Preprint on constraints available: Bayesian inference with projected densities

Sparse Bayesian Inference with Regularized Gaussians

Jasper M. Everink, Technical University of Denmark, jmev@dtu.dk

Joint work with Martin S. Andersen and Yiqiu Dong

DTU Compute
Department of Applied Mathematics and Computer Science

Sparsity

$$\{x \,|\, x_i = 0, i \in I\}$$

$$\{x \,|\, x_i - x_{i+1} = 0, i \in I\}$$

The importance of
low-dimensional subspaces

Deterministic setting

Ingredients:

- True signal: $x_0 \in \mathbb{R}^n$
- *Linear* forward operator: $A : \mathbb{R}^n \to \mathbb{R}^m$
- Noisy measurements: $b = Ax_0 + e$ with $e \sim \mathcal{N}(0, \lambda^{-1} I)$

Solve:

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} \frac{\lambda}{2}\|Ax - b\|_2^2 + \begin{cases} \frac{\delta}{2}\|Dx\|_2^2 & \text{(Tikhonov)} \\ \gamma\|Lx\|_1 & \text{(Sparsity/TV)} \\ \gamma\sum_i \|L_i x\|_2 & \text{(Group Lasso/TV)} \\ \chi_C(x) & \text{(Constraints)} \end{cases}$$

This talk: $$\operatorname*{argmin}_{x\in\mathbb{R}^n} \frac{\lambda}{2}\|Ax - b\|_2^2 + \gamma\|Lx\|_1 + \chi_C(x)$$

Bayesian setting

Ingredients:

- Likelihood: $\pi(b|x) \propto \exp\left(-\frac{\lambda}{2}\|Ax-b\|_2^2\right)$
- Prior: $\pi(x) \propto \mathbf{1}_C(x)\exp\left(-\gamma\|Lx\|_1\right)$
- Posterior: $\pi(x|b) \propto \mathbf{1}_C(x)\exp\left(-\frac{\lambda}{2}\|Ax-b\|_2^2-\gamma\|Lx\|_1\right)$

MAP estimate: $$\operatorname*{argmax}_{x\in\mathbb{R}^n} \pi(x|b) = \operatorname*{argmin}_{x\in\mathbb{R}^n} \frac{\lambda}{2}\|Ax-b\|_2^2+\gamma\|Lx\|_1+\chi_C(x)$$

The MAP estimate can be sparse, but
the posterior $\pi(x|b)$ assigns **zero** probability to sparse vectors.
This talk: A method to bring the deterministic regularization effects into a Bayesian setting.

Sampling Gaussian posteriors

Gaussian posterior: $\pi(x|b) \propto \exp\left(-\frac{\lambda}{2}\|Ax-b\|_2^2 - \frac{\delta}{2}\|Dx\|_2^2\right)$

Samples from this posterior can be obtained by solving

$$\underset{x}{\operatorname{argmin}} \frac{\lambda}{2}\|Ax-\hat{b}\|_2^2 + \frac{\delta}{2}\|Dx-\hat{c}\|_2^2,$$

with $\hat{b} \sim \mathcal{N}(b, \lambda^{-1}I)$ and $\hat{c} \sim \mathcal{N}(0, \delta^{-1}I)$.

What if we add regularization and/or constraints?

Regularized Gaussians

Formulation one: **Reguralized Gaussian posterior**
Define a new posterior distribution by

$$x \mid b := \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \frac{\lambda}{2}\|Ax-\hat{b}\|_2^2 + \frac{\delta}{2}\|Dx-\hat{c}\|_2^2 + f(x), \text{ with } \hat{b}\sim\mathcal{N}(b,\lambda^{-1}I) \text{ and } \hat{c}\sim\mathcal{N}(0,\delta^{-1}I).$$

Formulation two: **Proximal post-processed posterior**
Obtain a sample $x^\star$ from the unregularized posterior and apply the oblique proximal

