A METROPOLIS-HASTINGS METHOD FOR LINEAR INVERSE PROBLEMS WITH POISSON LIKELIHOOD AND GAUSSIAN PRIOR

Johnathan M. Bardsley¹,∗ & Aaron Luttman²

¹Department of Mathematical Sciences, The University of Montana, Missoula, MT 59812-0864, USA
²Applied Mathematics and Physics, National Security Technologies, LLC, P.O. Box 98521, M/S NLV078, Las Vegas, NV, 89193-8521, USA

Original Manuscript Submitted: 03/27/2015; Final Draft Received: 03/27/2015

Poisson noise models arise in a wide range of linear inverse problems in imaging. In the Bayesian setting, the Poisson likelihood function together with a Gaussian prior yields a posterior density function that is not of a well known form and is thus difficult to sample from, especially for large-scale problems. In this work, we present a method for computing samples from posterior density functions with Poisson likelihood and Gaussian prior, using a Gaussian approximation of the posterior as an independence proposal within a Metropolis-Hastings framework. To define our priors, we use Gaussian and Laplace-distributed Markov random fields, which are the Bayesian analogues of smoothness and total variation regularization, respectively. For the Laplace prior, a Gaussian approximation is used, whereas for the Gaussian prior, the scaling (or regularization) parameter is sampled using a hierarchical Gibbs sampler, eliminating the need to choose a regularization parameter a priori. The results are demonstrated on synthetic data—including a synthetic X-ray radiograph generated from a radiation transport code—and on real images used to calibrate a pulsed power high-energy X-ray source at a U.S. Department of Energy X-ray radiography facility.

KEY WORDS: Inverse Problems, Bayesian inference, Markov Chain Monte Carlo, Image Deblurring

1. INTRODUCTION

Image deconvolution is a fundamental problem in both quantitative and qualitative image analysis. Whereas qualitative imaging focuses on computing image reconstructions that “look good,” the purpose of quantitative imaging is to determine true values of the locations of image features or of properties of the objects in the scene, such as density. For accurate quantitative analysis, such as that done with X-ray radiography in the security sciences, deconvolution is essential for accurate numerical results, and it is also important to be able to estimate the uncertainties in the calculations. The purpose of this work is to present an approach to image deconvolution with uncertainty quantification that can be used in quantitative imaging when the Poisson negative-log likelihood is used as a fit-to-data function.

We begin with the standard convolution model,

\[ b(s) = Ax(s) \overset{\text{def}}{=} \int_{\Omega} a(s-t)x(t)dt, \quad s \in \Omega, \]

where \( b \) represents the measured image, defined on the field of view (FOV) \( \Omega \); \( a \) denotes the point spread function (PSF); and \( x \) denotes the unknown image, defined on an extended FOV \( \bar{\Omega} \), which is determined by the support of \( a \). For example, if \( \Omega = [0, 1] \) and \( a \) has support contained within \([-\xi, \xi]\), for \( \xi > 0 \), then \( \bar{\Omega} = [-\xi, 1 + \xi] \).

Since \( \Omega \subseteq \bar{\Omega} \), discretizing (1) yields a non-square linear system of equations \( \mathbf{b} = \mathbf{A} \mathbf{x} \), where \( \mathbf{b} \in \mathbb{R}^M \), \( \mathbf{x} \in \mathbb{R}^N \).

---

*Correspond to: Johnathan M. Bardsley, E-mail: bardsleyjm@mso.umt.edu, URL: http://www.math.umt.edu/bardsley/
and $\mathbf{A} \in \mathbb{R}^{M \times N}$, with $N > M$. Greater detail on the discretization of (1) is provided below. In practice, the values contained in $\mathbf{b}$ correspond to measurements, and we make the assumption that the measurement noise is Poisson distributed, which is a reasonable assumption in many quantitative imaging applications (see Section 4). Specifically, given a known background $\mathbf{g} \in \mathbb{R}^M$, we assume the statistical model

$$\mathbf{b} = \text{Poiss}(\mathbf{A} \mathbf{x} + \mathbf{g}),$$

(2)

where $\text{Poiss}(\mathbf{v})$ denotes an independent Poisson random vector with Poisson parameter vector (mean and variance) $\mathbf{v}$. The likelihood function for (2) is given by

$$p(\mathbf{b}|\mathbf{x}) \propto \exp \left( -\sum_{j=1}^{M} \{([\mathbf{A}\mathbf{x}]_j + g_j) - b_j \ln([\mathbf{A}\mathbf{x}]_j + g_j) \} \right),$$

(3)

where ‘$\propto$’ denotes proportionality. A standard approach for estimating $\mathbf{x}$ given $\mathbf{b}$ is to maximize the likelihood, or equivalently, to solve the minimization problem

$$\mathbf{x}_{\text{ML}} = \arg\min_{\mathbf{x} \geq 0} \sum_{j=1}^{M} \{([\mathbf{A}\mathbf{x}]_j + g_j) - b_j \ln([\mathbf{A}\mathbf{x}]_j + g_j) \}.$$  

(4)

This is ill-posed due to the facts that (i) $M < N$, making estimating $\mathbf{x}$ from $\mathbf{b}$ an under-determined problem, and (ii) $\mathbf{A}$ is an ill-conditioned matrix with singular values clustering at zero. The ill-posedness of (4) can be overcome using regularization, which in most cases amounts to solving, instead, an optimization problem of the form

$$\mathbf{x}_\delta = \arg\min_{\mathbf{x} \geq 0} \sum_{j=1}^{M} \{([\mathbf{A}\mathbf{x}]_j + g_j) - b_j \ln([\mathbf{A}\mathbf{x}]_j + g_j) \} + \delta J(\mathbf{x}),$$

(5)

where $\delta$ is called the regularization parameter and $J$ the regularization function. Regularization has been treated extensively in the Poisson noise case; see, e.g., [5–7, 9–11, 19, 23, 33, 35].

