Bayesian Spatially Varying Multi-Regularization Image Deblurring

Jordan Pillow[1], Matthias Morzfeld[2], Marylesa Howard[1], Jesse Adams[1] and Matthew Kupinski[3]

[1]Mission Support and Test Services, LLC, Las Vegas, NV, USA
[2]Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, CA, USA
[3]Department of Optical Sciences, The University of Arizona, Tucson, AZ, USA

December 15, 2021

Abstract

Many scientific experiments such as those found in astronomy, geology, microbiology, and X-ray radiography require the use of high-energy instruments to capture images. Since blur and noise are inevitably present in any imaging system, the images must be "deblurred" to extract the full information content. Mathematically, image deblurring is an ill-posed inverse problem that requires regularization. The regularization, in turn, has a large effect on the deblurred image: different regularization strengths, and types, lead to drastically different reconstructions. Moreover, many images contain a mixture of smooth and sharp features which suggests the use of multi-regularization, i.e., varying the type of regularization (e.g. Tikhonov or total variation) across the image. We address these issues by formulating the image deblurring problem within a hierarchical Bayesian framework in which we spatially adapt the strength of the regularization and also vary the regularization type across the image. In this way, the image itself, along with corresponding regularization strength at each pixel, are described jointly by a posterior distribution which we can sample by Markov chain Monte Carlo (MCMC) methods. We illustrate our techniques on simplified test problems and apply them to high-energy X-ray images taken at the Nevada National Security Site. Numerical tests show that our new method is robustly applicable and increases the quality of the image reconstruction when compared to other (Bayesian) methods.

1 Introduction

Deblurring images is necessary in many applications, including astronomy [1], geology [2, 3], microscopy [4], and X-ray radiography [5, 6]. Depending on the application, blurred images can result from atmospheric turbulence, radiation, the object of interest being in motion, or the imaging system itself [7–10]. Deblurring images is challenging for two reasons: (i) typical reconstruction methods require tuning of regularization parameters; and (ii) the images are likely to be high-resolution, making the computations costly due to large dataset sizes. In this work, we are interested in deblurring as it applies to images captured by high-energy X-ray systems, such as *Cygnus*, located at the Nevada National Security Site (NNSS) [11]. Cygnus is used to observe dynamic subcritical experiments at a high resolution. In [12], the author addresses the second issue of deblurring high-resolution images by using a scalable blocking scheme to deblur Cygnus images in a Bayesian matrix-free framework. However, the first issue of tuning regularization parameters remains. In this work, we focus on the first issue and develop a new Bayesian method that features flexible and adaptive regularization.

In brief, the Bayesian sampler we create varies the regularization strength and regularization type throughout the image – which enables vastly superior reconstructions for images of mixed scenes with sharp and smooth objects. We focus on the conceptual feasibility and applicability of our ideas and leave computational aspects, addressing difficulty (ii), for the future. Scaling to images with dimensions 4K×4K pixels, for example, or larger, can be constructed with a "matrix-free" implementation to avoid computational bottlenecks in high-resolution imaging, see, e.g., [13].

More specifically, the technique we created spatially adapts the regularization strengths to the image and utilizes different regularization types in different parts of the image, to accurately deblur mixed scenes with smooth and sharp features. Here, we make use of analogies between regularization and prior distributions in a Bayesian framework to let the image itself determine, or adapt, the regularization strength. This idea is not new and is known as hierarchical sampling: the Bayesian posterior distribution is defined jointly over the image itself and the parameters that define the regularization, or prior. For further explanations and applications of hierarchical sampling in imaging, see, e.g., [5, 6, 14].

One of our contributions is to extend the Bayesian hierarchical framework to allow varied regularization strengths at different features within the image. This is similar in spirit to [15], where conditional Gaussian distributions with spatially varying widths are used within iterative solvers, rather than hierarchical Bayesian sampling. Instead, we consider even more flexible regularization by extending the hierarchical framework to allow for different regularization *types* (Tikhonov or total variation) at different regions within the image. This mixed-regularization is applied with the regularization *type* being predetermined at each location within the image, and with regularization *strength* being spatially varying and determined by the Bayesian method. We test all our ideas and algorithms on synthetic test problems and on an X-ray sub-image taken with Cygnus. All numerical experiments indicate that the technique we created is robustly applicable, efficient in sampling the constructed posterior distribution, and leads to reconstructions that are superior (qualitatively and quantitatively) to reconstructions obtained from other (Bayesian) techniques.

Finally, we note that the hierarchical Bayesian framework, in principle, allows for an uncertainty quantification (UQ) by viewing the hierarchical Bayesian posterior distribution, not a single reconstruction, as the solution of the deblurring problem. The UQ, however, is problematic because it is sensitive to the choice in a layer of hyperparameters that define the hyperpriors for the regularization parameters. While UQ in imaging is a timely and important issue, the conceptual and numerical framework we describe here is problematic for physically-appropriate UQ due to the observed sensitivities in the hyperparameters. Future work should address these issues and make reliable UQ feasible in image deblurring.

The rest of this paper is organized as follows. In Section 2, we describe the necessary background for image deblurring and the hierarchical Bayesian approach. In Section 3, we describe how to extend the hierarchical Bayesian approach to spatially varying regularization strengths, and noise on the image, at different locations within the image. We illustrate our ideas with a simple and synthetic numerical example. Section 4 contains our main contribution and describes how to extend the hierarchical Bayesian approach to account a spatial mix of regularization types within an image, specifically Tikhonov and total variation. We again use a simple synthetic example to illustrate the main ideas. In Section 5 we apply our techniques to synthetic images and radiographs obtained from Cygnus. A summary and conclusions are contained in Section 6.

2 Background

In this section we describe the deblurring problem, the hierarchical Bayesian framework, and the Gibbs sampler. We also set up the notation used throughout this paper. For further details on the Bayesian approach to image deblurring, see, e.g., [14, 16].

2.1 Bayesian Formulation of the Image Deblurring Problem

Image *blurring* can be represented by the linear model

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\mathbf{x} \in \mathbb{R}^{N\times 1}$ is the true image, $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the blurring matrix, $\boldsymbol{\varepsilon} \in \mathbb{R}^{N\times 1}$ is a random variable, representing noise, and where $\mathbf{b} \in \mathbb{R}^{N\times 1}$ is the blurred image we are given. Typically, the matrix $\mathbf{A}$ and the images $\mathbf{x}$ and $\mathbf{b}$ are obtained via column stacking an image with $m \times n$ pixels, in which case $N = mn$. The goal of image *deblurring* is to compute $\mathbf{x}$ given $\mathbf{b}$ and $\mathbf{A}$, along with assumptions about the statistics of the error $\boldsymbol{\varepsilon}$, see, e.g., [9, 14, 17]. Throughout this paper, we assume that $\boldsymbol{\varepsilon}$ is Gaussian, and we write Gaussians as $\mathcal{N}(\mathbf{m}, \mathbf{C})$, where $\mathbf{m}$ is the mean and $\mathbf{C}$ is the covariance matrix.

For $\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\mathbf{0}, \lambda^{-1}\mathbf{I}\right)$, where $\lambda > 0$ is given and $\mathbf{I}$ is the identity matrix, (2.1) implies that $\mathbf{b}|\mathbf{x} \sim \mathcal{N}\left(\mathbf{A}\mathbf{x}, \lambda^{-1}\mathbf{I}\right)$, so that the likelihood is given by

$$p_{\mathrm{LH}}\left(\mathbf{b}|\mathbf{x}\right) \propto \exp\left(-\frac{\lambda}{2}\left|\left|\mathbf{A}\mathbf{x}-\mathbf{b}\right|\right|_2^2\right), \tag{2.2}$$

where $||\cdot||$ is the two-norm ($||\mathbf{x}||_2 = \sqrt{\mathbf{x}^T\mathbf{x}}$, where superscript T denotes a transpose). In this paper, we consider a Gaussian prior, in particular:

$$p_0(\mathbf{x}) \propto \exp\left(-\frac{\delta}{2}\left|\left|\mathbf{L}\mathbf{x}\right|\right|_2^2\right), \tag{2.3}$$

where $\delta > 0$ is given and $\mathbf{L}$ is a given matrix (details on choosing $\mathbf{L}$ are located in section 2.2). A posterior distribution now follows from Bayes' rule:

$$p(\mathbf{x}|\mathbf{b}) \propto p_{\mathrm{LH}}(\mathbf{b}|\mathbf{x})p_0(\mathbf{x}) \propto \exp\left(-\frac{\lambda}{2}\left|\left|\mathbf{A}\mathbf{x}-\mathbf{b}\right|\right|_2^2 - \frac{\delta}{2}\left|\left|\mathbf{L}\mathbf{x}\right|\right|_2^2\right). \tag{2.4}$$

The posterior is a distribution of image reconstructions $\mathbf{x}$, and attached to each $\mathbf{x}$ is a probability. The assigned probability is based on how well the reconstruction $\mathbf{x}$ relates to the given data, $\mathbf{b}$, and prior assumptions.

