Rafael Flock, Yiqiu Dong, Technical University of Denmark
Olivier Zahm, INRIA Grenoble
Felipe Uribe, Lappeenranta-Lahti University of Technology

Certified Coordinate Selection for large-dimensional Bayesian Inversion

Motivation

UQ for the reconstruction of large dimensional sparse signals

Data generating process, for example:

$$A(x_{\mathrm{true}}) + \varepsilon = y, \quad x_{\mathrm{true}} \in \mathbb{R}^d, y \in \mathbb{R}^m, \varepsilon \sim \mathcal{N}(0, \Sigma_{\mathrm{obs}}).$$

Motivation

UQ for the reconstruction of large dimensional sparse signals

Data generating process, for example:

$$A(x_{\mathrm{true}}) + \varepsilon = y, \quad x_{\mathrm{true}} \in \mathbb{R}^d, y \in \mathbb{R}^m, \varepsilon \sim \mathcal{N}(0, \Sigma_{\mathrm{obs}}).$$

Posterior density in x:

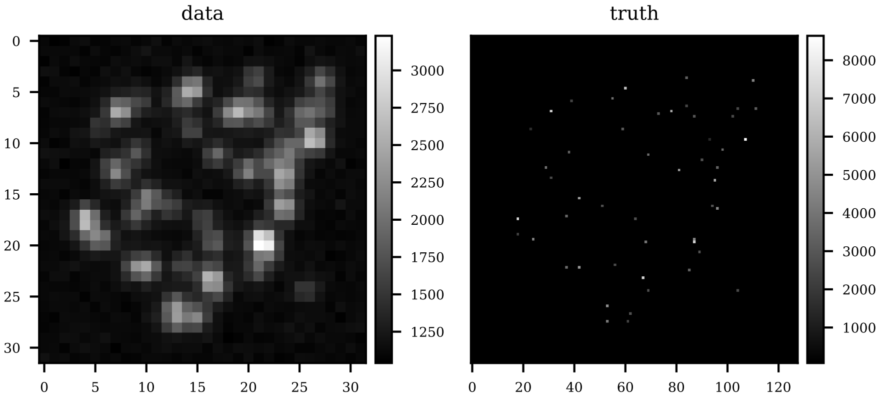
$$\pi(x|y) = \frac{1}{Z}\pi(y|x)\pi(x), \quad Z = \int \pi(y|x)\pi(x)\mathrm{d}x$$

In this talk: $\pi(x) \propto \mathcal{L}(x)\pi_0(x)$

where for example $\mathcal{L}(x) \propto \exp\left(-\frac{1}{2}\|y - A(x)\|^2_{\Sigma^{-1}_{\mathrm{obs}}}\right).$

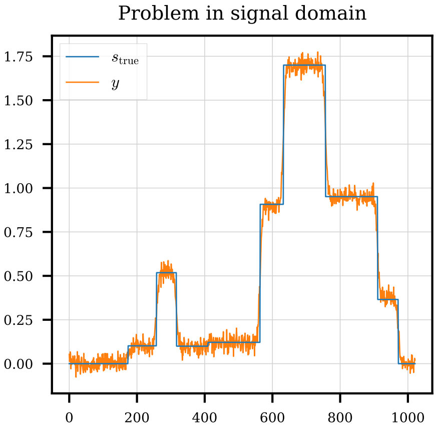
Motivation

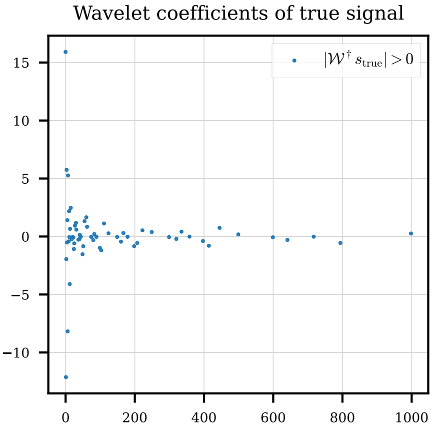
UQ for the reconstruction of large dimensional sparse signals



