Bayesian inference and uncertainty quantification for medical image reconstruction with Poisson data

Zhou, Qingping; Yu, Tengchao; Zhang, Xiaoqun; Li, Jinglai

DOI:
10.1137/19M1248352

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Zhou, Q, Yu, T, Zhang, X & Li, J 2020, 'Bayesian inference and uncertainty quantification for medical image reconstruction with Poisson data', *SIAM Journal on Imaging Sciences*, vol. 13, no. 1, pp. 29–52. https://doi.org/10.1137/19M1248352

SIAM J. IMAGING SCIENCES
Vol. 13, No. 1, pp. 29–52

Bayesian Inference and Uncertainty Quantification for Medical Image Reconstruction with Poisson Data*

Qingping Zhou†, Tengchao Yu†, 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 Markov chain Monte Carlo 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 combining the model discrepancy method and maximum likelihood estimation 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 example to demonstrate the effectiveness of the proposed method.

Key words. Poisson distribution, Bayesian inference, image reconstruction, uncertainty quantification, Markov chain Monte Carlo, positron emission tomography

AMS subject classifications. 68Q25, 68R10, 68U05

DOI. 10.1137/19M1248352

1. Introduction.

Image reconstruction involves constructing interpretable images of objects of interest from the data recorded by an imaging device [18]. 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 [25]. A very important example of such problems is positron emission tomography (PET) [31, 4], a nuclear medicine imaging technique that is

*Received by the editors March 6, 2019; accepted for publication (in revised form) September 16, 2019; published electronically January 7, 2020.
https://doi.org/10.1137/19M1248352
Funding: The work of the authors was supported by the National Natural Science Foundation of China under grants 11771288 and 11771289.

†School of Mathematical Sciences, Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China (zhou2015@sjtu.edu.cn, tengchaoyu@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, 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).

widely used in early detection and treatment follow-up of many diseases, including cancer. In PET, the detection of signal is essentially a photon counting process, and as a result the data is well modeled by a Poisson distribution [4, 25]. The problem has attracted considerable research interest, and a number of methods have been developed to recover the image, e.g., [39, 42, 17], 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 [27], 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., [24, 29, 21]. We note, however, that most of the works in the early years focused on computing a point estimate, which is usually the maximum a posteriori 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, because 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 [6], and a variational Gaussian approximation of the posterior is proposed in [3].

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 discretization variant or dimension dependent. To address the issue, Stuart [40] proposes an infinite dimensional framework, formulating the Bayesian inference problem in function spaces. Under the infinite dimensional framework, the inference results will converge with respect to discretization dimensionality, which is an important property for the numerical implementation. For example, it allows one to use multigrid strategy, e.g. [44, 32], to accelerate the sampling of the posterior. 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 total variation Gaussian (TG) hybrid prior distribution proposed in [43] 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 [37, 9], degenerate with respect to resolution refinement. In [14] the authors introduce a MCMC algorithm termed 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 [14] which accelerates the sampling procedure by incorporating the local gradient information of the likelihood function. In our problem, the pCN–Langevin (pCNL) algorithm cannot be used directly because the prior used here has the total variation (TV) term which can be nondifferentiable. 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 [34, 15] 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 [34, 15] 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 [20]. 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 by combining the realized discrepancy model fit assessment approach developed in [19] and the stochastic proximal gradient method developed in [16]. 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 follows. 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 $y \in Y \subset R^d$ and $\mathbf{y}$ related to u via the likelihood function $\pi(\mathbf{y}|u)$, i.e., the distribution of $\mathbf{y}$ is 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_{pr} 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:

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

which can be interpreted as Bayes' rule in the infinite dimensional setting. The posterior distribution μ^y thus can be computed from (2.1) with, for example, an 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 $A : X \to Y$,

$$\boldsymbol{\theta} = \mathrm{A}u. \tag{2.2a}$$

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

$$\theta_i = (\mathrm{A}u)_i = K \int_{L_i} u(\mathbf{x})|d\mathbf{x}| \tag{2.2b}$$

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

$$\pi_{\text{P}}(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \frac{(\theta_i)^{y_i} \exp(-\theta_i)}{y_i!}. \tag{2.3}$$

In the PET problem, there is an additional restriction: the unknown function u must be positive. The reason is twofold: 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 undefined. 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

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

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

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

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

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

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

$$a(b-1) \le u(\mathbf{x}) \le a(b+1). \tag{2.4}$$

Moreover, as the behavior of the error function is well understood (for example, its derivative is simply the Gaussian distribution), which allows us determine the parameters conveniently. That said, it is worth noting here that the methods presented here do not rely on this specific reparametrization formulation. 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 (2.3). Following [40], we can write the likelihood function $\pi(\mathbf{y}|z)$ in the form of

$$\pi(\mathbf{y}|z) \propto \exp(-\Phi(z, \mathbf{y})), \tag{2.5a}$$

where

$$\Phi(z; \mathbf{y}) = \sum_{i=1}^{d} (\mathrm{A}f(z))_i - y_i \ln(\mathrm{A}f(z))_i. \tag{2.5b}$$

For simplicity we can rewrite (2.5b) as

$$\Phi(z; \mathbf{y}) = \langle \mathrm{A}f(z), \mathbf{1}\rangle - \langle \mathbf{y}, \ln(\mathrm{A}f(z))\rangle = \langle \boldsymbol{\theta}, \mathbf{1}\rangle - \langle \mathbf{y}, \ln \boldsymbol{\theta}\rangle, \tag{2.6}$$

where $\mathbf{1}$ is a d-dimensional vector whose components are all one, $\boldsymbol{\theta} = \mathrm{A}f(z)$ 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 this does not cause 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., $\mu_{\text{pr}} = N(\xi, C_0)$. 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 cannot model functions with sharp jumps well.

To address the issue, here we use the TG prior proposed in [43]:

$$\frac{d\mu_{\rm pr}}{d\mu_0}(z) \propto \exp(-R(z)), \quad R(z) = \lambda \|z\|_{\rm TV}, \tag{2.7}$$

where $\mu_0 = N(\xi, C_0)$ is the Gaussian reference prior defined on X with mean ξ and covariance C_0 and $\|\cdot\|_{\rm TV}$ is the TV seminorm,

$$\|z\|_{\rm TV} = \int_\Omega \|\nabla u\|_2 dx, \tag{2.8}$$

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

$$\frac{d\mu^y}{d\mu_0}(z) \propto \exp(-\Phi(z) - R(z)), \tag{2.9}$$

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. We note that it is natural to directly apply the TV seminorm to the original image u; if we do so, however, Proposition 2.1 may no longer hold. For this technical reason we here choose to impose the TV seminorm on the new variable z. Nonetheless, it can be shown that

$$\|z\|_{\rm TV} \geq \frac{a}{c}\|u\|_{\rm TV}.$$

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 (2.6) satisfies certain important conditions, as stated by Proposition 2.1.

Proposition 2.1. *The functional Φ given in (2.6) 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 (2.7) and the log-likelihood function (2.6) yield a well-behaved posterior measure given by (2.9) in the infinite dimensional setting, as is summarized in the following theorem.