In the Bayesian setting, the analogue of regularization is the choice of a prior probability density $p(\mathbf{x}|\delta)$, yielding through Bayes’ Law the posterior density

$$p(\mathbf{x}|\mathbf{b}, \delta) \propto p(\mathbf{b}|\mathbf{x})p(\mathbf{x}|\delta).$$

(6)

One then obtains a regularized estimator $\mathbf{x}_{\delta}$ by computing the maximizer of $p(\mathbf{x}|\mathbf{b}, \delta)$, or equivalently, the minimizer of $-\ln p(\mathbf{x}|\mathbf{b}, \delta)$ given by

$$\mathbf{x}_{\delta} = \arg\min_{\mathbf{x} \geq 0} \sum_{j=1}^{M} \{([\mathbf{A}\mathbf{x}]_j + g_j) - b_j \ln([\mathbf{A}\mathbf{x}]_j + g_j) \} - \ln p(\mathbf{x}|\delta).$$

(7)

Note (7) and (5) are equivalent provided $p(\mathbf{x}|\delta) \propto \exp(\delta J(\mathbf{x}))$. To define $p(\mathbf{x}|\delta)$, we will use two different Markov random field (MRF) priors: a Gaussian distributed MRF yielding a Gaussian prior, and a Laplace distributed MRF yielding total variation regularization.

We emphasize that $\mathbf{x}_{\delta}$, defined by (5) or (7), is only an estimator, containing no information about the uncertainties inherent in the probability model defined by the posterior $p(\mathbf{x}|\mathbf{b}, \delta)$. The primary focus of the paper is to characterize the uncertainties inherent in (6) via sampling. In the case that both the measurement noise and prior are Gaussian, a Gaussian posterior density function results, from which we can sample directly by solving a linear system of equations (see, e.g., [4, 8, 20], in the context of large-scale linear inverse problems). In the case of Poisson measurement noise and Gaussian prior, the posterior density function $p(\mathbf{x}|\mathbf{b}, \delta)$ is a Poisson-Gaussian mixture, and no formulas exist for computing direct samples. In this paper, we introduce a Markov chain Monte Carlo (MCMC) method for sampling from $p(\mathbf{x}|\mathbf{b}, \delta)$ that makes use of a fixed Gaussian approximation of the posterior (which is based on a non-standard
Taylor series approximation of the negative-log of the posterior) as an independence proposal within a Metropolis-Hastings (MH) method. Other sampling schemes that make use of Gaussian proposals that incorporate derivative information about the negative-log of the posterior include the stochastic Newton method [24, 29] and the Hessian-preconditioned Metropolis Adjusted Langevin Algorithm (MALA) [30]. Finally, we note that in the Laplace prior case, we approximate the prior by a Gaussian, using the MAP estimator, and then sample from the resulting Poisson-Gaussian mixture. In all cases, the MCMC samples are used to obtain the sample mean (or median) estimator of \( \mathbf{x} \), as well as 95\% credibility intervals for each of the elements of \( \mathbf{x} \).

We make the assumption in this paper that our data \( \mathbf{b} \) (and hence unknown \( \mathbf{x} \)) are sufficiently positive that proposed samples are always positive, and hence, that the nonnegativity constraint (which appears in (7)) can be ignored. Fortunately, in the real-data applications that we consider in Section 4, this assumption holds, i.e., \( \mathbf{x}_{\text{MAP}} > 0 \), and our proposal density always yields samples with positive components.

The remainder of the text begins with preliminaries in Section 2, which include details on the numerical discretization of (1) and a brief discussion on the modeling of the prior probability density using MRFs. In Section 3, we present the MCMC method that is the main focus of our paper. The work concludes with various numerical experiments in Section 4 and with some closing remarks in Section 5.

2. PRELIMINARIES

In this section, we present several numerical and statistical preliminaries. We start by briefly discussing the numerical discretization of (1), providing some details that are relevant for quantitative imaging problems. We then derive the prior probability densities that we will use via Markov Random fields. In the descriptions, we focus on the one-dimensional case, in which \( \Omega \) is a finite interval in \( \mathbb{R} \); extensions to two- and higher-dimensions are either straightforward or can be found elsewhere.

2.1 Numerical Discretization of the Convolution Equation

In practice, the measurement device provides an image \( \mathbf{b} = [b_1, \ldots, b_M]^T \in \mathbb{R}^M \) and a measured or estimated discrete PSF \( \mathbf{a} = [a_{-K}, \ldots, a_0, \ldots, a_K]^T \in \mathbb{R}^{2K+1} \), both with the same uniform spatial sample rate \( h \). Thus we have measurements of \( \mathbf{b} \) on \( \Omega = [0, Mh] \) and of \( \mathbf{a} \) on \( [-Kh, Kh] \). Assuming \( \mathbf{a} = 0 \) outside of \( [-Kh, Kh] \) and \( \Omega = [-Kh, (M + K)h] \), equation (1) takes the form

\[
b(s) = \int_{-Kh}^{(M+K)h} a(s - t)x(t) \, dt, \quad 0 \leq s \leq Mh.
\]

To discretize this equation on \( \Omega = [0, Mh] \), we choose the grid \( s_i = (i - 1/2)h \), for \( i = 1, \ldots, M \), while on \( \Omega = [-Kh, (M + K)h] \), we choose \( t_j = (j - 1/2)h \), for \( j = -K + 1, \ldots, 0, \ldots, M + K \). Using midpoint quadrature then gives

\[
b(s_i) = h \sum_{j=i-K}^{i+K} a(s_i - t_j)x(t_j), \quad i = 1, \ldots, M.
\]

