A METROPOLIS-HASTINGS METHOD FOR LINEAR INVERSE PROBLEMS WITH POISSON LIKELIHOOD AND GAUSSIAN PRIOR*  
JOHNATHAN M. BARDISLEY AND AARON LUTTMAN  

Abstract. 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 method. 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, image deblurring, Bayesian inference, Metropolis-Hastings.  

AMS subject classifications. 15A29, 44A35, 62H35, 60J10  

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,  
\begin{equation}
\label{eq:1}
b(s) = Ax(s) \overset{\text{def}}{=} \int_{\Omega} a(s-t)x(t)dt, \quad s \in \Omega,
\end{equation}

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 $\Omega = [-\xi, 1+\xi]$. We will focus on one-dimensional signals in this paper.  

Since $\Omega \subseteq \bar{\Omega}$, discretizing (1.1) yields a non-square linear system of equations $b = Ax$, where $b \in \mathbb{R}^M$, $x \in \mathbb{R}^N$, and $A \in \mathbb{R}^{M \times N}$, with $N > M$. We will provide more detail on the discretization of (1.1) below. In practice, the values contained in $b$ correspond to measurements, and we make the assumption that the measurement  

\*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
noise is Poisson distributed, which is a reasonable assumption in many quantitative imaging applications (see Section 4). Specifically, given a known background \( g \in \mathbb{R}^M \), we assume the statistical model

\[
(1.2) \quad b = \text{Poiss}(Ax + g),
\]

where \( \text{Poiss}(v) \) denotes an independent Poisson random vector with Poisson parameter vector (mean and variance) \( v \). The likelihood function for (1.2) is given by

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

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

\[
(1.4) \quad x_{ML} = \arg\min_{x \geq 0} \sum_{j=1}^{M} \left\{ \ln([Ax]_j + g_j) - b_j \ln([Ax]_j + g_j) \right\}.
\]

Unfortunately, (1.4) is an ill-posed optimization problem due to the facts that (i) \( M < N \), making estimating \( x \) from \( b \) an under-determined problem, and (ii) \( A \) is an ill-conditioned matrix with singular values clustering at zero. The ill-posedness of (1.4) can be overcome using regularization, which has been treated extensively in the literature in the Poisson noise case when \( M = N \); see, e.g., [5, 6, 7, 10, 12]. In the extended FOV case, when \( \Omega \subset \subset \Omega \) in (1.1) and \( M < N \), computing regularized solutions for Poisson data was treated in [11], where regularization was implemented via the early truncation of Richardson-Lucy iterations. Here we implement regularization, instead, via the choice of a prior probability density function, or, equivalently, a regularization function. This is the first such work in the case that \( \Omega \subset \subset \Omega \) in (1.1) and \( M < N \).

A classical inverse problems approach to overcoming the difficulties with (1.4) is to add a regularization function of the form \( \delta J(x) \) – where \( \delta \) is a scaling parameter – to the negative-log likelihood in (1.4), to obtain the well-posed optimization problem

\[
(1.5) \quad x_{\delta} = \arg\min_{x \geq 0} \sum_{j=1}^{M} \left\{ \ln([Ax]_j + g_j) - b_j \ln([Ax]_j + g_j) \right\} + \delta J(x).
\]

In the Bayesian setting, on the other hand, one chooses a prior probability density \( p(x|\delta) \) and constructs the posterior density function through Bayes’ Law:

\[
(1.6) \quad p(x|b, \delta) \propto p(b|x)p(x|\delta).
\]

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

\[
(1.7) \quad x_{\delta} = \arg\min_{x \geq 0} \sum_{j=1}^{M} \left\{ \ln([Ax]_j + g_j) - b_j \ln([Ax]_j + g_j) \right\} - \ln p(x|\delta).
\]

Note (1.7) and (1.5) are equivalent provided \( p(x|\delta) \propto \exp (-\delta J(x)) \). To define \( p(x|\delta) \), we will use two different Markov random field (MRF) priors: a Gaussian distributed...
MRF yielding a Gaussian prior, and a Laplace distributed MRF yielding a regularization similar to total variation.

