Computationally efficient inference for sparsity-promoting hierarchical Bayesian models[a]

Bayesian inverse problems and UQ @ ICERM

Jan Glaubitz

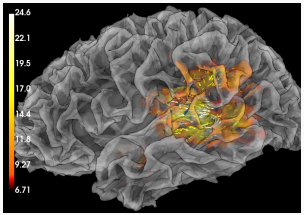
Joint work with Youssef Marzouk (MIT)

March 04, 2026

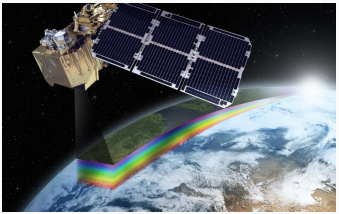
Department of Mathematics
Linköping University
Linköping, Sweden

[a]This work was supported by the DOD (ONR MURI) grant #N00014-20-1-2595, the Swedish Research Council (VR) Starting Grant #2025-05370, the Zenith Career Development Grant #26.07, and the National Academic Infrastructure for Supercomputing in Sweden (NAISS) grants #2024/22-1207 and #2025/22-1599.

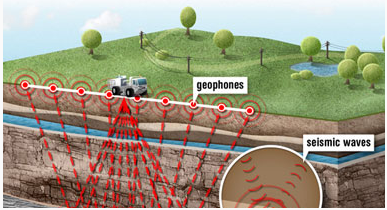
Inverse problem $\mathbf{y} = F(\mathbf{x}) + \mathbf{e}$



Medical imaging (MEG, CT, MRI)

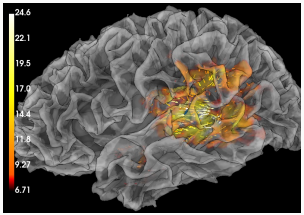


Remote sensing

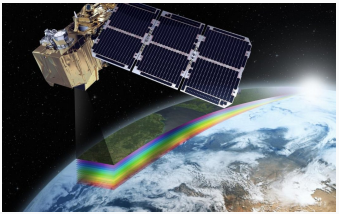


Geoscience

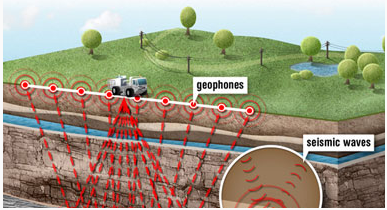
Inverse problem $\mathbf{y} = F(\mathbf{x}) + \mathbf{e}$



Medical imaging (MEG, CT, MRI)



Remote sensing

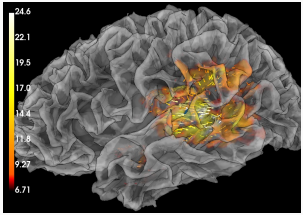


Geoscience

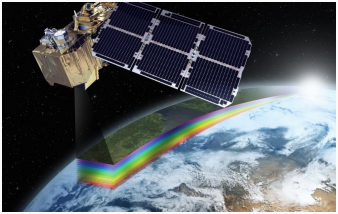
Belief. $\mathbf{x}$ or $\Psi\mathbf{x}$ is sparse

Bayesian inverse problems & Sparsity

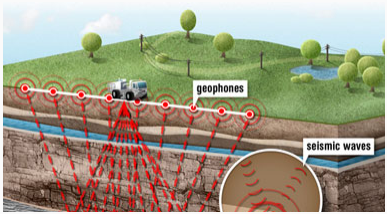
Inverse problem $\mathbf{y} = F(\mathbf{x}) + \mathbf{e}$



Medical imaging (MEG, CT, MRI)



Remote sensing



Geoscience

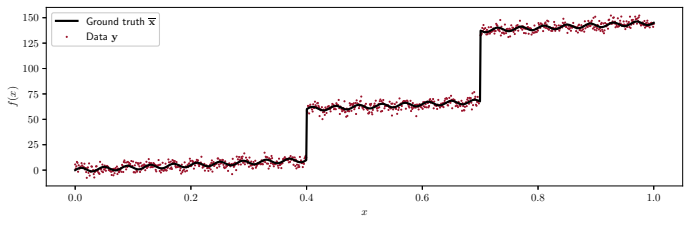
Belief. $\mathbf{x}$ or $\Psi\mathbf{x}$ is sparse

Bayesian approach[1,2]: $\pi^{y}(\mathbf{x}) \propto f(\mathbf{x};\mathbf{y})\ \pi^{0}(\mathbf{x})$