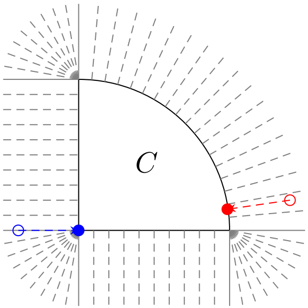
$$\text{prox}_f^{\Sigma^{-1}}(x^\star) := \underset{x\in\mathbb{R}^n}{\operatorname{argmin}} \|x-x^\star\|_{\Sigma^{-1}}^2 + f(x), \quad \text{with} \quad \Sigma^{-1} = \lambda A^TA + \delta D^TD.$$

Example - Constrained

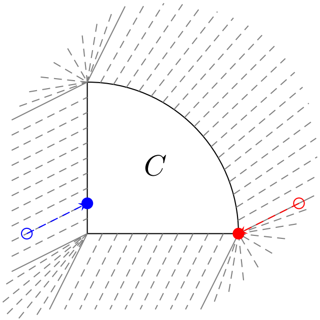
If $f(x) = \chi_C(x)$, then $\mathrm{prox}_f^{\Sigma^{-1}}(x^\star) = \Pi_C^{\Sigma^{-1}}(x^\star)$ is an oblique projection.

$$\Pi_C^I(x^\star) := \underset{x \in C}{\operatorname{argmin}} \|x - x^\star\|_2^2 \quad \text{or} \quad \Pi_C^{\Sigma^{-1}}(x^\star) := \underset{x \in C}{\operatorname{argmin}} \|x - x^\star\|_{\Sigma^{-1}}^2$$