We emphasize that $x_\delta$, defined by (1.5) or (1.7), is only an estimator, containing no information about the uncertainties inherent in the probability model defined by Bayes’ Law (1.6). The primary focus of the paper is to characterize the uncertainties inherent in (1.6) 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, 19], in the context of linear inverse problems). In the case of Poisson measurement noise and Gaussian prior, the posterior density function $p(x|b, \delta)$ is a Poisson-Gaussian mixture and no formulas exist for computing direct samples. In this paper, we introduce a method for sampling from $p(x|b, \delta)$ in such cases. It uses the MAP estimator $x_\delta$ to construct a Gaussian approximation of $p(x|b, \delta)$, which is in turn used as an independence proposal within a Metropolis-Hastings (MH) method, yielding a Markov chain Monte Carlo (MCMC) sampling scheme. In the Laplace prior case, after the MAP estimator $x_\delta$ has been computed, we use it to construct a Gaussian approximation of the prior. The resulting approximate posterior is a Poisson-Gaussian mixture, and hence we can apply our MCMC method. The problem of sampling from the full posterior in the Laplace prior (total variation regularization) case is quite challenging and is left for a later work. In all cases, the MCMC samples are used to obtain the sample mean estimator of $x$, as well as 95% credibility intervals for each of the elements of $x$.

Throughout the paper, we focus on one-dimensional signals, where our MCMC sampling scheme works well. We note that the applications and real data that we analyze in the numerical experiments are all one-dimensional signal reconstruction problems. The most natural extension of our MCMC method to two-dimensional signal reconstruction is, in principle, simple, but we have found it to be numerically unstable and requiring further work.

The remainder of the text begins with preliminaries in Section 2, which include details on the numerical discretization of (1.1), on the modeling of the prior probability density using MRFs, on the optimization algorithm for computing the MAP estimator, and on methods for estimating the regularization/prior scaling parameter $\delta$. 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 techniques and algorithms needed later. We start by briefly discussing the numerical discretization of (1.1). We then derive the required prior probability densities via Markov Random fields. In both of these descriptions, we focus on the one-dimensional case, in which $\Omega$ is a finite interval in $\mathbb{R}$. Next, we present the numerical optimization method that is needed for computing the MAP estimator $x_\delta$ defined by (1.7), and we briefly discuss the method used for estimating the regularization (or prior scaling) parameter $\delta$.

2.1. Numerical Discretization of the Convolution Equation. In practice, the measurement device provides a discrete PSF $a = [a_{-K}, \ldots, a_0, \ldots, a_K]^T \in \mathbb{R}^{2K+1}$ and an image $b = [b_1, \ldots, b_M]^T \in \mathbb{R}^M$, both with the same uniform spatial sample rate $h$. Thus we have measurements of $b$ on $\Omega = [0, Mh]$ and of $a$ on $[-Kh, Kh]$. Assuming $a = 0$ outside of $[-Kh, Kh]$ and $\hat{\Omega} = [-Kh, (M + K)h]$, equation (1.1)
takes the form
\[ b(s) = \int_{-K}^{(M+K)h} a(s-t)x(t)\, dt, \quad 0 \leq s \leq Mh. \]

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 \( \hat{\Omega} = [-Kh, (M + K)h] \), 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
\[ (2.1) \quad b_i = h \sum_{j=i-K}^{i+K} a_{i-j}x_j, \quad i = 1, \ldots, M, \]
which can be written \( \mathbf{b} = \mathbf{A}\mathbf{x} \), where \( \mathbf{A} \in \mathbb{R}^{M+2K \times N} \), \( N = M + 2K \), and \( \mathbf{x} = [x_{-K+1}, \ldots, x_0, \ldots, x_{M+K}]^T \in \mathbb{R}^{M+2K} \). Note that this is an under-determined system. The discretization of (1.1) in two or higher dimensions is similar.

