Bayesian inference and uncertainty quantification for image reconstruction with Poisson data[*]

Qingping Zhou[†], Xiaoqun Zhang[‡], and Jinglai Li[§]

Abstract. We provide a complete framework for performing infinite-dimensional Bayesian inference and uncertainty quantification for image reconstruction with Poisson data. In particular, we address the following issues to make the Bayesian framework applicable in practice. We first introduce a positivity-preserving reparametrization, and we prove that under the reparametrization and a hybrid prior, the posterior distribution is well-posed in the infinite dimensional setting. Second we provide a dimension-independent MCMC algorithm, based on the preconditioned Crank-Nicolson Langevin method, in which we use a primal-dual scheme to compute the offset direction. Third we give a method based on the model discrepancy to determine the regularization parameter in the hybrid prior. Finally we propose to use the obtained posterior distribution to detect artifacts in a recovered image. We provide an application example to demonstrate the effectiveness of the proposed method.

Key words. Bayesian inference, image reconstruction, Markov chain Monte Carlo, Poisson distribution, total variation, uncertainty quantification.

AMS subject classifications. 45Q05, 62F15, 65C05, 68U10, 94A08

1. Introduction. Image reconstruction involves constructing interpretable images of objects of interest from the data recorded by an imaging device [16]. Image reconstruction is usually cast as an inverse problem as one wants to determine the input to a system from the output of it. In most practical image reconstruction problems, the measurement and recording process is inevitably corrupted by noise, which renders the obtained data random. The statistical properties of the data have significant impact to the reconstruction results. In this work we shall focus on a special type of medical image reconstruction problems where the recorded data follows a Poisson distribution. The Poisson data usually arises in imaging problems where the unknown quantity of interest is an object which interacts with some known incident beam of photons or electrons [21]. A very important example of such problems is the Positron emission tomography(PET) [27, 3], a nuclear medicine imaging technique that is widely used in early detection and treatment follow up of many diseases, including cancer. In PET, the detection of signal is essential a photon counting process and as a result the data is well modeled by a Poisson distribution [3, 21]. The problem has attracted considerable research interests, and a number of methods have been developed to recover the image, e.g.,

[†]School of Mathematical Sciences, Institute of Natural Sciences, Shanghai Jiao Tong University, 800 Dongchuan Rd, Shanghai 200240, China, (zhou2015@sjtu.edu.cn).

[‡]Institute of Natural Sciences, School of Mathematical Sciences, and the MOE Key Laboratory of Scientific and Engineering Computing, Shanghai Jiao Tong University, 800 Dongchuan Rd, Shanghai 200240, China, (xqzhang@sjtu.edu.cn)..

[§]Corresponding Author, Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, UK (jinglai.li@liverpool.ac.uk).

[32, 35, 15], just to name a few.

On the other hand, the stochastic nature of the data also introduces uncertainty into the image reconstruction process, and as a result the image obtained is unavoidably subject to uncertainty. In practice, many important decisions such as diagnostics have to be made based on the images obtained. It is thus highly desirable to have methods that can not only compute the image but also quantify the uncertainty in the image obtained. To this end, the Bayesian inference method has become a popular tool for image reconstruction [23], largely thanks to its ability to quantify uncertainty in the obtained image. The Bayesian formulation has long been used to solve image reconstruction problems with Poisson data, e.g., [20, 25, 19]. We note, however, that most of the works in the early years focus on computing a point estimate, which is usually the maximum a posteriori (MAP) estimate in the Bayesian setting, because of the limited computational power available then. More recently, mounting interest has been directed to the computation of the complete posterior distribution, rather than a point estimate, of the image, for that it can provide the important uncertainty information of the reconstruction results. For example, a Markov chain Monte Carlo (MCMC) algorithm is developed to sample the posterior distribution of the image in [5], and a variational Gaussian approximation of the posterior is proposed in [2].

A serious challenge in the numerical implementation of the Bayesian image reconstruction is that in certain circumstances the inference results diverge with respect to resolution/discretization refinement, which is known as to be discretization variant or dimension dependent. To address the issue, Stuart [33] proposes an infinite dimensional framework, formulating the Bayesian inference problem in function spaces. Building on several existing works, we aim to provide in this work a complete framework for performing infinite dimensional Bayesian inference and uncertainty quantification for medical image reconstruction with Poisson data, while providing treatments of several issues surrounding the problem. Specifically we summarize the key ingredients of our Bayesian framework as the following. First, in the usual setup, the function of interest can be both positive and negative valued. However, in the Poisson problem, when the function is negative valued, it may cause the Poisson likelihood function to be undefined (see Section 2.3 for more details), which renders the posterior distribution ill-posed in the infinite dimensional setting. To tackle the issue, we introduce a reparametrization of the unknown image which ensures that the function of interest is always positive valued. Moreover, medical images are often subject to sharp jumps and here we use the TV-Gaussian (TG) hybrid prior distribution proposed in [36] to model the jumps in the function/image. Using the positivity-preserving reparametrization and the TG prior, we are able to show that the resulting posterior distribution is *well-posed* in the infinite dimensional setting, which, to the best of our knowledge, has not yet been done for the Poisson data model. Second, we consider the numerical implementation of the Bayesian inference. A main difficulty here is that many standard MCMC algorithms such as the well-known Metropolis-Hastings (MH) [31, 8], degenerate with respect to resolution refinement. In [13] the authors introduce a MCMC algorithm termed as the preconditioned Crank-Nicolson (pCN) method, the performance of which is independent of discretization dimensionality. The authors also provide a Langevin variant of the pCN algorithm in [13] which accelerates the sampling procedure by incorporating the local gradient information of the likelihood function. In our problem, the pCN-Langevin (pCNL) algorithm can not be used directly because the prior

used here has the total variation (TV) term which can be non-differentiable. To overcome this difficulty we modify the pCNL method by replacing the gradient direction with one computed by the primal-dual algorithm. We note that a similar problem is considered in [29, 14] where a proximal method is used to approximate the gradient direction. Other than that the directions are computed with different approaches, another main difference between the aforementioned works and the present one is that we use the pCN framework here so the algorithm is dimension independent, while the works [29, 14] concern finite dimensional problems where discretization refinement is not an issue. Third, an important issue in the TG hybrid prior is to determine the value of the regularization parameter of the TV term. In the Bayesian framework, such parameters are often determined with the hierarchical Bayes or the empirical Bayes method [18]. As discussed in Section 4, these methods, however, are computationally intractable in our problem as we do not know the normalization constant of the TG prior. Thus, in this work we provide a method to determine the value of the TV regularization parameter based on the realized discrepancy model fit assessment approach developed in [17]. Finally, we provide an application of the uncertainty information obtained in the Bayesian framework, where we use the posterior distribution to detect possible artifacts in any reconstructed image.

The rest of the paper is organized as the following. In Section 2, we present the infinite dimensional Bayesian formulation of the image reconstruction problem with Poisson data, and we prove that under the reparametrization the resulting posterior is well-posed in the function space. In Section 3, we describe the primal-dual pCN algorithm to sample the posterior distribution of the present problem. Section 4 provides a method to determine the value of the regularization parameter in the TG prior. Section 5 discusses how to use the posterior distribution to detect artifacts in a reconstructed image. Finally numerical experiments of the proposed Bayesian framework are performed in Section 6.

2. Infinite dimensional Bayesian image reconstruction with Poisson data. In this section, we formulate the image reconstruction with Poisson data in an infinite dimensional Bayesian framework.

2.1. The Bayesian inference formulation for functions. We start by presenting a generic Bayesian inference problem for functions. Let X be a separable Hilbert space of functions with inner product $\langle \cdot, \cdot \rangle_X$. Our goal is to infer $u \in X$ from data $y \in Y \subseteq R^d$ and $\mathbf{y}$ is related to u via the likelihood function $\pi(\mathbf{y}|u)$, i.e., the distribution of $\mathbf{y}$ conditional on the value of u. In the Bayesian setting, we first assume a prior distribution μ_{pr} of the unknown u, which represents one's prior knowledge on the unknown. In principle μ_0 can be any probabilistic measure defined on the space X. The posterior measure μ^y of u conditional on data y is provided by the Radon-Nikodym(R-N) derivative:

$$(2.1) \qquad \frac{d\mu^y}{d\mu_{\mathrm{pr}}}(u) = \pi(\mathbf{y}|u),$$

which can be interpreted as the Bayes' rule in the infinite dimensional setting. The posterior distribution μ^y thus can be computed from Eq. (2.1) with, for example, a MCMC simulation.

