NONLOCAL TV-GAUSSIAN PRIOR FOR BAYESIAN INVERSE PROBLEMS WITH APPLICATIONS TO LIMITED CT RECONSTRUCTION

DIDI LV AND QINGPING ZHOU
School of Mathematical Sciences and Institute of Natural Sciences
Shanghai Jiao Tong University, Shanghai 200240, China

JAE KYU CHOI
School of Mathematical Sciences, Tongji University
Shanghai 200082, China

JINGLAI LI
Department of Mathematical Sciences, University of Liverpool
Liverpool L69 6ZL, United Kingdom

XIAOQUN ZHANG*
School of Mathematical Sciences, MOE-LSC, and Institute of Natural Sciences
Shanghai Jiao Tong University
Shanghai 200240, China

(Communicated by Bin Dong)

ABSTRACT. Bayesian inference methods have been widely applied in inverse problems due to the ability of uncertainty characterization of the estimation. The prior distribution of the unknown plays an essential role in the Bayesian inference, and a good prior distribution can significantly improve the inference results. In this paper, we propose a hybrid prior distribution on combining the nonlocal total variation regularization (NLTV) and the Gaussian distribution, namely NLTG prior. The advantage of this hybrid prior is two-fold. The proposed prior models both texture and geometric structures present in images through the NLTV. The Gaussian reference measure also provides a flexibility of incorporating structure information from a reference image. Some theoretical properties are established for the hybrid prior. We apply the proposed prior to limited tomography reconstruction problem that is difficult due to severe data missing. Both maximum a posteriori and conditional mean estimates are computed through two efficient methods and the numerical experiments validate the advantages and feasibility of the proposed NLTG prior.

2010 Mathematics Subject Classification. Primary: 58F15, 58F17; Secondary: 53C35.

Key words and phrases. Bayesian inverse problems, limited tomography, nonlocal total variation, Gaussian measure, uncertainty quantification.

The work is supported by NSFC grants (No. 11771288, 11301337, 91630311) and National key research and development program No. 2017YFB0202902.

* Corresponding author: Xiaoqun Zhang.
1. Introduction. Bayesian inference methods [12, 17] have been a popular tool for inverse problems. Such popularity is mainly due to its ability to quantify the solution uncertainties. A typical Bayesian treatment consists of assigning a prior distribution to the unknown data and then update the distribution based on the observed data, yielding the posterior distribution. It is clear that, as most practical inverse problems are highly ill-posed, the performance of the Bayesian inference depends critically on the choice of the prior distribution. Recently considerable attentions have been paid to the studies of infinite dimensional Bayesian inverse problems, where the unknowns are functions of space or time, for example, medical images. For infinite dimensional problems, the Gaussian measures are arguably the most popular choice of prior distributions, as it has many theoretical and computational advantages in the infinite dimensional setting [37].

However, in many imaging problems, the images that one wants to recover are often subject to sharp jumps or discontinuities while the Gaussian prior distributions are not suitable for modeling such functions [43]. To this end several non-Gaussian priors have been proposed to model images, e.g., [39]. Since these prior distributions differ significantly from Gaussian, many sampling schemes based on the Gaussian prior can not be used directly. To address the issue, a hybrid prior motivated by the total variation (TV) regularization [33] in the deterministic setting was proposed in [43]; It has been proven in [21] that the TV based prior does not converge to a well-defined infinite-dimensional measure as the discretization dimension increases. The hybrid prior is a combination of the TV term and the Gaussian distribution: it uses a TV term to capture the sharp jumps in the functions and a Gaussian distribution as a reference measure to make sure that the resulting prior does converge to a well-defined probability measure in the function space in the infinite dimensional limit.

In this paper, we consider a typical imaging inverse problem: X-Ray computerized tomography (CT) problem with limited data. X-ray computed tomography plays an important role in clinical diseases diagnosis. Let $u$ denote the image to be reconstructed. Throughout this paper, we assume $u \in L_2(\mathbb{R}^2)$ is supported in a domain $\Omega$, and we only consider the two dimensional parallel beam CT for simplicity. Then the sinogram (or the projection data) $f$ is obtained by the following Radon transform [32]:

\begin{equation}
y(\alpha, s) = Pu(\alpha, s) = \int_{-\infty}^{\infty} u(sn + z^\perp)dz, \quad (\alpha, s) \in [0, 2\pi) \times \mathbb{R}
\end{equation}

where $n = (\cos \alpha, \sin \alpha)$ and $n^\perp = (-\sin \alpha, \cos \alpha)$. Given a sinogram $y = Pu$ on $[0, \pi) \times \mathbb{R}$, the tomographic image $u$ is reconstructed via the inverse Radon transform [1, 27]:

\begin{equation}
u(x) = \frac{1}{2\pi^2} \int_0^\pi \int_{-\infty}^{\infty} \frac{1}{x \cdot n - s} \left[ \frac{\partial}{\partial s} y(\alpha, s) \right] ds d\alpha
\end{equation}

where $x = (x_1, x_2) \in \mathbb{R}^2$.

