Simulation and optimisation in imaging inverse problems: Part 1.

Marcelo Pereyra
http://www.stats.bris.ac.uk/~mp12320/

University of Bristol

7th of July 2016, Peyresq, France.

Images are challenging physical measurements, not pictures!

Outline

1. Bayesian inference in imaging inverse problems
2. Proximal Markov chain Monte Carlo
3. Experiments
4. Conclusion

Outline

The Bayesian framework

- We are interested in an unknown image $\mathbf{x} \in \mathbb{R}^d$.
- We observe data $\mathbf{y}$, related to $\mathbf{x}$ by a statistical model $p(\mathbf{y}|\mathbf{x})$.
- The recovery of $\mathbf{x}$ from $\mathbf{y}$ is ill-posed or ill-conditioned.

- We address this difficulty by using a prior distribution $p(\mathbf{x})$.
- The posterior distribution of $\mathbf{x}$ given $\mathbf{y}$

$$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})/p(\mathbf{y})$$

models our knowledge about $\mathbf{x}$ after observing $\mathbf{y}$.

Inverse problems in mathematical imaging

Many imaging inverse problems involve models of the form

$$\pi(\mathbf{x}|\mathbf{y}) \propto \exp\{-g_1(\mathbf{x}) - g_2(\mathbf{x})\} \tag{1}$$

where $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ are lower semicontinuous convex functions from $\mathbb{R}^d \to (-\infty, +\infty]$. Typically g_1 is L-Lipschitz differentiable, e.g.,

$$g_1(\mathbf{x}) = \frac{1}{2\sigma^2}\|\mathbf{y} - A\mathbf{x}\|_2^2$$

for some observation $\mathbf{y} \in \mathbb{R}^p$ and linear operator $A \in \mathbb{R}^{p\times n}$, and

$$g_2(\mathbf{x}) = \alpha\|B\mathbf{x}\|_\dagger + \mathbf{1}_{\mathcal{S}}(\mathbf{x})$$

for some norm $\|\cdot\|_\dagger$, dictionary $B \in \mathbb{R}^{n\times n}$, and convex set $\mathcal{S}$. Often, $g_2 \notin \mathcal{C}^1$.

Maximum-a-posteriori (MAP) estimation

The predominant Bayesian approach in imaging is MAP estimation

$$\begin{aligned} \hat{\mathbf{x}}_{MAP} &= \operatorname*{argmax}_{\mathbf{x}\in\mathbb{R}^d} \pi(\mathbf{x}|\mathbf{y}), \\ &= \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^d} g_1(\mathbf{x}) + g_2(\mathbf{x}), \end{aligned} \tag{2}$$

which can be computed very efficiently (e.g. within milliseconds), even for large n, by using optimisation algorithms based on the following mapping:

Definition 1.1 (Proximity mappings (Moreau, 1962))

For $\lambda > 0$, the λ-proximity mapping of g convex l.s.c. is defined as

$$\operatorname{prox}_g^{\lambda}(\mathbf{x}) \triangleq \operatorname*{argmin}_{\mathbf{u}\in\mathbb{R}} g(\mathbf{u}) + \frac{1}{2\lambda}\|\mathbf{u}-\mathbf{x}\|^2.$$

See Combettes and Pesquet (2011) for list of proximity mappings.

MAP estimation by convex optimisation

Proximal gradient (forward-backward) algorithm

$$\mathbf{x}^{m+1} = \mathrm{prox}_{g_2}^{L^{-1}}\{\mathbf{x}^m + L^{-1}\nabla g_1(\mathbf{x}^m)\},$$

converges to $\hat{\mathbf{x}}_{MAP}$ at rate $O(1/m)$, with poss. acceleration to $O(1/m^2)$.

Alternating direction method of multipliers (ADMM) algorithm

$$\begin{aligned}
\mathbf{x}^{m+1} &= \mathrm{prox}_{g_1}^{\lambda}\{\mathbf{z}^m - \mathbf{u}^m\},\\
\mathbf{z}^{m+1} &= \mathrm{prox}_{g_2}^{\lambda}\{\mathbf{x}^{m+1} + \mathbf{u}^m\},\\
\mathbf{u}^{m+1} &= \mathbf{u}^m + \mathbf{x}^{m+1} - \mathbf{z}^{m+1},
\end{aligned}$$

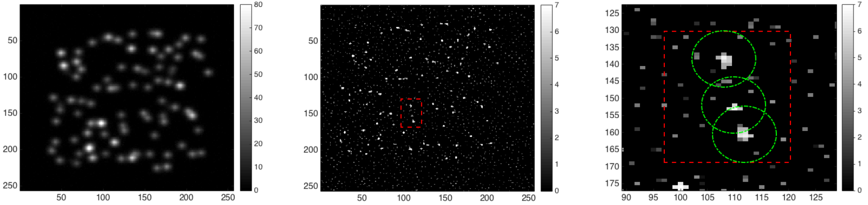
also converges to $\hat{\mathbf{x}}_{MAP}$, and does not require g_1 to be smooth.