2.2. Poisson data model and the positivity-preserving reparametrization. To perform the Bayesian inference, we first need to specify the likelihood function, which can be derived

from the underlying mathematical model relating the data and to the unknown image. We assume that the image is first projected to the noise-free observable via a mapping $A : X \to Y$,

$$(2.2) \qquad \boldsymbol{\theta} = \mathrm{A}u.$$

While noting that the proposed framework is rather general, here for simplicity we restrict ourselves in the case where A is a bounded linear transform. For example in the PET imaging problems, the mapping A is the Radon transform, where each θ_i is computed by integrating $u(\mathbf{x})$ alone a line L_i:

$$(2.3) \qquad \theta_i = (\mathrm{A}u)_i = K \int_{L_i} u(\mathbf{x})|d\mathbf{x}|,$$

for $i = 1, ..., d$, where K is a positive constant describing the noise level. The Radon transform is a bounded linear transform [26]. Poisson noise is then applied to the *projected observable* $\boldsymbol{\theta}$, yielding the likelihood function $\pi(\mathbf{y}|u) = \pi_\mathrm{P}(\mathbf{y}|\boldsymbol{\theta} = \mathrm{A}u)$, where $\pi_\mathrm{P}(\mathbf{y}|\boldsymbol{\theta})$ is the d-dimensional Poisson distribution:

$$(2.4) \qquad \pi_\mathrm{P}(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \frac{(\theta_i)^{y_i} \exp(-\theta_i)}{y_i!}.$$

In the PET problem, there is an additional restriction: the unknown function u must be positive. The reason is two-fold: first from the physical point of view, the unknown u represents the density of the medium, which is positive; from a technical point of view, if u is not constrained to be positive, it may yield some negative components of the predicted data $\boldsymbol{\theta}$, which renders the Poisson likelihood un-defined. To this end, we need to introduce a transformation to preserve positivity of the unknown u. To impose the positivity constraint, we reparameterize the unknown u as:

$$u(\mathbf{x}) = f(z(\mathbf{x})) = \frac{a}{2}(\mathrm{erf}(z(\mathbf{x})) + b),$$

where a and b are two constants satisfying $a > 0$ and $b > 1$, and $\mathrm{erf}(\cdot)$ is the error function defined as:

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt.$$

With the new parametrization, it is easy to see that for any $\mathbf{x} \in \Omega$, we have

$$(2.5) \qquad \frac{1}{2}a(b-1) \le u(\mathbf{x}) \le \frac{1}{2}a(b+1).$$

Now we can infer the new unknown z and once z is known u can be computed accordingly. In this setup, the likelihood function for z becomes

$$\pi(\mathbf{y}|z) = \pi_P(\mathbf{y}|\boldsymbol{\theta} = \mathrm{A}f(z))$$

where $\pi_P(\mathbf{y}|\boldsymbol{\theta})$ is the d-dimensional Poisson distribution given by Eq. (2.4). Following [33], we can write the likelihood function $\pi(\mathbf{y}|z)$ in the form of

$$(2.6a) \qquad \pi(\mathbf{y}|z) \propto \exp(-\Phi(z, \mathbf{y})),$$

where

$$(2.6b) \qquad \Phi(z; \mathbf{y}) = \sum_{i=1}^{d} (Af(z))_i - y_i \ln(Af(z))_i.$$

For simplicity we can rewrite Eq. (2.6b) as,

$$(2.7) \qquad \Phi(z; \mathbf{y}) = \langle Af(z), \mathbf{1} \rangle - \langle \mathbf{y}, \ln(Af(z)) \rangle = \langle \boldsymbol{\theta}, \mathbf{1} \rangle - \langle \mathbf{y}, \ln \boldsymbol{\theta} \rangle,$$

where $\mathbf{1}$ is a d-dimensional vector whose components are all one, $\boldsymbol{\theta} = Af(z)$ is the predicted observable, and $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product. In what follows we often omit the argument y and simply use $\Phi(z)$, when not causing ambiguity. This notation will be used often later.

2.3. Bayesian framework with the hybrid prior for PET imaging.

We now describe how the infinite dimensional Bayesian inference framework is applied to the PET problem. First we assume that the unknown function z is a function defined on Ω, a bounded open subset of R^2. In particular, we set the state space X to be the Sobolev space $H^1(\Omega)$:

$$X = H^1(\Omega) = \{z(\mathbf{x}) \in L_2(\Omega) \, | \, \partial_{x_1} z, \, \partial_{x_2} z \in L_2(\Omega) \text{ for all } \mathbf{x} = (x_1, \, x_2) \in \Omega \,\}.$$

The associated norm $\| \cdot \|_X = \| \cdot \|_{H^1}$ is

$$\|z\|_{H^1}^2 = \|z\|_{L_2(\Omega)}^2 + \|\partial_{x_1} z\|_{L_2(\Omega)}^2 + \|\partial_{x_2} z\|_{L_2(\Omega)}^2.$$

Choosing a good prior distribution is one of the most important issues in Bayesian inference. Conventionally one often assumes that the prior on z, is a Gaussian measure defined on X with mean ξ covariance operator C_0, i.e. $\mu_{\mathrm{pr}} = \mu_0 = N(\xi, C_0)$. Note that C_0 is symmetric positive and of trace class. The Gaussian prior has many theoretical and practical advantages, but a major limitation of the Gaussian prior is that it can not model functions with sharp jumps well.

To address the issue, here we use the TV-Gaussian prior proposed in [36]:

$$(2.8) \qquad \frac{d\mu_{\mathrm{pr}}}{d\mu_0}(z) \propto \exp(-R(z)), \quad R(z) = \lambda \|z\|_{\mathrm{TV}}.$$

where $\| \cdot \|_{\mathrm{TV}}$ is the TV seminorm,

$$(2.9) \qquad \|z\|_{\mathrm{TV}} = \lambda \int_\Omega \|\nabla u\|_2 dx,$$

and λ is a positive constant. It follows immediately that the R-N derivative of μ^y with respect to μ_0 is

$$(2.10) \qquad \frac{d\mu^y}{d\mu_0}(z) \propto \exp(-\Phi(z) - R(z)),$$

which returns to the conventional formulation of inference with a Gaussian prior. Thus all the methods developed for inference problems with Gaussian priors can be directly applied to our formulation.

Next we shall show that the formulated Bayesian inference problem is well defined in the infinite dimensional setting. We first show that $\Phi(z)$ given by Eq. (2.7) satisfies certain important conditions, as is stated by Proposition 2.1.

Proposition 2.1. *The functional Φ given in Eq. (2.7) has the following properties:*
1. *For every $r > 0$, there are constants $M(r) \in \mathbb{R}$ and $N(r) > 0$ such that, for all $z \in X$, and $y \in Y$ with $\|y\|_2 < r$,*

$$M \leq \Phi(z) \leq N;$$

2. *For every $r > 0$ there is a constant $M(r) > 0$ such that, for all $z, v \in X$ with $\max\{\|z\|_X, \|v\|_X\} < r$,*

$$|\Phi(z) - \Phi(v)| \leq M\|z - v\|_X;$$

3. *There exists a constant $M > 0$ such that for any $y, y' \in Y$, we have*

$$|\Phi(z; y) - \Phi(z; y')| \leq M\|y - y'\|_2.$$

A detailed proof of the proposition is provided in Appendix A. Following Proposition 2.1, we can conclude that the hybrid prior (2.8), and the log-likelihood function Eq. (2.7) yield a well-behaved posterior measure given by Eq. (2.10) in the infinite-dimensional setting, as is summarized in the following theorem:

Theorem 2.2. *For $\Phi(z)$ given by Eq. (2.7) and prior measure μ_{pr} given by Eq. (2.8), we have the following results:*
1. *μ^y given by Eq. (2.10) is a well-defined probability measure on X.*
2. *μ^y given by Eq. (2.10) is Lipschitz in the data y, with respect to the Hellinger distance: if μ^y and $\mu^{y'}$ are two measures corresponding to data y and y' the there exists $C = C(r)$ such that, for all y, y' with $\max\{\|y\|_2, \|y'\|_2\} < r$,*

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y'}) \leq C\|y - y'\|_2.$$

3. *Let*

$$(2.11) \qquad \frac{d\mu^y_{N_1, N_2}}{d\mu_0} = \exp(-\Phi_{N_1}(z) - R_{N_2}(z)),$$

where $\Phi_{N_1}(z)$ is a $N_1 \in \mathbb{N}$ dimensional approximation of $\Phi(z)$ and $R_{N_2}(z)$ is a $N_2 \in \mathbb{N}$ dimensional approximation of $R(z)$. Assume that Φ_{N_1} satisfies the three properties of Proposition 2.1 with constants uniform in N_1, and R_{N_2} satisfy Assumptions A.2 (i) and (ii) in [36] with constants uniform in N_2. Assume also that for $\forall \epsilon > 0$, there exist two positive sequences $\{a_{N_1}(\epsilon)\}$ and $\{b_{N_2}(\epsilon)\}$ converging to zero, such that $\mu_0(X_\epsilon) \geq 1 - \epsilon$ for $\forall N_1, N_2 \in \mathbb{N}$, where

$$X_\epsilon = \{z \in X \,|\, |\Phi(z) - \Phi_{N_1}(z)| \leq a_{N_1}(\epsilon), \, |R(z) - R_{N_2}(z)| \leq b_{N_2}(\epsilon)\}.$$

Then we have

$$d_{\mathrm{Hell}}(\mu^y, \mu^y_{N_1, N_2}) \to 0 \quad \text{as} \quad N_1, N_2 \to +\infty.$$

Theorem 2.2 is a direct consequence of Proposition 2.1, and the proof of the theorem can be found in [36] and is omitted here.

