Variationally Bayesian Medical Image Reconstruction Applied to X-ray Tomography*

Jonas Latz[1][0000−0002−4600−0247], Sergio Bacallado[1][0000−0002−7193−6450], Claire Delplancke[2][0000−0001−7483−0419], Matthias J. Ehrhardt[2][0000−0001−8523−353X], and Carola-Bibiane Schönlieb[1][0000−0003−0099−6306]

[1] Centre for Mathematical Sciences, University of Cambrige, Wilberforce Road, Cambridge CB3 0WA, United Kingdom
[2] Department of Mathematical Sciences, University of Bath, North Rd, Claverton Down, Bath BA2 7AY, United Kingdom
{jl2160, sb2116, cbs31}@cam.ac.uk, {cd902, me549}@bath.ac.uk

Abstract. We present a Bayesian statistical approach to medical image reconstruction. Here, we obtain not a point estimate for the image of interest, but a posterior distribution that also quantifies the reconstruction uncertainty. We approximate the posterior variationally. We illustrate the effectiveness of our approach in numerical experiments in which we study variationally Bayesian X-ray tomography.

Keywords: Medical Image Reconstruction · Statistical Learning

1 Statistical Image Reconstruction

Medical image reconstruction problems can often be formulated in the following way: Identify an *image* $\theta^\dagger \in \mathbb{R}^{N_{\text{im}}}$ given a *data set* $y^\dagger \in \mathbb{R}^{N_{\text{dt}}}$ which satisfies

$$y^\dagger = A\theta^\dagger + \eta^\dagger,$$

where $A \in \mathbb{R}^{N_{\text{dt}} \times N_{\text{im}}}$ is a matrix and $\eta^\dagger$ is noise. In, e.g., X-ray tomography A represents the Radon transform and $y^\dagger$ is the sinogram acquired by the scanner. This image reconstruction problem is usually approached by a (regularised) maximum likelihood method, which gives an estimate $\theta_* \in \mathbb{R}^{N_{\text{im}}}$ that approximates the true $\theta^\dagger$.

Due to noise and/or ill-conditioning of A, the estimate θ_* may contain a considerable amount of remaining uncertainty concerning the true image $\theta^\dagger$. Unfortunately, this uncertainty cannot be quantified through θ_* itself. Instead, a *Bayesian statistical approach* can be employed; see [3] for details. Here, the image $\theta^\dagger$ is represented by a random variable θ that has a certain probability distribution $\mu_{\text{prior}} := \mathbb{P}(\theta \in \cdot)$. This so-called *prior* represents knowledge and

* The authors acknowledge support from the EPSRC grant EP/S026045/1 "PET++: Improving Localisation, Diagnosis and Quantification in Clinical and Medical PET Imaging with Randomised Optimisation".

uncertainty in the image before the data $y^\dagger$ is observed. Then, Bayes' formula is employed to determine the *posterior* $\mu_{\text{post}} := \mathbb{P}(\theta \in \cdot | A\theta + \eta^\dagger = y^\dagger)$, which gives an accurate description of the knowledge gained through learning the data and of the remaining uncertainty in the reconstruction. In practice, this posterior usually needs to be approximated computationally.

2 Variational Bayes

To approximate the posterior in our medical image reconstruction problem we employ the so-called *variational Bayes* approach; see, e.g., Chapter 10 of [1]: we define a set M of probability distributions on $\mathbb{R}^{N_{\text{im}}}$ and find

$$\mu_{\text{var}} := \text{argmin}_{\mu \in M} D_{\text{KL}}(\mu \| \mu_{\text{post}}), \tag{1}$$

where D_{KL} denotes the *Kullback–Leibler Divergence*.

For the computational approximation of μ_{var}, it is vital that $D_{\text{KL}}(\mu \| \mu_{\text{post}})$ can be evaluated without first approximating μ_{post}. This is the case, if, e.g., μ_{prior} and $\mu \in M$ describe the distribution of some $f(\varphi)$, where f is a given function and φ is a non-degenerate Gaussian random vector.

