CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Enhancing industrial X-ray tomography by data-centric statistical methods

Jarkko Suuronen[1], Muhammad Emzir[2], Sari Lasanen[3], Simo Särkkä[2] and Lassi Roininen[1] (ID)

[1]School of Engineering Science, Lappeenranta-Lahti University of Technology, Lappeenranta, Finland
[2]Department of Electrical Engineering and Automation, Aalto University, Aalto, Finland
[3]Sodankylä Geophysical Observatory, University of Oulu, Oulu, Finland
*Corresponding author. E-mail: lassi.roininen@lut.fi

Abstract

X-ray tomography has applications in various industrial fields such as sawmill industry, oil and gas industry, as well as chemical, biomedical, and geotechnical engineering. In this article, we study Bayesian methods for the X-ray tomography reconstruction. In Bayesian methods, the inverse problem of tomographic reconstruction is solved with the help of a statistical prior distribution which encodes the possible internal structures by assigning probabilities for smoothness and edge distribution of the object. We compare Gaussian random field priors, that favor smoothness, to non-Gaussian total variation (TV), Besov, and Cauchy priors which promote sharp edges and high- and low-contrast areas in the object. We also present computational schemes for solving the resulting high-dimensional Bayesian inverse problem with 100,000–1,000,000 unknowns. We study the applicability of a no-U-turn variant of Hamiltonian Monte Carlo (HMC) methods and of a more classical adaptive Metropolis-within-Gibbs (MwG) algorithm to enable full uncertainty quantification of the reconstructions. We use maximum a posteriori (MAP) estimates with limited-memory BFGS (Broyden–Fletcher–Goldfarb–Shanno) optimization algorithm. As the first industrial application, we consider sawmill industry X-ray log tomography. The logs have knots, rotten parts, and even possibly metallic pieces, making them good examples for non-Gaussian priors. Secondly, we study drill-core rock sample tomography, an example from oil and gas industry. In that case, we compare the priors without uncertainty quantification. We show that Cauchy priors produce smaller number of artefacts than other choices, especially with sparse high-noise measurements, and choosing HMC enables systematic uncertainty quantification, provided that the posterior is not pathologically multimodal or heavy-tailed.

Impact Statement

Industrial X-ray tomography reconstruction accuracy depends on various factors, like the equipment, measurement geometry, and constraints of the target. For example, dynamical systems are harder targets than static ones. The harder and noisier the setting becomes, the more emphasis goes on mathematical modeling of the targets. Bayesian statistical inversion is a common choice for difficult measurement settings, and its limitations mainly come from the choice of the a priori models. Gaussian models are widely studied, but they provide smooth reconstructions. Total variation priors are not invariant under mesh changes, so doing systematic uncertainty quantification, like data-centric sensor optimization, cannot be done with them. Besov and Cauchy priors

CrossMark

however provide systematic non-Gaussian random field models, which can be used for contrast-boosting tomography. The drawback is higher computational cost. Hence, the techniques developed here are useful for non–time-critical applications with difficult measurement settings. In these cases, the methods developed may provide significantly better reconstructions than the traditional methods, like filtered back-projection.

1. Introduction

X-ray tomography has applications in various industrial fields such as sawmill industry, where it can be used for detecting knots, rotten parts, and foreign objects in logs (Shustrov et al., 2019; Zolotarev et al., 2019). In oil and gas industry, X-ray tomography can be used to analyze drill-core rock samples for identifying pore structures (Mendoza et al., 2019) and other structural properties of rock. In chemical engineering, X-ray microtomography can be used to measure the internal structure of substances at the micrometer level (Ou et al., 2017). In manufacturing, it can be used in nondestructive testing (Garcea et al., 2018; Rotella et al., 2018), endurance testing (Piao et al., 2019), and dimensional metrology (Kruth et al., 2011; Villarraga-Gómez et al., 2019). In geotechnical engineering, X-ray microtomography can be used to measure soil properties in laboratories, while a closely related travel-time tomography can be used to measure the structure of soil and rocks from cross-borehole measurements (Ernst et al., 2007). In medical engineering, applications include dental imaging (Kolehmainen et al., 2006) and tumor tracking in radiotherapy (Shieh et al., 2017).

The principle of X-ray tomography is that we transmit X-ray radiation to the object, which penetrates the object of interest over a collection of propagation paths, and the attenuated X-ray radiation is measured in a detector system (Toft, 1996). This allows us to estimate internal properties of an unknown object given the noise-perturbed indirect measurements. However, typically, we can transmit and measure the X-rays only from a limited number of angles around the object which makes it harder to reconstruct the internal structure of the object from the measurements. The reconstruction problem is thus an inverse problem where the measurements give only a limited amount of information on the object of interest (Siltanen et al., 2003).

In order to successfully reconstruct the internals of the object from the limited number of measurements, we need to introduce additional information to the reconstruction process. In this article, we study so-called Bayesian methods (Kaipio and Somersalo, 2005), where we introduce a statistical prior model for the possible internal structures in form of a probability distribution. This prior model encodes the information on what kind of structures are more likely and which are less likely than others. For example, a Gaussian random field prior puts higher probability for smoother structures, whereas total variation (TV), Besov, and Cauchy priors favor structures that can have sharper edges between the substructures (González, 2017). The advantage of the Bayesian formulation is that it provides uncertainty quantification mechanism to the reconstruction problem, as it not only provides single reconstruction but also the error bars for the reconstruction in the form of a probability distribution (Bardsley, 2012).

The Bayesian formulation of the X-ray tomography transforms the solution of the associated inverse problem to a Bayesian inference problem, and we need to use computational methods from Bayesian statistics to solve it. We use Markov chain Monte Carlo (MCMC) and optimization methods for computing conditional mean (CM) and maximum a posteriori (MAP) estimates, respectively. We consider a sawmill industry log tomography application for detecting knots, rotten parts, and foreign objects in sawmills (Shustrov et al., 2019; Zolotarev et al., 2019). As an oil and gas industry application, we have drill-core rock sample tomography for identifying pore structures (Mendoza et al., 2019).

1.1. X-ray tomography as a Bayesian statistical inverse problem

Reducing the number of measurements tends to add artefacts to the tomographic reconstruction. This means that we need to carefully evaluate the accuracy of the reconstruction for different levels of sparsity.

For this kind of problems, a typical approach is to deploy Bayesian statistical inversion techniques in the sense of Kaipio and Somersalo (2005). Bayesian inversion is the theory and practical data analysis of noisy indirect measurements within the Bayesian estimation framework. Following Toft (1996), X-ray radiation measured at single detector pixel over a given propagation path is

$$y_{\theta,s} = \iint \mathcal{X}(x_1, x_2)\delta(s - x_1\cos\theta - x_2\sin\theta)dx_1 dx_2 + e_{\theta,s} =: \mathcal{A}_{\theta,s}\mathcal{X} + e_{\theta,s}, \tag{1}$$

where $y_{\theta,s}$ is measured value at angle θ with translation s and $\mathcal{X}$ is varying X-ray attenuation coefficient field inside the object of interest. For practical computations, we discretize the operation $\mathcal{A}_{\theta,s}\mathcal{X}$ by approximating $\mathcal{X}$ with a piecewise defined function with constant values on square pixels and obtain a finite-dimensional presentation $y = AX + e$, where $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times m}$, $e \in \mathbb{R}^n$, and $X \in \mathbb{R}^m$ represent the approximated values of $\mathcal{X}$ on pixels.

In X-ray tomography, the noise process e is often modeled as a compound Poisson distributed noise (Whiting, 2002; Thanh et al., 2019). We make a simplification and assume that e is zero-mean Gaussian white noise with covariance $C = \sigma^2 I$. As our main aim is in contrast-boosting priors, we assume that this simplification does not affect the generality of the results. The solution of the tomography problem is then an a posteriori probability density via the Bayes formula

$$\pi(X|y) = \frac{\pi(X)\pi(y|X)}{\pi(y)} \propto \pi(X)\pi(y|X) = \pi(X)\exp\left(-\frac{1}{2}(y - AX)^T C^{-1}(y - AX)\right), \tag{2}$$

where $\pi(X)$ is the a priori density, that is, the probabilistic description of the unknown we know before any measurements are taken, $\pi(y|X)$ is the likelihood density, and $\pi(y)$ is a normalization constant which we can omit in computations.

1.2. Literature review

Sparse-angle tomography is an ill-posed inverse problem (Natterer, 2001), and thus in order to have stable solutions, we need to set the prior $\pi(X)$. A standard choice is to use Gaussian random field priors (Bardsley, 2013). They are easy to use through their analytic properties, and they are fully defined by their means and covariances. The drawback of Gaussian priors is that they are locally smoothing and thus cause contrast artefacts near sharp edges. As our objective is in reconstructing unknown objects with sharp edges, we need to choose a prior density that can model them. Such an edge-preserving, contrast-boosting prior needs to be non-Gaussian.

