Variational semi-blind sparse deconvolution with orthogonal kernel bases and its application to MRFM

Se Un Park, Nicolas Dobigeon, Alfred O. Hero

Variational semi-blind sparse deconvolution with orthogonal kernel bases and its application to MRFM ☆

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

[a] University of Michigan, Department of EECS, Ann Arbor, MI 48109-2122, USA
[b] University of Toulouse, IRIT/INP-ENSEEIHT, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

A B S T R A C T

We present a variational Bayesian method 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. Importantly, the proposed Bayesian deconvolution algorithm does not require hand tuning. Simulation results clearly demonstrate that the semi-blind deconvolution algorithm compares favorably with previous Markov chain Monte Carlo (MCMC) version of myopic sparse reconstruction. It significantly outperforms mismatched non-blind algorithms that rely on the assumption of the 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

The standard and popular image deconvolution techniques generally assume that the space-invariant instrument response, i.e., the point spread function (PSF), is perfectly known. However, in many practical situations, the true PSF is either unknown or, at best, partially known. For example, in an optical system a perfectly known PSF does not exist because of light diffraction, apparatus/lense aberration, out-of-focus, or image motion [1,2]. Such imperfections are common in general imaging systems including MRFM, where there exist additional model PSF errors in the sensitive magnetic resonance condition [3]. 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 [4,5] or myopic deconvolution [6–8].

In this paper, we formulate the semi-blind deconvolution task as an estimation problem in a Bayesian setting. Bayesian estimation has the great advantage of offering 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, via posterior variance and posterior confidence intervals. Extending our previous work, we propose a variational estimator for the parameters as contrasted to the Monte Carlo approach in [9]. This extension is non-trivial. Our variational Bayes algorithm iterates on a hidden variable domain associated with the mixture coefficients. This algorithm is faster, more scalable for equivalent image reconstruction qualities in [9].

☆ This research was partially supported by a grant from ARO, Grant no. W911NF-05-1-0403.

* Corresponding author. Tel.: +1 734 763 4497; fax: +1 734 763 8041.
E-mail addresses: seunpark@umich.edu (S.U. Park), nicolas.dobigeon@enseeiht.fr (N. Dobigeon), hero@umich.edu (A.O. Hero).

http://dx.doi.org/10.1016/j.sigpro.2013.06.013

Like in [9], the PSF uncertainty is modeled as the deviation of the a priori known PSF from the true PSF. Applying an eigendecomposition to 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 that the desired image is sparse, corresponding to the natural sparsity of the molecular image. The image prior is a weighted sum of a sparsity inducing part and a continuous distribution; a positive truncated *Laplacian and atom at zero* (LAZE) prior[1] [10]. Similar priors have been applied for estimating mixtures of densities [11–13] and sparse, nonnegative hyperspectral unmixing [14]. Here we introduce a hidden label variable for the contribution of the discrete mass (empty pixel) and a continuous density function (non-empty pixel). Similar to our 'hybrid' mixture model, inhomogeneous Gamma-Gaussian mixture models have been proposed in [15].

Bayesian inference of parameters from the posterior distribution generally requires challenging computations, such as functional optimization 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 can be accomplished using Markov chain Monte Carlo (MCMC) methods [16]. MCMC has been successfully adopted in numerous imaging problems such as image segmentation, denoising, and deblurring [17,16]. Recently, to solve blind deconvolution, two promising semi-blind MCMC methods have been suggested [9,18]. However, these sampling methods have the disadvantage that convergence may be slow.

An alternative to Monte Carlo integration is a variational approximation to the posterior distribution, and this approach is adopted in this paper. These approximations have been extensively exploited to conduct inference in graphical models [19]. If properly designed, they can produce an analytical posterior distribution from which Bayesian estimators can be efficiently computed. Compared to MCMC, variational methods are of lower computational complexity, since they avoid stochastic simulation. However, variational Bayes (VB) approaches have intrinsic limits; the convergence to the true distribution is not guaranteed, even though the posterior distribution will be asymptotically normal with mean equal to the maximum likelihood estimator under suitable conditions [20]. In addition, variational Bayes approximations can be easily implemented for only a limited number of statistical models. For example, this method is difficult to apply when latent variables have distributions that do not belong to the exponential family (e.g. a discrete distribution [9]). For mixture distributions, variational estimators in Gaussian mixtures and in exponential family converge locally to maximum likelihood estimator [21,22]. The theoretical convergence properties for sparse mixture models, such as our proposed model, are as yet unknown. This has not hindered the application of VB to sparse models to problems in our sparse image mixture model. Another possible intrinsic limit of the variational Bayes approach, particularly in (semi)-blind deconvolution, is that the posterior covariance structure cannot be effectively estimated nor recovered, unless the true joint distributions have independent individual distributions. This is primarily because VB algorithms are based on minimizing the KL-divergence between the true distribution and the VB approximating distribution, which is assumed to be factorized with respect to the individual parameters.

However, despite these limits, VB approaches have been widely applied with success to many different engineering problems [23–26]. A principal contribution of this paper is the development and implementation of a VB algorithm for mixture distributions in a hierarchical Bayesian model [27]. Similarly, the framework permits a Gaussian prior [28] or a Student's-t prior [29] for the PSF. We present comparisons of our variational solution to other blind deconvolution methods. These include the total variation (TV) prior for the PSF [30] and natural sharp edge priors for images with PSF regularization [31]. We also compare to basis kernels [29], the mixture model algorithm of Fergus et al. [32], and the related method of Shan et al. [33] under a motion blur model.

To implement variational Bayesian inference, prior distributions and the instrument-dependent likelihood function are specified. Then the posterior distributions are estimated by minimizing the Kullback–Leibler (KL) distance between the model and the empirical distribution. Simulations conducted on synthetic images show that the resulting myopic deconvolution algorithm outperforms previous mismatched non-blind algorithms and competes with the previous MCMC-based semi-blind method [9] with lower computational complexity.

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 [34–36]. Recently, MRFM has successfully demonstrated imaging [37,38] of a tobacco mosaic virus [39]. The 3D image reconstruction problem for MRFM experiments was investigated with Wiener filters [40,41,38], iterative least square reconstruction approaches [42,39], and recently the Bayesian estimation framework [10,43,8,9]. The drawback of these approaches is that they require prior knowledge on 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 [3]. The proposed semi-blind reconstruction method accounts for this partial knowledge.

The rest of this paper is organized as follows. Section 2 formulates the imaging deconvolution problem in a hierarchical Bayesian framework. Section 3 covers the variational methodology and our proposed solutions. Section 4 reports simulation results and an application to the real MRFM data. Section 6 discusses our findings and concludes.

2. Formulation

2.1. Image model

As in [9,43], the image model is defined as

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

[1] A Laplace distribution as a prior distribution acts as a sparse regularization using ℓ_1 norm. This can be seen by taking negative logarithm on the distribution.

where $\mathbf{y}$ is a $P \times 1$ vectorized measurement, $\mathbf{x} = [x_1, \ldots, x_N]^T \geqslant 0$ 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,[2] $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_P)$. 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 $\boldsymbol{\kappa}_0$.

2.2. PSF basis expansion

The nominal PSF $\boldsymbol{\kappa}_0$ is assumed to be generated with known parameters (gathered in the vector $\boldsymbol{\zeta}_0$) tuned during imaging experiments. However, due to model mismatch and experimental errors, the true PSF $\boldsymbol{\kappa}$ may deviate from the nominal PSF $\boldsymbol{\kappa}_0$. If the generation model for PSFs is complex, direct estimation of a parameter deviation, $\Delta\boldsymbol{\zeta} = \boldsymbol{\zeta}_{\boldsymbol{true}} - \boldsymbol{\zeta}_{\mathbf{0}}$, is difficult.

We model the PSF $\boldsymbol{\kappa}$ (resp. $\{\mathbf{H}\}$) as a perturbation about a nominal PSF $\boldsymbol{\kappa}_0$ (resp. $\{\mathbf{H}^0\}$) with K basis vectors $\boldsymbol{\kappa}_k$, $k = 1, \ldots, K$, that span a subspace representing possible perturbations $\Delta\boldsymbol{\kappa}$. We empirically determined this basis using the following PSF variational eigendecomposition approach. A number of PSFs $\tilde{\boldsymbol{\kappa}}$ are generated following the PSF generation model with parameters $\boldsymbol{\zeta}$ randomly drawn according to the Gaussian distribution[3] centered at the nominal values $\boldsymbol{\zeta}_0$. Then a standard principal component analysis (PCA) of the residuals $\{\tilde{\boldsymbol{\kappa}}_j - \boldsymbol{\kappa}_0\}_{j=1,\ldots}$ is used to identify K principal axes that are associated with the basis vectors $\boldsymbol{\kappa}_k$. The necessary number of basis vectors, K, is determined empirically by detecting a knee at the scree plot. The first few eigenfunctions, corresponding to the first few largest eigenvalues, explain major portion of the observed perturbations. If there is no PSF generation model, then we can decompose the support region of the true (suspected) PSF to produce an orthonormal basis. The necessary number of the bases is again chosen to explain most support areas that have major portion/energy of the desired PSF. This approach is presented in our experiment with Gaussian PSFs.

We use a basis expansion to present $\boldsymbol{\kappa}(\mathbf{c})$ as 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 $\{c_i\}$ 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] $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a Gaussian random variable with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

[3] The variances of the Gaussian distributions are carefully tuned so that their standard deviations produce a minimal volume ellipsoid that contains the set of valid PSFs.

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

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

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" [10,43,9]. 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, $w = \mathrm{P}(x_i \neq 0)$ is the prior probability of a non-zero pixel and $g(x_i|a) = (1/a)\exp(-x_i/a)\mathbf{1}_{\mathbb{R}_+^*}(x_i)$ 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 $\mathbb{E}$

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

A distinctive property of the image prior (4) is that it can be expressed as a latent variable model

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

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

$$z_i = \begin{cases} 1 & \text{if } x_i \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

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 as

$$p(x_i|z_i = 0) = \delta(x_i),$$
$$p(x_i|a, z_i = 1) = g(x_i|a),$$

which can be summarized in the following factorized form:

$$p(x_i|a, z_i) = \delta(x_i)^{1-z_i} g(x_i|a)^{z_i}. \tag{8}$$

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

$$p(\mathbf{x}|a, \mathbf{z}) = \prod_{i=1}^{N} [\delta(x_i)^{1-z_i} g(x_i|a)^{z_i}] \tag{9}$$

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.

