Sample-based Bayesian Inversion

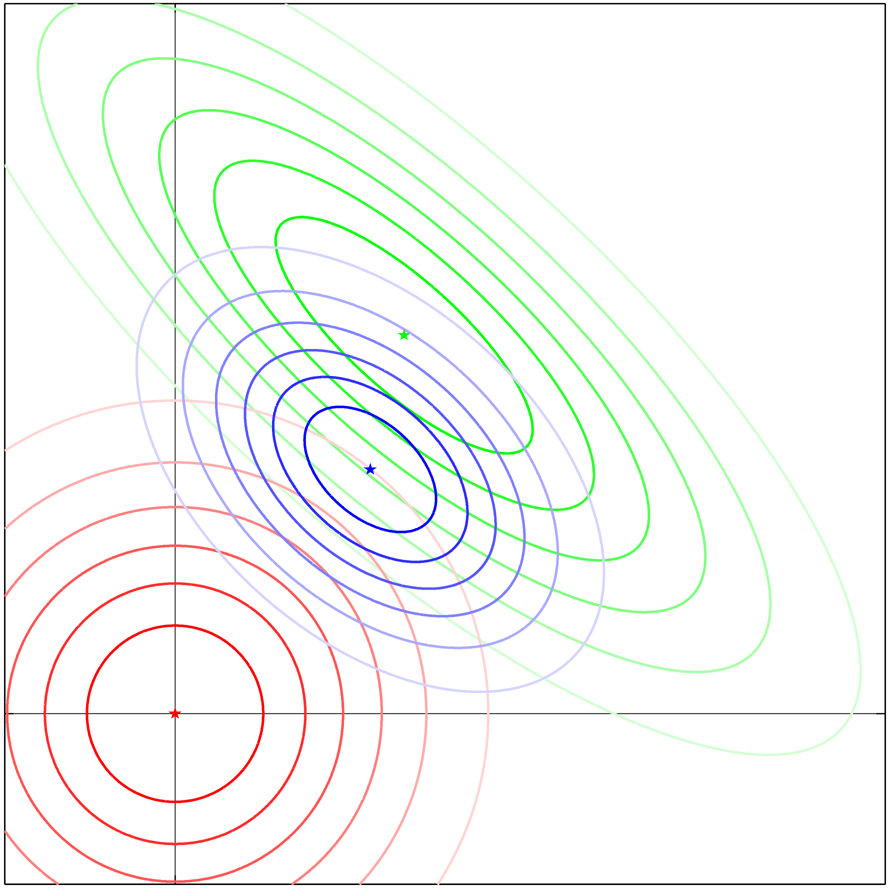
Felix Lucka
University College London

Centre for Medical Image Computing

Inverse Days
Tampere, Finland, 9.12.2014

Linear, ill-posed inverse problems with additive Gaussian noise:

$$f = Au + \varepsilon$$



$p_{like}(f|u) \propto$

$$\exp\left(-\tfrac{1}{2}\|f - Au\|^2_{\Sigma_\varepsilon^{-1}}\right)$$

$p_{prior}(u) \propto$

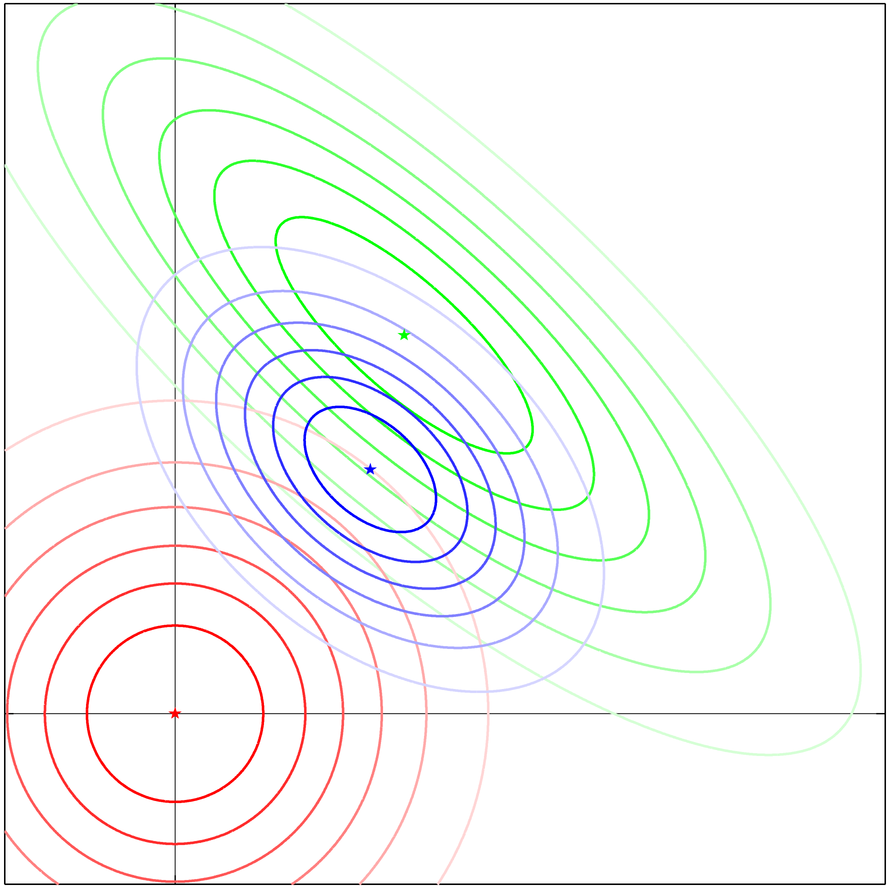
$$\exp\left(-\lambda\|D^T u\|_2^2\right)$$

$p_{post}(u|f) \propto$

$$\exp\left(-\tfrac{1}{2}\|f - Au\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\|D^T u\|_2^2\right)$$

Linear, ill-posed inverse problems with additive Gaussian noise:

$$\boxed{f = Au + \varepsilon}$$



Increased interest in Bayesian inversion:

- Probabilistic representation of solution allows for a rigorous quantification of its uncertainties.
- Advantage in applications where solution is subject to further analysis procedures.
- New computational tools for sampling high dimensional posterior distributions.

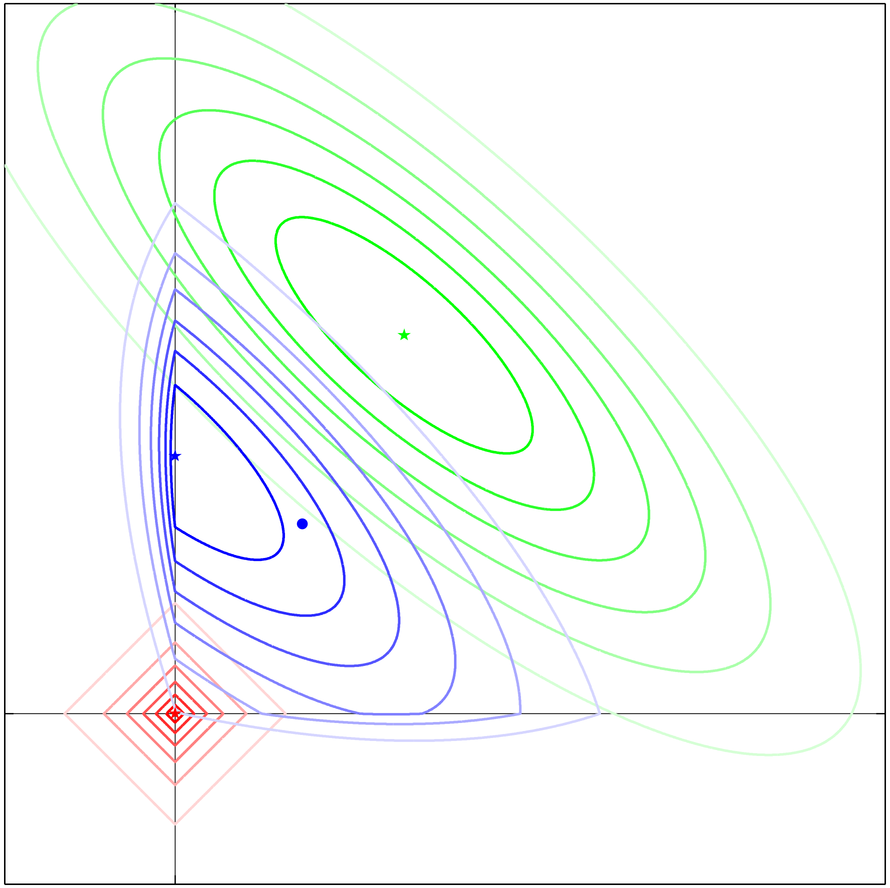
Inverse Days 2012: “Sparsity Constraints in Bayesian Inversion”

$$\boxed{f = Au + \varepsilon}$$

$$p_{like}(f|u) \propto \exp\left(-\tfrac{1}{2}\|f - Au\|^2_{\Sigma_\varepsilon^{-1}}\right)$$

