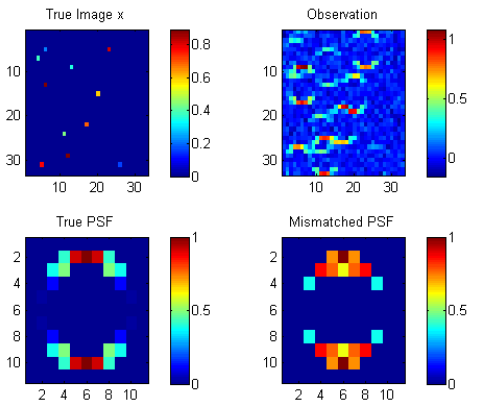


(b) Restored PSF and image with variance and weight map. $\hat{\boldsymbol{\kappa}} = \mathbb{E}\boldsymbol{\kappa}$ is close to the true one.

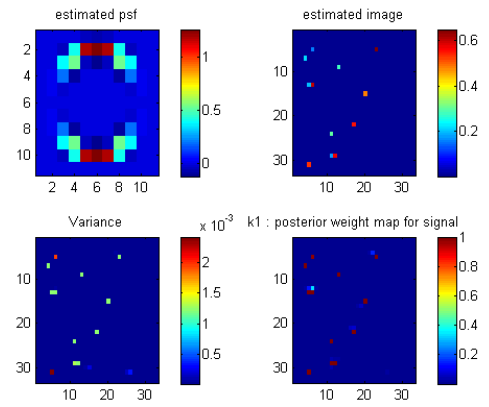


(c) PSF bases, $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_4$, for Gaussian PSF.

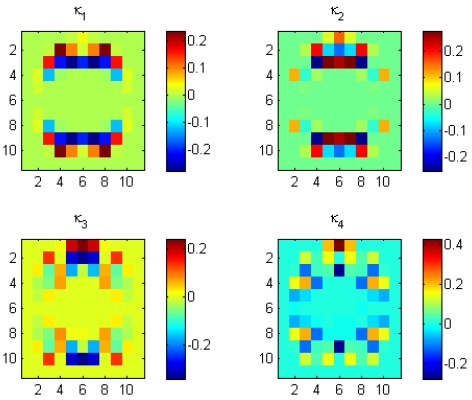
Figure 2. Experiment with Gaussian PSF



(a) true image, observation, true PSF, and mismatched PSF ($\boldsymbol{\kappa}_0$).



(b) Restored PSF and image with variance and weight map. $\hat{\boldsymbol{\kappa}} = \mathbb{E}\boldsymbol{\kappa}$ is close to the true one.



(c) PSF bases, $\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_4$, for MRFM PSF.

Figure 3. Experiment with simplified MRFM PSF

4.2 Simulation with MRFM type PSFs

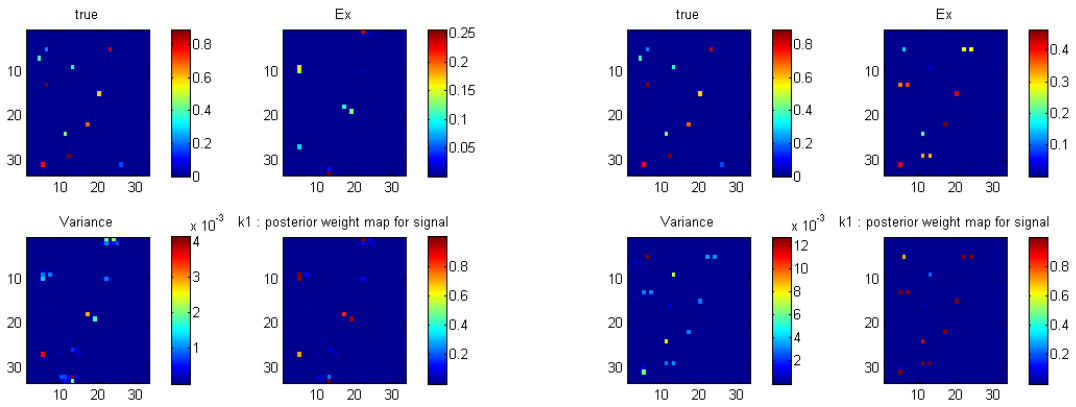
The true image $\mathbf{x}$ used to generate the data, observation $\mathbf{y}$, the true PSF, and the initial, mismatched PSF are shown in Fig. 3(a). The PSF models the PSF of the MRFM instrument, derived by Mamin *et al.*[26] The convergence is achieved after the 10th iteration. The reconstructed image can be compared to the true image in Fig. 3(b), where the pixel-wise variances and posterior probability of $z_i \neq 0$ are rendered. The PSF bases are obtained by the procedure proposed by Park *et al.*[5] The resulting $K = 4$ principal basis vectors are depicted in Fig.3(c). The reconstructed PSF with the bases clearly matches the true one, as seen in Fig. 3(b) and Fig. 3(a).