Finally, we note that, in the numerical implementations, we use the truncated Karhunen-Loève (KL) expansion [24] to represent the unknown z. Namely, we write z as

$$(2.12) \qquad z(\mathbf{x}) = \sum_{i=1}^{N} z_i \sqrt{\eta_i} e_i(\mathbf{x}),$$

where $\{\eta_i, e_i(\mathbf{x})\}$ are the eigenvalue-eigenfunction pair of the covariance operator C_0, and $(z_1, ..., z_N)$ are independent with each following a standard normal distribution. In the KL representation, the number of KL modes (eigenfunctions) N corresponds to the discretization dimensionality.

3. The primal-dual preconditioned Crank-Nicolson MCMC algorithm. In most practical image reconstruction problems, the posterior distribution can not be analytically calculated. Instead, one usually represent the posterior by samples drawn from it using MCMC algorithms. It is demonstrated in [13] that standard MCMC algorithms may become problematic in the infinite dimensional setting: its acceptance probability will generate to zero as the discretization dimensionality increases. Here we adopt the pCN MCMC algorithm particularly developed for the infinite dimensional problems [13]. An important feature of the pCN MCMC algorithm is that its sampling efficiency is independent of discretization dimensionality up to the numerical errors in the evaluation of the functionals $R(\cdot)$ and $\Phi(\cdot)$, which makes it particular useful for sampling the posterior distribution defined in function spaces. We start with a brief introduction of the pCN algorithm following the presentation of [13]. We denote $\Phi(z) + R(z)$ of Eq. (2.10) as $\Psi(z)$. Simply speaking the algorithms are derived by applying the Crank-Nicolson (CN) discretization to a stochastic partial differential equation whose invariant distribution is the posterior. We here omit the derivation details while referring interested readers to [13], and jump directly to the pCN proposal:

$$(3.1) \qquad v = (1 - \beta^2)^{\frac{1}{2}} z + \beta w,$$

where z and v are the present and the proposed positions respectively, $w \sim N(\xi, C_0)$ and $\beta \in [0, 1]$ is the parameter controlling the stepsize of the algorithm. The proposed sample v is then accepted or rejected according to the acceptance probability:

$$(3.2) \qquad a(v, z) = \min\{1, \exp[\Psi(z) - \Psi(v)]\},$$

which is independent of discretization dimensionality up to numerical errors.

The pCN proposal in Eq. (3.1) can be improved by incorporating the data information in the proposal, and following the idea of Langevin MCMC for the finite dimensional problems, one can derive the preconditioned Crank-Nicolson Langevin (pCNL) proposal:

$$(3.3) \qquad (2 + \delta)v = (2 - \delta)z - 2\delta \mathbf{C}_0 \mathcal{D}\Psi(z) + \sqrt{8\delta} w,$$

where $\delta \in [0, 2]$, $w \sim N(\xi, \mathbf{C}_0)$ and $\mathcal{D}$ is the gradient operator. If we define $\rho(z, v)$ as following:

$$(3.4) \qquad \rho(z, v) = \Psi(z) + \frac{1}{2}\langle v - z, \mathcal{D}\Psi(z)\rangle + \frac{\delta}{4}\langle z + v, \mathcal{D}\Psi(z)\rangle + \frac{\delta}{4}\|\mathbf{C}_0^{1/2}\mathcal{D}\Psi(z)\|^2,$$

then the acceptance probability is given by:

$$(3.5) \qquad a(z, v) = \min\{1, \exp\left(\rho(z, v) - \rho(v, z)\right)\}.$$

The pCNL algorithm is usually more efficient than the standard pCN algorithm as it takes advantage of the gradient information of the $\Psi(z)$. However, the pCNL algorithm can not be used directly in our problem as $\Psi(z)$ includes the TV term which is not differentiable. It is important to note that $-2\delta\mathbf{C}_0\mathcal{D}\Psi(z)$ is the offset term only affecting the mean of the proposal, and we can replace $\mathcal{D}\Psi(z)$ with another direction $g(z)$, yielding proposal:

$$(3.6) \qquad (2 + \delta)v = (2 - \delta)z - 2\delta\mathbf{C}_0 g + \sqrt{8\delta}w,$$

where $\delta \in [0, 2]$ and $w \sim N(\xi, \mathbf{C}_0)$. And it can be verified that the detailed balance is maintained if we change Eq. (3.4) to

$$(3.7) \qquad \rho(z, v) = \Psi(z) + \frac{1}{2}\langle v - z, g\rangle + \frac{\delta}{4}\langle z + v, g\rangle + \frac{\delta}{4}\|\mathbf{C}_0^{1/2}g\|^2,$$

while the acceptance probability is given by Eq. (3.5). Now we need to find a good direction $g(z)$. In [29] the authors use Moreau approximation to approximate the TV term in the Langevin MCMC algorithm. Here we shall provide an alternative approach, determining the offset direction in the MCMC iteration using the primal dual algorithm. The primal-dual algorithms are known to be very effective in solving optimization problems involving TV regularization [11, 12, 10], and we hereby give a brief description of the primal dual method applied to our problem. Suppose that we want to solve

$$(3.8) \qquad \min_{z \in X} \Psi(z) = \Phi(z) + \lambda\|z\|_{\mathrm{TV}}.$$

Introducing a new variable $\boldsymbol{\phi}(x) = [\phi_1(x), \phi_2(x)]$ with $\phi_1(\mathbf{x}), \phi_2(\mathbf{x}) \in L_2(\Omega)$ (we denote this as $\boldsymbol{\phi} \in L_2^2(\Omega)$), and we then rewrite the optimization problem (3.8) as

$$(3.9) \qquad \min_{z \in X, \phi \in L_2^2(\Omega)} \Psi(z, \phi) = \Phi(z) + \lambda\|\phi\|_{2,1}$$

$$\text{s.t.} \qquad \mathcal{D}z = \phi,$$

where $\|\phi\|_{2,1} = \left(\|\phi_1(\mathbf{x})\|_{L_2(\Omega)}^2 + \|\phi_2(\mathbf{x})\|_{L_2(\Omega)}^2\right)^{1/2}$. The augmented Lagrangian for Eq.(3.9) is

$$(3.10) \qquad \max_{\eta \in L_2^2(\Omega)} \min_{z \in X, \phi \in L_2^2(\Omega)} L_\rho(z, \phi, \eta) = \Phi(z) + \lambda\|\phi\|_{2,1} + \langle\eta, \mathcal{D}z - \phi\rangle + \frac{\rho}{2}\|\mathcal{D}z - \phi\|_2^2,$$

where $\eta \in L_2^2(\Omega)$ is the dual variable or Lagrange multiplier, and $\rho > 0$ is a constant called the penalty parameter. The resulting problem is then solved with the Alternating Direction Method of Multipliers (ADMM) [7]:

$$(3.11\text{a}) \qquad z^{k+1} = \arg\min_{z \in X} \ L_\rho(z, \phi^k, \eta^k),$$

$$(3.11\text{b}) \qquad \phi^{k+1} = \arg\min_{\phi \in L_2^q(\Omega)} \ L_\rho(z^{k+1}, \phi, \eta^k),$$

$$(3.11\text{c}) \qquad \eta^{k+1} = \eta^k + \rho(\phi^{k+1} - Dz^{k+1}).$$

The algorithm consists of a z-minimization step (3.11a), a ϕ–minimization step (3.11b), and a dual ascent step (3.11c). When applied to the MCMC algorithm, the iterations proceed as described in in Algorithm 3.1. It should be noted here that, in Step 3 of the algorithm, it is required to solve a minimization problem. However, for efficiency's sake, we choose not to solve this optimization accurately. In the numerical implementations in this work, we only perform one gradient descent iteration, and so the computational cost for each MCMC iteration does not increase much compared to standard pCN.