So far, we assumed the hyperparameters λ and δ are known, but this is rarely the case. We thus treat λ and δ as unknowns and define the hierarchical posterior distribution as

$$p(\mathbf{x}, \lambda, \delta|\mathbf{b}) \propto p_{\mathrm{LH}}(\mathbf{b}|\mathbf{x}, \lambda)p_0(\mathbf{x}|\delta)p_{\mathrm{H1}}(\lambda)p_{\mathrm{H2}}(\delta), \tag{2.5}$$

where p_{H1} and p_{H2} are hyperprior distributions for λ and δ, and where both $p_0(\mathbf{x}|\delta)$ and the likelihood, $p_{\mathrm{LH}}(\mathbf{b}|\mathbf{x}, \lambda)$ are as before. Note that the hierarchical posterior assigns probabilities to image reconstructions $\mathbf{x}$ *and* parameters λ and δ, in view of the blurred image $\mathbf{b}$. For numerical convenience [14], we choose Gamma distributions for the priors

$$p_{\mathrm{H1}}(\lambda) \propto \lambda^{\alpha_\lambda - 1}\exp(-\beta_\lambda\lambda), \quad p_{\mathrm{H2}}(\delta) \propto \delta^{\alpha_\delta - 1}\exp(-\beta_\delta\delta), \tag{2.6}$$

where $\alpha_\lambda, \alpha_\delta$ and $\beta_\lambda, \beta_\delta$ are shape and scale parameters, respectively. With these choices, the posterior distribution becomes

$$p(\mathbf{x}, \lambda, \delta|\mathbf{b}) \propto \lambda^{N/2+\alpha_\lambda-1}\delta^{N/2+\alpha_\delta-1}\exp\left(-\frac{\lambda}{2}\left|\left|\mathbf{A}\mathbf{x}-\mathbf{b}\right|\right|_2^2 - \frac{\delta}{2}\left|\left|\mathbf{L}\mathbf{x}\right|\right|_2^2 - \beta_\lambda\lambda - \beta_\delta\delta\right). \tag{2.7}$$

Throughout this paper, we choose $\alpha_\lambda = \alpha_\delta = 1$ and $\beta_\lambda = \beta_\delta = 10^{-4}$. The rationale for these choices is that the Gamma distribution, which we write as $\Gamma(\alpha, \beta) \propto s^{\alpha-1}\exp(-\beta s)$, behaves like a uniform distribution for small β and order one α, thus imposing only minor prior restrictions on λ and δ (see [14] for more details). The rest of this paper focuses on hierarchical Bayesian deblurring problems and their extensions.

2.2 Tikhonov and TV Regularization

We note that having a Gaussian prior in the Bayesian formulation is equivalent to enforcing Tikhonov regularization in deterministic inversion/deblurring. Using our notation, the Tikhonov-regularized solution would be

$$\mathbf{x}_{\mathrm{TIK}} := \arg\min_{\mathbf{x}}\left(\left|\left|\mathbf{A}\mathbf{x}-\mathbf{b}\right|\right|_2^2 + \alpha\left|\left|\mathbf{L}\mathbf{x}\right|\right|_2^2\right), \tag{2.8}$$

where $\alpha > 0$ is a regularization parameter and $\mathbf{L}$ is a positive semidefinite matrix, commonly chosen to be the identity matrix or powers of the discretized first/second derivative. Tikhonov regularization enforces smoothing, and is therefore appropriate only if all features in the image are smooth. If the important image features are sharp edges, then total variation (TV) regularization can be used [18]. TV regularization is

given by

$$\mathbf{x}_{\text{TV}} := \arg\min_{\mathbf{x}} \left(||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \alpha \, ||\mathbf{D}\mathbf{x}||_1 \right), \tag{2.9}$$

where $\alpha > 0$ and $\mathbf{D}$ is the discretized first derivative. Due to the 1-norm, the Bayesian interpretation of TV regularization yields a non-Gaussian prior. However, we can build a Gaussian prior that approximates the TV prior by following the methods used in [17] when constructing the Lagged Diffusivity Fixed Point Iteration algorithm. The Gaussian approximated TV prior is defined as

$$p_0(\mathbf{x}) \propto \exp\left(-\frac{\delta}{2} \, ||\mathbf{L}_{\text{TV}}\mathbf{x}||_2^2 \right), \tag{2.10}$$

where $\mathbf{L}_{\text{TV}} := \Psi(\mathbf{x})\mathbf{D}$, with $\Psi(\mathbf{x}) := \text{diag}\left(\left((\mathbf{D}\mathbf{x})^2 + \eta \right)^{-1/4} \right)$ where $0 < \eta << 1$. We refer the reader to [5, 6, 14, 16] for more information on the construction of $\mathbf{L}_{\text{TV}}$. We use the TV-approximated Gaussian prior in (2.10) throughout this work, and we label the algorithms and results as TV. Strictly speaking, however, these reconstructions are approximations of "true" TV regularized solutions.

2.3 Numerical Solution and Gibbs Sampling

The solution of a Bayesian deblurring problem is the posterior distribution. A numerical solution of the problem thus amounts to computing an approximation of this posterior distribution which we do via Markov chain Monte Carlo (MCMC), specifically, Gibbs sampling [19]. The basic idea of Gibbs sampling is to use conditional distributions to propose samples within the Markov chain. For the hierarchical deblurring problem, we chose the conditionals

$$\lambda|\mathbf{b}, \mathbf{x}, \delta \sim \Gamma\left(\frac{N}{2} + \alpha_\lambda, \frac{1}{2} \, ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \beta_\lambda \right), \tag{2.11}$$

$$\delta|\mathbf{b}, \mathbf{x}, \lambda \sim \Gamma\left(\frac{N}{2} + \alpha_\delta, \frac{1}{2} \, ||\mathbf{L}\mathbf{x}||_2^2 + \beta_\delta \right), \tag{2.12}$$

$$\mathbf{x}|\mathbf{b}, \lambda, \delta \sim \mathcal{N}\left(\lambda \left(\lambda \mathbf{A}^T\mathbf{A} + \delta \mathbf{L}^T\mathbf{L} \right)^{-1} \mathbf{A}^T\mathbf{b}, \left(\lambda \mathbf{A}^T\mathbf{A} + \delta \mathbf{L}^T\mathbf{L} \right)^{-1} \right). \tag{2.13}$$

Note that all conditionals are standard distributions and, therefore, relatively easy to sample.

Specifically, given an initial guess $\mathbf{x}^{(0)}$, we first sample $\lambda^{(1)} \sim p\left(\lambda|\mathbf{b}, \mathbf{x}^{(0)}\right)$ by drawing from a gamma distribution. We then proceed and sample $\delta^{(1)} \sim p\left(\delta|\mathbf{b}, \mathbf{x}^{(0)}\right)$, again by drawing from a gamma distribution. Finally, we sample $\mathbf{x}^{(1)} \sim p\left(\mathbf{x}|\mathbf{b}, \lambda^{(1)}, \delta^{(1)}\right)$, by drawing from a Gaussian. This procedure is then repeated for K iterations to produce $\lambda^1, \ldots, \lambda^K$, $\delta^1, \ldots, \delta^K$ and $\mathbf{x}^1, \ldots, \mathbf{x}^K$. This set of K samples, with each sample consisting of a λ, a δ and an $\mathbf{x}$, is the numerical solution of the deblurring problem. Note that the order in which we sample from the conditionals does not matter [14, 20, 21], i.e., we obtain similar results by initializing $\lambda^{(0)}$ and $\delta^{(0)}$, and then sampling $\mathbf{x}^{(k)}$, or when switching the order of λ and δ. These results, however, are asymptotic (large sample limit with $k \to \infty$) and the numerical efficiency of the sampler does depend on the order, or "blocking scheme", used [13, 22].

From a practical and computational perspective, one should discard the first few iterations, the burn-in period, because the sampler requires a few steps before reaching a statistically stationary equilibrium state. Moreover, the samples are correlated due to the Markov property of the Gibbs sampler – the current sample depends on the previous sample. The correlation between samples has a large effect on the efficiency of the Gibbs sampler. Specifically, the correlation of the samples causes the statistical error to be a factor of τ_{int} larger than that of independent sampling [21], and the factor τ_{int} is the integrated autocorrelation time (IACT). This suggests the definition of an effective number of samples:

$$K_{\text{eff}} = \frac{K}{\tau_{\text{int}}}. \tag{2.14}$$

In other words, the closer τ_{int} is to 1, the more efficient the MCMC method. In our numerical tests below, we will report the IACT (estimated as in [12]) associated with the sampler as measure of how effective the

sampling scheme is and also report how we defined the burn-in period.

3 Spatially Varying Hyperparameters

The quality of image reconstructions depends on the values of the hyperparameters, which define the anticipated noise level in the image and the strength of the regularization – λ and δ in (2.5). Finding workable and useful hyperparameter values has proven difficult in applications. The hierarchical Bayesian approach described above determines the hyperparameters adaptively from the blurred image, and other methods have been developed with similar goals in mind [23]. It is, however, possible that none of the existing methods yield "useful" hyperparameter combinations, because such a combination does not exist globally. For example, some features in the image may require more regularization than others and the noise level is also rarely constant throughout an image [13, 24]. It is thus natural to consider an extended framework, in which the hyperparameters are not scalars, but are actually vector values and varying across the image. We now describe an extension of the hierarchical framework that allows for different values of regularization strength (δ) and noise level (λ) at each pixel in the image. To keep things simple computationally, we implement this extension via independent priors for the various hyperparameters.