Theorem 2.2. *For $\Phi(z)$ given by (2.6) and prior measure $\mu_{\rm pr}$ given by (2.7), we have the following results:*

1. *μ^y given by (2.9) is a well-defined probability measure on X.*
2. *μ^y given by (2.9) 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' then there exists $C = C(r)$ such that, for all y, y' with $\max\{\|y\|_2, \|y'\|_2\} < r$,*

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

3. *Let*

$$\frac{d\mu^y_{N_1,N_2}}{d\mu_0} = \exp(-\Phi_{N_1}(z) - R_{N_2}(z)), \tag{2.10}$$

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} satisfies Assumption A.2(i) and (ii) in [43] with constants uniform in N_2. Assume also that for any $\epsilon > 0$, there exist two positive sequences $\{a_{N_1}(\epsilon)\}$ and $\{b_{N_2}(\epsilon)\}$ converging to zero such that $\mu_0(X_\epsilon) \ge 1 - \epsilon$ for any $N_1, N_2 \in \mathbb{N}$, where

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

Then we have

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

Theorem 2.2 is a direct consequence of Proposition 2.1, and the proof of the theorem can be found in [43] and is omitted here.

Finally, we note that, in the numerical implementations, we use the truncated Karhunen–Loève (KL) expansion [28] to represent the unknown z. Namely, we write z as

$$z(\mathbf{x}) = \sum_{i=1}^{N} z_i \sqrt{\eta_i} e_i(\mathbf{x}), \tag{2.11}$$

where $\{\eta_i,\, e_i(\mathbf{x})\}$ are the eigenvalue-eigenfunction pair of the covariance operator C_0, and $(z_1, \ldots, z_N)$ 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 pCN MCMC algorithm. In most practical image reconstruction problems, the posterior distribution cannot be analytically calculated. Instead, one usually represent the posterior by samples drawn from it using MCMC algorithms. It is demonstrated in [14] 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 [14]. 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 particularly useful for

sampling the posterior distribution defined in function spaces. We start with a brief introduction of the pCN algorithm following the presentation of [14]. We denote $\Phi(z) + R(z)$ of (2.9) as $\Psi(z)$. Simply speaking the algorithms are derived by applying the Crank–Nicolson discretization to a stochastic partial differential equation whose invariant distribution is the posterior. We here omit the derivation details while referring interested readers to [14], and jump directly to the pCN proposal:

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

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 step size of the algorithm. The proposed sample v is then accepted or rejected according to the acceptance probability:

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

which is independent of discretization dimensionality up to numerical errors.

The pCN proposal in (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 pCNL proposal:

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

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

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

then the acceptance probability is given by

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

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 cannot 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 an alternative direction $g(z)$, yielding the proposal

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

where $\delta \in [0, 2]$ and $w \sim N(\xi, \mathbf{C}_0)$. Regarding the proposal given by (3.6), we have the following theorem.

Theorem 3.1. *Assume that Ψ satisfies Assumption* 6.1 *in* [14], *and $g(z)$ is in the Cameron–Martin space associated with the Gaussian measure μ_0. Let $q(z, dv)$ be the conditional distribution defined by* (3.6), *and define*

$$\eta(dz, dv) = q(z, dv)\mu^y(dz), \quad \eta^\perp(dz, dv) = q(v, dz)\mu(dv)$$

on $X \times X$. We have that $\eta^\perp$ is equivalent to η and

$$\frac{d\eta^\perp}{d\eta}(z, v) = \exp(\rho(z, v) - \rho(v, z)), \tag{3.7}$$

where

$$\rho(z, v) = \Psi(z) + \frac{1}{2}\langle v - z, g\rangle + \frac{\delta}{4}\langle z + v, g\rangle + \frac{\delta}{4}\|\mathbf{C}_0^{1/2} g\|^2. \tag{3.8}$$

The proof of the theorem is provided in Appendix B. It should be clear that Theorem 3.1 implies that the MCMC algorithm with proposal (3.6) yields a well-defined acceptance probability in the function space, and as a result the chain satisfies the detailed balance condition in the function space and thus is ergodic. Another very important theoretical issue here is to estimate the spectral gaps and prove the geometric ergodicity of the algorithms in the infinite dimensional setting. We note that there are some results on the spectral gaps of the standard pCN [22] and the generalized pCN [38]. It is an interesting problem to analyze if similar results can also be obtained for the present algorithm.

Now we need to find a good direction $g(z)$. In [34] 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 [12, 13, 11], and we hereby give a brief description of the primal-dual method applied to our problem. Suppose that we want to solve

$$\min_{z \in X} \Psi(z) = \Phi(z) + \lambda \|z\|_{\mathrm{TV}}. \tag{3.9}$$

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 $\boldsymbol{\phi} \in L_2^2(\Omega)$), then we rewrite the optimization problem (3.9) as

$$\begin{aligned} \min_{z \in X, \phi \in L_2^2(\Omega)} \quad & \Psi(z, \phi) = \Phi(z) + \lambda \|\phi\|_{2,1} \\ \text{such that} \quad & \nabla z = \phi, \end{aligned} \tag{3.10}$$

where $\|\phi\|_{2,1} = (\|\phi_1(\mathbf{x})\|^2_{L_2(\Omega)} + \|\phi_2(\mathbf{x})\|^2_{L_2(\Omega)})^{1/2}$. The augmented Lagrangian for (3.10) is

$$\max_{\eta \in L_2^q(\Omega)} \min_{z \in X, \phi \in L_2^q(\Omega)} \quad L_\rho(z, \phi, \eta) = \Phi(z) + \lambda \|\phi\|_{2,1} + \langle \eta, \nabla z - \phi\rangle + \frac{\rho}{2}\|\nabla z - \phi\|_2^2, \tag{3.11}$$

where $\eta \in L_2^q(\Omega)$ is the dual variable or Lagrange multiplier and $\rho > 0$ is a constant called the penalty parameter. The resulting dual problem is then solved with the alternating direction method of multipliers (ADMM) [8]:

$$z^{k+1} = \arg\min_{z \in X} \quad L_\rho(z, \phi^k, \eta^k), \tag{3.12a}$$

$$\phi^{k+1} = \arg\min_{\phi \in L_2^q(\Omega)} \quad L_\rho(z^{k+1}, \phi, \eta^k), \tag{3.12b}$$

$$\eta^{k+1} = \eta^k + \rho(\phi^{k+1} - \nabla z^{k+1}). \tag{3.12c}$$

Algorithm 3.1 The primal-dual pCN (PD-pCN) algorithm.

Solve (3.11) and denote the solution as $(z^\star, \phi^\star, \eta^\star)$
Let $z^0 = z^\star$
for $k = 0, 1, 2, \cdots$ **do**
    Propose v using (3.6) and (3.13)
    Draw $\theta \sim U[0, 1]$;
    Compute $a(z, v)$ with (3.5) and (3.8);
    if $\theta \leq a$ **then**
        $z^{k+1} = v$;
    else
        $z^{k+1} = z^k$;
    end if
end for

The algorithm consists of a z-minimization step (3.12a), a ϕ-minimization step (3.12b), and a dual ascent step (3.12c).