Assuming \( b_i = b(s_i) \), for \( i = 1, \ldots, M \); \( a_{\ell} = a(\ell h) \), for \( \ell = -K, \ldots, K \); and \( x_j = x(t_j) \), for \( j = -K + 1, \ldots, 0, \ldots, M + K \), this system of equations takes the form

\[
b_i = h \sum_{j=i-K}^{i+K} a_{i-j}x_j, \quad i = 1, \ldots, M,
\]

which can be written as the underdetermined system \( \mathbf{b} = \mathbf{Ax} \), where \( \mathbf{A} \in \mathbb{R}^{M \times N} \), \( N = M + 2K \), and \( \mathbf{x} = [x_{-K+1}, \ldots, x_0, \ldots, x_{M+K}]^T \in \mathbb{R}^N \).
Finally, it is common in practice to ‘normalize the PSF,’ which amounts to replacing $a$ by $\hat{a} \overset{\text{def}}{=} a / \sum_{\ell=-K}^{K} a_{\ell}$. To motivate, suppose we know that the PSF conserves energy, which is to say that it integrates to 1. This can be imposed by dividing both sides of (8) by $c_{h} \overset{\text{def}}{=} h \sum_{\ell=-K}^{K} a_{\ell}$, which yields

$$\hat{b}_{i} = \sum_{j=1-K}^{i+K} \hat{a}_{i-j} x_{j}, \quad i = 1, \ldots, M,$$

(9)

where $\hat{b}_{i} = b_{i}/c_{h}$. The $h$'s cancel on the right when (8) is divided by $c_{h}$, leaving the normalized PSF $\hat{a}$ defined above, but, on the left, the sample rate $h$ remains in the rescaling. Moreover, if the PSF is normalized and the original measurement vector $b$ is used, the estimates of $x$ will be incorrectly scaled (by $c_{h}$). This is not an issue in qualitative imaging, but in quantitative imaging applications, the scaling of $x$ matters, and hence, $\hat{b}$ should be used.

Finally, we note that in two- and higher-dimensions, the discretization of (1) using mid-point quadrature on a uniform computational grids proceeds similarly, and the comments in the previous paragraph regarding the normalization of the PSF remain relevant.

### 2.2 Prior Modeling using Markov Random Fields

An appropriate prior $p(x|\delta)$ can be formulated from statistical assumptions about the increments, which in one-dimension are defined $\Delta x_{i} = x_{i+1} - x_{i}$, for $i = 1, \ldots, N-1$, following the discussion in [2, 3, 31], where the reader can find more detail as well as the analogous formulations for two-dimensional signals. We assume Neumann boundary conditions on the extended FOV, so that $x_{N+1} = x_{N}$, but the formulation presented here carries over to other boundary conditions as well, such as periodic or Dirichlet.

#### 2.2.1 Gaussian distributed increments.

First, we assume that the increments are independent Gaussian distributed, i.e.,

$$\Delta x_{i} \sim N(0, (\delta w_{i})^{-1}), \quad i = 1, \ldots, N-1.$$

Note that the increment variance is spatially dependent due to the presence of the $w_{i}$’s and that an edge in the image $x$ at spatial location $i$ corresponds to a relatively small value of $w_{i}$, which leads to a relatively large increment variance $(\delta w_{i})^{-1}$. With this assumption, the prior is an intrinsic Gaussian of the form (see [2, 31] for details)

$$p(x|\delta) \propto \delta^{(N-1)/2} \exp \left( -\frac{\delta}{2} x^{T} L x \right),$$

(10)

where

$$L = D^{T} W D, \quad W = \text{diag}(w_{1}, \ldots, w_{N-1}),$$

(11)

with

$$D = \begin{bmatrix}
-1 & 1 & \cdots & \cdots & 1 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
-1 & 1 & \cdots & \cdots & 1 \\
\end{bmatrix}_{(N-1) \times N}.$$  

(12)

Note that periodic or Dirichlet boundary conditions are implemented by modifying $D$. Motivated from a pixel-level probability assumption about the increments, the Gaussian prior (10) is known as a Gaussian Markov random field [31].
2.2.2 Laplace distributed increments.

An alternative prior that also allows for the formation of edges arises if we assume independent and identically distributed (iid) Laplace increments, i.e.,

$$\Delta x_i \sim \text{Laplace}(0, \delta^{-1}), \quad i = 1, \ldots, N - 1,$$

where if $$x \sim \text{Laplace}(0, \delta)$$, then

$$p(x|\delta) \propto \exp(-\delta|x|).$$

In one-dimension, this assumption yields the prior (see [3] for details)

$$p(x|\delta) \propto \exp(-\delta\|Dx\|_1),$$

which is also the discrete total variation prior, a connection also noted in [15]. The Laplace prior, when motivated from the pixel level in this way, is known as a Markov random field.

A difficulty with the Laplace increment model is that $$-\ln p(x|\delta)$$ is not differentiable. Hence, we use the following differentiable approximation for one-dimensional signals, also frequently used for total variation (see e.g. [35]):

$$p(x|\delta) \propto \exp\left(-\delta \frac{1}{2} \sum_{i=1}^{N-1} \psi((x_{i+1} - x_i)^2)\right),$$

where $$\psi(t) = 2\sqrt{t + \beta^2}$$. For more detailed derivations, also for two-dimensional signals, see [3].

