An MCMC Method for Uncertainty Quantification in Nonnegativity Constrained Inverse Problems

Johnathan M. Bardsley

Department of Mathematical Sciences
University of Montana
Missoula, Montana 59812
USA

E-mail: bardsleyj@mso.umt.edu

Colin Fox

Department of Physics
University of Otago
Dunedin 9054
New Zealand

E-mail: fox@physics.otago.ac.nz

MSC numbers: 15A29, 62F15, 65F22, 94A08

Keywords: inverse problems, image reconstruction, nonnegativity constrained optimization, Bayesian inference, Markov chain Monte Carlo, uncertainty quantification.

Abstract. The development of computational algorithms for solving inverse problems is, and has been, a primary focus of the inverse problems community. Less studied, but of increased interest, is uncertainty quantification for solutions of inverse problems obtained using computational methods. In this paper, we present a method of uncertainty quantification for linear inverse problems with nonnegativity constraints. Our approach utilizes a Bayesian statistical framework, and we present a simple Markov chain Monte Carlo (MCMC) method for sampling from a particular posterior distribution. From the posterior samples, estimation and uncertainty quantification for both the unknown (image in our case) and regularization parameter are performed. The primary challenge of the approach is that for each sample a large-scale nonnegativity constrained quadratic minimization problem must be solved. We perform numerical tests on both one- and two-dimensional image deconvolution problems, as well as on a computed tomography test case. Our results show that our nonnegativity constrained sampler is effective and computationally feasible.

1. Introduction

In this paper, we focus on linear models with independent and identically distributed Gaussian noise; that is we assume the data-noise model

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{b} \in \mathbb{R}^m$ corresponds to observed data; $\mathbf{x}$ is the $n \times 1$ vector of unknowns; $\mathbf{A}$ is the $m \times n$ forward model matrix obtained via a numerical discretization of the forward map; and $\boldsymbol{\eta}$ is an $m \times 1$ independent and identically distributed (iid) Gaussian random vector with variance σ^2 across all pixels.

We are interested in linear models of the form (1) that arise from the numerical discretization of an ill-posed linear operator equation $b = Ax$, where b and x are functions, and A is a compact operator. Such equations arise in a number of important image processing applications, such as computed tomography (CT) and image deconvolution.

The goal is to obtain an estimate of the unknown $\mathbf{x}$, given the data $\mathbf{b}$, the forward map $\mathbf{A}$, and statistical model (1). The most common approach for obtaining such an estimate is to write down a variational problem, or corresponding Euler-Lagrange equation, that is then solved using a numerical method, the solution being the desired estimate. For general discussions of ill-posed problems and regularization, see one of the many excellent tests on the subject, e.g., [7, 15].

However, applications exist in which the values of estimated parameters have financial, or other, consequences attached to them. In such cases, the problem of uncertainty quantification (UQ) is extremely important. In the recent paper [1], a Bayesian hierarchical model is presented, and an efficient MCMC method is introduced for sampling from the resulting posterior distribution. The samples from the posterior density are then used to compute an estimator (reconstructed image), e.g., the sample median, as well as for uncertainty quantification, e.g., by computing credibility intervals (Bayesian confidence intervals) for each sampled parameter. Similar approaches have

been studied by researches in the field of inverse problems and imaging; see, e.g., [3, 4, 9, 10, 13, 16].

In this paper, we extend the MCMC method from [1] to the case in which the unknown $\mathbf{x}$ has a nonnegativity constraint. This constraint arises in many applications, e.g., in CT, where the unknown is tissue density, and in image deblurring, where the unknown is light intensity. Nonnegativity constrained sampling is different from positivity constrained sampling. In the later, the probability of the unknown being zero is zero, whereas in the former the unknown is allowed to be zero with some positive probability, in our case determined by the likelihood and prior. For imaging examples, where zero intensity values are common, a nonnegativity constraint seems natural. Nevertheless, we compare results obtained with both approaches.

We will show that nonnegative samples of $\mathbf{x}$ can be computed by replacing the quadratic minimization problem (or equivalent linear system) that must be solved when computing unconstrained samples of $\mathbf{x}$ by the corresponding nonnegativity constrained quadratic minimization problem. We will provide a mathematical justification for, and interpretation of, this modification. While this modification is simple in theory, efficiently solving the resulting nonnegativity constrained quadratic program is a nontrivial problem. For this, we use the gradient projection–conjugate gradient (GPCG) method of [12]. GPCG is extremely efficient for large-scale problems and inherits finite convergence from CG (when exact arithmetic is assumed).

The paper is organized as follows. In Section 2 we derive the posterior density function from Bayes' Law and present the MCMC method of [1] for computing samples from the posterior density. Then in Section 3, we focus on sampling with constraints, and present methods for both positivity and nonnegativity constrained sampling. Finally, in Section 4, we present a method for determining MCMC chain convergence and then apply our nonnegativity constrained MCMC method to examples from image deconvolution in both one- and two-dimensions and from computed tomography. Conclusions are given in Section 5.

2. The posterior density function

Our MCMC method draws samples from a particular posterior density function, which we derive using Bayes' Law, as in [1, 8], and which has the form

$$p(\mathbf{x}, \lambda, \delta|\mathbf{b}) \propto p(\mathbf{b}|\mathbf{x}, \lambda)p(\lambda)\, p(\mathbf{x}|\delta)p(\delta). \tag{2}$$

The first term on the right is called the likelihood function, which is determined by (1) and has the form

$$p(\mathbf{b}|\mathbf{x}, \lambda) \propto \lambda^{n/2} \exp\left(-\frac{\lambda}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2\right),$$

where $\lambda = 1/\sigma^2$ is the noise precision.

The third term on the right in equation (2) is the prior probability density function and is assumed to have the form

$$p(\mathbf{x}|\delta) \propto \delta^{n/2} \exp\left(-\frac{\delta}{2}\mathbf{x}^T\mathbf{C}\mathbf{x}\right), \tag{3}$$

where δ is a scaling parameter for the prior precision matrix $\delta\mathbf{C}$. Here δ and $\mathbf{C}$ correspond to the regularization parameter and matrix, respectively. Note that (3) is also used in Bayesian spacial statistics [8, 14].

Finally, the hyper-prior probability densities for λ and δ are taken to be Gamma distributions:

$$p(\lambda) \propto \lambda^{\alpha_\lambda - 1} \exp(-\beta_\lambda \lambda), \tag{4}$$

$$p(\delta) \propto \delta^{\alpha_\delta - 1} \exp(-\beta_\delta \delta). \tag{5}$$

Note that the Gamma distribution $\Gamma(\alpha, \beta)$ has a probability density function satisfying $g(t|\alpha, \beta) \propto t^{\alpha-1} \exp(-\beta\, t)$. Following [8], we take $\alpha_\lambda = \alpha_\delta = 1$ and $\beta_\lambda = \beta_\delta = 10^{-4}$. Then the hyper-priors can be deemed to be "uninformative", since the mean and variance of the corresponding Gamma distributions is $\alpha/\beta = 10^4$ and $\alpha/\beta^2 = 10^8$, respectively. Uninformative hyper-priors are chosen so that their effect on the sampled values for λ and δ are negligible. Note that these choices of parameters require no tuning whatsoever and work on a variety of examples. Moreover, no other parameters remain to be defined. However, if one has a reasonable *a priori* notion of what λ and/or δ should be, the corresponding hyper-prior parameter choices could be modified accordingly.

Thus (2) has the form

$$p(\mathbf{x}, \lambda, \delta|\mathbf{b}) \propto \lambda^{n/2+\alpha_\lambda-1} \delta^{n/2+\alpha_\delta-1} \exp\left(-\frac{\lambda}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 - \frac{\delta}{2}\mathbf{x}^T\mathbf{C}\mathbf{x} - \beta_\lambda\lambda - \beta_\delta\delta\right). \tag{6}$$

Our MCMC method will sample from $p(\mathbf{x}, \lambda, \delta|\mathbf{b})$. In [1, 8], a MCMC method is introduced for sampling from (6), and uncertainty quantification is performed [1] for image deconvolution examples in one- and two-dimensions.