Our primal-dual pCN (PD-pCN) algorithm is designed as follows. First we solve (3.11) with the ADMM algorithm 3.1 obtaining the solution $(z^\star, \phi^\star, \eta^\star)$, and then we define

$$g(z) = \mathcal{T}_K \mathcal{D} L_\rho(z, \phi^\star, \eta^\star), \tag{3.13}$$

where operator $\mathcal{T}_K$ is the projection of its input function onto the space spanned by the KL models $\{e_1, \ldots c_K\}$ for a prescribed positive integer K. K should be no greater than the discretization dimensionality N. It should be clear that the function $g(z)$ computed with (3.13) is in the the Cameron–Martin space of μ_0. The complete algorithm is given in Algorithm 3.1. We note here that the proposed MCMC algorithm involves an optimization problem at the beginning, and the computational cost for solving this optimization is usually an order of magnitude lower that of the MCMC iterations. A main limitation of this algorithm is that, when some hyperparameters change, the optimization problem needs to be solved again, which makes it incompatible with the Metropolis within Gibbs [2] type of methods. It is also worth noting here that the main purpose of the proposed algorithm is to improve the sampling efficiency of the standard pCN algorithm while maintaining its dimension independence property. To this end a very interesting problem here is to incorporate the pCN framework with the aforementioned proximity based algorithm and compare the performance with the primal-dual based one.

4. Determining the hyperparameters. 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, the regularization parametter can be determined by the empirical Bayes (EB) approach [20]. Namely, the EB method seeks to maximize

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

where $Z(\lambda)$ is the normalization constant. A difficulty here is that $Z(\lambda)$ is usually not known in advance and needs to be evaluated with another Monte Carlo integration. To address this issue, a stochastic proximal gradient method was proposed in [36, 16]. The method can efficiently estimate the regularization parameter λ without the knowledge of Z. On the other hand, the method does require a suitable admissible set for λ. Here we provide a statistical approach to determine the admissible set of λ, which is derived from the realized discrepancy method for model assessment proposed in [19].

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 (2.3). In particular, for the discrepancy function given in (4.1), the p-value is simply

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

where $F_{\chi^2_d}(\cdot)$ 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 (2.7) is specified by the parameter λ, and as a result the posterior also depends on λ, and here we write the posterior as $\mu^y_\lambda(du)$ to emphasize its dependence on λ. In the Bayesian setting, one can compute the posterior predictive p-value:

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

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 cannot 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 suitable values of p_b are certainly problem

dependent, and in the present problem we suggest to choose λ so that the resulting value of p_b is *approximately* in the range of $0.1 \sim 0.7$. Based on this, we choose the admissible set for λ to be

$$\Lambda = \{\lambda \geq 0 | 0.1 \leq p_b(y, \lambda) \leq 0.7\}.$$

The optimal value of λ is then determined by using the method in [16] within Λ. 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 [5, 7] for further details. It is important to note that the discrepancy principle may lead to oversmoothing in certain problems [23], which, however, may not cause issues in the proposed method as it just uses the discrepancy method to identify the admissible set of the regularization parameter while the actual value of it is determined with EB. Finally we also note that, in addition to λ, the Gaussian distribution may also be subject to hyperparameters, and in principle these hyperparameters 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 hyperparameters 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 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 [35]

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

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 HPDI 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., $\alpha = 5\%$) [33]: that is, $\hat{u}(\mathbf{x})$ is regarded as 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 [41] that performing hypothesis test with

HPDI may cause certain theoretical issues, and so here we choose not to use the hypothesis test formulation.

It should be noted here that, in [35, 15], the authors utilize the highest posterior density 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 (Figure 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. In the numerical experiments, we consider two different noise levels: $K = 0.5$ corresponding to a higher noise level and $K = 1$ corresponding to a lower noise level. The test data, shown in Figure 1, are randomly simulated by plugging the true image into the Radon transform and the Poisson distribution (2.3) with the two aforementioned noise levels K. In the Bayesian inference, we use the hybrid prior distribution where the Gaussian part is taken to be zero mean and covariance

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

where d is taken to be 10^{-3} and γ is 2. The regularization parameter λ is determined by using the method presented in section 4, and details will be discussed in next section.

6.1. 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 for $K = 0.5$: $\lambda = 0, 1, 2, 3, 4, 5$, 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. Similarly we also test 5 values of λ for $K = 1$, and the results are shown in Table 2. 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

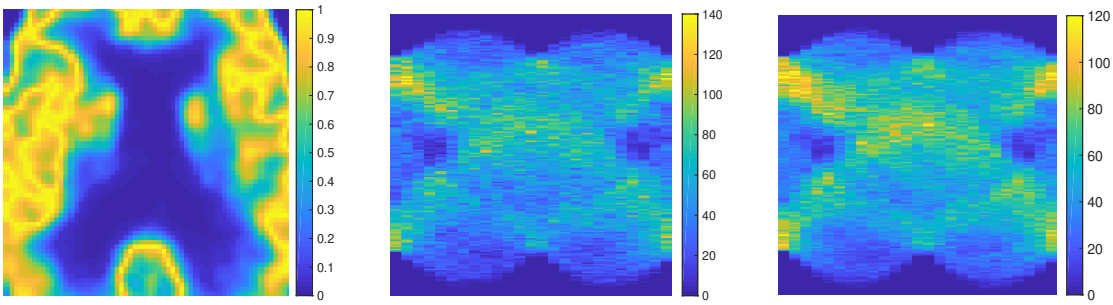


Figure 1. *Left: the true image. Middle: the simulated data for $K = 0.5$. Right: the simulated data for $K = 1$.*

Table 1

($K = 0.5$) The posterior predictive p-value (p_b) and the PSNR of the resulting posterior mean for different values of λ.

λ	0	1	2	3	4	5
p_b	0.99	0.74	0.32	0.08	0.0074	0.0004
PSNR	15.69	18.85	20.31	20.04	18.79	18.58

Table 2

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

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	21.99	20.66	19.76	19.27

compute the peak signal-to-noise ratio (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. Bases on the test results, for $K = 0.5$ we choose $\Lambda = [1, 3]$ and for $K = 1$ we choose $\Lambda = [0.5, 2]$. By optimizing λ within the identified intervals we obtain $\lambda = 2.4$ for $K = 0.5$ and $\lambda = 1.2$ for $K = 1$.