$$p_{prior}(u) \propto \exp\left(-\lambda\|D^T u\|_1\right)$$

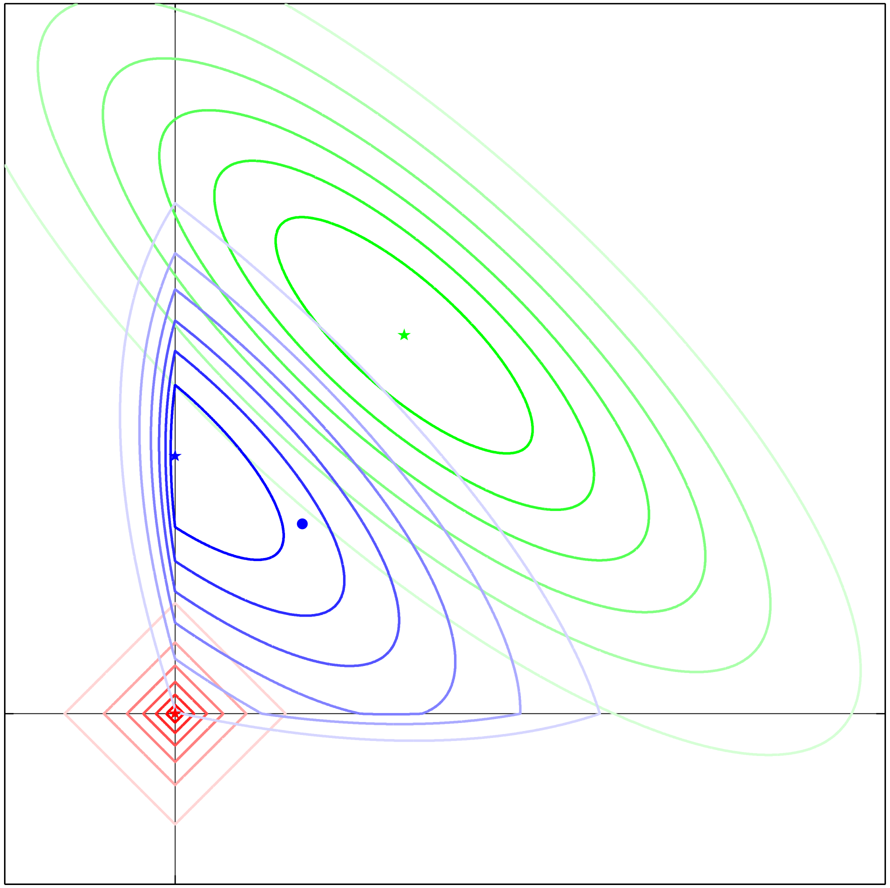
$$p_{post}(u|f) \propto \exp\left(-\tfrac{1}{2}\|f - Au\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\|D^T u\|_1\right)$$



Inverse Days 2012: "Sparsity Constraints in Bayesian Inversion"

$$\boxed{f = Au + \varepsilon}$$

$$p_{prior}(u) \propto \exp\left(-\lambda \|D^T u\|_1\right)$$



Starting point:

M. Lassas, S. Siltanen, 2004.
Can one use total variation prior for edge-preserving Bayesian inversion?
Inverse Problems 20

Aims: Sample based inference in high dimensions (infinite dimensional limits!)

Problem: Lack of suitable posterior samplers.

F. L., 2012.
Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors.
Inverse Problems 28(12); arXiv:1206.0262v2

Single component Gibbs sampler:

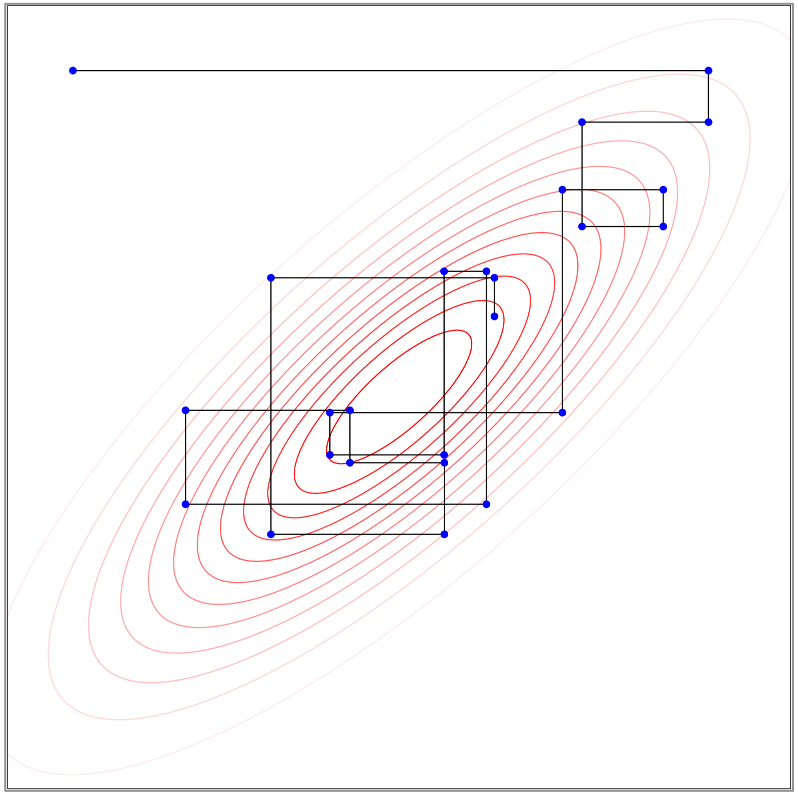
1. Fast computation of explicitly parameterized 1D cond. densities:

$$p(x) \propto \exp\left(-ax^2 + bx - c|x|\right)$$

application specific!

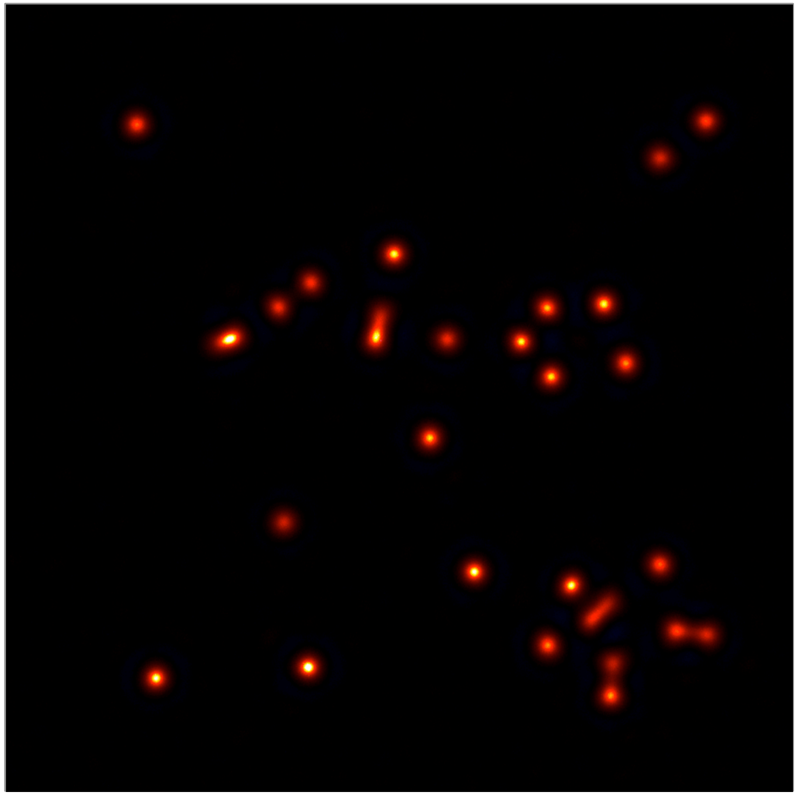
2. Fast, robust and exact sampling from 1D densities:

tedious implementation of inverse cumulative distribution (icd) sampling.



F. L., 2012.
Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors.
Inverse Problems 28(12); arXiv:1206.0262v2

- Scenarios:
  - 1D TV-deblurring
  - 2D deconvoloution
- Comparison to conventional Metropolis-Hastings sampling by visual convergence, burn-in and autocorrelation analysis
- Superior results, still efficient in high-dimensional scenarios.



CM estimate, n = 1023x1023

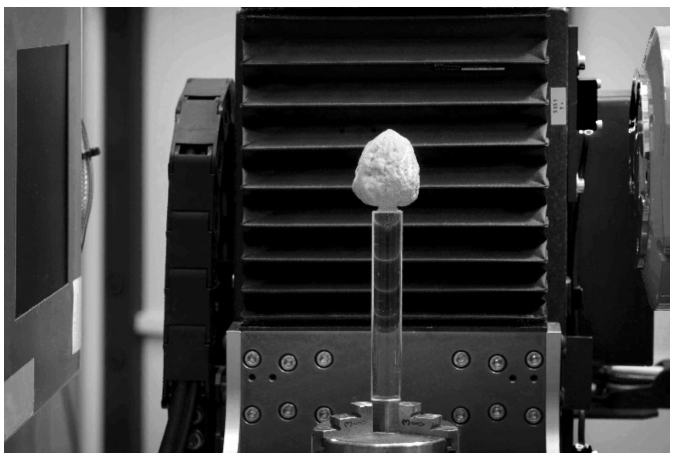
Subsequent Work

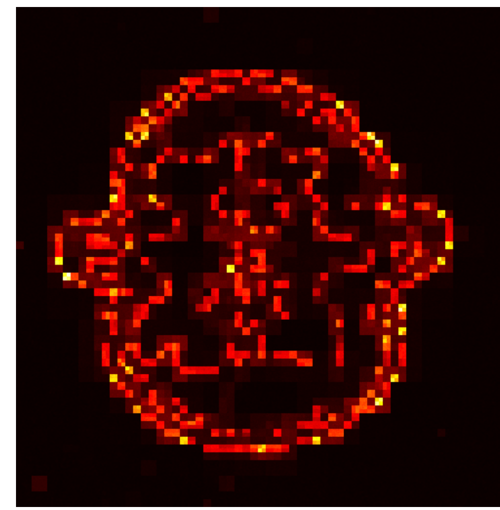
Applying the explicit ℓ_1 sampler to

- ✓ verify theoretical predictions about the infinite dimensional limits of TV priors.
- ✓ examine CT inversion with Besov space priors (cf. **Kolehmainen et al., 2012, Hämäläinen et al., 2013)**:

$$p_{prior}(u) \propto \exp\left(-\lambda \|WV^T u\|_1\right)$$
$$= \prod \exp\left(-\lambda w_i |\left\langle v_i, u \right\rangle|\right)$$

- ✓ demonstrate feasibility of high-dim. Bayesian inversion of experimental data





- ✓ develop new theoretical ideas:

! Thanks to Esa Niemi and Samuli Siltanen!