When computing posterior estimates, the unavoidable non-Gaussianity of any contrast-boosting prior leads to computational complexity and high computational cost. In two-dimensional industrial tomography, with even small or moderate mesh sizes, a typical problem has around 100,000–1,000,000 unknown parameters. High dimensionality in industry poses problems, especially in time-critical applications, where speed is valued. Point estimates computed with optimization methods are natural starting points as they are computationally light. For example, the MAP estimate is fast to compute but does not unfortunately provide full uncertainty quantification.

To fully quantify uncertainty, we need to explore the whole posterior with MCMC techniques, such as Metropolis-within-Gibbs, (MwG), (see, e.g., Markkanen et al., 2019). Alternatively, we could use variational Bayes methods (Tzikas et al., 2008) which aim to approximate a probability distribution with a simpler one, and they are computationally cheaper than MCMC. However, variational Bayes methods have inherent approximation errors in their estimates, whereas MCMC methods theoretically do not. That is why MCMC is often preferred for uncertainty quantification. There is a growing number of MCMC algorithms for high-dimensional problems (Cotter et al., 2013; Law, 2014; Beskos et al., 2017). Unfortunately, non-Gaussianity of the contrast-boosting priors breaks many of the assumptions of the function-space MCMC methods. MwG is a simple choice, but its convergence can be slow, requiring significantly long MCMC chains for accurate approximations. As an alternative, we use Hamiltonian Monte Carlo (HMC), (Neal, 2011). It has been used for a wide range of applications, and it is well-suited

for large-dimensional problems (Beskos et al., 2011). In particular, we use the HMC variant with no-U-turn sampling (HMC-NUTS), (Hoffman and Gelman, 2014).

There are several priors which promote contrast-boosting or edge-preserving inversion. A standard choice for edge-preserving inversion is the TV prior, which is based on using L^1-norms. In statistics, this method is called LASSO (least absolute shrinkage and selection operator). This method, however, has a drawback, which arises from the finiteness of the prior moments. When we make the discretization denser and denser, the TV priors converge to Gaussian priors in the discretization limit. This behavior was studied by Lassas and Siltanen (2004), and they showed that the estimators are not consistent under mesh refinement for Bayesian statistical inverse problems. This, naturally, means that doing uncertainty quantification under TV-prior assumption is not consistent with respect to the change of the mesh.

Recently, several hierarchical models, which promote more versatile behaviors, have been developed. These include, for example, deep Gaussian processes (Dunlop et al., 2018; Emzir et al., 2020), level-set methods (Dunlop et al., 2017), mixtures of compound Poisson processes and Gaussians (Hosseini, 2017), and stacked Matérn fields via stochastic partial differential equations (Roininen et al., 2019). The problem with hierarchical priors is that in the posteriors, the parameters and hyperparameters may become strongly coupled, which means that vanilla MCMC methods become problematic and, for example, reparameterizations are needed for sampling the posterior efficiently (Chada et al., 2019; Monterrubio-Gómez et al., 2020). In level-set methods, the number of levels is usually low because experiments have shown that the method deteriorates when the number of levels is increased.

Here, we utilize a different approach and make the prior non-Gaussian by construction without hierarchical modeling or compromising on uncertainty quantification. The first choice is to utilize priors based on Besov $\beta^s_{p,q}$ norms (Lassas et al., 2009). These priors are constructed on wavelet basis, typically Haar wavelets, and they have well-defined non-Gaussian discretization limit behavior. Besov priors have been utilized in a number of studies for Bayesian inversion (see, e.g., Niinimäki, 2013). The problem with Besov priors is the structure of wavelets and the truncation of the wavelet series. The coefficients of the wavelets are random but actually not interchangeable, since their decay is necessary for convergence of the series. That typically means that we make an unnecessary and strong prior assumption for edge locations based on wavelet properties, for example, consider a one-dimensional inverse problem on domain (0,1), we prefer an edge at 1/4 over an edge at 1/3 for Haar wavelets.

One possibility is to use α-stable prior fields, see Samorodnitsky and Taqqu (1994). Another type of α-stable priors can be constructed similarly to Besov priors. They use Karhunen-Loève (KL) type expansions instead of wavelet expansions but replace Gaussian coefficients in KL expansions with α-stable coefficients. For KL expansions, see Berlinet and Thomas-Agnan (2004), and for stable field expansions, see Sullivan (2017). Third option is to use pixel-based approaches (Bolin, 2014; Markkanen et al., 2019; Mendoza et al., 2019), which cover also some α-stable random walks. The family of α-stable processes contains both KL expansions and pixel-based approaches, but one should note that the different approaches lead to different statistical objects. We stress that both approaches lead to well-posedness of the inverse problem. In this paper, we shall utilize certain pixel-based approaches and limit the discussion to the Cauchy difference priors, which is a special case $\alpha = 1$.

1.3. Contribution and organization of this work

The contributions of this paper are twofold, first of all, we make a large-scale numerical comparison for the X-ray tomography problem with different Gaussian and non-Gaussian (TV, Besov, Cauchy) prior assumptions. For the log and drill-core tomography applications, our paper shows the tomographic imaging limits, that is, when building the actual industrial tomography systems, we know the minimum number of measurements and noise levels with which we can attain reconstructions with predefined accuracy. This minimizes the number of components, and thus the cost, needed in the imaging system but, at the same time, ensures operational accuracy. Secondly, we apply a carefully designed HMC algorithm to the high-dimensional non-Gaussian inverse problem. In particular, we show its applicability for the sparse-angle tomography problem. Also, studying HMC chain mixing and sampling from over high-dimensional non-Gaussian posteriors is a new contribution.

The rest of this paper is organized as follows: In Section 2, we review Gaussian, TV, Besov, and Cauchy priors. In Section 3, we introduce the necessary MwG and HMC tools. In Section 4, we have two synthetic case studies: Three dimensional-imaging of logs in sawmills and drill-core tomography problem. Finally, in Section 5, we conclude the study and make some notes on future developments.

2. Random field priors

For Gaussian, TV, and Besov priors, we write the priors as $\pi(X) \propto \exp(-G(X))$, where $G(X)$ is L^2-norm of differences of X for the Gaussian case, L^1-norm of differences for TV, and Besov $\mathcal{B}_{p,q}^s$-norm for a wavelet expansion with random coefficients $X_1, \ldots, X_m$ for Besov case. We will describe these distributions below and their respective $G(X)$. In the following, we will describe separately the Cauchy prior, as it is based on another type of construction.

2.1. Gaussian prior

Let us assume a zero-mean Gaussian prior, which is fully defined with one of the following choices,

$$G(X) = \frac{1}{2} X^T \Sigma^{-1} X = \frac{1}{2} X^T Q X = \frac{1}{2} X^T L^T L X = \frac{1}{2} (LX)^T LX,$$

where Σ is the covariance matrix, Q is the precision matrix, and L is a square-root (e.g., Cholesky factor) of the precision matrix. Common choices are exponential and squared exponential covariances, Matérn covariances, and Brownian motion covariance. Different choices lead to different kinds of presentations, for example, squared exponential leads to full matrices for all Σ, Q, L, but with certain choices, the Matérn covariances have sparse Q, L matrices due to the Markov property (Roininen et al., 2019).

Two-dimensional difference priors can be obtained by choosing L to be a discretization of the following operator equation $\nabla \mathcal{X} = \mathcal{w}$, where $\mathcal{w} = (\mathcal{w}_1, \mathcal{w}_2)^T$ with two statistically independent white noise random fields $\mathcal{w}_1$ and $\mathcal{w}_2$. Thus, we have $LX = W$ for the discrete presentation, which we will solve in the least-squares sense. We note that the precision matrix $Q = L^T L$ is not invertible, thus this is an improper prior, but if we impose zero-boundary conditions, we have a proper prior. For X-ray tomography, zero-boundary conditions are natural choices, as typically the domain of interest is larger than the object of interest, and thus putting zero-boundary is justified. Now, let us denote by $X_{i,j}$, the unknown at pixel (i,j). As a summation formula with zero-boundary conditions, we can write the prior then as

$$G_{\text{Gauss}}(X) = G_{\text{pr}}(X) + G_{\text{boundary}}(X) = \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\frac{(X_{i,j} - X_{i-1,j})^2}{\sigma_{\text{pr}}^2} + \frac{(X_{i,j} - X_{i,j-1})^2}{\sigma_{\text{pr}}^2} \right) + \sum_{i,j \in \partial\Omega} \frac{X_{i,j}^2}{\sigma_{\text{boundary}}^2},$$

where by $\partial\Omega$, we mean the boundary indices of the discretized domain Ω, σ_{pr}^2 is regularization parameter, and $\sigma_{\text{boundary}}^2$ zero-boundary parameter.