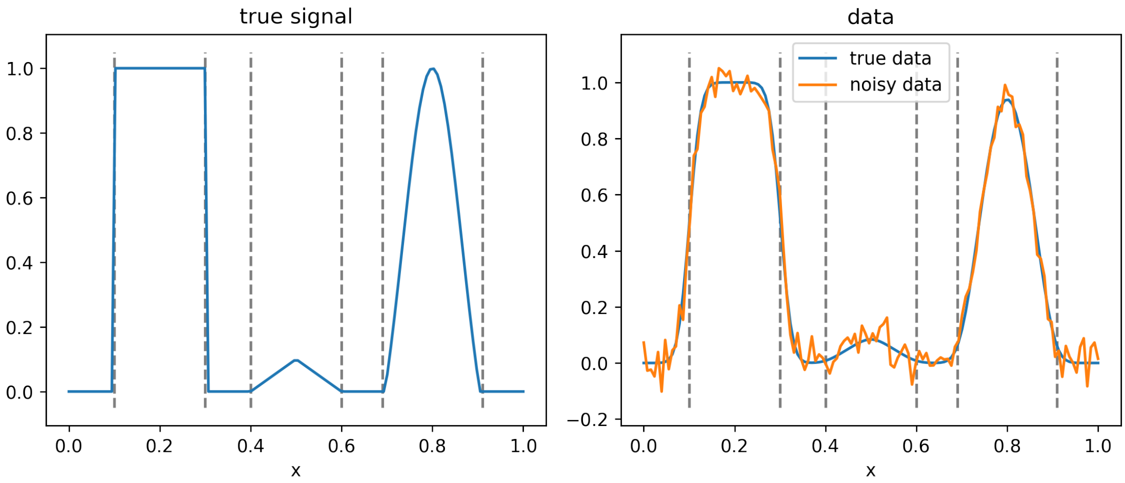
Π_C^I



$\Pi_C^{\Sigma^{-1}}$

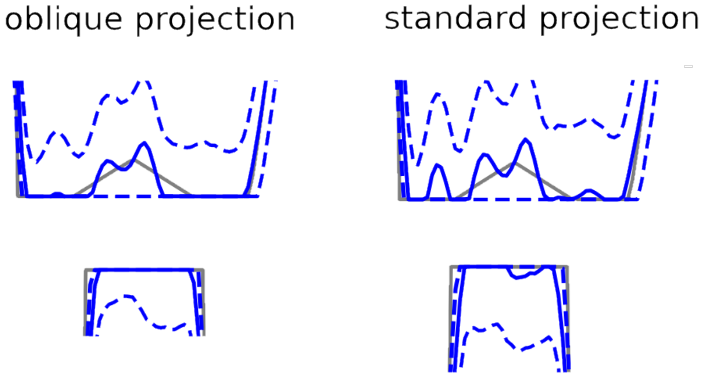


Deblurring example



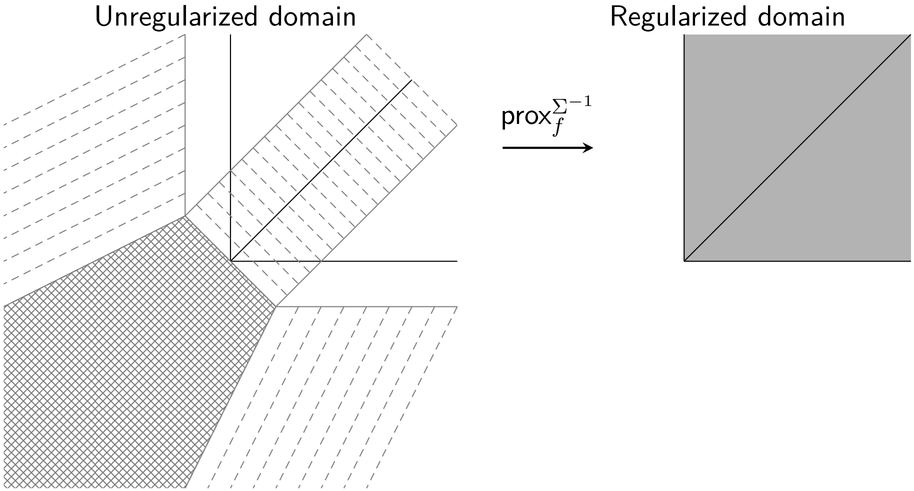
Deblurring example

Componentwise median with 95% credibility bands



Example - Nonnegative Total Variation

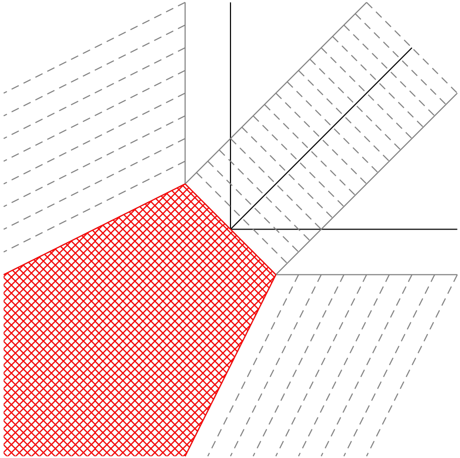
$$f(x) = |x_2 - x_1| + \chi_{\mathbb{R}^2_+}(x)$$



Example - Nonnegative Total Variation

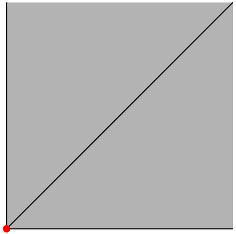
$$f(x) = |x_2 - x_1| + \chi_{\mathbb{R}^2_+}(x)$$