M. Burger, F. L., 2014.
Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators.
Inverse Problems 30(11); arXiv:1402.5297v2

$$p_{prior}(u) \propto \exp\left(-\lambda \|D^T u\|_1\right)$$

- Limited to D which can be diagonalized by basis V:

$$D^T V = W = \mathrm{diag}(w), \quad u = V\xi$$

$$p_{post}(\xi|f) \propto \exp\left(-\tfrac{1}{2}\|f - AV\xi\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\|W\xi\|_1\right)$$

- D a frame, dictionary etc.? TV in 2D/3D?
- ℓ_p^q-prior: $\exp\left(-\lambda\|D^T u\|_p^q\right)$
- Incorporate hard constraints? Non-negativity?

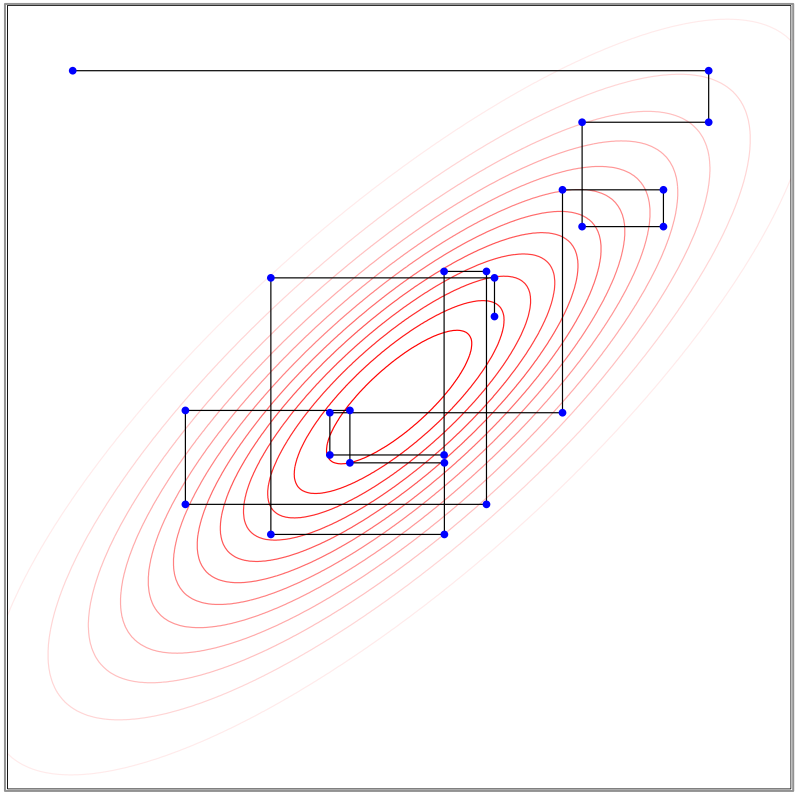
Generalizing SC Gibbs Sampling

Single component Gibbs sampler:

1. Fast computation of explicitly parameterized 1D cond. densities.
2. Fast, robust and exact sampling from 1D densities.

Simple parametrizations of SC densities for

- ℓ_p^q - priors
- ℓ_p^q - block/group priors
- Cauchy/Students-*t* priors
- TV in 2D/3D



Single component Gibbs sampler:

1. Fast computation of explicitly parameterized 1D cond. densities.
2. Fast, robust ~~and exact~~ sampling from 1D densities.

Metropolis-within-Gibbs sampling?

- SC densities vary to much, adapting MH kernel is difficult.

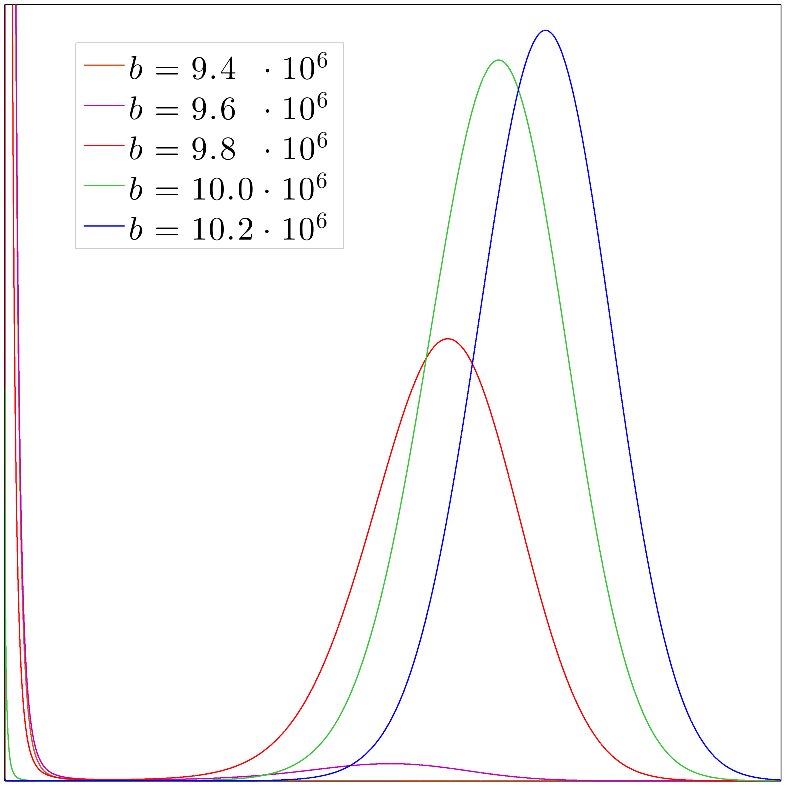
Gibbs-within-Gibbs sampling?

- Gibbs sampler automatically adapts, but cannot be applied to 1D densities!

Trick: Slice sampling

Neal, R.M., 2003. *Slice Sampling* Annals of Statistics 31(3)

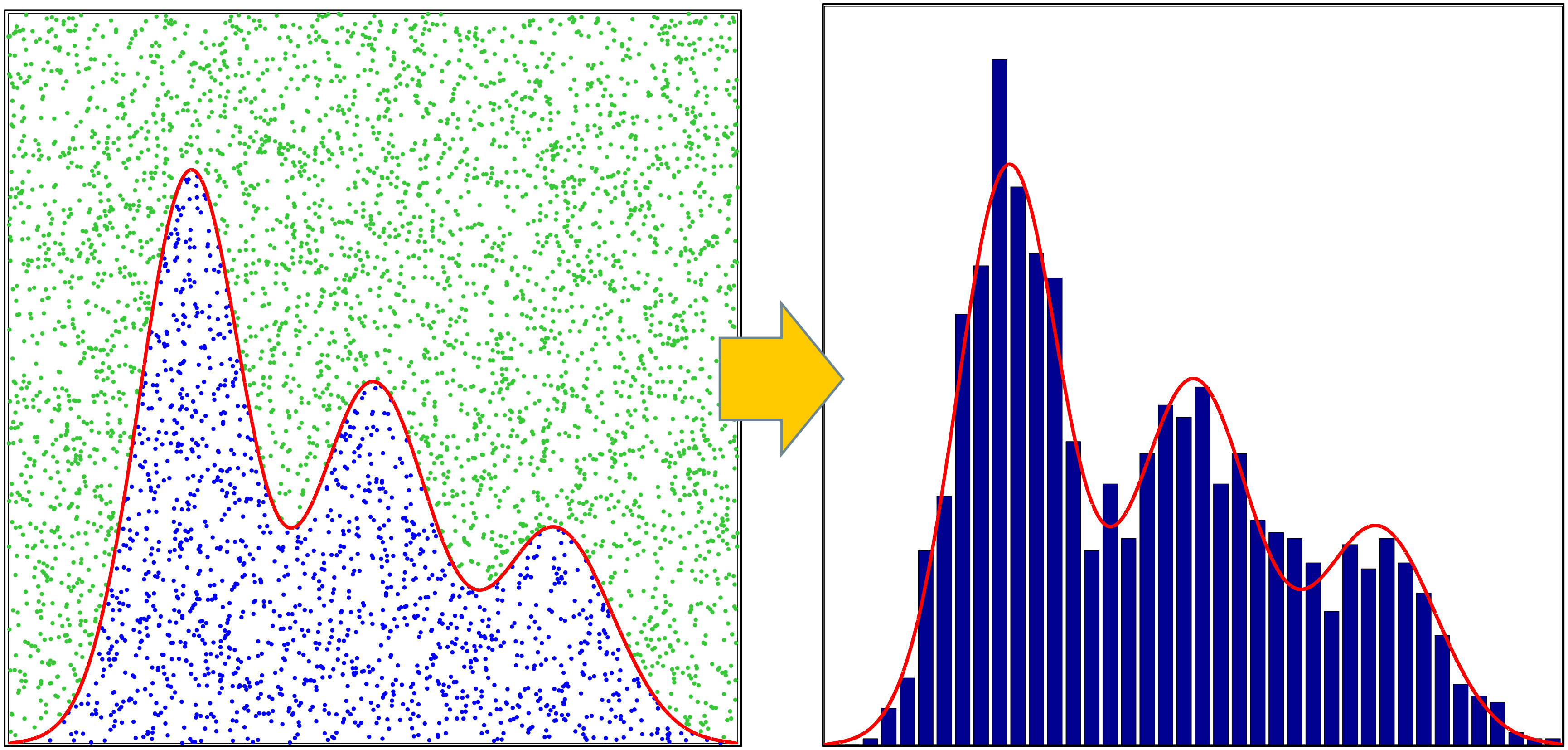
$$p(x) \propto \exp\left(-ax^2 + bx - c|x|^{0.8}\right)$$