[1] Stuart, "Inverse problems: a Bayesian perspective," 2010

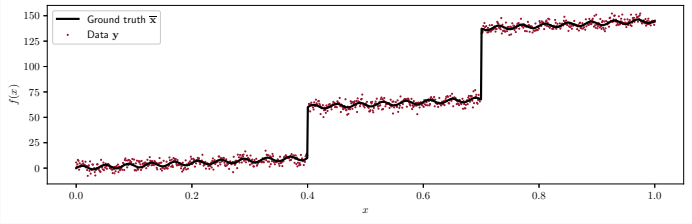
[2] Calvetti & Somersalo, "Bayesian Scientific Computing," 2023

Hierarchical sparsity-promoting priors



Assumption: $\Psi \mathbf{x} \in \mathbb{R}^K$ is sparse—e.g., $[\Psi \mathbf{x}]_k = x_{k+1} - x_k$

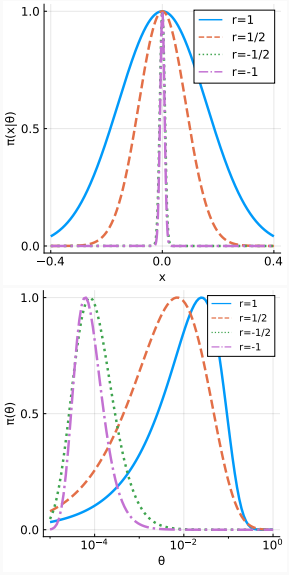
Hierarchical sparsity-promoting priors



Assumption: $\Psi \mathbf{x} \in \mathbb{R}^K$ is sparse—e.g., $[\Psi \mathbf{x}]_k = x_{k+1} - x_k$

Sparse Bayesian learning (SBL) priors[ab]

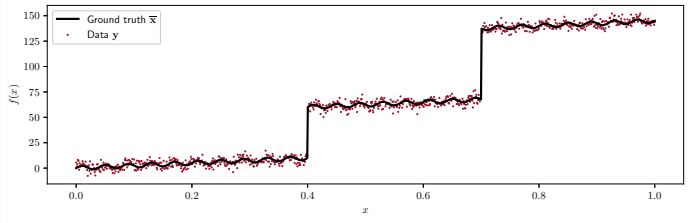
$$\begin{cases} [\Psi \mathbf{x}]_k \sim \mathcal{N}(0, \theta_k), \\ \theta_k \sim \mathcal{GG}(r, \beta, \vartheta), \end{cases} \qquad k = 1, \ldots, K$$



[a] Tipping, "Sparse Bayesian learning and the relevance vector machine," 2001

[b] Calvetti, Pragliola, Somersalo, Strang, "Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors," 2020

Hierarchical sparsity-promoting priors

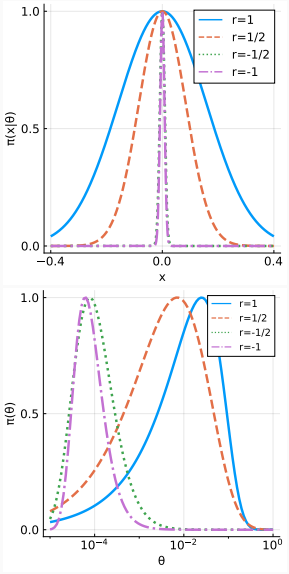


Assumption: $\Psi x \in \mathbb{R}^K$ is sparse—e.g., $[\Psi x]_k = x_{k+1} - x_k$

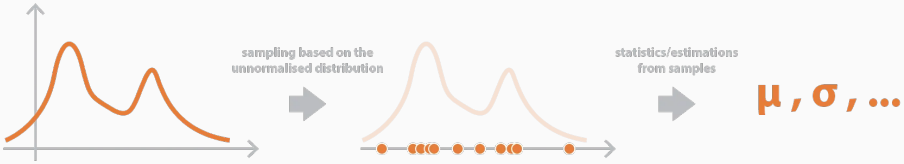
Sparse Bayesian learning (SBL) priors[ab]

$$\begin{cases} [\Psi x]_k \sim \mathcal{N}(0, \theta_k), \\ \theta_k \sim \mathcal{GG}(r, \beta, \vartheta), \end{cases} \qquad k = 1, \ldots, K$$

- $\mathcal{GG}$ makes $\theta_k \approx 0$ likely, while allowing for outliers
- If $\theta_k \approx 0$, then $[\Psi x]_k \approx 0$ is most likely → no jump!
- If $\theta_k \gg 0$, then $|[\Psi x]_k| \gg 0$ more likely → jump!