The problem (1) becomes ill-posed whenever the limited data is available in the subset of $[0, \pi) \times \mathbb{R}$ due to the reduced size of detector [40, 41] and/or the reduced number of projections [28, 36, 5]. In particular, if the projection data $f$ is available on $S_r \in [0, \pi) \times \mathbb{R}$:

\begin{equation}
S_r = \{ (\alpha, s) \in [0, \pi) \times \mathbb{R} : |s| \leq r \},
\end{equation}
then there exists a nontrivial function \( g \) called the ambiguity of \( P \), i.e. \( Pg = 0 \) in \( S_r \) \[40\]. As this ambiguity \( g \) is nonconstant in the region of interest (ROI) \( B(0, r) \) \[40, 42\], the reconstructed image via (2) using available \( g \) only on \( S_r \) will be deteriorated by \( g \). In the literature, numerous studies have been proposed to remove the ambiguity \( g \) due to the restriction of \( y \) on \( S_r \). These studies can be classified into the following two categories: the known subregion based approaches related to the restoration of signal from the truncated Hilbert transform, and the sparsity model based approaches, such as total variation regularization. See \[34, 8, 16\] for the detailed surveys.

In the literature, nonlocal regularization has been popular for imaging inverse problems, as an improvement of total variation regularization. They are originally proposed for natural image processing to restore repetitive patterns and textures, for example a heuristic copy-paste technique was firstly proposed for texture synthesis in \[9\], a more systematic nonlocal means filter was proposed in \[3\] and a nonlocal variational framework was established in \[14\]. The main idea of the nonlocal methods is to utilize the similarities present in an image as a weight for restoring, smoothing or regularization. As an extension of TV, nonlocal total variation (NLTV) regularization method is among those popular variational regularization tools due to its flexibility of recovering both texture and geometry patterns for diverse imaging inverse problems, see \[44, 45, 30, 23\] for the applications. It has been demonstrated that in many practical problems, the NLTV method has better performance than the standard TV, especially for recovering textures and structures of images. More definitions and details are present in Section 2.2. NLTV has been also applied in the context of CT reconstruction, for example in \[44, 24, 18\].

Inspired by the success of nonlocal regularization, we propose to improve the hybrid TV-Gaussian (TG) prior proposed in \[43\] for image reconstruction by incorporating the nonlocal methods. We shall replace the TV term in the hybrid prior with a NLTV term, and theoretically we prove that the resulting new hybrid prior can also lead to a well defined posterior distribution in the infinite dimensional setting. To demonstrate the effectiveness of the hybrid NLTV-Gaussian (NLTG) prior, we apply it to the limited tomography problems, where only limited projection data are available. In particular, we consider the two common types of point estimation in the Bayesian framework: maximum a posterior (MAP) and conditional mean (CM) with the NLTG prior. The MAP estimate consists of solving an optimization problem, while the computing of CM involves evaluating a high-dimensional integration problem \[17, 26\]. Our numerical results show that the NLTG prior defined with a reference image can greatly improve the reconstruction quality compared to other conventional method. Furthermore, we also provide the quantification of the uncertainty of the CM estimator, which is consistent with the noise level. The advantages of our model are the following: The NLTV term can better recover textures and structures of images, especially for highly ill-posed or severely data-missing inverse problems such as the limited tomography reconstruction in this paper; Theoretically, the function space is extended to a larger function space compared to \( H^1 \) considered in TG prior, for dealing with larger images class; Finally, the Gaussian measure provides some freedom to incorporate other prior information through the covariance matrix, such as structures from a reference image, which is particularly usefully for limited CT problem.

The remainder of this paper is structured as follows: In Section 2, we describe the proposed NLTG priors on a Hilbert space and present the related theoretical
properties. In Section 3 we apply the proposed NLTG prior to the limited tomography problem and present the MAP and CM estimators. The numerical results of both MAP and CM estimates are shown in section 4. Finally, we draw our conclusions in the last section.

2. The NLTG priors.

2.1. The Bayesian framework and the TG prior. We first give a brief introduction to the basic setup of the Bayesian inference methods for inverse problems. We consider the forward model of the following generic form:

\[ y = F(u) + n, \]

where \( u \) is an unknown function (in this work we shall restrict ourselves to the situation where \( u \) is a real-valued function defined in \( \mathbb{R}^2 \), i.e., an image), \( y \in \mathbb{R}^m \) is the measured data and \( n \) is a \( m \)-dimensional zero mean Gaussian noise with covariance matrix \( \Sigma_0 \). Our goal here is to estimate the unknown function \( u \) from the measured data \( y \).

First we assume that the unknown \( u \) lies in a Hilbert space of functions, say \( X \). We then choose a probabilistic measure defined on \( X \), denoted by \( \mu_{pr} \), to be the prior measure of \( u \). The posterior measure of \( u \), denoted as \( \mu_y \), is represented as the Radon-Nikodym (R-N) derivative with respect to \( \mu_{pr} \):

\[ \frac{d\mu_y}{d\mu_{pr}} \propto \exp \left( -\Phi(u) \right), \]

where \( \Phi(u) = \frac{1}{2} \| F(u) - y \|_{\Sigma_0}^2 = \frac{1}{2} \langle \Sigma_0^{-\frac{1}{2}} (F(u) - y), \Sigma_0^{-\frac{1}{2}} (F(u) - y) \rangle \) is the data fidelity term in deterministic inverse problem. In the Bayesian framework, the posterior distribution depends on the information from data and the prior knowledge represented by the prior distribution, and so the choice of the prior distribution plays an essential role in the Bayesian method.