Fundamental Theorem of Simulation

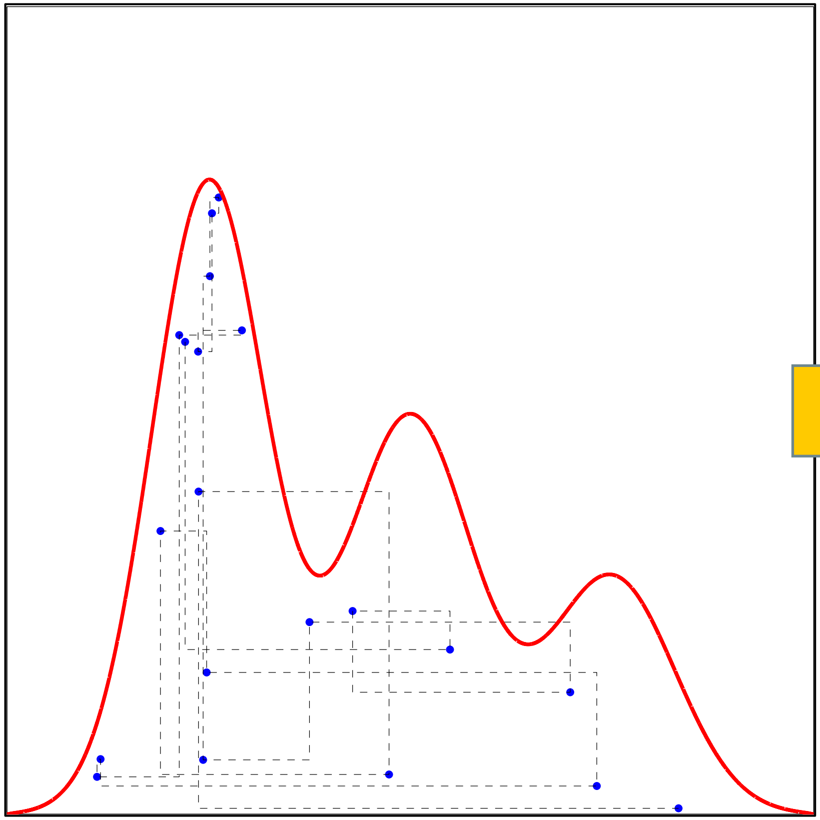
Sampling from $p(x)$ ⟺ sampling uniformly from area below its graph, $\mathcal{G}_p := \{(x,y) | 0 \leqslant y \leqslant p(x)\}$

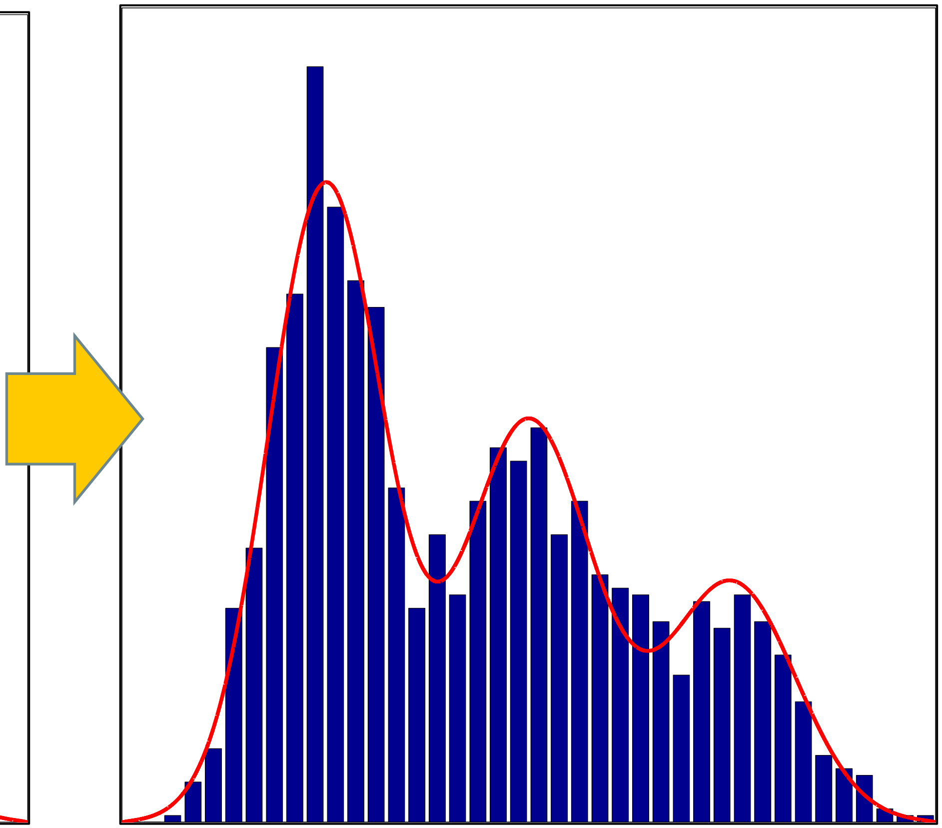
- Basis for *accept-reject* sampling methods
  ! needs uniform sampler for area enclosing $\mathcal{G}_p$!



OR: Run a Gibbs sampler $\tilde{p}(x, y) \propto \mathbf{1}_{\mathcal{G}_p}(x, y)$ and only keep the x samples:

1. draw y uniformly from $[0, p(x^i)]$ (vertical move)
2. draw x^{i+1} uniformly from $S^y := \{z \mid p(z) \geqslant y\}$ (horizontal move)

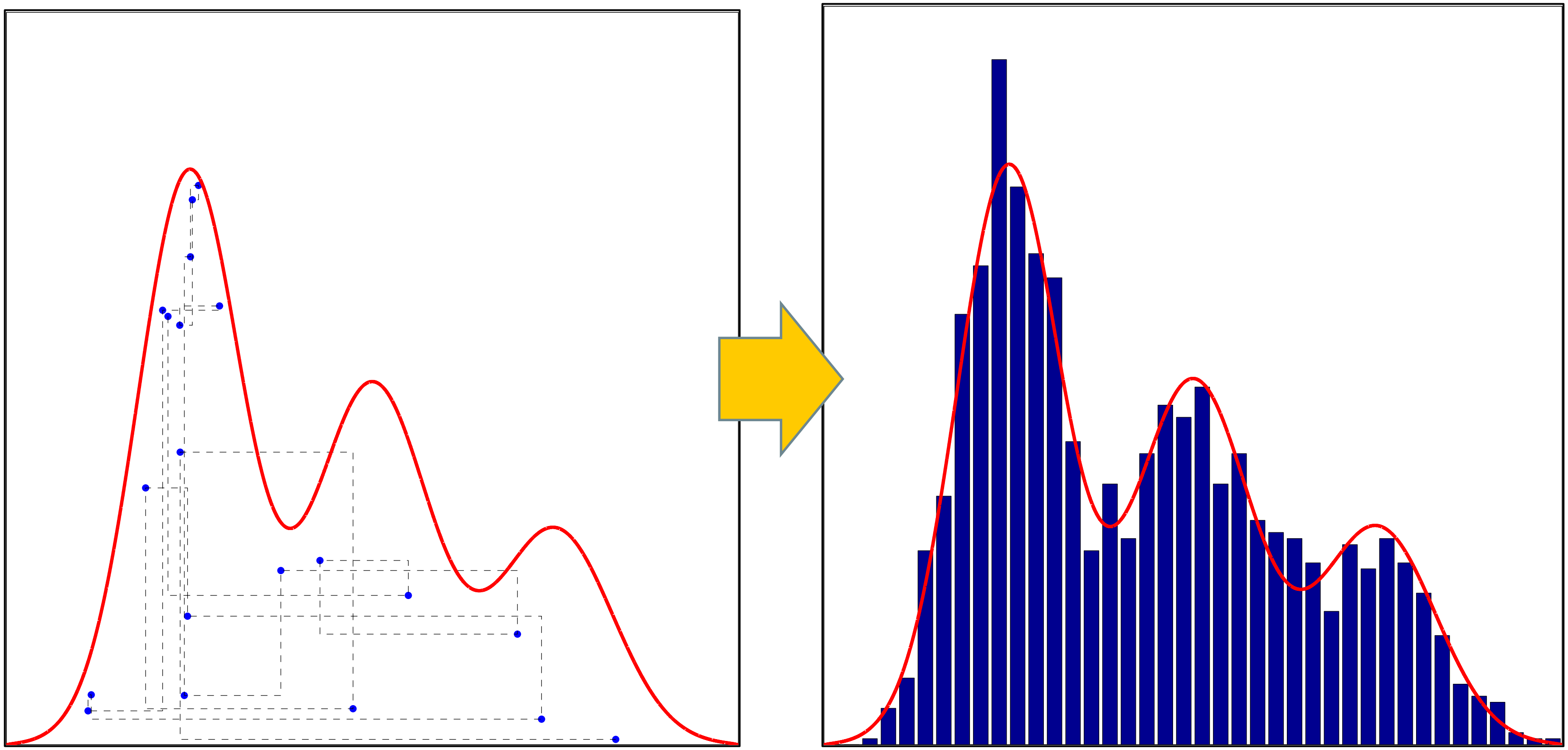




OR: Run a Gibbs sampler $\tilde{p}(x,y) \propto \mathbf{1}_{\mathcal{G}_p}(x,y)$ and only keep the x samples:

1. draw y uniformly from $[0, p(x^i)]$ (vertical move)
2. draw x^{i+1} uniformly from $S^y := \{z \mid p(z) \geqslant y\}$ (horizontal move)

Horizontal move not easy to implement for our problems, but...