2.2. Total variation prior

Similarly to the Gaussian prior, we can write a two-dimensional TV prior with

$$G_{\text{TV}}(X) = G_{\text{pr}}(X) + G_{\text{boundary}}(X) = \sum_{i=1}^{I} \sum_{j=1}^{J} \alpha \left(|X_{i,j} - X_{i-1,j}| + |X_{i,j} - X_{i,j-1}| \right) + \sum_{i,j \in \partial\Omega} \alpha_{\text{boundary}} |X_{i,j}|,$$

where $\alpha, \alpha_{\text{boundary}}$ are regularization parameters. We note that in statistics, these distribution are often called Laplace distributions. This form of TV prior is anisotropic. According to González et al. (2017), it can be shown that an isotropic form is obtained when we choose

$$G_{\text{pr}}(X) = \sum_{i=1}^{I} \sum_{j=1}^{J} \alpha \left((X_{i,j} - X_{i-1,j})^2 + (X_{i,j} - X_{i,j-1})^2 \right)^{1/2}.$$

Using higher-order TV priors would be possible (Liu et al., 2015), but often the basic first-order TV prior is selected because it preserves sharp edges within the reconstruction much better. The major drawback of TV prior is that due to its finite moments, the prior is not discretization-invariant, and it resembles Gaussian difference prior on dense enough meshes. It is notable that MAP estimate and the CM estimate differ drastically of each other in the limit (Lassas and Siltanen, 2004). While the inconsistency is not desirable, the prior is still used in many practical applications. In the numerical experiments, we use anisotropic TV prior. We apply Charbonnier approximation to make the gradient of the posterior differentiable everywhere (Charbonnier et al., 1997).

2.3. Besov prior

Gaussian and TV priors are based on Gaussian and Laplace distributions. Besov priors are based on the wavelet coefficients of the reconstruction. The coefficients are calculated by discrete wavelet transform (DWT) (Lassas et al., 2009). The discrete wavelet transform decomposes the reconstruction into approximation and detail coefficients. For detailed treatise on Besov priors, see (Niinimäki, 2013). Besov prior for X derives from wavelet expansion for function $\mathcal{X}$

$$\mathcal{X} = \sum_{d_1=0}^{2^k-1}\sum_{d_2=0}^{2^k-1}\langle\mathcal{X},\phi_{k,d_1,d_2}\rangle\phi_{k,d_1,d_2} + \sum_{r=k}^{\infty}\sum_{d_1=0}^{2^r-1}\sum_{d_2=0}^{2^r-1}\sum_{t=1}^{3}\langle\mathcal{X},\psi_{r,d_1,d_2,t}\rangle\psi_{r,d_1,d_2,t}, \tag{3}$$

and its Besov norm is

$$\|\mathcal{X}\|_{\mathscr{B}^s_{p,q}} = \left(\sum_{d_1=0}^{2^k-1}\sum_{d_2=0}^{2^k-1}|\langle\mathcal{X},\phi_{k,d_1,d_2}\rangle|^p + \sum_{r=k}^{\infty}2^{rq(s+1-2/p)}\sum_{d_1=0}^{2^r-1}\sum_{d_2=0}^{2^r-1}\sum_{t=1}^{3}|\langle\mathcal{X},\psi_{r,d_1,d_2,t}\rangle|^p\right)^{1/p}.$$

In Equation (3), ϕ is the father wavelet and ψ is the mother wavelet. The subindices refer to different scales and translations of the functions (Meyer, 1992). We utilize the prior for Besov space $\mathscr{B}^1_{1,1}$, which reduces the norm into absolute sum of wavelet coefficients. Other Besov spaces than $\mathscr{B}^1_{1,1}$ would be doable as well, but in those cases, the wavelet coefficients would have different weights or they would be raised to other powers than one. We use Haar wavelets as our father and mother wavelets. Similarly to the TV prior, we use Charbonnier approximation for the gradient.

In practise, we compute the two-dimensional DWT of X, which is now represented as a reconstruction of size of $2^n \times 2^n$, $n \in \mathbb{N}$ square pixels. The Besov prior is defined by

$$G_{\text{Besov}}(X) = \sum_{d_1=0}^{2^k-1}\sum_{d_2=0}^{2^k-1}|a_{k,d_1,d_2}| + \sum_{r=k}^{k_{\max}}\sum_{d_1=0}^{2^r-1}\sum_{d_2=0}^{2^r-1}\sum_{t=1}^{3}|b_{r,d_1,d_2,t}|.$$

Terms a_{k,d_1,d_2} refer to the approximation coefficients of the DWT of X and terms $b_{r,d_1,d_2,t}$ to the corresponding detail coefficients. As DWT algorithm, we choose a matrix operator method described by Wang and Vieira (2010). A more common and faster way to calculate the DWT would be to use fast wavelet transform, but we want to keep track of how changing the value of an individual pixel alters the wavelet coefficients and therefore the prior, too. This property is needed in the MwG algorithm because we propose new samples for each component of X in a sequential manner and calculating the whole DWT every time for all the components would be too time consuming.

Besov priors are discretization-invariant and therefore should remain consistent if the resolution is increased (Lassas et al., 2009). Their drawback is the tendency to produce block artefacts into the point estimates: the wavelet coefficients at certain low scale are different, even though they are adjacent to each other. This makes Besov priors location-dependent, that is, they are not translation-invariant.

2.4. Cauchy prior

While the previous three priors were constructed based on different types of norms, Cauchy priors are constructed based on Cauchy walk or more profoundly general α-stable random walk. These one-

dimensional objects have well-defined limits, and the generalization in (Markkanen et al., 2019) to two-dimensional setting is given as a prior density

$$\pi(X) \propto \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{h\lambda}{(h\lambda)^2 + (X_{i,j} - X_{i-1,j})^2} \frac{h\lambda}{(h\lambda)^2 + (X_{i,j} - X_{i,j-1})^2},$$

where h is discretization step in both coordinate directions, and λ is the regularization parameter. These priors have been used for X-ray tomography in oil and gas industry (Mendoza et al., 2019) and subsurface imaging (Muhumuza et al., 2020).

A second formulation of the Cauchy difference prior was derived in (Chada et al., 2019). In that, the starting point was α-stable sheets, which lead to difference approximations of the form

$$\pi(X) \propto \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{h^2\lambda}{(h^2\lambda)^2 + (X_{i,j} - X_{i-1,j} - X_{i,j-1} + X_{i-1,j-1})^2}.$$

We note that Cauchy difference prior is theoretically justified in the discretization limit, which is not the case with TV prior. Unlike Gaussian difference prior, Cauchy difference prior favors small increments and steep transitions relatively much more. This means that Cauchy difference prior preserves edges within the lattice and does not smooth them out. A disadvantage of Cauchy difference prior is its anisotropic nature.

3. Posterior estimates

The two most common estimators drawn from the posterior density (Equation 2) are the MAP and CM estimates. For MAP estimation, we will use limited-memory BFGS (L-BFGS) optimization algorithm, which belongs to the family of quasi-Newtonian methods. For CM estimation, we use MCMC methods, as they also enable uncertainty quantification of the estimators by generating samples from the target distribution Kaipio and Somersalo (2005). For a detailed MCMC review, see Luengo et al. (2020).

3.1. Adaptive Metropolis-within-Gibbs

We use single-component adaptive Metropolis-Hastings (SCAM) (Haario et al., 2005), an adaptive variant of the MwG, (Brooks et al., 2011) algorithm, to generate samples from the posterior distribution. At each iteration, each component of the target distribution is treated as a univariate probability distribution, while all the other component values of the distribution are kept constant. After a new value for the component is proposed, for example, using a Gaussian distributed transition from the previous value of the component, the new value is either accepted or rejected according to the Metropolis-Hastings rule. A new sample is added to the chain after all the components have been looped over. The process then continues until a fixed number of samples have been generated.

Adaptive MwG uses a Gaussian proposal distribution and adapts its proposal covariance by using previous samples. In high-dimensional distributions, it is more efficient than a Metropolis-Hastings algorithm which proposes the new values for all components at once (Raftery et al., 2001).