As is discussed in Section 1, probably the most popular prior in the infinite dimensional setting is the Gaussian measure, i.e. \( \mu_{pr} = \mu_0 = N(0, C_0) \), a Gaussian measure defined on \( X \) with zero mean and covariance operator \( C_0 \). Note that \( C_0 \) is symmetric positive operator of trace class [22]. To better model functions with sharp jumps, the hybrid prior in [43] takes the form

\[ \frac{d\mu_{pr}}{d\mu_0} \propto \exp \left( -R(u) \right), \]

where \( R(u) \) represents additional prior information (or regularization) on \( u \). In this case, one can write the R-N derivative of \( \mu_y \) with respect to \( \mu_0 \):

\[ \frac{d\mu_y}{d\mu_0} \propto \exp \left( -\Phi(u) - R(u) \right), \]

which in turn returns to the conventional formulation with Gaussian priors. Specifically, TG prior chooses the state space \( X \) to be Sobolev space \( H_1 \) and \( R(u) = \lambda \| u \|_{TV} = \lambda \int \| \nabla u \|_2 dx \), as introduced in [43, Section 2.3].

2.2. Nonlocal total variation. Here we provide the formulation of the NLTG prior and we start with the notation for nonlocal regularization. One may refer to [14, 13, 46, 19] for the details of the variational framework based nonlocal operators and [25, 44, 45, 29] for the applications to inverse problems.
Let $\Omega \subseteq \mathbb{R}^2$ be a bounded set, and $x \in \Omega$. Given a reference image $f : \Omega \to \mathbb{R}$, we can define a nonnegative symmetric weight function $\omega : \Omega \times \Omega \to \mathbb{R}$ as follows:

$$\omega(x, y) = \exp \left\{ -\frac{\langle G_a, |f(x + \cdot) - f(y + \cdot)|^2 \rangle}{h^2} \right\}$$

where $G_a$ is a Gaussian kernel with the standard deviation $a$, $h$ is a filtering parameter and $\langle \cdot, \cdot \rangle$ is the standard inner product in $L_2(\Omega)$. Note that in general, $h$ corresponds to the noise level; conventionally we set it to be the standard deviation of the noise [25].

Let $u : \Omega \to \mathbb{R}$. Using the weight function $\omega : \Omega \times \Omega \to \mathbb{R}$ in (8), we define the nonlocal (NL) gradient $\nabla_{\omega} u : \Omega \times \Omega \to \mathbb{R}$ as

$$\nabla_{\omega} u(x, y) = (u(y) - u(x)) \sqrt{\omega(x, y)}.$$ 

For a given $p : \Omega \times \Omega \to \mathbb{R}$, its NL divergence is defined by the standard adjoint relation with the NL gradient operator as follows:

$$\langle \nabla_{\omega} u, p \rangle = -\langle u, \text{div}_{\omega} p \rangle,$$

which leads to the following explicit formula:

$$\text{div}_{\omega} p(x) = \int_{\Omega} (p(y, x) - p(x, y)) \sqrt{\omega(x, y)} \, dy.$$

Now, we design the following functional based on the nonlocal operators:

$$J_{\text{NLTV}}(u) = \int_{\Omega} \|\nabla_{\omega} u(x, \cdot)\|_{L_2(\Omega)} \, dx = \int_{\Omega} \left( \int_{\Omega} |u(x) - u(y)|^2 \omega(x, y) \, dy \right)^{1/2} \, dx.$$

We can see that the functional in (11) is analogous to the total variation (TV) seminorm, and the NLTG prior can be constructed similarly. We first need to specify the state space $X$, which should be desirably a separable Hilbert space. Recall that $\Omega$ is a bounded set in $\mathbb{R}^2$. As convention, we choose $\Omega = [0, 1]^2$ throughout this paper. (Note, however, that our analysis is valid for any bounded domain with $C^1$ boundary). For a given reference image $f : \Omega \to \mathbb{R}$ and a weight function $\omega : \Omega \times \Omega \to \mathbb{R}$ defined as (8), we introduce

$$(X = \{ u \in L_2(\Omega) : \nabla_{\omega} u \in L_2(\Omega \times \Omega) \},$$

where the NL gradient $\nabla_{\omega} u$ is defined as (9). Obviously, we have $X \subseteq L_2(\Omega)$.

In addition, we present the following lemma which in turn shows that the NLTV functional defined as (11) is well defined on $L_2(\Omega)$.

**Lemma 2.1.** Let $f \in L_2(\Omega)$ be a given reference image, and let $\omega : \Omega \times \Omega \to \mathbb{R}$ be a weight function defined as (8). For $u \in L_2(\Omega)$, we have

$$\|\nabla_{\omega} u\|_{L_2(\Omega \times \Omega)} \leq 2 \|u\|_{L_2(\Omega)}$$

and

$$J_{\text{NLTV}}(u) \leq 2 \|u\|_{L_2(\Omega)}.$$ 

As a consequence, $\nabla_{\omega} : L_2(\Omega) \to L_2(\Omega \times \Omega)$ is a continuous linear operator.

**Proof.** First we note that, from the definition (8) of $\omega$, we have

$$\omega(x, y) \in (0, 1], \forall x, y \in \Omega.$$ 

**Inverse Problems and Imaging** Volume 14, No. 1 (2020), 117–132
Hence, using Hölder’s inequality (e.g. [11]), we have
\[
\|\nabla \omega u\|_{L^2(\Omega \times \Omega)}^2 = \int_\Omega \int_\Omega |\nabla \omega u(x, y)|^2 \, dy \, dx \\
= \int_\Omega \int_\Omega |u(x) - u(y)|^2 \omega(x, y) \, dy \, dx \\
\leq \int_\Omega \int_\Omega |u(x) - u(y)|^2 \omega(x, y) \, dy \, dx \\
\leq \int_\Omega \int_\Omega |u(y)|^2 \, dy \, dx + \int_\Omega \int_\Omega |u(x)|^2 \, dy \, dx \\
+ 2 \int_\Omega \int_\Omega |u(x)u(y)| \, dy \, dx \\
\leq 2 \|u\|_{L^2(\Omega)}^2 + 2 \|u\|_{L^2(\Omega)}^2 = 4 \|u\|_{L^2(\Omega)}^2,
\]
where we used the fact that $|\Omega| = 1$. This concludes (13).