2.1. Defining the matrix $\mathbf{C}$ *via Gaussian Markov random fields*

It remains to define the prior precision (regularization) matrix $\mathbf{C}$. Staying with the statistical flavor of the presentation, we use Gaussian Markov random fields (MRF's), as in [8, 14]. Some of the notations set in this subsection will be used later in the paper.

A MRF on the pixel grid is defined by specifying a neighborhood system at each pixel, as well as a set of n conditional densities $\{p(x_i|\mathbf{x}_{\partial_i}),\ i = 1, \ldots, n\}$, where $\mathbf{x}_{\partial_i} = \{x_j | j \in \partial_i\}$, with $j \in \partial_i$ if $i \neq j$ and pixels i and j are neighbors. We note that MRF's have the property that for all i, the full conditional density $p(x_i|\mathbf{x}_{-i}) = p(x_i|\mathbf{x}_{\partial_i})$, where $\mathbf{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, i.e. it depends only upon it's neighbors.

For a given neighborhood system, we define the conditional densities to be Gaussian and of the form

$$x_i|\mathbf{x}_{\partial_i} \sim N(\bar{x}_{\partial_i}, 1/n_i),$$

where n_i is the number of neighbors belonging to x_i and $\bar{x}_{\partial_i}$ is the average of the x_j's neighboring x_i. The resulting joint Gaussian density for $\mathbf{x}$ is then given by [8, 14]

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

where

$$[\mathbf{C}]_{ij} = \begin{cases} n_i & i = j, \\ -1 & j \in \partial_i, \\ 0 & \text{otherwise.} \end{cases}$$

Thus if we define $\mathbf{C}_{-i}$ to be the matrix that results if the ith column of $\mathbf{C}$ is removed, we have $n_i x_{\partial_i} = -[\mathbf{C}_{-i}\mathbf{x}_{-i}]_i$.

Note that if zero boundary pixels are assumed and a standard first-order neighborhood is used – which in 1D includes the pixels to the left and right, and in 2D those to the left, right, above, and below – the discrete negative-Laplacian with Dirichlet boundary conditions results. If, however, periodic boundary conditions are assumed, the discrete negative-Laplacian with periodic boundary conditions results. Finally, recall that in practice, we include the scaling parameter δ (see (3)), which is necessary in any practical application [8, 14].

2.2. MCMC sampling of the posterior distribution

In this section, we present the MCMC method of [1]. Our choice of prior (3) on $\mathbf{x}$, and hyper-priors (4) on λ and (5) on δ were made with conjugacy relationships in mind [6], i.e. so that the full conditional densities have the same form as the corresponding prior/hyper-prior. To see this, note that the full conditional densities have the form

$$\begin{aligned} p(\mathbf{x}|\lambda,\delta,\mathbf{b}) &\propto \exp\left(-\frac{\lambda}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2 - \frac{\delta}{2}\mathbf{x}^T\mathbf{C}\mathbf{x}\right), \\ p(\lambda|\mathbf{x},\delta,\mathbf{b}) &\propto \lambda^{n/2+\alpha_\lambda-1}\exp\left(\left[-\frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2 - \beta_\lambda\right]\lambda\right), \\ p(\delta|\mathbf{x},\lambda,\mathbf{b}) &\propto \delta^{n/2+\alpha_\delta-1}\exp\left(\left[-\frac{1}{2}\mathbf{x}^T\mathbf{C}\mathbf{x} - \beta_\delta\right]\delta\right), \end{aligned}$$

and hence,

$$\mathbf{x}|\lambda,\delta,\mathbf{b} \sim N\left((\lambda\mathbf{A}^T\mathbf{A}+\delta\mathbf{C})^{-1}\lambda\mathbf{A}^T\mathbf{b}, (\lambda\mathbf{A}^T\mathbf{A}+\delta\mathbf{C})^{-1}\right), \tag{7}$$

$$\lambda|\mathbf{x},\delta,\mathbf{b} \sim \Gamma\left(n/2+\alpha_\lambda, \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2+\beta_\lambda\right), \tag{8}$$

$$\delta|\mathbf{x},\lambda,\mathbf{b} \sim \Gamma\left(n/2+\alpha_\delta, \frac{1}{2}\mathbf{x}^T\mathbf{C}\mathbf{x}+\beta_\delta\right). \tag{9}$$

The following Gibbs sampler follows immediately from (7)-(9).

An Unconstrained MCMC Method for Sampling from $p(\mathbf{x},\delta,\lambda|\mathbf{b})$.

0. Initialize δ_0 and λ_0, and set $k = 0$;

1. Compute $\mathbf{x}^k \sim N\left((\lambda_k \mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})^{-1}\lambda_k\mathbf{A}^T\mathbf{b}, (\lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})^{-1}\right)$;
2. Compute $\lambda_{k+1} \sim \Gamma\left(n/2 + \alpha_\lambda, \frac{1}{2}\|\mathbf{A}\mathbf{x}^k - \mathbf{b}\|^2 + \beta_\lambda\right)$;
3. Compute $\delta_{k+1} \sim \Gamma\left(n/2 + \alpha_\delta, \frac{1}{2}(\mathbf{x}^k)^T\mathbf{C}\mathbf{x}^k + \beta_\delta\right)$;
4. Set $k = k + 1$ and return to Step 1.

Since the parameters λ and δ are scalar, the scalar random draws required in Steps 2 and 3 are very efficient and easy to compute given the appropriate random number generator. Note that this is the MCMC method presented in [1].

For inverse problems, the computational bottleneck occurs in Step 1, where the following linear system must be solved at each iteration for $\mathbf{x}^k$:

$$(\lambda_k\mathbf{A}^T\mathbf{A}+\delta_k\mathbf{C})\mathbf{x}^k = \lambda_k\mathbf{A}^T\mathbf{b}+\mathbf{w}, \quad \text{where} \quad \mathbf{w} \sim N(\mathbf{0}, \lambda_k\mathbf{A}^T\mathbf{A}+\delta_k\mathbf{C}). \tag{10}$$

Thus in order for the MCMC method to be effecient, solutions of (10) must be efficiently computable. In [1], for example, a two-dimensional image deconvolution test case with periodic boundary conditions is considered. In this instance, $(\lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})$ can be diagonalized via the discrete Fourier transform, and hence, (10) can be efficiently solved.

3. MCMC sampling with constraints

Computing Gaussian samples as in (10) clearly does not impose constraints on the computed samples. However, in many imaging applications a nonnegativity constraint is natural.

The problem of sampling from a Gaussian with a positivity constraint has been studied by other researchers; see, e.g., [3] for the scalar, standard normal case. The negative portions of the Gaussian are simply truncated, yielding no mass at the boundary of the constraints. As we will see, positivity constrained sampling can be incorporated into (10), within a Gibbsian framework, in a straightforward manner.

Nonnegativity constrained sampling, on the other hand, results from projecting the negative portion of the distribution onto the boundary of the constraints, yielding mass at the boundary. In some imaging applications, e.g. looking at dark regions of the night sky in astronomy, one can imagine a positive probability that certain pixels have a zero value, yielding mass at the boundary.

In what follows, we show how to perform both types of sampling schemes and compare the two methods later in the paper.

3.1. Positivity constrained sampling

A convergent sampler also results if we replace Step 1 of the above MCMC method by a component-wise Gibbs sampler (see [8]), as follows:

(i) Compute a sample x_i^k from the random variable with probability density $p(x|\mathbf{x}_{-i}^k, \lambda_k, \delta_k)$ proportional to

$$\exp\left\{-\frac{\lambda_k[\mathbf{A}^T\mathbf{A}]_{ii} + \delta_k\mathbf{C}_{ii}}{2}\left(x - \frac{\lambda_k([\mathbf{A}^T\mathbf{b}]_i - [\mathbf{A}^T\mathbf{A}_{-i}\mathbf{x}_{-i}^k]_i) - \delta_k\mathbf{C}_{-i}\mathbf{x}_{-i}}{\lambda_k[\mathbf{A}^T\mathbf{A}]_{ii} + \delta_k\mathbf{C}_{ii}}\right)^2\right\}, \tag{11}$$

for $i = 1, \ldots, n$, where $\mathbf{A}_{-i}$ and $\mathbf{C}_{-i}$ are the matrices that result after the ith column of $\mathbf{A}$ and $\mathbf{C}$, respectively, are removed.