Finally, it is common in practice to ‘normalize the PSF’, which amounts to replacing \( \mathbf{a} \) by \( \hat{\mathbf{a}} \equiv \mathbf{a}/\sum_{\ell=-K}^{K} a_\ell \). To motivate, suppose we know that the PSF preserves energy, which is to say that it integrates to 1. We can impose this constraint by dividing both sides of (2.1) by \( c_h \equiv h \sum_{\ell=-K}^{K} a_\ell \), which yields
\[ (2.2) \quad \hat{b}_i = \sum_{j=i-K}^{i+K} \hat{a}_{i-j}x_j, \quad i = 1, \ldots, M, \]
where \( \hat{b}_i = b_i/c_h \). Note that the \( h \)'s cancels on the right when (2.1) is divided by \( c_h \), leaving the normalized PSF \( \hat{\mathbf{a}} \) 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 \( \mathbf{b} \) is used, the estimates of \( \mathbf{x} \) will be incorrectly scaled (by \( c_h \)). This is not an issue in qualitative imaging, but in quantitative imaging applications, the scaling of \( \mathbf{x} \) matters, and hence, \( \hat{\mathbf{b}} \) should be used.

**2.2. Prior Modeling using Markov Random Fields.** An appropriate prior \( p(\mathbf{x}|\delta) \) 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, 26], 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.,
\[ \Delta x_i \sim \mathcal{N}(0, (\delta w_i)^{-1}), \quad i = 1, \ldots, N-1. \]
Note that the increment variance is spatially dependent due to the presence of the \( w_i \)'s and that an edge in the image \( \mathbf{x} \) 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, 26] for details)

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

(2.3)

where

$$L = D^T W D, \quad W = \text{diag}(w_1, \ldots, w_{N-1}),$$

(2.4)

with

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

(2.5)

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 (2.3) is known as a Gaussian Markov random field [26]. For a detailed derivation of the GMRF prior in two-dimensions, see [2].

### 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

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

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

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

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. [29]):

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

(2.6)

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

### 2.3. Numerical Optimization Algorithm for computing the MAP Estimator.

Next, we describe the numerical optimization algorithm for computing the MAP estimator $x_d$ defined by (1.7). In both the Gaussian and Laplace increment cases, we need to solve an optimization problem of the form

$$x_d = \arg\min_{x \geq 0} \ell(x|b, \delta),$$

(2.7)
Poisson Sampling

where

\begin{equation}
(2.8) \quad \ell(x|b, \delta) = \sum_{j=1}^{M} \left\{ (Ax_j + g_j) - b_j \ln((Ax_j + g_j)) \right\} + \delta J(x),
\end{equation}

with \( \delta J(x) = -\ln p(x|\delta) \). The gradient of \( \ell(x|b, \delta) \) is given by

\begin{equation}
(2.9) \quad \nabla_x \ell(x|b, \delta) = \mathbf{A}^T \text{diag} \left( \frac{1}{Ax + g} \right) (Ax - (b - g)) + \delta \mathbf{L}(x)x.
\end{equation}

In the Gaussian increment case, for one-dimensional signals, \( \mathbf{L}(x) \) is the matrix defined by (2.4), whereas in the Laplace prior case,

\begin{equation}
(2.10) \quad \mathbf{L}(x) = \mathbf{D}^T \text{diag}(\psi'((Dx)^2)) \mathbf{D}.
\end{equation}

The extension to two-dimensional signals is straightforward [3].

Analogous to the lagged-diffusivity (LD) fixed point iteration [29], we can build a fixed point iteration from \( \nabla_x \ell(x|b) = 0 \) by lagging \( x \) to be equal to \( x_k \) in (2.9), except where it appears in the linear terms, and then solving for \( x \) to obtain:

\begin{equation}
(2.11) \quad x_{k+1} = \left( \mathbf{A}^T \text{diag} \left( \frac{1}{Ax_k + g} \right) \mathbf{A} + \delta \mathbf{L}(x_k) \right)^{-1} \mathbf{A}^T \left( \frac{b - g}{Ax_k + g} \right).\nonumber
\end{equation}

Iteration (2.11) can be alternatively written in the form of a quasi-Newton iteration:

\begin{equation}
(2.12) \quad x_{k+1} = x_k - \mathbf{H}(x_k)^{-1} \nabla \ell(x_k|b),
\end{equation}

where

\begin{equation}
\mathbf{H}(x_k) = \mathbf{A}^T \text{diag} \left( \frac{1}{Ax_k + g} \right) \mathbf{A} + \delta \mathbf{L}(x_k).
\end{equation}

We call this method lagged Poisson (LP) fixed point iteration, and to our knowledge it has not appeared elsewhere in the literature.