3.2. Hamiltonian Monte Carlo

In Hamiltonian, or Hybrid, Monte Carlo (HMC), the negative logarithm of the probability density function of the target distribution is interpreted as a potential energy function of a physical system (Neal, 2011; Betancourt, 2017; Barp et al., 2018) with the target variable being the position variable of the system. Samples are generated by simulating trajectories from the physical system defined by the potential energy function and an additional momentum variable. The energy of the physical system is the sum of the potential energy and kinetic energy which is proportional to the square of the momentum. One initializes a random momentum vector, sets the position to be the last sample in the chain, and

simulates a trajectory of the system using so-called symplectic integrator (Barp et al., 2018). The final position of the trajectory is then accepted or rejected into the chain using a Metropolis-Hastings rule.

An advantage of HMC is that if trajectory simulation parameters are suitable, it can sample from the posterior distribution efficiently. It also takes the correlations of the distribution components into account. Its computational complexity also scales better with the dimensionality than that of a simple Metropolis-Hastings (Neal, 2011; Beskos et al., 2013) algorithm.

The trajectory simulation of HMC requires manually specifying a set of parameters, which greatly affect the efficiency of the sample generation. To cope with this problem, there is an automatic HMC algorithm called no-U-turn (NUTS) HMC (Hoffman and Gelman, 2014), which eliminates the need for manual tuning. For these reasons, we use NUTS as an alternative to MwG in our numerical experiments. However, even with automatic tuning, HMC may still struggle at sampling from multimodal (Mangoubi et al., 2018) or heavy-tailed distributions (Livingstone et al., 2017). To cope with these problems, it would be possible to use methods such as Riemannian manifold HMC (Girolami and Calderhead, 2011) or continuously tempered HMC (Graham and Storkey, 2017) instead of the plain NUTS HMC.

4. Numerical experiments

We have two numerical experiments, the log tomography and the drill-core tomography. In log tomography, we have a foreign object inside, thus it is a high-contrast target. The second experiment is a lower-contrast target with lots of distinct regions with different attenuation coefficient. The measurement setting is the same for both test cases: We shall use X-ray tomography with parallel beam geometry. The computational grid is 512×512, so we have 262,144-dimensional posterior distributions. We have additive white noise models, according to Equation (1), we fix the standard deviation of the noise to 1.5% of the maximum value of the line integrals. The parameters of each prior distribution at each measurement scenario are set using a grid search. For each prior, we select the parameter value, which gives approximately a minimum L^2-error between the MAP estimate and the ground truth. In the log tomography case, we use a different log slice without a high attenuation object to select the parameters. In real-life applications, we should select the parameters other means, such as setting a hyperprior distribution for the prior parameter—this will naturally introduce computational burden especially due to coupling of the unknown (Chada et al., 2019; Monterrubio-Gómez et al., 2020). For Besov prior, we use eight levels in the DWT.

In simulations, we shall use the four priors, and utilise L-BFGS, adaptive MwG, and HMC-NUTS. The initial points of MCMC chains are set to be the MAP estimate obtained with L-BFGS. As MCMC chain lengths, we use 500,000 samples in MwG to adapt proposal variances for each component and use those variances to generate 400,000 samples. To save memory, we thin the MwG chains by storing every 300th sample. For HMC-NUTS, we run 100 times for the dual averaging algorithm to tune the Störmer–Verlet symplectic integrator, and then we generate 4,000 samples for uncertainty quantification. For MCMC convergence analysis, we use Geweke z-score test for measuring stationarity of the chain and autocorrelation function and time τ for checking chain mixing (Geweke, 1992; Roy, 2020). We also convert the z-score into a two-sided p-value, for certain significance level comparisons. For good MCMC convergence, τ should be small and p-value close to 1.

Depending on the prior, and because in log tomography, we have a high-contrast target, the marginal and conditional posteriors turn out to be wide or even multimodal. The low-contrast drill-core tomography is in this sense an easier target, hence we quantify uncertainty only the log tomography as it is a more demanding test scenario. Thus for the drill-core tomography phantom, we only compute MAP estimators as the log tomography already showcases the capabilities of the uncertainty quantification.

4.1. Log tomography

Our test case is visualized in Figure 1, with metallic piece in black and knots in light color, and a small rotten part in near white with more smooth behavior. Besides the ground truth, we have also plotted the
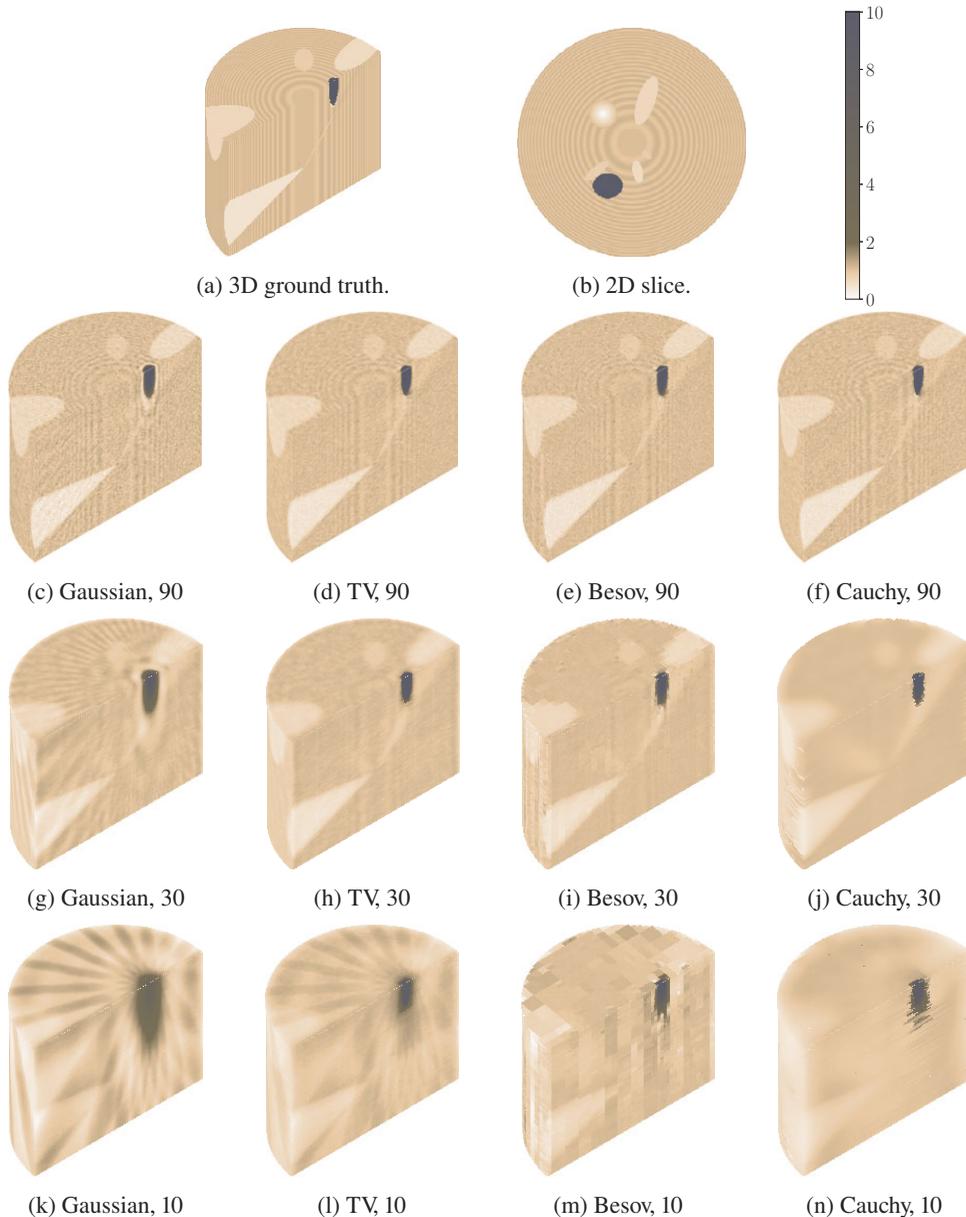
Figure 1. Log tomography: Three-dimensional and two-dimensional ground truths with knots (light material) and metallic piece (black). Three-dimensional reconstructions obtained by stacking two-dimensional maximum a posteriori (MAP) estimates with different angles and priors.

three-dimensional estimates, obtained by stacking two-dimensional MAP estimate slices computed with L-BFGS. We have all the four different priors, and the number of equispaced measurement angles is 10, 30, or 90. We have tabulated the peak signal-to-noise ratio (PSNR) of the reconstructions in Table 1. While all the methods seem to capture the main features in densest-data case, reducing the number of measurement causes severe artefacts with Gaussian and TV priors, and the Besov prior promotes blocky estimate. Cauchy prior captures the metallic piece even in the most difficult case, but it struggles with knots. The parameters used in the reconstructions are: For Gaussian prior, $\sigma = 10$ for all test cases; for TV, $\alpha = 0.72$ for the 10 angle case, 1.40 in 30 angle case, and 2.68 in 90 angle case; for Besov, $\alpha = 1.40$ in the

Table 1. PSNR values in decibels for the MAP and CM estimates with different priors in log tomography.