Specifically, let λ_i and δ_i, $i = 1, \ldots, N$, be defined at each pixel of the image and assume independent gamma distributions as priors:

$$p(\lambda_i) \sim \text{Gamma}(\alpha_\lambda, \beta_\lambda) \propto \lambda_i^{\alpha_\lambda - 1} \exp(-\beta_\lambda \lambda_i), \tag{3.1}$$

$$p(\delta_i) \sim \text{Gamma}(\alpha_\delta, \beta_\delta) \propto \delta_i^{\alpha_\delta - 1} \exp(-\beta_\delta \delta_i). \tag{3.2}$$

Here, $\alpha_\lambda, \alpha_\delta$ are Gamma shape parameters and $\beta_\lambda, \beta_\delta$ are Gamma rate parameters and for simplicity we chose the same shape/rate parameters at each pixel. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)$ be N-dimensional vectors whose elements λ_i and δ_i at each pixel. Because of independence of λ_i and δ_i, we have

$$p(\boldsymbol{\lambda}) = \prod_{i=1}^{N} p(\lambda_i) \ , \qquad p(\boldsymbol{\delta}) = \prod_{i=1}^{N} p(\delta_i) \ , \qquad p(\boldsymbol{\lambda}, \boldsymbol{\delta}) = p(\boldsymbol{\lambda}) p(\boldsymbol{\delta}) \ . \tag{3.3}$$

Thus, the likelihood and prior distributions are

$$p(\mathbf{b}|\mathbf{x}, \boldsymbol{\lambda}) \propto \left(\prod_{i=1}^{N} \lambda_i \right)^{1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \lambda_i \left[(\mathbf{A}\mathbf{x})_i - b_i \right]^2 \right), \tag{3.4}$$

$$p(\mathbf{x}|\boldsymbol{\delta}) \propto \left(\prod_{i=1}^{N} \delta_i \right)^{1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \delta_i \left(\mathbf{L}\mathbf{x} \right)_i^2 \right). \tag{3.5}$$

Combining the likelihood (3.4), the prior (3.5), and the hyperprior (3.3), we can write the posterior distribution as

$$\begin{aligned} p(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\delta}|\mathbf{b}) &\propto p(\mathbf{b}|\mathbf{x}, \boldsymbol{\lambda}) p(\mathbf{x}|\boldsymbol{\delta}) p(\boldsymbol{\lambda}) p(\boldsymbol{\delta}), \\ &\propto \prod_{i=1}^{N} \lambda_i^{\alpha_\lambda - 1/2} \delta_i^{\alpha_\delta - 1/2} \\ &\quad \times \exp\left(-\sum_{i=1}^{N} \lambda_i \left(\frac{1}{2} \left[(\mathbf{A}\mathbf{x})_i - b_i \right]^2 + \beta_\lambda \right) - \sum_{i=1}^{N} \delta_i \left(\frac{1}{2} \left(\mathbf{L}\mathbf{x} \right)_i^2 + \beta_\delta \right) \right). \end{aligned} \tag{3.6}$$

Our goal is now to create a MCMC sampler that can draw samples from this posterior distribution.

Algorithm 1: HGSV Sampler

input: Initial approximation $\mathbf{x}^{(0)}$; maximum number of iterations K; and conditional distributions $p(\mathbf{x}|\mathbf{b},\boldsymbol{\lambda},\boldsymbol{\delta}), p(\lambda_i|\mathbf{b},\mathbf{x})$, and $p(\delta_i|\mathbf{x})$

for $k = 1,\ldots,K$ **do**
  for $i = 1,\ldots,N$ **do**
    $\lambda_i^{(k)} \sim p(\lambda_i|\mathbf{b},\mathbf{x}^{(k-1)})$ (a gamma distribution)
    $\delta_i^{(k)} \sim p(\delta_i|\mathbf{x}^{(k-1)})$ (a gamma distribution)
  end
  $\mathbf{x}^{(k)} \sim p(\mathbf{x}|\mathbf{b},\boldsymbol{\lambda}^{(k)},\boldsymbol{\delta}^{(k)})$ (a normal distribution)
end

3.1 The HGSV Sampler

Here, we use a Gibbs sampler to draw samples from the posterior distribution (3.6). This requires that we sample the conditional distributions $p(\lambda_i|\mathbf{b},\mathbf{x})$, $p(\delta_i|\mathbf{b},\mathbf{x})$ for $i = 1,\ldots,N$, and $p(\mathbf{x}|\mathbf{b},\boldsymbol{\lambda},\boldsymbol{\delta})$. Per our choices and assumptions, these conditional distributions are standard distributions. In particular, we have:

$$\lambda_i|\mathbf{b},\mathbf{x} \sim \text{Gamma}\left(\frac{1}{2}+\alpha_\lambda, \frac{1}{2}\left[(\mathbf{A}\mathbf{x})_i - b_i\right]^2 + \beta_\lambda\right), \tag{3.7}$$

$$\delta_i|\mathbf{x} \sim \text{Gamma}\left(\frac{1}{2}+\alpha_\delta, \frac{1}{2}\left(\mathbf{L}\mathbf{x}\right)_i^2 + \beta_\delta\right), \tag{3.8}$$

which are standard gamma distributions that are easy to sample. For the conditional distribution over the image $\mathbf{x}$, we have that

$$p(\mathbf{x}|\mathbf{b},\boldsymbol{\lambda},\boldsymbol{\delta}) \propto \exp\left(-\sum_{i=1}^{N}\frac{\lambda_i}{2}\left[(\mathbf{A}\mathbf{x})_i - b_i\right]^2 - \sum_{i=1}^{N}\frac{\delta_i}{2}\left(\mathbf{L}\mathbf{x}\right)_i^2\right), \tag{3.9}$$

$$= \exp\left(-\frac{1}{2}\left|\left|\boldsymbol{\Lambda}^{1/2}\left(\mathbf{A}\mathbf{x}-\mathbf{b}\right)\right|\right|_2^2 - \frac{1}{2}\left|\left|\boldsymbol{\Delta}^{1/2}\mathbf{L}\mathbf{x}\right|\right|_2^2\right), \tag{3.10}$$

where $\boldsymbol{\Lambda}^{1/2} = \text{diag}\left(\boldsymbol{\lambda}^{1/2}\right)$ and $\boldsymbol{\Delta}^{1/2} = \text{diag}\left(\boldsymbol{\delta}^{1/2}\right)$ with $\boldsymbol{\lambda}^{1/2} = \left(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_N^{1/2}\right)^T$ and $\boldsymbol{\delta}^{1/2} = \left(\delta_1^{1/2}, \delta_2^{1/2}, \ldots, \delta_N^{1/2}\right)^T$. Here and below, $\text{diag}(\mathbf{v})$ is a $n \times n$ diagonal matrix whose diagonal elements are the n components of the vector $\mathbf{v}$. It is now easy to see that $p(\mathbf{x}|\mathbf{b},\boldsymbol{\lambda},\boldsymbol{\delta})$ is the normal distribution

$$p(\mathbf{x}|\mathbf{b},\boldsymbol{\lambda},\boldsymbol{\delta}) \sim \mathcal{N}\left(\mathbf{H}^{-1}\mathbf{A}^T\boldsymbol{\Lambda}\mathbf{b}, \mathbf{H}^{-1}\right), \tag{3.11}$$

where $\mathbf{H} = \mathbf{A}^T\boldsymbol{\Lambda}\mathbf{A} + \mathbf{L}^T\boldsymbol{\Delta}\mathbf{L}$ with $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ and $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta})$.

With all conditional distributions in place, it is now straightfoward to perfrom Gibbs sampling and we call this sampler the HGSV sampler, for a (h)ierarchical (G)ibbs sampler for (s)patially (v)arying hyperparameters. Algorithm 1 details how the HGSV works and shows that the spatially varying parameters are sampled from their conditional distributions, conditioned on the previous sample $\mathbf{x}^{(k-1)}$ (but other sequences are equally valid). Notice that the HGSV sampler requires the matrix square root of the precision matrix $\mathbf{H}$ which we implement via Cholesky factorization (but other strategies are also valid).

3.2 1D Illustration of the HGSV Sampler

Here we provide a 1D illustration of deblurring using the HGSV sampler (2D, i.e., image problems are discussed in section 5). We use a 1D setup here for simplicity and because it already captures the main ideas. Given a signal with $n = 256$ pixels (orange dotted line in Figure 3.1), the blurred signal is generated

Table 3.1: RMSE's of HGSVM samplers with Tikhonov or TV regularization.

	HGSV (TIK)	HGSV (TV)
RMSE$(\mathbf{x}, \cdot)$	0.06	0.05
% Decrease in RMSE	25%	37.5%

by convolving the true signal with a discrete version $\mathbf{a}$ of the kernel

$$a(s) = \frac{1}{\gamma\sqrt{2\pi}} \exp\left(-\frac{s^2}{2\gamma^2}\right), \tag{3.12}$$

with $\gamma = 0.02$. Assuming zero boundary conditions, the blurring matrix $\mathbf{A}$ is defined by a Toeplitz matrix of the discrete kernel $\mathbf{a}$, scaled by the step size $h = 1/256$ [9, 14]. The blurred, noisy signal in Figure 3.1 is given by

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}, \tag{3.13}$$

where $\mathbf{x}$ is the true signal, and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ is a 256×1 random vector with variance σ^2 chosen so that the signal-to-noise ratio (SNR), defined in [14] as

$$||\mathbf{A}\mathbf{x}||_2 / \sqrt{n\sigma^2}, \tag{3.14}$$

is 25. We run the HGSV sampler, first with a Tikhonov prior and then with an (approximate) TV prior, initializing $\mathbf{x}^{(0)} = \mathbf{b}$ and setting the maximum number of iterations to be $K = 10{,}000$. We discard the first 1,000 samples to account for burn-in, and take the mean of the remaining samples to get the reconstructions in Figures 3.2(a) and 3.2(b). To assess the effectiveness of the HGSV sampler, we compute the mean IACT for the $\boldsymbol{\lambda}$- and $\boldsymbol{\delta}$-chains and we compute values very near one, indicating a high statistical efficiency of HGSV on this problem.

We measure the error between the true signal, $\mathbf{x}$, and a reconstruction, $\overline{\mathbf{x}}$, by computing the root mean square error (RMSE) defined as

$$\text{RMSE}\left(\mathbf{x}, \overline{\mathbf{x}}\right) = \frac{||\mathbf{x} - \overline{\mathbf{x}}||_2}{\sqrt{N}}. \tag{3.15}$$

The root mean square error (RMSE) between the true signal and the blurred, noisy signal is about 0.08. Table 3.1 shows the computed RMSE values along with the percentage the RMSE has decreased from that of the blurred, noisy signal. The RMSE between the HGSV Tikhonov reconstruction and the truth is approximately 0.06, for an error reduction of 25%; and the RMSE between the HGSV TV reconstruction and the truth is approximately 0.05, for an error reduction of 37.5%.