6.2. 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. We here use the case $K = 1$ as an example. 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 $i = 1 \dots 6$. 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 step size 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 $i = 1 \dots 6$:

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

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 Figure 2. One can see from the figure that the difference decreases as N increases 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 PSNR [26], a commonly used metric of the quality of a reconstructed image. We show the PSNR results in Figure 2 (right), and the figure shows that the PSNR increases as N increases from 1000 to 4000 and remains approximately

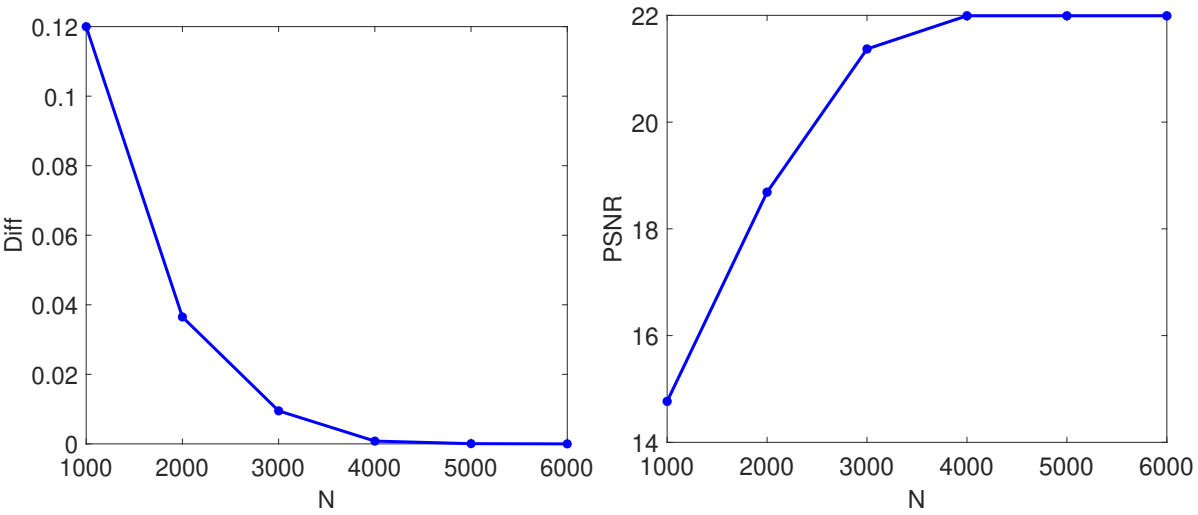


Figure 2. *Left: the convergence of the posterior mean. Right: the PSNR of the posterior mean as a function of N.*

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. Next, to further demonstrate that the proposed PD-pCN MCMC algorithm is independent of discretization dimensionality, we perform the MCMC simulation with different values of δ which is the parameter controlling the step size of the algorithm. In Figure 3 we plot the average acceptance probability as a

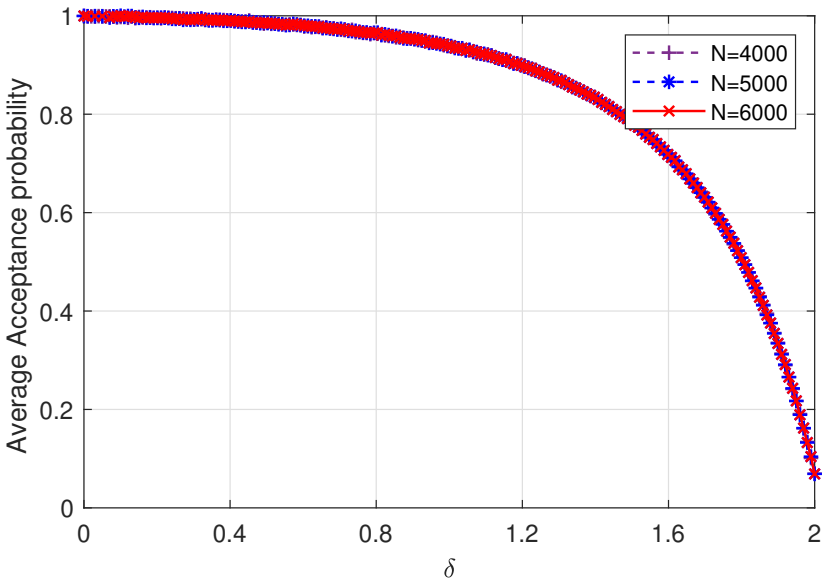


Figure 3. *The average acceptance probability plotted against δ for $N = 4000, 5000, 6000$.*

function of δ for three different values of N, and one can see that the acceptance probabilities under different discretization dimensionality agree well with either other, indicating that the acceptance probability of the algorithm is independent of the discretization dimensionality N. In the rest of the work, we fix $N = 6000$.

6.3. 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 algorithms. We reinstate that in both algorithms we have chosen the step size so that the resulting acceptance probability is around 25%. In particular, to achieve the sought acceptance probability, the values of the step size parameter β in pCN are taken to be 0.04 (for $K = 1$) and 0.09 (for $K = 0.5$); the values of the step size parameter δ in PD-pCN are 0.18 (for $K = 1$) and 0.23 (for $K = 0.5$). The total computational time is around 12 hours in a workstation with a 6-core 2.50 GHZ processor. We compute the autocorrelation function (ACF) of the samples generated by the two methods at all the grid points, and we show the ACF at the points with the fastest and the slowest convergence rates in Figure 4 ($K = 0.5$) and Figure 5 ($K = 1$). One can see from the figures that at both points the ACF of the proposed PD-pCN method decays much faster than that of the standard pCN. To further compare the performance, we compute the effective sample size (ESS), which is defined as

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

where τ is the integrated autocorrelation time and N is total sample size. In Figure 6, we compare the ESS at three chosen rows from left to right in the image, namely, rows 1, 64, and 128, for $K = 1$. 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 for $K = 0.5$, where the results are qualitatively similar to those three shown in Figure 6, and so we omit those results.

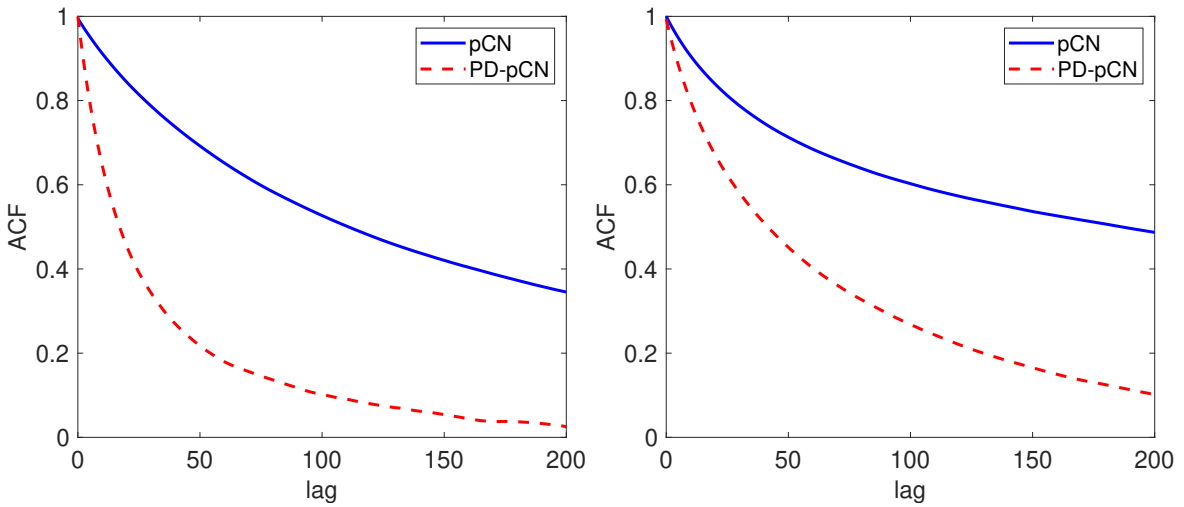


Figure 4. *($K = 0.5$) The ACF of the fastest (left) and the slowest (right) components of the samples drawn by the pCN and the PD-pCN methods.*