Estimator	Number of angles	Gaussian	TV	Besov	Cauchy
MAP	90	33.7	36.6	35.0	36.0
MAP	30	31.2	35.7	34.2	33.5
MAP	10	26.6	32.4	31.1	31.0
CM (MwG)	90	—	36.3	34.6	35.8
CM (MwG)	30	—	33.2	32.5	34.8
CM (HMC)	90	—	36.2	34.6	35.6
CM (HMC)	30	—	33.0	32.5	34.2

Abbreviations: CM, conditional mean; HMC, Hamiltonian Monte Carlo; MAP, maximum a posteriori; MwG, Metropolis-within-Gibbs; PSNR, peak signal-to-noise ratio; TV, total variation.

10 and 30 angle cases and 3.16 in the 90 angle case; for Cauchy, $(\lambda h)^2 = 0.0014$ in the 10 angle case, 0.0040 in the 30 angle case, and 0.0317 in the 90 angle case.

It is easier to see the effects of different priors in the two-dimensional slices in Figure 2. The densest scenario, with 90 angles, TV, and Cauchy priors produce very good MAP estimates recovering the edges. MAP estimate with Besov prior looks blocky. In the MAP estimate with Gaussian prior, the edges of the metallic object are smoothed. Reducing the number of measurement angles causes visible artefacts into MAP estimates with TV and Gaussian priors, but TV prior preserves the edges much better. Meanwhile, the rotation-dependency of Cauchy difference prior becomes apparent with coordinate-axis-aligned sharp artefacts. The Besov prior MAP estimate remains more or less consistent even if the number of angles is decreased. However, in practice, the block artefacts will spoil the MAP estimates and that is not useful in practical applications. We could decrease the number of wavelet transform levels in order to reduce blockiness or use another wavelet family.

In Figure 3, we have plotted the CM estimates computed with MwG and HMC. Here we skip Gaussian priors, as the CM and MAP estimates with Gaussian priors are the same. The PSNR values are tabulated in Table 1. It is notable that the CM estimate with TV prior differs noticeably from its MAP estimate (Figures 1 and 2). For high-dimensional inverse problems, it has been demonstrated by Lassas and Siltanen (2004, figure 4) that CM estimates might be wildly oscillatory, even though the same prior produces reasonable MAP estimates. Also, the great number of scanning artefacts is notable. On the other hand, CM estimates with Cauchy difference prior have less coordinate-axis-related artefacts, and CM estimates of Besov prior are less blocky than their corresponding MAP estimates.

In Figure 4, we have logarithmic pixel-wise variances estimated from MwG and HMC-NUTS MCMC chains. In Table 2, we have tabulated MCMC chain convergence statistics for two different pixel locations, and the autocorrelation functions are plotted in Figure 5. For pixel $(256, 256)$, MwG has good convergence in all cases. HMC performs otherwise well, with the exception of a poor autocorrelation for the 30-angle case with Cauchy prior. The Geweke test p-values are good in all cases. Pixel $(343, 237)$ is located at the edge of the high-attenuation object and the wood. For the 90-angle case with Cauchy prior, the marginal distribution of the posterior of that pixel is bimodal, and HMC cannot jump between the modes efficiently. This causes bad autocorrelation and p-values, that is, we have serious convergence issues. 30-angle case is better in the sense of convergence of the chain, as HMC does not explore the other mode at all. In general, MwG chains with Cauchy prior mix better and have better autocorrelation values than HMC chains. However, MwG may not find all the posterior modes. This is because some of the pixels at the boundary of the high-attenuation object have much higher variance than the other boundary pixels, as seen from the variance estimates.

The posteriors with Cauchy prior lead to multimodal and heavy-tailed distributions, and this induces mixing issues for both methods, but particularly for HMC as it is known to underperform at sampling from
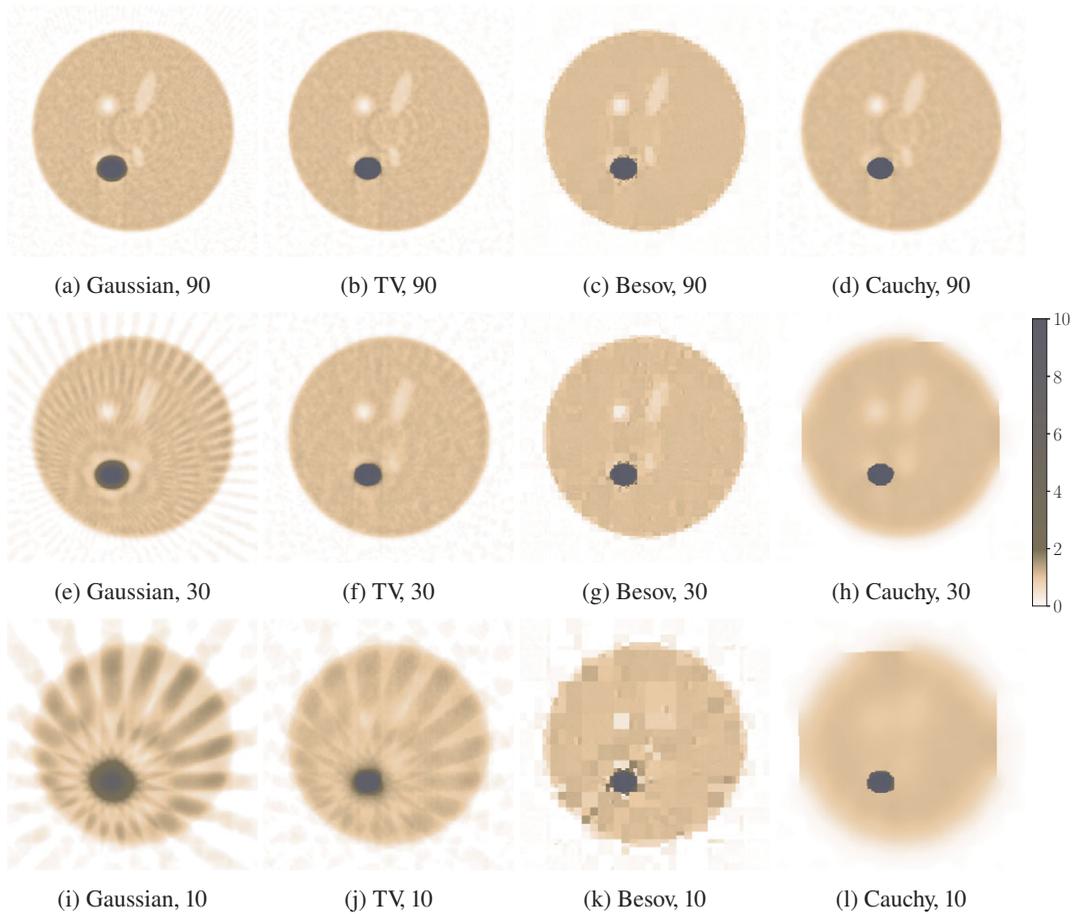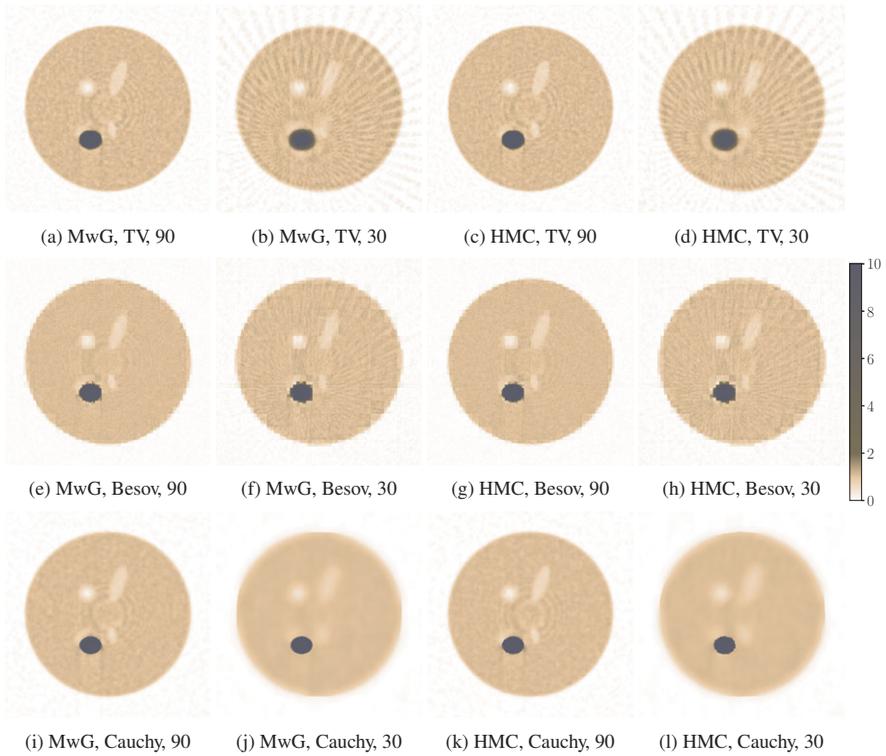
Figure 2. *Log tomography: Two-dimensional maximum a posteriori (MAP) estimates with different measurement angles and prior assumptions.*