For (14), we first note that for each $x \in \Omega$, the mapping
\[
(15) \quad x \mapsto \|\nabla \omega u(x, \cdot)\|_{L^2(\Omega)}
\]
is in $L^2(\Omega, dx)$, by (13), where we specified the Lebesgue measure $dx$ for clarity. Then applying the Hölder’s inequality again, we have
\[
J_{NLTV}(u) \leq \left( \int_\Omega \|\nabla \omega u(x, \cdot)\|_{L^2(\Omega)}^2 \, dx \right)^{\frac{1}{2}} = \|\nabla \omega u\|_{L^2(\Omega \times \Omega)} \leq 2 \|u\|_{L^2(\Omega)},
\]
and this concludes Lemma 2.1.

Combining the definition (12) and Lemma 2.1, we can see that $\mathcal{X} = L^2(\Omega)$. Now by taking
\[
(16) \quad R(u) = \lambda J_{NLTV}(u),
\]
in (7), we obtain the NLTG prior distribution on $L^2(\Omega)$.

2.3. **Theoretical properties of the NLTG prior.** In this section, we show that the NLTG prior leads to a well-behaved posterior distribution in $\mathcal{X}$, where the proofs follow the similar line as [43].

Following [37], we assume that the forward operator $F : L^2(\Omega) \to \mathbb{R}^m$ satisfies the following conditions:

A1. For every $\epsilon > 0$, there exists $M = M(\epsilon) > 0$ such that
\[
\|F(u)\|_{\Sigma_0^{-1}} \leq \exp \left( \epsilon \|u\|_{L^2(\Omega)}^2 + M \right), \quad \forall u \in L^2(\Omega).
\]

A2. For every $\delta > 0$, there exists $L = L(\delta) > 0$ such that for all $u_1, u_2 \in L^2(\Omega)$ with
\[
\max \left\{ \|u_1\|_{L^2(\Omega)}, \|u_2\|_{L^2(\Omega)} \right\} \leq \delta,
\]
we have
\[
\|F(u_1) - F(u_2)\|_{\Sigma_0^{-1}} \leq L \|u_1 - u_2\|_{L^2(\Omega)}.
\]

We have the following proposition on $\mu_{pr}$ in (6).

**Proposition 1.** Let $R : L^2(\Omega) \to \mathbb{R}$ be defined as (16). Then we have the followings:
1. For all \( u \in L_2(\Omega) \), we have \( R(u) \geq 0 \).
2. For every \( \delta > 0 \), there exists \( M = M(\delta) > 0 \) such that for all \( u \in L_2(\Omega) \) with \( \| u \|_{L_2(\Omega)} \leq \delta \), we have \( R(u) \leq M \).
3. For every \( \delta > 0 \), there exists \( L = L(\delta) > 0 \) such that for all \( u_1, u_2 \in L_2(\Omega) \) with
\[
\max \left\{ \| u_1 \|_{L_2(\Omega)}, \| u_2 \|_{L_2(\Omega)} \right\} \leq \delta,
\]
we have
\[
|R(u_1) - R(u_2)| \leq L \| u_1 - u_2 \|_{L_2(\Omega)}.
\]

**Proof.** Since (i) is trivial, we focus on (ii) and (iii). For (ii), note that (13) of Lemma 2.1 gives
\[
J_{NLTV}(u) \leq 2 \| u \|_{L_2(\Omega)}.
\]
For a given \( \delta > 0 \), we choose \( M = 2\lambda \delta \). Then whenever \( \| u \|_{L_2(\Omega)} \leq \delta \), we have
\[
R(u) = \lambda J_{NLTV}(u) \leq 2\lambda \| u \|_{L_2(\Omega)} \leq M,
\]
which concludes (ii). For (iii), first note that for \( u_1, u_2 \in L_2(\Omega) \), we have
\[
|J_{NLTV}(u_1) - J_{NLTV}(u_2)| = \left| \int_\Omega \| \nabla \omega u_1(x, \cdot) \|_{L_2(\Omega)} \, dx - \int_\Omega \| \nabla \omega u_2(x, \cdot) \|_{L_2(\Omega)} \, dx \right|
\]
\[
\leq \int_\Omega \| \nabla \omega (u_1 - u_2)(x, \cdot) \|_{L_2(\Omega)} \, dx
\]
\[
= J_{NLTV}(u_1 - u_2) \leq 2 \| u_1 - u_2 \|_{L_2(\Omega)}.
\]
where the last inequality follows from (14) of Lemma 2.1. Then by letting \( L = 2\lambda \), we conclude (iii). This completes the proof. \( \square \)

Theorem 2.2 states that \( \mu^y \) in (7) is a well-defined probability measure on \( L_2(\Omega) \) and it is Lipschitz continuous in the data \( y \). Since the theorem is a direct consequence of the fact that \( \Phi + R \) satisfies [38, Assumptions 2.6.], we omit the proof.