4.3 Comparison with PSF-mismatched reconstruction

The results from the variational deconvolution algorithm with a mismatched Gaussian PSF and a MRFM type PSF are presented in Fig. 4(a) and Fig. 4(b), respectively; the relevant PSFs and observations are presented in Fig. 2(a) in Section 4.1 and in Fig. 3(a) in Section 4.2, respectively. Compared with the results of our VB semi-blind algorithm (Algorithm 1), shown in Fig. 2(b) and Fig. 3(b), the reconstructed images in Fig. 4(a) and Fig. 4(b), respectively, inaccurately estimate signal locations and blur most of the non-zero values.

4.4 Comparison with the MCMC semi-blind algorithm

To compare the proposed method with the previous MCMC semi-blind deconvolution method proposed by Park *et al.*,[5] we present the approximated posterior density $q(x_i)$ for a non-zero pixel x_i estimated from both



(a) Non-myopic result with a mismatched Gaussian PSF.

(b) Non-myopic result with a mismatched MRFM type PSF.

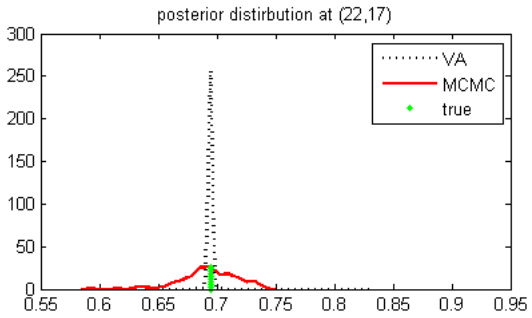


Figure 4. Two approximate posterior densities at (22,17). The variance of the variational density is much smaller than that of the previous method. (Because of non-uniqueness in scale in blind deconvolution, distributions from variational and MCMC methods are scaled to have the peak at the true value for a fair comparison of their envelopes.)

methods. We picked the nonzero pixel at the coordinate $(22, 17)$ of the true sparse image (top-left in Fig. 3(a)) and used a MRFM type PSF (bottom in Fig. 3(a)) for the comparison. The posterior distribution estimated by the MCMC method is approximated by averaging the posteriors of 180 samples generated after the burn-in period. Conversely, the posterior distribution approximated by the proposed variational approach is the Bernoulli truncated-Gaussian distribution in (26), whose parameters are computed from the final outputs of the iterative algorithm. The approximated densities are superimposed in Fig. 4[§]. In this figure, the precision obtained in our method is evident, whereas the previous sampling based method seems to suffer from additional Monte Carlo errors.

To quantify the comparison, we performed experiments with the same set of four sparse images as used in Park *et al.*[5] By generating 100 different noise realizations for 100 independent trials with each true image, we measured errors according to various criteria. We tested four sparse images with sparsity levels $\|\mathbf{x}\|_0 = 6, 11, 18, 30$. Fig. 5 visualizes reconstruction error performance for several measures of error[¶] . From these figures we conclude that the VB semi-blind algorithm performs at least as well as the previous MCMC semi-blind algorithm. In addition, the VB method outperforms AM[27] and non-blind MCMC[6] methods. In terms of PSF estimation, the VB semi-

[§]The delta function in the variational density is not visible since the estimated weight for the function is negligible to be rendered.

[¶]Note that the ℓ_0 norm has been normalized. The true image has value 1; $\|\hat{\mathbf{x}}\|_0/\|\mathbf{x}\|_0$ is used for MCMC method; $\mathbb{E}(w) \times N/\|\mathbf{x}\|_0$ for variational method since this method does not produce zero pixels but $\mathbb{E}(w)$.
Note also that, for our simulated data, the (normalized) true noise levels are $\|\mathbf{n}\|^2/\|\mathbf{x}\|_0 = 0.1475, 0.2975, 0.2831, 0.3062$ for $\|\mathbf{x}\|_0 = 6, 11, 18, 30$, respectively.