The direct method (10) is preferable to this component-wise Gibbs sampler provided it is computationally feasible. However, (11) has the benefit that a positivity constraint is easily added. This can be done if Step 1 is replaced by

(i) Compute a sample x_i^k from the random variable with probability density

$$p_+(x|\mathbf{x}_{-i}^k, \lambda_k, \delta_k) \propto p_+(x)p(x|\mathbf{x}_{-i}^k, \lambda_k, \delta_k) \tag{12}$$

for $i = 1, \ldots, n$, where $p(x|\mathbf{x}_{-i}^k, \lambda_k, \delta_k)$ is given in (11), and $p_+(x)$ is 1 for positive values of x and 0 otherwise.

The resulting sampler for the posterior probability density $p(\mathbf{x}, \delta, \lambda|\mathbf{b})$ is provably convergent. Moreover, we can directly sample from (12) mimicking the approach of [3] for sampling from the positivity constrained standard normal density. There the Inverse Cumulative Distribution Function (CDF) Method is used.

Inverse CDF Method for sampling from a random variable x given its CDF ϕ:

(i) Compute $t \sim \text{Uniform}([0, 1])$;

(ii) Calculate $x = \phi^{-1}(t)$.

To implement the Inverse CDF Method, we need ϕ^{-1} for x with probability density function (12).

To simplify notation, let μ and γ be the mean and precision of the Gaussian random variable (11). Then the cumulative density of (12) has the form

$$\begin{aligned}
\phi(z) &= C \int_0^z \exp\left\{-\frac{\gamma}{2}(x-\mu)^2\right\} dx, \\
&= C\sqrt{\frac{2}{\gamma}} \left(\int_0^{\sqrt{\frac{\gamma}{2}}(z-\mu)} e^{-s^2}\, ds + \int_0^{\sqrt{\frac{\gamma}{2}}\mu} e^{-s^2}\, ds \right), \\
&= C\sqrt{\frac{\pi}{2\gamma}} \left(\operatorname{erf}\left(\sqrt{\frac{\gamma}{2}}(z-\mu)\right) + \operatorname{erf}\left(\sqrt{\frac{\gamma}{2}}\mu\right) \right),
\end{aligned}$$

where the first inequality follows from the change of variables $s = \frac{\gamma}{2}(x - \mu)$, and $\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-s^2}\, ds$. Now since $\lim_{z\to\infty} \phi(z) = \lim_{z\to\infty} \operatorname{erf}(z) = 1$, we can solve for the normalizing constant C:

$$C = \left(\sqrt{\frac{\pi}{2\gamma}} \left(1 + \operatorname{erf}\left(\sqrt{\frac{\gamma}{2}}\mu\right) \right) \right)^{-1}$$

Hence the cumulative density function for (12) is given by

$$\phi(z) = \frac{\operatorname{erf}\left(\sqrt{\frac{\gamma}{2}}(z-\mu)\right) + \operatorname{erf}\left(\sqrt{\frac{\gamma}{2}}\mu\right)}{1 + \operatorname{erf}\left(\sqrt{\frac{\gamma}{2}}\mu\right)}.$$

Finally, solving the equation $t = \phi(z)$ for z yields

$$\phi^{-1}(t) = \mu + \sqrt{\frac{2}{\gamma}}\,\text{erfinv}\left\{t\left(1 + \text{erf}\left(\sqrt{\frac{\gamma}{2}}\mu\right)\right) - \text{erf}\left(\sqrt{\frac{\gamma}{2}}\mu\right)\right\}. \tag{13}$$

Built-in functions for evaluating erf and erfinv are contained in MATLAB. Thus we have shown how to sample from (12), yielding a positivity constrained, component-wise Gibbs sampler for the posterior density $p(\mathbf{x}, \delta, \lambda|\mathbf{b})$. For large-scale problems, given the component-wise nature of this sampler, the resulting MCMC will be computationally demanding.

3.2. Nonnegativity Constrained Sampling

Note that for the conditional distributions defined by (12), the probability of $x_i = 0$ is zero for all i. However, one can imagine applications in image processing where $x_i = 0$ with some positive probability. For example, in the application of astronomy, we would expect a reconstruction of an image with a dark background to have a large number of zero valued pixels. Similarly, in medical imaging, e.g. computed tomography, we would expect zero tissue density values for pixels in regions where there is no tissue, and hence, where the x-ray signal is not attenuated.

We now introduce a probability model that will have such characteristics. We begin by noting, as in [1], that (10) can be equivalently written as the solution of the variational problem

$$\mathbf{x}^k = \arg\min_{\mathbf{x}} \left\{\frac{1}{2}\mathbf{x}^T(\lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})\mathbf{x} - \mathbf{x}^T(\lambda_k\mathbf{A}^T\mathbf{b} + \mathbf{w})\right\}, \tag{14}$$

where $\mathbf{w} \sim N(\mathbf{0}, \lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})$. This suggests the following obvious modification of (14) for computing nonnegativity constrained samples:

$$\mathbf{x}^k = \arg\min_{\mathbf{x}\geq\mathbf{0}} \left\{\frac{1}{2}\mathbf{x}^T(\lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})\mathbf{x} - \mathbf{x}^T(\lambda_k\mathbf{A}^T\mathbf{b} + \mathbf{w})\right\}, \tag{15}$$

where $\mathbf{w} \sim N(\mathbf{0}, \lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})$. It is important to ask: What sort of probability model does (15) define? To see this, we note that (15) can be shown to be equivalent to

$$\mathbf{x}^k_{\text{CON}} = \arg\min_{\mathbf{x}\geq\mathbf{0}} \|\mathbf{x} - \mathbf{x}^k_{\text{UC}}\|^2_E, \tag{16}$$

where $\mathbf{x}^k_{\text{UC}}$ is an unconstrained sample defined by (10), and

$$\|\mathbf{w}\|^2_E = \langle\mathbf{w}, \mathbf{w}\rangle_E = \mathbf{w}^T(\lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})\mathbf{w}. \tag{17}$$

Thus we have proven the following result.

Theorem 1. *Given the norm $\|\cdot\|_E$ defined by (17) and an unconstrained sample $\mathbf{x}^{k+1}_{\text{UC}}$ drawn from*

$$N\left((\lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})^{-1}\lambda_k\mathbf{A}^T\mathbf{b}, (\lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})^{-1}\right),$$

the corresponding vector $\mathbf{x}^k_{\text{CON}}$ defined by (15) is the minimum-norm projection of $\mathbf{x}^{k+1}_{\text{UC}}$ onto the set of all nonnegative vectors, i.e. it satisfies (16).

It is clear that the probability density/mass function defined by (15) will have positive mass along the boundaries of $\mathbf{x} \geq \mathbf{0}$. We believe that these mass values can likely be analytically computed, but we do not pursue such formulae here.

3.2.1. A nonnegativity constrained MCMC method One additional modification must be made to the unconstrained MCMC method presented in Section 2.2 to obtain our nonnegativity constrained MCMC method.