such distributions (Livingstone et al., 2017; Mangoubi et al., 2018). HMC-NUTS is an almost plug-and-play alternative for MwG for all other cases, except Cauchy, for which special HMC methods should be considered. HMC chain mixing can be improved by adapting a global mass matrix for the momentum by running a preliminary run for posterior variance estimation. For our high-dimensional case, local mass matrix is computationally infeasible. Tempered HMC may also enhance the algorithm (Graham and Storkey, 2017). For MwG tuning, we have limited chances for improvement, but, that is, robust adaptive Metropolis-Hastings could provide some further enhancements (Vihola, 2011).
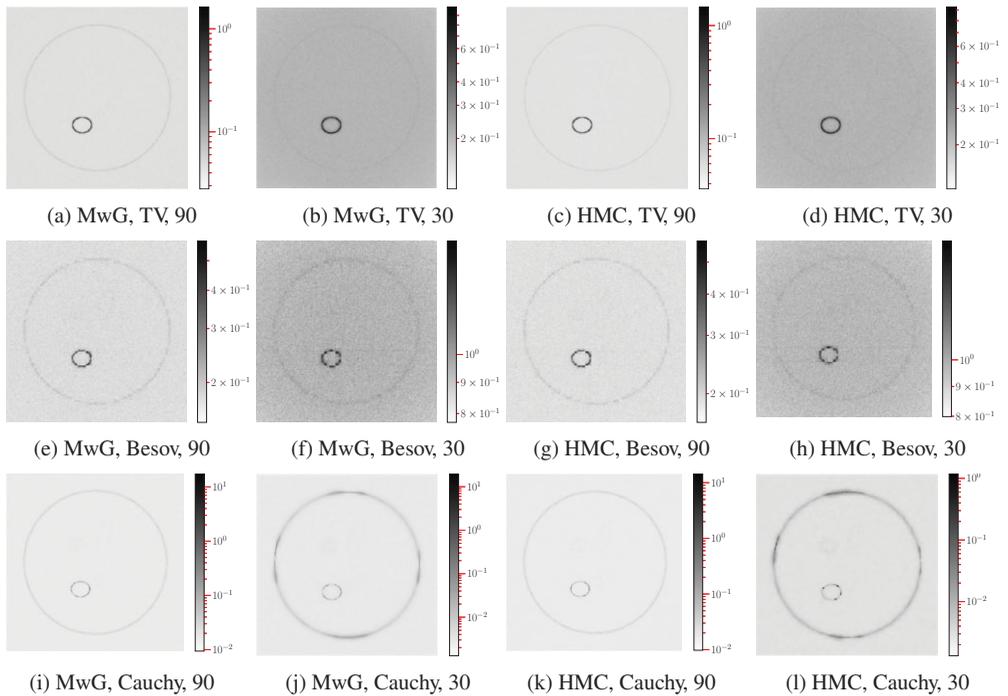
4.2. Drill-core tomography

Here we take a simpler setting for a lower-contrast target, and we detect pores and mineralized soil within a core sample. Unlike in the log tomography experiment, the objects do not have drastically different absorption coefficient levels. However, the drill-core sample has lots of pores and soil cobs and their sizes also vary.

We demonstrate only MAP estimation, and the results are in Figure 6. At the first glance, there are no significant differences within the MAP estimates if we compare TV, Cauchy, and Gaussian difference priors to each other at 90 measurement angle scenario. Nevertheless, we consider the Cauchy and TV priors as the best ones, since they have the best PSNR values and least amount of noise, and they preserve the edges of the pores with low-and high-attenuation regions of the core sample well. If we wanted to

(a) MwG, TV, 90　　(b) MwG, TV, 30　　(c) HMC, TV, 90　　(d) HMC, TV, 30

(e) MwG, Besov, 90　　(f) MwG, Besov, 30　　(g) HMC, Besov, 90　　(h) HMC, Besov, 30

(i) MwG, Cauchy, 90　　(j) MwG, Cauchy, 30　　(k) HMC, Cauchy, 90　　(l) HMC, Cauchy, 30

Figure 3. *Log tomography: Two-dimensional conditional mean (CM) estimates with different measurement angles, prior assumptions, and samplers.*

(a) MwG, TV, 90　　(b) MwG, TV, 30　　(c) HMC, TV, 90　　(d) HMC, TV, 30

(e) MwG, Besov, 90　　(f) MwG, Besov, 30　　(g) HMC, Besov, 90　　(h) HMC, Besov, 30

(i) MwG, Cauchy, 90　　(j) MwG, Cauchy, 30　　(k) HMC, Cauchy, 90　　(l) HMC, Cauchy, 30

Figure 4. *Log tomography pixel-wise variance estimates.*

Table 2. Log tomography chain convergence estimators for two pixels with different priors and varying number of angles.

| Estimator	Pixel	TV 90	TV 30	Besov 90	Besov 30	Cauchy 90	Cauchy 30
MwG, τ	(256,256)	1.246	1.140	1.091	1.495	1.271	1.385
MwG, Geweke p	(256,256)	0.958	0.995	0.977	0.992	0.996	0.999
HMC, τ	(256,256)	2.414	4.200	1.061	0.566	3.436	26.65
HMC, Geweke p	(256,256)	0.988	0.991	0.944	0.972	0.994	0.991
MwG, τ	(343,237)	1.046	1.082	1.605	1.102	1.545	5.229
MwG, Geweke p	(343,237)	0.958	0.998	0.976	0.999	0.925	0.884
HMC, τ	(343,237)	34.56	3.596	3.907	1.288	1371	35.67
HMC, Geweke p	(343,237)	0.973	0.995	0.985	0.970	0.066	0.962

Abbreviations: HMC, Hamiltonian Monte Carlo; MwG, Metropolis-within-Gibbs; TV, total variation.
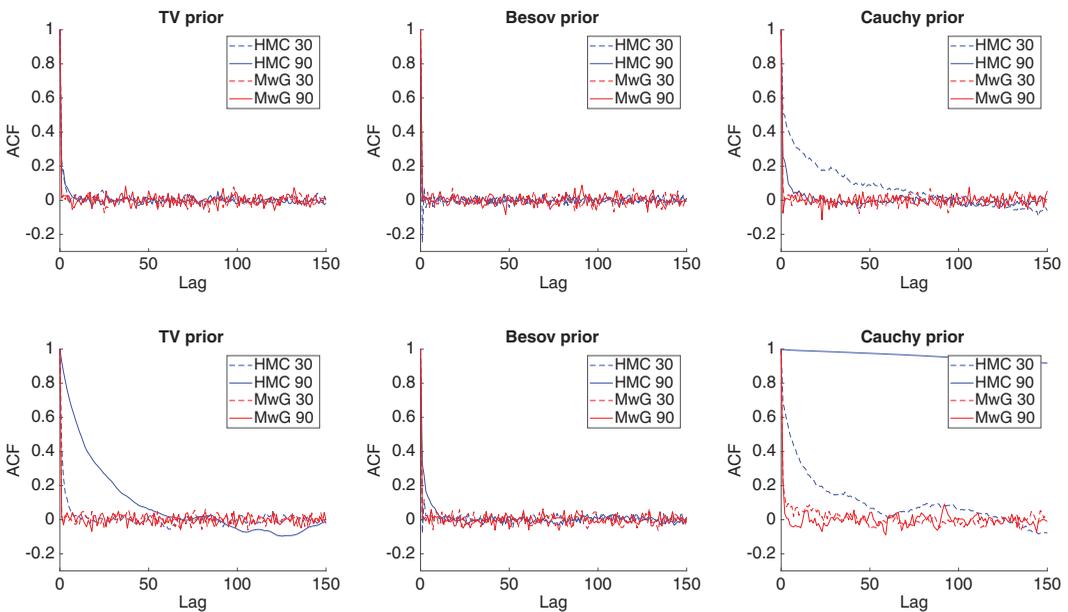
Figure 5. *Log tomography Markov chain Monte Carlo (MCMC) chain autocorrelation functions for pixels* (256,256) *(top row) and* (343,237) *(bottom row).*

achieve the same level of noise with Gaussian difference prior, we would end up having clearly over-smoothed MAP estimates. The PSNR values are in Table 3. Again, we use grid search to set the parameters: For Gaussian prior $\sigma = 0.37$ and Besov prior $\alpha = 0.61$ in all cases; for 10-angle case with TV prior $\alpha = 0.1$ and 0.37 in 30-angle case and 0.72 in 90-angle case; for 10-angle case with Cauchy prior $(\lambda h)^2 = 1.66$ and 0.18 in the 30-angle case and 0.55 in the 90-angle case.

