11 1st, 2020

Bayesian inference and uncertainty quantification for image reconstruction with Poisson data

Qingping Zhou(周清平)

joined work with Tengchao Yu, Xiaoquan Zhang and Jinglai Li

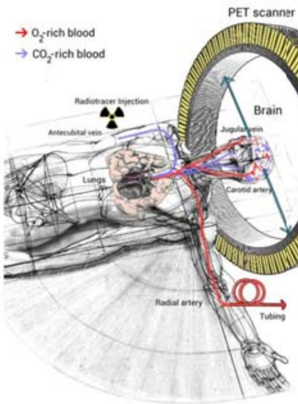
School of Mathematics and Statistics

CENTRAL SOUTH UNIVERSITY

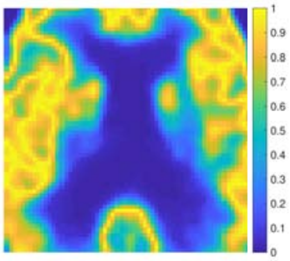
Image reconstruction

- Image reconstruction can often be formulated as mathematical model $\mathbf{y} = A\mathbf{u} \otimes \eta$, the operator $\otimes$ denote addition(Gaussian noise) or nonlinear operator(Poisson noise):
  - Denoising: A is an identity operator
  - Deblurring: A is a convolution operator
  - Positron emission tomography (PET): A is a linear projection operator
  - ...
- It can be modeled as an inverse problem: $\mathbf{y} \Rightarrow \mathbf{u}$
- Various sources of uncertainty: observation noise, model error, numerical error, ill-posedness...
  The estimation results have uncertainty!

Positron Emission Tomography(PET)

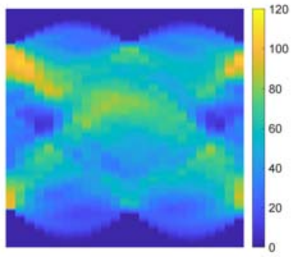


- $u(x)$: unknown
- $\mathcal{L}(\theta, t)$: positron beam
- Sinogram(noisy free data):
  $\psi(t, \theta) = Af = \int_{\mathcal{L}(\theta,t)} u(x)|dx|$.
- Data y follows a Poisson distribution:
  $$\boxed{y|\psi \sim \pi_{\mathrm{P}}(\cdot|K\psi) = \prod_{i=1}^{d} \frac{(K\psi_i)^{y_i} \exp(-K\psi_i)}{y_i!}}$$
- Goal: reconstruct $u(x)$



Ground truth

$\xRightarrow[Transform]{Radon}$



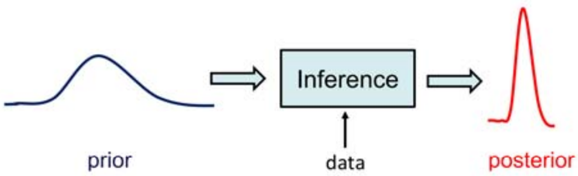
Simulated data

Bayesian formula

- Bayes' Theorem:

$$\overbrace{\pi(u|y)}^{\text{posterior}} = \frac{\overbrace{\pi(y|u)}^{\text{likelihood}}\ \overbrace{\pi(u)}^{\text{prior}}}{\underbrace{\pi(y)}_{\text{evidence}}}$$

$$\pi(u|y) \propto \pi(y|u)\ \pi(u)$$



  prior distribution: knowledge on $\mathbf{u}$ before $\mathbf{y}$ observed.

  posterior distribution: knowledge on $\mathbf{u}$ after $\mathbf{y}$ observed.

  likelihood function: the impact of $\mathbf{y}$ on the degree of belief on $\mathbf{u}$.

  evidence: the probability of $\mathbf{y}$, also known as normalizaiton constant.

- The prior is required to yield a "useful" estimate of the unknown.
- The posterior summarizes information from both the data and the prior.