3 Eulerian and Lagrangian approaches

Inspired by [2], we consider two different models for μ_{prior} and M: a Eulerian and a Lagragian model. In both cases, we aim to represent the uncertain image θ through a system $(\psi_k)_{k=1}^{N_{\text{ba}}} \in \mathbb{R}^{N_{\text{im}} \times N_{\text{ba}}}$, i.e.

$$\theta = \sum_{k=1}^{N_{\text{ba}}} \xi_k \psi_k$$

for suitable random prefactors $\xi_1, \dots, \xi_{N_{\text{ba}}}$. The difference of Eulerian and Lagrangian approach appears in the choice of the system $(\psi_k)_{k=1}^{N_{\text{ba}}}$: In the *Eulerian approach*, we choose $(\psi_k)_{k=1}^{N_{\text{ba}}}$ to be a high-dimensional fixed frame or basis, e.g., a system of wavelets or curvelets, and the $(\xi_k)_{k=1}^{N_{\text{ba}}}$ to be sparsity-inducing. In the *Lagragian approach*, we assume to have a low-dimensional system $(\psi_k)_{k=1}^{N_{\text{ba}}}$ that can be adapted. Hence, not only the prefactors $(\xi_k)_{k=1}^{N_{\text{ba}}}$ are uncertain, but also the vectors in $(\psi_k)_{k=1}^{N_{\text{ba}}}$.

3.1 Particular modelling

Let $S(x) = (1 + \exp(-x))^{-1}$ be a non-linearity acting element-wise on vectors. Then, we choose the following:
Eulerian approach. We set $\xi_k := \zeta_k S(\zeta_k'), k = 1, \dots, N_{\text{ba}}$. In the prior, we choose $\zeta_1, \dots, \zeta_{N_{\text{ba}}}, \zeta_1', \dots, \zeta_{N_{\text{ba}}}' \sim \text{N}(0, 1^2)$ independent. In M we consider the same structure but assume that the means of the $(\zeta_k, \zeta_k')_{k=1}^{N_{\text{ba}}}$ and the variances of the

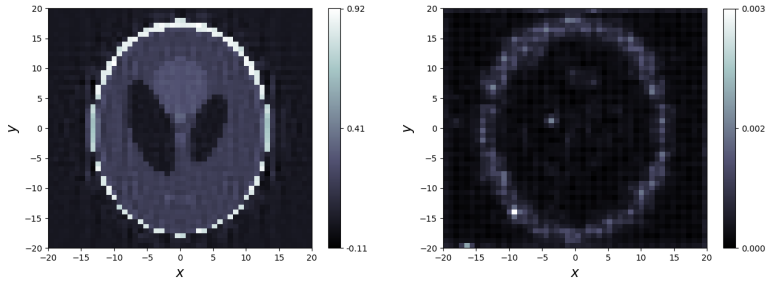

Fig. 1. Eulerian variational posterior: Point-wise mean (left column) and variance (right column).

$(\zeta_k)_{k=1}^{N_{\mathrm{ba}}}$ are degrees of freedom. The resulting prior of the prefactors $(\xi_k)_{k=1}^{N_{\mathrm{ba}}}$ is a smooth spike-and-slab prior that enforces sparsity.

Lagrangian approach. We set $\xi_k := \zeta_k, \psi_k = S(\Sigma\zeta' - m_k), k = 1, \ldots, N_{\mathrm{ba}}$, where $\Sigma \in \mathbb{R}^{N_{\mathrm{im}} \times N_{\mathrm{in}}}$ is a fixed matrix and $m_1 < \ldots < m_{N_{\mathrm{ba}}}$ are fixed scalars. In the prior, we choose $\zeta_1, \ldots, \zeta_{N_{\mathrm{ba}}}, \zeta'_1, \ldots, \zeta'_{N_{\mathrm{in}}} \sim \mathrm{N}(0, 1^2)$ independent. In M we consider the same structure but assume that mean and covariance matrix of $(\zeta_k)_{k=1}^{N_{\mathrm{ba}}}$, as well as the mean and variances of the $(\zeta'_n)_{n=1}^{N_{\mathrm{in}}}$ are degrees of freedom. The basis functions ψ_k are obtained through a thresholding of the random object $\Sigma\zeta'$, which represents a discretisation of a (smooth) Gaussian process.