In several numerical experiments, (2.11) has been found to be convergent, suggesting that as with LD [29], LP may be a convergent fixed point iteration under certain restrictions. However, we do not pursue a convergence analysis here. To make (2.11) a globally convergent method for general problems with a nonnegativity constraint, a projected line search can be implemented, as described in the appendix.

2.4. Regularization Parameter Selection. Finally, we describe how to estimate the regularization/prior scaling parameter \( \delta \). Let \( L(x|b) \) \( \overset{\text{def}}{=} -\ln p(b|x) \), where \( p(b|x) \) is the Poisson likelihood function defined by (1.3). Then in [6], the following Taylor series approximation is derived:

\begin{equation}
(2.13) \quad L(x|b) = L(x_0|b) + L_{\delta}^{\text{wls}}(x|b) + \mathcal{E},\nonumber
\end{equation}

where \( \mathcal{E} \) denotes the error in the Taylor series approximation, and

\begin{equation}
(2.14) \quad L_{\delta}^{\text{wls}}(x|b) = \frac{1}{2} \frac{\| \mathbf{Ax} - (b - g) \|}{\sqrt{\| \mathbf{Ax} + g \|}}^2,
\end{equation}

with the vector division and square root taken component-wise.
In [6], the approximation (2.14) is used to extend standard regularization parameter selection methods, which are appropriate for penalized least squares problems, to the Poisson negative-log likelihood case. For example, the discrepancy principle (DP) presented in [6] chooses the value of \( \delta \) satisfying

\[
L^\text{wls}_\delta(x_\delta | \b) = \frac{1}{2} \left\| \frac{Ax_\delta - (b - g)}{\sqrt{Ax_\delta + g}} \right\|^2 = \frac{M}{2},
\]

where \( x_\delta \) is the MAP estimator defined by (1.7). In [6], \( L^\text{wls}_\delta(x | \b) \) is also used to extend the methods of generalized cross validation and unbiased predictive risk estimation for regularization parameter selection to the Poisson negative-log likelihood case. In the numerical experiments below, we implement DP by choosing \( \delta \) to satisfy

\[
\delta_{\text{DP}} = \arg\min_{\delta \geq 0} \left( \frac{2}{M} L^\text{wls}_\delta(x_\delta | \b) - 1 \right)^2.
\]

Note that (2.16) agrees with (2.15) when (2.15) is solvable with respect to \( \delta \).

Before continuing, we note that (2.13, 2.14) also suggest the following Gaussian approximation of Poisson likelihood function:

\[
p(x | \b) \approx c \cdot \exp \left( -\frac{1}{2} x^T L(x_\delta) x \right),
\]

where \( c \) is a normalizing constant. This approximation will be used in the Metropolis-Hastings algorithm presented next.

**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.,

\[
p(x | b, \delta) \propto \exp \left( -\sum_{j=1}^{M} \left\{ [Ax_j + g_j] - b_j \ln([Ax_j + g_j]) - \frac{\delta}{2} x^T L(x_\delta) x \right\} \right).
\]

Note that when the Laplace prior \( p(x | \delta) \) defined by (2.6) is assumed, the posterior density function does not have the form (3.1). In this case, we use the following Gaussian approximation: \( p(x | \delta) \approx c \cdot \exp \left( -\frac{1}{2} x^T L(x_\delta) x \right) \) in (3.1), where \( c \) is a normalizing constant, \( x_\delta \) is the MAP estimator defined by (1.7), and \( L(x_\delta) \) is defined by (2.10) for one-dimensional signals. We have been unsuccessful in developing a convergent sampling scheme when we do not use this Gaussian approximation of the Laplace prior.