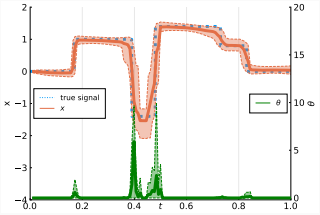
Sampling & UQ



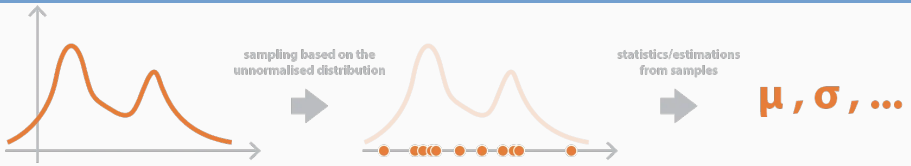
Goal. Sample from $\pi^y(\mathbf{x},\boldsymbol{\theta}) \propto f(\mathbf{x};\mathbf{y})\pi^0(\mathbf{x},\boldsymbol{\theta})$

Enables approxating posterior expectations:

$$\int \mathcal{G}(\mathbf{x},\boldsymbol{\theta})\pi^y(\mathbf{x},\boldsymbol{\theta})\mathrm{d}(\mathbf{x},\boldsymbol{\theta}) \approx \frac{1}{I}\sum_{i=1}^{I}\mathcal{G}(\mathbf{x}_i,\boldsymbol{\theta}_i)$$



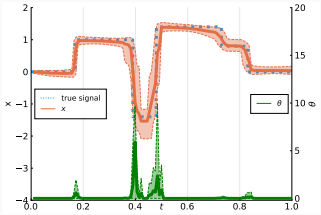
Sampling & UQ



Goal. Sample from $\pi^y(\mathbf{x}, \boldsymbol{\theta}) \propto f(\mathbf{x}; \mathbf{y})\pi^0(\mathbf{x}, \boldsymbol{\theta})$

Enables approxating posterior expectations:

$$\int \mathcal{G}(\mathbf{x}, \boldsymbol{\theta})\pi^y(\mathbf{x}, \boldsymbol{\theta})\mathrm{d}(\mathbf{x}, \boldsymbol{\theta}) \approx \frac{1}{I}\sum_{i=1}^{I} \mathcal{G}(\mathbf{x}_i, \boldsymbol{\theta}_i)$$

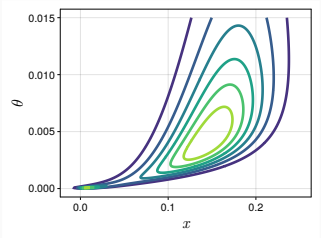


Challenges for sparsity-promoting priors.[a]

(i) multi-modal

(ii) strong correlation between $\mathbf{x}$ and $\boldsymbol{\theta}$

(iii) high-dimensional

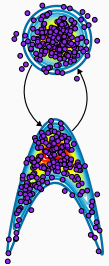


[a] Calvetti, Somersalo, "Computationally efficient sampling methods for sparsity promoting hierarchical Bayesian models," 2024

Our approach: Hierarchical prior normalization[3]

Idea. Transform the SBL prior $\pi^0(\mathbf{x}, \boldsymbol{\theta})$ into a standard normal one $\phi^0(\mathbf{u}, \boldsymbol{\tau})$

Need. $S : (\mathbf{x}, \boldsymbol{\theta}) \mapsto (\mathbf{u}, \boldsymbol{\tau})$ s. t. $(\mathbf{u}, \boldsymbol{\tau})$ follows ϕ^0 whenever $(\mathbf{x}, \boldsymbol{\theta})$ follows π^0



[3] G. & Marzouk, "Efficient sampling for sparse Bayesian learning using hierarchical prior normalization," 2025

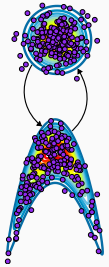
Our approach: Hierarchical prior normalization[3]

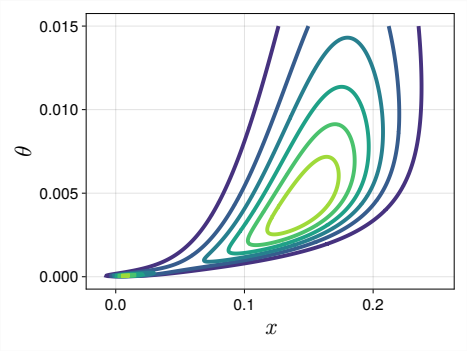
Idea. Transform the SBL prior $\pi^0(\mathbf{x}, \boldsymbol{\theta})$ into a standard normal one $\phi^0(\mathbf{u}, \boldsymbol{\tau})$