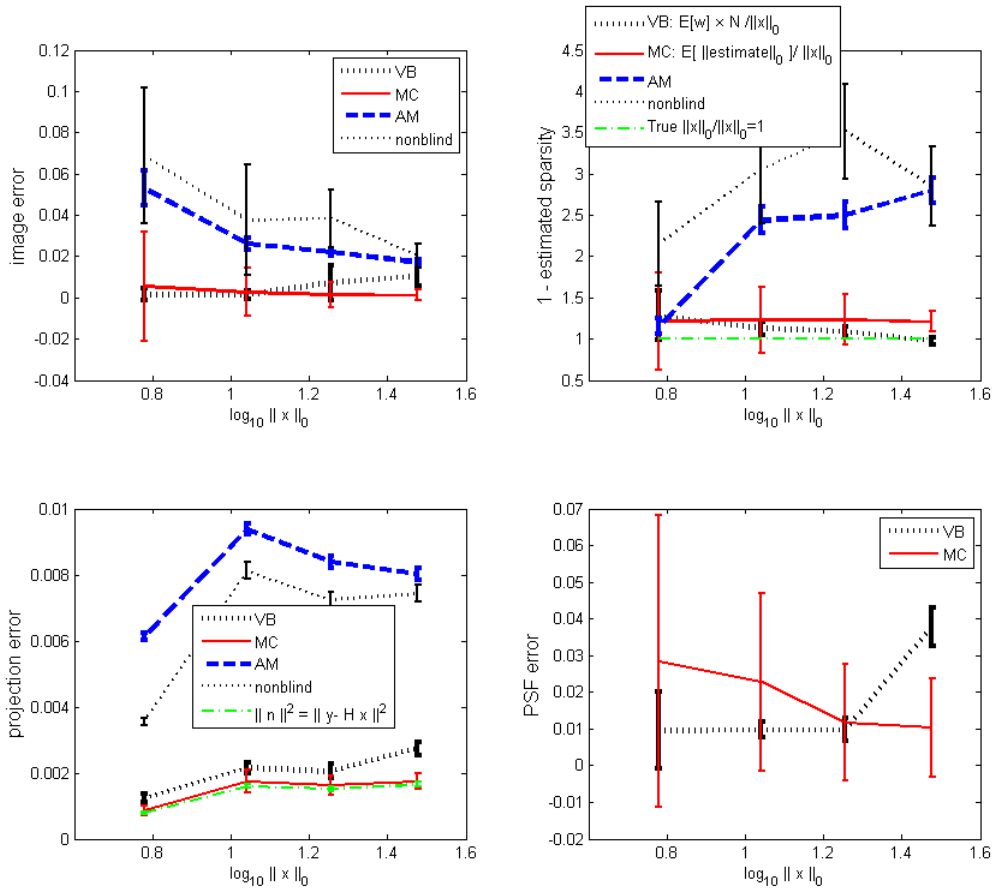


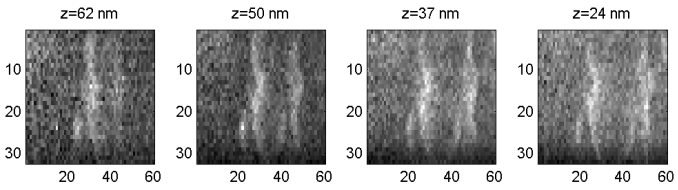
Figure 5. Error bar graphs of results from two semi-blind deconvolution algorithm. For several images of different sparsity levels, errors are illustrated with standard deviations. Quantities for 'image error' (top-left), 'projection error' (bottom-left), and 'PSF error' (bottom-right) are $\mathbb{E}[\|\mathbf{x}/\|\mathbf{x}\| - \hat{\mathbf{x}}/\|\hat{\mathbf{x}}\|\|^2]/\|\mathbf{x}\|_0$, $\mathbb{E}[\|\mathbf{y} - \hat{\mathbf{y}}\|^2]/\|\mathbf{x}\|_0$, and $\mathbb{E}\|\boldsymbol{\kappa}/\|\boldsymbol{\kappa}\| - \hat{\boldsymbol{\kappa}}/\|\hat{\boldsymbol{\kappa}}\|\|^2$, respectively. 'VB', 'MC', 'AM', and 'nonblind' are the semi-blind variational method (proposed here), semi-blind MCMC method,[5] alternating minimization method,[27] and non-blind method,[6] respectively.