In the 10-angle and 30-angle scenarios, MAP estimates with Besov prior are severely ruined with block artefacts—the phantom consists of many spherical objects which are close to each other, but the wavelet coefficients of Haar family do not approximate well such details. Unlike in the log tomography, there are not so many artefacts present in the MAP estimates with Gaussian prior. In the 10-angle case, MAP estimates calculated with Gaussian, TV, and Cauchy priors are very close to each other, and neither of them can be considered superior over the others. We could try to fix the coordinate-axis dependency of
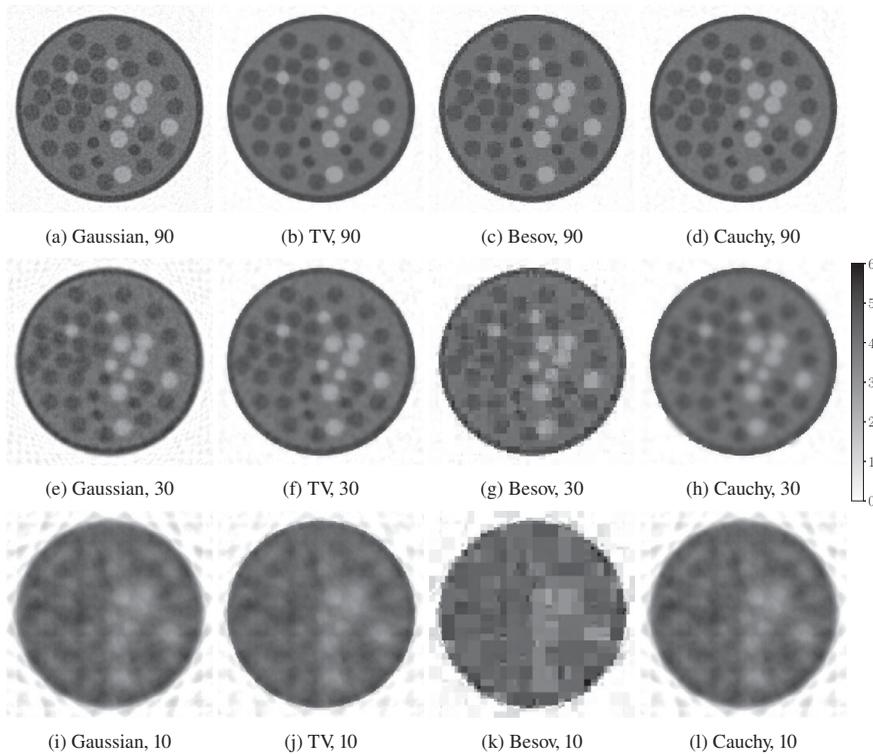
Figure 6. *Drill-core maximum a posteriori (MAP) estimates for the with different measurement angles and prior assumptions.*

Table 3. Drill-core tomography PSNR values in decibels for the MAP estimates with different priors.

No. of angles/Prior	Gaussian	TV	Besov	Cauchy
90	25.0	26.8	24.3	26.0
30	23.2	25.6	22.0	23.6
10	18.4	20.8	17.5	18.5

Abbreviations: MAP, maximum a posteriori; PSNR, peak signal-to-noise ratio; TV, total variation.

Cauchy prior by introducing isotropic Cauchy prior so that we set a bivariate Cauchy distribution for differences of each pixel to their neighbors.

5. Conclusion

We have made a Bayesian inversion industrial tomography comparison study with four different priors. The targets have different contrasts, that is, sharp edges were to be reconstructed. This is a classical problem, where one needs either hierarchical or non-Gaussian models, and hence inherently computational complexity becomes a problem. The chosen methods, to use optimization or MCMC, is a balance between fast computations and uncertainty quantification. It should be noted that for many industrial applications, the computational cost is still too high, like for many dynamical systems. However, for static objects and case studies where computational time is not critical, the correctly chosen methods can significantly enhance the reconstruction results for difficult measurement geometries and targets.

Systematic choice of different parameters, including mesh refinement parameters, dictate the reconstruction accuracy. In order to show the maximum possible reconstruction capabilities, we have used simple grid searches. In reality, parameter choice requires implementing some kind of a hierarchical model. This would naturally further increase computational complexity.

In all the scenarios considered, MAP estimates with Cauchy and TV priors were the best—their artefacts were mild in general and edge-preserving performance excellent. The estimates using Besov prior with Haar wavelets underperformed compared to the other priors in most cases due to the major block artefacts it tends to produce. We could decrease the block artefacts by reducing the number of DWT levels or using another wavelet family but even that cannot fix the location dependency of the prior. Likewise, we could try replacing the Cauchy prior with a bivariate Cauchy distribution to make it isotropic. For further development, applying acceptance-ratio-based adaptation in MwG and using tempered HMC would improve sampling. Other prior distributions, like general α-stable priors and using hierarchical priors with possibly many layers might potentially improve the edge-preserving properties. A future avenue is also in reducing the instability by combining non-Gaussian priors to the machine-learning-based priors Adler and Öktem (2018), hence possibly providing more robust methods against noise and difficult measurement geometries.

Competing Interests. The authors declare no competing interests exist.

Data Availability Statement. Codes and the synthetic data used are available openly in GitHub https://github.com/suurj/tomo/.

Authorship Contributions. Conseptualization, S.L., S.S., and L.R.; Funding acquisition, S.S. and L.R.; Methodology, J.S., M.E., S.L., S.S., and L.R.; Data visualization, J.S.; Software, J.S. and M.E.; Original draft, J.S., M.E., S.L., and L.R. All authors approved the final submitted draft.

Ethical Standards. We have followed the ethical code of conduct for research and publication of the Academy of Finland.

References

Adler J and Öktem O (2018) *Deep Bayesian Inversion*. Available at arXiv:1811.05910.

Bardsley J (2012) MCMC-based image reconstruction with uncertainty quantification. *SIAM Journal on Scientific Computing 34* (3), A1316–A1332.

Bardsley J (2013) Gaussian Markov random field priors for inverse problems. *Inverse Problems and Imaging 7*, 397–416.

Barp A, Briol F-X, Kennedy AD and Girolami M (2018) Geometry and dynamics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application 5*(1), 451–471.

Berlinet A and Thomas-Agnan C (2004) *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Boston, MA: Kluwer Academic Publishers.

Beskos A, Girolami M, Lan S, Farrell PE and Stuart AM (2017) Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics 335*, 327–351.

Beskos A, Pillai N, Roberts G, Sanz-Serna J-M and Stuart A (2013) Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli 19*(5A), 1501–1534.

Beskos A, Pinski FJ Sanz-Serna JM and Stuart A (2011) Hybrid Monte Carlo on Hilbert spaces. *Stochastic Processes and their Applications 121*(10), 2201–2230.

Betancourt M (2017) *A Conceptual Introduction to Hamiltonian Monte Carlo*. Avavilable at arXiv:1701.02434.

Bolin D (2014) Spatial Matérn fields driven by non-Gaussian noise. *Scandinavian Journal of Statistics 41*, 557–579.

Brooks S, Gelman A, Jones G and Meng X-L (2011) *Handbook of Markov Chain Monte Carlo*. CRC press.

Chada NK, Lasanen S and Roininen L (2019) *Posterior Convergence Analysis of α-Stable Sheets*. Avaialble at arXiv: 1907.03086.

Charbonnier P, Blanc-Feraud L, Aubert G and Barlaud M (1997) Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing 6*(2), 298–311.

Cotter SL, Roberts GO, Stuart AM and White D (2013) MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science 28*(3), 424–446.

Dunlop MM, **Girolami MA**, **Stuart AM and Teckentrup AL** (2018) How deep are deep Gaussian processes? *Journal of Machine Learning Research 19*(54), 1–46.

Dunlop MM, **Iglesias MA and Stuart AM** (2017) Hierarchical Bayesian level set inversion. *Statistics and Computing 27*, 1555–1584.