Recall that in the unconstrained case, the prior had the form (3). However, for $\mathbf{x} \geq \mathbf{0}$, if $\mathbf{D_x}$ is a diagonal matrix with diagonal values $[\mathbf{D_x}]_{ii} = 1$ for $x_i > 0$ and $[\mathbf{D_x}]_{ii} = 0$ for $x_i = 0$, then

$$\frac{\delta}{2}\mathbf{x}^T\mathbf{C}\mathbf{x} = \frac{\delta}{2}\mathbf{x}^T(\mathbf{D_x}\mathbf{C}\mathbf{D_x})\mathbf{x},$$

which corresponds to a degenerate precision matrix $\delta\mathbf{D_x}\mathbf{C}\mathbf{D_x}$ if $x_i = 0$ for any i. Because of this degeneracy, the prior changes slightly to

$$p(\mathbf{x}|\delta) = \delta^{n_p/2} \exp\left(\frac{\delta}{2}\mathbf{x}^T\mathbf{C}\mathbf{x}\right),$$

where n_p is the number of positive elements in $\mathbf{x}$. Thus the full conditional density for δ also changes slightly, from (7) to

$$p(\delta|\mathbf{x}, \lambda, \mathbf{b}) \propto \delta^{n_p/2+\alpha_\delta-1} \exp\left(\left[-\frac{1}{2}\mathbf{x}^T\mathbf{C}\mathbf{x} - \beta_\delta\right]\delta\right).$$

Thus Step 3 of our MCMC method will need to be modified.

Nonnegativity Constrained MCMC Method for Sampling from $p(\mathbf{x}, \delta, \lambda|\mathbf{b})$.

0. Initialize δ_0 and λ_0, and set $k = 0$.
1. Compute $\mathbf{x}^k$ defined by (15) with $\mathbf{w} \sim N(\mathbf{0}, \lambda_k\mathbf{A}^T\mathbf{A} + \delta_k\mathbf{C})$.
2. Compute $\lambda_{k+1} \sim \Gamma\left(n/2 + \alpha_\lambda, \frac{1}{2}\|\mathbf{A}\mathbf{x}^k - \mathbf{b}\|^2 + \beta_\lambda\right)$;
3. Compute $\delta_{k+1} \sim \Gamma\left(n_p^{k+1}/2 + \alpha_\delta, \frac{1}{2}(\mathbf{x}^k)^T\mathbf{C}\mathbf{x}^k + \beta_\delta\right)$, where n_p^{k+1} is the number of nonzero entries in $\mathbf{x}^k$;
4. Set $k = k + 1$ and return to Step 1.

In our experience, the resulting MCMC method will yield a credibililty interval for λ that contains the true value of $1/\sigma^2$, whereas without the change in Step 3, $1/\sigma^2$ will lie outside the credibility interval for λ.

It remains to present a method for computing $\mathbf{x}^k$ in Step 1. We do this in the next sub-section.

3.2.2. An Iterative Method for Computing Nonnegative Samples Finally, to complete our nonnegativity constrained MCMC method, we present a highly efficient iteration for solving (15), which is required in Step 1.

In our experience, the most efficient bound constrained quadratic minimization method is the gradient projection-conjugate gradient (GPCG) algorithm of Moré and Toreldo [12]. It can be shown that, just as for CG, GPCG has the property of finite convergence in exact arithmetic [12]. Moreover, for large-scale problems, the use of gradient projection iterations allows for large numbers of active constraints to be added, or dropped, at each iteration. In our experience, it is this latter property that distinguishes GPCG's performance when compared with other extensions of CG to bound constrained quadratic minimization.

For the sake of completeness, we now provide a brief description of the GPCG algorithm for solving problems of the form

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\geq\mathbf{0}} \left\{ q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{B}\mathbf{x} - \mathbf{x}^T\mathbf{c} \right\}, \tag{18}$$

where $\mathbf{B} \in \mathbb{R}^{n\times n}$ is symmetric positive definite, and $\mathbf{c} \in \mathbb{R}^n$. The reader interested in more detail should see [12].

GPCG Algorithm for solving (18):

Step 0. Set $j = 0$ and choose initial guess $\mathbf{x}^0 = \mathbf{1}$.

Step 1. Apply gradient projection iterations (see below for more detail) to (18) with initial guess $\mathbf{x}^j$ until stopping criteria are satisfied. Output the updated $\mathbf{x}^j \geq 0$.

Step 2. Compute a quadratic Taylor series approximation q_j of q centered at $\mathbf{x}^j$ and restricted to the indices r such that $x_r^j > 0$ (see below for more detail). Use the conjugate gradient (CG) iteration to minimize q_j until stopping criteria are satisfied. Use the most recent CG iterate $\mathbf{p}^j$ as a search direction in a projected backtracking line search. Output $\mathbf{x}^{j+1}$.

Step 3. If the outer iteration stopping criteria have been met, end the GPCG iterations. Otherwise, set $j = j + 1$ and return to Step 1.

Step 1, gradient projection iteration: The gradient projection (GP) iteration [2, 11] used in Step 1 of GPCG is defined as follows: given $\mathbf{x}^i \geq 0$ compute $\mathbf{x}^{i+1}$ via

$$\begin{aligned} &\mathbf{p}^i = \mathbf{c} - \mathbf{B}\mathbf{x}^i, \\ &\alpha^i = \arg\min_{\alpha>0} q(\mathcal{P}(\mathbf{x}^i + \alpha\mathbf{p}^i)), \\ &\mathbf{x}^{i+1} = \mathcal{P}(\mathbf{x}^i + \delta^i\mathbf{p}^i). \end{aligned} \tag{19}$$

Here $\mathcal{P}(\mathbf{x}) = \max\{\mathbf{x}, \mathbf{0}\}$, where the maximum is computed component-wise. In practice, the subproblem (19) is solved inexactly using a projected backtracking line search algorithm.

GP is stopped once a sufficient decrease condition, given in [12], is satisfied. For more detail on the GP implementation within GPCG, see [12].

Step 2, conjugate gradient iteration: The quadratic Taylor series approximation of q used in Step 2 of iteration k of the GPCG algorithm is of the form

$$q_j(\mathbf{p}) = q(\mathbf{x}^j) + \langle \nabla_{\text{red}} q(\mathbf{x}^j), \mathbf{p}\rangle + \frac{1}{2}\langle \nabla^2_{\text{red}} q(\mathbf{x}^j)\,\mathbf{p}, \mathbf{p}\rangle, \tag{20}$$

where

$$[\nabla_{\text{red}} q(\mathbf{x}^j)]_r = \begin{cases} [\mathbf{B}\mathbf{x}^j - \mathbf{c}]_i, & x_r > 0 \\ 0, & x_r = 0, \end{cases}$$

and

$$[\nabla^2_{\text{red}} q(\mathbf{x}^j)]_{rs} = \begin{cases} [\mathbf{B}]_{rs}, & \text{if } x_r > 0 \text{ and } x_s > 0 \\ \delta_{rs}, & \text{otherwise.} \end{cases}$$

where $\delta_{rs} = 1$ if $r = s$ and is 0 otherwise.

After an approximate minimizer $\mathbf{p}^j$ of q_j has been computed by the CG iterative method, equipped with the stopping rule of [12], a backtracking line search is performed, with $\mathbf{p}^j$ as the search direction, to guarantee that $q(\mathbf{x}^{j+1}) < q(\mathbf{x}^j)$.

4. Numerical Experiments

In this section, we show how the output from our MCMC algorithm can be used, both to give point estimates as well as to allow for the quantification of uncertainty in those estimates. First, however, we need a means for assessing MCMC chain convergence.

4.1. Assessing MCMC chain convergence

As in [1], we follow the recommended approach presented in [6], which requires the computation of multiple MCMC chains, with radonomly chosen starting points. With multiple chains in hand, a statistic for each sampled parameter is then computed, whose value provides a measure of convergence.