Bayesian formula in function space

- Let X bet the Sobolev space $H^1(\Omega)$ in which Ω is a bounded open subset of $\mathbb{R}^d$. The posterior measure μ^y is provided by the Radon-Nikodym derivative:

$$\frac{d\mu^y}{d\mu_0}(u) \propto \exp[-\Phi(u, y)],$$

  where

  - μ_0 is a prior measure, a.k.a reference measure
  - $\Phi(u, \mathbf{y})$ is a potential function

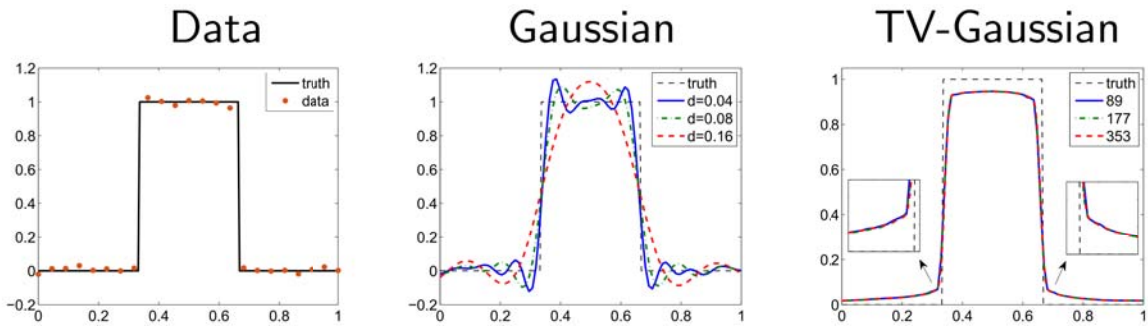
- Positivity-preserving reparametrization
  - $u = f(z(\mathbf{x})) = \frac{a}{2}[\mathrm{erf}(z(\mathbf{x})) + b], a > 0, b > 1.$

- Infer new unkown z

$$\frac{d\mu^y}{d\mu_0}(z) \propto \exp[-\Phi(z, y)].$$

Prior

- Total variation-Gaussian (TV-Gaussian) prior can detect edges or jumps in the images or functions



- The Radon-Nikodym derivative of μ_{pr} w.r.t. μ_0 is given by:

$$\frac{d\mu_{\mathrm{pr}}}{d\mu_0}(z) \propto \exp(-R(z)),$$

where

- $R(z) = \lambda \|z\|_{\mathrm{TV}} = \lambda \int_\Omega \|\nabla z\|_2 dx$
- $\mu_0 = N(\xi, C_0)$, $C_0(\mathbf{x}, \mathbf{x}') = \gamma \exp\left[-\frac{\|\mathbf{x}-\mathbf{x}'\|_1}{d}\right]$

Likelihood and posterior

- Likelihood function(noise level $K = 1$)

$$\pi(\mathbf{y}|u) = \pi_{\mathrm{P}}(\mathbf{y}|K\psi) = \prod_{i=1}^{d} \frac{(\psi_i)^{y_i}\exp(-\psi_i)}{y_i!}.$$

- Posterior measure
  Radon-Nikodym derivative of μ^y w.r.t. μ_0 can be written as:

$$\frac{d\mu^y}{d\mu_0}(z) = \frac{d\mu^y}{d\mu_{\mathrm{pr}}}\frac{d\mu_{\mathrm{pr}}}{d\mu_0}(z) \propto \exp[-\Phi(z,\mathbf{y}) - R(z)],$$

  we define $\Psi(z) := \Phi(z,\mathbf{y}) + R(z)$.

Primal-dual(PD) preconditioned Crank-Nicolson Langevin method(pCNL) algorithm

- pCNL is a dimension-independent sampling method.

$$(2+\delta)v = \begin{cases} (2-\delta)z - 2\delta\mathcal{C}_0\mathcal{D}\Psi(z) + \sqrt{8\delta}w & \text{pCNL} \\ (2-\delta)z - 2\delta\mathcal{C}_0 g(z) + \sqrt{8\delta}w & \text{PD-pCN,} \end{cases}$$

  where $\delta \in [0,2]$, $w \sim N(\xi, \mathcal{C}_0)$.