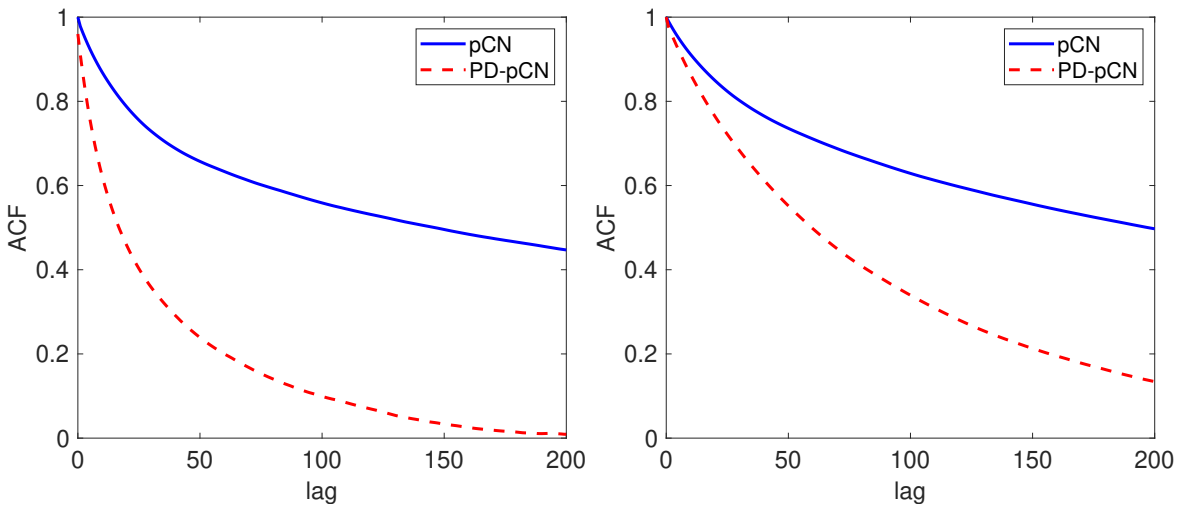


Figure 5. *($K = 1$) The ACF of the fastest (left) and the slowest (right) components of the samples drawn by the pCN and the PD-pCN methods.*

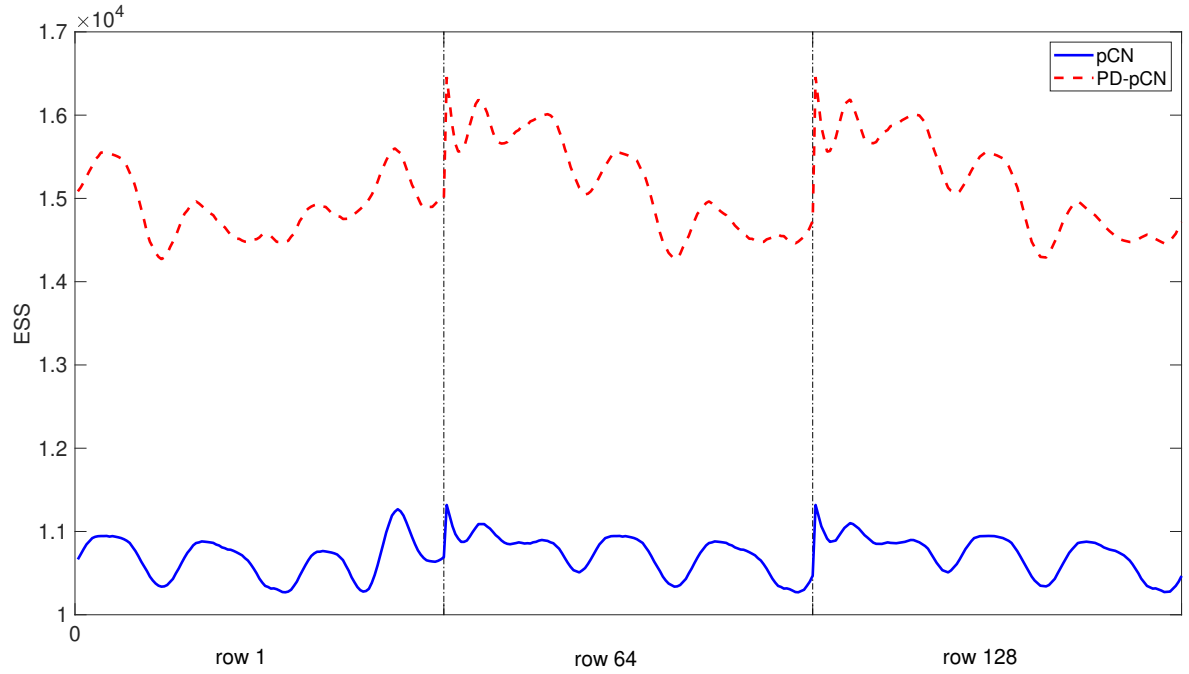
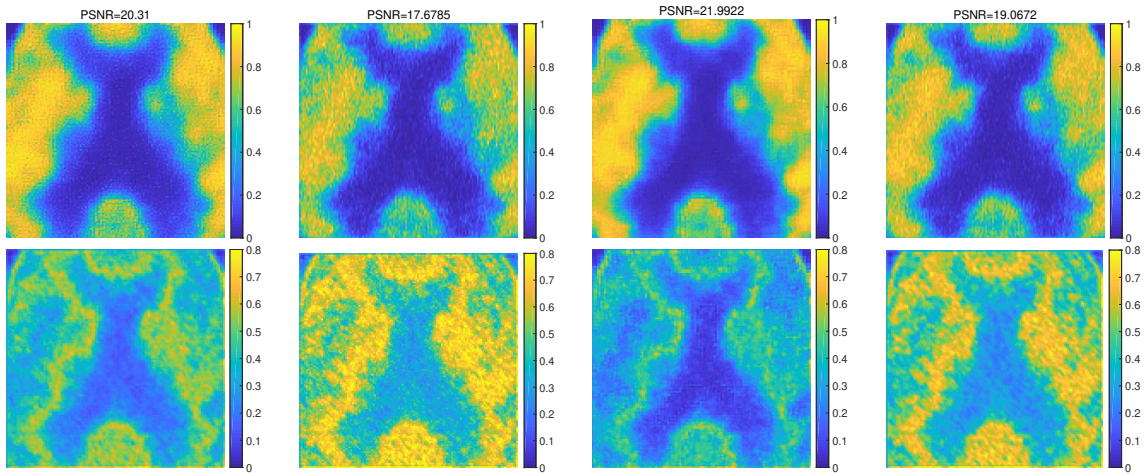


Figure 6. *ESS of the pCN and the PD-pCN algorithms.*

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 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 (pointwise) 95% HPDI as a metric of the posterior uncertainty (intuitively speaking the wider the HPDI is, the more