**3.1. A Gaussian approximation of the posterior density function.** Our MCMC method is a Metropolis-Hastings method with independence proposal

\[
p_{x_\delta}(x | b, \delta) \propto \delta^{(N-1)/2} \exp \left( -\frac{1}{2} \left\| \frac{Ax - (b - g)}{\sqrt{Ax_\delta + g}} \right\|^2 - \frac{\delta}{2} x^T L(x_\delta) x \right),
\]

where \( x_\delta \) is the MAP estimator. Note that this is the Gaussian approximation of the posterior density (3.1) obtained by replacing the Poisson likelihood (1.3) with the Gaussian approximation (2.14, 2.17). The distribution for \( p_{x_\delta}(x | b, \delta) \) defined by (3.2) is Gaussian of the form

\[
x | b, \delta \sim N \left( H(x_\delta)^{-1} A^T \left( \frac{b - g}{Ax_\delta + g} \right), H(x_\delta)^{-1} \right),
\]
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 (3.3) require the computation of \( H(x_\delta)^{-1/2} \), which is computationally intractable for very large-scale problems. An alternative, which is more computationally efficient, is to solve the stochastic linear system
\[ (3.4) \quad H(x_\delta)x = A^T \left( \frac{b - g}{Ax_\delta + g} \right) + w, \quad w \sim \mathcal{N}(0,H(x_\delta)), \]

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 \( v_1 \sim \mathcal{N}(0,I_{M \times M}) \), \( v_2 \sim \mathcal{N}(0,I_{N-1 \times N-1}) \).

3.2. Sampling from \( p(x|b,\delta) \) using Metropolis-Hastings. We now present our first MCMC method, which uses the approximate Gaussian density \( p_{x_\delta}(x|b,\delta) \), defined by (3.3), as an independence proposal within a Metropolis-Hastings (MH) algorithm. The MH correction of the samples from \( p_{x_\delta}(x|b,\delta) \) yields theoretically correct samples from the full posterior \( p(x|b,\delta) \) defined by (3.1). 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 (3.1).

MH Algorithm for sampling from \( p(x|b,\delta) \) defined by (3.1)

0. Compute \( x_\delta \). Set \( k = 0 \) and \( x_0 = x_\delta \).
1. Compute a candidate sample \( x_* \sim p_{x_\delta}(x|b,\delta) \) by solving (3.4) for a given \( w \).
2. Compute the acceptance ratio
\[ (3.5) \quad r = \min \left\{ 1, \frac{p(x_*|b,\delta)p_{x_\delta}(x_\delta|x_\delta,b,\delta)}{p(x_\delta|b,\delta)p_{x_\delta}(x_*|x_\delta,b,\delta)} \right\} \]

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:
1. Note that since \( p(x|b,\delta) \) and \( p_{x_\delta}(x|b,\delta) \) have the same prior, the priors cancel in (3.5). Indeed, this is why we use the Gaussian approximation in the Laplace prior case; otherwise, the priors don’t cancel and the computation of (3.5) becomes unstable. Moreover, we have found it to be numerically advantageous to use the following equivalent definition of (3.5): taking into account the cancelation of the priors:
\[ (3.6) \quad r = \min \left\{ 1, e^{-c} \right\}, \]
\[ c = -\ln p(b|x_*) + \ln p(b|x_k) - L^\text{wis}_\delta(x_k|b) + L^\text{wis}_\delta(x_*|b). \]
2. It is possible to sample from the local Gaussian \( p_{x_k}(x|b, \delta) \), instead of \( p_{x_k}(x|b, \delta) \), in Step 1, but this requires a modification of the prior ratio in (3.5) from \( p_{x_k}(x_k|b, \delta)/p_{x_k}(x, b, \delta) \) to \( p_{x_k}(x_k|b, \delta)/p_{x_k}(x, b, \delta) \). A major drawback of this approach is that since the proposals are different, the normalization constants do not cancel, making this approach more computationally demanding and potentially numerically unstable.

### 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.,

\[
p(\delta) \propto \delta^{\alpha-1} \exp(-\beta \delta).
\]

We choose the parameter values \( \alpha = 1 \) and \( \beta = 10^{-4} \), since then (3.7) is essentially flat over the domain of possible values of \( \delta \). Given (3.7), Bayes’ law yields the posterior density function

\[
p(x, \delta|b) \propto p(b|x)p(x|\delta)p(\delta)
= \delta^{(N-1)/2 + \alpha - 1} \exp\left(-\sum_{j=1}^{M} \{([Ax]_j + g_j) - b_j \ln([Ax]_j + g_j)\}\right)
\times \exp\left(-\frac{\delta}{2}x^TLx - \beta \delta\right).
\]

The conditional density for \( x|b, \delta \) then has the form (3.1), with \( L(x_\delta) = L \), and hence, we can use \( p_{x_k}(x|b, \delta) \), as defined in (3.2), 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