- Find offset direction $g(z)$ using primal-dual method
  - Define problem: $\min_{z\in X} \Psi(z) = \Phi(z) + \lambda\|z\|_{\mathrm{TV}}$ s.t. $\mathcal{D}z = \phi$
  - The augmented Lagrangian:

$$(z^*, \phi^*, \eta^*) = \max_{\eta\in L_2^q(\Omega)} \min_{z\in X, \phi\in L_2^q(\Omega)} \quad L_\rho(z,\phi,\eta)$$

$$= \Phi(z) + \lambda\|\phi\|_{H^1} + \langle \eta, \mathcal{D}z - \phi\rangle + \frac{\rho}{2}\|\mathcal{D}z - \phi\|_2^2,$$

    where $\boldsymbol{\phi}(x) = [\phi_1(x), \phi_2(x)]$, $\|\phi\|_{H^1} = \left(\|\phi_1(\mathbf{x})\|_{L_2(\Omega)}^2 + \|\phi_2(\mathbf{x})\|_{L_2(\Omega)}^2\right)^{1/2}$.
  - $g = \mathcal{D}L_\rho(z^k, \phi^*, \eta^*)$

PD-pCN MCMC algorithm

1. $(z^*, \phi^*, \eta^*) = \max_{\eta \in L_2^q(\Omega)} \min_{z \in X, \phi \in L_2^q(\Omega)} \quad L_\rho(z, \phi, \eta)$
2. Let $z^0 = z^*$
3. For $k = 0, 1, 2...$
4. $\quad g = \mathcal{D}L_\rho(z^k, \phi^*, \eta^*)$
5. $\quad$ Propose v using $(2+\delta)v = (2-\delta)z - 2\delta\mathcal{C}_0 g + \sqrt{8\delta}w$
6. $\quad$ Compute accept probability $a(z, v)$[1]
7. $\quad$ if $a \sim U[0,1] < a(z, v)$ then
8. $\qquad z^{k+1} = v$
9. $\quad$ else
10. $\qquad z^{k+1} = z^k$

[1] The accept probability

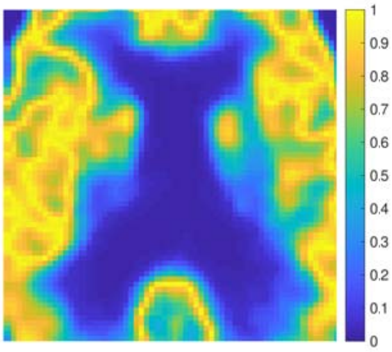
$$a(z, v) = \min\{1, \exp(\rho(z, v) - \rho(v, z))\},$$

where

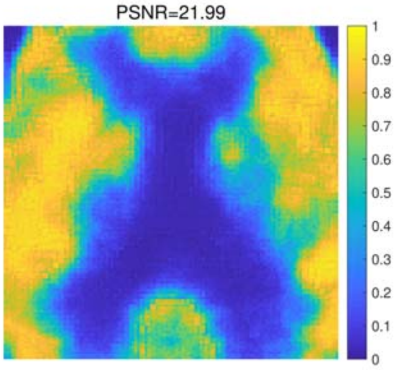
$$\rho(z, v) = \Psi(z) + \frac{1}{2}\langle v - z, g\rangle + \frac{\delta}{4}\langle z + v, g\rangle + \frac{\delta}{4}\|\mathcal{C}_0^{1/2} g\|^2.$$