For comparison, we also ran the hierarchical Gibbs sampler *without* spatially varying parameters, i.e., the sampler aims to find global parameters λ and δ as in equations (2.11)–(2.13). The reconstructions are shown in Figures 3.2(c) and 3.2(d). This type of sampler is used in [5, 6, 14], and we abbreviate this sampler as HG. The RMSE between the HG Tikhonov reconstruction and the truth is also approximately 0.06, and the RMSE between the HG TV reconstruction and the truth is approximately 0.04. Though the HG reconstructions and the HGSV reconstructions have similar RMSEs, the HGSV reconstructions better minimize Gibbs phenomenon artifacts, i.e., oscillations with a Tikhonov prior and staircasing with a TV prior.

3.3 Discussion and Computational Aspects

Our choice of hyperpriors, in particular their independence, is motivated by computational efficiency. The Gibbs sampler is known to be an effective tool (no samples are rejected) and has found application in a wide range of topics. Moreover, due to (conditional) independence, the sampling of the hyperparameters can be done in parallel (each λ_i is independent of the other λ's or all δ_i). The computational difficulties arise when the images are very large - large number of pixels means that the dimension N is large. In the high-dimensional case, even drawing samples from a Gaussian can be computationally demanding, but there are effective tools that can be put to use [13, 25–30].

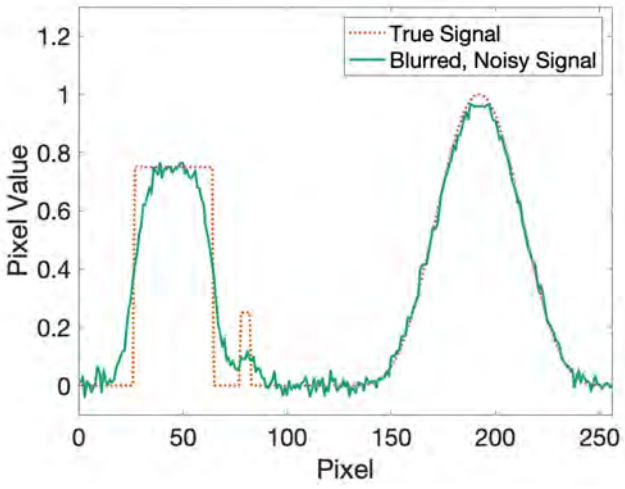


Figure 3.1: A signal (dotted orange) and its blurred, noisy version (solid green).

Here, we focus on the conceptual foundation of varying regularization throughout the image, rather than on large-scale image deblurring. In particular, how to assemble, or better avoid to assemble the blurring matrix $\mathbf{A}$, depends on the specific image and assumed boundary conditions. We leave scaling of the algorithm to large or extreme scale for future work.

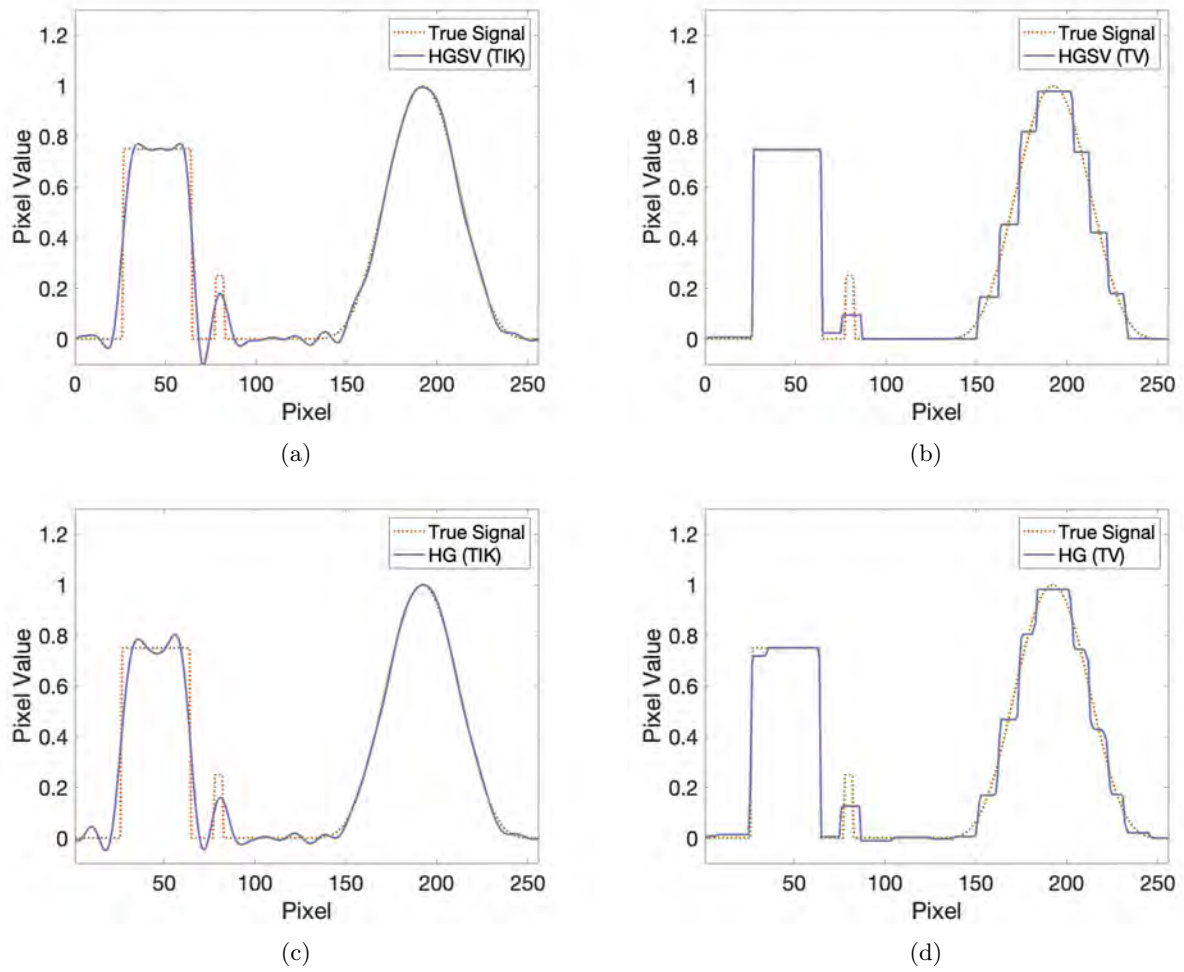


Figure 3.2: Subfigures (a) and (b) show HGSV mean reconstructions using a smoothing Tikhonov prior (a) and a TV prior (b). Subfigures (c) and (d) show standard HG mean reconstructions using a Tikhonov prior (c) and a TV prior (d) In all cases 10,000 samples were drawn, the first 1,000 samples were discarded for burn-in.

4 Varying Hyperparameters and Regularization Type

We now further extend the hierarchical, Bayesian approach to image deblurring and incorporate a spatial mixing of two priors and regularization types. Specifically, we use the "usual" Tikhonov regularization and Gaussian prior for some parts of the image, but use a Gaussian prior that approximates a TV regularization (see section 2) in other parts of the image. For now, we assume there exists a partitioning that defines where each type of regularization/prior is used, but we also outline heuristic guidelines for how to compute such a partition. We call the resulting algorithm the (h)ierarchical (G)ibbs sampler with (s)patially (v)arying parameters and (m)ixed regularization, or the HGSVM sampler. We develop this sampler because most images – and especially Cygnus images – combine smooth and sharp features, so that effectively combining the two regularization techniques results in better reconstructions.

Specifically, for an image with N pixels, we assume that it has been determined that certain regions should use Tikhonov regularization, while the remaining regions should use TV. Let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)^T$ be an $N \times 1$ logical vector that determines such a partitioning where for each $i = 1, \ldots, N$, $\pi_i \in \{0, 1\}$. Let the Tikhonov regions correspond to $\pi_i = 0$ and the TV regions correspond to $\pi_i = 1$. We now incorporate the partition $\boldsymbol{\pi}$ into the hierarchical appriach with spatially varying parameters.

For $i = 1, \ldots, N$, let λ_i be the inverse variance of the added Gaussian noise, and δ_i be the strength of the imposed regularization. These parameters are unknown and therefore, in the Bayesian paradigm, are treated as random variables. We chose independent priors for λ_i and δ_i, $i = 1, \ldots, N$; and we assign Gamma distributions to the two sets of parameters:

$$p(\lambda_i) \sim \text{Gamma}(\alpha_\lambda, \beta_\lambda) \propto \lambda_i^{\alpha_\lambda - 1} \exp(-\beta_\lambda \lambda_i), \tag{4.1}$$

$$p(\delta_i) \sim \text{Gamma}(\alpha_\delta, \beta_\delta) \propto \delta_i^{\alpha_\delta - 1} \exp(-\beta_\delta \delta_i), \tag{4.2}$$

where $\alpha_\lambda, \alpha_\delta$ are Gamma shape parameters and $\beta_\lambda, \beta_\delta$ are Gamma rate parameters as before. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^T$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)^T$. The conditonal prior for the image now depends on the partition $\boldsymbol{\pi}$:

$$p(\mathbf{x}|\boldsymbol{\delta}, \boldsymbol{\pi}) \propto \left(\prod_{i=1}^{N} \delta_i \right)^{1/2} \exp\left(-\frac{\delta_i}{2} \sum_{i=1}^{N} \left[\, (1 - \pi_i) \left(\mathbf{L}_{\text{TIK}} \mathbf{x} \right)_i^2 + \pi_i \left(\mathbf{L}_{\text{TV}} \mathbf{x} \right)_i^2 \, \right] \right), \tag{4.3}$$

where $\mathbf{L}_{\text{TV}}$ is as defined in section 2.2, and $\mathbf{L}_{\text{TIK}}$ is the positive definite matrix used for Tikhonov regularization (typically a first or second order derivative).

