Sampling-Based Inference for Linearised Deep Image Prior with Application to Cone-Beam CT Reconstruction

Riccardo Barbano

Computer Science Department, University College London

A joint work with Shreyas Padhy and Javier Antoran

Extending [APB+22] to Cone-Beam CT reconstruction

Overview

1. Bayesian linear regression and the EM algorithm
2. Evidence maximisation using stochastic approximation
3. NN uncertainty quantification as linear model inference
4. Tomographic reconstruction & uncertainty quantification
5. Results

Bayesian Gaussian regression and the EM algorithm I

Bayesian Gaussian linear regression

Consider a Bayesian conjugate Gaussian linear regression with multidimensional outputs. We observe inputs $x_1, \ldots, x_n \in \mathbb{R}^d$ and corresponding outputs $y_1, \ldots, y_n \in \mathbb{R}^m$, modelled as

$$y_i = \phi(x_i)\theta + \eta_i, \tag{1}$$

where $\phi \colon \mathbb{R}^d \mapsto \mathbb{R}^m \times \mathbb{R}^{d'}$ is a known embedding function. The parameters θ are assumed sampled from $\mathcal{N}(0, A^{-1})$ with an unknown precision matrix $A \in \mathbb{R}^{d' \times d'}$, and for each $i \leq n$, $\eta_i \sim \mathcal{N}(0, B_i^{-1})$ are additive noise vectors with precision matrices $B_i \in \mathbb{R}^{m \times m}$ relating the m output dimensions.
We then use the stacked notation: $Y \in \mathbb{R}^{nm}$ for concatenation of $y_1, \ldots, y_n$; $\mathrm{B} \in \mathbb{R}^{nm \times nm}$ for a block diagonal matrix with blocks $B_1, \ldots, B_n$; and $\Phi = [\phi(X_1)^T; \ldots; \phi(X_n)^T]^T \in \mathbb{R}^{nm \times d'}$ for the embedded design matrix.

Bayesian Gaussian regression and the EM algorithm II

Posterior inference & EM algorithm [Bis06]

Our goal is to infer the posterior distribution for the parameters θ given our observations, under the setting of A of the form $A = \alpha I$ for $\alpha > 0$ most likely to have generated the observed data.

We use the iterative procedure of [Mac92], which alternates computing the posterior for θ, denoted Π, for a given choice of A, and updating A, until the pair (A, Π) locally converge.
EM algorithm starts with some initial $A \in \mathbb{R}^{d' \times d'}$, and iterates:

- (E step) Given A, the posterior for θ, denoted Π, is computed exactly as

$$\Pi = \mathcal{N}(\bar{\theta}, H^{-1}), \tag{2}$$

where $H = M + A$, where $M = \Phi^T \mathrm{B} \Phi$, and $\bar{\theta} = H^{-1} \Phi^T \mathrm{B} Y$.

Bayesian Gaussian regression and the EM algorithm III

- (M step) We lower bound the log-probability density of the observed data, i.e. the evidence, for the model with posterior Π and precision A' as

$$\log p(Y; A') \geq -\frac{1}{2}\|\bar{\theta}\|_{A'}^2 - \frac{1}{2}\log\det(I + A'^{-1}M) + C \qquad (3)$$
$$:= \mathcal{M}(A'),$$

for C independent of A'. We choose an A that improves this lower bound.

Limited scalability

The above inference and hyperparameter selection procedure for Π and A is futile when both d' and nm are large.

Evidence maximisation using stochastic approximation I

M-step via stochastic "Mackay update"

Let's assume we can efficiently obtain samples $\zeta_1, \ldots, \zeta_k \sim \Pi^0$ at each step, where Π^0 is a zero-mean version of the posterior Π, and access to $\bar{\theta}$, the mean of Π.

[APB+22] show that $\alpha = \hat{\gamma}/\|\bar{\theta}\|^2$, where

$$\gamma = \mathrm{Tr}\left\{H^{-1}M\right\} = \mathrm{Tr}\left\{H^{-\frac{1}{2}}MH^{-\frac{1}{2}}\right\} \tag{4}$$

$$= \mathbb{E}[\zeta_1^T M \zeta_1] \approx \frac{1}{k}\sum_{j=1}^{k} \zeta_j^T \Phi^T \mathrm{B} \Phi \zeta_j := \hat{\gamma}. \tag{5}$$

where the quantity γ is the effective dimension of the regression problem. It can be interpreted as the number of directions in which the weights θ are strongly determined by the data.