2.3 Computing $$x_\delta$$ and estimating $$\delta$$

Given the prior probability densities, the MAP estimator $$x_\delta$$ defined by (7) must be computed. In both the Gaussian and Laplace increment cases, this requires solving an optimization problem of the form

$$x_\delta = \arg\min_x \ell(x|b, \delta),$$

where

$$\ell(x|b, \delta) = \sum_{j=1}^{M} \left\{ ([Ax]_j + g_j) - b_j \ln([Ax]_j + g_j) \right\} + \delta J(x),$$

with $$\delta J(x) = -\ln p(x|\delta)$$. For this problem, we use the algorithm described in [7], which is a quasi-Newton iteration with a line search. The optimization algorithm requires both the gradient and Hessian of $$\ell(x|b, \delta)$$, which we will also make use of later in the paper and so present now. The gradient has the form

$$\nabla \ell(x|b, \delta) = A^T \text{diag}\left( \frac{1}{Ax + g} \right) (Ax - (b - g)) + \delta L(x)x,$$

where in the Guassian prior case $$L(x)$$ is the matrix defined by (11), whereas in the Laplace prior case,

$$L(x) = D^T \text{diag}(\Psi'(\|Dx\|^2))D.$$

The Hessian of $$\ell(x|b)$$ has the form

$$\nabla^2 \ell(x|b, \delta) = A^T \text{diag}\left( \frac{b}{(Ax + g)^2} \right) A.$$

To estimate the regularization/prior scaling parameter $$\delta$$, we make use of the second-order Taylor series approximation of $$\ell(x|b, \delta)$$ derived in [6], which has the form:

$$L(x|b) = L(x_\delta|b) + L_{\delta}^{\text{lin}}(x|b) + \mathcal{E},$$
where $\mathcal{E}$ denotes the error in the Taylor series approximation, and
\[
L_{\delta}^{\text{wls}}(x|b) = \frac{1}{2} \left\| \frac{Ax - (b - g)}{\sqrt{Ax_b + g}} \right\|^2,
\] (20)
or
\[
L_{\delta}^{\text{wls}}(x|b) = \frac{1}{2} \left\| \frac{Ax - (b - g)}{\sqrt{b}} \right\|^2,
\] (21)
with the vector division and square root taken component-wise. We note that both (20) and (21) yield Taylor series approximations with the same order of error and result in proposals with very similar performance. In [6], the approximation (20) is used to extend standard regularization parameter selection methods, which are appropriate for penalized least squares problems, to the Poisson negative-log likelihood case. Here, we use the discrepancy principle (DP) method presented in [6] to estimate $\delta$.

We present (19-21) because we will use the resulting Gaussian approximations of the Poisson likelihood, given by
\[
p(b|x) \approx c \cdot \exp\left(-L_{\delta}^{\text{wls}}(x|b)\right),
\] (22)
as our proposal density in the next section, where $c$ in (22) is a normalization constant.

3. A METROPOLIS-HASTINGS ALGORITHM FOR UNCERTAINTY QUANTIFICATION

In this section, we present our MCMC method for sampling from posterior density functions with Poisson likelihood and Gaussian prior, i.e.,
\[
p(x|b, \delta) \propto \exp\left(-\sum_{j=1}^{M}\{([Ax]_j + g_j) - b_j \ln([Ax]_j + g_j)\} - \frac{\delta}{2}x^TL(x_{\delta})x\right).
\] (23)
Note that when the Laplace prior $p(x|\delta)$ defined by (13) is assumed, the posterior density function does not have the form (23). In this case, we use the following Gaussian approximation: $p(x|\delta) \approx c \cdot \exp\left(-\frac{1}{2}xL(x_{\delta})x\right)$ in (23), where $c$ is a normalizing constant, $x_{\delta}$ is the MAP estimator defined by (7), and $L(x_{\delta})$ is defined by (17) for one-dimensional signals.

3.1 A Gaussian approximation of the posterior density function

The Metropolis-Hastings independence proposal for the negative-log likelihood approximation (20) (the modifications required when (21) is used instead are straightforward) is
\[
p_{x_{\delta}}(x|b, \delta) \propto \delta^{(N-1)/2}\exp\left(-\frac{1}{2} \left\| \frac{Ax - (b - g)}{\sqrt{Ax_b + g}} \right\|^2 - \frac{\delta}{2}x^TL(x_{\delta})x\right),
\] (24)
where $x_{\delta}$ is the MAP estimator. Note that this is the Gaussian approximation of the posterior density (23) obtained by replacing the Poisson likelihood (3) with the Gaussian approximation (20,22). The distribution for $p_{x_{\delta}}(x|b, \delta)$ defined by (24) is Gaussian of the form
\[
x|b, \delta \sim \mathcal{N}\left(H(x_{\delta})^{-1}A^T\left(\frac{b - g}{Ax_{\delta} + g}\right), H(x_{\delta})^{-1}\right),
\] (25)
where
\[
H(x_{\delta}) = A^T\text{diag}\left(\frac{1}{Ax_{\delta} + g}\right)A + \delta L(x_{\delta}).
\]
Standard approaches for sampling from (25) require the computation of $H(x_\delta)^{-1/2}$, which is intractable for very large-scale problems. An alternative, which is more computationally efficient, is to solve the stochastic linear system

$$H(x_\delta)x = A^T \left( \frac{b - g}{Ax_\delta + g} \right) + w, \quad w \sim \mathcal{N}(0, H(x_\delta)), \quad (26)$$

where $w$ is computed via $w = (H(x_\delta)^{1/2})^T v$, for $v \sim \mathcal{N}(0, I)$. Here, $H(x_\delta)^{1/2}$ is a matrix square root of $H(x_\delta)$, such as the Cholesky factorization, satisfying $(H(x_\delta)^{1/2})^T H(x_\delta)^{1/2} = H(x_\delta)$. (Note that we have used the result, and will again below, that if $v \sim \mathcal{N}(0, I_{q \times q})$, $B \in \mathbb{R}^{p \times q}$ has rank $p$, and $w = Bv$, then $w \sim \mathcal{N}(0, BB^T)$.) Unfortunately, $H(x_\delta)^{1/2}$ may not be easily computed, in which case we can take advantage of the structure of $H(x_\delta)$. Specifically,

$$w = A^T \text{diag} \left( \frac{1}{\sqrt{Ax_\delta + g}} \right) v_1 + \sqrt{\delta} D^T W^{1/2} v_2,$$

where $v_1 \sim \mathcal{N}(0, I_{M \times M})$, $v_2 \sim \mathcal{N}(0, I_{N-1 \times N-1})$. A similar structure can be exploited in two-dimensions; see [8] for details.