The likelihood does not depend on regularization terms and, therefore, does not depend on $\boldsymbol{\pi}$. Hence, the likelihood is as in the previous sections:

$$p(\mathbf{b}|\mathbf{x}, \boldsymbol{\lambda}) \propto \left(\prod_{i=1}^{N} \lambda_i \right)^{1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \lambda_i \left[(\mathbf{A}\mathbf{x})_i - \mathbf{b}_i \right]^2 \right). \tag{4.4}$$

The posterior distribution is defined by product of the priors and the likelihood:

$$\begin{aligned} p(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\delta}|\mathbf{b}, \boldsymbol{\pi}) &\propto p(\mathbf{b}|\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\pi}) p(\mathbf{x}|\boldsymbol{\delta}, \boldsymbol{\pi}) p(\boldsymbol{\lambda}) p(\boldsymbol{\delta}), \qquad (4.5) \\ &\propto \left(\prod_{i=1}^{N} \lambda_i \right)^{\alpha_\lambda - 1/2} \left(\prod_{i=1}^{N} \delta_i \right)^{\alpha_\delta - 1/2} \\ &\quad \times \exp\left(-\sum_{i=1}^{N} \lambda_i \left(\frac{1}{2} \left[(\mathbf{A}\mathbf{x})_i - \mathbf{b}_i \right]^2 + \beta_\lambda \right) \right) \\ &\quad \times \exp\left(-\sum_{i=1}^{N} \delta_i \left(\frac{(1 - \pi_i)}{2} \left(\mathbf{L}_{\text{TIK}} \mathbf{x} \right)_i^2 + \frac{\pi_i}{2} \left(\mathbf{L}_{\text{TV}} \mathbf{x} \right)_i^2 + \beta_\delta \right) \right). \qquad (4.6) \end{aligned}$$

This posterior describes the image reconstructions ($\mathbf{x}$), along with hyperparameters that describe the noise level ($\boldsymbol{\lambda}$),and the regularization strength/prior covariance for two different types of priors ($\boldsymbol{\delta}$).

4.1 The HGSVM Sampler

We draw samples from the posterior distribution (4.6) by using a Gibbs sampler. This requires that we draw samples from the conditionals, which is largely as before. What has changed, however, is the incorporation of the partition $\boldsymbol{\pi}$. The $\boldsymbol{\lambda}^{(k)}$ samples do not depend on the partition, so they are sampled as before. We note that δ_i determines the strength of TV regularization whenever $\pi_i = 1$ and Tikhonov regularization whenever $\pi_i = 0$. The $\boldsymbol{\delta}^{(k)}$ samples are sampled by

$$\delta_i^{(k)} \sim (1 - \pi_i) p_{\text{TIK}} \left(\delta_i | \mathbf{x}^{(k-1)} \right) + \pi_i p_{\text{TV}} \left(\delta_i | \mathbf{x}^{(k-1)} \right), \tag{4.7}$$

where $p_{\text{TIK}}(\delta_i|\mathbf{x})$ and $p_{\text{TV}}(\delta_i|\mathbf{x})$ are the conditionals with only Tikhonov or TV regularization, respectively. The sampled images $\mathbf{x}^{(k)}$ are sampled from the normal distribution

$$\mathcal{N}\left(\mathbf{m}^{(k)}, \mathbf{H}^{(k)} \right), \tag{4.8}$$

Algorithm 2: HGSVM Sampler

input: Initial approximation $\mathbf{x}^{(0)}$; desired number of samples K;
logical vector $\boldsymbol{\pi}$; and conditional distributions $p(\mathbf{x}|\mathbf{b},\boldsymbol{\lambda},\boldsymbol{\delta}), p(\lambda_i|\mathbf{b},\mathbf{x})$, $p(\delta_i|\mathbf{x})$

for $k = 1,\ldots,K$ **do**
  for $i = 1,\ldots,N$ **do**
    $\lambda_i^{(k)} \sim p\left(\lambda_i|\mathbf{b},\mathbf{x}^{(k-1)}\right)$ (gamma distribution)
    $\delta_i^{(k)} \sim p\left(\delta_i|\mathbf{x}^{(k-1)},\boldsymbol{\pi}\right)$ (gamma distribution)
  end
  $\mathbf{x}^{(k)} \sim p\left(\mathbf{x}|\mathbf{b},\boldsymbol{\lambda}^{(k)},\boldsymbol{\delta}^{(k)},\boldsymbol{\pi}\right)$ (normal distribution)
end

with

$$\mathbf{H}^{(k)} := \left(\mathbf{A}^T\boldsymbol{\Lambda}^{(k)}\mathbf{A} + \mathbf{L}_{\text{TIK}}^T\boldsymbol{\Pi}\boldsymbol{\Delta}^{(k)}\boldsymbol{\Pi}\mathbf{L}_{\text{TIK}} + \mathbf{L}_{\text{TV}}^T(\mathbf{I}-\boldsymbol{\Pi})\boldsymbol{\Delta}^{(k)}(\mathbf{I}-\boldsymbol{\Pi})\mathbf{L}_{\text{TV}}\right)^{-1}, \tag{4.9}$$

$$\mathbf{m}^{(k)} := \mathbf{H}^{(k)}\mathbf{A}^T\boldsymbol{\Lambda}^{(k)}\mathbf{b}, \tag{4.10}$$

where $\boldsymbol{\Lambda}^{(k)} = \text{diag}\left(\boldsymbol{\lambda}^{(k)}\right)$; $\boldsymbol{\Delta}^{(k)} = \text{diag}\left(\boldsymbol{\delta}^{(k)}\right)$; and $\boldsymbol{\Pi} = \text{diag}\left(\boldsymbol{\pi}\right)$. The HGSVM sampler is summarized in Algorithm 2. Along with the conditional distributions (using (3.1) and replacing (3.2) with (4.7)), the algorithm requires an initial approximation $\mathbf{x}^{(0)}$ to the true signal, the desired number of samples K, and the partitioning vector $\boldsymbol{\pi}$. We first sample the vectors of hyperparameters $\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$ from the gamma distributions in (3.1)–(3.2) – but, again, other sequences are equally valid. We then use the sampled values of $\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$ to sample from the conditional distribution $p(\mathbf{x}|\mathbf{b},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\pi})$. Once K samples have been drawn and the number c of initial samples to discard for burn-in has been determined, the optimal reconstruction is the mean of the remaining $K-c$ samples.

4.2 1D Illustration of the HGSVM Sampler

We consider the same 1D deblurring example as before, using the blurred and noisy signal in Figure 3.1. We observed from the reconstructions in Figure 3.2, that the Tikhonov-based reconstruction experienced oscillations on the left half of the signal, while the TV-based reconstruction had a "staircasing" effect on the right half of the signal. We conclude that an effective regularization partitioning would be to use TV on the left half of the signal (up to about pixel 145), and Tikhonov on the right half. Therefore, in the HGSVM sampler, the vector $\boldsymbol{\pi}$ would contain 1's up to the 145th entry, and 0's everywhere else. Figure 4.1 shows the mean reconstruction of 10,000 samples with a burn-in of 1,000 samples.

The mean IACT's for the $\boldsymbol{\lambda}$- and $\boldsymbol{\delta}$-chains are $\overline{\tau}_{\boldsymbol{\lambda}} = 1.0115$ and $\overline{\tau}_{\boldsymbol{\delta}} = 1.1328$, indicating a high statistical efficiency of the HGSVM sampler.The RMSE between the HGSV Tikhonov reconstruction and the truth is 0.0017, for an error reduction of 66%. Thus, the HGSVM sampler produces results that are both qualitatively and quantitatively better than HGSVM or the traditional hierarchical Bayesian method.

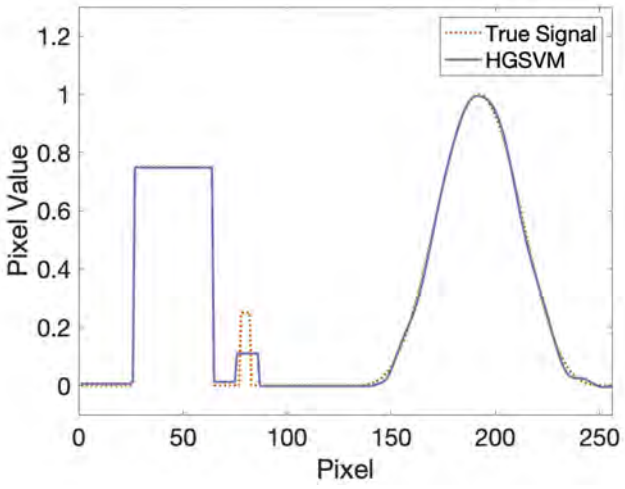


Figure 4.1: A mean reconstruction using the HGSVM sampler. The left half of the signal uses a TV prior, while the right half uses a Tikhonov prior.

5 Numerical Examples

In this section, we apply HGSV and HGSVM to two-dimensional test problems. We first use two synthetic test images, and we conclude by applying the methods to a test image from Cygnus.

5.1 Synthetic Deconvolution Examples in 2D

Figures 5.1 and 5.2 show deblurring results of two 128×128 simulated test images. The true images are shown in Figures 5.1(a) and 5.2(a). Figure 5.1(a) shows a smooth sphere-like object in the left center region of the scene and a solid rectangle in the top right. We will refer to this image as the "sphere/rectangle image". Figure 5.2(a) shows a cartoon picture of a satellite with stars and planets in the background. This image is more complex, as the body of the satellite contains several interconnected pieces, with a glaring effect on the top left and bottom right panels of the satellite. We will refer to this image as the "satellite image."