Inference results

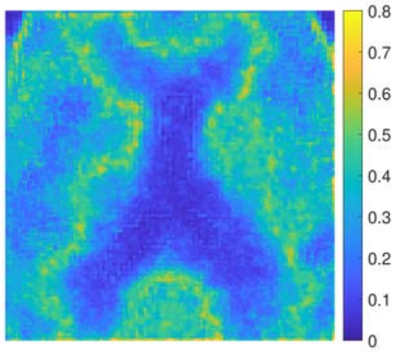
Posterior mean and its highest posterior density interval (HPDI)



truth



posterior mean

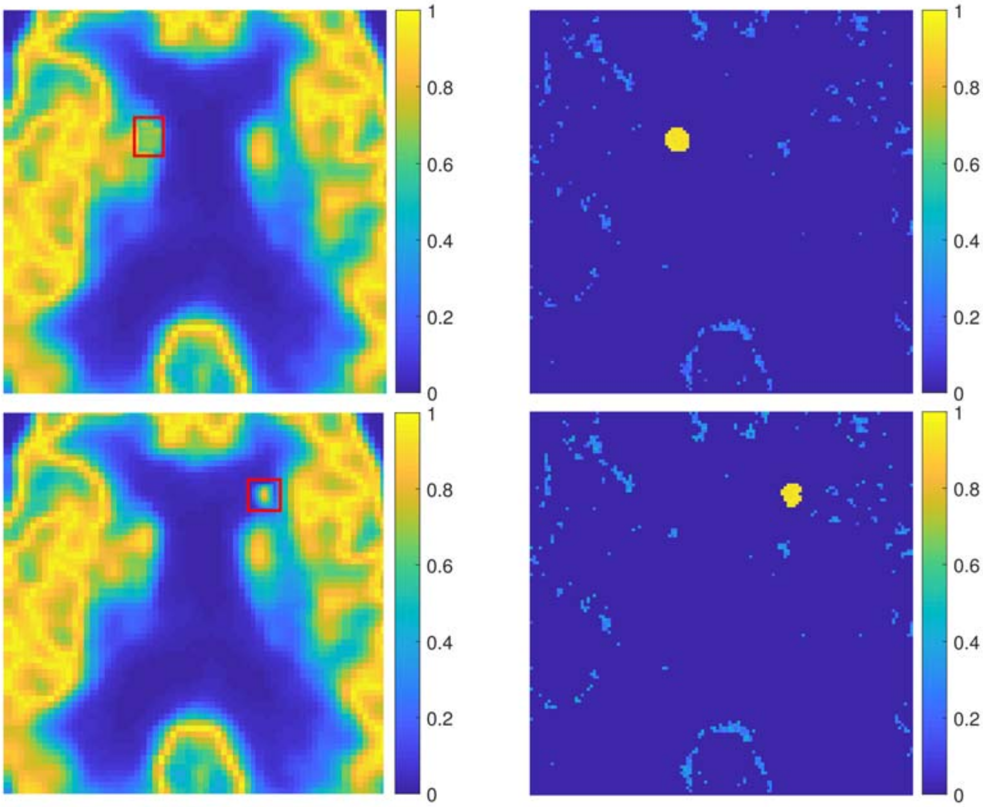


HPDI(95%)

Artifacts detection using posterior distribution

- In image reconstruction, due to the imperfection of the reconstruction method, the recovered image often contains features which are not present in the true imaged object, and such features are called artifacts.
- Let $\pi_{\mathbf{x}}(u(\mathbf{x})|\mathbf{y})$ be the posterior of $u(\mathbf{x})$, π_α is the largest constant satisfying
$$P(u(\mathbf{x})|\pi_{\mathbf{x}}(u(\mathbf{x})|\mathbf{y}) > \pi_\alpha) \geq (1-\alpha).$$
The $100(1-\alpha)\%$ highest posterior density intrval(HPDI) is
$$C_\alpha = \{u(\mathbf{x})|\pi_{\mathbf{x}}(u(\mathbf{x})|\mathbf{y}) > \pi_\alpha\}.$$
- To test a pixel value $\hat{u}(\mathbf{x})$, we compute the smallest HPDI(largest α that contains $\hat{u}(\mathbf{x})$. The larger the HPDI is, the less likely the pixel value is correct.