Motivation

UQ for the reconstruction of large dimensional sparse signals





Motivation

UQ for the reconstruction of large dimensional sparse signals

Laplace prior to enforce sparsity:

$$\pi_0(x) \propto \exp\left(-\sum \delta_i |x_i|\right), \quad \delta_i > 0$$

Posterior density in x:

$$\pi(x) \propto \mathcal{L}(x) \exp\left(-\sum \delta_i |x_i|\right)$$

Motivation

UQ for the reconstruction of large dimensional sparse signals

Laplace prior to enforce sparsity:

$$\pi_0(x) \propto \exp\left(-\sum \delta_i |x_i|\right), \quad \delta_i > 0$$

Posterior density in x:

$$\pi(x) \propto \mathcal{L}(x) \exp\left(-\sum \delta_i |x_i|\right)$$

How can we find the 'most important coordinates' and how can we approximate the posterior density with this knowledge?

Outline

- A posterior approximation
- Certifying the approximation
- Sampling options
- Numerical example 1: 1D piece-wise constant signal
- Numerical example 2: 2D super-resolution microscopy
- Conclusions

A posterior approximation

Coordinate splitting

Replace the likelihood by a ridge approximation:

$$\pi(x) \propto \mathcal{L}(x)\pi_0(x)$$
$$\tilde{\pi}(x) \propto \widetilde{\mathcal{L}}(x_{\mathcal{I}})\pi_0(x_{\mathcal{I}})\pi_0(x_{\mathcal{I}^c}) = \tilde{\pi}(x_{\mathcal{I}})\pi_0(x_{\mathcal{I}^c}),$$

given a coordinate splitting $x := (x_{\mathcal{I}}, x_{\mathcal{I}^c})$, $x_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|}$, $x_{\mathcal{I}^c} \in \mathbb{R}^{|\mathcal{I}^c|}$.

Ideally $|\mathcal{I}| \ll |\mathcal{I}^c|$

A posterior approximation

Optimal reduced likelihood

Proposition

For $\pi_0(x) \propto \exp\left(-\sum \delta_i |x_i|\right),\ \delta_i > 0$ the optimal reduced likelihood which minimizes the (squared) Hellinger distance

$$\mathcal{D}_{\mathrm{H}}\left(\pi || \tilde{\pi}\right)^2 = \frac{1}{2} \int_{\mathbb{R}^d} \left(\sqrt{\pi(x)} - \sqrt{\tilde{\pi}(x)} \right)^2 \mathrm{d}x \,,$$

is given by

$$\widetilde{\mathcal{L}}^*(x_{\mathcal{I}}) = \left(\int_{\mathbb{R}_{\mathcal{I}^c}} \sqrt{\mathcal{L}(x_{\mathcal{I}}, x_{\mathcal{I}^c})} \pi_0(x_{\mathcal{I}^c}) \mathrm{d}x_{\mathcal{I}^c} \right)^2 .$$

Optimal reduced likelihood

Proposition

For $\pi_0(x) \propto \exp\left(-\sum \delta_i |x_i|\right),\ \delta_i > 0$ the optimal reduced likelihood which minimizes the (squared) Hellinger distance

$$\mathcal{D}_{\mathrm{H}}\left(\pi || \tilde{\pi}\right)^2 = \frac{1}{2}\int_{\mathbb{R}^d} \left(\sqrt{\pi(x)} - \sqrt{\tilde{\pi}(x)}\right)^2 \mathrm{d}x \,,$$

is given by

$$\widetilde{\mathcal{L}}^*(x_{\mathcal{I}}) = \left(\int_{\mathbb{R}_{\mathcal{I}^c}} \sqrt{\mathcal{L}(x_{\mathcal{I}}, x_{\mathcal{I}^c})}\pi_0(x_{\mathcal{I}^c})\mathrm{d}x_{\mathcal{I}^c}\right)^2 .$$

How to select $\mathcal{I}$?