Generalized Slice Sampling

...Slice sampling is a variant of auxiliary variables algorithms:

- Introduce additional variable y with suitable $p(y|x)$.
- Sample $p(x,y) = p(y|x)p(x)$ by a Gibbs sampler (needs $p(y|x)$, $p(x|y)$)
- Keep only $\{x^i\}$

Generalized Slice Sampling

...Slice sampling is a variant of auxiliary variables algorithms:

- Introduce additional variable y with suitable $p(y|x)$.
- Sample $p(x,y) = p(y|x)p(x)$ by a Gibbs sampler (needs $p(y|x)$, $p(x|y)$)
- Keep only $\{x^i\}$

Basic slice sampler:

$$p(x)$$

$$p(y|x) = \frac{\mathbf{1}_{[0,p(x)]}(y)}{p(x)}$$

$$\Longrightarrow p(x,y) = p(x)\frac{1}{p(x)}\mathbf{1}_{[0,p(x)]}(y)$$

$$\Longrightarrow p(x|y) \propto \mathbf{1}_{[0,p(x)]}(y) = \mathbf{1}_{\{x|p(x)\geqslant y\}}(x)$$

Generalized Slice Sampling

...Slice sampling is a variant of auxiliary variables algorithms:

- Introduce additional variable y with suitable $p(y|x)$.
- Sample $p(x,y) = p(y|x)p(x)$ by a Gibbs sampler (needs $p(y|x)$, $p(x|y)$)
- Keep only $\{x^i\}$

Basic slice sampler:

$$p(x)$$

$$p(y|x) = \frac{\mathbf{1}_{[0,p(x)]}(y)}{p(x)}$$

$$\Longrightarrow p(x,y) = p(x)\frac{1}{p(x)}\mathbf{1}_{[0,p(x)]}(y)$$

$$\begin{aligned}\Longrightarrow p(x|y) &\propto \mathbf{1}_{[0,p(x)]}(y) \\ &= \mathbf{1}_{\{x|p(x)\geqslant y\}}(x)\end{aligned}$$

Factorizing slice sampler:

$$p(x) = p_1(x)p_2(x)$$

$$p(y|x) = \frac{\mathbf{1}_{[0,p_2(x)]}(y)}{p_2(x)}$$

$$\Longrightarrow p(x,y) = p_1(x)\mathbf{1}_{[0,p_2(x)]}(y)$$

$$\Longrightarrow p(x|y) \propto p_1(x)\mathbf{1}_{\{x|p_2(x)\geqslant y\}}(x)$$

Slice-Within-Gibbs Sampler

For linear inverse problems, most SC densities have the form

$$p(x) \propto \exp\left(-ax^2 + bx - \phi(x)\right) = \underbrace{\exp\left(-ax^2 + bx\right)}_{p_1(x)} \underbrace{\exp\left(-\phi(x)\right)}_{p_2(x)}$$

For $i = 1, \ldots, K$ do:

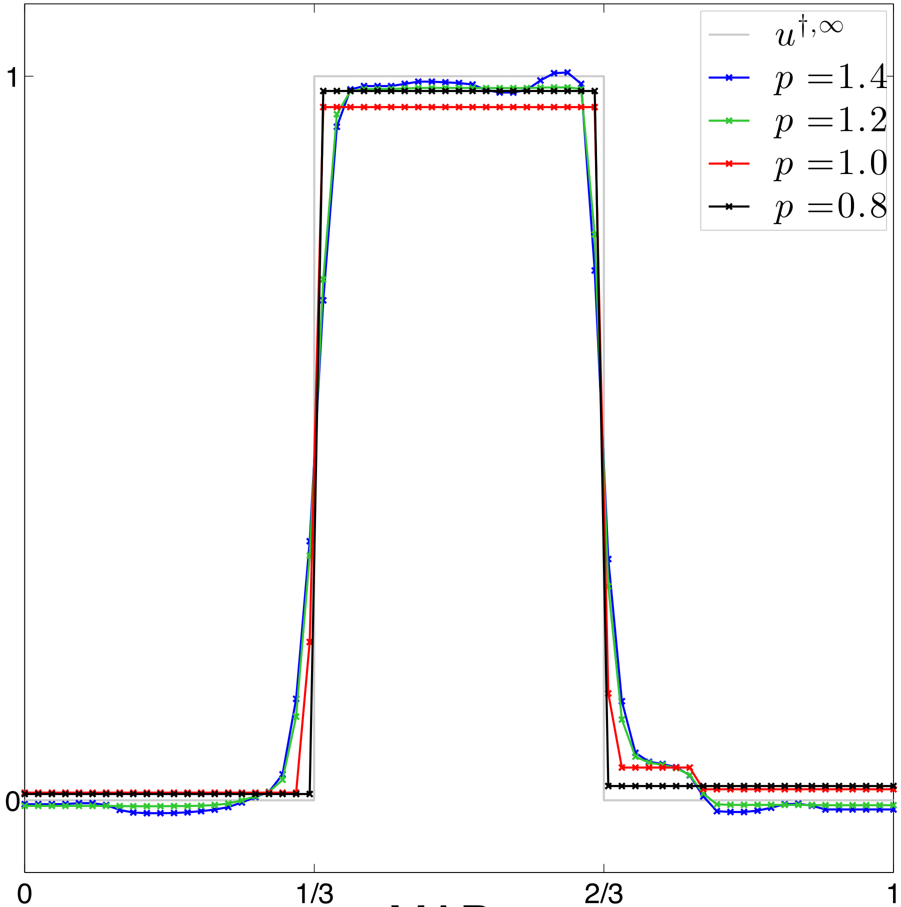
1. draw y uniformly from $[0, p_2(x^i)]$ (vertical move)
2. draw x^{i+1} from $p_1(x)$ truncated to S_2^y (weighted horizontal move)

Return x^K as update to Gibbs sampler.

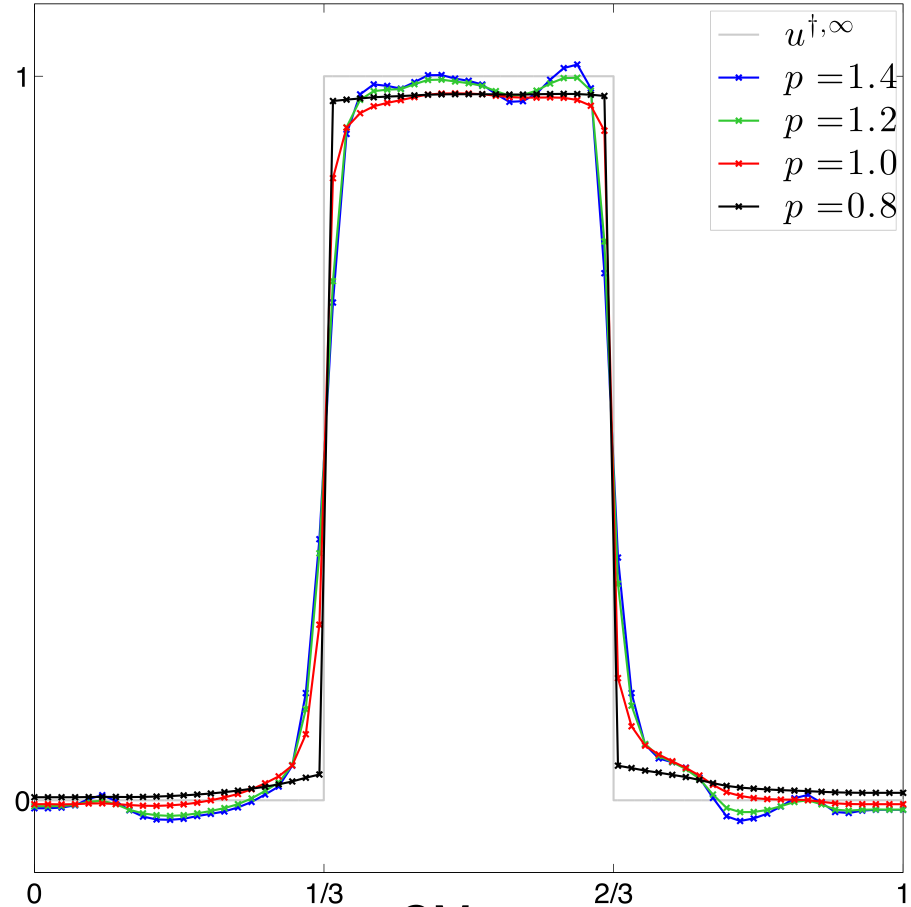
Involves

- Explicit or numerical computation of $S_2^y := \{z \mid p_2(z) \geqslant y\}$
- Fast and robust sampling from truncated Gaussians (**Chopin, 2010)**
- Hard-constraints can be easily incorporated.