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

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

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

0. Compute \( x_\delta \). Set \( k = 0, \delta_0 = \delta, \) and \( x_0 = x_\delta \).
1. Compute a candidate sample \( x_* \sim p_{x_\delta}(x|b, \delta_k) \) by solving (3.4).
2. Compute the acceptance ratio

\[
r = \min\left\{ 1, \frac{p(x_*, b, \delta_k) p_{x_\delta}(x_k|b, \delta_k)}{p(x_k|b, \delta_k) p_{x_\delta}(x_*|b, \delta_k)} \right\}
\]

and a random draw \( u \sim U(0, 1) \). If \( u < r \), set \( x_{k+1} = x_* \), else set \( x_{k+1} = x_k \).
3. Compute a conditional sample \( \delta_{k+1} \sim p(\delta|b, x_{k+1}) \) defined by (3.9).
4. Set \( k = k + 1 \) and return to Step 1.

**Remarks:**

1. 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 (3.8) is very similar, and so likely also degrades in the infinite limit.
2. Just as above, we compute (3.10) using an expression analogous to (3.6).

3. 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 (3.10) 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.

4. Finally, if the Laplace prior is used, one can fix \( L = L(x_\delta) \) defined in (2.10) at the outset, where \( x_\delta \) is the MAP estimator (1.7) 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. 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 [23, 30, 21, 27]. A typical setup is shown in Fig. 1. 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 through the scene. The transmitted photons are absorbed by a scintillator, which approximately linearly emits visible photons in response. 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 transmission for a given radiograph, i.e. to determine the function mapping object areal density onto image intensity, it is common to include a calibration object with known areal densities, such as the step wedge shown in Fig. 1.

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 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.

4.1. 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. 2 (a). The blurred edge, the mean of the Metropolis-Hastings samples, and 95% confidence bands are shown in Fig. 2 (b). No noise is added to the blurred edge; this example is merely shown to demonstrate that the sampling scheme accurately reproduces uncontaminated edges.

Fig. 3 (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 Lead (Pb) step wedge. A horizontal cross section through the step wedge is shown in Fig. 3 (b), along with the Poisson MAP estimate, again with Laplace MRF prior. Though the data is synthetic, this is not an “inverse crime” in the sense of [20], as the data is generated using a forward model (Monte Carlo simulation) that is different from the convolution model used for inversion. We know the correct PSF, because we know the exact shape of the radiation source used in the simulation. In this case, the source is a Gaussian X-ray spot with $\sigma = 7.5$ pixels. Nonetheless, we do not know the “correct answer” for the deconvolution, since the intensity in the image is a function of how many particles are run in the simulation and their stochastic properties. For this data set, 1000 samples were computed and last 500 used for analysis. The mean of the Metropolis-Hastings samples is shown in Fig. 3 (c) with 95% credibility bands. The mean reconstruction sharpens the signal as would be expected given the size of the PSF, and the credibility bands give tight error estimates.

4.2. Edge Reconstruction in X-ray Radiography. Fig. 4 (a) shows a radiograph from the Cygnus dual-beam radiographic imaging facility at the U.S. Department of Energy’s Nevada National Security Site [27]. Cygnus is a 2.25 MeV end-point energy pulsed power X-ray source with a rod-pinch diode [23, 25]. The image is a
composite target with several objects, each used for a different system calibration. 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), can be computed by a method involving the “L” rolled edge (backwards Pac-man) in the lower right of the image in Fig. 4 (a). (See [9] for the method used to compute the PSF.) 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. 4 (b). The step wedge 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. 5 (a) shows a one-dimensional vertical cross section, or line-out, of the step wedge. We use the Laplace increment (total variation) prior and the Poisson lagged-diffusivity iteration as described in Section 2.3 (without projected line search), with the regularization parameter $\delta = .0075$ and MAP estimator $x_{\text{MAP}} \overset{\text{def}}{=} x_s$ computed as described in Section 2.3. A plot of $x_{\text{MAP}}$ is shown with the data in Fig. 5 (a). We compute 2500 samples and use the last half for analysis: the conditional mean and 95% credibility bounds are plotted in Fig. 5 (b). The credibility bounds show small spikes at the jump discontinuities in the signal, indicating higher levels of uncertainty near the discontinuities.