Unregularized domain



$\xrightarrow{\text{prox}_f^{\Sigma^{-1}}}$

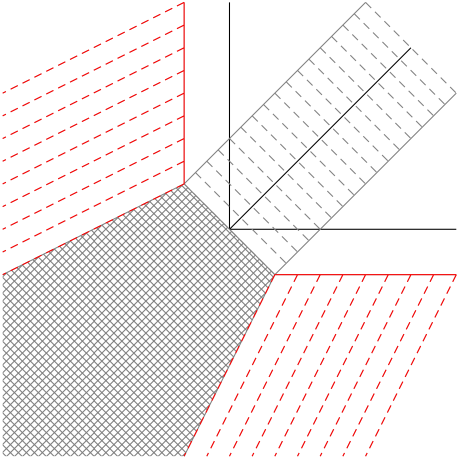
Regularized domain



Example - Nonnegative Total Variation

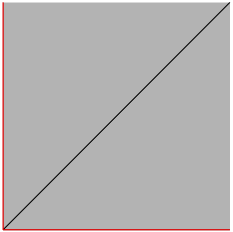
$$f(x) = |x_2 - x_1| + \chi_{\mathbb{R}^2_+}(x)$$

Unregularized domain



$$\xrightarrow{\operatorname{prox}_f^{\Sigma^{-1}}}$$

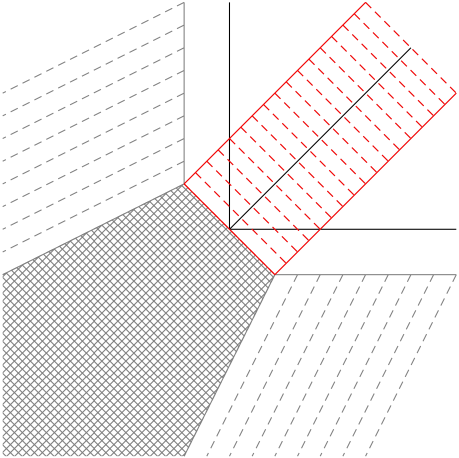
Regularized domain



Example - Nonnegative Total Variation

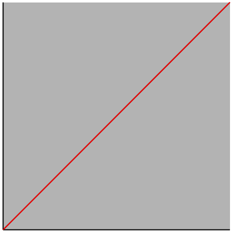
$$f(x) = |x_2 - x_1| + \chi_{\mathbb{R}^2_+}(x)$$

Unregularized domain



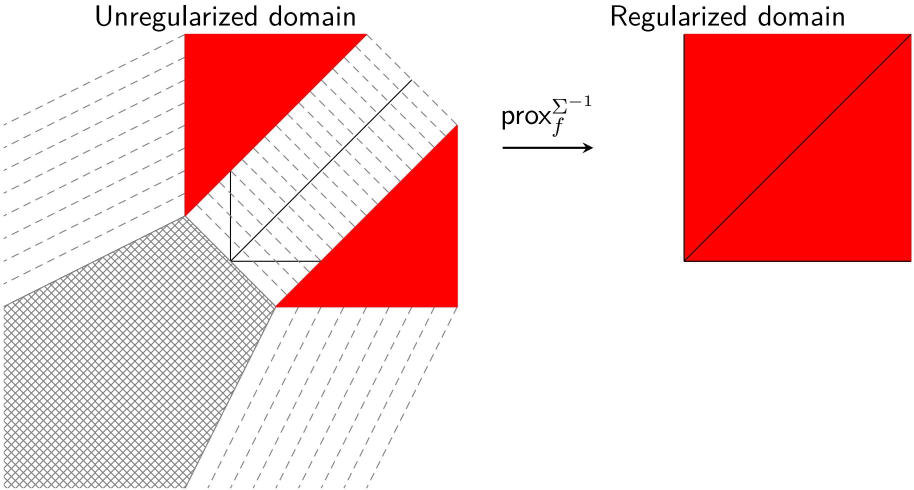
$\xrightarrow{\text{prox}_f^{\Sigma^{-1}}}$