Certifying the approximation

Upper bound on the Hellinger distance

Proposition

For $\widetilde{\pi}^*(x) \propto \widetilde{\mathcal{L}}^*(x)\pi_0(x)$, we can control the Hellinger distance with

$$\mathcal{D}_{\mathrm{H}}\left(\pi || \widetilde{\pi}^*\right)^2 \leq 4 \sum_{i \in \mathcal{I}^c} h_i,$$

where the entries of the *diagnostic* vector $h \in \mathbb{R}^d$ are

$$h_i = \frac{1}{\delta_i^2} \int_{\mathbb{R}^d} (\partial_i \log \mathcal{L}(x))^2 \pi(x) \mathrm{d}x \, .$$

Certifying the approximation

Additive Gaussian noise and linear forward model

For problems of the form

$$\pi(x) \propto \exp\left(-\frac{1}{2}\|Ax - y\|^2_{\Sigma_{\mathrm{obs}}^{-1}} - \sum \delta_i \|x_i\|\right)$$

we have

$$h = \mathrm{diag}\left(A^\mathsf{T}\Sigma_{\mathrm{obs}}^{-1} A \Sigma A^\mathsf{T}\Sigma_{\mathrm{obs}}^{-1} A\right) + (A^\mathsf{T}\Sigma_{\mathrm{obs}}^{-1} y - A^\mathsf{T}\Sigma_{\mathrm{obs}}^{-1}\mu)^{\circ 2},$$

where Σ and μ are the posterior covariance and mean, respectively.

Certifying the approximation

Additive Gaussian noise and linear forward model

$$h = \operatorname{diag}\left(A^\intercal \Sigma_{\mathrm{obs}}^{-1} A \Sigma A^\intercal \Sigma_{\mathrm{obs}}^{-1} A\right) + (A^\intercal \Sigma_{\mathrm{obs}}^{-1} y - A^\intercal \Sigma_{\mathrm{obs}}^{-1} \mu)^{\circ 2},$$

where Σ and μ are the posterior covariance and mean, respectively.

Additive Gaussian noise and linear forward model

$$h = \operatorname{diag}\left(A^\intercal \Sigma_{\mathrm{obs}}^{-1} A \Sigma A^\intercal \Sigma_{\mathrm{obs}}^{-1} A\right) + (A^\intercal \Sigma_{\mathrm{obs}}^{-1} y - A^\intercal \Sigma_{\mathrm{obs}}^{-1} \mu)^{\circ 2},$$

where Σ and μ are the posterior covariance and mean, respectively.

In practice, we can approximate μ and Σ by, e.g.,

- A Gaussian posterior approximation at the maximum-a-posteriori probability (MAP) estimate, i.e.,
  $\mu \approx x_{\mathrm{MAP}}$ and $\Sigma^{-1} \approx -\nabla^2 \log \pi(x_{\mathrm{MAP}})$

Certifying the approximation

Additive Gaussian noise and linear forward model

$$h = \mathrm{diag}\left(A^\intercal \Sigma_{\mathrm{obs}}^{-1} A \Sigma A^\intercal \Sigma_{\mathrm{obs}}^{-1} A\right) + (A^\intercal \Sigma_{\mathrm{obs}}^{-1} y - A^\intercal \Sigma_{\mathrm{obs}}^{-1} \mu)^{\circ 2},$$

where Σ and μ are the posterior covariance and mean, respectively.