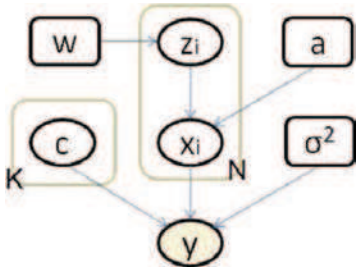

Fig. 1. Conditional relationships between variables. A node at an arrow tail conditions the node at the arrow head.

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

$$\mathcal{S}_k = [-\Delta c_k, \Delta c_k]. \tag{10}$$

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

$$p(\mathbf{c}) = \prod_{k=1}^{K} \frac{1}{2\Delta c_k} \mathbf{1}_{\mathcal{S}_k}(c_k). \tag{11}$$

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 (see Appendix A.1 for the details of this distribution):

$$\sigma^2 | \varsigma_0, \varsigma_1 \sim \mathcal{IG}(\varsigma_0, \varsigma_1). \tag{12}$$

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 [10,43], the values of the hyperparameters $\{a, w\}$ greatly impact the quality of the deconvolution. Following the approach in [9], 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

$$a | \boldsymbol{\alpha} \sim \mathcal{IG}(\alpha_0, \alpha_1), \tag{13}$$

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), which are iteratively updated in accordance with data fidelity. The parameter values will reflect the degree of prior knowledge and we set $\beta_0 = \beta_1 = 1$ to obtain a non-informative prior (see Appendix A.2 for the details of this distribution)

$$w \sim \mathcal{B}(\beta_0, \beta_1). \tag{14}$$

2.5. Posterior distribution

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

$$p(\mathbf{x}, a, \mathbf{z}, w, \mathbf{c}, \sigma^2 | \mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}, \mathbf{c}, \sigma^2) \times p(\mathbf{x}|a, \mathbf{z}) p(\mathbf{z}|w) p(w) p(a) p(\mathbf{c}) p(\sigma^2). \tag{15}$$

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.

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

$$\ln p(\mathbf{y}|\mathcal{M}) = \mathcal{L}(q) + \mathrm{KL}(q \| p) \tag{16}$$

with

$$\mathcal{L}(q) = \int q(\mathbf{U}|\mathcal{M}) \ln \left(\frac{p(\mathbf{y}, \mathbf{U}|\mathcal{M})}{q(\mathbf{U}|\mathcal{M})} \right) d\mathbf{U} \tag{17}$$

$$\mathrm{KL}(q \| p) = - \int q(\mathbf{U}|\mathcal{M}) \ln \left(\frac{p(\mathbf{U}|\mathbf{y}, \mathcal{M})}{q(\mathbf{U}|\mathcal{M})} \right) d\mathbf{U}. \tag{18}$$

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

$$\ln q_j^*(\mathbf{U}_j) = \mathrm{E}_{\setminus \mathbf{U}_j}[\ln p(\mathbf{U}, \mathbf{y})] + (\text{const}), \quad \forall j \tag{19}$$

where $\mathrm{E}_{\setminus \mathbf{U}_j}$ denotes the expectation[4] with respect to all factors $\mathbf{U}_i$ except $i = j$. We will call $q^*(\mathbf{U})$ the posterior surrogate for p.

[4] In the sequel, we use both $\mathrm{E}[\cdot]$ 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. $\mathrm{E}_{\setminus \mathbf{U}_j}$.

3.2. Suggested factorization

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

$$q(\mathbf{U})=q(\mathbf{x},a,\mathbf{z},w,\mathbf{c},\sigma^2)=q(\mathbf{x},\mathbf{z},\mathbf{c})q(a,w,\sigma^2). \tag{20}$$

Ignoring constants,[5] (19) leads to

$$\ln q(a,w,\sigma^2)=\underbrace{\mathrm{E}_{\setminus a}\ln p(\mathbf{x}|a,\mathbf{z})p(a)}_{\ln q(a)} + \underbrace{\mathrm{E}_{\setminus w}\ln p(\mathbf{z}|w)p(w)}_{\ln q(w)} + \underbrace{\mathrm{E}_{\setminus\sigma^2}\ln p(\mathbf{y}|\mathbf{x},\sigma^2)p(\sigma^2)}_{\ln q(\sigma^2)} \tag{21}$$

which induces the factorization

$$q(\boldsymbol{\phi})=q(a)q(w)q(\sigma^2). \tag{22}$$

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

$$q(\boldsymbol{\theta})=\left[\prod_i q(x_i|z_i)\right]q(\mathbf{z})q(\mathbf{c}) \tag{23}$$

leading to the fully factorized distribution

$$q(\boldsymbol{\theta},\boldsymbol{\phi})=\left[\prod_i q(x_i|z_i)\right]q(a)q(\mathbf{z})q(w)q(\mathbf{c})q(\sigma^2) \tag{24}$$

3.3. Approximating distribution q

In this section, we specify the marginal distributions in the approximated posterior distribution required in (24). More details are described in Appendix B. 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

$$q(a)=\mathcal{IG}(\tilde{\alpha}_0,\tilde{\alpha}_1), \tag{25}$$

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

$$q(w)=\mathcal{B}(\tilde{\beta}_0,\tilde{\beta}_1), \tag{26}$$

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

$$q(\sigma^2)=\mathcal{IG}(\tilde{\varsigma}_0,\tilde{\varsigma}_1), \tag{27}$$

with $\tilde{\varsigma}_0=P/2+\varsigma_0$, $\tilde{\varsigma}_1=\langle\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\rangle/2+\varsigma_1$, and $\langle\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\rangle = \|\mathbf{y}-\langle\mathbf{H}\rangle\langle\mathbf{x}\rangle\|^2+\sum\mathrm{var}[x_i][\|\langle\boldsymbol{\kappa}\rangle\|^2+\sum_l\sigma_{c_l}\|\boldsymbol{\kappa}_l\|^2]+\sum_l\sigma_{c_l}\|\mathbf{H}^l\langle\mathbf{x}\rangle\|^2$, where σ_{c_l} is the variance of the Gaussian distribution $q(c_l)$ given in (33) and $\mathrm{var}[x_i]$ is computed under the distribution $q(x_i)$ defined in the next section and described in Appendix B.3.

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

3.3.4. Posterior surrogate for x

We first note that

$$\ln q(\mathbf{x},\mathbf{z})=\ln q(\mathbf{x}|\mathbf{z})q(\mathbf{z})=\mathrm{E}[\ln p(\mathbf{y}|\mathbf{x},\sigma^2)p(\mathbf{x}|a,\mathbf{z})p(\mathbf{z}|w)]. \tag{28}$$

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

$$q(x_i|z_i=0)=\delta(x_i), \tag{29}$$

$$q(x_i|z_i=1)=\phi_+(\mu_i,\eta_i), \tag{30}$$

where $\phi_+(\mu,\sigma^2)$ is a positively truncated-Gaussian density function with the hidden mean μ and variance σ^2, $\eta_i=1/[\langle\|\mathbf{h}_i\|^2\rangle\langle1/\sigma^2\rangle]$, $\mu_i=\eta_i[\langle\mathbf{h}_i^T\mathbf{e}_i\rangle\langle1/\sigma^2\rangle-\langle1/a\rangle]$, $\mathbf{e}_i=\mathbf{y}-\mathbf{H}\mathbf{x}_{-i}$, $\mathbf{x}_{-i}$ is $\mathbf{x}$ except for the ith entry replaced with 0, and $\mathbf{h}_i$ is the ith column of $\mathbf{H}$. Therefore

$$q(x_i)=q(z_i=0)\delta(x_i)+q(z_i=1)\phi_+(\mu_i,\eta_i), \tag{31}$$

which is a Bernoulli truncated-Gaussian density.

3.3.5. Posterior surrogate for z

For $i=1,\ldots,N$

$$q(z_i=1)=1/[1+C'_i] \quad\text{and}\quad q(z_i=0)=1-q(z_i=1), \tag{32}$$

with $C'_i=\exp(C_i/2\quad\times\tilde{\varsigma}_0/\tilde{\varsigma}_1+\mu_i\tilde{\alpha}_0/\tilde{\alpha}_1+\ln\tilde{\alpha}_1-\psi(\tilde{\alpha}_0)+\psi(\tilde{\beta}_0)-\psi(\tilde{\beta}_1))$. 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 $j=1,\ldots,K$

$$q(c_j)=\phi(\mu_{c_j},\sigma_{c_j}), \tag{33}$$

where $\phi(\mu,\sigma)$ is the probability density function for the normal distribution with the mean μ and variance σ

$$\mu_{c_j}=\frac{\langle\mathbf{x}^T\mathbf{H}^{j^T}\mathbf{y}-\mathbf{x}\mathbf{H}^{j^T}\mathbf{H}^0\mathbf{x}-\sum_{l\neq j}\mathbf{x}^T\mathbf{H}^{j^T}\mathbf{H}^l c_l\mathbf{x}\rangle}{\langle\mathbf{x}^T\mathbf{H}^{j^T}\mathbf{H}^j\mathbf{x}\rangle},$$

and $1/\sigma_{c_j}=\langle1/\sigma^2\rangle\langle\mathbf{x}^T\mathbf{H}^{j^T}\mathbf{H}^j\mathbf{x}\rangle$.

Algorithm 1. VB semi-blind image reconstruction algorithm.