**Theorem 2.2.** Assume that \( F : L_2(\Omega) \to \mathbb{R}^m \) satisfies A1-A2. Let \( R : L_2(\Omega) \to \mathbb{R} \) be defined as (16). For a given \( y \in \mathbb{R}^m \), we define \( \mu^y \) as in (7). Then we have the followings:

1. \( \mu^y \) is a well-defined measure on \( L_2(\Omega) \).
2. \( \mu^y \) is Lipschitz continuous in the data \( y \) with respect to the Hellinger distance. More precisely, if \( \mu^y \) and \( \mu^{y'} \) are two measures corresponding to data \( y \) and \( y' \) respectively, then for every \( \delta > 0 \), there exists \( Z = Z(\delta) > 0 \) such that for all \( y, y' \in \mathbb{R}^m \) with
\[
\max \{ \| y \|_2, \| y' \|_2 \} \leq \delta,
\]
we have
\[
d_{Hell}(\mu^y, \mu^{y'}) \leq Z \| y - y' \|_{\Sigma_0^{-1}}.
\]
As a result, the expectation of any polynomially bounded function \( g : L_2(\Omega) \to K \) is continuous in \( y \).
For practical concerns, it is important to consider the finite dimensional approximation of $\mu^y$. In particular we consider the following finite dimensional approximation:

\[(17)\]

\[
d\mu^y_{N_1,N_2} \propto \exp\left( -\Phi_{N_1}(u) - R_{N_2}(u) \right),
\]

where $\Phi_{N_1}(u)$ is the $N_1$ dimensional approximation of $\Phi(u)$ with $F_{N_1}$ being the $N_1$ dimensional approximation of $F$ and $R_{N_2}(u)$ is the $N_2$ dimensional approximation of $R(u)$. Theorem 2.3 provides the convergence property of $\mu^y_{N_1,N_2}$.

**Theorem 2.3.** Assume that $F$ and $F_{N_1}$ satisfies A1 with constants independent of $N_1$ and $R_{N_2}$ satisfies Proposition 1 (i)-(ii) with constants independent of $N_2$. Assume further that for all $\epsilon > 0$, there exist two sequences \( \{a_{N_1}(\epsilon)\} \) and \( \{b_{N_2}(\epsilon)\} \), both of which converge to 0, such that

\[(18)\]

\[
\Phi(u) - \Phi_{N_1}(u) \leq a_{N_1}(\epsilon), \quad R(u) - R_{N_2}(u) \leq b_{N_2}(\epsilon),
\]

then we have

\[(19)\]

\[
d_{Hell}(\mu^y, \mu^y_{N_1,N_2}) \to 0,
\]

as $N_1, N_2 \to \infty$.

In particular, noting that $L_2(\Omega)$ is a separable Hilbert space, we can consider the finite dimensional approximation of $u$, as present in the following corollary.

**Corollary 1.** Let \( \{e_k : k \in \mathbb{N}\} \) be a complete orthonormal basis of $L_2(\Omega)$. For $N \in \mathbb{N}$, we define

\[(20)\]

\[
u_N = \sum_{k=1}^{N} \langle u, e_k \rangle e_k,
\]

and

\[(21)\]

\[
d\mu^y_{N_1,N_2} \propto \exp\left( -\Phi(u) - R(u_N) \right).
\]

If $F$ satisfies A1-A2, then we have

\[(22)\]

\[
d_{Hell}(\mu^y, \mu^y_{N_1,N_2}) \to 0,
\]

as $N \to \infty$.

The proof of Theorem 2.3 and Corollary 1 can be directly followed by [43, Theorem 2.3., Corollary 2.4], hence we omit here.

### 3. Bayesian estimators.

In this section, we discuss how to compute the two popular point estimators in the Bayesian setting. The first often used point estimator is the MAP estimator, and following the same steps as in [8], we can show that the MAP estimator with the NLTG prior is equivalent to solve

\[(23)\]

\[
\min_{u \in L_2(\Omega)} I(u) := \Phi(u) + \lambda J_{NLTV}(u) + \frac{1}{2} \|C_0^{-\frac{1}{2}}u\|_X,
\]

where $\Phi(u) = \frac{1}{2}\|Fu - y\|_2^2$ with $F$ being the linear operator in the limited tomography problem and the last term is the Cameron-Martin norm [37, 43] with respect to a Gaussian measure $\mu_0 = N(0, C_0)$ defined on the Hilbert space $\mathcal{X}$ with zero mean and covariance matrix $C_0$.

To minimize (23), we adopt the widely used split Bregman method [15] which is equivalent to the alternating direction method of multipliers (ADMM). For the
sake of clarity, we present the split Bregman method for (23) in Algorithm 1. The subproblem (24) is solved by conjugate gradient method and the subproblem (25) is solved directly by soft-shrinkage.

Algorithm 1 Split Bregman Method for MAP estimate with NLTG prior.