In practice, we can approximate μ and Σ by, e.g.,

- A Gaussian posterior approximation at the maximum-a-posteriori probability (MAP) estimate, i.e.,
  $\mu \approx x_{\mathrm{MAP}}$ and $\Sigma^{-1} \approx -\nabla^2 \log \pi(x_{\mathrm{MAP}})$
- Prior mean and covariance, i.e,
  $\mu \approx 0$ and $\Sigma \approx 2\,\mathrm{diag}\left(\delta_i^{-2}\right)$

Sampling Options

Option 1: Sampling the approximated posterior

1. Compute diagnostic h and perform coordinate splitting $x := (x_{\mathcal{I}}, x_{\mathcal{I}^c})$
2. Draw N samples $\{x_{\mathcal{I}}\}_{k=1}^{N}$ from the reduced marginal posterior $\widetilde{\pi}^*(x_{\mathcal{I}}) \propto \widetilde{\mathcal{L}}^*(x_{\mathcal{I}})\pi_0(x_{\mathcal{I}})$.
3. Draw N samples $\{x_{\mathcal{I}^c}\}_{k=1}^{N}$ from $\pi_0(x_{\mathcal{I}^c})$.
4. Reassemble samples from (2) and (3): $\{x\}_{k=1}^{N} = \{(x_{\mathcal{I}}^{(k)}, x_{\mathcal{I}^c}^{(k)})\}_{k=1}^{N}$.

Option 2: Sampling the exact posterior

Pseudo-marginal MCMC[1], delayed acceptance MCMC[2]

[1] Christophe, A., Roberts, G. O.: The pseudo-marginal approach for efficient Monte Carlo computations. (2009)

[2] Liu, Jun S., Rong Chen: Sequential Monte Carlo methods for dynamic systems. Journal of the American statistical association 93.443 (1998).

Option 2: Sampling the exact posterior

Pseudo-marginal MCMC[1], delayed acceptance MCMC[2]

Pseudo-marginal MCMC

1. Compute diagnostic h and perform coordinate splitting $x := (x_{\mathcal{I}}, x_{\mathcal{I}^c})$

[1] Christophe, A., Roberts, G. O.: The pseudo-marginal approach for efficient Monte Carlo computations. (2009)

[2] Liu, Jun S., Rong Chen: Sequential Monte Carlo methods for dynamic systems. Journal of the American statistical association 93.443 (1998).

Sampling Options

Option 2: Sampling the exact posterior

2. for $i = 1 \dots N$:
   - A. Draw $z_{\mathcal{I}} \sim q(\cdot | x_{\mathcal{I}}^{(i-1)})$
   - B. Draw M i.d. samples $z_{\mathcal{I}^c}^{(j)} \sim \pi_0(\cdot)$
   - C. Compute $\widetilde{\mathcal{L}}(z_{\mathcal{I}}) \approx \frac{1}{M} \sum_{j=1}^{M} \mathcal{L}(z_{\mathcal{I}}, z_{\mathcal{I}^c}^{(j)})$
   - D. Set $\{x_{\mathcal{I}}^{(i)}, \{x_{\mathcal{I}^c}^{(i,j)}\}_{j=1}^{M}\} = \{z_{\mathcal{I}}, \{z_{\mathcal{I}^c}^{(j)}\}_{j=1}^{M}\}$ with acceptance probability

$$\alpha = \min \left\{ 1, \frac{\pi_0(z_{\mathcal{I}}) \widetilde{\mathcal{L}}(z_{\mathcal{I}}) q(x_{\mathcal{I}} | z_{\mathcal{I}})}{\pi_0(x_{\mathcal{I}}) \widetilde{\mathcal{L}}(x_{\mathcal{I}}) q(z_{\mathcal{I}} | x_{\mathcal{I}})} \right\}$$