1: % *Initialization*:
2: Initialize estimates $\langle\mathbf{x}^{(0)}\rangle$, $\langle\mathbf{z}^{(0)}\rangle$, and $w^{(0)}$, and set $\mathbf{c}=\mathbf{0}$ to have $\hat{\boldsymbol{\kappa}}^{(0)}=\boldsymbol{\kappa}_0$,
3: % *Iterations*:
4: **for** $t=1,2,\ldots$, **do**
5: Evaluate $\tilde{\alpha}_0^{(t)},\tilde{\alpha}_1^{(t)}$ in (25) by using $\langle\mathbf{x}^{(t-1)}\rangle,\langle\mathbf{z}^{(t-1)}\rangle$,
6: Evaluate $\tilde{\beta}_0^{(t)},\tilde{\beta}_1^{(t)}$ in (26) by using $\langle\mathbf{z}^{(t-1)}\rangle$,
7: Evaluate $\tilde{\varsigma}_0^{(t)},\tilde{\varsigma}_1^{(t)}$ in (27) from $\langle\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2\rangle$,
8: **for** $i=1,2,\ldots,N$ **do**
9: Evaluate necessary statistics (μ_i,η_i) for $q(x_i|z_i=1)$ in (29),
10: Evaluate $q(z_i=1)$ in (32),
11: Evaluate $\langle x_i\rangle,\mathrm{var}[x_i]$,
12: For $l=1,\ldots,K$, evaluate $\mu_{c_l},1/\sigma_{c_l}$ for $q(c_l)$ in (33),
13: **end for**
14: **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 Figs. 2 and 6, respectively. Then the proposed variational algorithm is applied to a tobacco virus MRFM data set. There are many possible approaches for selecting hyperparameters, including the non-informative approach of [9] and the expectation–maximization approach of [12]. In our experiments, hyperparameters ς_0, ς_1, α_0, and α_1 for the densities are chosen based on the framework advocated in [9]. This leads to the vague priors corresponding to selecting small values $\varsigma_0 = \varsigma_1 = \alpha_0 = \alpha_1 = 1$. For w, the noninformative initialization is made by setting $\beta_0 = \beta_1 = 1$, which gives flexibility to the surrogate posterior density for w. The resulting prior Beta distribution for w is a uniform distribution on $[0, 1]$ for the mean proportion of non-zero pixels.

$$w \sim \mathcal{B}(\beta_0, \beta_1) \sim \mathcal{U}([0, 1]). \tag{34}$$

The initial image used to initialize the algorithm is obtained from one Landweber iteration [44].

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. Some quantities of interest, computed from the outputs of the variational algorithm, are depicted as functions of the iteration number in Fig. 3. These plots indicate that

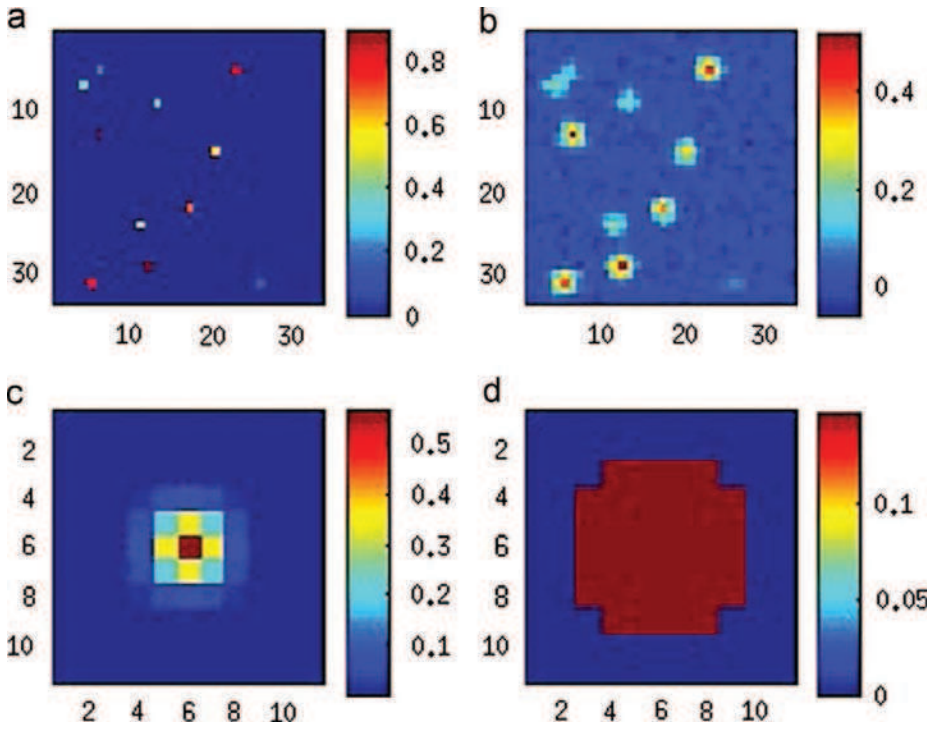

Fig. 2. Experiment with Gaussian PSF: true image (a), observation (b), true PSF (c) and mismatched PSF (κ_0) (d).

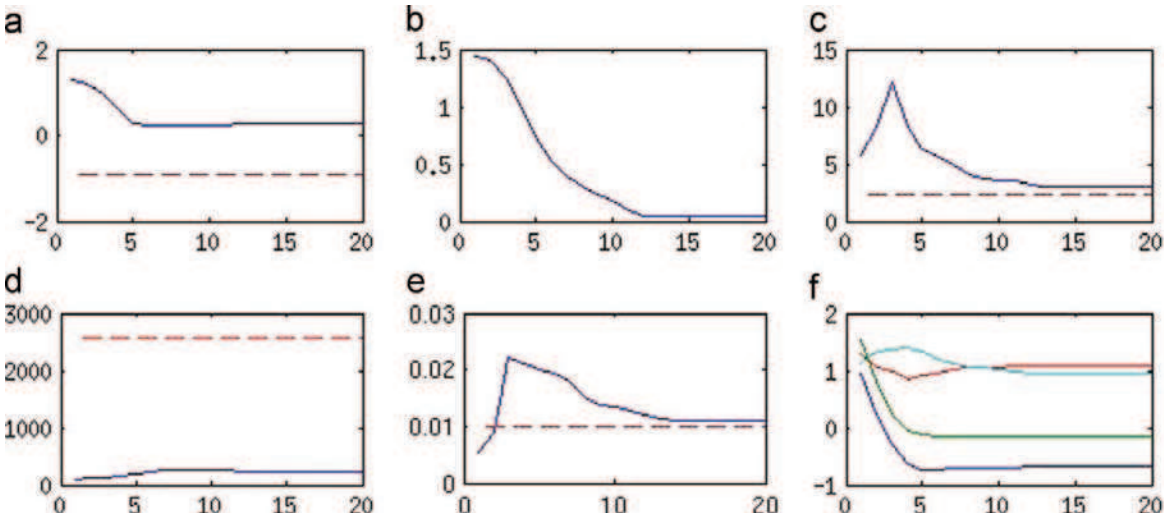

Fig. 3. Result of Algorithm 1: curves of residual, error, $\mathrm{E}[1/a]$, $\mathrm{E}[1/\sigma^2]$, $\mathrm{E}[w]$, $\mathrm{E}[\mathbf{c}]$, as functions of the number of iterations. These curves show how fast the convergence is achieved. (a) $\log\|\mathbf{y}-\mathrm{E}\mathbf{H}\mathrm{E}\mathbf{x}\|^2$ (solid line) and noise level (dashed line), (b) $\log\|\mathbf{x}_{true}-\mathrm{E}\mathbf{x}\|^2$, (c) $\mathrm{E}[1/a]$ (solid line) and true value (dashed line), (d) $\mathrm{E}[1/\sigma^2]$ (solid line) and true value (dashed line), (e) $\mathrm{E}[w]$ (solid line) and true value (dashed line) and (f) $\mathrm{E}[\mathbf{c}]$. Four PSF coefficients.

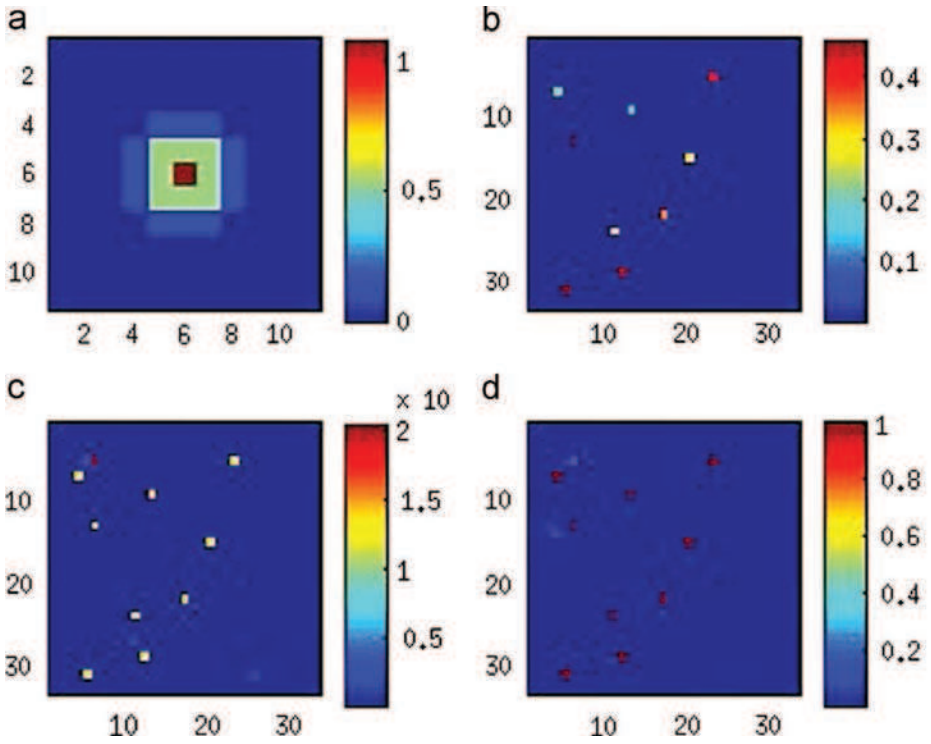

Fig. 4. (a) Restored PSF, (b) image, (c) map of pixel-wise (posterior) variance, and (d) weight map. $\hat{\boldsymbol{\kappa}} = \mathrm{E}\boldsymbol{\kappa}$ is close to the true one. A pixel-wise weight shown in (d) is the posterior probability of the pixel being a nonzero signal.

convergence to the steady state is achieved after few iterations. In Fig. 3, E[w] and E[$1/a$] get close to the true level but E[$1/\sigma^2$] shows a deviation from the true values. This large deviation implies that our estimation of noise level is conservative; the estimated noise level is larger than the true level. This relates to the large deviation in projection error from noise level (Fig. 3(a)). The drastic changes in the initial steps seen in the curves of E[$1/a$], E[w] are due to the imperfect prior knowledge (initialization). The final estimated PSF and reconstructed image are depicted in Fig. 4, along with the reconstructed variances and posterior probability of $z_i \neq 0$. We decomposed the support region of the true PSF to produce orthonormal bases $\{\boldsymbol{\kappa}_i\}_i$ shown in Fig. 5. We extracted 4 bases because these four PSF bases clearly explain the significant part of the true Gaussian PSF. In other words, little energy resides outside of this basis set in PSF space.