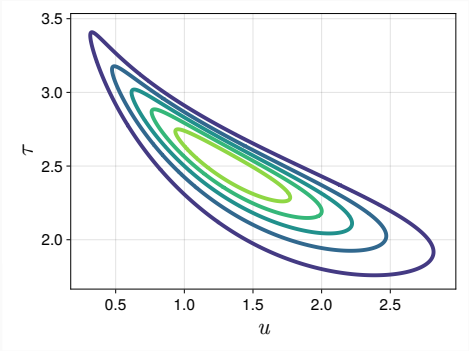
Need. $S : (\mathbf{x}, \boldsymbol{\theta}) \mapsto (\mathbf{u}, \boldsymbol{\tau})$ s. t. $(\mathbf{u}, \boldsymbol{\tau})$ follows ϕ^0 whenever $(\mathbf{x}, \boldsymbol{\theta})$ follows π^0

Get. Simpler "prior-normalized" posterior

$$\phi^y(\mathbf{u}, \boldsymbol{\tau}) = (S_{\#}\pi^y)(\mathbf{u}, \boldsymbol{\tau}) = \frac{1}{Z} f(S^{-1}(\mathbf{u}, \boldsymbol{\tau}); \mathbf{y})\, \phi^0(\mathbf{u}, \boldsymbol{\tau})$$







[3] G. & Marzouk, "Efficient sampling for sparse Bayesian learning using hierarchical prior normalization," 2025

Product-like form. ($\Psi = I$, i.e., $\mathbf{x}$ sparse)

$$\pi^0(\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \mathcal{N}(x_i|0, \theta_i)\mathcal{GG}(\theta_i|r, \beta, \vartheta) \implies S(\mathbf{x}, \boldsymbol{\theta}) = \begin{bmatrix} s_1(x_1, \theta_1) \\ \vdots \\ s_n(x_n, \theta_n) \end{bmatrix}$$

with $s_i : (x_i, \theta_i) \mapsto (u_i, \tau_i)$ transforming $\mathcal{N}(x_i|0, \theta_i)\mathcal{GG}(\theta_i|r, \beta, \vartheta)$ into $\mathcal{N}(\mathbf{0}_2, I_2)$

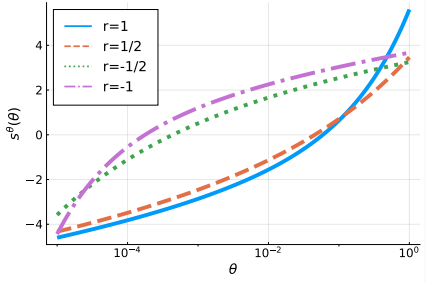
Product-like form. ($\Psi = I$, i.e., $\mathbf{x}$ sparse)

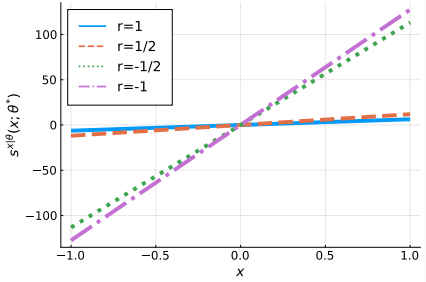
$$\pi^0(\mathbf{x},\boldsymbol{\theta}) = \prod_{i=1}^{n} \mathcal{N}(x_i|0,\theta_i)\mathcal{GG}(\theta_i|r,\beta,\vartheta) \implies S(\mathbf{x},\boldsymbol{\theta}) = \begin{bmatrix} s_1(x_1,\theta_1) \\ \vdots \\ s_n(x_n,\theta_n) \end{bmatrix}$$

with $s_i : (x_i,\theta_i) \mapsto (u_i,\tau_i)$ transforming $\mathcal{N}(x_i|0,\theta_i)\mathcal{GG}(\theta_i|r,\beta,\vartheta)$ into $\mathcal{N}(\mathbf{0}_2, I_2)$

Knothe–Rosenblatt (KR) rearrangements.

$$s(x,\theta) = \begin{bmatrix} s^{\theta}(\theta) \\ s^{x|\theta}(x;\theta) \end{bmatrix}, \quad \begin{cases} s^{\theta} = \left(\Psi^0\right)^{-1} \circ \mathcal{P}^{\theta}, \\ s^{x|\theta} = x/\sqrt{\theta}. \end{cases}$$

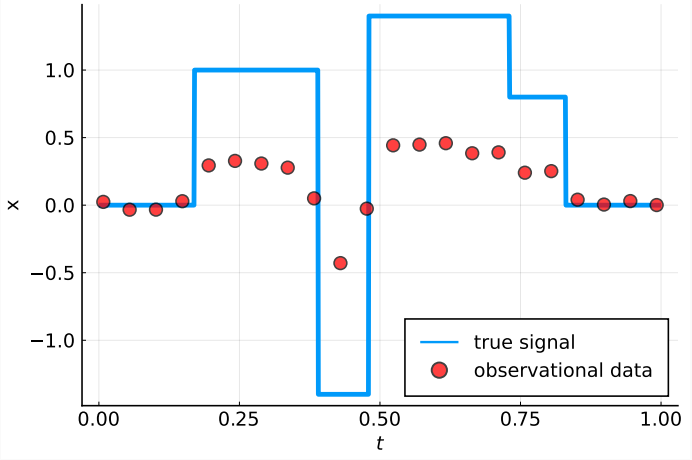




Example 1: Signal deblurring

Goal. Recover $n = 128$ nodal values from $m = 22$ noisy blurry observations:

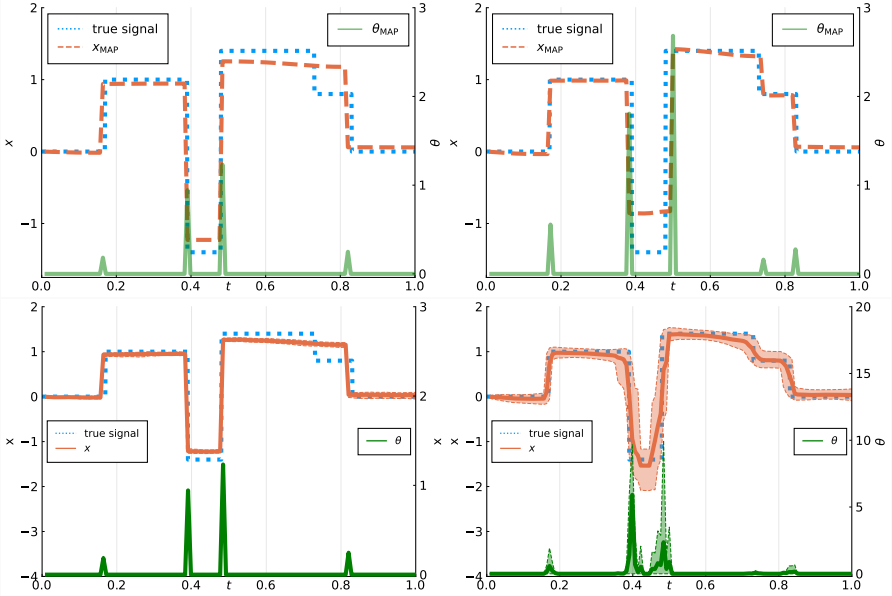
$$\mathbf{y} = F\mathbf{x} + \mathbf{e} \quad \text{where} \quad [\Psi\mathbf{x}]_i = x_{i+1} - x_i \text{ is sparse}$$



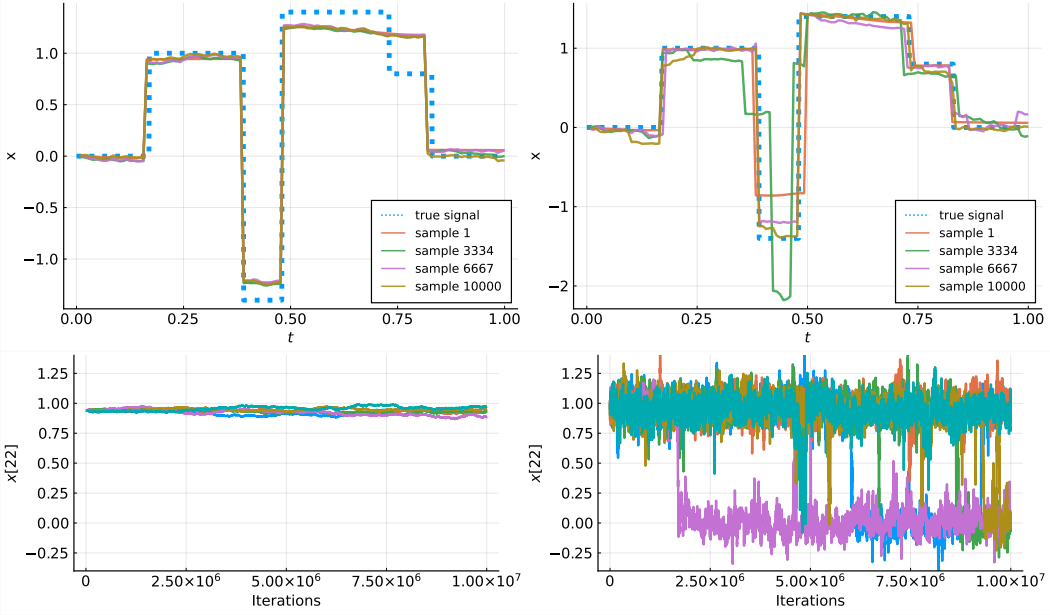
Compare. AM sampler for original and prior-normalized posterior with $r = -1$