### 3.2 Sampling from $p(x|b, \delta)$ using Metropolis-Hastings

We now present our first MCMC method, which uses the approximate Gaussian density $p_{xkl}(x|b, \delta)$, defined by (25), as an independence proposal distribution within a Metropolis-Hastings (MH) algorithm. The MH correction of the samples from $p_{xkl}(x|b, \delta)$ yields theoretically correct samples from the full posterior $p(x|b, \delta)$ defined by (23). Note, again, that in the Laplace prior case, we are using a Gaussian approximation of the prior to obtain an approximate posterior density function of the form (23).

**MH Algorithm for sampling from $p(x|b, \delta)$ defined by (23)**

0. Compute $x_\delta$. Set $k = 0$ and $x_0 = x_\delta$.

1. Compute a candidate sample $x_* \sim p_{xkl}(x|b, \delta)$ by solving (26) for a given $w$.

2. Compute the acceptance ratio

$$r = \min \left\{ \frac{p(x_*|b, \delta) p_{xkl}(x_\delta|b, \delta)}{p(x_\delta|b, \delta) p_{xkl}(x_*|b, \delta)} \right\} \quad (27)$$

and a random draw $u \sim U(0, 1)$. If $u < r$, set $x_{k+1} = x_*$, else set $x_{k+1} = x_k$.

4. Set $k = k + 1$ and return to Step 1.

**Remarks:**

i. Note that since $p(x|b, \delta)$ and $p_{xkl}(x|b, \delta)$ have the same prior, the priors cancel in (27). Indeed, this is why we use the Gaussian approximation in the Laplace prior case; otherwise, the priors don’t cancel and the computation of (27) becomes unstable. Moreover, we have found it to be numerically advantageous to use the following equivalent definition of (27): taking into account the cancelation of the priors:

$$r = \min \left\{ 1, e^{-c} \right\}, \quad (28)$$

$$c = -\ln p(b|x_\delta) + \ln p(b|x_k) - L^\text{ws}(x_k|b) + L^\text{ws}(x_*|b).$$

ii. It is possible to sample from the local Gaussian $p_{xkl}(x|b, \delta)$, instead of $p_{xkl}(x|b, \delta)$, in Step 1, but this requires a modification of the prior ratio in (27) from $p_{xkl}(x_k|b, \delta) / p_{xkl}(x_*|b, \delta)$ to $p_{xkl}(x_k|b, \delta) / p_{xkl}(x_*|b, \delta)$. A major drawback of this approach, as for the SN sampler, is that since the proposals are different, the normalization constants do not cancel, making this approach more computationally demanding.

Volume 1, Number 1, 2015
3.3 Hierarchical Modeling and Sampling $\delta$

In the case that the original prior is Gaussian, we can assume a probability model, or hyper-prior, on $\delta$, and sample $\delta$ in addition to $x$. When $\delta$ is the scaling parameter for a Gaussian inverse-covariance matrix, as it is here, it is standard to assume that it is Gamma-distributed [14], i.e.,

$$p(\delta) \propto \delta^{\alpha-1} \exp(-\beta \delta).$$ \hspace{1cm} (29)

We choose the parameter values $\alpha = 1$ and $\beta = 10^{-4}$, since then (29) is essentially flat over the domain of possible values of $\delta$. Given (29), Bayes’ law yields the posterior density function

$$p(x, \delta | b) \propto p(b | x)p(x | \delta)p(\delta)$$

$$= \delta^{(N-1)/2+\alpha-1} \exp \left( - \sum_{j=1}^{M} \left( (|Ax|_j + g_j) - b_j \ln(|Ax|_j + g_j) \right) \right)$$

$$\times \exp \left( - \frac{\delta}{2} x^T L x - \beta \delta \right).$$ \hspace{1cm} (30)

The conditional density for $x | b, \delta$ then has the form (23), with $L(x_b) = L$, and hence, we can use $p_{x_k}(x | b, \delta)$, as defined in (24), as an independence proposal with a MH correction. The conditional density for $\delta | b, x$, on the other hand, is Gamma distributed with probability density function

$$p(\delta | b, x) \propto \delta^{(N-1)/2+\alpha-1} \exp \left( - \left( x^T L x / 2 + \beta \right) \delta \right),$$ \hspace{1cm} (31)

from which we can easily sample using standard software. Combining $p(x | b, \delta)$ and $p(\delta | b, x)$ within a Metropolis-within-Gibbs sampling framework, we obtain the following method.

**Metropolis-within-Gibbs Algorithm for sampling from $p(x, \delta | b)$**.

0. Compute $x_0$. Set $k = 0$, $\delta_0 = \delta$, and $x_0 = x_0$.

1. Compute a candidate sample $x_\ast \sim p_{x_k}(x | b, \delta_k)$ by solving (26).

2. Compute the acceptance ratio

$$r = \min \left\{ 1, \frac{p(x_\ast | b, \delta_k) p_{x_k}(x_\ast | b, \delta_k)}{p(x_k | b, \delta_k) p_{x_k}(x_k | b, \delta_k)} \right\}$$ \hspace{1cm} (32)

and a random draw $u \sim U(0, 1)$. If $u < r$, set $x_{k+1} = x_\ast$, else set $x_{k+1} = x_k$.