The reconstructed PSF clearly matches the true one, as seen in Figs. 2 and 4. Note that the restored image is slightly attenuated while the restored PSF is amplified because of intrinsic scale ambiguity.

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. 6. The PSF models the PSF of the MRFM instrument, derived by Mamin et al. [3]. The convergence of the algorithm is achieved after the 10th iteration. The reconstructed image can be compared to the true image in Fig. 7, where the pixel-wise variances and posterior probability of $z_i \neq 0$ are rendered. The PSF bases are obtained by the procedure proposed in Section 2.2 with the simplified MRFM PSF model and the nominal parameter values [10]. Specifically, by detecting a knee $K=4$ at the scree plot, explaining more than 98.69% of the observed perturbations (Fig. 3 in [9]), we use the first four eigenfunctions, corresponding to the first four largest eigenvalues. The resulting $K=4$ principal basis vectors are depicted in Fig. 8. The reconstructed PSF with the bases clearly matches the true one, as seen in Figs. 6 and 7.

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 Figs. 9 and 10, respectively; the relevant PSFs and observations are presented in Fig. 2 in Section 4.1 and in Fig. 6 in Section 4.2, respectively. Compared with the results of our VB semi-blind algorithm (Algorithm 1), shown in Figs. 4 and 7, the reconstructed images from the mismatched non-blind VB algorithm in Figs. 9 and 10, respectively, inaccurately estimate signal locations and blur most of the non-zero values.

Additional experiments (not shown here) establish that the PSF estimator is very accurate when the algorithm is initialized with the true image.

4.4. Comparison with other algorithms

To quantify the comparison, we performed experiments with the same set of four sparse images and the MRFM type PSFs as used in [9]. 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 $\|\mathbf{x}\|_0 = 6, 11, 18, 30$.

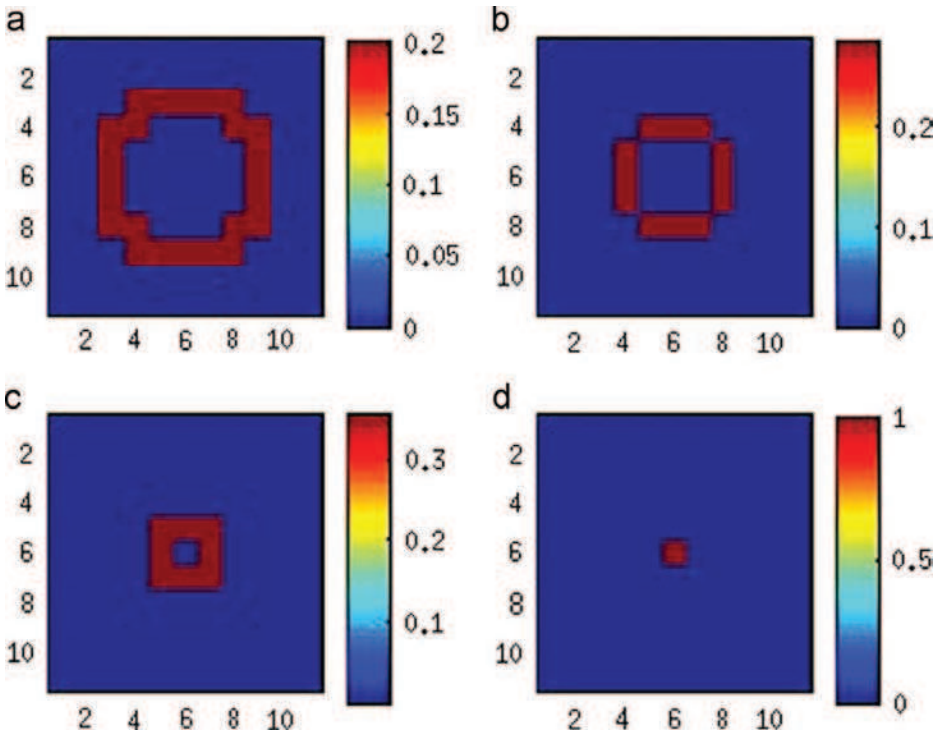

Fig. 5. PSF bases, $\kappa_1, \ldots, \kappa_4$, for Gaussian PSF. (a) The first basis κ_1, (b) the second basis κ_2, (c) the third basis κ_3, (d) the fourth basis κ_4.

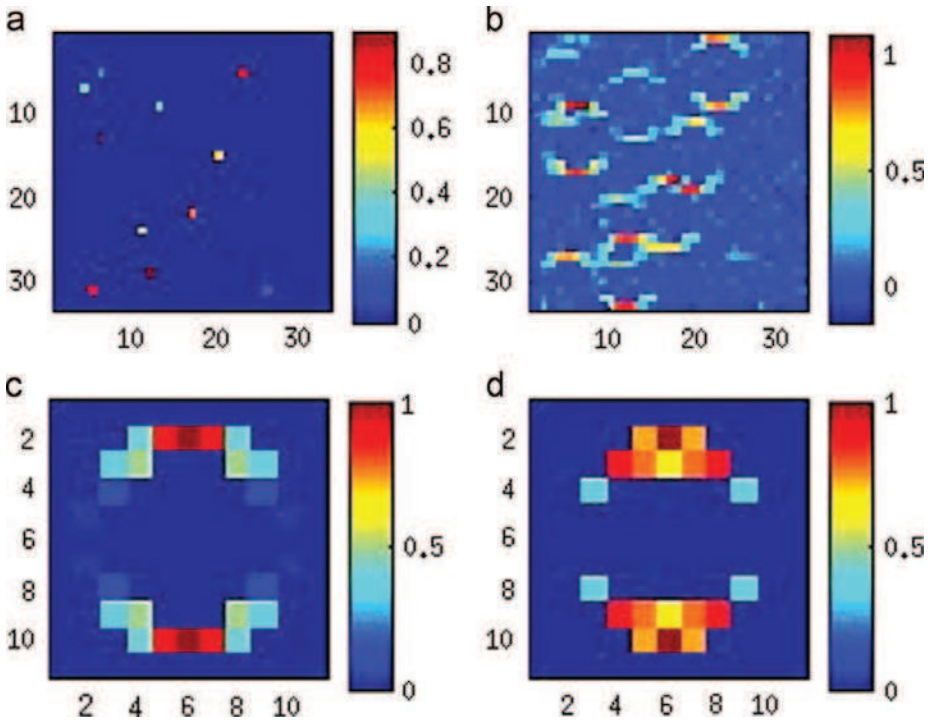

Fig. 6. Experiment with simplified MRFM PSF: true image (a), observation (b), true PSF (c), and mismatched PSF (κ_0) (d).

Under these criteria,[6] Fig. 11 visualizes the 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 [45] and the mismatched non-blind MCMC [43] methods. In terms of PSF estimation, for very sparse images the VB semi-blind method seems to outperform the MCMC method. Also, the proposed VB semi-blind method converges more quickly and requires fewer iterations. For example, the VB

[6] 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; $\mathrm{E}[w] \times N/\|\mathbf{x}\|_0$ for variational method since this method does not produce zero pixels but $\mathrm{E}[w]$.

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.

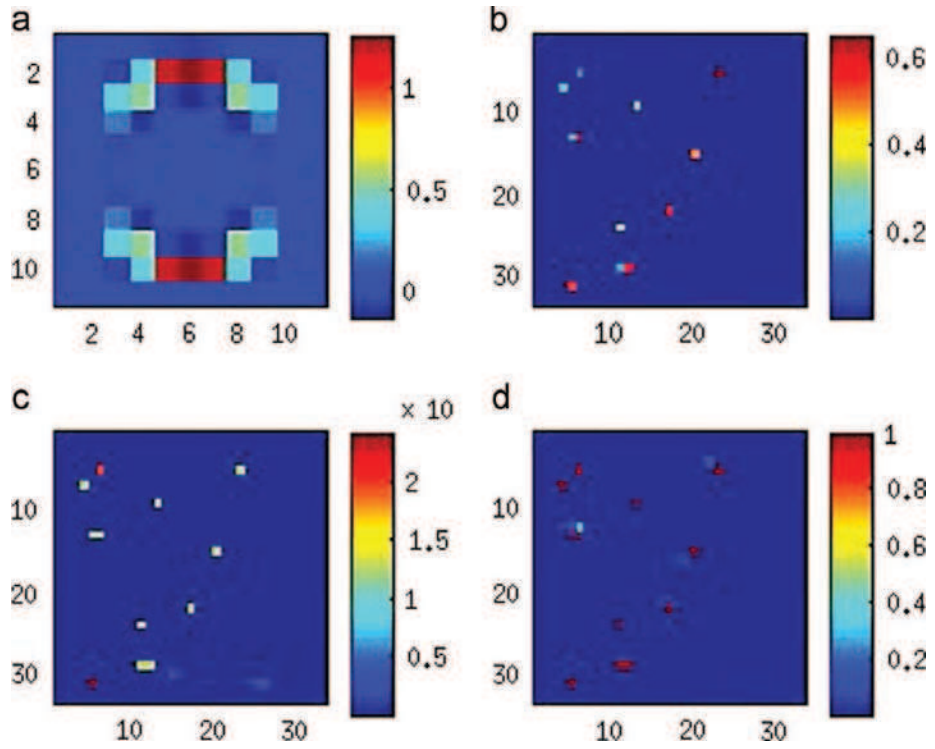

Fig. 7. Restored PSF and image with pixel-wise variance and weight map. $\hat{\kappa} = E\kappa$ is close to the true one: (a) Estimated PSF, (b) estimated image, (c) variance map, and (d) weight map.