This statistic is defined as follows. Suppose we compute n_r parallel chains, each of length n_s (after discarding the first half of the simulations), and that $\{\psi_{ij}\}$, for $i = 1, \ldots, n_s$ and $j = 1, \ldots, n_r$, is the collection of samples of a single parameter. Then we define

$$B = \frac{n_s}{n_r - 1} \sum_{j=1}^{n_r} (\overline{\psi}_{\cdot j} - \overline{\psi}_{\cdot\cdot})^2, \quad \text{where} \quad \overline{\psi}_{\cdot j} = \frac{1}{n_s} \sum_{i=1}^{n_s} \psi_{ij}, \quad \text{and} \quad \overline{\psi}_{\cdot\cdot} = \frac{1}{n_r} \sum_{j=1}^{n_r} \overline{\psi}_{\cdot j}\,;$$

and

$$W = \frac{1}{n_r} \sum_{j=1}^{n_r} s_j^2, \quad \text{where} \quad s_j^2 = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} (\psi_{ij} - \overline{\psi}_{\cdot j})^2.$$

Note that $\overline{\psi}_{\cdot j}$ and $\overline{\psi}_{\cdot\cdot}$ are the individual chain mean and overall sample mean, respectively. Thus B provides a measure of the variance between the n_r chains, while W provides a measure of the variance within individual chains.

The marginal posterior variance $\text{var}(\psi|\mathbf{b})$ can then be estimated by

$$\widehat{\text{var}}^+(\psi|\mathbf{b}) = \frac{n_s - 1}{n_s} W + \frac{1}{n_s} B,$$

which is an unbiased estimate under stationarity [6]. The statistic of interest to us, however, is

$$\widehat{R} = \sqrt{\frac{\widehat{\text{var}}^{+}(\psi|\mathbf{b})}{W}}, \tag{21}$$

which declines to 1 as $n_s \to \infty$.

Once $\widehat{R}$ is 'near' 1 for all sampled parameters, the $n_s n_r$ samples from the last half of all of the sequences together can be treated as samples from the target distribution [6]. We will see that in our examples, $\widehat{R}$ will be much nearer to 1 than 1.1, which is deemed acceptable in [6].

4.2. A one-dimensional image deblurring example

We begin with a one-dimensional example (from [15]), because in one-dimensions it is more straightforward to visualize uncertainty from the MCMC samples. However, it is also the case that one-dimensional inverse problems arise in applications.

The data model is of the form (1), with $\mathbf{b} = \mathbf{A}\mathbf{x}$ obtained via mid-point quadrature applied to the convolution equation

$$b(s) = \int_0^1 A(s - s')x(s')ds',$$

with a Gaussian convolution kernel $A(s) = \exp(-s^2/(2\gamma^2))/\sqrt{\pi\gamma^2}$, $\gamma > 0$. Then $\mathbf{A}$ has the form

$$[\mathbf{A}]_{ij} = h \exp\left(-((i-j)h)^2/(2\gamma^2)\right)/\sqrt{\pi\gamma^2}, \quad 1 \le i, j \le n, \tag{22}$$

where $h = 1/n$ with n the number of grid points in $[0, 1]$. We use $n = 80$, and the resulting $\mathbf{A}$ has full column rank with condition number on the order of 10^{16}, resulting in a severely ill-conditioned problem. The true image is given by the solid line in Figure 1, and the data $\mathbf{b}$ is generated using (1), with the noise variance σ^2 chosen so that the noise strength is 2% that of the signal strength.

In each of the three tests below – unconstrained, positivity constrained, and nonnegativity constrained MCMC – we sample from the posterior density $p(\mathbf{x}, \lambda, \delta|\mathbf{b})$ by computing 20 MCMC chains each of length 500. Moreover, the initial values δ_0 and λ_0 in Step 0 are chosen randomly from the uniform distributions $U(0, 0.5)$ and $U(2, 8)$, respectively.

We present the unconstrained results from [1] for comparison. The computation of the samples in this case took approximately 36 seconds. The maximum $\widehat{R}$ value at the end of the run was 1.011. Direct image sampling (10) was used in Step 1 of the (unconstrained) MCMC algorithm via a Cholesky factorization of $\lambda_k \mathbf{A}^T\mathbf{A} + \delta_k \mathbf{C}$. From the image samples, on the left in Figure 2, we plot the median (or 0.5 quantile) as our reconstruction, and 95% credibility images given by the 0.025 and 0.975 quantiles of the samples at each pixel; these three quantiles were computed using MATLAB's `quantile` function. From the samples for λ and δ, on the right in Figure 2, we plot histograms

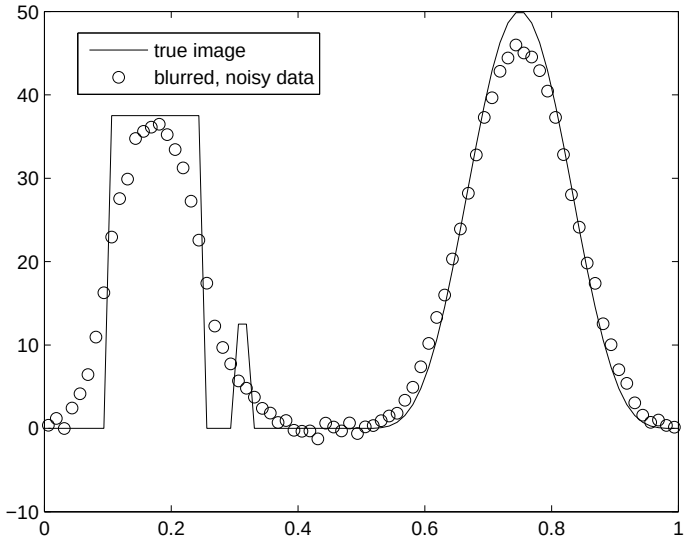

Figure 1. The one-dimensional true image and blurred noisy data.

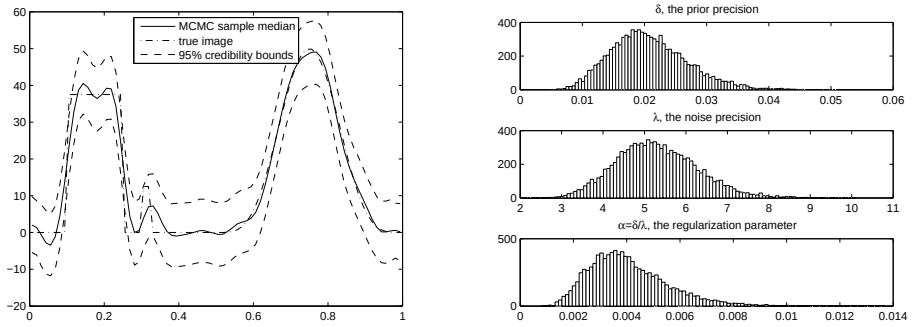

Figure 2. One-dimensional unconstrained example. On the left are plots of the median, and the 0.025 and 0.975 quantiles of the image samples. On the right are histograms of the samples of the precision parameters δ and λ, as well as of the regularization parameter $\alpha = \delta/\lambda$.

for λ, δ, and the regularization parameter $\alpha = \delta/\lambda$. Finally, as a verification of our statistical model, we note that the true noise precision $1/\sigma^2 \approx 5.35$ is contained within the 95% credibility interval for λ, $[3.94, 8.35]$, computed using MATLAB's `quantile` function.

For the positivity constrained case, the computation of the samples took approximately 52 seconds. The maximum $\widehat{R}$ value at the end of the run was 1.006. From the image samples, on the left in Figure 3, we plot the median image, the 0.025 quantile image, and the 0.975 quantile image. From the samples for λ and δ, on the right in Figure 3, we plot histograms for λ, δ, and the regularization parameter $\alpha = \delta/\lambda$. The reconstructed image appears to be reasonably accurate, however, for this example, we note that the true noise precision $1/\sigma^2 \approx 5.35$ is *not* contained within the 95%

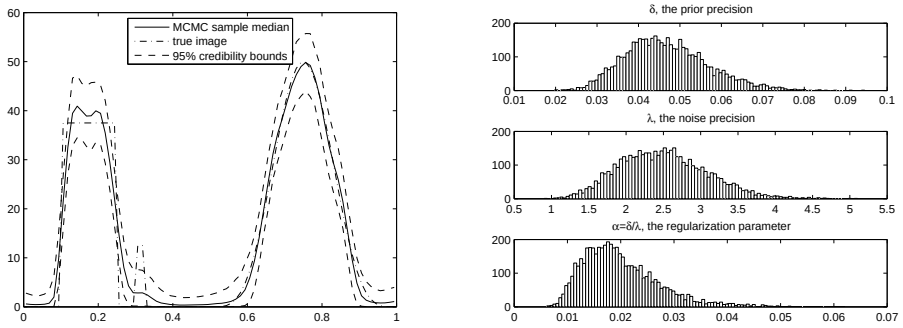

Figure 3. One-dimensional positivity constrained example. On the left are plots of the median, and the 0.025 and 0.975 quantiles of the image samples. On the right are histograms of the samples of the precision parameters δ and λ, as well as of the regularization parameter $\alpha = \delta/\lambda$.