Artifacts detection examples



test image credible level (1-α)

Determining λ

- Posterior

$$\frac{d\mu^y}{d\mu_0}(z) \propto \exp[-\Phi(z, \mathbf{y}) - \underbrace{\lambda \|z\|_{TV}}_{R(z)}]$$

- The parameter λ controls the strength of the prior.
- Full Bayes: treat λ as additional random variable and infer $\pi(u, \lambda)$
- Empirical Bayes: maximize the marginal likelihood $\max_\lambda \pi(\mathbf{y}|\lambda)$
- Both methods are computationally intractable as we do not know the normalization constant of prior.

The realized discrepancy

- Introduce a function that measures the discrepancy between data $\mathbf{y}$ and the unknown u, say $D(\mathbf{y}, u)$
- For a given u, let $\tilde{\mathbf{y}}$ be data realized from $\pi(\cdot|u)$, we define the probability

$$p_c(\mathbf{y}, u) = \mathbb{P}(D(\mathbf{y}, u) > D(\tilde{\mathbf{y}}, u))$$

  as the likelihood for $\mathbf{y}$ is drawn from $\pi(\cdot|u)$.
- p_c can assess the fitness of a specific value of u to the data
- We choose $D(\mathbf{y}, \psi) = \sum_{i=1}^{n} \frac{(y_i - \psi_i)^2}{\psi_i^2}$, then

$$p_c(\mathbf{y}, \psi) = P(\chi_n^2 > D(\mathbf{y}, \psi))$$

Posterior predictive p-value

- The posterior predictive p-value is define as the average of realized discrepancy over all possible u:

$$p_b(\mathbf{y}) = \int p_c(\mathbf{y}, u)\pi(u|\mathbf{y})du.$$

- The posterior predictive p_b assess the fitness between the posterior distribution and the data $\mathbf{y}$.
- p_b is too small: the data is not well fitted(underfitting)
- p_b is too big: the posterior fits the data "too well"(overfitting)
- We should choose λ in a way that the effects of prior and data are well balanced, which is indicated by an appropriate value of p_b.
- The posterior predictive p_b and the PSNR of the resulting MAP

λ	0	0.5	1	2	3	4
p_b	0.99	0.82	0.28	0.04	0.0034	0.0005
PSNR	18.21	21.90	22.0	20.66	19.76	19.27

Theorem results

- μ^y is a well-defined probability measure on X.
- μ^y is Lipschitz in the data y, with respect to the Hellinger distance: if μ^y and $\mu^{y'}$ are two measures corresponding to data y and y', then there exists $C = C(r)$ such that, for all y, y' with $\max\{\|y\|_2, \|y'\|_2\} < r$,

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

theorem results(cont'd)

- Let

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

where $\Phi_{N_1}(z)$ is a $N_1 \in \mathbb{N}$ dimensional approximation of $\Phi(z)$ and $R_{N_2}(z)$ is a $N_2 \in \mathbb{N}$ dimensional approximation of $R(z)$. Assume that Φ_{N_1} satisfies Proposition 5.1, and R_{N_2} satisfis Assumptions A.2 (i) and (ii) in (Yao et al, IP, 2016). Assume also that for any $\epsilon > 0$, there exist two positive sequences $\{a_{N_1}(\epsilon)\}$ and $\{b_{N_2}(\epsilon)\}$ converging to zero, such that $\mu_0(X_\epsilon) \geq 1 - \epsilon$ for any $N_1, N_2 \in \mathbb{N}$, where

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

Then we have

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

Summary

We provide a complete framework for performing infinite dimensional and uncertainty quantification Bayesian inference for image reconstruction with Poisson data.

1. Computation: dimensional-independent sampling method
2. Theorem: well-defined posterior in functional space
3. Application: artifacts detection

Reference

- Yao Z, *et al.*. A TV-Gaussian prior for infinite-dimensional Bayesian inverse problems and its numerical implementations[J]. *Inverse Problems*, (2016).

- Zhou Q, Yu T, Zhang X, *et al.* Bayesian Inference and Uncertainty Quantification for Medical Image Reconstruction with Poisson Data[J]. SIAM Journal on Imaging Sciences, 2020, 13(1): 29-52.

Thank you for your attention