3. Return Markov chain $\{x_{\mathcal{I}}^{(i)}, \{x_{\mathcal{I}^c}^{(i,j)}\}_{j=1}^{M}\}_{i=1}^{N}$

Sampling Options

Option 2: Sampling the exact posterior

4. Recycling step for $i = 1 \dots N$:
   - Set $x_{\mathcal{I}^c}^{(i)} = x_{\mathcal{I}^c}^{(i,j)}$ with probability

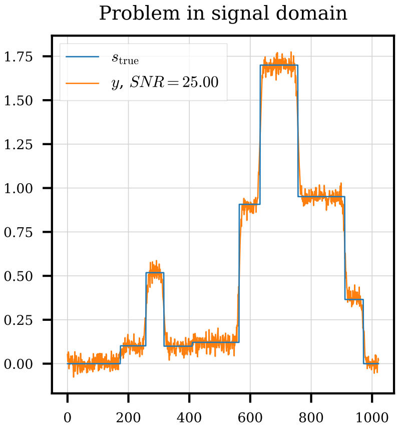
$$\mathbb{P}(X_{\mathcal{I}^c}^{(i)} = x_{\mathcal{I}^c}^{(i,j)} | x_{\mathcal{I}}^{(i)}, \{x_{\mathcal{I}^c}^{(i,j)}\}_{j=1}^{M}) = \frac{\mathcal{L}(x_{\mathcal{I}}^{(i)}, x_{\mathcal{I}^c}^{(i,j)})}{\sum_{j=1}^{M} \mathcal{L}(x_{\mathcal{I}}^{(i)}, x_{\mathcal{I}^c}^{(i,j)})}.$$

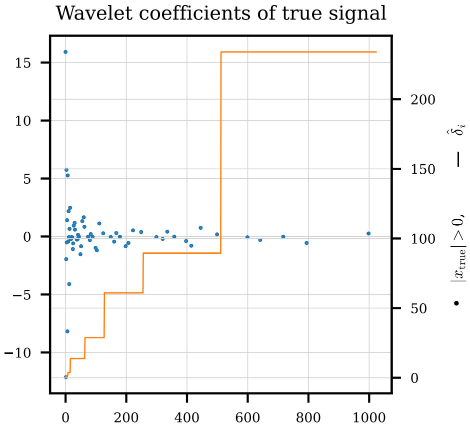
   - Reassemble $x^{(i)} = (x_{\mathcal{I}}^{(i)}, x_{\mathcal{I}^c}^{(i)})$
5. Return Markov chain $\{x^{(i)}\}_{i=1}^{N}$

DTU

Numerical example 1: 1D piece-wise constant signal

Problem description





Numerical example 1: 1D piece-wise constant signal

Problem description

The data is generated as

$$y = Rs_{\mathrm{true}} + \varepsilon,$$

where R is a Gaussian blur operator and $\varepsilon \sim \mathcal{N}(0, \sigma_{\mathrm{obs}}^2)$.

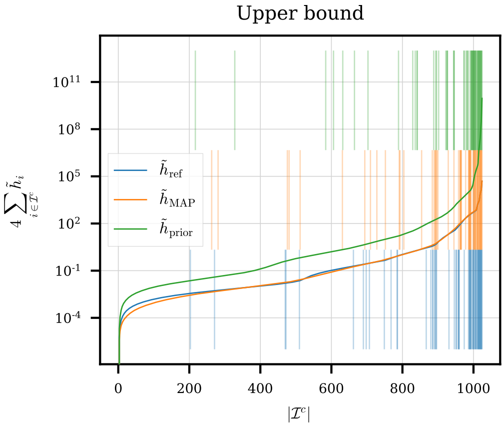
The posterior density in coefficient space (synthesis formulation) is

$$\pi(x) \propto \exp\left(-\frac{1}{2\sigma_{\mathrm{obs}}^2}\|y - RWx\|_2^2 - \sum_{i=1}^{d} \delta_i |x_i|\right),$$

where a W is the synthesis operator of a 10-level Haar wavelet basis.

Numerical example 1: 1D piece-wise constant signal

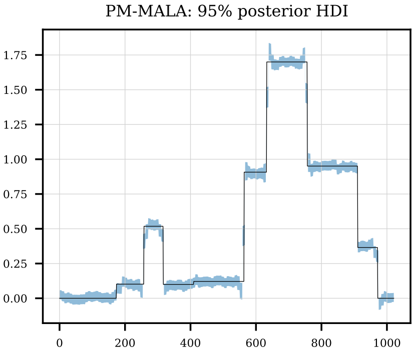
Selection of coordinates



The vertical lines indicate the indices $\{i : x_{\mathrm{true},i} \neq 0\}$.