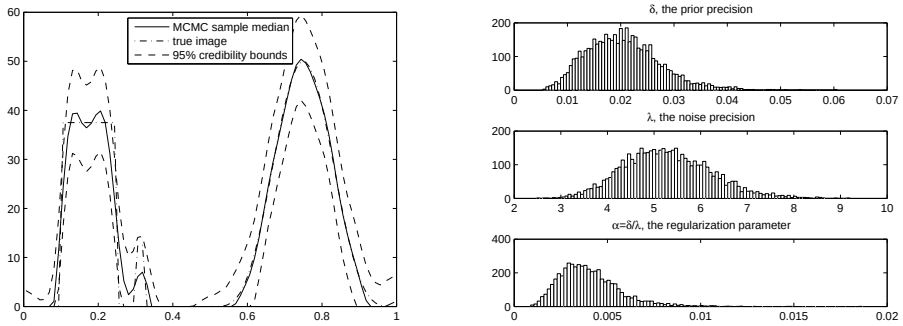

Figure 4. One-dimensional nonnegativity constrained example. On the left are plots of the median, and the 0.025 and 0.975 quantiles of the image samples. On the right are histograms of the samples of the precision parameters δ and λ, as well as of the regularization parameter $\alpha = \delta/\lambda$.

credibility interval for λ, $[1.48, 3.73]$, computed using MATLAB's `quantile` function, suggesting that the positivity constrained statistical model is incorrect. In particular, it seems reasonable that, for this example, there is a positive probability of certain pixel values being zero.

For the nonnegativity constrained case, the computation of the samples took approximately 49 seconds. The maximum $\widehat{R}$ value at the end of the run was 1.002. We use the GPCG method to approximately solve the nonnegativity constrained quadratic programming problem in Step 1, with a maximum of 5 gradient projection and 20 CG iterations per outer GPCG iteration, and stopping tolerances of 10^{-6} for the relative decrease in the norm of the project gradient, and a maximum of 20 outer GPCG iterations. Output analogous to the previous two examples is given in Figure 4. Note that the value of $1/\sigma^2 \approx 5.35$ is within the 95% credibility interval [3.61, 7.28], which validates the approach. In addition, the reconstruction appears to be of a higher resolution than in the positivity constrained case.

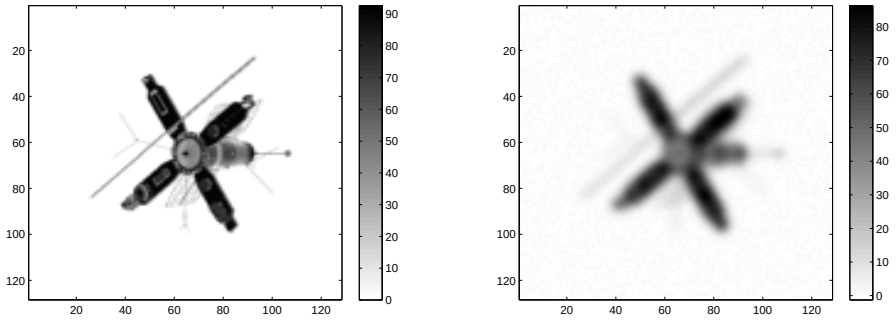

Figure 5. On the left is the true image, and on the right is the blurred noisy data.

4.3. A two-dimensional image deblurring example

In many inverse problems applications the spatial domain is two-dimensional. Thus we must show that our method is also effective on two-dimensional problems. Two-dimensional convolution has the form

$$b(s,t) = \int_0^1 \int_0^1 A(s-s', t-t')x(s',t')ds'dt'.$$

As above, we choose a Gaussian convolution kernel A, and discretize using mid-point quadrature on an 128×128 uniform computational grid over [0,1]×[0,1]. Periodic boundary conditions for the image are assumed, so that $\mathbf{A}$ is an $n^2 \times n^2$ block circulant with circulant blocks matrix, and hence is diagonalizable by the two-dimensional discrete Fourier transform (DFT) [7, 15]. Finally, the data $\mathbf{b}$ is generated using (1) with the noise variance σ^2 chosen so that the noise strength is 2% that of the signal strength. The true image and data are shown in Figure 5.

We sample from the posterior density $p(\mathbf{x}, \lambda, \delta|\mathbf{b})$ by computing 20 MCMC chains each of length 200. Moreover, the initial values δ_0 and λ_0 in Step 0 are chosen randomly from the uniform distributions $U(0, 0.5)$ and $U(5, 10)$, respectively.

For the unconstrained constrained sampler of [1], the computation took approximately 30 seconds. The maximum $\widehat{R}$ value at the end of the run was $\widehat{R} = 1.074$. The diagonalization of $\lambda_k \mathbf{A}^T\mathbf{A} + \delta_k \mathbf{C}$ obtained via the discrete Fourier transform was used to make direct image sampling (10) efficient. We plot the median of the sampled images as the reconstruction in Figure 6. From the samples for λ and δ, on the right in Figure 6, we plot histograms for λ, δ, and the regularization parameter $\alpha = \delta/\lambda$. The approach is validated by the fact that the true noise precision $1/\sigma^2 = 8.09$ is contained within the sample 95% credibility interval [7.59, 8.63].

For the nonnegativity constrained sampler, the computation took approximately 2845 seconds. The maximum $\widehat{R}$ value at the end of the run was $\widehat{R} = 1.026$. We implement GPCG for the nonnegativity constrained image sampling step with the same stopping criteria as in the one-dimensional case. We plot the median of the sampled images as the reconstruction in Figure 7. From the samples for λ and δ, on the right in Figure 7, we plot histograms for λ, δ, and the regularization parameter $\alpha = \delta/\lambda$.

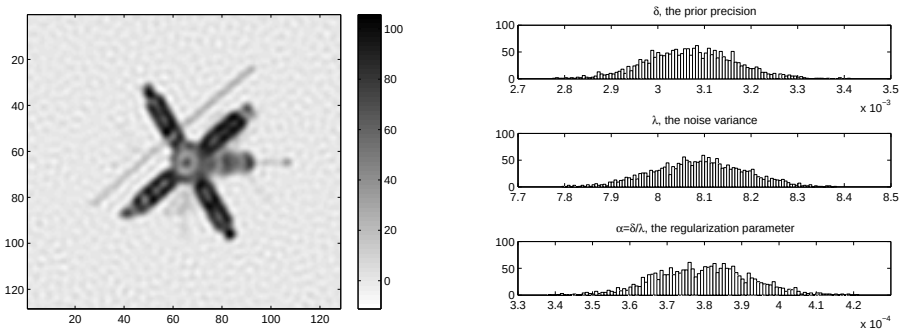

Figure 6. Two-dimensional unconstrained example. On the left is the median image. On the right are histograms of the samples of the precision parameters δ and λ, as well as of the regularization parameter $\alpha = \lambda/\delta$.