$$p_{prior}(u) \propto \exp\left(-\lambda \|\nabla u\|_p^p\right)$$



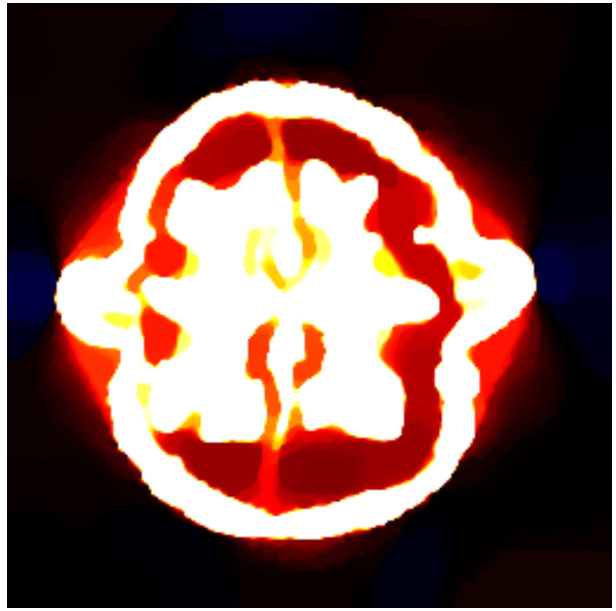
MAP



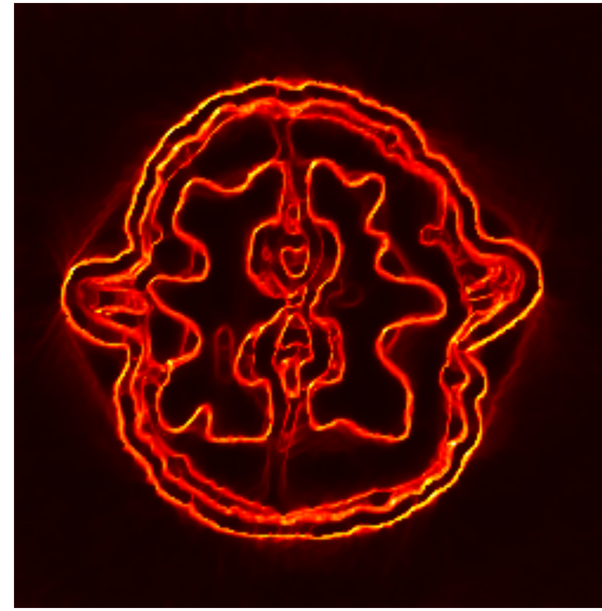
CM

Experimental CT Data with TV Priors I

MAP, 256x256

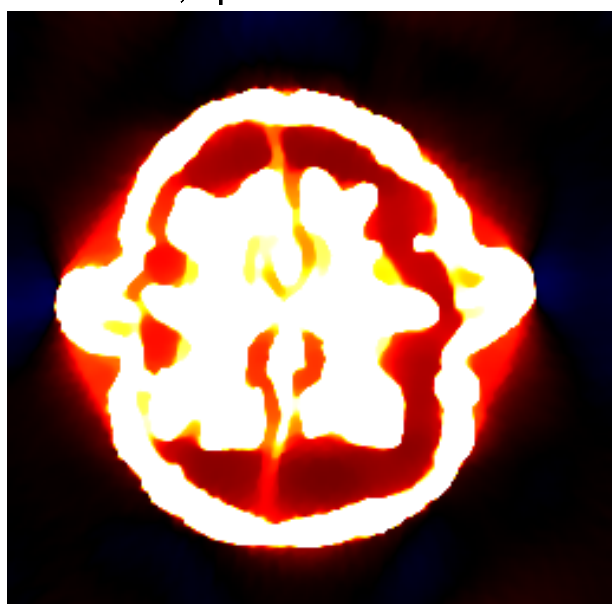


MAP, special color scale



CStd, 256x256

CM, 256x256



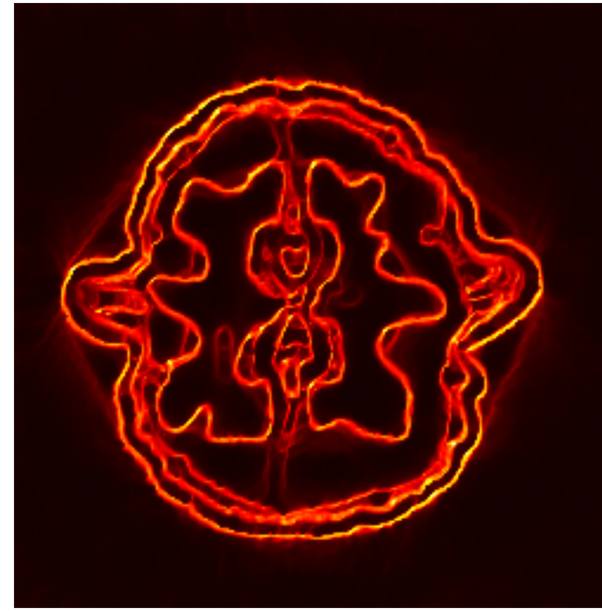
CM, special color scale

CM of $\|\nabla u\|_2$

Experimental CT Data with TV Priors II

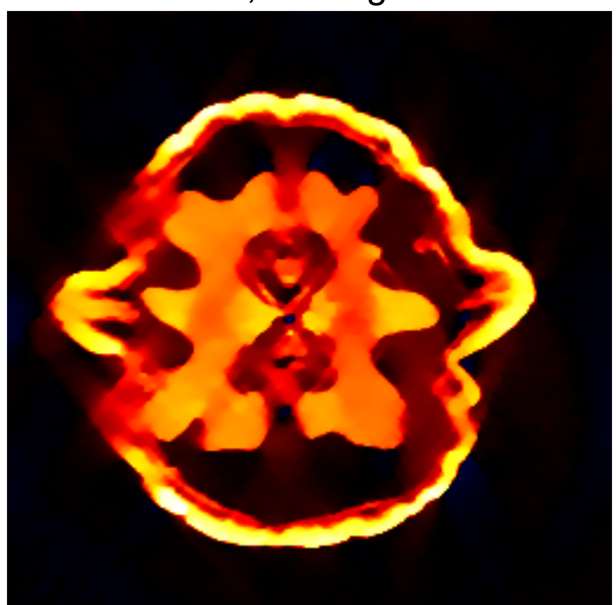
MAP, full angle

CM, full angle

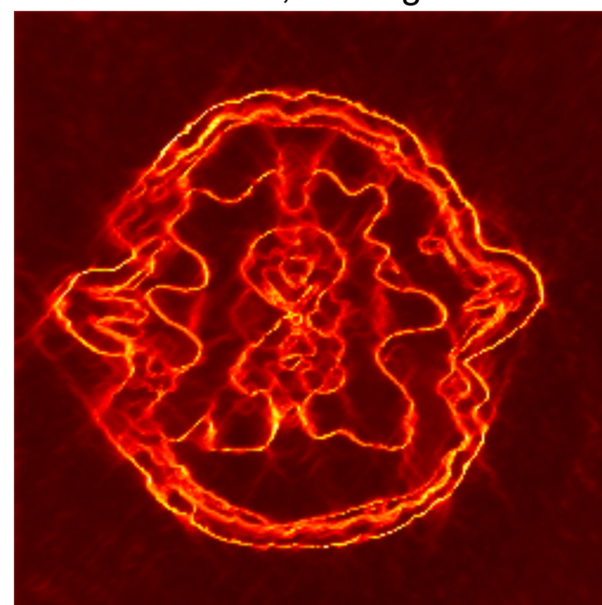


CStd, full angle

MAP, limited angle

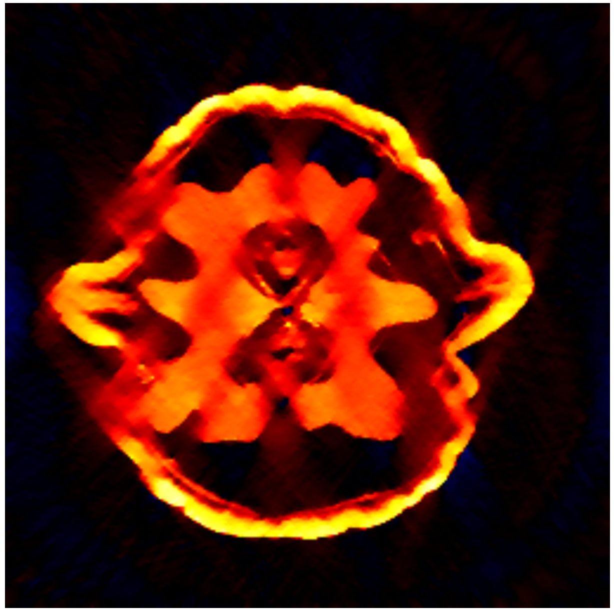


CM, limited angle

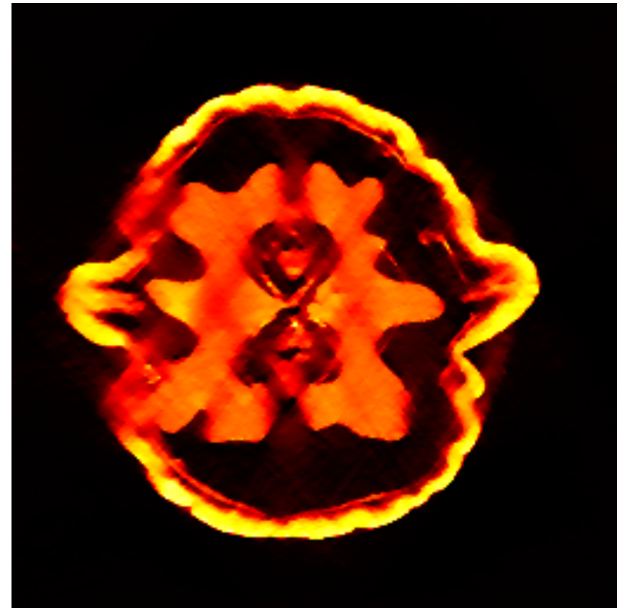


CStd, limited angle

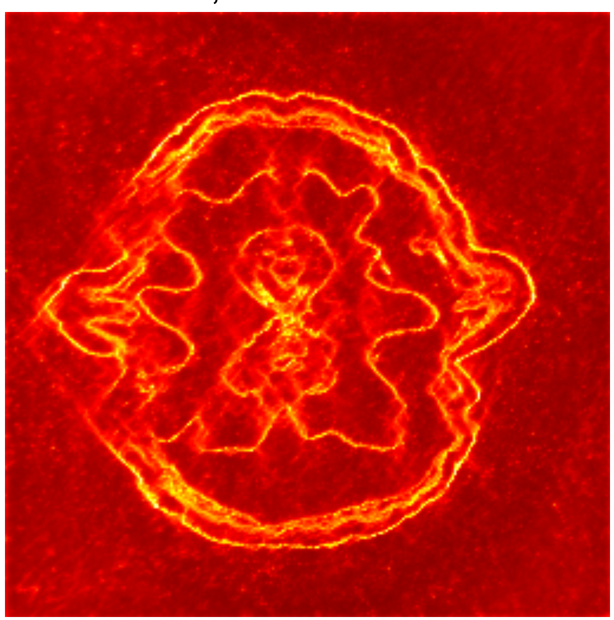
Experimental CT Data with TV Priors III



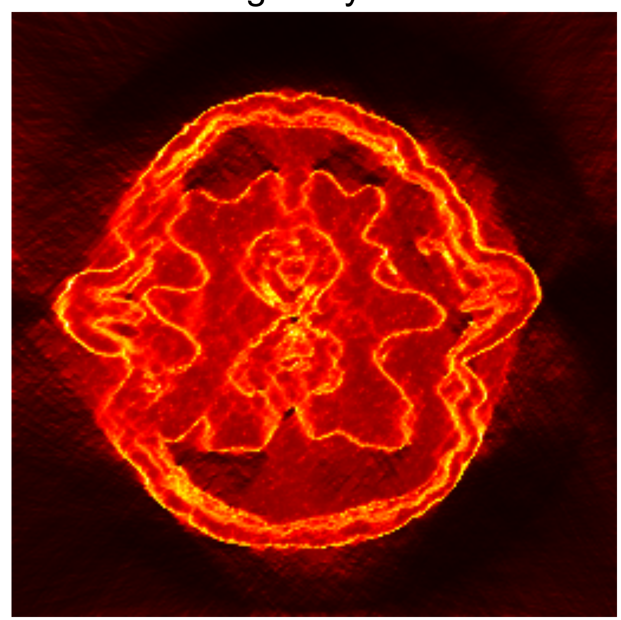
CM, no constraints



CM non-negativity constraints



CStd, no constraints



CStd, non-negativity constraints

Discussion, Conclusion & Outlook

Methods:

- Sample-based Bayesian inversion in high dimensions is feasible if tailored samplers are developed.
- Auxiliary variables/splitting schemes like slice sampling also promising tools for sampling (cf. splitting in optimization).
- Potential generalizations:
  - Parallelization
  - Adaptive Gibbs sampling
  - Multi-grid/multi-resolution schemes
  - Simulated annealing
  - Non-linear problems and non-Gaussian noise

Theory:

- TV(-p) in 2D/3D?
- Non log-concave priors

Applications:

- Proof-of-concept character up to now
- Specific imaging task to explore full potential

Thanks for your attention!

F. L., 2014.
Bayesian Inversion in Biomedical Imaging.
PhD thesis (**submitted today**)

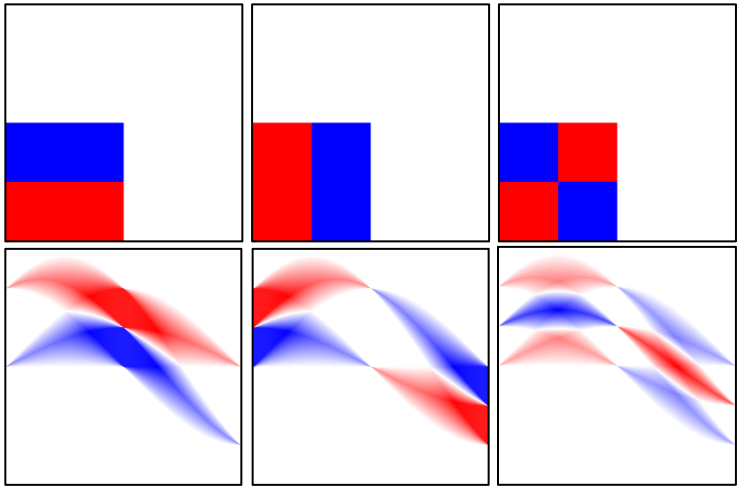
M. Burger, F. L., 2014.
Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators.
Inverse Problems 30(11); arXiv:1402.5297v2

F. L., 2012.
Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors.
Inverse Problems 28(12); arXiv:1206.0262v2

Subsequent Work: Application to CT

V. Kolehmainen, M. Lassas, K. Niinimäki, S. Siltanen, 2012.
Sparsity-promoting Bayesian inversion. Inverse Problems 28(2)

K. Hämäläinen, A. Kallonen & the above, 2013.
Sparse Tomography. SIAM J Sci Comput 35(3)



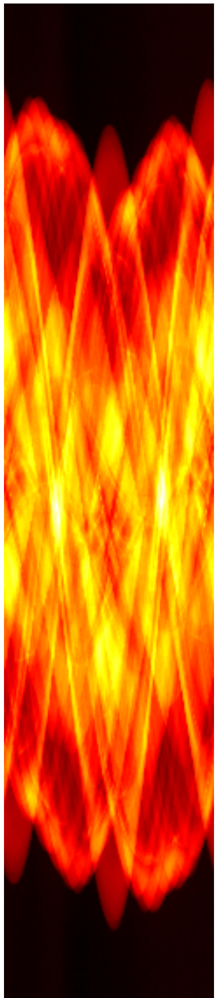
Besov space priors:

$$p_{prior}(u) \propto \exp\left(-\lambda \|WV^T u\|_1\right) = \prod \exp\left(-\lambda w_i |\langle v_i, u\rangle|\right)$$

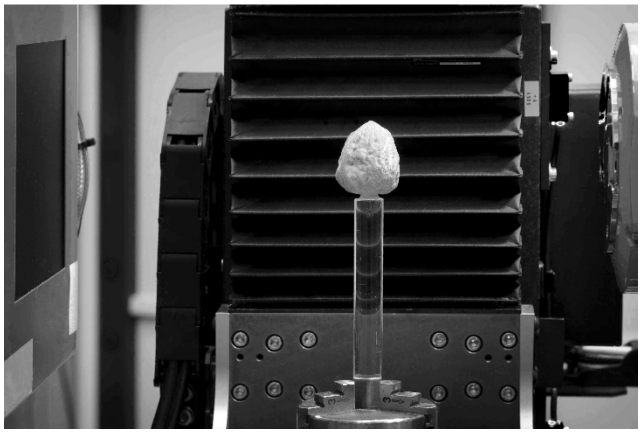
- Compute Av_i in a fast (matrix-free) way.