Algorithm 3.1 The Primal-dual pCN(PD-pCN) algorithm
___

1: Initialize z^0, ϕ^0, η^0
2: **for** $k = 0, 1, 2, \cdots$ **do**

3:      $z^* = \arg\min_{z \in X} \quad L_\rho(z, \phi^k, \eta^k)$;
4:      $g = z^* - z^k$;
5:      Propose v using Eq. (3.6)
6:      Draw $\theta \sim U[0, 1]$;
7:      Compute $a(z, v)$ with Eq. (3.5) and Eq. (3.7);
8:      **if** $\theta \leq a$ **then**
9:          $z^{k+1} = v$;
10:         $\phi^{k+1} = \arg\min_{\phi \in L_2^q(\Omega)} \quad L_\rho(z^{k+1}, \phi, \eta^k)$;
11:         $\eta^{k+1} = \eta^k + \rho(\phi^{k+1} - \mathcal{D}z^k)$;
12:     **else**
13:         $z^{k+1} = z^k$;
14:         $\phi^{k+1} = \phi^k$;
15:         $\eta^{k+1} = \eta^k$;
16:     **end if**
17: **end for**

---

4. Determining the TV regularization parameter. Just like the deterministic inverse problems, it is an important issue to determine the TV regularization parameter λ in the hybrid prior. In the Bayesian setting, standard methods such as the empirical Bayes (EB) approach [18] can be too computationally intensive in the present problem, especially as we have no knowledge of the normalization constant of the hybrid prior. Namely, recall that the EB method seeks to maximize

$$\pi(\mathbf{y}|\lambda) = \int \pi(\mathbf{y}|u)\mu_{\mathrm{pr}}(du) = \int \pi(\mathbf{y}|u)\frac{1}{Z(\lambda)}\exp(-\lambda\|u\|_{\mathrm{TV}})\mu_0(du)$$

where $Z(\lambda)$ is not known in advance and needs to be evaluated with another Monte Carlo integration. Here we provide a statistical approach to determine the value of λ, which is derived from the realized discrepancy method for model assessment proposed in [17].

The basic idea of the method is to choose a function $D(\mathbf{y}, \boldsymbol{\theta})$ that measures the discrepancy between the measured data $\mathbf{y}$ and the projected observable $\boldsymbol{\theta}$, and for the present problem we

use the χ^2 discrepancy,

$$D(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{d} \frac{(y_i - \theta_i)^2}{\theta_i^2}. \tag{4.1}$$

Now knowing that $\theta = \mathrm{A}u$, we can use this discrepancy to assess how well a specific choice of u fits the data. The classical p-value based on the discrepancy $D(\mathbf{y}, \boldsymbol{\theta})$ is

$$p_c(\mathbf{y}, \boldsymbol{\theta}) = \mathbb{P}[D(\mathbf{y}, \boldsymbol{\theta}) > D(\tilde{\mathbf{y}}, \boldsymbol{\theta})] \tag{4.2}$$

where $\tilde{\mathbf{y}}$ is the simulated data from model (2.4). In particular, for the discrepancy function given in Eq. (4.1), the p-value is simply,

$$p_c(\mathbf{y}, \boldsymbol{\theta}) = 1 - F_{\chi_d^2}(D(\mathbf{y}, \boldsymbol{\theta})), \tag{4.3}$$

where $F_{\chi_d^2}(\cdot)$ is the cumulative distribution function of the χ^2 distribution with the degree of freedom d. The classic p-value computed this way provides an assessment of how well a single estimate of u fits the data $\mathbf{y}$. The method can be extended to the Bayesian setting to assess the fitness of the posterior distribution to data. First recall that our prior distribution given by Eq. (2.8) is specified by the parameter λ, and as a result the posterior also depends on λ and here we write the posterior as $\mu_\lambda^y(du)$ to emphasize its dependence on λ. In the Bayesian setting, one can compute the posterior predictive p-value:

$$p_b(\mathbf{y}, \lambda) = \int p_c(\mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}, \lambda) d\boldsymbol{\theta} = \int p_c(\mathbf{y}, \mathrm{A}u) \mu_\lambda^y(du), \tag{4.4}$$

which is essentially the classical p-value averaged over the posterior distribution. The posterior predictive p-value assesses the fitness of the posterior distribution to the data: intuitively speaking, larger value of p_b indicates better fitness of the posterior to the data $\mathbf{y}$. However, one can not simply choose the value of λ that yields the largest value of p_b, or, equivalently the best fitness to the data, as that may cause overfitting. In other words, if the posterior fits the data "too well", it often implies that the effect of the prior distribution is so weak that the posterior is dominated by the data. In the image reconstruction problem, this situation is greatly undesirable, as the problem is highly ill-posed and we need significant contribution from the prior distribution to obtain good estimates of the unknown u. In this respect, we should choose λ in a way that the effects of the prior and data are well balanced, which should be indicated by an appropriate value of p_b. The ideal value of p_b is certainly problem dependent, and in the present problem we suggest to choose λ so that the resulting value of p_b is in the range of $0.1 \sim 0.5$. It is worth mentioning that methods using the data discrepancy to determine regularization parameters for Poisson data model have also been developed in the deterministic setting, and we refer to [4, 6] for further details. Finally we also note that, in addition to λ, the Gaussian distribution may also be subject to hyper-parameters, and in principle these hyper-parameters can be determined along with λ using the proposed approach. However, here we choose not to do so for two reasons: first determining multiple parameters may significantly increase the computational cost; second, as the Gaussian distribution is merely used as a reference measure in our hybrid prior, the posterior distribution is not sensitive to it, and it usually suffices to choose these hyper-parameters based upon certain prior information (for example, historical data).

5. Artifact detection using the posterior distribution. In practical image reconstruction problems, due to the imperfection of methods or devices, a reconstructed image may contain artifacts what are not present in the original imaged object. In this section we describe an application of the posterior distribution to detect artifacts in a reconstructed image.

Specifically, we consider the posterior distribution of the image at any given point $\mathbf{x}$, which is denoted as $u_{\mathbf{x}}$. Consequently we can write the posterior distribution of $u_{\mathbf{x}}$ as $\pi_{\mathbf{x}}(u_{\mathbf{x}}|\mathbf{y})$. Next we consider the highest posterior density interval (HPDI) which is essentially the narrowest interval corresponding to a given confidence level. More precisely, for an $\alpha \in [0,1]$, the $100(1-\alpha)\%$ HPDI is defined as (cite)

$$C_\alpha = \{u(\mathbf{x})|\pi_{\mathbf{x}}(u(\mathbf{x})|\mathbf{y}) > \pi_\alpha\}$$

where π_α is the largest constant satisfying $\mathbb{P}[u_{\mathbf{x}}|\pi_{\mathbf{x}}(u_{\mathbf{x}}|\mathbf{y}) > \pi_\alpha] = 1 - \alpha$. Now suppose that we have a reconstruct image $\hat{u}$ and we also write its value at $\mathbf{x}$ as $\hat{u}_{\mathbf{x}}$. Next we shall estimate how large the credible level $(1-\alpha)$ must be so that the associated HDPI may contain $\hat{u}_{\mathbf{x}}$. That is, we compute the smallest value of $(1-\alpha)$ such that,

$$\hat{u}_{\mathbf{x}} \in C_\alpha.$$

Intuitively speaking, the larger the computed credible level $(1-\alpha)$ is, the more likely the considered $\hat{u}_{\mathbf{x}}$ is an artifact. And we thus use the credible level $(1-\alpha)$ to measure how likely a point is an artifact, and we can do this test for any point $\mathbf{x} \in \Omega$. Alternatively, the problem can also be formulated as a Bayesian hypothesis test with a fixed α (e.g. $\alpha = 5\%$) [28]: that is, $\hat{u}(\mathbf{x})$ is regarded to be an artifact if it is not contained in the $(1-\alpha)$ HPDI for the prescribed value of α. However, it has been pointed out in [34] that performing hypothesis test with HPDI may cause certain theoretical issue and so here we choose not to use the hypothesis test formulation.

It should be noted here that, in [30, 14], the authors utilize the highest posterior density (HDP) region to test if a candidate image is likely to be a solution to the reconstruction problem. The purpose of the present work differs from the aforementioned ones in that we want to identify regions or pixels which are unlikely to be present in the original image, rather than to assess the entire image.