3. Compute a conditional sample $\delta_{k+1} \sim p(\delta | b, x_{k+1})$ defined by (31).

4. Set $k = k + 1$ and return to Step 1.

**Remarks:**

i. The behavior of the analogous hierarchical sampling scheme in the Gaussian measurement error case is studied in [1], where it is found that as $N$ gets larger (and the discretization finer), the performance of the sampling scheme degrades. The above Metropolis-within-Gibbs method for sampling from (30) is very similar, and so likely also degrades in the infinite limit.

ii. Just as above, we compute (32) using an expression analogous to (28).

iii. Also just as above, the local Gaussian $p_{x_k}(x | b, \delta_k)$ can be used in Step 1, and this requires the same change to the proposal ratio in (32) that was mentioned in the previous Remark (ii). Moreover, just as before, if local Gaussian proposals are used, the normalization constants do not cancel, making this a more computationally demanding approach.
iv. Finally, if the Laplace prior is used, one can fix $L = L(x_\delta)$ defined in (17) at the outset, where $x_\delta$ is the MAP estimator (7) for the original value of $\delta$. However, as new values of $\delta$ are sampled, the question of whether $L(x_\delta)$ and $x_\delta$ should change as well is relevant. We do not address this question here, using the hierarchical sampling method in the numerical experiments only when the prior is Gaussian.

4. NUMERICAL RESULTS

4.1 Image Deblurring: One- and Two-Dimensional Synthetic Test Cases

We begin with a synthetic one-dimensional deconvolution problem of the form (1) with $\tilde{\Omega} = [0, 1]$ and $\Omega = [21/120, 100/120]$. We follow Section 2.1 in our discretization with Gaussian kernel $a(s) = \exp(-s^2/(2\gamma^2))/\sqrt{2\pi\gamma^2}$ and $n = 120$ grid points, yielding an $80 \times 120$ linear system $b = Ax$. The true image $x$ used to generate the data is plotted in Fig. 1 (a), as is the data $b$ generated using the Poisson noise model (2) with a noise strength that is approximately 3.5% that of the signal strength. We compute 5 parallel chains of the Metropolis-within-Gibbs algorithm, each with 5000 MCMC samples, and used the last half of each chain for our analysis. The acceptance rate for our proposal was approximately 50%. In Fig. 1 (b), the sample median and 95% credibility bounds are plotted; note that the credibility bounds become much wider where the model is unspecified and the prior dominates, i.e. $0 \le t < 1/6$ and $5/6 < t \le 1$. Fig. 1 (c) shows a histogram of the $\delta$ samples. Finally, we note that we also implemented the stochastic Newton algorithm [24] on this example and it performed poorly, exhibiting extremely low acceptance rates.

Next we test the algorithm on a two-dimensional deconvolution problem, assuming model

$$b(s, t) = \int_0^1 \int_0^1 a(s - s', t - t')x(s', t')ds'dt', \quad 0 \le s, t \le 1,$$

FIG. 1: (a) The one-dimensional image used to generate the data and blurred noisy data, (b) plots of the true image, sample median, and 95% credibility bounds computed from the samples, and (c) a histogram of the $\delta$ samples.
discretized using mid-point quadrature on a $128 \times 128$ uniform computational grid over $[0,1] \times [0,1]$. The $128^2 \times 1$ vectors $x$ and $b$ can be created from $128 \times 128$ image arrays $X$ and $B$, yielding a $128^2 \times 128^2$ system of linear equations $b = Ax$. For computational simplicity, we assume that $x$ is a periodic function on $[0,1] \times [0,1]$, so that $A$ has block circulant with circulant block structure and can be diagonalized by the 2D-DFT, for low storage requirements and efficient matrix-vector multiplication. The true image $X$ used to generate the data is plotted in Fig. 2 (a), and the blurred, noisy data $B$ is shown in (b). The data was generated using noise model (2) with a noise strength that is approximately 5% that of the signal strength. For estimation and uncertainty quantification, we compute 2 parallel chains, each with 3000 MCMC samples, and we use the last half of each chain for our analysis. Again, the acceptance rate for our proposal was approximately 50%. Fig. 2 (c) gives the sample mean, and (d) shows the pixel-wise standard deviation. Fig. 2 (e) is a histogram of the $\delta$ samples.

4.2 Results for Synthetic and Real Deconvolution Examples in X-ray Radiography

The sampling schemes detailed above are appropriate for any application modeled as a Fredholm operator equation that satisfies the positivity requirements of the Poisson formulation. In this section we focus on an application in X-ray radiography that is fundamental to diagnostic imaging in the security sciences. We note that for the real data test cases, the signals are one dimensional but still quite large-scale.

Most applications of imaging are qualitative, in that the goal is to produce image reconstructions that look as good as possible. In the security sciences, however, pulsed power X-ray radiography is a quantitative imaging diagnostic used to calculate the true locations of image features in 3D space or to calculate the densities of the objects being imaged [22, 26, 32, 36]. A typical setup is shown in Fig. 3, where a pulsed power, flash X-ray source generates X-rays that are pulsed through the scene, some of which are absorbed by the objects in the scene and some of which are transmitted. The transmitted photons are absorbed by a scintillator, which approximately linearly emits visible photons in response to the absorbed X-rays. The visible photons are focused through a lens system and collected on a CCD array, resulting in a radiograph. For such systems, it is often possible to accurately determine the sources and
magnitude of noise, and such images are often dominated by Poisson noise due to particle counting at the scintillator and on the CCD. In this case, it is natural to use the Poisson formulation.