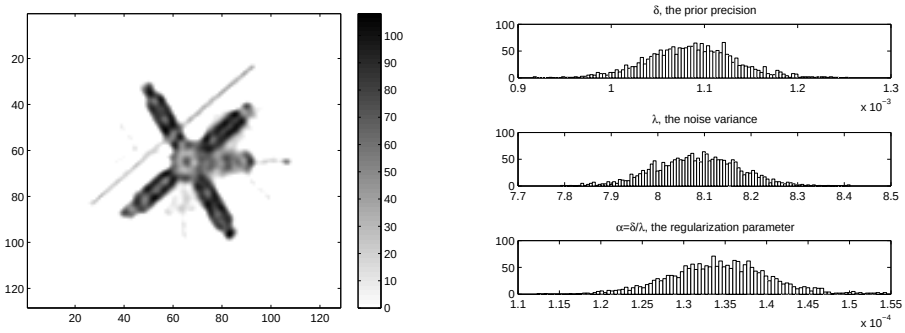

Figure 7. Two-dimensional nonnegativity constrained example. On the left is the median image. On the right are histograms of the samples of the precision parameters δ and λ, as well as of λ/δ.

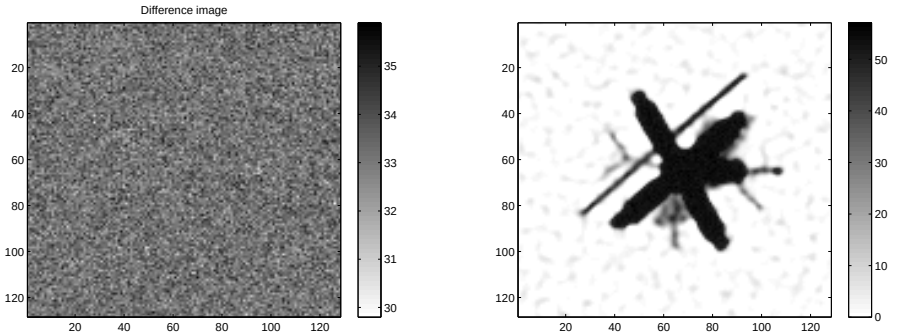

Figure 8. Plot of the difference between the upper and lower 95% credibility images. On the left is the plot for the unconstrained case, and on the right is a plot for the nonnegativity constrained case.

The approach is validated by the fact that the true noise precision $1/\sigma^2 = 8.09$ is contained within the sample 95% credibility interval [7.59, 8.63]. Note that visually, the nonnegativity constrained median image seems to be of higher resolution.

It remains to quantify the uncertainty in the median image estimates for both of the two-dimensional examples. This is more difficult than in one-dimension because it is not possible to plot the median image together with the 95% confidence images $\mathbf{x}^{95}_{\text{upper}}$ and $\mathbf{x}^{95}_{\text{lower}}$, which as in one-dimension, look like the median image shifted up and down, respectively. Thus, for both the unconstrained and nonnegativity constrained problems, we plot $\mathbf{x}^{95}_{\text{upper}} - \mathbf{x}^{95}_{\text{lower}}$ in Figure 8. Note in the unconstrained case (on the left), the uncertainty is uniform throughout the image, whereas in the nonnegativity constrained case (on the right), the uncertainty is higher in the higher intensity regions, and is very small (in fact, in many cases 0) in the zero-intensity regions of the true image.

4.4. Computed tomography

Computed tomography (CT) involves the reconstruction of the mass absorption function x of a body from one-dimensional projections of that body. A particular one-dimensional projection is obtained by integrating x along all parallel lines making a given angle ω with an axis in a fixed coordinate system. Each line L can be uniquely represented in this coordinate system by ω together with its perpendicular distance d to the origin. Suppose $L(\omega, d) = \{z(s) \mid 0 \leq s \leq S\}$, with an X-ray source located at $z(0)$ and a sensor at $z(S)$. The Radon transform model for CT has the form

$$b(\omega, d) = \int_{L(\omega,d)} x(z(s))\, ds. \tag{23}$$

A discretized version of (23) is what is solved in the CT inverse problem, where b corresponds to collected data, and x is the unknown. The discretization occurs both in the spatial domain, where x is defined, as well as in the Radon transform domain, where b is defined and the independent variables are ω and d. We will use a uniform $n \times n$ spatial grid and a uniform $n_\omega \times n_d$ grid in the Radon transform domain. Then, after column-stacking the resulting two-dimensional arrays, we obtain a matrix-vector system of the form (1), where the data vector $\mathbf{b} \in \mathbb{R}^{n_\omega n_d}$, the unknown vector $\mathbf{x} \in \mathbb{R}^{n^2}$, and the forward model matrix $\mathbf{A} \in \mathbb{R}^{n^2 \times (n_\omega n_d)}$. We generate data, once again, using (1) with $n = n_\omega = n_d = 100$ and σ^2 chosen so that the noise power is 2% that of the image. The truth image used is on the left in Figure 9 (generated using MATLAB's `phantom` function) and the data (called a sinogram) is given on the right.

We sample from the posterior density $p(\mathbf{x}, \lambda, \delta|\mathbf{b})$ by computing 5 MCMC chains each of length 150 in both the nonnegativity constrained and unconstrained cases. The maximum $\widehat{R}$ value at the end of the run was $\widehat{R} = 1.053$ in the unconstrained case and $\widehat{R} = 1.042$ in the constrained case. In both cases, the initial values δ_0 and λ_0 were chosen in Step 0 from the uniform distributions $U(0, 0.5)$ and $U(5, 10)$, respectively. In the unconstrained case, the conjugate gradient method was used for solving (14) with a stopping tolerance of 10^{-6} for the gradient norm and a maximum of 60 iterations; the later stopping rule was satisfied in all cases, suggesting that for CT, standard CG has has a difficult time converging in a reasonable number of iterations. This could have the effect that the Markov property of the chain is lost. However, in the constrained

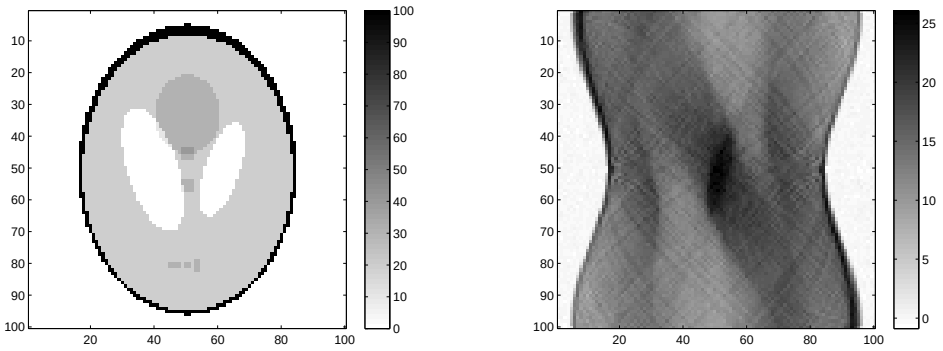

Figure 9. On the left is the two-dimensional true image, and on the right is the blurred noisy data.

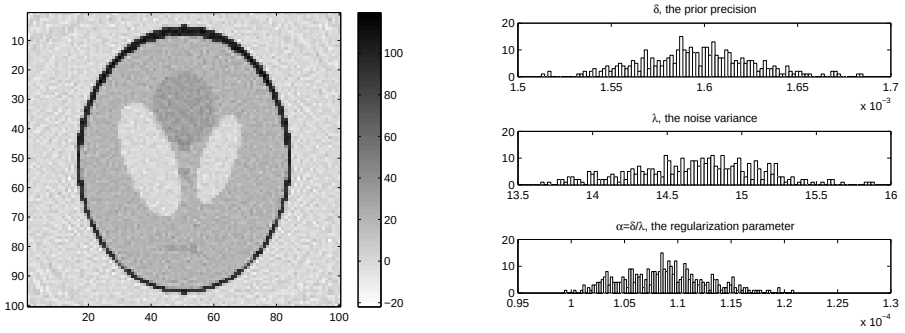

Figure 10. Computed tomography unconstrained example. On the left is the median image. On the right are histograms of the samples of the precision parameters δ and λ, as well as of λ/δ.

case, the projected gradient norm stopping tolerance of 10^{-6} is satisfied in every case by GPCG (with the iteration parameter values as in the above image deblurring examples), suggesting that the method is, indeed, converging.