6. Numerical results. In this section we demonstrate the performance of the proposed Bayesian framework, by applying it to a PET image reconstruction problem with synthetic data. In particular the ground truth image (Fig. 1, left) is chosen from the Harvard whole brain atlas [1]. We let $\Omega = (0,1)^2$ and set the image size to be 128×128. In the Radon transform we use 60 projections equilaterally sampled from 0 to π. The test data, shown in the right figure of Fig. 1, is randomly simulated by plugging the true image into the Radon transform and the Poisson distribution (2.4) where the noise level K is taken to be 1. In the Bayesian inference, we use the hybrid prior distribution where the Gaussian part is taken to be zero mean and covariance:

$$(6.1) \qquad K(\mathbf{x}, \mathbf{x}') = \gamma \exp\left[-\frac{\|\mathbf{x}-\mathbf{x}'\|_1}{d},\right]$$

with $d = 1/128$ and $\gamma = 5$. The regularization parameter λ is chosen to be 1 (it is determined by using the proposed realized discrepancy method and details will be discussed later).
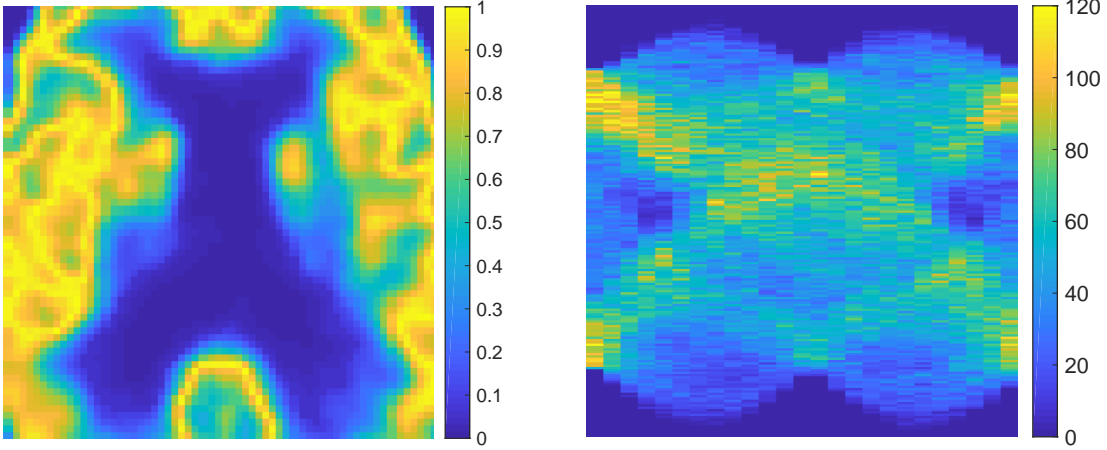
Figure 1: Left: the true image. Right: the noisy data.

6.1. Convergence with respect to discretization dimensionality.

In the numerical implementation we represent the unknown z using the truncated KL expansion with N KL-modes. First we shall demonstrate that the posterior distributions converges with respect to the discretization dimensionality N. To do so, we perform the proposed PD-pCN MCMC simulation and compute the posterior means with six different values of N: $N_i = i \times 10^3$ for $i = 1...6$. We note here that, in all the MCMC simulations performed in this section, we fix the number of samples to be 5×10^5 with additional 0.5×10^5 samples used in the burn-in step, and also, in all the simulations the stepsize β has been chosen in a way that the resulting acceptance probability is around 25%. We then compute the L_2 norm of the difference between the posterior mean with $N = N_i$ and that with $N = N_{i+1}$ for each $i = 1...6$:

$$(6.2) \qquad \mathrm{Diff} = \int_\Omega (\hat{u}_{N_i}(\mathbf{x}) - \hat{u}_{N_{i+1}}(\mathbf{x}))^2 d\mathbf{x}$$

where $\hat{u}_{N_i}$ is the posterior mean of u computed with N_i KL modes. We plot the L_2 difference against the discretization dimensionality N in Fig. 2 (left). One can see from the figure that the difference decrease as N increase and the difference becomes approximately zero for $N = 5000$ and $N = 6000$, indicating the convergence of the posterior mean with respect to N. For each posterior mean $\hat{u}_{N_i}$, we also compute its peak signal-to-noise ratio (PSNR) [22], a commonly used metric of the quality of a reconstructed image. We show the PSNR results in Fig. 2 (right), and the figure shows that the PSNR increases as N increases from 1000 to 4000, and remains approximately constant from 4000 to 6000, suggesting that increasing the discretization dimensionality can improve the inference accuracy until the posterior converges, and so it is important to use sufficiently large discretization dimensionality in such problems. In the rest of the work, we fix $N = 6000$.

6.2. Sampling efficiency of the PD-pCN algorithm.

Next, we shall compare the performance of the proposed PD-pCN algorithm and the standard pCN. We draw 5×10^5 samples from the posterior distribution using both the standard pCN and the proposed PD-pCN al-
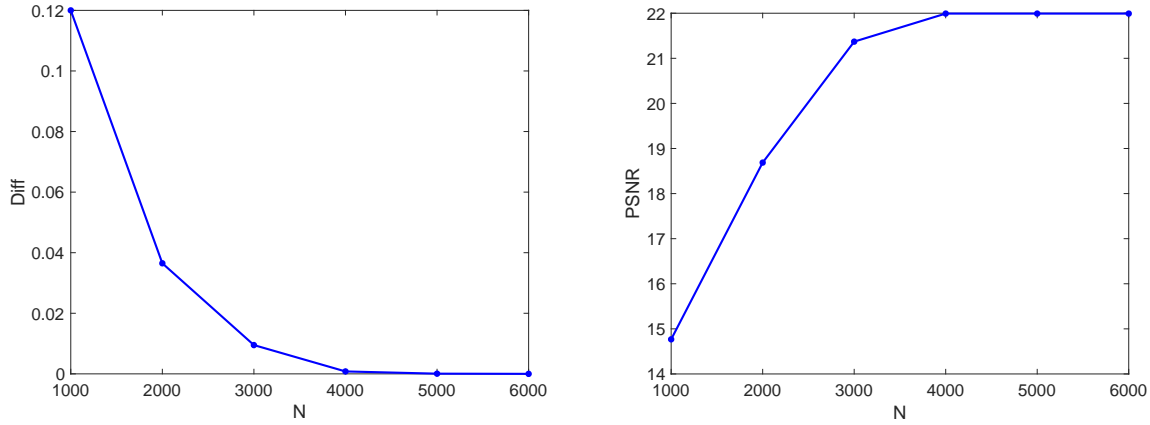
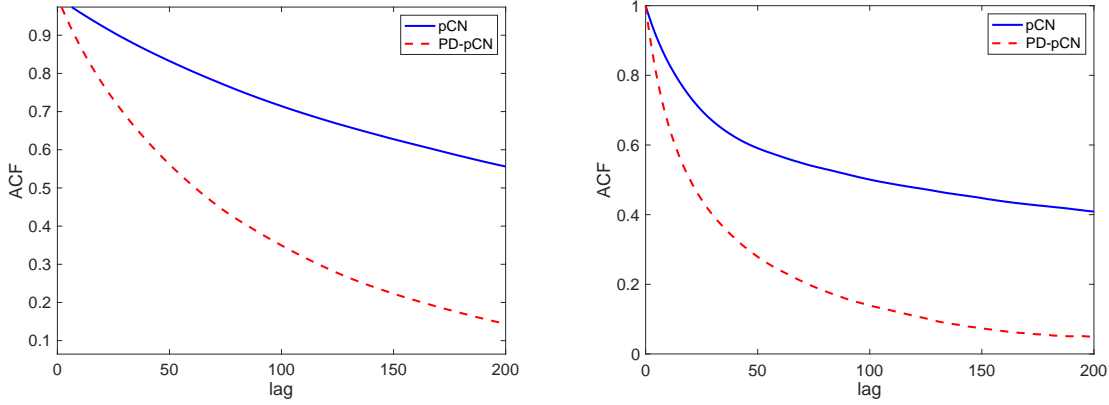Figure 2: Left: the convergence of the posterior mean. Right: the PSNR of the posterior mean as a function of N.

Figure 3: The ACF of the samples drawn by the pCN and the PD-pCN methods.

gorithms. We reinstate that in both algorithms we have chosen the step size so that the resulting acceptance probability is around 25%. We compute the auto-correlation function (ACF) of the samples generated by the two methods at two points: one in the center of the image, $\mathbf{x} = [0.5, 0.5]$ and the other near the boundary of it $\mathbf{x} = [0.05, 0.05]$. We plot the ACF at the two points in Fig. 3, and one can see from the figure that at both points the ACF of the propose PD-pCN method decays much faster than that of the standard pCN. To further compare the performance, we compute the effective sample size(ESS):

$$\text{ESS} = \frac{N}{1 + 2\tau},$$

where τ is the integrated auto-correlation time and N is total sample size. In Fig. 4, we compare the ESS at three chosen rows from left to right in the image, namely row 1, 64 and
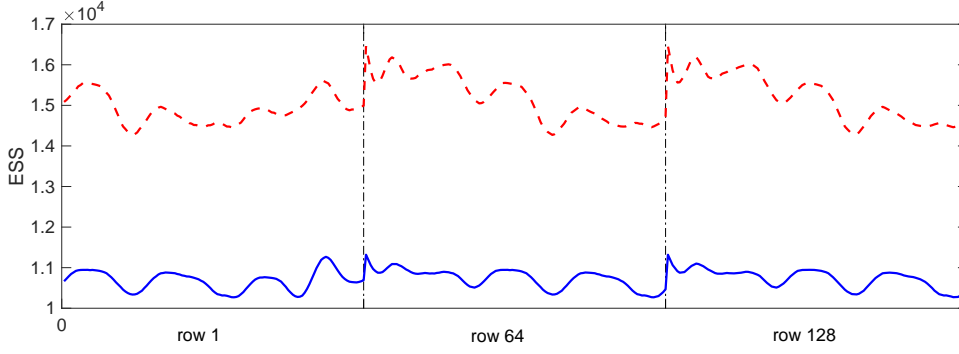
Figure 4: ESS of pCN and pCNL algorithms

λ	0	0.5	1	2	3	4
p_b	0.99	0.82	0.28	0.04	0.0034	0.0005
PSNR	18.21	21.90	22.0	20.66	19.76	19.27

Table 1: The posterior predictive p-value (p_b) and the PSNR of the resulting posterior mean for different values of λ.