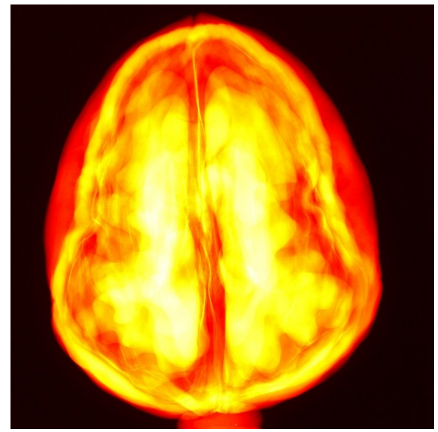
Subsequent Work: Application to CT

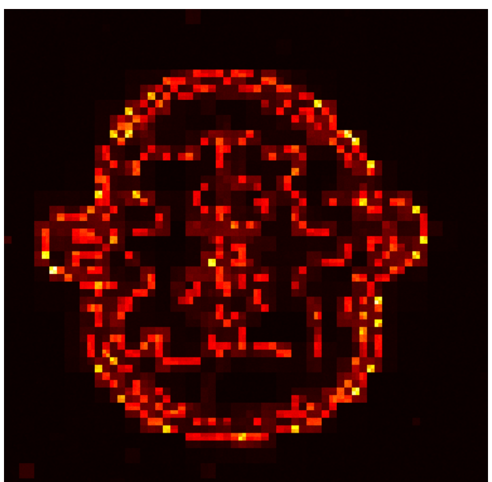
K. Hämäläinen, A. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen, 2013. *Sparse Tomography.* SIAM J Sci Comput 35(3)



Challenge: Fan beam geometry (...and other real-world stuff)





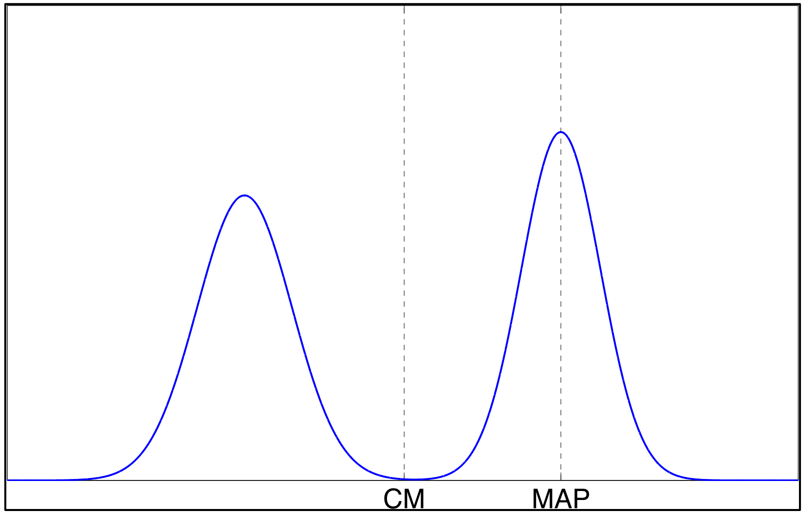


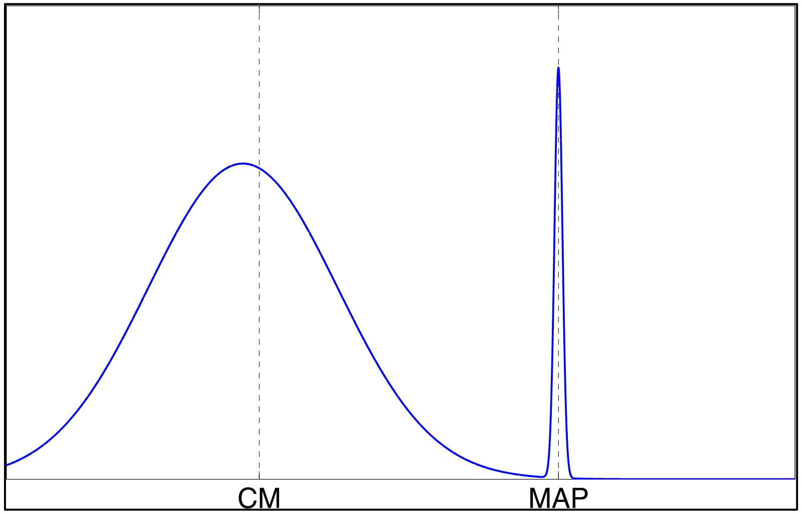
! Thanks to Esa Niemi and Samuli Siltanen for data and pics !

Subsequent Work: MAP vs. CM Studies

M. Burger, F. L., 2014.
Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators.
Inverse Problems 30(11); arXiv:1402.5297v2

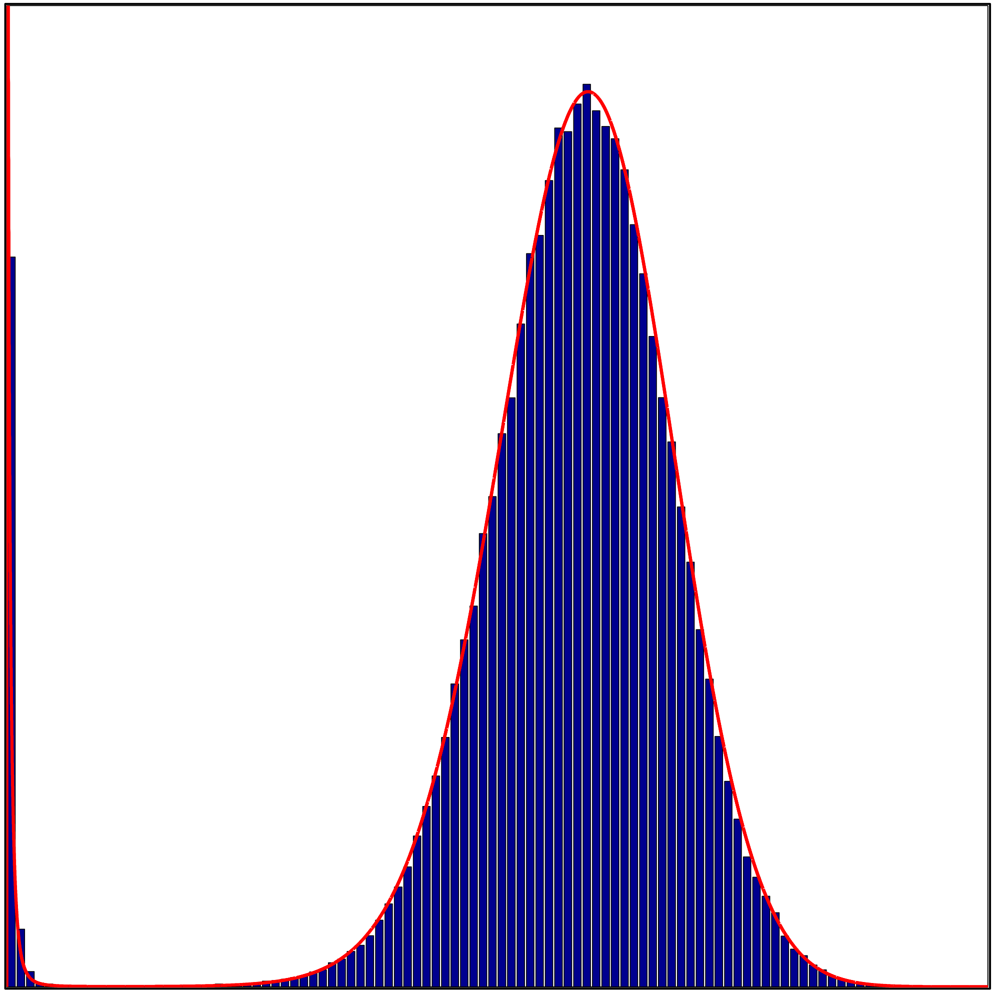
- "MAP or CM?" = "Variational regularization vs. Bayesian inference"?
- Contrasts classical believes with recent computational results.
- Develops new Bayes cost functions based on Bregman distances.
- Both MAP and CM are proper Bayes estimators for convex cost functions.





Slice-Within-Gibbs Sampler

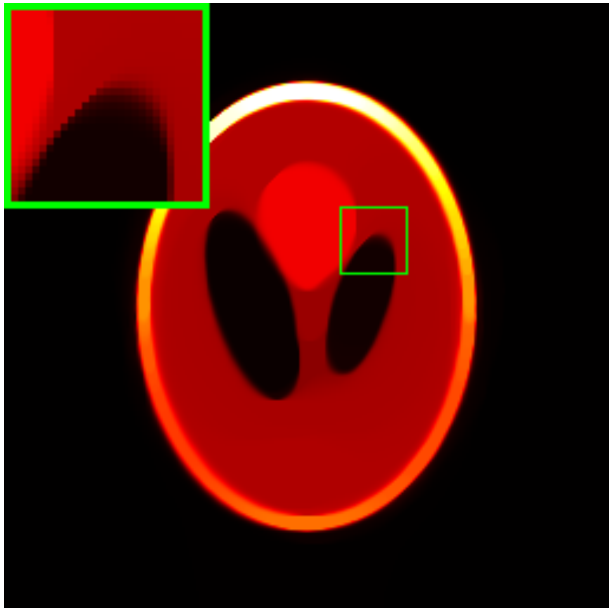
- ✓ Compare histograms with pdf to verify that it works
- ✓ Compute autocorrelation dependent on number of slice sampling steps to validate performance:
  - Good performance for log-concave priors.



TV Discretization Invariance Studies in 2D

MAP, 64x64

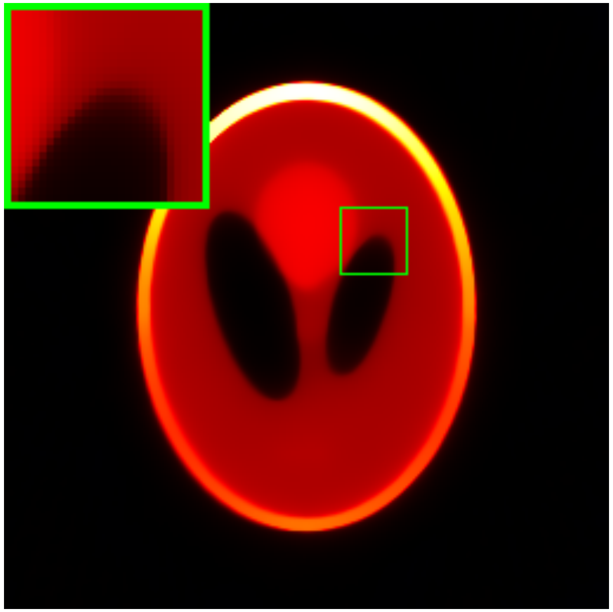
MAP, 128x128



MAP, 256x256

CM, 64x64

CM, 128x128



CM, 256x256