We blur both test images with a discrete version of the two-dimensional Gaussian kernel

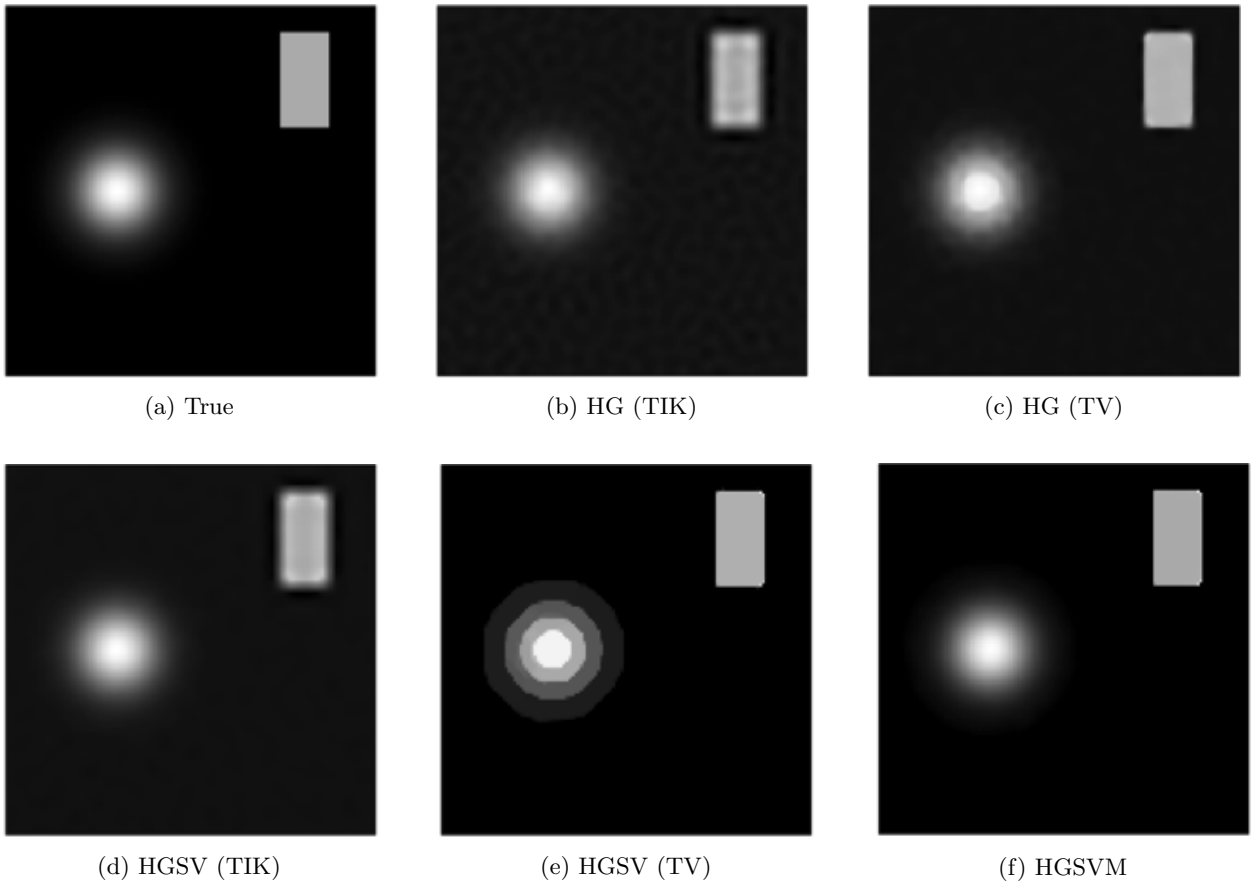
$$a(s,t) = \frac{1}{2\pi\gamma^2} \exp\left(-\frac{s^2}{2\gamma^2} - \frac{t^2}{2\gamma^2}\right), \tag{5.1}$$

assuming zero boundary conditions. We add Gaussian noise with SNR equal to 25 to the blurred image and it is this blurred and noisy image we attempt to deblur.

5.1.1 Sphere/Rectangle Test Image

Figure 5.1 shows the true sphere/rectangle test image along with five reconstructions. The reconstructions labeled HG (TIK/TV) use the standard non-spatially varying parameters hierarchical Gibbs sampler found in section 2. The bottom row shows two HGSV reconstructions with Tikhonov and TV-type regularizations, and the final image is the HGSVM reconstruction. In all cases, 1,000 samples were drawn with the initial 100 samples discarded for burn-in. We note that for the HG reconstructions, there is only one λ-chain and δ-chain for the entire image; whereas for the HGSV and HGSVM reconstructions, there are 128^2 λ-chains and δ-chains. Table 5.1 shows the mean (in the HG case, single) IACT values for the chains, along with the RMSE values between the reconstructions and the true image.

The HG (TIK) and HG (TV) reconstructions are still experiencing the effects of blur and noise. We see oscillations in the background of the HG (TIK) reconstruction and blurred edges in the HG (TV)



(a) True (b) HG (TIK) (c) HG (TV)

(d) HGSV (TIK) (e) HGSV (TV) (f) HGSVM

Figure 5.1: (Top row) The true sphere/rectangle test image (a); the mean reconstruction using the standard hierarchical Gibbs sampler with Tikhonov regularization (b); the mean reconstruction using the standard hierarchical Gibbs sampler with TV regularization (c). (Bottom row) HGSV reconstruction with Tikhonov regularization (d); HGSV reconstruction with TV regularization (e); HGSVM reconstruction with Tikhonov regularization applied to the sphere and TV regularization applied everywhere else (f).

reconstruction. The IACT's for the δ-parameters are high, at 21.9 for the HG (TIK) sampler and 56.5 for the HG (TV) sampler. This means for one sample of the δ hyperparameter, the HG (TIK) sampler would need to generate approximately 22 more samples before obtaining an independent sample. The HG (TV) sampler would need to generate about 57 samples before obtaining an effectively independent sample. The reconstructions, however, have RMSE's with the true image that have decreased by 25% for HG (TIK) and 50% for HG (TV) compared to the initial error between the blurred, noisy image and the truth.

The HGSV (TIK), HGSV (TV), and HGSVM samplers produce λ- and δ-chains with average IACT's close to one, indicating that almost every sample is independent. The HGSV (TIK) reconstruction has an RMSE value about the same amount as the HG (TIK) reconstruction. The HGSV (TV) reconstruction successfully sharpens the edges of the solid rectangle; however, it also forces the smooth bump to transform into concentric circles. Despite this, the RMSE is about 50% less than that of the initial RMSE between the blurred, noisy image and the true image. Finally, the HGSVM sampler produces a reconstruction that is the closest qualitatively and quantitatively to the true image. The sampler uses a partitioning scheme where Tikhonov regularization is used for the sphere, and TV regularization is used everywhere else. The RMSE is about 75% less than the original RMSE between the blurred, noisy image and the true image, which is the largest error reduction of all methods we tried.

Table 5.1: A comparison of RMSE's and mean IACT's for the parameter vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$ in the sphere/rectangle reconstructions. The reconstructions HG (TIK/TV) refers to those obtained from the standard hierarchical Gibbs sampler, and HGSV (TIK/TV) refers to the reconstructions obtained from the HGSV sampler. The original error between the truth and observed blurred and noisy image is $\text{RMSE}(\mathbf{x}, \mathbf{b}) \approx 0.04$.

	HG (TIK)	HGSV (TIK)	HG (TV)	HGSV (TV)	HGSVM
$\overline{\tau}_{\boldsymbol{\lambda}}$	2.59	1.02	2.11	1.01	1.00
$\overline{\tau}_{\boldsymbol{\delta}}$	21.9	1.11	56.5	1.93	1.19
$\text{RMSE}(\mathbf{x}, \cdot)$	0.03	0.03	0.02	0.02	0.01
% Decrease in RMSE	25%	25%	50%	50%	75%

Table 5.2: A comparison of RMSE's and mean IACT's for the parameter vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$ in the satellite reconstructions. The reconstructions HG (TIK/TV) refers to those obtained from the standard hierarchical Gibbs sampler, and HGSV (TIK/TV) refers to the reconstructions obtained from the HGSV sampler. The original error between the truth and observed blurred and noisy image is $\text{RMSE}(\mathbf{x}, \mathbf{b}) \approx 0.09$.