Evidence maximisation using stochastic approximation II

E-step via sampling from the linear model's posterior

We turn to sampling from $\Pi^0 = \mathcal{N}(0, H^{-1})$.
It is known that for $\mathcal{E} \in \mathbb{R}^{nm}$ the concatenation of $\epsilon_1, \ldots, \epsilon_n$ with $\epsilon_i \sim \mathcal{N}(0, B_i^{-1})$ and $\theta^0 \sim \mathcal{N}(0, A^{-1})$, the minimiser of is a random variable ζ with distribution Π^0:

$$L(z) = \frac{1}{2}\|\Phi z - \mathcal{E}\|_{\mathrm{B}}^2 + \frac{1}{2}\|z - \theta^0\|_A^2. \tag{6}$$

This is called the "sample-then-optimise" method [PY10]. We may thus obtain a posterior sample by optimising this quadratic loss for a given sample pair $(\mathcal{E}, \theta^0)$.

NN uncertainty quantification as linear model inference I

Linearised Laplace method

We train a NN of the form $f\colon \mathbb{R}^{d'} \times \mathbb{R}^{d} \mapsto \mathbb{R}^{m}$, obtaining weights $\bar{w} \in \mathbb{R}^{d'}$ optimising

$$\mathcal{L}(f(w, \cdot)) = \sum_{i=1}^{n} \ell(y_i, f(w, x_i)) + \mathcal{R}(w) \tag{7}$$

We then resort to the linearised Laplace method:

- We take a first-order Taylor expansion of f around $\bar{w}$, yielding the surrogate model

$$h(\theta, x) = f(\bar{w}, x) + \phi(x)(\theta - \bar{w}), \tag{8}$$

for $\phi(x) = \nabla_w f(\bar{w}, x)$. This is an affine model in the features $\phi(x)$ given by the network Jacobian at x.

Algorithm I

https://github.com/educating-dip/bayes_dip

Algorithm 1: Sampling-based linearised Laplace

Inputs: initial $\alpha > 0$; $k, k' \in \mathbb{N}$, number of samples for stochastic EM and prediction, respectively.

Sample random $\theta_1^n, \ldots, \theta_k^n$

while α has not converged **do**

- Find $\bar{\theta}$ by optimising linear model loss $\mathcal{L}(h(\theta, \cdot))$
- Draw posterior samples $\zeta_1 \ldots \zeta_k$ by optimising objective L with $\theta_1^n, \ldots, \theta_k^n$
- Estimate effective dimension $\hat{\gamma}$, using samples $\zeta_1 \ldots \zeta_k$
- Update prior precision $\alpha \leftarrow \hat{\gamma}/\|\bar{\theta}\|_2^2$

Output: posterior samples $\zeta'_1, \ldots, \zeta'_k$

Tomographic reconstruction & UQ I

Problem Setup: Cone-Beam CT reconstruction

CT reconstruction consists in solving a linear inverse problem in imaging. We observe a set of measurements $y \in \mathbb{R}^m$, which we assume to be generated as

$$y = Ux^* + \eta \tag{9}$$

for $U \in \mathbb{R}^{m \times d}$ the discrete Radon transform, $x^* \in \mathbb{R}^d$ the image to reconstruct and $\eta \sim \mathcal{N}(0, I)$ random noise. We have $m \ll d$, making the problem under-constrained.

Tomographic reconstruction & UQ II

Reconstruction with the Deep Image Prior

We reconstruct x^* with Deep Image prior [UVL20], which trains $w \in \mathbb{R}^{d'}$ of a fully convolutional U-Net autoencoder $f : \mathbb{R}^{d'} \to \mathbb{R}^{d}$, where the input is fixed.

To estimate the uncertainty in this reconstruction, we linearise the U-Net around $\bar{w}$.

This leaves us with a model affine in the parameters and with design matrix $U\Phi \in \mathbb{R}^{m \times d'}$.

Results I

We consider 3D reconstruction of a Walnut with a downscaled image resolution of $(167px)^3$, from 20 equally distributed angles, sub-sampling projection rows and columns by a factor of 3. Here $n = 1, m = 1.6M, d' \approx 5M$.