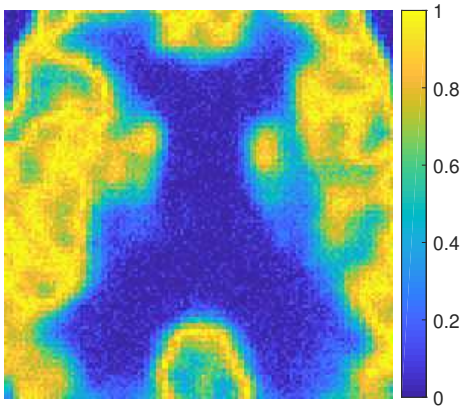


(a) $K = 0.5$, TG (b) $K = 0.5$, Gaussian (c) $K = 1$, TG (d) $K = 1$, Gaussian

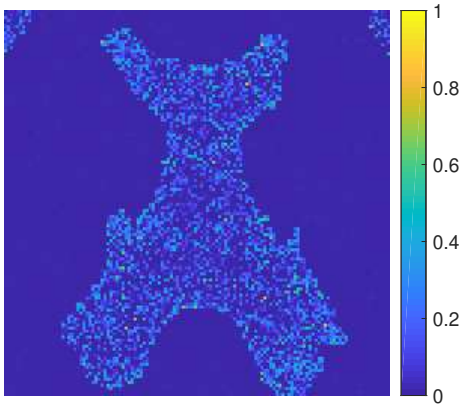
Figure 7. *The posterior results of the TG prior and the Gaussian prior. The top figures show the posterior mean and the bottom ones shows the corresponding interval width of 95% HPDI.*

uncertainty there is). We plot these posterior results in Figure 7: the posterior mean and the 95% HPDI for $K = 0.5$ and $K = 1$ are shown in Figure 7(a) and Figure 7(c), respectively. 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 Figure 7(b) and Figure 7(d). 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, [10].

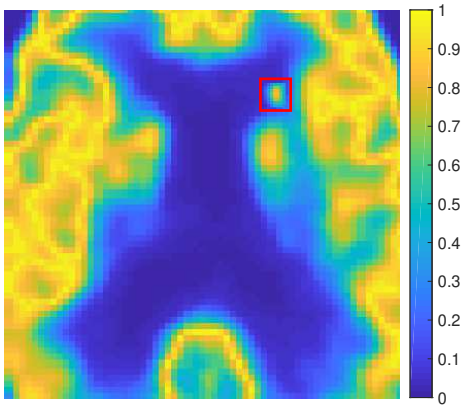
6.5. Identifying artifacts using HPDI. As is discussed in section 5, an important application of the proposed Bayesian framework is that the resulting posterior distribution can be used to detect artifacts in a reconstructed image. We now demonstrate this application with three surrogate test images which are generated by making certain modifications of the ground truth. Figure 8 summarizes the results for $K = 1$. Specifically the first image shown in Figure 8(a) is generated by adding some random noise to the ground truth without any structural changes, the second one shown in Figure 8(c) is generated by adding some artificial components to the ground truth, and the third one shown in Figure 8(e) 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 Figure 8(b) (for test image 1), 8(d) (for test image 2), and 8(f) (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



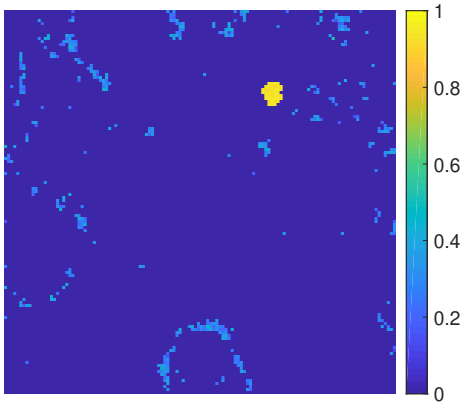
(a) Surrogate test image one.



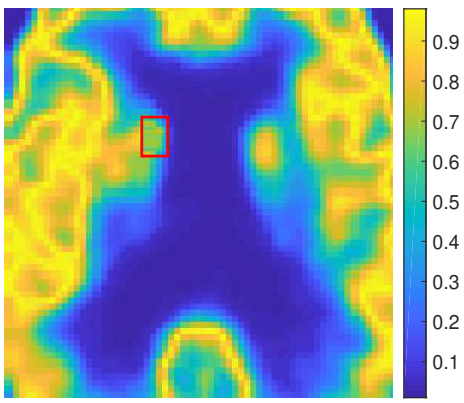
(b) The credible level $(1-\alpha)$ computed for test image one.



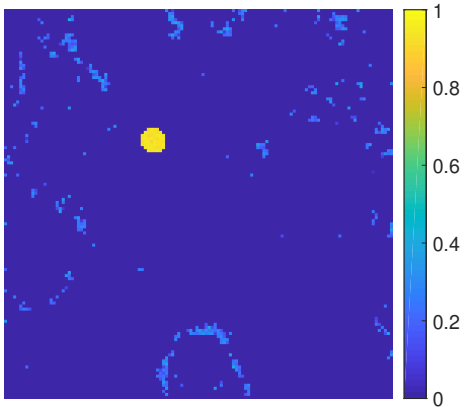
(c) Surrogate test image two.



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

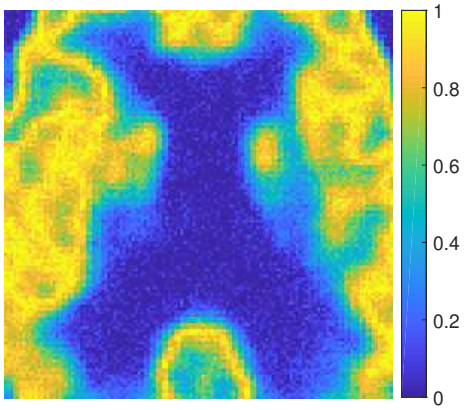


(e) Surrogate test image three.

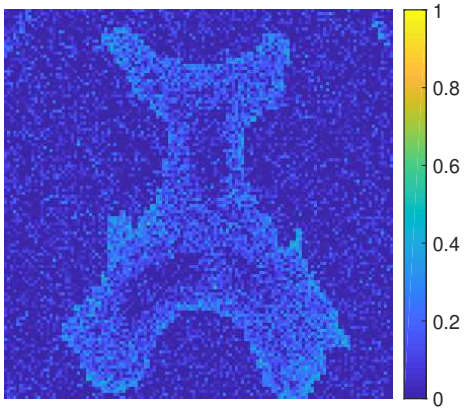


(f) The credible level $(1-\alpha)$ computed for test image three.

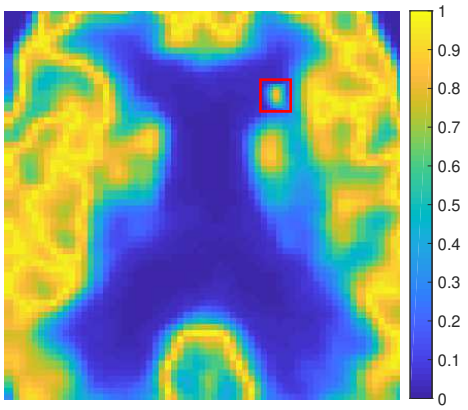
Figure 8. *$(K = 1)$ The credible level for three surrogate test images.*



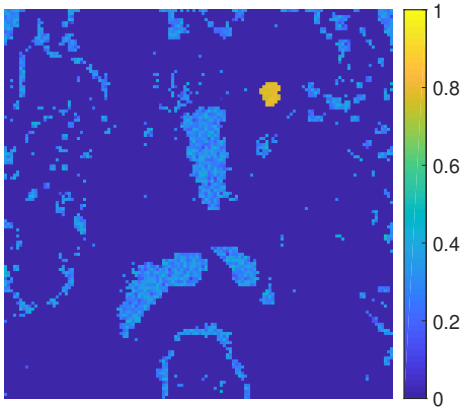
(a) Surrogate test image one.