1: Initialization: $u^0 = u_{\text{ref}}, b^0 = d^0 = 0, y, \lambda, \mu$.
2: for $k = 0, 1, 2, \cdots$ do
3: Update $u^{k+1}$:
   \[ u^{k+1} = \arg\min_u \Phi(u) + \frac{\mu}{2} \| d_k - \nabla \omega u - b_k \|_2^2 + \frac{1}{2} \| C_0^{-\frac{1}{2}} u \|_{\mathcal{X}} \]
4: Update $d^{k+1}$:
   \[ d^{k+1} = \arg\min_d \lambda \| d \|_1 + \frac{\mu}{2} \| d - \nabla \omega u^{k+1} - b_k \|_2^2 \]
5: Update $b^{k+1}$:
   \[ b^{k+1} = b_k + \nabla \omega u^{k+1} - d^{k+1} \]
6: end for

Another often used point estimator in the Bayesian setting is the conditional mean (CM), or the posterior mean, which is usually evaluated using the samples drawn from the posterior distribution with the Markov Chain Monte Carlo (MCMC) methods. In this work we use the preconditioned Crank-Nicolson (pCN) algorithm developed in [7] for its property of being independent of discretization dimensionality. Simply speaking, the pCN algorithm proposes according to,

\[ v = (1 - \beta^2)^{\frac{1}{2}} u + \beta w, \]

where $u$ is the present position, $v$ is the proposed position, and $\beta \in [0, 1]$ is the parameter controlling the stepsize, and $w \sim \mathcal{N}(0, C_0)$. The associated acceptance probability is

\[ a(u, v) = \min \left\{ 1, \exp \left[ \Phi(u) + J_{\text{NLTV}}(u) - \Phi(v) - J_{\text{NLTV}}(v) \right] \right\}. \]

We describe the complete pCN algorithm in Algorithm 2.

Algorithm 2 The preconditioned Crank-Nicolson (pCN) algorithm

1: Initialization: $u^{(0)} \in \mathcal{X}$
2: for $n = 0$ to $N - 1$ do
3: Propose $v$ using Eq. (27);
4: Draw $\theta \sim U[0, 1]$
5: if $\theta \leq a(u^{(n)}, v)$ then
6: $u^{(n+1)} = v$;
7: else
8: $u^{(n+1)} = u^{(n)}$;
9: end if
10: end for
4. **Numerical results.**

4.1. **Experimental setup.** In this section, we present some experimental results to demonstrate the performance of the proposed NLTG prior. In particular we compare the results (both MAP and CM estimates) of the proposed method with those of the Filtered back projection (FBP) method, the NLTV regularization model and the TG method. We use the $128 \times 128$ XCAT image [35] taking integer values in $[0, 255]$ as the original image $u_{ori}$. Then the ground truth image $u_{gt}$ is generated by adding one round shaped object which stands for the tumor in lung and further adding a sinusoidal wave as an inhomogeneous background. Finally, the reference images $u_{ref}$, that can be considered as the previous CT image of the same patient taken by the same CT scanner, is obtained by a direct reconstruction from 500 projections of $u_{ori}$ with Gaussian noise of standard derivation of 5 and 20. The two images are denoted as $u_{ref}^1$ and $u_{ref}^2$ respectively and the above mentioned four images are shown in Figure 1.

![Figure 1. XCAT images: origin, ground truth and reference with different noise levels.](image)

Given a reference image $u_{ref}$, the covariance matrix $C_0$ for the Gaussian measure term is computed as

$$C_0(i, j) = \exp\left(-\frac{\|u_{ref}(i) - u_{ref}(j)\|^2}{h^2}\right)$$

following the idea of radial basis function kernel in machine learning [4] and the similarity weight in nonlocal means filter [3]. Here $h$ is a tuning parameter of Gaussian prior. We note that this choice of covariance matrix is different from usual Gaussian measure, as the correlation between the pixels value of $u_{ref}$ is used instead of spatial distance. Such choice aims to bring structures and edge information of the reference image to the to-be-reconstructed image, which is especially important for reconstruction from highly missing data. We adopt this covariance matrix for TG prior as well for a fair comparison. For the NLTV weight (as well as used NLTV regularization), we only use the 10 largest weights and the 4 closest neighbors for each pixel, as adopted in [44].

To synthesize the limited projection data $y$, the forward operator $F$ is taken as the discrete Radon transform followed by the restriction onto the discretization of $[0, \pi] \times [-\frac{1}{2}, \frac{1}{2}]$. In this experiments, we use 50 equally spaced projections with Gaussian noise of standard variation 1 and 10. Finally, all the reconstructed images are improved by imposing the intensity bound constraint $[0, 255]$ using

$$u_{rec} = \min\{\max(u^*, 0), 255\}$$
where $u^*$ denotes the output of all the methods in comparison.

### 4.2. MAP results.

In solving (23), the filtering parameter $h$ in both (8) and (29) is chosen as in [44]. The maximum outer iteration number is set as 80 and the regularization parameter $\lambda$ is manually chosen so that we can obtain the optimal restoration results.

Tables 1 summarizes the PSNR and SSIM values of all the methods. As we can see from the tables, our NLTG prior consistently outperforms the other reconstruction methods, namely the FBP, TV, TGV [2], TG and the NLTV priors. The reconstructed images with different methods are present in Figures 2 and 3. As can be expected, TV, TGV and FBP can not successfully reconstruct reasonable results due to limited projections. The TG prior based MAP estimate shows better performance since our proposed Gaussian covariance matrix (29) provides some structure information from the reference image as prior. We can also see that, compared to the TG prior, the NLTG prior shows the advantage of NLTV by using the similarity in the image. In addition, compared to the NLTV prior, the NLTG prior can obtain better recovery result, thanks to the presence of the Gaussian term which extracts more structure information by the covariance matrix computed from the reference image. As we can see from the figures, the visual improvements are consistent with the improvements in the indices. We also note that in terms of computation time, the hybrid priori, TG and NLTG are more time-consuming due to the extra covariance matrix term. In practice, this can be accelerated using a low dimensional approximation of the inverse of covariance matrix, as discussed in [47].