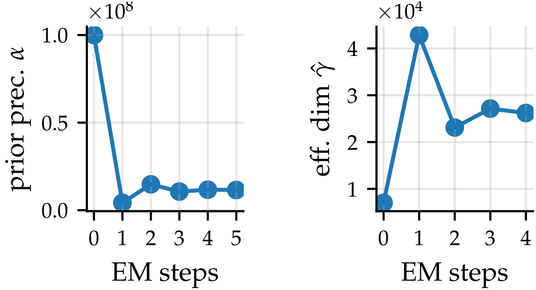
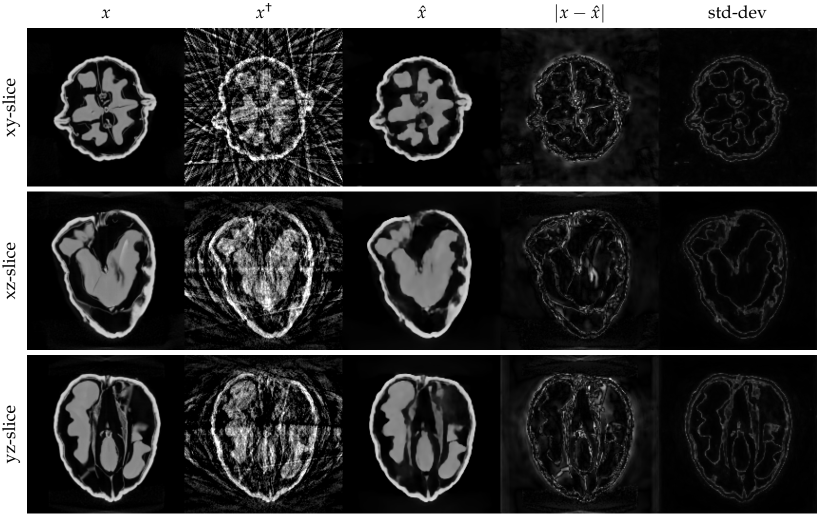


Figure: Traces of prior precision α and eff. dim. $\hat{\gamma}$ vs EM steps for the tomographic reconstruction task.

Results II



Results III

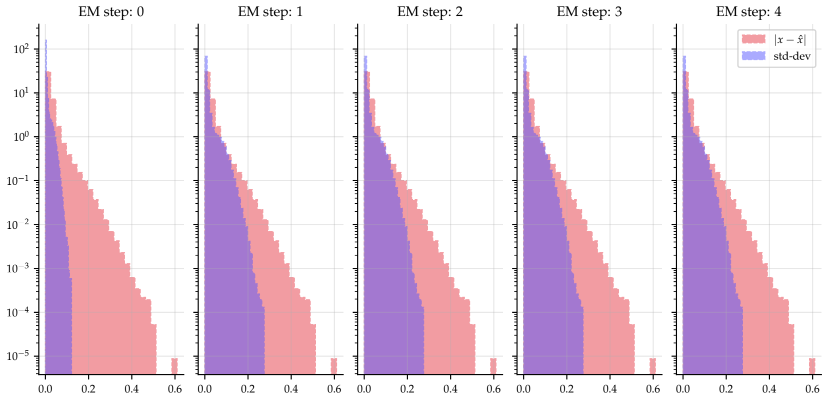


Figure: Histograms of the voxel-wise error computed between the reconstructed walnut and the ground-truth along with the histograms of predictive standard deviations across voxels.

References I

Javier Antorán, Shreyas Padhy, Riccardo Barbano, Eric Nalisnick, David Janz, and José Miguel Hernández-Lobato, Sampling-based inference for large linear models, with application to linearised laplace, arXiv preprint arXiv:2210.04994 (2022).

Christopher M Bishop, Pattern recognition and machine learning, springer, 2006.

David John Cameron Mackay, Bayesian methods for adaptive models, Ph.D. thesis, USA, 1992.

George Papandreou and Alan L. Yuille, Gaussian sampling by local perturbations, Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada (John D. Lafferty, Christopher K. I. Williams, John Shawe-Taylor, Richard S. Zemel, and Aron Culotta, eds.), Curran Associates, Inc., 2010, pp. 1858–1866.

Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky, Deep image prior, Int. J. Comput. Vis. **128** (2020), no. 7, 1867–1888.