blind method seems to outperform the MCMC method, especially for very sparse images. Also, the proposed VB semi-blind method converges more quickly and requires fewer iterations. For example, the VB semi-blind algorithm converges in approximately 9.5 seconds after 12 iterations, but the previous MCMC algorithm takes more than 19.2 seconds after 40 iterations to achieve convergence[‖].

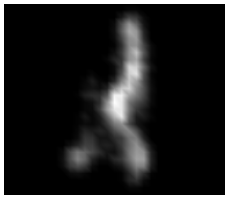
4.5 Application to tobacco mosaic virus (TMV) data

We applied the proposed variational semi-blind sparse deconvolution algorithm to the tobacco mosaic virus data, made available by our IBM collaborators,[20] shown in Fig. 6(a). Our algorithm is easily modifiable to these 3D raw image data and 3D PSF with an additional dimension in dealing with basis functions to evaluate each voxel value x_i. The reconstruction of the 6th layer is shown in Fig. 6(b), and is consistent with the results obtained by other methods.[5,6] The estimated deviation in PSF is small, as predicted in Park *et al.*[5]

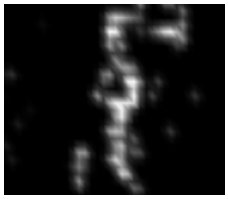
[‖]Matlab is used under Fedora Release 12 and Intel Xeon E5506 CPU in Dell Precision T5500 platform.



(a) TMV raw data.



(b) VB: Estimated virus image.



(c) MCMC: Estimated virus image from Park *et al.*[5]

Figure 6. TMV data from MRFM experiment and estimates.

5. CONCLUSION

We suggested a novel variational solution to a semi-blind sparse image reconstruction problem. Our method uses Bayesian inference for image and PSF restoration with a sparsity-inducing image prior via the variational Bayes approximation. From the simulation results, we conclude that the performance of the VB method competes with MCMC methods in sparse image estimation, while requiring fewer computations. Compared to a non-myopic algorithm whose mismatched PSF leads to imprecise and blurred signal locations in the restored image, the VB semi-blind algorithm correctly produces sparse image estimates. The benefits of this solution compared to the previous MCMC solution are: 1) faster convergence, 2) stability of the method, and 3) memory efficiency.

REFERENCES

[1] Makni, S., Ciuciu, P., Idier, J., and Poline, J.-B., "Semi-blind deconvolution of neural impulse response in fMRI using a Gibbs sampling method," in [*Proc. IEEE Int. Conf. Acoust., Speech, and Signal (ICASSP)*], **5**, 601–604 (May 2004).

[2] Pillonetto, G. and Cobelli, C., "Identifiability of the stochastic semi-blind deconvolution problem for a class of time-invariant linear systems," *Automatica* **43**, 647–654 (April 2007).

[3] Sarri, P., Thomas, G., Sekko, E., and Neveux, P., "Myopic deconvolution combining Kalman filter and tracking control," in [*Proc. IEEE Int. Conf. Acoust., Speech, and Signal (ICASSP)*], **3**, 1833–1836 (1998).

[4] Chenegros, G., Mugnier, L. M., Lacombe, F., and Glanc, M., "3D phase diversity: a myopic deconvolution method for short-exposure images: application to retinal imaging," *J. Opt. Soc. Am. A* **24**, 1349–1357 (May 2007).

[5] Park, S. U., Dobigeon, N., and Hero, A. O., "Myopic sparse image reconstruction with application to MRFM," in [*Proc. Computational Imaging Conference in IS&T SPIE Symposium on Electronic Imaging Science and Technology*], Bouman, C. A., Pollak, I., and Wolfe, P. J., eds., *Computational Imaging IX* **7873**, SPIE (Jan. 2011).

[6] Dobigeon, N., Hero, A. O., and Tourneret, J.-Y., "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Processing* **18**, 2059–2070 (Sept. 2009).