128. Just as the ACF, the results show that the PD-pCN algorithm achieves much higher ESS than the standard pCN. We have also examined the ESS of other rows where the results are qualitatively similar to those three shown in Fig. 4, and so we omit those results.

6.3. Determining parameter λ. As is discussed at the beginning of the section, the prior parameter λ is determined with the realized discrepancy method discussed in Section 4. We here provide some details on the issue. Specifically we test five different values of λ: $\lambda = 0, 0.5, 1, 2, 3$, and using the method discussed in Section 4 we compute the corresponding posterior predictive p-value for each value of λ, shown in Table 1. We can see from the table that, as λ increases, the resulting p-value decays. These results agree well with our expectation that as λ becomes larger, the prior distribution becomes stronger, and as a result the p-value which assesses the fitness of the posterior to the data becomes smaller. We also compute the PSNR of the posterior distribution computed with all these λ values, and the results are also given in the table. We can see here that, for both very large and very small p-values, the associated posterior means are of rather poor quality in terms of PSNR. That is, when the p-value is too large, the posterior distribution overfits the data, and when it is too small, the posterior underfits the data; both cases lead to a poor performance of the inference, and so we must choose a proper p-value that represents a good balance of the prior and the data. Based on our numerical tests, we suggest that for the present problem it is suitable to use λ where the associated p-value is in [0.1, 0.5], and so here we choose $\lambda = 1$ where the resulting p-value is 0.28.

6.4. The inference results. To illustrate the inference results, we compute the posterior mean of the TG prior, which is regarded as a point estimate of the image. As is mentioned

(a) The posterior mean of the TG prior

(b) The interval width of 95% HPDI of the TG prior.

(c) The posterior mean of the Gaussian prior

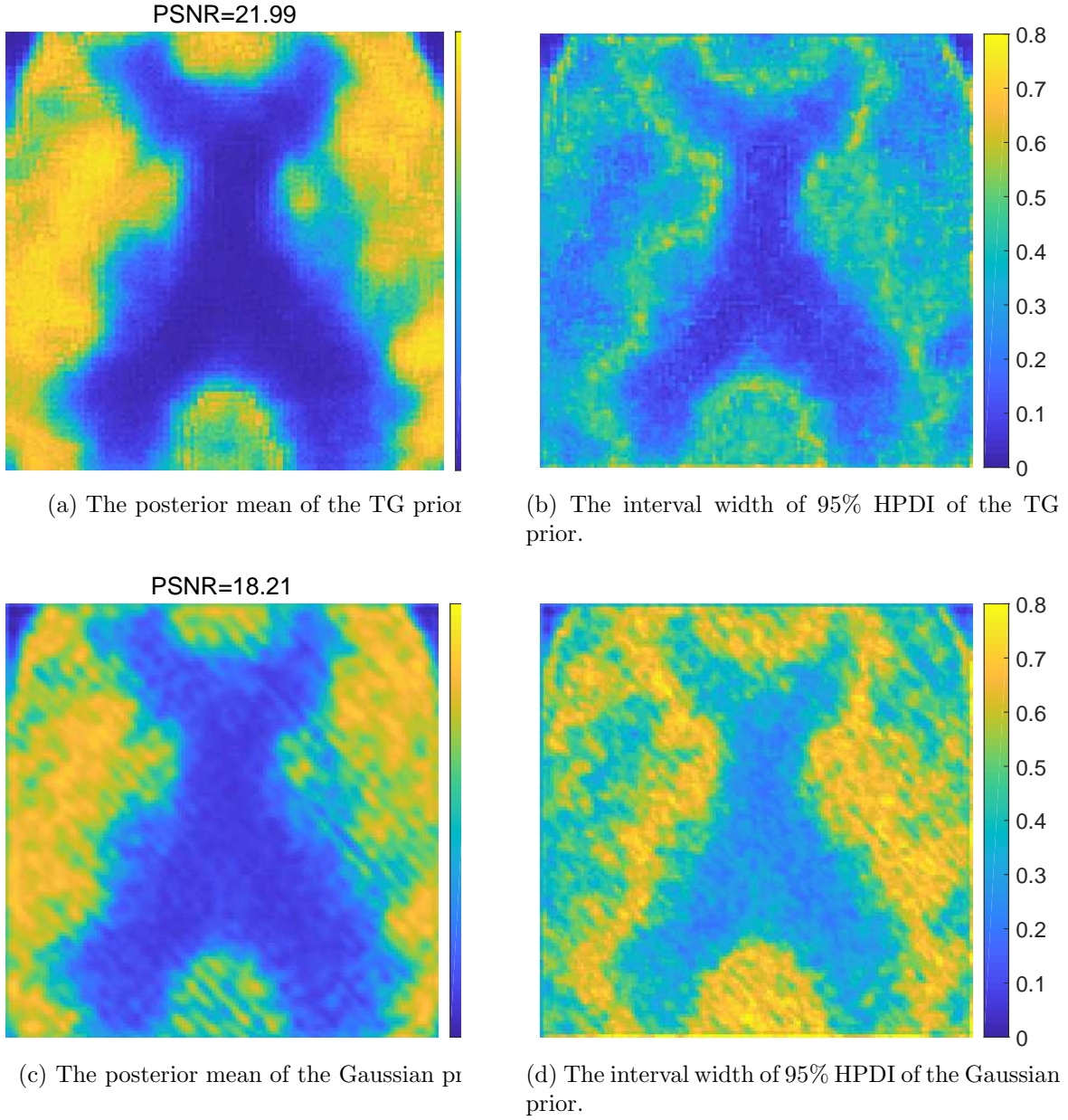(d) The interval width of 95% HPDI of the Gaussian prior.

Figure 5: A comparison of the posterior statistics of the TG and the Gaussian priors.

earlier, a main advantage of the Bayesian method is its ability to quantify the uncertainty in the reconstruction and to this end, we use the width of the (point-wise) 95% HPDI as a metric of the posterior uncertainty (intuitively speaking the wider the HPDI is, the more uncertainty there is). We plot these posterior results in Fig. 5 (Fig. 5a and Fig. 5b). As a comparison, we also compute the posterior mean as well as the 95% HPDI width, for the

Gaussian prior corresponding to setting $\lambda = 0$ in the TG prior, and the results are also shown in Figs. 5 (Fig. 5c and Fig. 5d). The figures show that the posterior mean obtained with the TG prior is clearly of better quality than that of the Gaussian prior, suggesting that including the edge-preserving TV term significantly improves the performance of the prior. It is worth noting here that, the Gaussian prior used here is not optimized for the best performance, and the performance can be potentially improved by using some carefully designed Gaussian priors, for example, [9].

6.5. The impact of Poisson likelihood function. In practice, the observation noise is often assumed to be Gaussian, for that it usually leads to great convenience for computation. In what follows we shall demonstrate that, if the data is generated truly from a Poisson distribution, and if the noise is strong, the use of a Gaussian noise model may yield very poor reconstruction results. Recall that the data $\mathbf{y}$ used in this example is drawn from the Poisson model, and here we perform the Bayesian inference using a Gaussian likelihood:

$$(6.3) \qquad \pi(\mathbf{y}|\theta) = \mathcal{N}(\boldsymbol{\theta}, \Theta),$$

where Θ is a $d \times d$ diagonal matrix with each diagonal component to be θ_i for $i = 1...d$. By design the Gaussian likelihood (6.3) has the same mean and covariance as the original Poisson likelihood. We compute the posterior distribution using the Gaussian likelihood and the same prior as before, and we plot the posterior mean in Fig. 6. It can be seen clearly from the figure that the posterior mean computed with the Gaussian likelihood is of much poor quality than that of the Poisson likelihood (shown in Fig. 5a). The results can be also compared quantitatively: the PSNR of the posterior mean of Gaussian likelihood is calculated to be 12.49, while that of the Poisson likelihood is 21.99. The significant difference between the results of the Gaussian likelihood and the Poisson likelihood suggests that the use of the Gaussian likelihood function may result in poor reconstruction performance in problems with Poisson data.