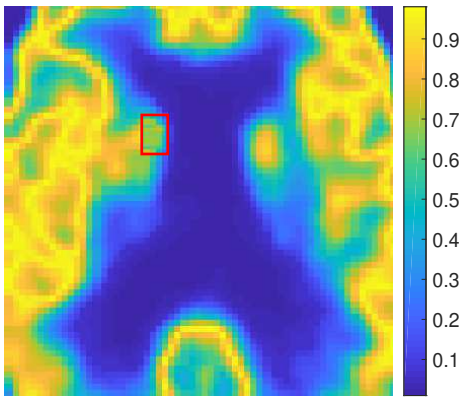
(b) The credible level $(1-\alpha)$ computed for test image one.



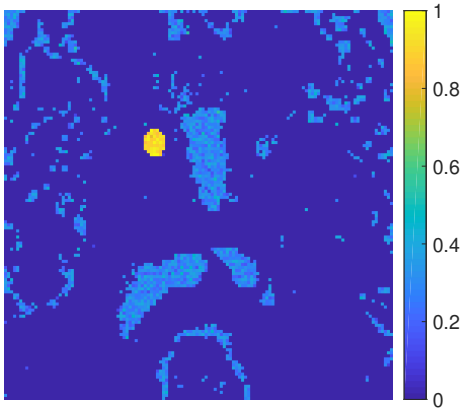
(c) Surrogate test image two.



(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 9. *($K = 0.5$) The credible level for three surrogate test images.*

truth, and so the credible level result in Figure 8(b) 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 contain artifacts. Figure 9 shows the same test results but for $K = 0.5$, and one can see that the figure exhibits qualitatively the same behaviors as those of $K = 1$. 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 medical 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 a reconstructed image. We believe the proposed Bayesian framework can be used to reconstruct images and evaluate the uncertainty associated to reconstruction in many practical medical imaging problems with Poisson data.

There are several problems related to this work that we plan to investigate in the future. In the future, we plan to apply the methods developed in this work to those real-world problems, especially the PET image reconstruction.

Appendix A. Proof of Proposition 2.1. We provide a proof of Proposition 2.1 here.

Proof. (1) From (2.2) and (2.4), we obtain directly that

$$0 < \underline{\boldsymbol{\theta}} \le \boldsymbol{\theta} \le \overline{\boldsymbol{\theta}}$$

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

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

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

$$\begin{aligned}\Phi(z) &= \langle \boldsymbol{\theta}, \mathbf{1}\rangle - \langle y, \ln \boldsymbol{\theta}\rangle \\ &\le \langle \boldsymbol{\theta}, \mathbf{1}\rangle + \|\ln \boldsymbol{\theta}\|_2 \|y\|_2\end{aligned}$$

$$\leq \langle \overline{\boldsymbol{\theta}}, \mathbf{1} \rangle + l_{\max} \|y\|_2$$
$$\leq \langle \overline{\boldsymbol{\theta}}, \mathbf{1} \rangle + l_{\max} r.$$

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

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

Since the Radon transform A is a bounded linear operator from the L_2 space to R^d [30], we have

$$\begin{aligned} \|Af(z) - Af(v)\|_2 &\leq \|A\| \|f(z) - f(v)\|_{L_2} \\ &= \|A\| \| \int_v^z e^{-t^2} dt \|_{L_2} \leq \|A\| \| \int_v^z dt \|_{L_2} = \|A\| \|z - v\|_{L_2}, \end{aligned}$$

which completes the proof.

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

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

Appendix B. Proof of Theorem 3.1. We define $\eta_0(z, v)$ to be the measure $\eta(z, v)$ on $X \times X$ with $\Psi \equiv 0$, and it is obvious that the measure $\eta_0(z, v)$ is Gaussian. Moreover we have

$$\eta(dz, dv) = q(z, dv)\mu(dz), \quad \eta_0(dz, dv) = q(z, dv)\mu_0(dz) \tag{B.1}$$

and that the measures μ and μ_0 are equivalent. It follows that η and η_0 are equivalent and

$$\frac{d\eta}{d\eta_0}(z, v) = \frac{d\mu}{d\mu_0}(z) = Z \exp(-\Psi(z)), \quad \frac{d\eta}{d\eta_0}(v, z) = Z \exp(-\Psi(v)). \tag{B.2}$$

Now we define

$$\eta_0^{\perp}(dz, dv) = q(v, dz)\mu_0(dv),$$

and by some elementary calculations we can derive

$$\begin{aligned} \frac{d\eta_0^{\perp}}{d\eta_0}(z, v) = \exp \Bigg(& -\frac{1}{2} \frac{||2\mathcal{C}_0^{-\frac{1}{2}}(z - v) + \delta \mathcal{C}_0^{-\frac{1}{2}}(v + z) + 2\delta \mathcal{C}_0^{\frac{1}{2}} g(v)||^2}{8\delta} - \frac{1}{2} \frac{||v||^2}{\mathcal{C}} \\ & + \frac{1}{2} \frac{||2\mathcal{C}_0^{-\frac{1}{2}}(v - z) + \delta \mathcal{C}_0^{-\frac{1}{2}}(z + v) + 2\delta \mathcal{C}_0^{\frac{1}{2}} g(z)||^2}{8\delta} + \frac{1}{2} \frac{||z||^2}{\mathcal{C}} \Bigg) \end{aligned} \tag{B.3}$$

$$\begin{aligned} = \exp \Bigg(& -\frac{1}{2} \langle z - v, g(v) \rangle - \frac{\delta}{4} \langle (z + v), g(v) \rangle - \frac{\delta}{4} \langle \mathcal{C}_0^{\frac{1}{2}} g(v), \mathcal{C}_0^{\frac{1}{2}} g(v) \rangle \\ & + \frac{1}{2} \langle v - z, g(z) \rangle + \frac{\delta}{4} \langle (v + z), g(z) \rangle + \frac{\delta}{4} \langle \mathcal{C}_0^{\frac{1}{2}} g(z), \mathcal{C}_0^{\frac{1}{2}} g(z) \rangle \Bigg). \end{aligned}$$

As $z, v \in X$ and $g(z)$ is in the Cameron–Martin space of μ_0, $\langle z - v, g(z)\rangle$, $\langle v + z, g(z)\rangle$, and $||\mathcal{C}_0^{\frac{1}{2}} g(z)||^2$ are finite, and $\frac{d\eta_0^\perp}{d\eta_0}$ is well defined. Now recall that

$$\frac{d\eta^\perp}{d\eta}(z, v) = \frac{d\eta}{d\eta_0}(v, z)\frac{d\eta_0^\perp}{d\eta_0}(z, v)\frac{d\eta_0}{d\eta}(z, v). \tag{B.4}$$

Substituting (B.2) and (B.3) into (B.4) yields

$$\frac{d\eta^\perp}{d\eta}(z, v) = \exp(\rho(z, v) - \rho(v, z)), \tag{B.5}$$

where

$$\rho(z, v) = \Phi(z) + \frac{1}{2}\langle v - z, g(z)\rangle + \frac{\delta}{4}\langle z + v, g(z)\rangle + \frac{\delta}{4}||\mathcal{C}_0^{\frac{1}{2}} g(z)||^2. \tag{B.6}$$

REFERENCES

[1] K. A Johnson and J. Becker, The Whole Brain Atlas, www.med.harvard.edu/aanlib/, accessed April 4, 2010.