Statistics (MAP initialization)

Setup. $r = -1$, $J = 6$ chains, 10^7 samples, MAP initialization

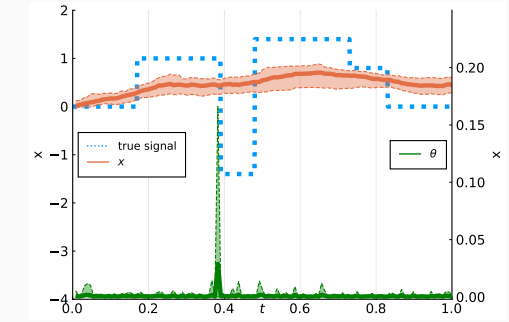


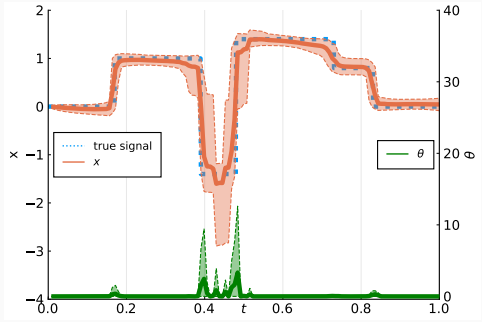
Samples & traces (MAP initialization)

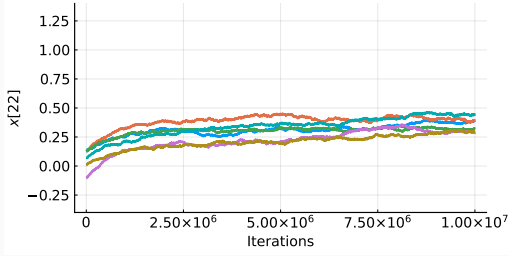


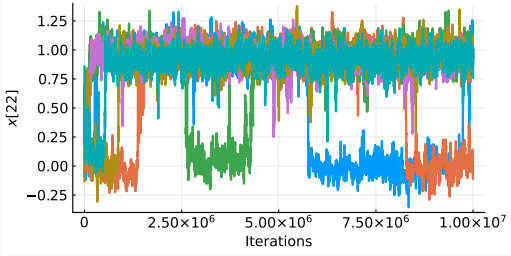
Statistics & samples (random initialization)

Setup. $r = -1$, $J = 6$ chains, 10^7 samples, random initialization

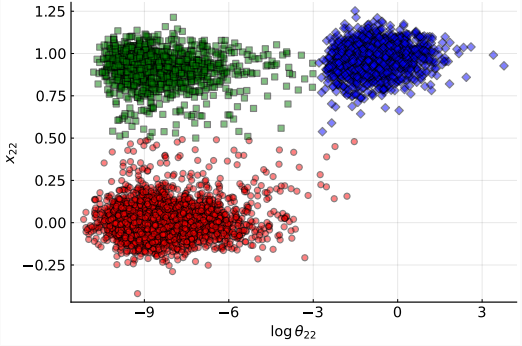




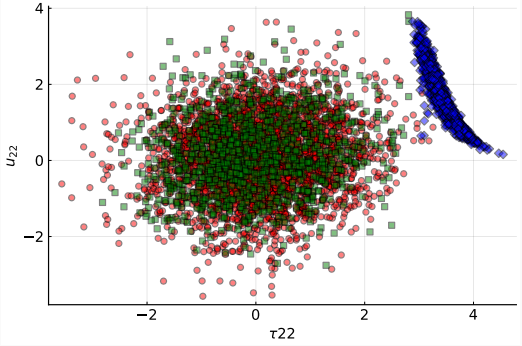




Scatter plots (random initialization)



(a) Original posterior



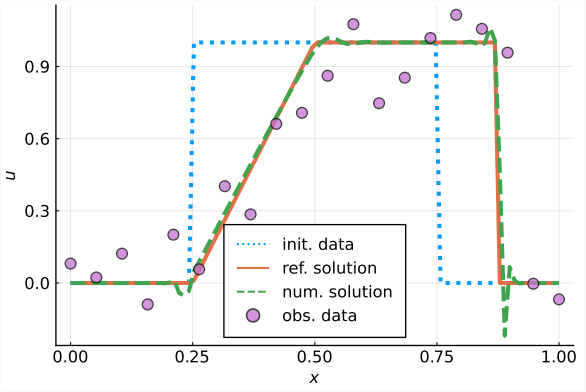
(b) Prior-normalized posterior

Computational example II: Inverse Burgers' equation

Burgers' equation. $\partial_t u(x,t) + \partial_x u(x,t)^2 = 0$ with $u(x,0) = u_0(x)$

