Bayesian Inference with Projected Densities

Martin S. Andersen

Department of Applied Mathematics and Computer Science
Technical University of Denmark

joint work with Jasper Everink and Yiqiu Dong

ELLIIT Focus Period Linköping 2022
November 8, 2022

Linear Inverse Problems

$$b = Ax + e$$

- observations $b \in \mathbb{R}^m$
- model/design/regressor matrix $A \in \mathbb{R}^{m \times n}$
- parameter vector $x \in \mathbb{R}^n$
- random measurement noise $e \in \mathbb{R}^m$

Bayesian approach

$$\underbrace{\pi(x|b)}_{\text{posterior}} \propto \underbrace{\pi(b|x)}_{\text{likelihood}} \underbrace{\pi(x)}_{\text{prior}}$$

What is a suitable noise model? What is a suitable prior?

Linear Inverse Problems (cont.)

Gaussian model

- Gaussian noise: $e \sim \mathcal{N}(0, \Sigma_e)$ with $\Sigma_e^{-1} = \lambda\, I$
- Gaussian prior: $x \sim \mathcal{N}(0, \Sigma_x)$ with $\Sigma_x^{-1} = \delta\, L^T L$
- Gaussian posterior

$$\pi(x|b) = \mathcal{N}(\hat{x}, \Sigma_{x|b}) \;\propto\; \exp\left(-\frac{\lambda}{2}\|Ax - b\|_2^2 - \frac{\delta}{2}\|Lx\|_2^2\right)$$

  where $\Sigma_{x|b} = (\delta L^T L + \lambda A^T A)^{-1}$ and $\hat{x} = \lambda \Sigma_{x|b} A^T b$
- posterior mean is maximum a posteriori estimate
- Bayesian perspective: $\pi(x|b)$ is *the* solution

Sampling from the posterior distribution

Objective: compute samples from posterior distribution $\mathcal{N}(\hat{x}, \Sigma_{x|b})$

- Direct method: factorize $\Sigma_{x|b} = R^T R$ and compute

$$x = \hat{x} + R^T y, \quad y \sim \mathcal{N}(0, I)$$

- Randomize-then-optimize: generate $\hat{b} \sim \mathcal{N}(b, \lambda^{-1} I)$ and $\hat{c} \in \mathcal{N}(0, \delta^{-1} I)$ and compute

$$x = \operatorname*{argmin}_u \left\{ \frac{\lambda}{2} \|Au - \hat{b}\|_2^2 + \frac{\delta}{2} \|Lu - \hat{c}\|_2^2 \right\}$$

Markov chain Monte Carlo and Plug & Play priors

- Unadjusted Langevin algorithm: samples from approximate posterior

$$x_{k+1} = x_k + \gamma \nabla \log \pi(y|x_k) + \gamma \nabla \log \pi(x_k) + \sqrt{2\gamma} u_{k+1}, \qquad u_{k+1} \in \mathcal{N}(0, I)$$

- Apply Moreau–Yosida regularization if prior is nonsmooth:

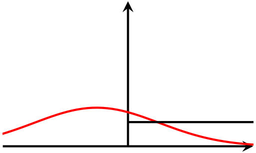
$$x_{k+1} = x_k + \gamma \nabla \log \pi(y|x_k) + \frac{\gamma}{\eta} [\mathrm{prox}_U^{\eta}(x_k) - x_k] + \sqrt{2\gamma} u_{k+1}, \qquad u_{k+1} \in \mathcal{N}(0, I)$$

- Replace $\eta^{-1}[\mathrm{prox}_U^{\eta}(x_k) - x_k]$ by $\epsilon^{-1}[D_\epsilon(x_k) - x_k]$ where D_ϵ is a neural network (Laumont et al. 2022)

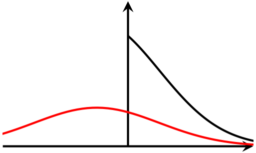
Additional prior information

Suppose x is known to be nonnegative, *i.e.*, $x \in \mathbb{R}^n_+$. What is a suitable prior?