	HG (TIK)	HGSV (TIK)	HG (TV)	HGSV (TV)	HGSVM
$\overline{\tau}_{\boldsymbol{\lambda}}$	3.17	1.06	123	1.04	1.01
$\overline{\tau}_{\boldsymbol{\delta}}$	19.9	1.35	40.0	1.88	1.41
$\text{RMSE}(\mathbf{x}, \cdot)$	0.06	0.07	0.21	0.04	0.02
% Decrease in RMSE	33.3%	22.2%	−133.3%	55.6%	77.8%

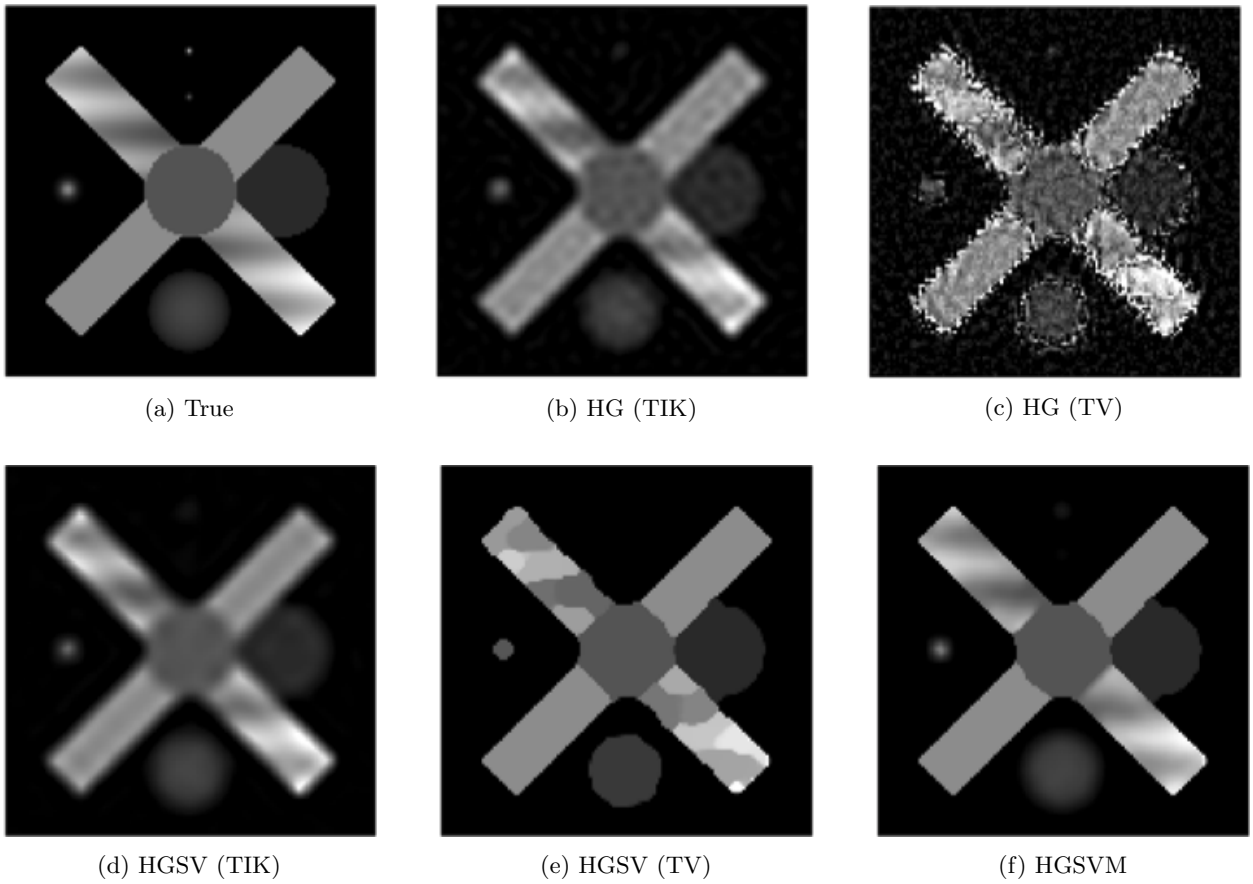
5.1.2 Satellite Test Image

Figure 5.2 shows the true satellite image in the left subfigure on the top row, along with HG (TIK/TV), HGSV (TIK/TV), and HGSVM reconstructions. For all samplers, 1,000 samples were drawn with the initial 100 discarded for burn-in. Table 5.2 shows the mean IACT values for the λ- and δ-chains, along with a comparison of the reconstructions' errors to that between the true and blurred images.

We observe essentially the same behavior in this more complex image than in the simpler example just above: the HGSVM sampler yields the best reconstruction in terms of RMSE and also in terms of qualitatively comparing the reconstruction with the true figure. All other sampler produce reconstructions that are polluted by artifacts. In particular, the HG (TV) reconstruction is noisy and fails to reduce the RMSE at all. The HG (TIK) sampler produces λ- and δ-chains with IACT's $\tau_\lambda = 3.17$ and $\tau_\delta = 19.9$, and the reconstruction reduces the initial RMSE between the true and blurred image by 25%, even though it is qualitatively poor. The HGSV (TIK), HGSV (TV), and HGSVM reconstructions again produce chains with IACT values close to one, indicating that almost every sample is independent after burn-in. The HGSVM reconstruction applies Tikhonov regularization to the planets in the background, as well as the top left and bottom right satellite panels. TV regularization is applied everywhere else. The HGSVM reconstruction qualitatively matches the true satellite image well, and reduces the original RMSE value by approximately 77.8%.

5.2 Deconvolution with Cygnus Test Data

We now apply the HGSVM sampler (and its simpler variants) to real data, obtained by the Cygnus X-ray system at the Nevada National Security Site. The image shown in Figure 5.3 is 4K×4K, and is used for calibrating the X-ray system. The image contains the following objects: a step wedge, an Abel cylinder, and an L-rolled edge. The step wedge is on the left, producing the stacked solid blocks. The Abel cylinder is in the top middle of the scene, and the cylinder consists of concentric cylinders of varying densities. The L-rolled edge or "Pacman"-shaped object on the right is made of tungsten, which blocks all incoming X-rays. This results in solid black in that portion of the image. Because we focus on describing the conceptual foundation of varying regularization and noise levels throughout the image, and because our code is not optimized to work at very large scales, we consider a smaller portion indicated by the smaller red box in the left panel of Figure 5.3. The box contains a 400×400 sub-image, which we downsample to an image of size 100×100. To deal with boundary conditions, we in fact deblur a larger image (outer red box in Figure 5.3(a)) with zero



(a) True (b) HG (TIK) (c) HG (TV)

(d) HGSV (TIK) (e) HGSV (TV) (f) HGSVM

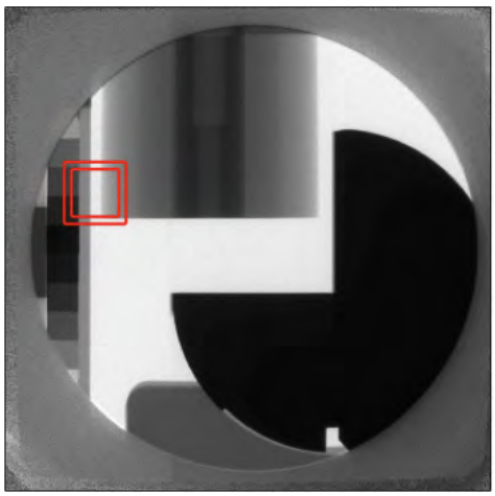
Figure 5.2: (Top row) The true satellite test image (a); the mean reconstruction using the standard hierarchical Gibbs sampler with Tikhonov regularization (b); the mean reconstruction using the standard hierarchical Gibbs sampler with TV regularization (c). (Bottom row) HGSV reconstruction with Tikhonov regularization (d); HGSV reconstruction with TV regularization (e); HGSVM reconstruction with Tikhonov regularization applied to the planets along with the top left and bottom right panels, and TV regularization applied everywhere else (f).

boundary conditions and crop the outer rim that is polluted by boundary artifacts (see [31] for a discussion on boundary artifacts in MCMC-based deconvolution). The blurring matrix is built by estimating the point spread function using the methods in [22] as the blurring kernel.

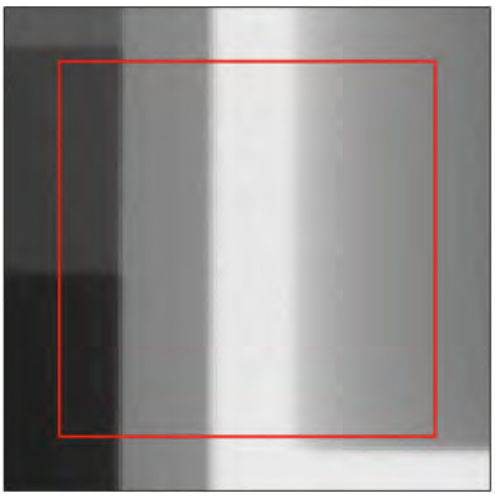
Figure 5.4 shows the original scaled-down sub-image of interest along with HGSV (TIK/TV) and HGSVM reconstructions. The HGSV (TIK/TV) samplers are successful at producing quality reconstructions. The HGSVM sampler uses Tikhonov in the region containing the Abel cylinder and TV everywhere else. This choice is based on prior knowledge about the calibration objects. The Abel cylinder is made up of concentric cylinders; therefore when X-rayed, the image should show smooth transitions. Table 5.3 shows the mean IACT's for the λ- and δ-chains. The HGSV (TIK/TV) and HGSVM samplers produce chains that are on average close to one, with the exception of the δ-chains for the HGSV (TV) sampler. The mean IACT for the δ-chains are approximately 2.15, but this is due to the sampler struggling within the Abel cylinder region.

We interject by noting that the poor performance of the HGSV (TV) reconstruction in the Abel cylinder area is an example of how one might heuristically determine a regularization-based partition. If we were given a corrupted image with no prior information about the true scene, we could run the HGSV sampler with Tikhonov and TV regularization separately, then analyze the reconstructions to determine regions of irregular behavior. Staircasing effects (artifacts from TV regularization) are more easily spotted in 2D, as sharp jumps from one color to another within a small area are easier to detect with the naked eye.