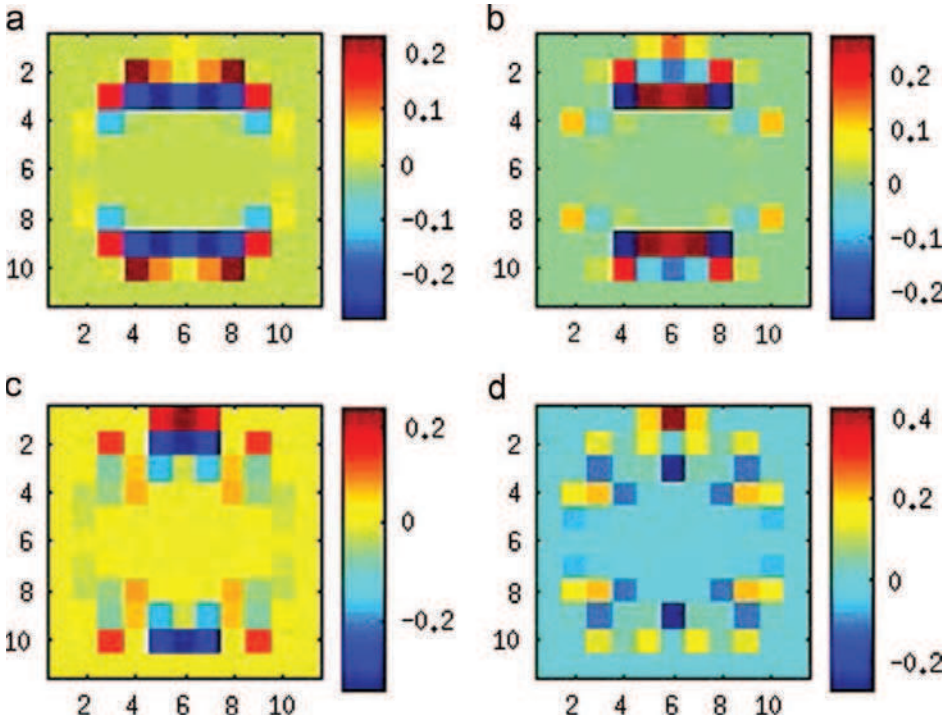

Fig. 8. PSF bases, $\kappa_1, \ldots, \kappa_4$, for MRFM PSF: (a) the first basis κ_1, (b) the second basis κ_2, (c) the third basis κ_3, and (d) the fourth basis κ_4.

semi-blind algorithm converges in approximately 9.6 s after 12 iterations, but the previous MCMC algorithm takes more than 19.2 s after 40 iterations to achieve convergence.[7]

[7] The convergence here is defined as the state where the change in estimation curves over time is negligible.

In addition, we made comparisons between our sparse image reconstruction method and other state-of-the-art blind deconvolution methods [28–33], as shown in our previous work [9]. These algorithms were initialized with the nominal, mismatched PSF and were applied to the same sparse image as our experiment above. For a fair comparison, we made a sparse prior modification in the image model of other algorithms, as

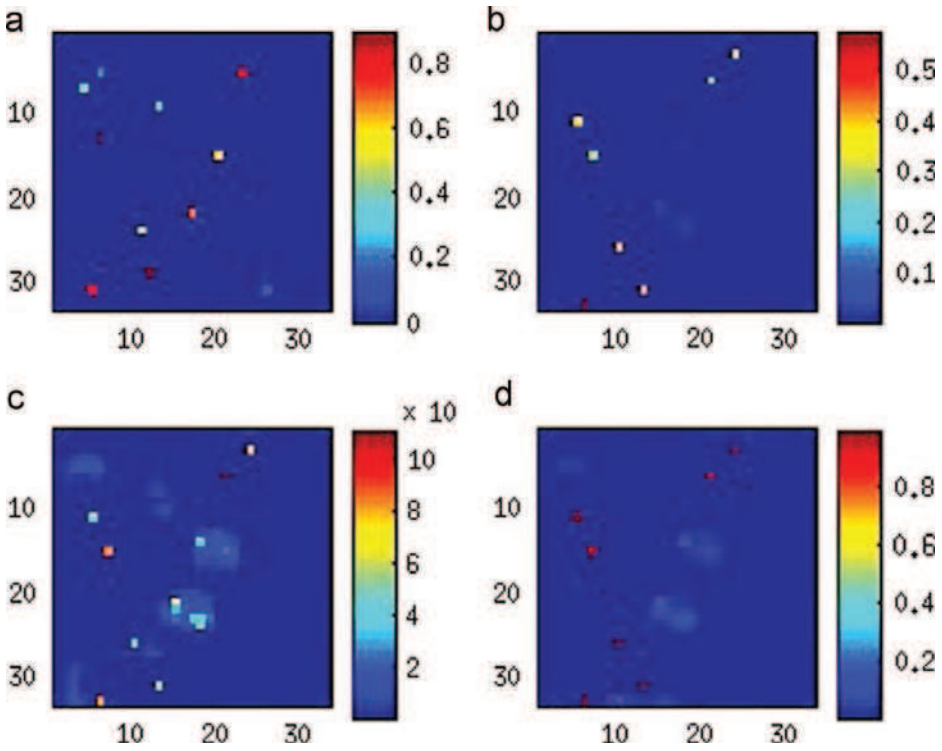

Fig. 9. (Mismatched) non-blind result with a mismatched Gaussian PSF: (a) true image, (b) estimated image, (c) variance map, and (d) weight map.

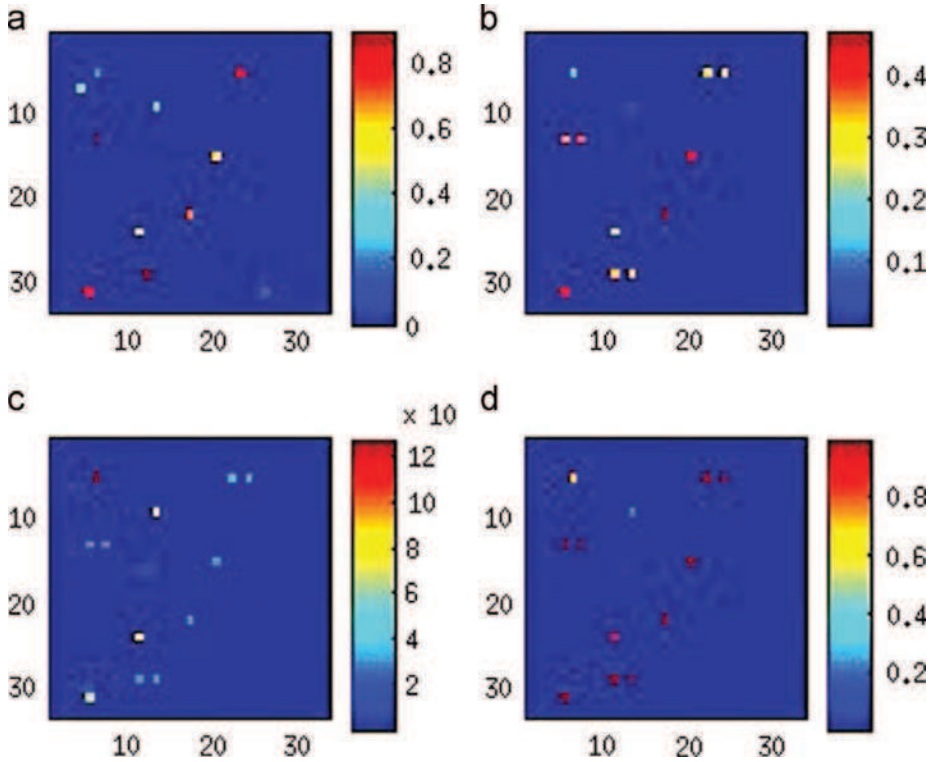

Fig. 10. (Mismatched) non-blind result with a mismatched MRFM type PSF: (a) true image, (b) estimated image, (c) variance map, and (d) weight map.

needed. Most of these methods do not assume or fit into the sparse model in our experiments, thus leading to poor performance in terms of image and PSF estimation errors. Among these tested algorithms, two of them, proposed by Tzikas et al. [29] and Almeida et al. [31], produced non-trivial and convergent solutions and the corresponding results are compared to ours in Fig. 11. By using basis kernels the method proposed by Tzikas et al. [29] uses a similar PSF model to ours. Because a sparse image prior is not assumed in their algorithm [29], we applied their suggested PSF model along with our sparse image prior for a fair comparison. The method proposed by Almeida et al. [31] exploits the sharp edge property in natural images and uses initial,