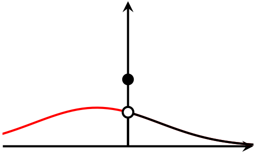
Uniform



Truncated Gaussian



Projected Gaussian



Sampling from projected Gaussian density

Oblique projection of $\pi_x = \mathcal{N}(\mu, \Sigma)$ onto a closed, convex set $C \subset \mathbb{R}^n$

$$z = \Pi_C^{\Sigma^{-1}}(x) = \operatorname*{argmin}_{u \in C} \frac{1}{2}\|u - x\|_{\Sigma^{-1}}^2, \quad x \sim \mathcal{N}(\mu, \Sigma)$$

- z is well-defined since Σ is positive definite and $\Pi_C^{\Sigma^{-1}}$ is continuous

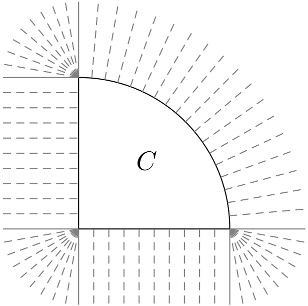
$$\mathbb{P}(\Pi_C^{\Sigma^{-1}}(x) \in E) = \mathbb{P}\left(x \in \left[\Pi_C^{\Sigma^{-1}}\right]^{-1}(E)\right)$$

- optimality condition for projection problem can be expressed as

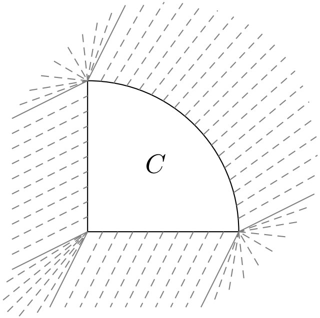
$$x \in z + \Sigma N_C(z)$$

Oblique projection

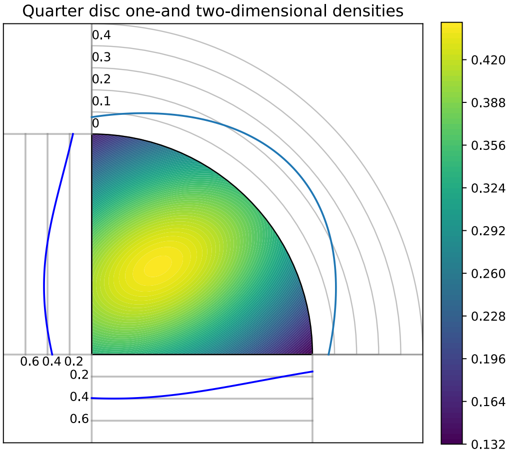
$z + N_C(z)$



$z + \Sigma N_C(z)$



Example


Quarter disc one-and two-dimensional densities
0.4
0.3
0.2
0.1
0
0.6 0.4 0.2
0.2
0.4
0.6
0.420
0.388
0.356
0.324
0.292
0.260
0.228
0.196
0.164
0.132

Non-empty, closed, convex set

- Positive probability on boundary and relative interior

$$\mathbb{P}(\Pi_C^{\Sigma^{-1}}(x) \in \mathbf{bd}(C)) > 0 \quad \text{and} \quad \mathbb{P}(\Pi_C^{\Sigma^{-1}}(x) \in \mathbf{relint}(C)) > 0$$

- Posterior mean is in the relative interior of C

$$\mathbb{E}[\Pi_C^{\Sigma^{-1}}(x)] \in \mathbf{relint}(C)$$

- MAP estimate is undefined

Polyhedral set

Suppose C is a polyhedral set and F is a face of C.

- For any measurable set $E \subseteq \mathbf{relint}(F)$, we have

$$\mathbb{P}\left(\Pi_C^{\Sigma^{-1}}(x) \in E\right) = \int_E \pi_F(z)\, dz$$

  where $\pi_F(z)$ is the $\dim(F)$-dimensional density

$$\pi_F(z) = \int_{\Sigma N_C(z)} \pi_x(z+v)\, dv.$$

- Projected density $\Pi_C^{\Sigma^{-1}}(x)$ is proportional to the density of x on the relative interior of any face of C.

Projected Gaussian posterior

Randomize-then-optimize: generate $\hat{b} \sim \mathcal{N}(b, \lambda^{-1} I)$ and $\hat{c} \in \mathcal{N}(0, \delta^{-1} I)$ and compute

$$x = \operatorname*{argmin}_{u \in C} \left\{ \frac{\lambda}{2} \|Au - \hat{b}\|_2^2 + \frac{\delta}{2} \|Lu - \hat{c}\|_2^2 \right\}$$

- every sample requires the solution of randomized constrained LS problem
- posterior is well-defined, but what does the prior look like?
- suppose C is polyhedral and F is a face of C

$$\pi_{x,F}(u) \propto \pi_x(x_0 + Cu), \qquad x_0 + Cu \in \mathbf{relint}\, F$$