Regularized domain



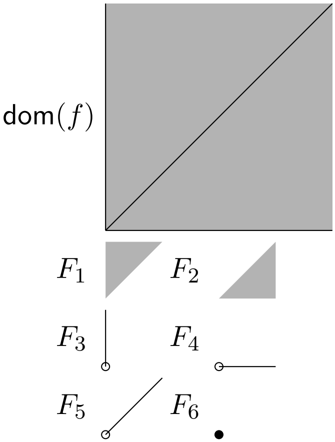
Example - Nonnegative Total Variation

$$f(x) = |x_2 - x_1| + \chi_{\mathbb{R}^2_+}(x)$$



Low-dimensional structure

$$f(x) = |x_2 - x_1| + \chi_{\mathbb{R}^2_+}(x) = \|Lx\|_1 + \chi_C(x)$$



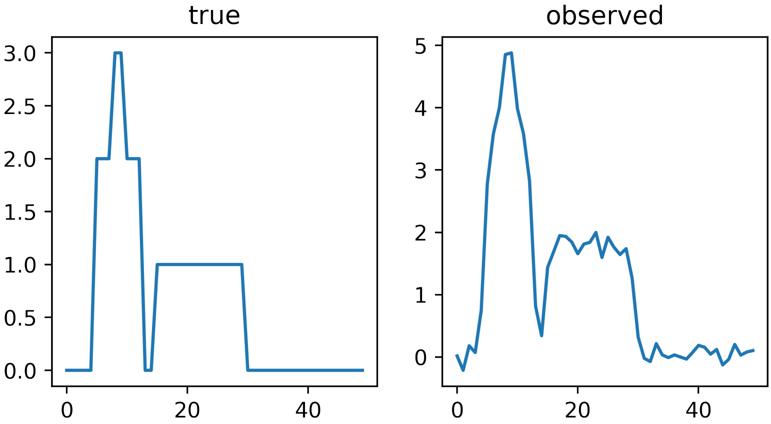
$\text{prox}_f^{\Sigma^{-1}}(x^\star)$:

- has positive probability on each face of C,
- has positive probability on sets of vectors that are sparse in Lx.

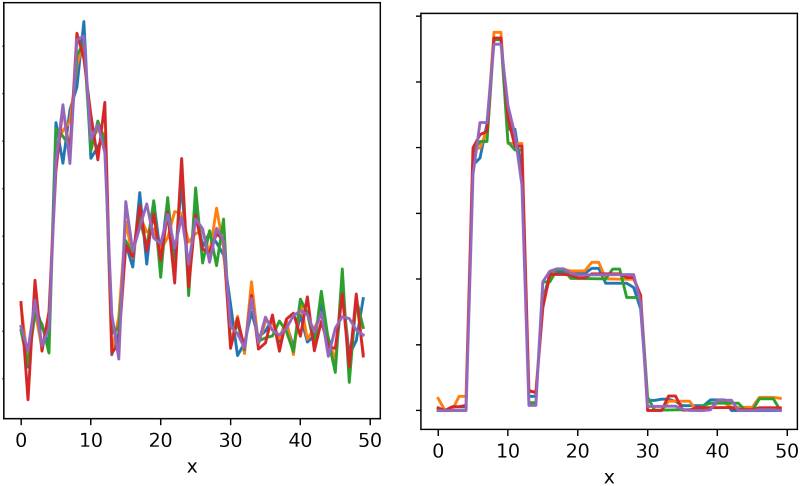
But,

- it has no density with respect to the Lebesgue measure,
- can currently only sample by solving the optimization problem.

Numerical example: 1D deblurring



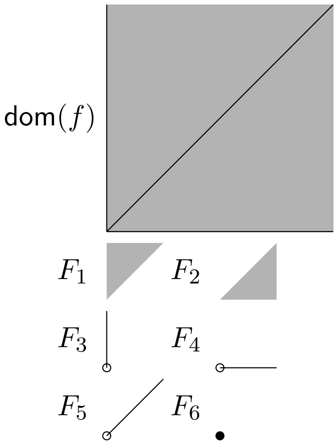
Numerical example: deblurring



Five samples from an unregularized posterior (left) and a regularized posterior (right).