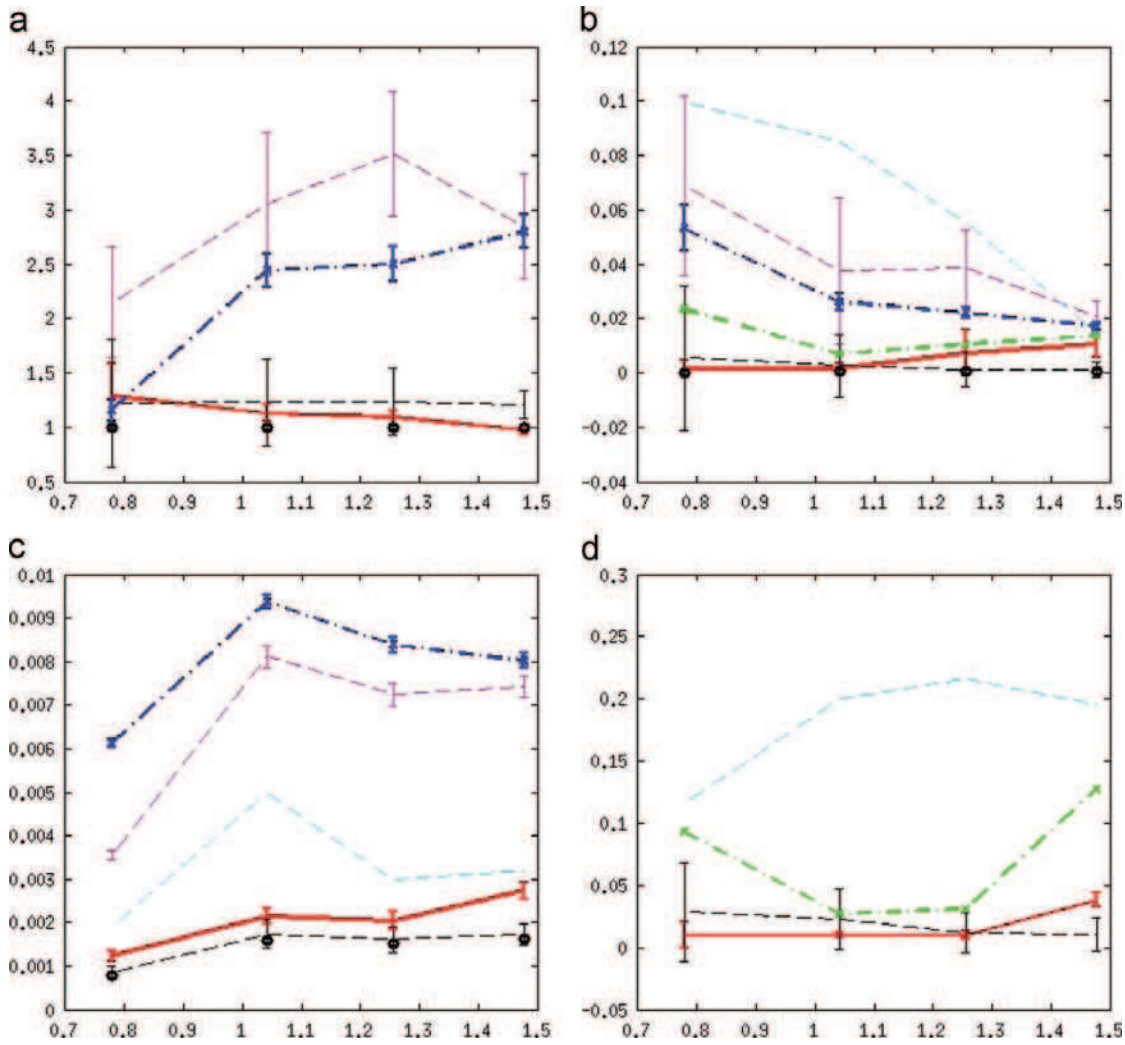

Fig. 11. For various image sparsity levels (x-axis: $\log_{10}\|\mathbf{x}\|_0$), performance of several blind, semi-blind, and non-blind deconvolution algorithms: the proposed method (red), AM (blue), Almeida's method (green), Tzikas's method (cyan), semi-blind MC (black), and mismatched non-blind MC (magenta). Errors are illustrated with standard deviations. (a) Estimated sparsity. Normalized true level is 1 (black circles). (b) Normalized error in reconstructed image. For the lower bound, information about the true PSF is only available to the oracle IST (black circles). (c) Residual (projection) error. The noise level appears in black circles. (d) PSF recovery error, as a performance gauge of our semi-blind method. At the initial stage of the algorithm, $\|\boldsymbol{\kappa}_0/\|\boldsymbol{\kappa}_0\| - \boldsymbol{\kappa}/\|\boldsymbol{\kappa}\|\|_2^2 = 0.5627$. (Some of the sparsity measure and residual errors are too large to be plotted together with results from other algorithms.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

high regularization for effective PSF estimation. Both of these perform worse than our VB method as seen in Fig. 11. The remaining algorithms [28,30,32,33], which focus on photo image reconstruction or motion blur, either produce a trivial solution ($\hat{\mathbf{x}} \approx \mathbf{y}$) or are a special case of Tzikas's model [29].

To show lower bound our myopic reconstruction algorithm, we used the Iterative Shrinkage/Thresholding (IST) algorithm with a true PSF. This algorithm effectively restores sparse images with a sparsity constraint [46]. We demonstrate comparisons of the computation time[8] of our proposed reconstruction algorithm to that of others in Table 1.

[8] Matlab is used under Windows 7 Enterprise and HP-Z200 (Quad 2.66 GHz) platform.

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 [39], shown in the first row in Fig. 12. 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 noise is assumed Gaussian [37,39] and the four PSF bases are obtained by the procedure proposed in Section 2.2 with the physical MRFM PSF model and the nominal parameter values [3]. The reconstruction of the sixth layer is shown in Fig. 12(b), and is consistent with the results obtained by other methods. (see [9,43].) The estimated deviation in PSF is small, as predicted in [9].

While they now exhibit similar smoothness, the VB and MCMC images are still somewhat different since each algorithm follows different iterative trajectories in the

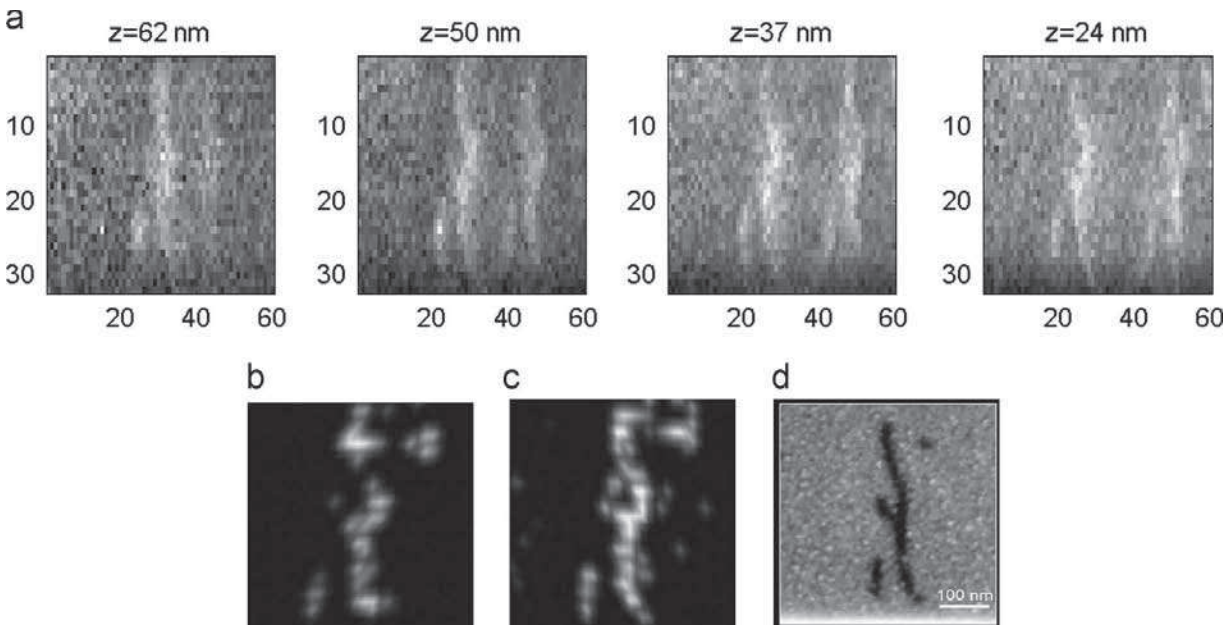

Fig. 12. (a) TMV raw data, (b) estimated virus image by VB, (c) estimated virus image by MCMC [9], and (d) virus image from electron microscope [39].

Table 1
Computation time of algorithms (in seconds), for the data in Fig. 6.

Our method	9.58
Semi-blind MC [9]	19.20
Bayesian non-blind [43]	3.61
AM [45]	0.40
Almeida's method [31]	5.63
Amizic's method [30]	5.69
Tzikas's method [29]	20.31
(oracle) IST [46]	0.09

high-dimensional space of 3D images, thus converging possibly to slightly different stopping points near the maximum of the surrogate distribution. We conclude that the two images from VB and MCMC are comparable in that both represent the 2D SEM image well, but VB is significantly faster.

5. Discussion

5.1. Solving scale ambiguity

In blind deconvolution, joint identifiability is a common issue. For example, because of scale ambiguity, the unicity cannot be guaranteed in a general setting. It is not proven in our solution either. However, the shift/time ambiguity issue noticed in [47] is implicitly addressed in our method using a nominal and basis PSFs. Moreover, our constraint on the PSF space using a basis approach effectively excludes a delta function as a PSF solution, thus avoiding the trivial solution. Secondly, the PSF solution is restricted to this linear spanning space, starting form the initial, mismatched PSF. We can, therefore, reasonably expect that the solution provided by the algorithm is close to the true PSF, away from the trivial solution or the initial PSF.

To resolve scale ambiguity in a MCMC Bayesian framework, stochastic samplers are proposed in [47] by imposing a fixed variance on a certain distribution.[9] Another approach to resolve the scale ambiguity is to assume a hidden scale variable that is multiplied to the PSF and dividing the image (or vice versa), where the scale is drawn along each iteration of the Gibbs sampler [48].

5.2. Exploiting spatial correlations

Our Bayesian hierarchical model (Fig. 1) does not account for possible spatial dependencies that might exist in the image. Spatial dependency can be easily incorporated in the model by adding a spatial latent variable with an associated prior distribution. This can be accomplished, for example, by adding a hidden Markov random field model to the vector $\mathbf{x}$ in Fig. 1. Examples of Markov random field models that have been applied to imaging problems similar to ours are Ising or Potts models [49], Gauss–Markov random fields [50], and Hierarchical Dirichlet processes [51]. Bayesian inference of the hidden parameters of such model is feasible using Monte Carlo and Gibbs sampling, as in [51,52], and using variational Bayes EM [53]. Spatial dependency extensions of our model is a worthwhile and interesting topic for future study but will not be pursued further in this paper.

6. Conclusion

We suggested a novel variational solution to a semi-blind sparse deconvolution problem. Our method uses Bayesian inference for image and PSF restoration with a sparsity-inducing image prior via the variational Bayes

[9] We note that this MCMC method designed for 1D signal deconvolution is not efficient for analyzing 2D and 3D images, since the grouped and marginalized samplers are usually slow to converge requiring hundreds of iterations [47].

approximation. Its power in automatically producing all required parameter values from the data merits further attention for the extraction of image properties and retrieval of necessary features.

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-blind 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 solution [9] are faster convergence and stability of the method.

Acknowledgments

The authors gratefully acknowledge Dr. Dan Rugar for providing the tobacco virus data and his insightful comments on this work.

Appendix A. Useful distributions

A.1. Inverse Gamma distribution

The density of an inverse Gamma random variable $X \sim \mathcal{IG}(a,b)$ is $(b^a/\Gamma(a))x^{-a-1}\exp(-b/x)$, for $x \in (0,\infty)$. $\mathrm{E}X^{-1} = a/b$ and $\mathrm{E}\ln(X) = \ln(b) - \psi(a)$.

A.2. Beta distribution

The density of a Beta random variable $X \sim \mathcal{B}(a,b)$ is $(\Gamma(a)\Gamma(b)/\Gamma(a+b))x^{b-1}(1-x)^{a-1}$, for $x \in (0,1)$, with $\Gamma(c) = \int_0^\infty t^{c-1}e^{-t}\,dt$. The mean of $\mathcal{B}(a,b)$ is $b/(a+b)$ and $\mathrm{E}\ln(\mathcal{B}(a,b)) = \psi(b) - \psi(a+b)$, where ψ is a digamma function.