Hierarchical Bayesian model

Hierarchical Bayesian model with hyper priors $\lambda \sim \Gamma(\alpha_\lambda, \beta_\lambda)$ and $\delta \sim \Gamma(\alpha_\delta, \beta_\delta)$

$$
\begin{aligned}
\pi(\lambda) &\propto \lambda^{\alpha_\lambda - 1} \exp(-\beta_\lambda \lambda), \qquad \text{for } \lambda > 0 \\
\pi(\delta) &\propto \delta^{\alpha_\delta - 1} \exp(-\beta_\delta \delta), \qquad \text{for } \delta > 0
\end{aligned}
$$

Full posterior

$$
\begin{aligned}
\pi_{x,\lambda,\delta|b}(x, \lambda, \delta) \propto\ & \lambda^{m/2+\alpha_\lambda-1} \delta^{n/2+\alpha_\delta-1} \\
& \times \exp\left(-\frac{\lambda}{2} \|Ax - b\|_2^2 - \frac{\delta}{2} \|Lx\|_2^2 - \beta_\lambda \lambda - \beta_\delta \delta \right),
\end{aligned}
$$

Polyhedral Cone Hierarchical Gibbs Sampler

Initialization: Choose $(x^0, \alpha_\lambda, \beta_\lambda, \alpha_\delta, \alpha_\lambda)$.
For $k = 1, 2, \ldots$
Compute $(\lambda_k, \delta_k) \sim \pi_{\lambda,\delta|x,b}$ as follows:
$\lambda_k \sim \Gamma\left(m/2 + \alpha_\lambda, \frac{1}{2}\|Ax_{k-1} - b\|_2^2 + \beta_\lambda\right)$
$\delta_k \sim \Gamma\left(\mathsf{dim}(F(x_{k-1}))/2 + \alpha_\delta, \frac{1}{2}\|Lx_{k-1}\|_2^2 + \beta_\delta\right)$
Compute $x_k \sim \pi_{x|b,\lambda,\delta}$ by solving instance of:

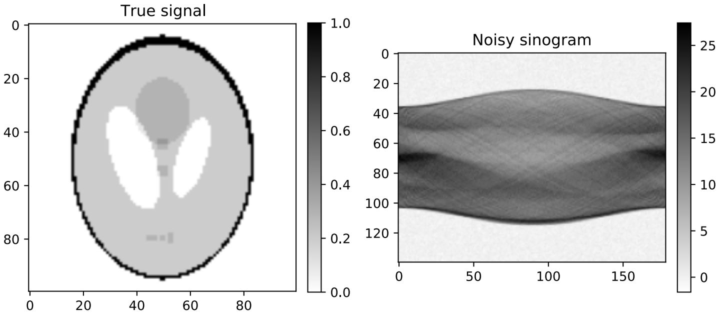
$$x_k = \operatorname*{argmin}_{u \in C} \left\{ \frac{\lambda_k}{2}\|Au - \hat{b}\|_2^2 + \frac{\delta_k}{2}\|Lu - \hat{c}\|_2^2 \right\}$$

End.

Extension: convex regularization functions with polyhedral epigraph

Numerical example: CT reconstruction

$$b = Ax + e$$



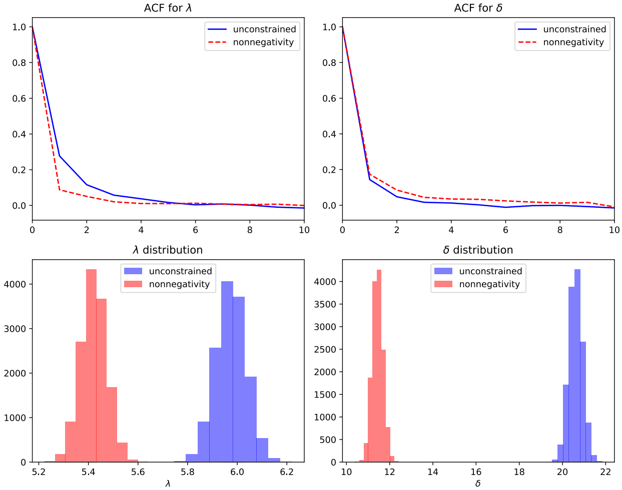
Numerical example: CT reconstruction (cont.)