[7] Ting, M., Raich, R., and Hero, A. O., "Sparse image reconstruction for molecular imaging," *IEEE Trans. Image Processing* **18**, 1215–1227 (June 2009).

[8] Bishop, C. M., [*Pattern Recognition and Machine Learning*], Springer, New York, NY, USA (2006).

[9] Nasios, N. and Bors, A., "Variational learning for Gaussian mixture models," *IEEE Trans. Systems, Man, Cybernet. Part B* **36**, 849–862 (Aug. 2006).

[10] Corduneanu, A. and Bishop, C. M., "Variational Bayesian model selection for mixture distributions," in [*Proc. Conf. Artificial Intelligence and Statistics*], (2001).

[11] Robert, C. P. and Casella, G., [*Monte Carlo Statistical Methods*], Springer, New York, NY, USA, 2 ed. (2004).

[12] Gilks, W. R., [*Markov Chain Monte Carlo In Practice*], Chapman and Hall/CRC (1999).

[13] Orieux, F., Giovannelli, J.-F., and Rodet, T., "Bayesian estimation of regularization and point spread function parameters for Wiener-Hunt deconvolution," *J. Opt. Soc. Am. A* **27**, 1593–1607 (July 2010).

[14] Attias, H., "A variational Bayesian framework for graphical models," in [*Proc. Advances in Neural Information Processing Systems (NIPS)*], 209–215, MIT Press (2000).

[15] Sidles, J. A., "Noninductive detection of single-proton magnetic resonance," *Appl. Phys. Lett.* **58**, 2854–2856 (June 1991).

[16] Sidles, J. A., "Folded stern-gerlach experiment as a means for detecting nuclear magnetic resonance in individual nuclei," *Phys. Rev. Lett.* **68**, 1124–1127 (Feb 1992).

[17] Sidles, J. A., Garbini, J. L., Bruland, K. J., Rugar, D., Züger, O., Hoen, S., and Yannoni, C. S., "Magnetic resonance force microscopy," *Rev. Mod. Phys.* **67**, 249–265 (Jan 1995).

[18] Rugar, D., Yannoni, C. S., and Sidles, J. A., "Mechanical detection of magnetic resonance," *Nature* **360**, 563–566 (Dec. 1992).

[19] Züger, O., Hoen, S. T., Yannoni, C. S., and Rugar, D., "Three-dimensional imaging with a nuclear magnetic resonance force microscope," *J. Appl. Phys.* **79**, 1881–1884 (Feb. 1996).

[20] Degen, C. L., Poggio, M., Mamin, H. J., Rettner, C. T., and Rugar, D., "Nanoscale magnetic resonance imaging," *Proc. Nat. Academy of Science* **106**, 1313–1317 (Feb. 2009).

[21] Züger, O. and Rugar, D., "First images from a magnetic resonance force microscope," *Applied Physics Letters* **63**(18), 2496–2498 (1993).

[22] Züger, O. and Rugar, D., "Magnetic resonance detection and imaging using force microscope techniques," *J. Appl. Phys.* **75**, 6211–6216 (May 1994).

[23] Chao, S., Dougherty, W. M., Garbini, J. L., and Sidles, J. A., "Nanometer-scale magnetic resonance imaging," *Review Sci. Instrum.* **75**, 1175–1181 (April 2004).

[24] Degen, C. L., Poggio, M., Mamin, H. J., Rettner, C. T., and Rugar, D., "Nanoscale magnetic resonance imaging. Supporting information," *Proc. Nat. Academy of Science* **106** (Feb. 2009).

[25] Landweber, L., "An iteration formula for Fredholm integral equations of the first kind," *Amer. J. Math.* **73**, 615–624 (July 1951).

[26] Mamin, J., Budakian, R., and Rugar, D., "Point response function of an MRFM tip," tech. rep., IBM Research Division (Oct. 2003).

[27] Herrity, K., Raich, R., and Hero, A. O., "Blind reconstruction of sparse images with unknown point spread function," in [*Proc. Computational Imaging Conference in IS&TSPIE Symposium on Electronic Imaging Science and Technology*], Bouman, C. A., Miller, E. L., and Pollak, I., eds., **6814**, SPIE, San Jose, CA, USA (Jan. 2008).