Numerical example 1: 1D piece-wise constant signal

Pseudo-marginal for exact inference



setting	PM-MALA	MALA
$\lvert\mathcal{I}\rvert$	200	-
max $\hat{R}$	1.00	1.19
$\mathrm{ESS}_{\mathcal{I}}$	3977	93
$\mathrm{ESS}_{\mathcal{I}^c}$	19756	204
time [min]	52.7	34.4
mean step size	9.8×10^{-3}	2.2×10^{-4}

20 000 samples in total, averages over 10 chains

Numerical example 1: 1D piece-wise constant signal

Sampling the approximate posterior

Optimal reduced likelihood:

$$\widetilde{\mathcal{L}}^*(x_{\mathcal{I}}) = \left(\int_{\mathbb{R}_{\mathcal{I}^c}} \sqrt{\mathcal{L}(x_{\mathcal{I}}, x_{\mathcal{I}^c})} \pi_0(x_{\mathcal{I}^c}) \mathrm{d}x_{\mathcal{I}^c} \right)^2 .$$

Numerical example 1: 1D piece-wise constant signal

Sampling the approximate posterior

Optimal reduced likelihood:

$$\widetilde{\mathcal{L}}^*(x_{\mathcal{I}}) = \left(\int_{\mathbb{R}_{\mathcal{I}^c}} \sqrt{\mathcal{L}(x_{\mathcal{I}}, x_{\mathcal{I}^c})} \pi_0(x_{\mathcal{I}^c}) \mathrm{d}x_{\mathcal{I}^c} \right)^2 .$$

Approximation:

$$\widetilde{\pi}(x) \propto \mathcal{L}(x_{\mathcal{I}}, x_{\mathcal{I}^c} = 0) \pi_0(x).$$

Numerical example 1: 1D piece-wise constant signal

Sampling the approximate posterior

setting	red-MALA	MALA
$\lvert\mathcal{I}\rvert$	400	-
max $\hat{R}$	1.00	1.06
$\mathrm{ESS}_{\mathcal{I}}$	2722.0	156.1
$\mathrm{ESS}_{\mathcal{I}^c}$	-	356.3
time [min]	62.3	62.9
mean step size	3.2×10^{-3}	1.8×10^{-4}

Numerical estimation of Hellinger distance:

$$\mathcal{D}_{\mathrm{H}}\left(\pi \| \widetilde{\pi}\right)^2 \leq 2 \int \left(\sqrt{\frac{\rho(x)}{\tilde{\rho}(x)}} - 1 \right)^2 \widetilde{\pi}(x)\mathrm{d}x$$

$$\approx \frac{2}{N} \sum_{i=1}^{N} \left(\sqrt{\frac{\rho(x^{(i)})}{\tilde{\rho}(x^{(i)})}} - 1 \right)^2$$

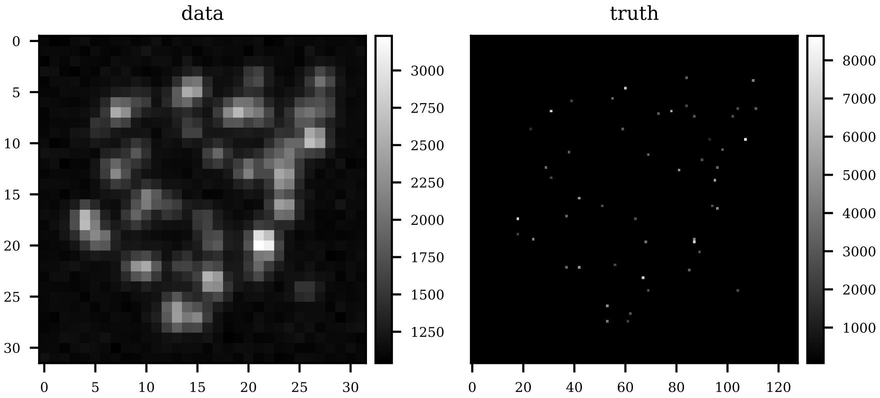
$x^{(i)} \sim \widetilde{\pi}(x)$, and ρ, $\tilde{\rho}$ are the unnormalized posterior densities.

red-MALA: $\mathcal{D}_{\mathrm{H}}\left(\pi \| \widetilde{\pi}\right)^2 \leq 3.42 \times 10^{-2} \pm 4.57 \times 10^{-4}$.