| Noise | Ref. | FBP  | TV   | TGV  | TG   | NLTV | NLTG |
|-------|------|------|------|------|------|------|------|
| 5     | $u_{\text{ref}}$ | 13.30/0.21 | 18.98/0.60 | 21.21/0.71 | 29.08/0.88 | 29.69/0.87 | **30.71/0.91** |
| 20    | $u_{\text{ref}}$ | 9.40/0.06   | 15.66/0.46 | 18.26/0.48 | 23.10/0.54 | 23.92/0.78 | **24.72/0.79** |

**Table 1.** MAP results: PSNR and SSIM for different sinogram noise levels and reference images

### 4.3. CM results.

The hyperparameters in the CM model are the same as the MAP model. For both TG prior and NLTG prior, we perform the pCN approach with $9.5 \times 10^5$ samples and another $0.5 \times 10^5$ samples as the pre-run. The stepwise $\beta$ has been chosen to make the acceptance probability is around 25%. The CM results are shown in Figure 4, and the PSNR and SSIM are shown in Table 2.

One can see from the figures and Table 2 that, the CM model with NLTG prior also outperformed the one with TG prior for different scenarios. Compared to MAP results, the PSNR of CM estimator is smaller under low noise level while higher under high noise level. Nonetheless, the SSIM of MAP estimator are consistently better than CM model. One can also see from the images that MAP estimator preserves better sharp edges on efficiently suppressing the noise in the low noise level while CM provides slightly more details on the high noise level. The main advantage of Bayesian techniques is that it can measure the uncertainty in the estimates. Figure 5 summarizes the 95% confidence interval (CI) gaps for each setting. It is easy to see that the CI gaps with TG prior are larger than that of NLTG prior, as one would expect as the similarity of structures extracted by NLTG prior improve the reconstruction than the local edge sharpness enhanced by TG prior. As expected, the uncertainty is higher with higher noise level. It is worth
noting that edges have high uncertainty than the smooth regions. This may be due to the fact that the edges are always more difficult to be preserved in CM estimator with both TG and NLTG prior, which leads to a high uncertainty interval.

5. Conclusions. In this paper, we consider the Bayesian inference methods for infinite dimensional inverse problems, and in particular we propose a prior distribution that combines the nonlocal method with a standard Gaussian distribution.
Theoretically, we show that the proposed prior distribution leads to a well-behaved posterior measure in the infinite dimensional setting. The proposed prior is applied to a severely data missing inverse problem: limited CT reconstruction. The numerical experiments demonstrate the performance of the proposed NLTG prior is superior to the existing and adapted methods. The comparison of MAP and CM estimators also provide some insights for solving high dimensional imaging inverse problems. We believe that the proposed NLTG prior distribution can be useful in a large class of image reconstruction problems, especially when reference images are available.

REFERENCES

[1] R. N. Bracewell and A. C. Riddle, Inversion of fan-beam scans in radio astronomy, *Astrophysical Journal*, **150** (1967).
[2] K. Bredies, K. Kunisch and T. Pock, Total generalized variation, *SIAM J. Imaging Sci.*, **3** (2010), 492–526.
[3] A. Buades, B. Coll and J. M. Morel, A review of image denoising algorithms, with a new one, *Multiscale Model. Simul.*, **4** (2005), 490–530.
(a) Noise level of 5 with $u_{ref}^{1}$ (b) Noisy level of 5 with $u_{ref}^{2}$ (c) Noisy level of 20 with $u_{ref}^{1}$ (d) Noisy level of 20 with $u_{ref}^{2}$

**Figure 4.** CM results with different sinogram data noise levels and references images. Upper: NLTG; Lower: TG.

(a) Noise level of 5 with $u_{ref}^{1}$ (b) Noisy level of 5 with $u_{ref}^{2}$ (c) Noisy level of 20 with $u_{ref}^{1}$ (d) Noisy level of 20 with $u_{ref}^{2}$

**Figure 5.** The 95% confidence interval for different sinogram data noise level and references images. The range of the values is from 0 (black) to 100 (whitest). Upper: NLTG; Lower: TG.

[4] C. Chang and C. Lin, LIBSVM: A library for support vector machines, *ACM Transac. Intelligent Systems Technology (TIST)*, 2 (2011).

[5] K. Choi, J. Wang, L. Zhu, T. S. Suh, S. Boyd and L. Xing, Compressed sensing based cone-beam computed tomography reconstruction with a first-order method, *Med. Phys.*, 37 (2010), 5113–5125.
[6] F. R. K. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, 92, American Mathematical Society, Providence, RI, 1997.

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

[8] M. Dashti, K. J. H. Law, A. M. Stuart and J. Voss, MAP estimators and their consistency in Bayesian nonparametric inverse problems, *Inverse Problems*, 29 (2013), 27pp.

[9] A. A. Efros and T. K. Leung, Texture synthesis by nonparametric sampling, IEEE International Conference on Computer Vision, Greece, 1999, 1033–1038.

[10] A. Elmoataz, O. Lezoray and S. Bougleux, Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing, *IEEE Trans. Image Process.*, 17 (2008), 1047–1060.

[11] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, Pure and Applied Mathematics, John Wiley & Sons, Inc., New York, 1999.

[12] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin, *Bayesian Data Analysis*, Texts in Statistical Science Series, CRC Press, Boca Raton, FL, 2014.

[13] G. Gilboa and S. Osher, Nonlocal linear image regularization and supervised segmentation, *Multiscale Model. Simul.*, 6 (2007), 595–630.

[14] G. Gilboa and S. Osher, Nonlocal operators with applications to image processing, *Multiscale Model. Simul.*, 7 (2008), 1005–1028.