[2] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, *An introduction to MCMC for machine learning*, Mach. Learn., 50 (2003), pp. 5–43.

[3] S. R. Arridge, K. Ito, B. Jin, and C. Zhang, *Variational Gaussian approximation for Poisson data*, Inverse Problems, 34 (2018), 025005.

[4] D. L. Bailey, M. N. Maisey, D. W. Townsend, and P. E. Valk, *Positron Emission Tomography*, Springer, New York, NY, 2005.

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

[6] J. M. Bardsley and A. Luttman, *A Metropolis-Hastings method for linear inverse problems with Poisson likelihood and Gaussian prior*, Int. J. Uncertain. Quantif., 6 (2016), pp. 35–55.

[7] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni, *A discrepancy principle for poisson data*, Inverse Problems, 26 (2010), 105004.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.

[9] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*, CRC Press, Boca Raton, FL, 2011.

[10] D. Calvetti and E. Somersalo, *A Gaussian hypermodel to recover blocky objects*, Inverse Problems, 23 (2007), 733.

[11] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

[12] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[13] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.

[14] S. L. Cotter, G. O. Roberts, A. Stuart, and D. White, *MCMC methods for functions: Modifying old algorithms to make them faster*, Statist. Sci., 28 (2013), pp. 424–446.

[15] A. Durmus, E. Moulines, and M. Pereyra, *Efficient Bayesian computation by proximal Markov chain Monte Carlo: When Langevin meets Moreau*, SIAM J. Imaging Sci., 11 (2018), pp. 473–506.

[16] A. Fernandez-Vidal and M. Pereyra, *Maximum likelihood estimation of regularisation parameters*, in Proceedings of the IEEE International Conference On Image Processing, 2018.

[17] J. A. Fessler, *Penalized weighted least-squares image reconstruction for positron emission tomography*, IEEE Trans. Med. Imaging, 13 (1994), pp. 290–300.

[18] J. A. Fessler, *Medical Image Reconstruction: A Brief Overview of Past Milestones and Future Directions*, arXiv preprint, arXiv:1707.05927, 2017.
[19] A. Gelman, *Posterior predictive assessment of model fitness via realized discrepancies (with discussion)*, Statist. Sinica, 6 (1996), pp. 733–760.
[20] A. Gelman, H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, Chapman and Hall/CRC, Boca Raton, FL, 2013.
[21] P. J. Green, *Bayesian reconstructions from emission tomography data using a modified EM algorithm*, IEEE Trans. Med. Imaging, 9 (1990), pp. 84–93.
[22] M. Hairer, A. M. Stuart, and S. J. Vollmer, *Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions*, Ann. Appl. Probab., 24 (2014), pp. 2455–2490.
[23] P. Hall and D. Titterington, *Common structure of techniques for choosing smoothing parameters in regression problems*, J. R. Stat. Soc. Ser. B Star. Methodol., 49 (1987), pp. 184–198.
[24] T. Hebert and R. Leahy, *A generalized em algorithm for 3-d Bayesian reconstruction from Poisson data using Gibbs priors*, IEEE Trans. Med. Imaging, 8 (1989), pp. 194–202.
[25] T. Hohage and F. Werner, *Inverse problems with Poisson data: Statistical regularization theory, applications and algorithms*, Inverse Problems, 32 (2016), 093001.
[26] A. Hore and D. Ziou, *Image quality metrics: PSNR vs. SSIM*, in Proceedings of the 2010 20th International Conference on Pattern Recognition, IEEE, 2010, pp. 2366–2369.
[27] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, New York, NY, 2005.
[28] J. Li, *A note on the Karhunen–Loève expansions for infinite-dimensional Bayesian inverse problems*, Statist. Probab. Lett., 106 (2015), pp. 1–4.
[29] E. Ü. Mumcuoglu, R. M. Leahy, and S. R. Cherry, *Bayesian reconstruction of PET images: Methodology and performance analysis*, Phys. Med. Biol., 41 (1996), 1777.
[30] F. Natterer, *The Mathematics of Computerized Tomography*, SIAM, Philadelphia, PA, 2001.
[31] J. M. Ollinger and J. A. Fessler, *Positron-emission tomography*, IEEE Signal Process. Mag., 14 (1997), pp. 43–55.
[32] T.-S. Pan and A. E. Yagle, *Numerical study of multigrid implementations of some iterative image reconstruction algorithms*, IEEE Trans. Med. Imaging, 10 (1991), pp. 572–588.
[33] C. A. d. B. Pereira, J. M. Stern, and S. Wechsler, *Can a significance test be genuinely Bayesian?*, Bayesian Anal., 3 (2008), pp. 79–100.
[34] M. Pereyra, *Proximal Markov chain Monte Carlo algorithms*, Stat. Comput., 26 (2016), pp. 745–760.
[35] M. Pereyra, *Maximum-a-posteriori estimation with Bayesian confidence regions*, SIAM J. Imaging Sci., 10 (2017), pp. 285–302.
[36] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, *Computing the Cramer–Rao bound of Markov random field parameters: Application to the Ising and the Potts models*, IEEE Signal Process. Lett., 21 (2013), pp. 47–50.
[37] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Science & Business Media, New York, NY, 2013.
[38] D. Rudolf and B. Sprungk, *On a generalization of the preconditioned Crank–Nicolson Metropolis algorithm*, Found. Comput. Math., 18 (2018), pp. 309–343.
[39] L. A. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, IEEE Trans. Med. Imaging, 1 (1982), pp. 113–122.
[40] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.
[41] M. Thulin, *Decision-theoretic justifications for Bayesian hypothesis testing using credible sets*, J. Statist. Plann. Inference, 146 (2014), pp. 133–138.
[42] Y. Vardi, L. Shepp, and L. Kaufman, *A statistical model for positron emission tomography*, J. Amer. statist. Assoc., 80 (1985), pp. 8–20.
[43] Z. Yao, Z. Hu, and J. Li, *A TV-Gaussian prior for infinite-dimensional Bayesian inverse problems and its numerical implementations*, Inverse Problems, 32 (2016), 075006.
[44] J. C. Ye, C. A. Bouman, K. J. Webb, and R. P. Millane, *Nonlinear multigrid algorithms for Bayesian optical diffusion tomography*, IEEE Trans. Image Process., 10 (2001), pp. 909–922.