The unconstrained sampler was also implemented on the tomography example in [1], though there more extensive tests were performed. The computation here took 2637 seconds. We plot the median of the sampled images as the reconstruction in Figure 10. From the samples for λ and δ, on the right in Figure 10, we plot histograms for λ, δ, and the regularization parameter $\alpha = \delta/\lambda$. As was also noted in the analogous experiment in [1], in this case $1/\sigma^2 = 13.24$ is not contained within the sample 95% credibility interval [13.85, 15.47].

For the constrained sampler, the computations took 8301 seconds, however as was stated above, the gradient stopping tolerance was met for each sample. We plot the median of the sampled images as the reconstruction in Figure 11. From the samples for λ and δ, on the right in Figure 11, we plot histograms for λ, δ, and the regularization parameter $\alpha = \delta/\lambda$. Contrary to the unconstrained case, the nonnegativity constrained sampler yields a 95% credibility interval [12.61, 13.53] for λ containing the true value

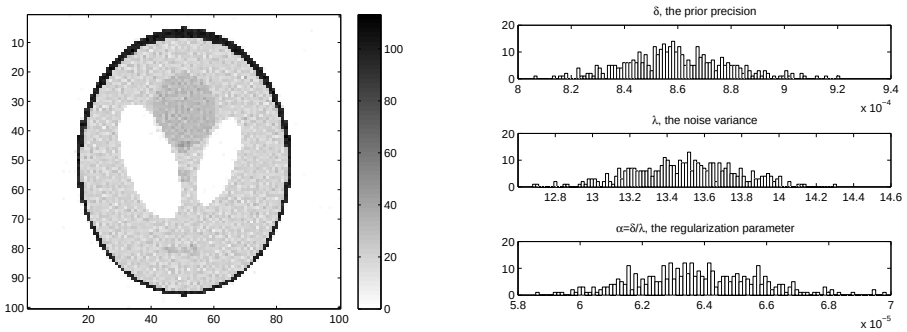

Figure 11. Computed tomography unconstrained example. On the left is the median image. On the right are histograms of the samples of the precision parameters δ and λ, as well as of λ/δ.

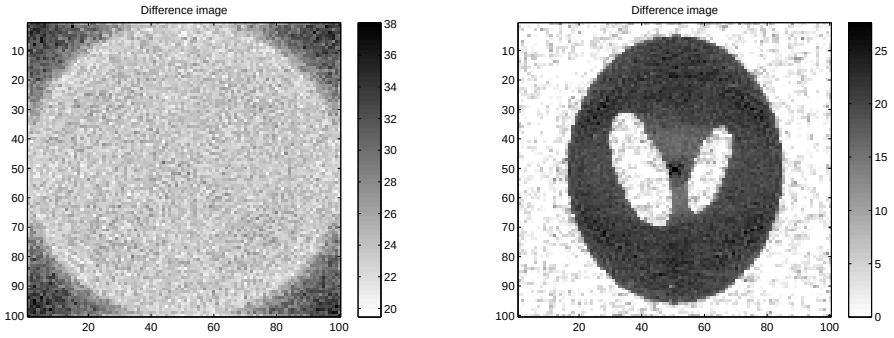

Figure 12. Plot of the difference between the upper and lower 95% credibility images. On the left is the plot for the unconstrained case, and on the right is a plot for the nonnegativity constrained case.

$1/\sigma^2 = 13.24$.

It remains to quantify the uncertainty in the median image estimates. As above, for both the unconstrained and nonnegativity constrained problems, we plot $\mathbf{x}^{95}_{\text{upper}} - \mathbf{x}^{95}_{\text{lower}}$ in Figure 8. Note in the unconstrained case (on the left), the uncertainty is uniform throughout the central disc of the image – and interestingly not in the corner regions – whereas in the nonnegativity constrained case (on the right), the uncertainty is higher in the higher intensity regions, and is very small (in fact, in many cases 0) in the zero-intensity regions of the true image, including in the corners and in the lungs. This result coincides, more or less, with the difference images in the two-dimensional image deblurring case.

5. Conclusions

We consider linear inverse problems with nonnegativity constraints and present an MCMC method for such problems, whose output (samples) can be used to obtain both estimates of unknowns, as well as for quantifying uncertainty in those estimates.

Significantly, *a priori* choice of tuning parameters, such as the regularization parameter in classical inverse problems, is not required. However, each sample (and hundreds must be computed) requires the solution of a large-scale nonnegativity constrained quadratic program, so the approach is computationally intensive. We advocate the use of the gradient projection-conjugate gradient method (GPCG) [12] for solving the large-scale nonnegativity constrained quadratic program; to our knowledge it is the most efficient method for such problems.

For comparison, we also present a positivity constrained MCMC sampler closely related to previous literature and show that the nonnegativity constrained sampler yields better results. Moreover, we provide a theorem interpreting the probability density implicity defined by the addition of the nonnegativity constraint.

Our tests focus on image deconvolution examples in both one- and two-dimensions, as well as on computed tomography. The results in one-dimension are used to argue that the nonnegativity constrained sampler gives better results than the positivity constrained sampler in the image deblurring case. Due to the component-wise nature of the positivity contrained sampler, it's use on two-dimensional problems is not feasible on the laptop used for the tests. In addition, in one-dimension, uncertainty quantification is more easily visualized.

In order to show that the approach is efficient enough to be used on large-scale problems, we present two-dimensional image deblurring and computed tomography results. In all cases, the approach is effective. In the computed tomography case, it is worth noting that the predictability intervals for the samples of the noise precision λ contain the true value in the nonnegativity constrained case, but not in the unconstrained case, and this is true over repeated experiments.

Acknowledgements

This work was supported by the National Science Foundation under grant DMS-0915107. The author would like to thank the University of Montana and the Department of Physics at the University of Otago, New Zealand, and Dr. Colin Fox in particular, for their support during his 2010-11 sabbatical year.

References

[1] J. M. Bardsley, *An MCMC Method for Estimation and Uncertainty Quantification in Linear Inverse Problems*, submitted, Math Sciences, University of Montana, Tech. Report #28, 2010.

[2] P. H. Calamai and J. J. Moré, *Projected Gradient Methods for Linearly Constrained Problems*, Mathematical Programming, **39** (1987), pp. 93–116.

[3] D. Calvetti and E. Somersalo, *Introduction to Bayesian Scientific Computing*, Springer 2007.

[4] D. Calvetti and E. Somersal, *Hypermodels in the Bayesian Imaging Framework*, Inverse Problems, **24(3)**, 2008, doi: 10.1088/0266-5611/24/3/034013.

[5] T. Cui and C. Fox, *Computational Inference for Inverse Problems*, International Society for Bayesian Analysis Bulletin, **17(3)**, 2010, pp. 12-15.

[6] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis, Second Edition*, Chapman & Hall/CRC, Texts in Statistical Science, 2004.

[7] P. C. Hansen, J. Nagy, and D. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, Philadelphi, 2006.

[8] Dave Higdon, *A primer on space-time modelling from a Bayesian perspective*, Los Alamos Nation Laboratory, Statistical Sciences Group, Technical Report, LA-UR-05-3097.

[9] J. P. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen, *Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography*, Inverse Problems, **16(5)**, 2000, pp. 14871522.

[10] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer 2005.

[11] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999.

[12] J. J. Moré and G. Toraldo, *On the Solution of Large Quadratic Programming Problems with Bound Constraints*, SIAM Journal on Optimization, **1** (1991), pp. 93–113.

[13] G. Nicholls and C. Fox, *Prior modelling and posterior sampling in impedance imaging*, Bayesian Inference for Inverse Problems, Proc. SPIE 3459, 1998, pp. 116-127.

[14] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*, Chapman and Hall/CRC, 2005.

[15] C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.

[16] D. Watzenig and C. Fox, *A review of statistical modeling and inference for electrical capacitance tomography*, Measurement Science and Technology, **20(5)**, 2009, doi:10.1088/0957-0233/20/5/052002.