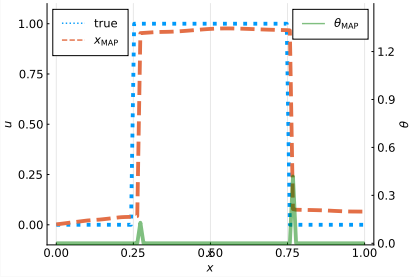
Goal. Recover u_0 at $n = 100$ points from $m = 20$ noisy solution observations:

$$\mathbf{b} = \mathcal{F}(\mathbf{u}) + \mathbf{e} \quad \text{where} \quad [\Psi \mathbf{u}]_i = u_{i+1} - u_i \text{ is sparse}$$

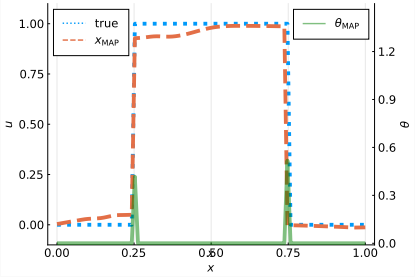


Compare. AM sampler (10^7) for original and prior-normalized posterior, $r = -1$

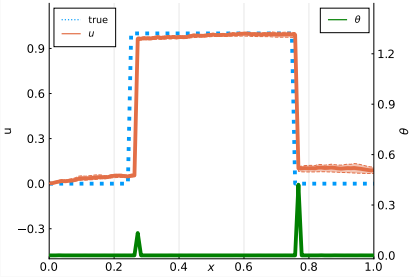
MAP estimates & Statistics



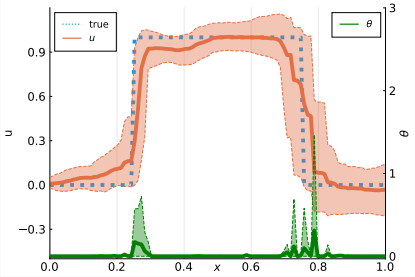
Original posterior, MAP estimate



Prior-normalized posterior, MAP estimate

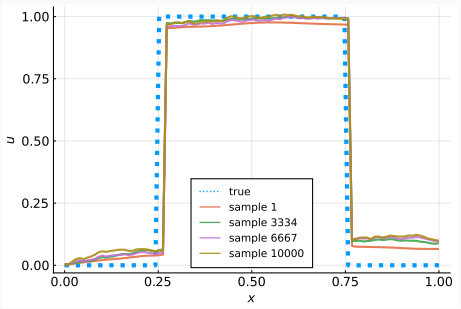


Original posterior, statistics

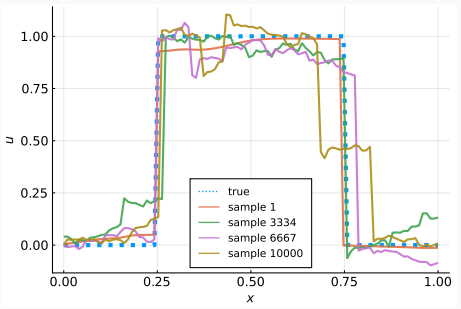


Prior-normalized posterior, statistics

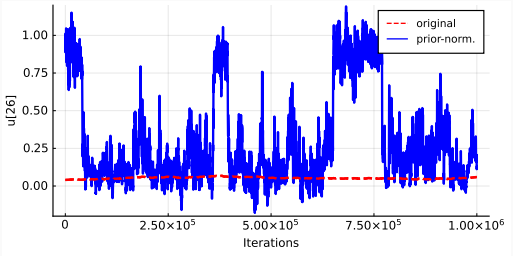
Samples & traces: MAP initialized



Original posterior, samples

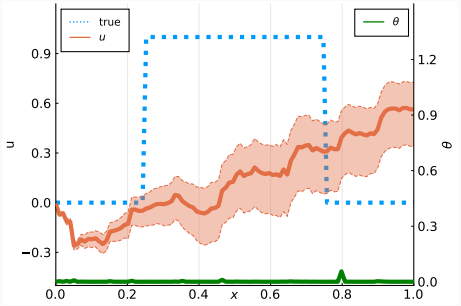


Prior-normalized posterior, samples

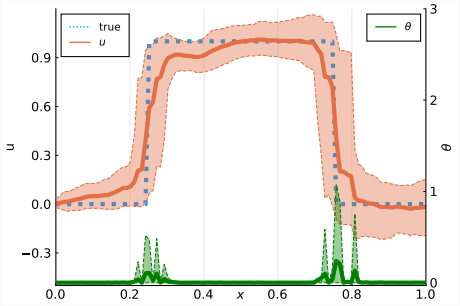


Traces

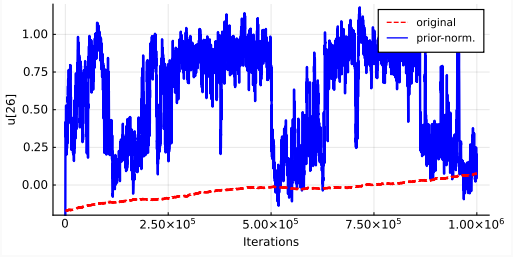
Statistics & samples: Randomly initialized