Recover $\mathbf{x} \in \mathbb{R}^d$ from low resolution and noisy measurements

$$\mathbf{y} = H\mathbf{x} + \mathbf{w},$$

where H is a circulant blurring matrix. We use the Bayesian model

$$\pi(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\|\mathbf{y} - H\mathbf{x}\|^2/2\sigma^2 - \beta\|\mathbf{x}\|_1\right). \tag{3}$$



$\mathbf{y}$ $\hat{\mathbf{x}}_{MAP}$ Uncertainty estimates

Figure : Resolution enhancement of the `Molecules` image of size 256×256 pixels.

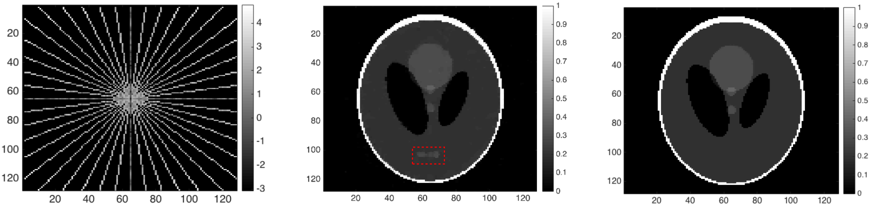
Recover $\mathbf{x} \in \mathbb{R}^d$ from partially observed and noisy Fourier measurements

$$\mathbf{y} = \Phi \mathcal{F} \mathbf{x} + \mathbf{w},$$

where Φ is a mask and $\mathcal{F}$ is the 2D Fourier operator. We use the model

$$\pi(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\|\mathbf{y} - \Phi\mathcal{F}\mathbf{x}\|^2/2\sigma^2 - \beta\|\nabla_d \mathbf{x}\|_{1-2}\right), \tag{4}$$

where ∇_d is the 2d discrete gradient operator and $\|\cdot\|_{1-2}$ the $\ell_1 - \ell_2$ norm.



$\mathbf{y}$ $\hat{\mathbf{x}}_{MAP}$ Possible solution?

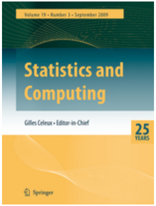
Figure : Tomographic reconstruction of the Shepp-Logan phantom image.

Limitations

- Proximal optimisation algorithms deliver accurate approximations of $\hat{\mathbf{x}}_{MAP}$ efficiently. However, $\hat{\mathbf{x}}_{MAP}$ provides very little about $\pi(\mathbf{x}|\mathbf{y})$.
- More advanced statistical analyses require other inference tools (e.g. MCMC algorithms) that are often very computationally expensive.
- High-dimensional MCMC methods rely strongly on differential calculus and may perform badly if $\pi(\mathbf{x}|\mathbf{y})$ is not sufficiently regular.
- This talk describes "proximal" MCMC algorithms (Pereyra, 2015; Durmus et al., 2016), which exploit convex analysis for simulation.

Modern Bayesian computation

Recent surveys on Bayesian computation...



25th anniversary special issue on Bayesian computation

P. Green, K. Latuszynski, M. Pereyra, C. P. Robert, "Bayesian computation: a perspective on the current state, and sampling backwards and forwards", Statistics and Computing, vol. 25, no. 4, pp 835-862, Jul. 2015.

Special issue on "Stochastic simulation and optimisation in signal processing"

M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. Hero, and S. McLaughlin, "A Survey of Stochastic Simulation and Optimization Methods in Signal Processing" IEEE Sel. Topics in Signal Processing, in press.

Outline

Inference by Markov chain Monte Carlo integration

Monte Carlo integration
Given a set of samples $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}$ distributed according to $p(\mathbf{x}|\mathbf{y})$, we approximate posterior expectations and probabilities

$$\frac{1}{M}\sum \phi(\mathbf{x}^{(m)}) \to \mathrm{E}\{\phi(\mathbf{x})|\mathbf{y}\}, \quad \text{as } M \to \infty$$

Guarantees from CLTs [e.g., $\frac{1}{\sqrt{M}}\sum \phi(\mathbf{x}^{(m)}) \sim \mathcal{N}(\mathrm{E}\{\phi(\mathbf{x})|\mathbf{y}\}, \Sigma)$].