Some high-energy pulsed power systems suffer from the drawback of not generating the exact same X-ray spectrum on every pulse, which means that the X-ray transmission through the scene will change from shot to shot, i.e., intensity values will systematically change from shot to shot when imaging objects of the same density. (This would not be the case if the spectrum and intensity were pulse-invariant.) In order to calculate the X-ray transmission function for a given radiograph, it is common to include a calibration object with known material properties, such as the step wedge shown in Fig. 3.

In the following subsections we demonstrate the deconvolution sampling scheme detailed above on synthetic examples meant to mimic the imaging of calibration step wedges, as well as on real radiographs of two different step wedges imaged by a U.S. Department of Energy high-energy, flash X-ray radiography system. We conclude the section with an example of deconvolving an X-ray radiograph using the Gaussian prior and hierarchical sampling scheme presented in Section 3.3.

4.2.1 Results with Synthetic Data

The first example is a synthetic binary edge, with lower intensity value of 15 and upper intensity value of 1015. The edge is blurred with a Gaussian kernel with $\sigma = 15$ pixels. The edge, the blurred edge, and the Poisson MAP estimate with Laplace MRF prior are shown in Fig. 4 (a). Given the Poisson MAP estimate, 10,000 MH samples were computed, with the first 5000 being discarded for burn-in. The true edge, the blurred edge, and the median of the final 5000 Metropolis-Hastings samples are shown in Fig. 4 (b), with the gray shaded region giving a 95% credibility band. No noise is added to the blurred edge; this example is merely shown to demonstrate that the sampling scheme
FIG. 4: (a) A binary edge (solid), blurred with a Gaussian PSF (dotted), and the Poisson MAP reconstruction (dash-dotted). (b) The binary edge (solid), blurred edge (dotted), and the median of 5000 MH samples (dashed) with 95% credibility bands (shaded region).

FIG. 5: (a) A step wedge simulated with a radiation transport code with a rolled edge (top black region) for computing the point spread function. (b) The PSF computed from the rolled edge.
Fig. 6 (a) A horizontal line out (cross section) through the step wedge simulated with MCNP6, along with the Poisson MAP estimate, and (b) zoomed in on one of the steps. (c) The same step wedge with the median of the MH samples (after burn-in) and (d) the same zoomed in region.

accurately reproduces uncontaminated edges.

Fig. 5 (a) shows a synthetic radiograph generated with the Monte Carlo-based radiation transport code MCNP6 [16]. A Gaussian X-ray source is simulated to emit photons through a scene containing a tungsten (W) step wedge (bottom center of the image), an object similar to the cartoon object shown in Fig. 3. The dark top region of Fig. 5 (a) is a so-called rolled edge, which is simply a thick object with an edge that is rolled off, rather than sharp. Though the data is synthetic, this is not an “inverse crime” in the sense of [21], as the data is generated using a forward model (Monte Carlo simulation) that is different from the convolution model used for inversion. Because it is a simulation, we do know the shape of the X-ray source and the cone of X-ray propagation. Nonetheless, we do not know the “correct answer” for the deconvolution, since the intensity in the simulated image is a function of how many particles are run in the simulation and their stochastic properties. The PSF used in the deconvolution is computed from horizontal edge between the white and black space [13] and is shown in Fig. 5 (b).

A horizontal cross section through the step wedge is shown in Fig. 6 (a), along with the Poisson MAP estimate of the deconvolved edge, again with Laplace MRF prior. Given the MAP estimate, 10,000 MH samples are computed, with the first 5,000 discarded for burn-in. Fig. 6 (c) shows the cross-section with the median of the MH samples and a gray shaded region corresponding to a 95% credibility band. Images (b) and (d) show the same as (a) and (c), respectively, zoomed in one of the jump discontinuities. The median reconstruction sharpens the signal as would be expected given the size of the PSF, and the credibility bands give tight error estimates.
4.2.2 Edge Reconstruction in X-ray Radiography

Fig. 7 (a) shows a radiograph (log scale) from the Cygnus dual-beam radiographic imaging facility at the U.S. Department of Energy’s Nevada National Security Site [32]. Cygnus is a 2.25 MeV end-point energy pulsed power X-ray source with a rod-pincher diode [26, 28]. This image shows a tungsten (W) step wedge, which consists of piecewise constant steps of different thicknesses; thicker regions of the object correspond to lower intensity values (due to greater X-ray attenuation) in the image. The line-out must be de-blurred before transmission can be accurately determined; see [18] for a simulation-based approach to computing transmission curves. Experimental approaches for computing transmission of high-energy X-ray systems are largely undocumented in the literature. Fig. 7 (b) shows the point spread function used in deconvolution, assuming circular symmetry.

A horizontal line-out from the image is shown in Fig. 8 (a), along with the Poisson-TV MAP estimate with Laplace prior. Again 10,000 samples were computed, using δ = 0.1, and the last half were used for analysis. Plot (c) shows the original data, the median of the MH samples, and 95% credibility bands. Images (b) and (d) show the same as (a) and (c), respectively, zoomed in on the right-most jump discontinuity.

**Remark:** Note that the deblurred signals in Fig. 8 (b),(d) are not as sharp as one might expect, given that the object has sharp edges. Nonetheless, the algorithm has performed as it should, as the width of the PSF is such that deconvolution of this PSF can only sharpen jumps on the order of approximately 10 pixels, i.e. a blurred jump with width 10 pixels could be sharpened to a jump with width 1 pixel, and plot (d) shows that the jumps have been sharpened on the order of 10 pixels. This particular step wedge, however, has jumps with widths much greater than 10 pixels, which cannot be deblurred completely with the measured PSF. This excess blur is caused by cone beam effects and X-ray scatter – not by scintillator or optical blur – and deconvolving it out of the image requires a different procedure for measuring the system response (PSF).