Our bound computed with h_{MAP}: $\mathcal{D}_{\mathrm{H}}\left(\pi \| \widetilde{\pi}\right)^2 \leq 1.0 \times 10^{-1}$!

Numerical example 2: 2D super-resolution microscopy

Problem description



Example from Zhu, L., Zhang, W., Elnatan, D., Huang, B.: Faster STORM using compressed sensing. Nature Methods 9(7), 721–723 (2012)

Numerical example 2: 2D super-resolution microscopy

Problem description

True super-resolution image: 50 molecules with photon count simulated as $x \sim \mathrm{lognormal}$ with mode 3000 and standard deviation 1700.

Data generation:

$$Ax_{\mathrm{true}} + \varepsilon = y,$$

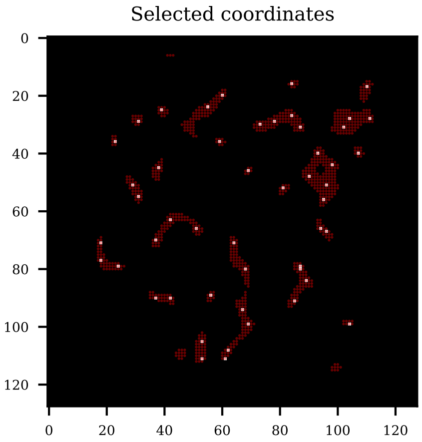
where $A : \mathbb{R}^{128\times128} \to \mathbb{R}^{32\times32}$ known blurring Kernel from the optical measurement instrument, and $\varepsilon \sim \mathcal{N}(0, \sigma^2_{\mathrm{obs}})$ ($SNR = 54$).

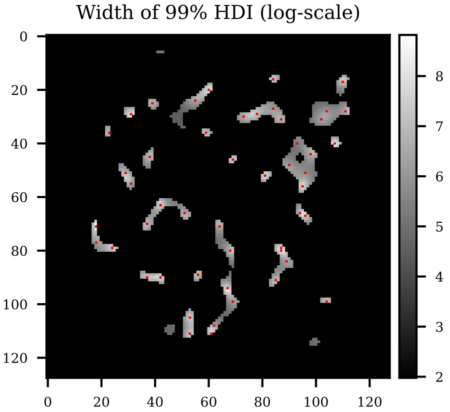
Posterior:

$$\pi(x) \propto \exp\left(-\frac{1}{2\sigma^2_{\mathrm{obs}}}\|y - Ax\|_2^2 - \delta\|x\|_1\right).$$

Numerical example 2: 2D super-resolution microscopy

Results





Conclusions

The method

- Selection of coordinates based on a bound on the Hellinger distance
- Sampling of exact posterior with specialized MCMC algorithms or approximate posterior
- Estimation of *diagnostic* in the case of linear forward model and Gaussian likelihood based on posterior mean and covariance

Conclusions

The method

- Selection of coordinates based on a bound on the Hellinger distance
- Sampling of exact posterior with specialized MCMC algorithms or approximate posterior
- Estimation of *diagnostic* in the case of linear forward model and Gaussian likelihood based on posterior mean and covariance

Outlook

- Exploration of other ways of computing the *diagnostic*
- Especially computing h_{MAP} problematic, since $\frac{\mathrm{d}}{\mathrm{d}x}|x|$ required

CUQI

THE VELUX FOUNDATIONS
VILLUM FONDEN VELUX FONDEN