Fig. 6 (a) shows another real radiograph from the Cygnus facility of a Lead (Pb) step wedge, which is also used to compute transmission curves for the system. This
Fig. 3. (a) A simulated step wedge generated by a radiation transport code. The corners are black due to hard collimation of the simulated X-ray beam. (b) A horizontal line out (cross section) through the step wedge, with the Laplace prior MAP estimate. (c) The step wedge cross section with the mean of the Metropolis-Hastings samples and 95% credibility bands.

is the step wedge that was simulated to generate the image and data in Fig. 3. A horizontal line-out from the image is shown in Fig. 6 (b), along with the Poisson-TV MAP estimate with Laplace prior. Again 2500 samples were computed, using $\delta = 0.1$, and the last half were used for analysis. Plot (c) shows the results of the sampling, with the original data, the mean of the Metropolis-Hastings samples, and 95% credibility bands.

Remark: Note that the deblurred signals in Fig. 6 (b),(c) 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. 4. (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. 5. (a) Lineout and Poisson-TV MAP estimate of the signal. (b) The mean of the Metropolis-Hastings samples, along with 95% credibility bands. The spikes near the edges indicate higher uncertainty near the jump discontinuities.

4.3. 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. 7 (a), is from the central cylinder object in the upper middle section of Fig. 4 (a). This object is used for testing algorithms for computing object densities from image data, so the lineout shown in (a) is inverted on the intensity scale (bright pixels have low values), because that is how the data is analyzed in density reconstruction. The PSF used is the same as that for the step wedge. Fig. 7 (b) shows the Gaussian MRF prior MAP estimate, which results in a deblurred signal with noise largely removed as an effect of the smoothing prior. In plot (c) of Fig. 7, we plot a histogram of the $\delta$ samples obtained within the hierarchical, Metropolis-within-Gibbs scheme.

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. The proposal requires that we first compute the MAP estimator, and so we also present an efficient algorithm for this step. We derive our priors using a Markov random field (MRF) approach with a focus on Gaussian and Laplace
Fig. 6. (a) A real X-ray radiograph of the Lead (Pb) step wedge simulated in Fig. 3 (a). (b) A vertical line out (cross section) through the object with the Poisson-TV MAP estimate. (c) The same line out with the Metropolis-Hastings sample mean and 95% credibility bands. (d) Zoomed in on one of the edges to show that the deconvolution has sharpened the edges by approximately the correct amount, given the computed PSF.

Fig. 7. (a) A horizontal cross section from the central cylinder in Fig. 4 (a), with the Gaussian MRF prior MAP estimate. The intensity has been inverted so that bright pixels have low values, because that is how the data is analyzed when calculating object density. (b) The mean of the Metropolis-within-Gibbs samples with 95% credibility bands. (c) Histogram of the samples for the scale parameter $\delta$.

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. Finally, we focus on one-dimensional signals, both because the applications we consider are one-dimensional, but also because our MCMC method appears to be unstable for two-dimensional signal reconstruction problems.

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 reconstruc-
Poisson Sampling

tion. 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.

Appendix: Projected, Poisson Lagged-Diffusivity Iteration. This discussion follows the description of the projected lagged-diffusivity iteration found in [7]. First, we define the active set for $x_0$, which is defined

$$\mathcal{A}(x) = \{ i \mid x_i = 0 \}.$$  

The complementary set of indices is called the inactive set and is denoted by $\mathcal{I}(x)$. The orthogonal projection of a vector $x \in \mathbb{R}^N$ onto $\{x \in \mathbb{R}^N \mid x \geq 0\}$ is given by $P(x) = \max\{x, 0\}$, where the maximum is computed component-wise. Finally, let $D_{\mathcal{A}}(x)$ denote the diagonal matrix with components

$$[D_{\mathcal{A}}(x)]_{ii} = \begin{cases} 1, & i \in \mathcal{I}(x) \\ 0, & i \in \mathcal{A}(x). \end{cases}$$

Then $D_{\mathcal{A}}(x) = I - D_{\mathcal{I}}(x)$, where $I$ is the $N \times N$ identity matrix.