6.6. Identifying artifacts using HPDI. As is discussed in Section 5, an important application of the proposed Bayesian framework is that the resulting posterior distribution can be used to detect artifacts in a reconstructed image. We now demonstrate this application with three surrogate test images which are generated by making certain modification of the ground truth. Specifically the first image shown in Fig. 7a is generated by adding some random noise to the ground truth without any structural changes, the second one shown in Fig. 7c is generated by adding some artificial components to the ground truth, and the third one shown in Fig. 7e, is the result of removing some components from the ground truth. Thus both the last two test images have structural changes from the ground truth, and in both figures, the regions in which components are altered from the ground truth are highlighted with red boxes. We compute the credible level $(1-\alpha)$ for all three images, and show the results in Figs. 7b (for test image 1), 7d (for test image 2) and 7f (for test image 3). It can be seen here that, though the first test image is visibly perturbed by random noise, it does not have structural difference from the ground truth and so credible level result in Fig. 7b does not suggest any region has high likelihood to contain artifacts. On the other hand, in the other two test images, at the locations where the original image is altered (i.e., artifacts introduced), the resulting credible level $(1-\alpha)$ is significantly higher than other regions, suggesting that these locations may
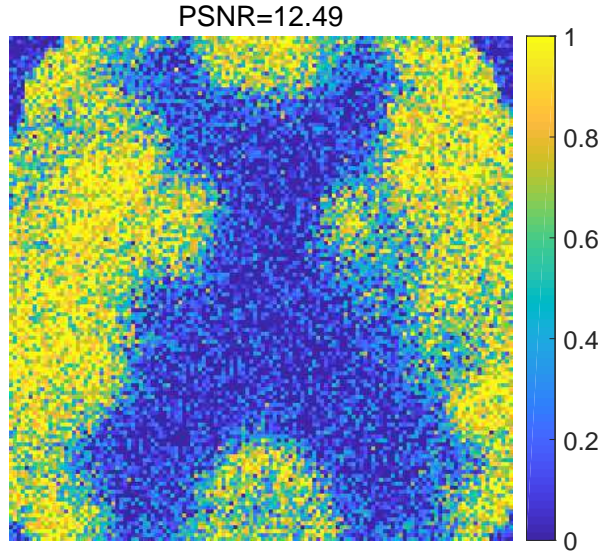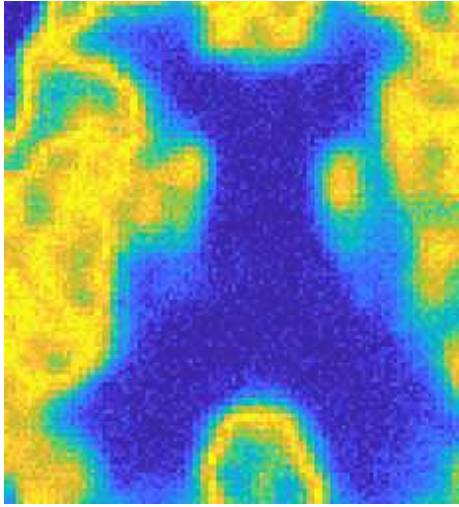
Figure 6: The posterior mean of the Gaussian likelihood.

contain artifacts. The results demonstrate that the proposed method can rather effectively detect the artifacts in a test image.

7. Conclusions. In this work, we have presented a complete treatment for performing Bayesian inference and uncertainty quantification for image reconstruction problems with Poisson data. In particular, we formulate the problem in an infinite dimensional setting and we prove that the resulting posterior distribution is well-posed in this setting. Second, to sample the unknown function/image, we provide a modified pCNL MCMC algorithm, the efficiency of which is independent of discretization dimensionality. Specifically the modified algorithm calculates the offset direction in the original pCNL algorithm by using a primal-dual method, to avoid computing the gradient of the TV term in our formulation. Third, we also give a method to determine the TV regularization parameter λ which is critical for the prior distribution. The method is based on the realized discrepancy method for assessing model fitness. Finally we provide an application of the uncertainty information obtained by the Bayesian framework, using the posterior distribution to identify possible artifacts in an image reconstructed. We believe the proposed Bayesian framework can be used to reconstruct images and evaluate the uncertainty or credibility of the images reconstructed in many practical imaging problems with Poisson data. In the future, we plan to further investigate the possibility of applying the methods developed in this work to those real-world problems, especially the PET image reconstruction.
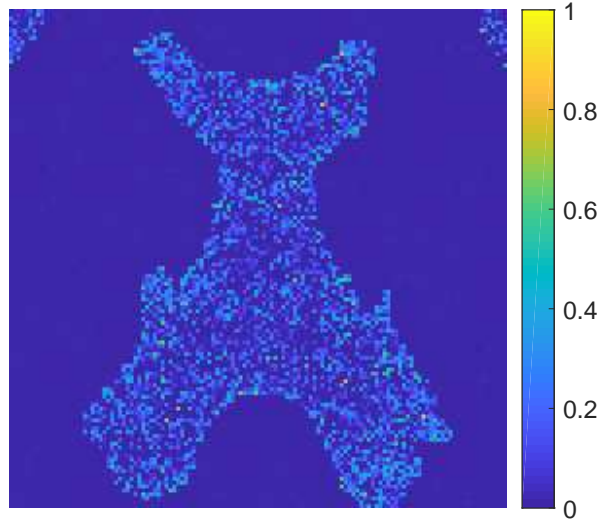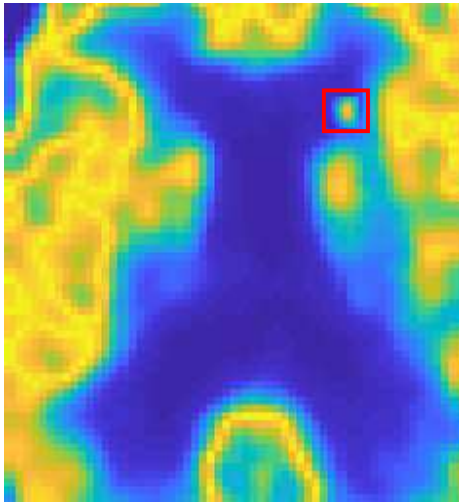
Appendix A. Proof of Proposition.   We provide a proof of Proposition 2.1 in the Appendix.

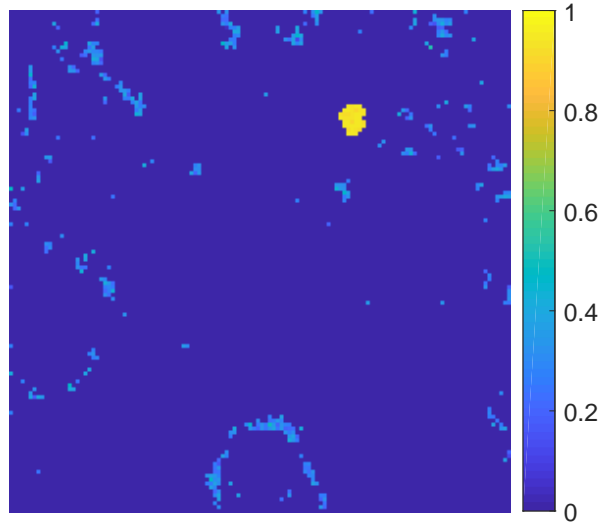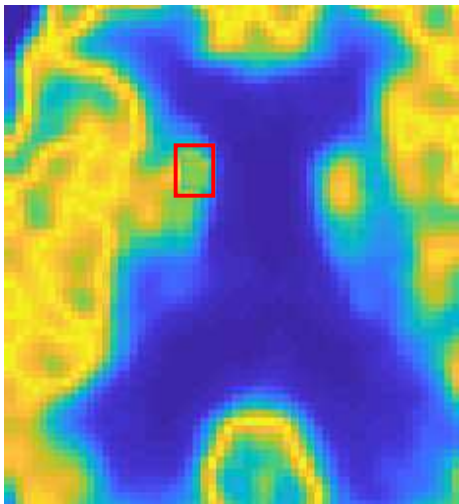*Proof.* (1) Since A is bounded, from Eq. (2.5), we obtain directly that

$$0 < \underline{\boldsymbol{\theta}} \leq \boldsymbol{\theta} \leq \overline{\boldsymbol{\theta}},$$

(a) Surrogate test image one

(b) The credible level $(1 - \alpha)$ computed for test image one.