Emzir M, **Lasanen S**, **Purisha Z**, **Roininen L and Särkkä S** (2020) *Non-stationary Multi-Layered Gaussian Priors for Bayesian Inversion*. Avaialble at arXiv:2006.15634.

Ernst JR, **Green AG**, **Maurer H and Holliger K** (2007) Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data. *Geophysics 72*(5), J53–J64.

Garcea S, **Wang Y and Withers P** (2018) X-ray computed tomography of polymer composites. *Composites Science and Technology 156*, 305–319.

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics*. University Press, pp. 169–193.

Girolami M and Calderhead B (2011) Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(2), 123–214.

González G (2017) *Edge-Promoting Priors in Electrical Impedance Tomography*. PhD thesis, University of Eastern Finland.

González G, **Kolehmainen V and Seppänen A** (2017) Isotropic and anisotropic total variation regularization in electrical impedance tomography. *Computers & Mathematics with Applications 74*, 564–576.

Graham MM and Storkey AJ (2017) *Continuously Tempered Hamiltonian Monte Carlo*. In *UAI*.

Haario H, **Saksman E and Tamminen J** (2005) Componentwise adaptation for high dimensional MCMC. *Computational Statistics 20*, 265–273.

Hoffman MD and Gelman A (2014) The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research 15*, 1593–1623.

Hosseini B (2017) Well-posed Bayesian inverse problems with infinitely divisible and heavy-tailed prior measures. *SIAM/ASA Journal on Uncertainty Quantification 5*, 1024–1060.

Kaipio J and Somersalo E (2005) *Statistical and Computational Inverse Problems*. Dordrecht: Springer.

Kolehmainen V, **Vanne A**, **Siltanen S**, **Järvenpää S**, **Kaipio J**, **Lassas M and Kalke M** (2006) Parallelized Bayesian inversion for three-dimensional dental X-ray imaging. *IEEE Transactions on Medical Imaging 25*(2), 218–228.

Kruth J-P, **Bartscher M**, **Carmignato S**, **Schmitt R**, **Chiffre L and Weckenmann A** (2011) Computed tomography for dimensional metrology. *Cirp Annals-manufacturing Technology 60*, 821–842.

Lassas M, **Saksman E and Siltanen S** (2009) Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Problems and Imaging 3*(1), 87–122.

Lassas M and Siltanen S (2004) Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Problems 20*(5), 1537–1563.

Law KJH (2014) Proposals which speed up function-space MCMC. *Journal of Computational and Applied Mathematics 262*, 127–138.

Liu J, **Huang T-Z**, **Seselnick IW**, **Lv X-G and Chen P-Y** (2015) Image restoration using total variation with overlapping group sparsity. *Information Sciences 295*, 232–246.

Livingstone S, **Faulkner MF and Roberts GO** (2017) *Kinetic Energy Choice in Hamiltonian/Hybrid Monte Carlo*. Available at arXiv:1706.02649.

Luengo D, **Martino L**, **Bugallo M**, **Elvira V and Särkkä S** (2020) A survey of Monte Carlo methods for parameter estimation. *EURASIP Journal on Advances in Signal Processing 2020*, 25.

Mangoubi O, **Pillai NS and Smith A** (2018) *Does Hamiltonian Monte Carlo Mix Faster than a Random Walk on Multimodal Densities?* Avaialble at arXiv:1808.03230.

Markkanen M, **Roininen L**, **Huttunen JMJ and Lasanen S** (2019) Cauchy difference priors for edge-preserving Bayesian inversion. *Journal of Inverse and Ill-posed Problems 27*(2), 225–240.

Mendoza A, **Roininen L**, **Girolami M**, **Heikkinen J and Haario H** (2019) Statistical methods to enable practical on-site tomographic imaging of whole-core samples. *Geophysics 84*(3), D89–D100.

Meyer Y (1992) Wavelets and operators. In *Cambridge Studies in Advanced Mathematics*, Vol *37*. Cambridge: Cambridge University Press.

Monterrubio-Gómez K, **Roininen L**, **Wade S**, **Damoulas T and Girolami M** (2020) Posterior inference for sparse hierarchical non-stationary models. *Computational Statistics & Data Analysis 148*, 106954.

Muhumuza K, **Roininen L**, **Huttunen JMJ and Lähivaara T** (2020) A Bayesian-based approach to improving acoustic Born waveform inversion of seismic data for viscoelastic media. *Inverse Problems 36*(7), 075010.

Natterer F. (2001). *The Mathematics of Computerized Tomography (Classics in Applied Mathematics)*. SIAM: Society for Industrial and Applied Mathematics.

Neal R. (2011) MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*, Chapter 5. CRC Press, pp. 113–162.

Niinimäki K (2013) *Computational Optimization Methods for Large-scale Inverse Problems*. PhD thesis, University of Eastern Finland.

Ou X, **Zhang X**, **Lowe T**, **Blanc R**, **Rad MN**, **Wang Y**, **Batail N**, **Pham C**, **Shokri N**, **Garforth AA**, **Withers PJ and Fan X** (2017) X-ray micro computed tomography characterization of cellular sic foams for their applications in chemical engineering. *Materials Characterization 123*, 20–28.

Piao Z-y, **Zhou Z-y**, **Xu J and Wang H-d** (2019) Use of X-ray computed tomography to investigate rolling contact cracks in plasma sprayed Fe-Cr-B-Si coating. *Tribology Letters 67*(1), 11.

Raftery A, **Tanner M and Wells M** (2001) *Statistics in the 21st Century. Chapman & Hall/CRC Monographs on Statistics & Applied Probability*. CRC Press.

Roininen L, **Girolami M**, **Lasanen S, and Markkanen M** (2019) Hyperpriors for Matérn fields with applications in Bayesian inversion. *Inverse Problems and Imaging 13*(1), 1–29.

Rotella A, **Nadot Y**, **Piellard M**, **Augustin R and Fleuriot M** (2018) Fatigue limit of a cast al-si-mg alloy (a357-t6) with natural casting shrinkages using astm standard x-ray inspection. *International Journal of Fatigue 114*, 177–188.

Roy V (2020) Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application 7*(1), 387–412.

Samorodnitsky G and Taqqu MS (1994) *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance*. New York: Chapman & Hall.

Shieh C-C, **Caillet V**, **Dunbar M**, **Keall PJ**, **Booth JT**, **Hardcastle N**, **Haddad C**, **Eade T and Feain IJ** (2017) A Bayesian approach for three-dimensional markerless tumor tracking using kV imaging during lung radiotherapy. *Physics in Medicine and Biology 62*(8), 3065–3080.

Shustrov D, **Eerola T**, **Lensu L**, **Kälviäinen H and Haario H** (2019) Fine-grained wood species identification using convolutional neural networks. In *Image Analysis*, Springer International Publishing, pp. 67–77.

Siltanen S, **Kolehmainen V**, **Järvenpää S**, **Kaipio J**, **Koistinen P**, **Lassas M**, **Pirttilä J and Somersalo E** (2003) Statistical inversion for medical X-ray tomography with few radiographs: I. General theory. *Physics in Medicine and Biology 48*(10), 1465–1490.

Sullivan TJ (2017) Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors. *Inverse Problems and Imaging 11*(5), 857–874.

Thanh D, **Prasath S and Le Minh H** (2019) A review on CT and X-ray images denoising methods. *Informatica 43*, 151–159.

Toft P (1996) *The Radon Transform — Theory and Implementation*. PhD thesis, Technical University of Denmark.

Tzikas D, **Likas A and Galatsanos N** (2008) The variational approximation for Bayesian inference life after the EM algorithm. *Signal Processing Magazine, IEEE 25*, 131–146.

Vihola M (2011) Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing 22*(5), 997–1008.

Villarraga-Gómez H, **Herazo EL and Smith ST** (2019) X-ray computed tomography: from medical imaging to dimensional metrology. *Precision Engineering 60*, 544–569.

Wang H and Vieira J (2010) 2-D wavelet transforms in the form of matrices and application in compressed sensing. In *2010 8th World Congress on Intelligent Control and Automation*, pp. 35–39.

Whiting BR (2002) Signal statistics in x-ray computed tomography. In *Medical Imaging 2002: Physics of Medical Imaging*, Vol *4682*. International Society for Optics and Photonics, pp. 53–60.

Zolotarev F, **Eerola T**, **Lensu L**, **Kälviäinen H**, **Haario H**, **Heikkinen J and Kauppi T** (2019) Timber tracing with multimodal encoder-decoder networks. In *Computer Analysis of Images and Patterns*, Springer International Publishing, pp. 342–353.