Conditional prior

$$f(x) = \|Lx\|_1 + \chi_C(x)$$



Probability distribution satisfies:

$$\pi(x \mid b, F_j) \propto \exp\left(-\frac{\lambda}{2}\|Ax - b\|_2^2 - \frac{\delta}{2}\|Dx\|_2^2 - f(x)\right).$$

Conditional prior:

$$\pi(x \mid F_i) \propto \frac{\pi(x \mid b, F_i)}{\pi(b \mid x, F_i)} \propto \exp\left(-\frac{1}{2}\|Dx\|_2^2 - f(x)\right).$$

Hierarchical model

$$x \,|\, b, \lambda, \delta := \operatorname*{argmin}_{x \in \mathbb{R}^n} \frac{\lambda}{2}\|Ax - \hat{b}\|_2^2 + \frac{\delta}{2}\|Dx - \hat{c}\|_2^2 + \chi_{\mathbb{R}^n_+}(x),$$

with $\hat{b} \sim \mathcal{N}(b, \lambda^{-1}I)$ and $\hat{c} \sim \mathcal{N}(0, \delta^{-1}I)$.

Add hyperpriors:

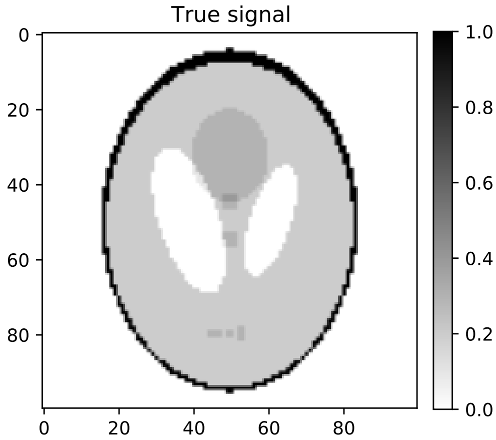
$$\begin{aligned} \pi(\lambda) &\propto \lambda^{\alpha_\lambda - 1}\exp(-\beta_\lambda \lambda), \qquad \text{for } \lambda > 0 \quad \text{and} \\ \pi(\delta) &\propto \delta^{\alpha_\delta - 1}\exp(-\beta_\delta \delta), \qquad \text{for } \delta > 0, \end{aligned}$$

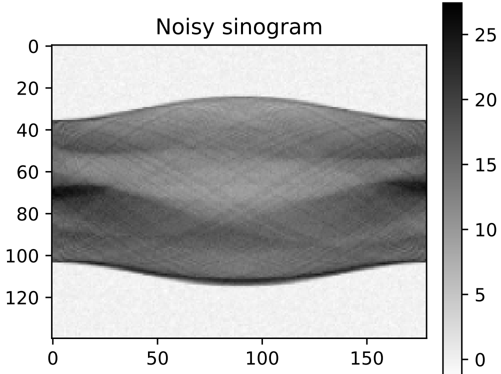
Gibbs sampler

$$f(x) = \chi_{\mathbb{R}^n_+}(x)$$

Input: $x^0, \alpha_\lambda, \beta_\lambda, \alpha_\delta, \alpha_\lambda, k_{\max}$
for $k = 1$ **to** $k_{\max}$ **do**
  Compute $(\lambda_k, \delta_k) \sim \pi(\lambda, \delta \mid x, b)$ as follows:
  $$\lambda_k \sim \Gamma\left(m/2 + \alpha_\lambda, \tfrac{1}{2}\|Ax^{k-1} - b\|_2^2 + \beta_\lambda\right),$$
  $$\delta_k \sim \Gamma\left(\|x\|_0/2 + \alpha_\delta, \tfrac{1}{2}\|Lx^{k-1}\|_2^2 + \beta_\delta\right).$$
  Compute $x^k \sim \pi_{x|b,\lambda^k,\delta^k}$.
end for
return $\{(x^k, \lambda_k, \delta_k)\}_{k=1,\ldots,k_{\max}}$