Fig. 9 (a) shows a composite target with several objects, each used for a different system calibration at the Cygnus facility. The X-ray beam is contained within a circular collimator, which is why the image features are contained within the black circle around the outside. The PSF for this system, shown in image (c), is computed using the “L” rolled edge (backwards Pac-man) in the lower-left of the image, using the same technique as is cited above. In this application we are interested in the tantalum step wedge on the left side of the image, a closeup of which is shown in Fig. 9 (b).

Fig. 10 (a) shows a one-dimensional vertical cross section through the step wedge, along with the Poisson MAP estimate with Laplace prior and δ = 0.25. Again, 10,000 MH samples were computed, with the first 5,000 discarded for burn-in. Image (c) shows the same line out with the median of the MH samples, and 95% credibility bands shaded in gray. Images (b) and (d) show the same as (a) and (c), respectively, zoomed in on one of the jump discontinuities.

4.2.3 Hierarchical Reconstruction with X-ray Radiographs

The final results we present are on the deconvolution of a signal that does not have jump discontinuities, so the hierarchical sampling scheme of Section 3.3 with Gaussian MRF prior is appropriate. The data, shown in Fig. 11, is from the central cylinder object in the upper middle section of Fig. 9 (a). The PSF used is the same as that for the step wedge. Fig. 12 (a) shows the Gaussian MRF prior MAP estimate, which results in a deblurred signal with noise largely removed as an effect of the smoothing prior.

Given the MAP estimate, 2000 hierarchical MH samples were computed, with the first 1000 being discarded for burn-in. Fig. 12 (c) shows the line out with the median of 1000 MH samples, and images (b) and (d) show the same as (a) and (c), respectively, zoomed in around the lowest point on the curve. Since the prior is not jump discontinuity-enhancing, the reconstruction is not as sharp as the Laplace prior reconstructions, but that is also undesirable in this case, since the signal does not have jump discontinuities.

Fig. 13 (a) and (b) show the time series and histogram of the 2000 δ samples. The time series does not mimic Gaussian white noise, but it is, nonetheless, quite stationary. The histogram shows that the δ samples are taken from a Gamma distribution that is nearly Gaussian with quite a small variance, showing that, even with only 2000 samples computed, the MCMC chain has already reached near convergence.
FIG. 7: (a) A radiograph (log-scale) of a step wedge calibration object from the Cygnus pulsed power X-ray facility at the Nevada National Security Site, along with (b) the measured point spread function.

FIG. 8: (a) A horizontal line out (cross section) through the real tungsten step wedge imaged at the Cygnus X-ray radiography facility, along with the Poisson Lagged Diffusivity estimate, and (b) zoomed in on one of the steps. (c) The same step wedge with the median of the MH samples (after burn-in) and (d) the same zoomed in region.
FIG. 9: (a) A composite target used for several different X-ray imaging system calibrations. The black portion around the outside is due to collimation of the X-ray beam. (b) Zoomed in on the Tantalum step wedge with 10 different thicknesses, used for the calculation of transmission curves. (c) The PSF computed from the composite image in (a), with a relative intensity scale on the vertical axis (normalized so that the area under the PSF is 1). The full width at half maximum (FWHM, a common method for measuring the width of the PSF) is approximately 8.9 pixels.

FIG. 10: (a) A vertical line out through the Tantalum step wedge shown in Fig. 9 (b), and the PLD reconstruction. The same line out and reconstruction, zoomed in around a couple of the edges. (c) The line out with the median of the Metropolis-Hastings samples, along with 95% credibility bands, and (d) zoomed in.
FIG. 11: (a) A zoom in on the central cylinder in Fig. 9 (a), along with (b) a horizontal line out from the central axis of symmetry.

FIG. 12: (a) A line out from the concentric cylinders, along with the Gaussian MRF MAP reconstruction, and the (b) same zoomed in around the lowest point. (c) The same line out with the median of the MH samples, and (d) zoomed in at the same point.
FIG. 13: (a) The time series plot of the $\delta$ samples from the hierarchical sampling, and (b) the histogram of the $\delta$ samples.

5. CONCLUSIONS

In this work, we have presented an MCMC method for sampling from posterior density functions arising in linear inverse problems with Poisson likelihood and Gaussian prior. The method uses a Gaussian approximation of the Poisson likelihood to construct a Gaussian proposal for use within a Metropolis-Hastings method. We derive our priors using a Markov random field (MRF) approach with a focus on Gaussian and Laplace MRF’s. In the Laplace MRF case, the prior is non-Gaussian, but we use the MAP estimator to obtain an accurate Gaussian approximation of the prior for the sampling step. In the Gaussian prior case, no approximation of the prior is needed, and a hierarchical, Metropolis-within-Gibbs sampling scheme is presented, which also samples the prior scaling parameter $\delta$; its primary advantage is that it does not require the regularization parameter to be chosen ahead of time. We present results on synthetic data for both one- and two-dimensional signals, as well as several results on one-dimensional real data examples.

These kinds of approaches are especially important in quantitative imaging, such as X-ray radiography in the security sciences, since we require not only reconstructions of the unknown but also estimates of the uncertainties associated with the reconstruction. To this end, the developed approach was demonstrated on synthetic data that mimics the kind of images we use for X-ray source calibration, and then we extended the results to real X-ray radiographs taken at the U. S. Department of Energy’s Nevada National Security Site. In all cases, the deconvolution approach presented here is an effective technique for deblurring the measured imagery.

ACKNOWLEDGMENTS

This work was supported by the Site-Directed Research and Development Program of National Security Technologies, LLC, under Contract No. DE-AC52-06NA25946 with the U.S. Department of Energy. 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 form of this manuscript, or allow others to do so, for United States Government purposes. DOE/NV/25946–2042.

REFERENCES

15. M. Green, Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising, and an improved denoising algorithm, CAM Report 02-55, UCLA, October 2002.