Importantly, both models represent non-linear transformations of Gaussian random variables. Thus, they are suitable for a computationally efficient variational approximation, as discussed in §2.

4 Examples of variationally Bayesian reconstructions

We now give computational examples in which we reconstruct a 50×50 Shepp–Logan phantom based on full-angle X-ray data. We assume to have perturbed the data with scaled Gaussian white noise and quantify the uncertainty in the image as if this noise had been applied. The results shown, however, are based on noisefree data. In each case, we employ suitable stochastic optimisation methods to solve the variational optimisation problems (1).

In the *Eulerian approach*, we choose $(\psi_k)_{k=1}^{N_{\mathrm{ba}}}$ to be a frame of 5184 coiflets. We show the results in Figure 1.

In the *Lagrangian approach*, we consider two settings. In both settings, we choose Σ to be a matrix the columns of which are $N_{\mathrm{in}} = 2500$ orthogonal sinusoidal basis vectors. The basis vectors are weighted with prefactors which go to zero as the frequency of oscillation of the basis vector increases. In the first setting ('long correlation length') the prefactors converge quickly, in the second case ('short correlation length') the prefactors converge slower. In both cases, we choose $N_{\mathrm{ba}} = 11$ adapted basis vectors. We show the results of our estimations in Figure 2.

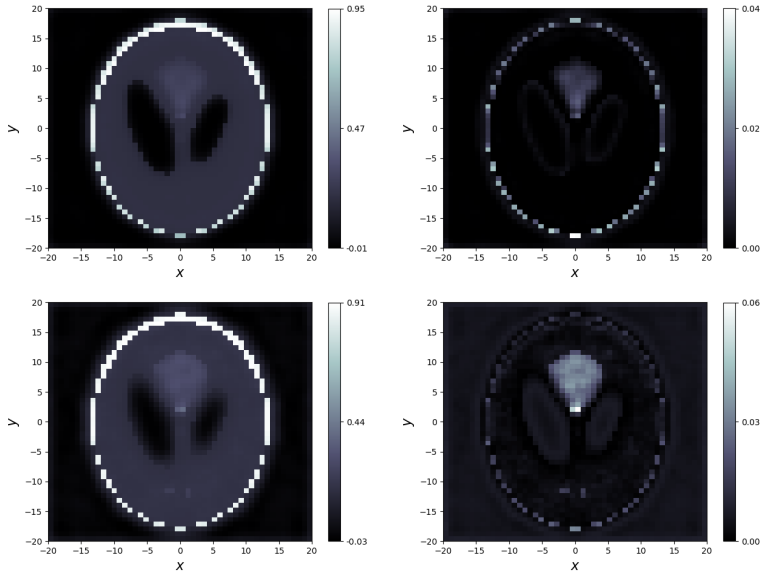

Fig. 2. Lagrangian variational posterior: Point-wise mean (left column) and variance (right column). Long correlation length (top row) and short correlation length (bottom row).

In the Eulerian approach, we see an overall rougher appearance of the piecewise constant phantom and a very small posterior variance. The Lagrangian posterior means are a much more accurate reconstruction of the phantom and the variances appear to highlight areas in which we expect a larger reconstruction uncertainty. Using the long correlation length appears to give us a smoother image and cleaner edges, but loses small details. The short correlation length blurs the edges slightly, but retains all fine details.

References

1. Bishop, C.: Pattern recognition and machine learning. Springer-Verlag, New York (2006).
2. Candès, E.J., Donoho, D.L.: Curvelets: A surprisingly effective nonadaptive representation for objects with edges. In: Cohen, A., Rabut, C., Schumaker, L.L. (eds.) Curve and Surface Fitting: Saint-Malo 1999, Vanderbilt Univ. Press, Nashville, TN (2000).
3. Stuart, A.M.: Inverse Problems: A Bayesian perspective. Acta Numerica **19**, 451–559 (2010).