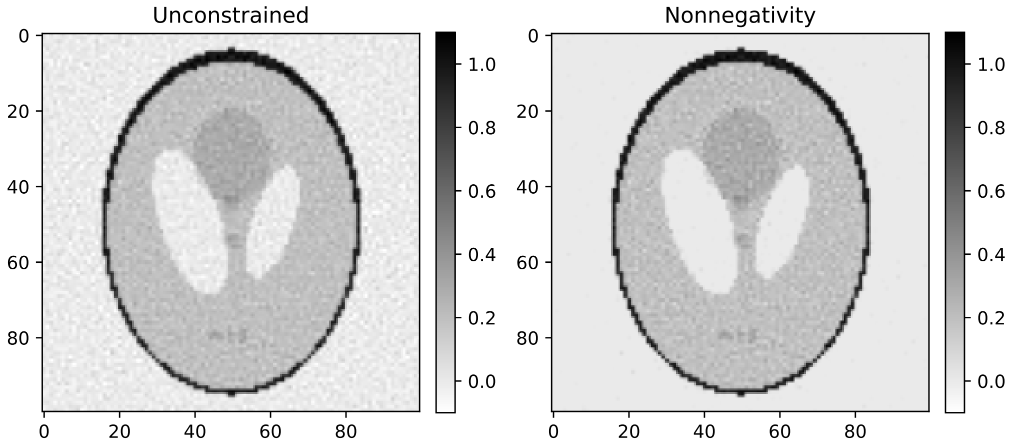
Numerical example: nonnegative CT





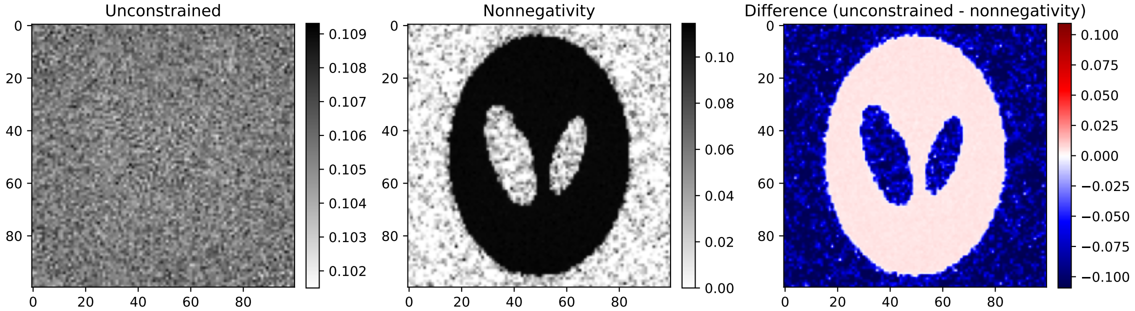
Numerical example: nonnegative CT

Componentwise median



Numerical example: nonnegative CT

Componentwise width of 95% credibility intervals



Other regularization functions

?

$\gamma \sum_i \|L_i x\|_p$

Polyhedral epigraph

$\chi_{[0,1]^n}(x)$ $\gamma \|Lx - c\|_1$

$\max_i\{a_i^T x + b_i\}$

Polyhedral cone epigraph

$\chi_{\mathbb{R}_+^n}(x)$ $\gamma \|Lx\|_1$

$\max_i\{a_i^T x\}$

Positive probability on subspaces

Conditional prior

Gibbs sampler for a hierarchical model

Conclusions and future work

Pros

- Posterior distribution with sparsity structure.
- Sample using existing methods from optimization theory.

Cons

- Solving an optimization problem accurately per sample is expensive.

Preprint on constraints available: Bayesian inference with projected densities

Sparse Bayesian Inference with Regularized Gaussians

Jasper M. Everink, Technical University of Denmark, jmev@dtu.dk

Joint work with Martin S. Andersen and Yiqiu Dong