We now present our computational method for solving (2.7). A general formulation of the method is as follows: given $x_k$, define $D_{\mathcal{A}}^k \equiv D_{\mathcal{A}}(x_k)$, $D_{\mathcal{I}}^k \equiv D_{\mathcal{I}}(x_k)$, and compute

$$v_k = - (D_{\mathcal{I}}^k H_k D_{\mathcal{I}}^k + D_{\mathcal{A}}^k)^{-1} \nabla \ell(x_k \mid b);$$

$$\lambda_k = \arg \min_{\lambda > 0} T_{\alpha}(P(x_k + \lambda v_k));$$

$$x_{k+1} = P(x_k + \lambda_k v_k).$$

Note that (5.2) can be equivalently expressed

$$v_k = - (D_{\mathcal{I}}^k H_k D_{\mathcal{I}}^k + D_{\mathcal{A}}^k)^{-1} D_{\mathcal{I}}^k \nabla \ell(x_k \mid b) - D_{\mathcal{A}}^k \nabla \ell(x_k \mid b),$$

so the linear system that needs to be solved is

$$(D_{\mathcal{I}}^k H_k D_{\mathcal{I}}^k + D_{\mathcal{A}}^k) v = -D_{\mathcal{I}}^k \nabla \ell(x_k \mid b).$$

For sufficiently large-scale problems, this system should be solved approximately. For this we use conjugate gradient (CG) iterations, with initial guess $v_0 = 0$ and stopping criterion

$$\| (D_{\mathcal{I}}^k H_k D_{\mathcal{I}}^k + D_{\mathcal{A}}^k) v_j^l + D_{\mathcal{I}}^k \nabla \ell(x_k \mid b) \| \leq \min \left\{ \frac{1}{2}, \| D_{\mathcal{I}}^k \nabla \ell(x_k \mid b) \| \right\} \cdot \| D_{\mathcal{A}}^k \nabla \ell(x_k \mid b) \|,$$

where $v_j^l$ is the $j$th CG iterate at outer iteration $k$; CG stopping criteria of this type are standard [24]. In addition, we use the stopping criteria $j = CG_{\text{max}}$. If either of the stopping criteria hold, CG iterations are stopped and $v_k$ in (5.2) is taken to be the most recent CG iterate. Then $v_k$ can be guaranteed to be a descent direction, making it appropriate for use in the line search subproblem (5.3).
Because $\ell(x|b)$ is not a quadratic function, the line search subproblem (5.3) must also be approximately solved. A standard approach is to do this iteratively using a backtracking line search algorithm. In our implementation, we use the quadratic backtracking line search scheme of [22], which generates a finite decreasing sequence of approximate solutions $\{l_k^m\}_{m=1}^M$, where $m$ is the smallest positive integer such that the sufficient decrease condition

$$\ell(\mathcal{P}(x_k + \lambda_k^m v_k | b)) \leq \ell(x_k | b) + \mu \langle \nabla \ell(x_k | b), \mathcal{P}(x_k + \lambda_k^m v_k) - x_k \rangle$$

holds. In our implementation, $\mu = 0.1$, which is what is used in [22].

Finally, we stop projected Newton iterations based on the value of the norm of the projected gradient, which is defined by

$$\nabla_{\text{proj}} \ell(x_k | b) \overset{\text{def}}{=} D_k^b \nabla \ell(x_k | b) + \min \{0, D_k^b \nabla \ell(x_k | b)_{g} \},$$

noting that since $\ell(x|b)$ is strictly convex, $x_*\text{solves (2.7)}$ if and only if $\nabla_{\text{proj}} \ell(x_0 | b) = 0$. This motivates the following stopping criteria for the outer iterations:

$$\|\nabla_{\text{proj}} T_\alpha(x_k)\|/\|\nabla_{\text{proj}} T_\alpha(x_0)\| < \text{GradTol},$$

where GradTol is a small positive number. We also stop iterations if $k = \text{LD}_{\text{max}}$.

Finally, we note that the projected Newton method [13] results if the full Hessian is used for $H_k$ instead, but the resulting method is less computationally efficient. Moreover, one possibility for improving the performance of the method is to precondition the CG iterations. However, preconditioning is made difficult in this case because the form of $H_k$ changes nontrivially as the active set changes.

REFERENCES