[15] T. Goldstein and S. Osher, The split Bregman method for $L_1$-regularized problems, *SIAM J. Imaging Sci.*, 2 (2009), 323–343.

[16] T. Heußer, M. Brehm, S. Marcus, S. Sawall and M. Kachelrieß, CT data completion based on prior scans, *IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, Anaheim, CA, 2012, 2969–2976.

[17] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Applied Mathematical Sciences, 160, Springer-Verlag, New York, 2005.

[18] H. Kim, J. Chen, A. Wang, C. Chuang, M. Held and J. Pouliot, Non-local total-variation NLTV minimization combined with reweighted L1-norm for compressed sensing CT reconstruction, *Phys. Med. Biol.*, 61 (2016).

[19] J. Li, A note on the Karhunen-Loève expansions for infinite-dimensional Bayesian inverse problems, *Statist. Probab. Lett.*, 106 (2015), 1–4.

[20] J. Liu, H. Ding, S. Molloi, X. Zhang and H. Gao, Nonlocal total variation based spectral CT image reconstruction, *Med. Phys.*, 42 (2015), 3570–3570.

[21] J. Liu, H. Ding, S. Molloi, X. Zhang and H. Gao, TICMR: Total image constrained material reconstruction via nonlocal total variation regularization for spectral CT, *IEEE Trans. Medical Imaging*, 35 (2016), 2578–2586.

[22] Y. Lou, X. Zhang, S. Osher and A. Bertozzi, Image recovery via nonlocal operators, *J. Sci. Comput.*, 42 (2010), 185–197.

[23] F. Lucka, S. Pursiainen, M. Burger and C. H. Wolters, Hierarchical Bayesian inference for the EEG inverse problem using realistic FE head models: Depth localization and source separation for focal primary currents, *Neuroimage*, 61 (2012), 1364–1382.

[24] F. Natterer, *The Mathematics of Computerized Tomography*, John Wiley & Sons, Ltd., Chichester, 1986.

[25] X. Pan, E. Y. Sidky and M. Vannier, Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?, *Inverse Problems*, 25 (2009), 36pp.

[26] G. Peyré, Image processing with nonlocal spectral bases, *Multiscale Model. Simul.*, 7 (2008), 703–730.

[27] G. Peyré, S. Bougleux and L. Cohen, Non-local regularization of inverse problems, in *ECCV 2008: Computer Vision*, Lecture Notes in Computer Science, 5304, Springer, Berlin, Heidelberg, 2008, 57–68.

[28] E. T. Quinto, Singularities of the X-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$, *SIAM J. Math. Anal.*, 24 (1993), 1215–1225.
[32] J. Radon, Uber die bestimmung von funktionen durch ihre integralwerte laengs geweisser mannigfaltigkeiten, Berichte Saechsishe Acad. Wissenschaft. Math. Phys., Klass, 69 (1917).

[33] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D, 60 (1992), 259–268.

[34] T. Schuster, B. Kaltenbacher, B. Hofmann and K. S. Kazimierski, Regularization Methods in Banach Spaces, Radon Series on Computational and Applied Mathematics, 10, Walter de Gruyter GmbH & Co. KG, Berlin, 2012.

[35] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes and B. M. W. Tsui, 4D XCAT phantom for multimodality imaging research, Med. Phys., 37 (2010), 4902–4915.

[36] E. Y. Sidky and X. Pan, Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization, Phys. Med. Biol., 53 (2008).

[37] A. M. Stuart, Inverse problems: A Bayesian perspective, Acta Numer., 19 (2010), 451–559.

[38] S. J. Vollmer, Dimension-independent MCMC sampling for inverse problems with non-gaussian priors, SIAM/ASA J. Uncertain. Quantif., 3 (2015), 535–561.

[39] S. J. Vollmer, Dimension-independent MCMC sampling for inverse problems with non-Gaussian priors, SIAM/ASA J. Uncertain. Quantif., 3 (2015), 535–561.

[40] G. Wang and H. Yu, The meaning of interior tomography, Phys. Med. Biol., 58 (2013), R161–186.

[41] J. P. Ward, M. Lee, J. C. Ye and M. Unser, Interior tomography using 1D generalized total variation, Part I: Mathematical foundation, SIAM J. Imaging Sci., 8 (2015), 226–247.

[42] J. Yang, H. Yu, M. Jiang and G. Wang, High-order total variation minimization for interior tomography, Inverse Problems, 26 (2010), 29pp.

[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), 19pp.

[44] X. Zhang, M. Burger, X. Bresson and S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, SIAM J. Imaging Sci., 3 (2010), 253–276.

[45] X. Zhang and T. F. Chan, Wavelet inpainting by nonlocal toral variation, Inverse Probl. Imaging, 4 (2010), 191–210.

[46] D. Zhou and B. Schölkopf, Regularization on discrete spaces, in Joint Pattern Recognition Symposium, Lecture Notes in Computer Science, 3663, Springer, Berlin, Heidelberg, 2005, 361–368.

[47] Q. Zhou, W. Liu, J. Li and Y. M. Marzouk, An approximate empirical Bayesian method for large scale linear Gaussian inverse problems, Inverse Problems, 34 (2018), 18pp.

Received April 2019; revised August 2019.

E-mail address: Eric2014_Lv@sjtu.edu.cn
E-mail address: zhou2015@sjtu.edu.cn
E-mail address: jayckjk@tongji.edu.cn
E-mail address: jinglai.li@liverpool.ac.uk
E-mail address: xqzhang@sjtu.edu.cn