Markov chain Monte Carlo:
Construct a Markov kernel $\mathbf{x}^{(m+1)}|\mathbf{x}^{(m)} \sim K(\cdot|\mathbf{x}^{(m)})$ such that the Markov chain $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}$ has $p(\mathbf{x}|\mathbf{y})$ as stationary distribution.

MCMC simulation in high-dimensional spaces is very challenging.

Unadjusted Langevin algorithm

- Suppose that $\pi \in \mathcal{C}^1$. We could simulate from π by mimicking a Langevin diffusion process that converges to π as $t \to \infty$

$$X: \quad dX(t) = \frac{1}{2}\nabla \log \pi\left(X(t)\right) dt + dW(t), \quad 0 \le t \le T, \quad X(0) = \mathbf{x}_0.$$

- Direct simulation from y is generally not possible. Instead, we use a forward Euler approximation of X ("unadjusted Langevin algorithm")

$$\mathrm{ULA}: \quad \mathbf{x}^{(m+1)} = \mathbf{x}^{(m)} + \delta\nabla \log \pi(\mathbf{x}^{(m)}) + \sqrt{2\delta}\mathbf{z}^{(m)}, \quad \mathbf{z}^{(m)} \sim \mathcal{N}(0, \mathbb{I}_d)$$

- However, ULA may perform badly if $\pi \notin \mathcal{C}^1$, or if $\nabla \log \pi$ is not Lipchitz continuous (e.g., if $\pi(x) \propto \exp\left(-\gamma x^{\beta}\right)$ with $\beta > 2$).

Metropolis-adjusted Langevin algorithm (MALA)

MALA combines ULA with a Metropolis-Hastings step that removes (asymptotically) the bias due to the discretisation:

1. 1) Use ULA to generate candidate

$$\mathbf{x}^* = \mathbf{x}^{(m)} + \delta \nabla \log \pi(\mathbf{x}^{(m)}) + \sqrt{2\delta}\mathbf{z}^{(m)}, \quad \mathbf{z}^{(m)} \sim \mathcal{N}(0, \mathbb{I}_d)$$

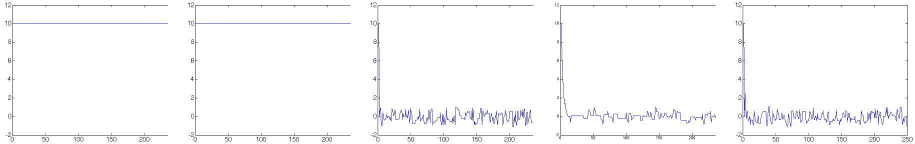
2. 2) With probability

$$\rho^{(m+1)} = 1 \wedge \frac{\pi(\mathbf{x}^*)}{\pi[\mathbf{x}^{(m)}]} \frac{p_{\mathcal{N}}[\mathbf{x}^{(m)} | \mathbf{x}^* + \delta \nabla \log \pi(\mathbf{x}^*), 2\delta \mathbb{I}_d]}{p_{\mathcal{N}}[\mathbf{x}^* | \mathbf{x}^{(m)} + \delta \nabla \log \pi(\mathbf{x}^{(m)}), 2\delta \mathbb{I}_d]}$$

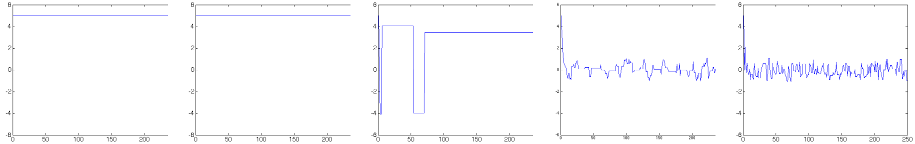
Set $\mathbf{x}^{(m+1)} = \mathbf{x}^*$. Otherwise, set $\mathbf{x}^{(m+1)} = \mathbf{x}^{(m)}$.

However, MALAs (and Hamiltonian MCs) often also perform badly if $\pi \notin \mathcal{C}^1$, or if $\nabla \log \pi$ is not Lipchitz continuous !!!

Illustrative example - $\pi(x) \propto \exp\{-x^4\}$:



(a) MALA (b) HMC (c) MALTA (d) SMMALA (f) Px-MALA



(a) MALA (b) HMC (c) MALTA (d) SMMALA (f) Px-MALA

Comparison: MALA, Hamiltonian MC (Neal, 2012), ϵ-truncated gradient MALA (MALTA) (Roberts and Tweedie, 1996), simplified manifold MALA (SMMALA) (Girolami and Calderhead, 2011) and proximal MALA (Pereyra, 2015).