Original posterior, statistics



Prior-normalized posterior, statistics

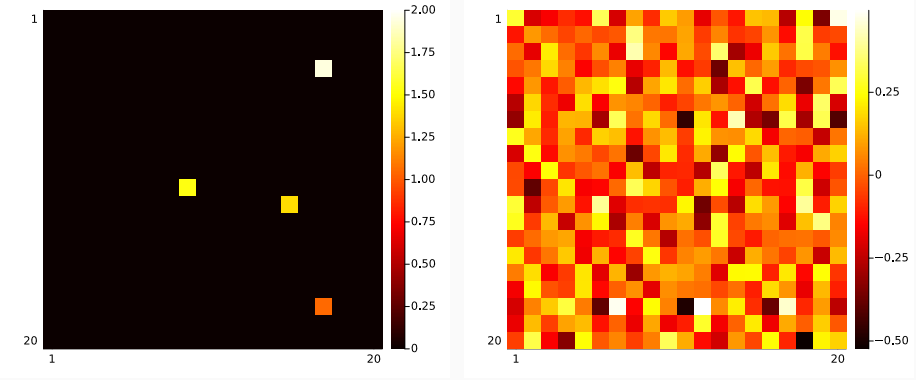


Traces

Example 3: Impulse image

Goal. Recover 20×20 impulse image from its noisy DCT:

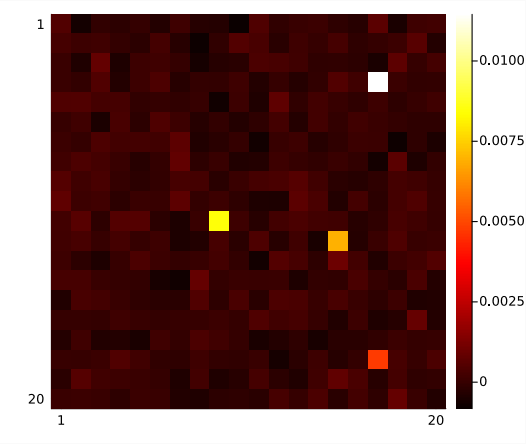
$$\mathbf{y} = F\mathbf{x} + \mathbf{e} \quad \text{where} \quad \mathbf{x} \text{ is sparse}$$



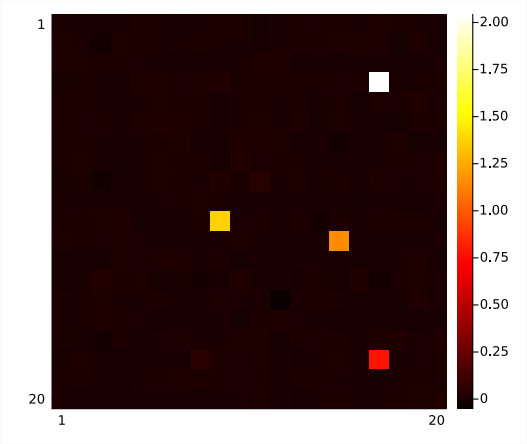
Setup. $r = -1$, $J = 4$ chains, random initialization

Compare. Original posterior (Gibbs sampler, 10^5 samples, $\approx$ 20m runtime) vs prior-normalized posterior (ESS, 10^4, $\approx$ 3m runtime)

Statistics – x-mean



(a) Original posterior



(b) Prior-normalized posterior

Outlook

- Generally applies to scale-mixtures-of-normal priors [**Jonathan Lindbloom**]
- Combine with geometry-exploiting (dimension-robust) samplers
- High-dimensional problems (dynamic inverse problems) [**Mirjeta Pasha**]
- Chemical reaction neural network [**Federica Milinanni**]
- Use as a preconditioner—combined with, e.g., flow matching or diffusion
- What else?

For more details:

Glaubitz and Marzouk. "Efficient sampling for sparse Bayesian learning using hierarchical prior normalization." To appear in SIAM-ASA UQ (2026).

Thank you & Acknowledgments

The End

Thank You!

VR Starting Grant #2025-05370

Zenith Career Development Grant

NAISS grant #2024/22-1207