The HGSVM sampler is best for this problem due to the prior knowledge of the calibration targets. The



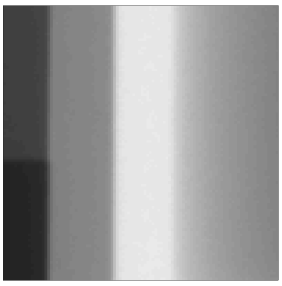
(a) Cygnus test image



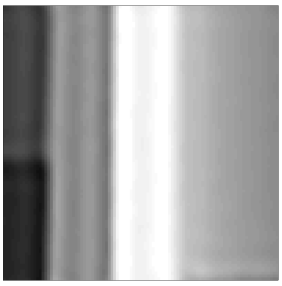
(b) Region of interest

Figure 5.3: (a) The 4K×4K Cygnus test image. The inner red box highlights a 400×400 sub-image of interest, and the outer red box highlights a 512×512 region containing the sub-image of interest. (b) The zoomed in 512×512 region with the 400×400 sub-image of interest within the inner red box. For our computations, we use the larger region which will incur artifacts along its boundaries. We then crop out the boundaries and take this as the reconstruction for the smaller sub-image.

true profile of the step wedge should be piecewise constant, while the Abel cylinder should be smooth due to the X-rays passing through the concentric cylinders. Hence, Tikhonov regularization should be applied to the region containing the Abel cylinder, and TV regularization should be enforced everywhere else. We also note how the HGSVM sampler successfully handles the transition from TV to Tikhonov regularization at the border of the Abel cylinder region. If we were to simply piece together the HGSV reconstructions in Figures 5.4(b) and 5.4(c) using the existing partitioning scheme, artifacts could appear along the transitional border.

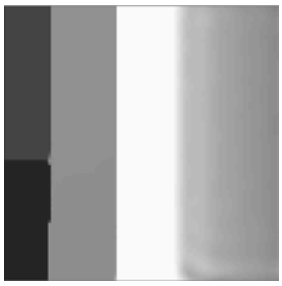


(a) Original



(b) HGSV (TIK)

(c) HGSV (TV)



(d) HGSVM

Figure 5.4: The scaled-down original 100×100 Cygnus sub-image (a); the HGSV reconstruction with Tikhonov regularization (b); HGSV reconstruction with TV regularization (c); HGSVM reconstruction with Tikhonov regularization applied to the Abel cylinder section and TV regularization applied everywhere else (d).

Table 5.3: A comparison of mean IACT's for the parameter vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$ in the Cygnus reconstructions.

	HGSV (TIK)	HGSV (TV)	HGSVM
$\overline{\tau}_{\boldsymbol{\lambda}}$	1.27	1.08	1.08
$\overline{\tau}_{\boldsymbol{\delta}}$	1.53	2.15	1.53

6 Conclusion

In this work, we have introduced a new method (the HGSVM method) that allows not only the strength, but the type of regularization to vary across an image. We do this by formulating the deblurring problem in a hierarchical Bayesian framework. Using the Bayesian paradigm, the unknown regularization parameters are random variables, eliminating the need to hand-tune the parameters or choose an appropriate parameter selection method. The spatially varying parameters, especially in the case of TV regularization, produce successful reconstructions because the added noise forces certain regions of the image to require stronger regularization than others. Having the parameters vary also allows us to effectively mix regularization methods, which is necessary in the common case of deblurring an image with both smooth features and edges. The mixing of Tikhonov and TV regularization in an image is done by creating a logical partitioning vector of 1's and 0's. The partitioning vector is built based on a combination of prior knowledge and heuristics, but we aim to automate this process in the future.

We illustrated the HGSVM sampler on simplified test problems and synthetic images (where the true image and noise statistics are known). In all test problems, we found that HGSVM is statistically efficient (small integrated autocorrelation time) and leads to reconstructions that are qualitatively and quantitatively better than reconstructions obtained with simpler methods. We also successfully applied HGSVM to a 100×100 scaled-down sub-image of a 4K×4K image captured by Cygnus. The reconstruction via HGSVM are superior to those obtained with simpler methods (HG, or HGSV). As in the synthetic numerical tests, we found that HGSVM is statistically efficient with integrated autocorrelation times between 1 and 2 (on average). This means that about every other sample generated from the HGSVM sampler is independent.

The HGSVM method has proven to be useful in deblurring Cygnus sub-images. Future work should address the scalability of HGSVM to larger images. To clarify, we are confident HGSVM can successfully deblur high-resolution images when given a large storage capacity, as with a supercomputer. We would, however, prefer the sampler to work for larger images using only a standard laptop.

Acknowledgements

This manuscript has been authored in part by Mission Support and Test Services, LLC, under Contract No. DE-NA0003624 with the U.S. Department of Energy, National Nuclear Security Administration (DOE-NNSA), NA-10 Office of Defense Programs, and supported by the Site-Directed Research and Development Program. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published content of this manuscript, or allow others to do so, for United States Government purposes. The U.S. Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government. DOE/NV/03624–1256.

Matthias Morzfeld was supported by the US Office of Naval Research (ONR) grant N00014-21-1-2309.

References

[1] A. Fiandrotti, S. M. Fosson, C. Ravazzi, and E. Magli, "GPU-accelerated algorithms for compressed signals recovery with application to astronomical imagery deblurring," *International Journal of Remote Sensing*, vol. 39, no. 7, pp. 2043–2065, 2018.

[2] J. Lee and P. Kitanidis, "Bayesian inversion with total variation prior for discrete geologic structure identification," *Water Resources Research*, vol. 49, pp. 7658–7669, 2013.

[3] M. Picozzi, S. Parolai, and D. Bindi, "Deblurring of frequency-wavenumber images from small-scale seismic arrays," *Geophysical Journal International*, vol. 181, no. 1, pp. 357–368, 2010.

[4] H. Zhao, Z. Ke, *et al.*, "A New Deep Learning Method for Image Deblurring in Optical Microscopic Systems," *Journal of Biophotonics*, vol. 13, no. 3, 2020.

[5] M. Howard, M. Fowler, A. Luttman, S. E. Mitchell, and M. C. Hock, "Bayesian Abel inversion in quantitative X-ray radiography," *SIAM J. Sci. Comput.*, vol. 38, no. 3, pp. B396–B413, 2016.

[6] M. Howard, A. Luttman, and M. Fowler, "Sampling-based uncertainty quantification in deconvolution of X-ray radiographs," *J. Comput. Appl. Math.*, vol. 270, pp. 43–51, 2014.

[7] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. IOP Publishing, Bristol, UK, 1998.

[8] T. F. Chan and J. Shen, *Image processing and analysis*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005. Variational, PDE, wavelet, and stochastic methods.

[9] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring images*, vol. 3 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006. Matrices, spectra, and filtering.

[10] R. L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*. The Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, 1991.

[11] D. S. Nelson, E. C. Ormond, I. Molina, *et al.*, "Cygnus Trigger System," *IEEE Transactions on Plasma Science*, vol. 36, no. 5, 2008.

[12] J. Adams, *Scalable block Gibbs sampling for image deblurring in X-ray radiography*. PhD thesis, The University of Arizona, 2019.

[13] J. Adams, M. Morzfeld, K. Joyce, M. Howard, and A. Luttman, "A blocking scheme for dimension-robust Gibbs sampling in large-scale image deblurring," *Inverse Problems in Science and Engineering*, 2021.

[14] J. M. Bardsley, *Computational uncertainty quantification for inverse problems*, vol. 19 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2018.

[15] D. Calvetti, M. Pragliola, E. Somersalo, and A. Strang, "Sparse reconstructions from few noisy data: Analysis of hierarchical Bayesian models with generalized gamma hyperpriors," *Inverse Problems*, vol. 36, no. 2, 2020.

[16] J. Pillow, *Bayesian Spatially Varying Multi-Regularization Image Deblurring*. PhD thesis, The University of Arizona, 2021.

[17] C. R. Vogel, *Computational methods for inverse problems*, vol. 23 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. With a foreword by H. T. Banks.

[18] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp. 259–268, 1992.

[19] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, 1984.

[20] G. Roberts and S. Sahu, "Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler," *J R Stat Soc Ser B*, vol. 59, pp. 291–317, 1997.

[21] A. Sokal, "Monte Carlo methods in statistical mechanics: foundations and new algorithms," in *Functional integration (Cargèse, 1996)*, vol. 361 of *NATO Adv. Sci. Inst. Ser. B Phys.*, pp. 131–192, Plenum, New York, 1997.

[22] K. T. Joyce, J. M. Bardsley, and A. Luttman, "Point spread function estimation in X-ray imaging with partially collapsed Gibbs sampling," *SIAM J. Sci. Comput.*, vol. 40, no. 3, pp. B766–B787, 2018.

[23] J. M. Bardsley, "Gaussian Markov Random Field Priors for Inverse Problems," *Inverse Problems and Imaging*, vol. 7, no. 2, pp. 397–416, 2013.

[24] L. Antonelli, V. De Simone, and D. di Serafino, "Spatially Adaptive Regularization in Image Segmentation," *Algorithms*, vol. 13, no. 9, 2020.

[25] C. Fox and R. A. Norton, "Fast Sampling in a Linear-Gaussian Inverse Problem," *SIAM/ASA J. Uncertainty Quantification*, vol. 4, pp. 1191–1218, 2016.

[26] C. Fox and A. Parker, "Accelerated Gibbs sampling of normal distributions using matrix splittings and polynomials," *Bernoulli*, vol. 23, p. 3711–3743, 2017.

[27] A. Parker and C. Fox, "Sampling Gaussian distributions in Krylov spaces with conjugate gradients," *SIAM J Sci Comput*, vol. 34, pp. B312–B334, 2012.

[28] A. Parker, B. Pitts, L. Lorenz, *et al.*, "Polynomial accelerated solutions to a large Gaussian model for imaging biofilms: in theory and finite precision," *J Am Stat Assoc*, vol. 113, pp. 1431–1442, 2018.

[29] J. Chen, M. Anitescu, and Y. Saad, "Computing f(a)b via least squares polynomial approximations," *SIAM J Sci Comput*, vol. 33, pp. 195–222, 2011.

[30] E. Chow and Y. Saad, "Preconditioned Krylov subspace methods for sampling multivariate Gaussian distributions," *SIAM J Sci Comput*, vol. 36, pp. A588–A608, 2014.

[31] J. M. Bardsley and A. Luttman, "Dealing with boundary artifacts in MCMC-based deconvolution," *Linear Algebra and its Applications*, vol. 473, pp. 339–358, 2015.