(c) Surrogate test image two

(d) The credible level $(1 - \alpha)$ computed for test image two.
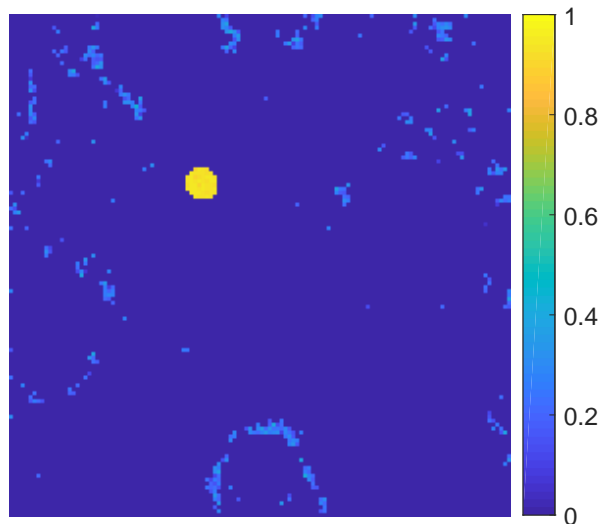

(e) Surrogate test image three

(f) The credible level $(1-\alpha)$ computed for test image three.

Figure 7: The credible level for three surrogate test images.

for two constant vectors $\underline{\boldsymbol{\theta}}$ and $\overline{\boldsymbol{\theta}}$. It follows directly that $\|\ln\boldsymbol{\theta}\|_2 \leq l_{\max}$ for a positive constant $l_{\max}$.

For every $r > 0$, we have $\|y\|_2 < r$. By the Cauchy-Schwarz inequality, we obtain the lower bound,

$$
\begin{aligned}
\Phi(z) = \langle\boldsymbol{\theta},\mathbf{1}\rangle - \langle y,\ln\boldsymbol{\theta}\rangle \\
\geq \langle\boldsymbol{\theta},\mathbf{1}\rangle - \|\ln\boldsymbol{\theta}\|_2\|y\|_2 \\
\geq \langle\underline{\boldsymbol{\theta}},\mathbf{1}\rangle - l_{\max}\|y\|_2 \\
\geq \langle\underline{\boldsymbol{\theta}},\mathbf{1}\rangle - l_{\max}r.
\end{aligned}
$$

For upper bound, once again we apply the Cauchy-Schwarz inequality to the functional Φ, obtaining

$$
\begin{aligned}
\Phi(z) = \langle\boldsymbol{\theta},\mathbf{1}\rangle - \langle y,\ln\boldsymbol{\theta}\rangle \\
\leq \langle\boldsymbol{\theta},\mathbf{1}\rangle + \|\ln\boldsymbol{\theta}\|_2\|y\|_2 \\
\leq \langle\overline{\boldsymbol{\theta}},\mathbf{1}\rangle + l_{\max}\|y\|_2 \\
\leq \langle\overline{\boldsymbol{\theta}},\mathbf{1}\rangle + l_{\max}r.
\end{aligned}
$$

(2) In this proof, we use M for positive constants. Let z and v be any two elements in X, and we have,

$$
\begin{aligned}
|\Phi(z) - \Phi(v)| = |\langle Af(z) - Af(v),\mathbf{1}\rangle - \langle y,\ln(Af(z)) - \ln(Af(v))\rangle| \\
\leq |\langle Af(z) - Af(v),\mathbf{1}\rangle| + |\langle y,\ln(Af(z)) - \ln(Af(v))\rangle| \\
\leq \|\mathbf{1}\|_2\|Af(z) - Af(v)\|_2 + \|y\|_2\|\ln(Af(z)) - \ln(Af(v))\|_2 \\
\leq \|\sqrt{d}\|\|Af(z) - Af(v)\|_2 + \|y\|_2 M\|Af(z) - Af(v)\|_2 \\
= (\sqrt{d} + \|y\|_2 M)\|Af(z) - Af(v)\|_2.
\end{aligned}
$$
(A.1)

Since the mapping A is a bounded linear transform from the X to Y, we have,

(A.2) $\quad \|Af(z) - Af(v)\|_2 \leq \|A\|\|f(z) - f(v)\|_{L_2}$

$$
= \|A\|\|\int_v^z e^{-t^2} dt\|_{L_2} \leq \|A\|\|\int_v^z dt\|_{L_2} = \|A\|\|z - v\|_{L_2},
$$

which completes the proof.

(3) For any y, $y' \in Y$, it is easy to show,

$$
|\Phi(z,y) - \Phi(z,y')| = |\langle y - y',\ln\boldsymbol{\theta}\rangle| \leq \|y - y'\|_2\|\ln\boldsymbol{\theta}\|_2 \leq l_{\max}\|y - y'\|_2. \quad \blacksquare
$$

REFERENCES

[1] Johnson K A and Becker J. www.med.harvard.edu/aanlib/. Accessed April 4, 2010.
[2] Simon R Arridge, Kazufumi Ito, Bangti Jin, and Chen Zhang. Variational gaussian approximation for poisson data. *Inverse Problems*, 34(2):025005, 2018.

[3] Dale L Bailey, Michael N Maisey, David W Townsend, and Peter E Valk. *Positron emission tomography*. Springer, 2005.

[4] Johnathan M Bardsley and John Goldes. Regularization parameter selection methods for ill-posed poisson maximum likelihood estimation. *Inverse Problems*, 25(9):095005, 2009.

[5] Johnathan M Bardsley and Aaron Luttman. A metropolis-hastings method for linear inverse problems with poisson likelihood and gaussian prior. *International Journal for Uncertainty Quantification*, 6(1), 2016.

[6] Mario Bertero, Patrizia Boccacci, Giorgio Talenti, Riccardo Zanella, and Luca Zanni. A discrepancy principle for poisson data. *Inverse problems*, 26(10):105004, 2010.

[7] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[8] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of markov chain monte carlo*. CRC press, 2011.

[9] Daniela Calvetti and Erkki Somersalo. A gaussian hypermodel to recover blocky objects. *Inverse problems*, 23(2):733, 2007.

[10] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004.

[11] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[12] Tony F Chan, Gene H Golub, and Pep Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM journal on scientific computing*, 20(6):1964–1977, 1999.

[13] Simon L Cotter, Gareth O Roberts, AM Stuart, David White, et al. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446, 2013.

[14] Alain Durmus, Eric Moulines, and Marcelo Pereyra. Efficient bayesian computation by proximal markov chain monte carlo: when langevin meets moreau. *SIAM Journal on Imaging Sciences*, 11(1):473–506, 2018.

[15] Jeffrey A Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE transactions on medical imaging*, 13(2):290–300, 1994.

[16] Jeffrey A Fessler. Medical image reconstruction: a brief overview of past milestones and future directions. *arXiv preprint arXiv:1707.05927*, 2017.

[17] A Gelman. Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, 6(4):733–760, 1996.

[18] Andrew Gelman, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

[19] Peter J Green. Bayesian reconstructions from emission tomography data using a modified em algorithm. *IEEE transactions on medical imaging*, 9(1):84–93, 1990.

[20] Tom Hebert and Richard Leahy. A generalized em algorithm for 3-d bayesian reconstruction from poisson data using gibbs priors. *IEEE transactions on medical imaging*, 8(2):194–202, 1989.

[21] Thorsten Hohage and Frank Werner. Inverse problems with poisson data: statistical regularization theory, applications and algorithms. *Inverse Problems*, 32(9):093001, 2016.

[22] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369. IEEE, 2010.

[23] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160. Springer, 2005.

[24] Jinglai Li. A note on the karhunen–loève expansions for infinite-dimensional bayesian inverse problems. *Statistics & Probability Letters*, 106:1–4, 2015.

[25] Erkan Ü Mumcuoglu, Richard M Leahy, and Simon R Cherry. Bayesian reconstruction of pet images: methodology and performance analysis. *Physics in Medicine & Biology*, 41(9):1777, 1996.

[26] Frank Natterer. *The mathematics of computerized tomography*. SIAM, 2001.

[27] John M Ollinger and Jeffrey A Fessler. Positron-emission tomography. *IEEE Signal Processing Magazine*, 14(1):43–55, 1997.

[28] Carlos A de B Pereira, Julio Michael Stern, Sergio Wechsler, et al. Can a significance test be genuinely bayesian? *Bayesian Analysis*, 3(1):79–100, 2008.

[29] Marcelo Pereyra. Proximal markov chain monte carlo algorithms. *Statistics and Computing*, 26(4):745–760, 2016.

[30] Marcelo Pereyra. Maximum-a-posteriori estimation with bayesian confidence regions. *SIAM Journal on Imaging Sciences*, 10(1):285–302, 2017.

[31] Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[32] Lawrence A Shepp and Yehuda Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE transactions on medical imaging*, 1(2):113–122, 1982.

[33] A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[34] Måns Thulin. Decision-theoretic justifications for bayesian hypothesis testing using credible sets. *Journal of Statistical Planning and Inference*, 146:133–138, 2014.

[35] Yehuda Vardi, LA Shepp, and Linda Kaufman. A statistical model for positron emission tomography. *Journal of the American statistical Association*, 80(389):8–20, 1985.

[36] Zhewei Yao, Zixi Hu, and Jinglai Li. A tv-gaussian prior for infinite-dimensional bayesian inverse problems and its numerical implementations. *Inverse Problems*, 32(7):075006, 2016.