Model

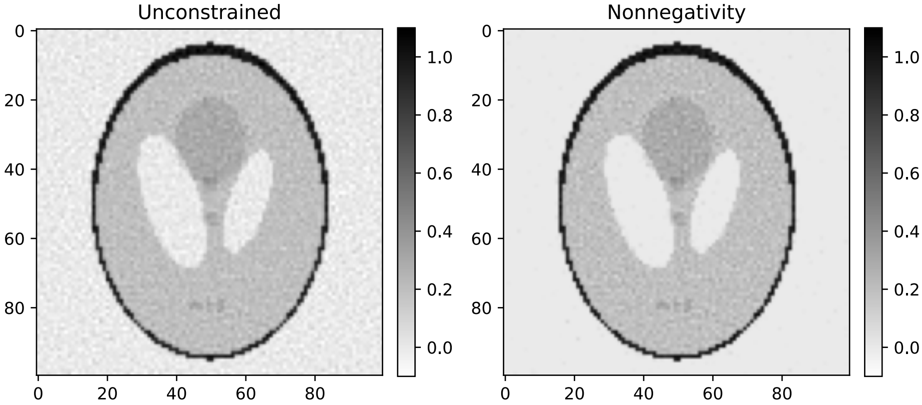
- L is first-order diff.
- $C = \mathbb{R}^n_+$

Gibbs sampler

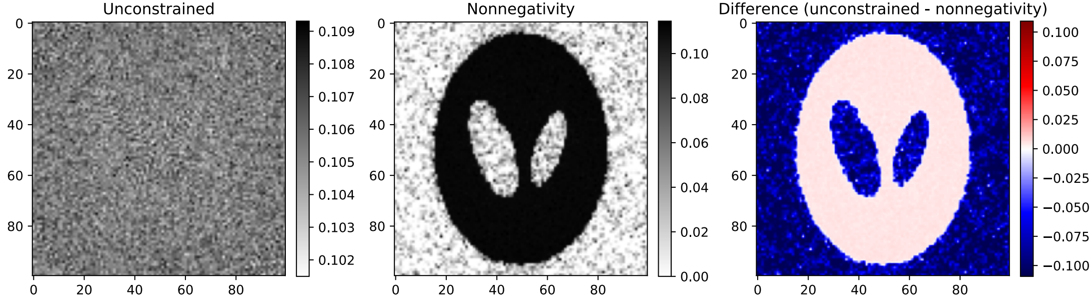
- 15000 samples (1000 burn-in)
- 1000 iterations with APG



Numerical example: CT reconstruction (cont.)



Numerical example: CT reconstruction (cont.)



Summary

- implicit prior defined through projected posterior
- positive probability on boundary
- Gibbs sampler when the set C is polyhedral
- preprint available on arXiv [2209.12481]

References

Bangsgaard, Katrine O and Martin S Andersen (2021). "A statistical reconstruction model for absorption CT with source uncertainty". In: *Inverse Problems* 37.8, p. 085009.

Bardsley, Johnathan M and Colin Fox (2012). "An MCMC method for uncertainty quantification in nonnegativity constrained inverse problems". In: *Inverse Problems in Science and Engineering* 20.4, pp. 477–498.

Bardsley, Johnathan M and Per Christian Hansen (2020). "MCMC algorithms for computational UQ of nonnegativity constrained linear inverse problems". In: *SIAM Journal on Scientific Computing* 42.2, A1269–A1288.

Beck, Amir and Marc Teboulle (2009). "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". In: *SIAM journal on imaging sciences* 2.1, pp. 183–202.

Everink, J. M., Y. Dong, and M. S. Andersen (2022). *Bayesian Inference with Projected Densities*. eprint: `arXiv:2209.12481`.

Geweke, John F (1996). "Bayesian inference for linear models subject to linear inequality constraints". In: *Modelling and Prediction Honoring Seymour Geisser*. Springer, pp. 248–263.

Hansen, Per Christian, Jakob Jørgensen, and William RB Lionheart (2021). *Computed Tomography: Algorithms, Insight, and Just Enough Theory*. SIAM.

Hansen, Per Christian and Jakob Sauer Jørgensen (2018). "AIR Tools II: algebraic iterative reconstruction methods, improved implementation". In: *Numerical Algorithms* 79.1, pp. 107–137.

References (cont.)

Laumont, R. et al. (May 2022). "Bayesian Imaging Using Plug & Play Priors: When Langevin Meets Tweedie". In: *SIAM Journal on Imaging Sciences* 15.2, pp. 701–737. DOI: 10.1137/21m1406349.