A.3. Positively truncated Gaussian distribution

The density of a truncated Gaussian random variable x_i is denoted by $x_i \sim \mathcal{N}_+(x_i;\mu,\eta)$, and its statistics used in the paper are

$$\mathrm{E}[x_i | x_i > 0] = \mathrm{E}[\mathcal{N}_+(x_i;\mu,\eta)] = \mu + \sqrt{\eta}\frac{\phi(-\mu/\sqrt{\eta})}{1-\Phi_0(-\mu/\sqrt{\eta})},$$

$$\mathrm{E}[x_i^2 | x_i > 0] = \mathrm{var}[x_i | x_i > 0] + (\mathrm{E}[x_i | x_i > 0])^2 = \eta + \mu(\mathrm{E}[x_i | x_i > 0]),$$

where Φ_0 is a cumulative distribution function for the standard normal distribution.

Appendix B. Derivations of $q(\cdot)$

In this section, we derive the posterior densities defined by variational Bayes framework in Section 3.

B.1. Derivation of $q(\mathbf{c})$

We denote the expected value of the squared residual term by $R = \mathrm{E}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$. For c_l, $l = 1,\ldots,K$

$$\begin{aligned} R &= \mathrm{E}\|\mathbf{y} - \mathbf{H}^0\mathbf{x} - \sum_{l\neq j}\mathbf{H}^l\mathbf{x}c_l - \mathbf{H}^j\mathbf{x}c_j\|^2 \\ &= c_j^2\langle \mathbf{x}^T\mathbf{H}^{jT}\mathbf{H}^j\mathbf{x}\rangle - 2c_j\langle \mathbf{x}^T\mathbf{H}^{jT}\mathbf{y} - \mathbf{x}\mathbf{H}^{jT}\mathbf{H}^0\mathbf{x} \\ &\quad - \sum_{l\neq j}\mathbf{x}^T\mathbf{H}^{jT}\mathbf{H}^l c_l\mathbf{x}\rangle + \mathrm{const}, \end{aligned}$$

where $\mathbf{H}^j$ is the convolution matrix corresponding to the convolution with $\boldsymbol{\kappa}_j$. For $i \neq j$ and $i,j > 0$, $\mathrm{E}(\mathbf{H}^i\mathbf{x})^T(\mathbf{H}^j\mathbf{x}) = \mathrm{tr}(\mathbf{H}^{iT}\mathbf{H}^j(\mathrm{cov}(\mathbf{x}) + \langle\mathbf{x}\rangle\langle\mathbf{x}^T\rangle)) = (\mathbf{H}^i\langle\mathbf{x}\rangle)^T(\mathbf{H}^j\langle\mathbf{x}\rangle)$, since $\mathrm{tr}(\mathbf{H}^{iT}\mathbf{H}^j\mathrm{cov}(\mathbf{x})) = \mathrm{tr}(\mathbf{H}^iD^T\mathbf{H}^jD) = \sum_k d_k^2\mathbf{h}_k^i\mathbf{h}_k^j = 0$. Here, $\mathrm{cov}(\mathbf{x})$ is approximated as a diagonal matrix $D^2 = \mathrm{diag}(d_1^2,\ldots,d_n^2)$. This is reasonable, especially when the expected recovered signal $\hat{\mathbf{x}}$ exhibits high sparsity. Likewise, $\mathrm{E}(\mathbf{H}^0\mathbf{x})^T(\mathbf{H}^j\mathbf{x}) = \boldsymbol{\kappa}_0^T\boldsymbol{\kappa}_j\sum_i\mathrm{var}[x_i] + (\mathbf{H}^0\langle\mathbf{x}\rangle)^T(\mathbf{H}^j\langle\mathbf{x}\rangle)$ and $\mathrm{E}(\mathbf{H}^j\mathbf{x})^T(\mathbf{H}^j\mathbf{x}) = \|\boldsymbol{\kappa}_j\|^2\sum_i\mathrm{var}[x_i] + \|\mathbf{H}^j\langle\mathbf{x}\rangle\|^2$.

Then, we factorize

$$\mathrm{E}\left[-\frac{R}{2\sigma^2}\right] = -\frac{(c_j - \mu_{c_j})^2}{2\sigma_{c_j}},$$

with

$$\mu_{c_j} = \frac{\langle \mathbf{x}^T\mathbf{H}^{jT}\mathbf{y} - \mathbf{x}\mathbf{H}^{jT}\mathbf{H}^0\mathbf{x} - \sum_{l\neq j}\mathbf{x}^T\mathbf{H}^{jT}\mathbf{H}^l c_l\mathbf{x}\rangle}{\langle \mathbf{x}^T\mathbf{H}^{jT}\mathbf{H}^j\mathbf{x}\rangle},$$

$$1/\sigma_{c_j} = \langle 1/\sigma^2\rangle\langle \mathbf{x}^T\mathbf{H}^{jT}\mathbf{H}^j\mathbf{x}\rangle.$$

If we set the prior, $p(c_j)$, to be a uniform distribution over a wide range of the real line that covers error tolerances, we obtain a normally distributed variational density $q(c_j) = \phi(\mu_{c_j},\sigma_{c_j})$ with its mean μ_{c_j} and variance σ_{c_j} defined above, because $\ln q(c_j) = \mathrm{E}[-R/2\sigma^2]$. By the independence assumption, $q(\mathbf{c}) = \prod q(c_j)$, so $q(\mathbf{c})$ can be easily evaluated.

B.2. Derivation of $q(\sigma^2)$

We evaluate R ignoring edge effects; $R = \|\mathbf{y} - \langle\mathbf{H}\rangle\langle\mathbf{x}\rangle\|^2 + \sum\mathrm{var}[x_i][\|\langle\boldsymbol{\kappa}\rangle\|^2 + \sum_l\sigma_{c_l}\|\boldsymbol{\kappa}_l\|^2] + \sum_l\sigma_{c_l}\|\mathbf{H}^l\langle\mathbf{x}\rangle\|^2$. $\|\boldsymbol{\kappa}\|^2$ is a kernel energy in ℓ_2 sense and the variance terms add uncertainty, due to the uncertainty in $\boldsymbol{\kappa}$, to the estimation of density. Applying (19) (ignoring constants)

$$\begin{aligned} \ln q(\sigma^2) &= \mathrm{E}_{\setminus\sigma^2}[\ln p(\mathbf{y}|\mathbf{x},\mathbf{c},\sigma^2)p(\sigma^2)p(\mathbf{x}|a,w)p(w)p(a)] \\ &= \mathrm{E}_{\mathbf{x},\mathbf{c}}[\ln p(\mathbf{y}|\mathbf{x},\sigma^2)] + \ln p(\sigma^2) \\ &= -\frac{\mathrm{E}_{\mathbf{x},\mathbf{c}}[\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2]}{2\sigma^2} - \frac{P}{2}\ln\sigma^2 + \ln p(\sigma^2). \end{aligned}$$

$$\mathcal{IG}(\tilde{\varsigma}_0,\tilde{\varsigma}_1) \triangleq q(\sigma^2) = \mathcal{IG}(P/2 + \varsigma_0, \langle\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2\rangle/2 + \varsigma_1)$$

where $\mathrm{E}_{\setminus\sigma^2}$ denotes expectation with respect to all variables except σ^2.

B.3. Derivation of $q(\mathbf{x})$

For x_i, $i = 1,\ldots,N$, $R = \mathrm{E}\|\mathbf{e}_i - \mathbf{h}_i x_i\|^2$ with $\mathbf{e}_i = \mathbf{y} - \mathbf{H}\mathbf{x}_{-i} = \mathbf{y} - \mathbf{H}^0\mathbf{x}_{-i} - \sum_l\mathbf{H}^l c_l\mathbf{x}_{-i}$,

$\mathbf{h}_i = [\mathbf{H}^0 + \sum \mathbf{H}^l c_l]_i = \mathbf{h}_i^0 + \sum \mathbf{h}_i^l c_l = (i\text{th column of}\mathbf{H})$. Ignoring constants, $R = \langle \|\mathbf{h}_i\|^2 \rangle x_i^2 - 2\langle \mathbf{h}_i^T \mathbf{e}_i \rangle x_i$.

Using the orthogonality of the kernel bases and uncorrelatedness of c_l's, we derive the following terms (necessary to evaluate R): $\langle \|\mathbf{h}_i\|^2 \rangle = \|\mathbf{h}_i^0\|^2 + \sum_l \sigma_{c_l} \|\mathbf{h}_i^l\|^2$ and, $\langle \mathbf{h}_i^T \mathbf{e}_i \rangle = \langle \mathbf{h}_i^T \rangle (\mathbf{y} - \langle \mathbf{H} \rangle \langle \mathbf{x}_{-i} \rangle) - \sum_l \mathrm{var}[c_l] \mathbf{h}_i^{lT} \mathbf{H}^l \langle \mathbf{x}_{-i} \rangle$.

Then, $\mathrm{var}[x_i] = w'_i \mathrm{E}[x_i^2 | x_i > 0] - w_i'^2 (\mathrm{E}[x_i | x_i > 0])^2$, $\mathrm{E}[x_i] = w'_i \mathrm{E}[x_i | x_i > 0]$, where $w'_i = q(z_i = 1)$ is the posterior weight for the normal distribution and $1 - w'_i$ is the weight for the delta function. The required statistics of x_i that are used to derive the distribution above are obtained by applying Appendix A.3.

B.4. Derivation of $q(\mathbf{z})$

To derive $q(z_i = 1) = \langle z_i \rangle$, we evaluate the unnormalized version $\hat{q}(z_i)$ of $q(z_i)$ and normalize it.

$$\ln \hat{q}(z_i = 1) = \mathrm{E}_{\setminus z_i} \left[-\|\mathbf{e}_i - \mathbf{h}_i x_i\|^2 2\sigma^2 - \ln a - \frac{x_i}{a} + \ln w \right]$$

with $x_i \sim N_+(\mu_i, \eta_i)$

and

$$\ln \hat{q}(z_i = 0) = \mathrm{E}_{\setminus z_i} \left[-\frac{\|\mathbf{e}_i\|^2}{2\sigma^2} + \ln(1 - w) \right] \text{ with } x_i = 0.$$

The normalized version of the weight is $q(z_i = 1) = 1/[1 + C'_i]$. $C'_i = \exp(\ln \hat{q}(z_i = 0) - \ln \hat{q}(z_i = 1)) = \exp(C_i/2 \times \langle 1/\sigma^2 \rangle + \mu \langle 1/a \rangle + \langle \ln a \rangle + \langle \ln(1-w) - \ln w \rangle = \exp(C_i/2 \times \tilde{\xi}_0 / \tilde{\xi}_1 + \mu \tilde{\alpha}_0 / \tilde{\alpha}_1 + \ln \tilde{\alpha}_1 - \psi(\tilde{\alpha}_0) + \psi(\tilde{\beta}_0) - \psi(\tilde{\beta}_1))$. ψ is a digamma function and $C_i = \langle \|\mathbf{h}_i\|^2 \rangle (\mu_i^2 + \eta_i) - 2\langle \mathbf{e}_i^T \mathbf{h}_i \rangle \mu_i$.

References

[1] R. Ward, B. Saleh, Deblurring random blur, IEEE Transactions on Acoustics, Speech, Signal Processing 35 (10) (1987) 1494–1498.

[2] D. Kundur, D. Hatzinakos, Blind image deconvolution, IEEE Signal Processing Magazine 13 (3) (1996) 43–64.

[3] J. Mamin, R. Budakian, D. Rugar, Point Response Function of an MRFM Tip, Tech. Rep., IBM Research Division, October 2003.

[4] S. Makni, P. Ciuciu, J. Idier, J.-B. Poline, Joint detection–estimation of brain activity in functional MRI: a multichannel deconvolution solution, IEEE Transactions on Signal Processing 53 (9) (2005) 3488–3502.

[5] G. Pillonetto, C. Cobelli, Identifiability of the stochastic semi-blind deconvolution problem for a class of time-invariant linear systems, Automatica 43 (4) (2007) 647–654.

[6] P. Sarri, G. Thomas, E. Sekko, P. Neveux, Myopic deconvolution combining Kalman filter and tracking control, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal (ICASSP), vol. 3, 1998, pp. 1833–1836.

[7] G. Chenegros, L.M. Mugnier, F. Lacombe, M. Glanc, 3D phase diversity: a myopic deconvolution method for short-exposure images. Application to retinal imaging, Journal of the Optical Society of America 24 (5) (2007) 1349–1357.

[8] S.U. Park, N. Dobigeon, A.O. Hero, Myopic sparse image reconstruction with application to MRFM, in: C.A. Bouman, I. Pollak, P.J. Wolfe (Eds.), Proceedings of the Computational Imaging Conference in IS&T SPIE Symposium on Electronic Imaging Science and Technology, vol. 7873, SPIE, 2011, pp. 787303/1–787303/14.

[9] S.U. Park, N. Dobigeon, A.O. Hero, Semi-blind sparse image reconstruction with application to MRFM, IEEE Transactions on Image Processing 21 (9) (2012) 3838–3849.

[10] M. Ting, R. Raich, A.O. Hero, Sparse image reconstruction for molecular imaging, IEEE Transactions on Image Processing 18 (6) (2009) 1215–1227.

[11] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, NY, USA, 2006.

[12] N. Nasios, A. Bors, Variational learning for Gaussian mixture models, IEEE Transactions on Systems, Man, and Cybernetics, Part B 36 (4) (2006) 849–862.

[13] A. Corduneanu, C.M. Bishop, Variational Bayesian model selection for mixture distributions, in: Proceedings of the Conference on Artificial Intelligence and Statistics, 2001, pp. 27–34.

[14] K. Themelis, A. Rontogiannis, K. Koutroumbas, A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing, IEEE Transactions on Signal Processing 60 (2) (2012) 585–599.

[15] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, P. Ciuciu, A fully Bayesian approach to the parcel-based detection–estimation of brain activity in fMRI, Neuroimage 41 (3) (2008) 941–969.

[16] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, 2nd edition, Springer, 2004.

[17] W.R. Gilks, Markov Chain Monte Carlo in Practice, Chapman and Hall/CRC, 1999.

[18] F. Orieux, J.-F. Giovannelli, T. Rodet, Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution, Journal of the Optical Society of America 27 (7) (2010) 1593–1607.

[19] H. Attias, A variational Bayesian framework for graphical models, in: Proceedings of the Advances in Neural Information Processing Systems (NIPS), MIT Press, 2000, pp. 209–215.

[20] A.M. Walker, On the asymptotic behaviour of posterior distributions, Journal of the Royal Statistical Society: Series B (Methodological) 31 (1) (1969) 80–88.

[21] B. Wang, D. Titterington, Convergence and asymptotic normality of variational Bayesian approximations for exponential family models with missing values, in: Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI), AUAI Press, 2004, pp. 577–584.

[22] B. Wang, M. Titterington, Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model, Bayesian Analysis 1 (3) (2006) 625–650.

[23] C.M. Bishop, J.M. Winn, C.C. Nh, Non-linear Bayesian image modelling, in: Proceedings of the European Conference on Computer Vision (EECV), Springer-Verlag, 2000, pp. 3–17.

[24] Z. Ghahramani, M.J. Beal, Variational inference for Bayesian mixtures of factor analysers, in: Proceedings of the Advances in Neural Information Processing Systems (NIPS), MIT Press, 2000, pp. 449–455.

[25] J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, M. West, The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures, in: Bayesian Statistics, vol. 7, 2003, pp. 453–464.

[26] J. Winn, C.M. Bishop, T. Jaakkola, Variational message passing, Journal of Machine Learning Research 6 (2005) 661–694.

[27] S.U. Park, N. Dobigeon, A.O. Hero, Variational semi-blind sparse image reconstruction with application to MRFM, in: C.A. Bouman, I. Pollak, P.J. Wolfe (Eds.), Proceedings of the Computational Imaging Conference in IS&T SPIE Symposium on Electronic Imaging Science and Technology, vol. 8296, SPIE, 2012, pp. 82960G–82960G-11.

[28] S. Babacan, R. Molina, A. Katsaggelos, Variational Bayesian blind deconvolution using a total variation prior, IEEE Transactions on Image Processing 18 (1) (2009) 12–26.

[29] D. Tzikas, A. Likas, N. Galatsanos, Variational Bayesian sparse kernel-based blind image deconvolution with Student's-t priors, IEEE Transactions on Image Processing 18 (4) (2009) 753–764.

[30] B. Amizic, S.D. Babacan, R. Molina, A.K. Katsaggelos, Sparse Bayesian blind image deconvolution with parameter estimation, in: Proceedings of the European Signal Processing Conference (EUSIPCO), Aalborg (Denmark), 2010, pp. 626–630.

[31] M. Almeida, L. Almeida, Blind and semi-blind deblurring of natural images, IEEE Transactions on Image Processing 19 (1) (2010) 36–52.

[32] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, W.T. Freeman, Removing camera shake from a single photograph, in: ACM SIGGRAPH 2006 Papers, SIGGRAPH '06, ACM, New York, NY, USA, 2006, pp. 787–794.

[33] Q. Shan, J. Jia, A. Agarwala, High-quality motion deblurring from a single image, ACM Transactions on Graphics 27 (3) (2008) 1.

[34] J.A. Sidles, Noninductive detection of single-proton magnetic resonance, Applied Physics Letters 58 (24) (1991) 2854–2856.

[35] J.A. Sidles, Folded Stern–Gerlach experiment as a means for detecting nuclear magnetic resonance in individual nuclei, Physical Review Letters 68 (8) (1992) 1124–1127.

[36] J.A. Sidles, J.L. Garbini, K.J. Bruland, D. Rugar, O. Züger, S. Hoen, C. S. Yannoni, Magnetic resonance force microscopy, Reviews of Modern Physics 67 (1) (1995) 249–265.

[37] D. Rugar, C.S. Yannoni, J.A. Sidles, Mechanical detection of magnetic resonance, Nature 360 (6404) (1992) 563–566.

[38] O. Züger, S.T. Hoen, C.S. Yannoni, D. Rugar, Three-dimensional imaging with a nuclear magnetic resonance force microscope, Journal of Applied Physics 79 (4) (1996) 1881–1884.
[39] C.L. Degen, M. Poggio, H.J. Mamin, C.T. Rettner, D. Rugar, Nanoscale magnetic resonance imaging, Proceedings of the National Academy of Sciences 106 (5) (2009) 1313–1317.
[40] O. Züger, D. Rugar, First images from a magnetic resonance force microscope, Applied Physics Letters 63 (18) (1993) 2496–2498.
[41] O. Züger, D. Rugar, Magnetic resonance detection and imaging using force microscope techniques, Journal of Applied Physics 75 (10) (1994) 6211–6216.
[42] S. Chao, W.M. Dougherty, J.L. Garbini, J.A. Sidles, Nanometer-scale magnetic resonance imaging, Review of Scientific Instruments 75 (5) (2004) 1175–1181.
[43] N. Dobigeon, A.O. Hero, J.-Y. Tourneret, Hierarchical Bayesian sparse image reconstruction with application to MRFM, IEEE Transactions on Image Processing 18 (9) (2009) 2059–2070.
[44] L. Landweber, An iteration formula for Fredholm integral equations of the first kind, American Journal of Mathematics 73 (3) (1951) 615–624.
[45] K. Herrity, R. Raich, A.O. Hero, Blind deconvolution for sparse molecular imaging, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, USA, 2008, pp. 545–548.
[46] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on Pure and Applied Mathematics 57 (11) (2004) 1413–1457.
[47] D. Ge, J. Idier, E.L. Carpentier, Enhanced sampling schemes for MCMC based blind Bernoulli–Gaussian deconvolution, Signal Processing 91 (4) (2011) 759–772.
[48] T. Vincent, L. Risser, P. Ciuciu, Spatially adaptive mixture modeling for analysis of fMRI time series, IEEE Transactions on Medical Imaging 29 (4) (2010) 1059–1074.
[49] J. Besag, P.J. Green, Spatial statistics and Bayesian computation, Journal of the Royal Statistical Society: Series B (Methodological) 55 (1) (1993) 25–37.
[50] M.A.T. Figueiredo, J.M.N. Leitao, Unsupervised image restoration and edge location using compound Gauss–Markov random fields and the MDL principle, IEEE Transactions on Image Processing 6 (1997) 1089–1102.
[51] R. Mittelman, N. Dobigeon, A. Hero, Hyperspectral image unmixing using a multiresolution sticky HDP, IEEE Transactions on Signal Processing 60 (4) (2012) 1656–1671.
[52] F. Forbes, G. Fort, Combining Monte Carlo and mean-field-like methods for inference in hidden Markov random fields, IEEE Transactions on Image Processing 16 (3) (2007) 824–837.
[53] F. Forbes, N. Peyrard, Hidden Markov random field model selection criteria based on mean field-like approximations, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (9) (2003) 1089–1101.