Moreau approximations

Idea: Regularise π to enable high-dimensional MCMC sampling.

Definition 2.1

Moreau approximations of π

We define the λ-Moreau approximation of π as the following density

$$\pi_{\lambda}(\mathbf{x}) \triangleq \sup_{\mathbf{u}\in\mathbb{R}} \frac{1}{\kappa'}\pi(\mathbf{u}) \exp\left[-\frac{1}{2\lambda}\|\mathbf{u}-\mathbf{x}\|^2\right] \tag{5}$$

with normalizing constant $\kappa' \in \mathbb{R}^+$ and regularisation parameter $\lambda > 0$.

Moreau approximations

Key properties:

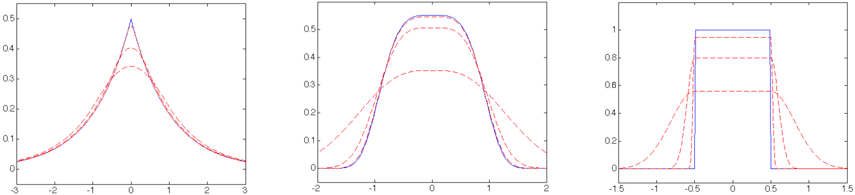
1. *Differentiability*:
   - $\pi_\lambda \in \mathcal{C}^1$ even if π not differentiable, with

$$\nabla \log \pi_\lambda(\mathbf{x}) = \{\mathrm{prox}_g^\lambda(\mathbf{x}) - \mathbf{x}\}/\lambda.$$

   - $\nabla \log \pi_\lambda(\mathbf{x})$ is $1/\lambda$-Lipchitz continuous.

2. *Convergence to* π:
   - $\lim_{\lambda\to 0} \|\pi_\lambda - \pi\|_{TV} = 0$.
   - If $g(\mathbf{x}) = -\log \pi(\mathbf{x})$ is L-Lipchitz, then $\|\pi_\lambda - \pi\|_{TV} \leq \lambda L^2$.

Examples of Moreau approximations:



$\pi(x) \propto \exp\left(-|x|\right)$ $\qquad$ $\pi(x) \propto \exp\left(-x^4\right)$ $\qquad$ $\pi(x) \propto \mathbf{1}_{[-0.5,0.5]}(x)$

Figure : True densities (solid blue) and Moreau approximations (dashed red).

Proximal ULA

Idea: Approximate X with a "regularised" auxiliary Langevin diffusion X_λ with ergodic measure

$$\pi_\lambda^*(\mathbf{x}) \propto \pi_1(\mathbf{x})\pi_{2,\lambda}(\mathbf{x})$$

using a factorisation $\pi(\mathbf{x}) = \pi_1(\mathbf{x})\pi_2(\mathbf{x})$ such that

$$\pi_1(\mathbf{x}) \propto \exp\{-g_1(\mathbf{x})\}$$

is "easy", i.e., with $g_1 \in \mathcal{C}^1$, convex, and ∇g_1 L_1-Lipschitz, and

$$\pi_2(\mathbf{x}) \propto \exp\{-g_2(\mathbf{x})\}$$

with g_2 l.s.c, convex, and with tractable proximity mapping.

We can make X_λ and π_λ^* arbitrarily close to X and π.

Proximal ULA

We use an Euler approximation of X_λ to simulate from $\pi_\lambda^* \approx \pi$

$$\mathbf{x}^{(m+1)} = \mathbf{x}^{(m)} + \delta \nabla \log \pi_\lambda \{\mathbf{x}^{(m)}\} + \sqrt{2\delta}\mathbf{z}^{(m)}, \quad \mathbf{z}^{(m)} \sim \mathcal{N}(0, \mathbb{I}_d).$$

Replacing $\nabla \log \pi_{2,\lambda}(\mathbf{x}) = \{\text{prox}_{g_2}^\lambda(\mathbf{x}) - \mathbf{x}\}/\lambda$ leads to the (Moreau-Yoshida regularised) proximal ULA

$$\text{MYULA}: \quad \mathbf{x}^{(m+1)} = (1 - \tfrac{\delta}{\lambda})\mathbf{x}^{(m)} - \delta \nabla g_1\{\mathbf{x}^{(m)}\} + \tfrac{\delta}{\lambda}\,\text{prox}_{g_2}^\lambda\{\mathbf{x}^{(m)}\} + \sqrt{2\delta}\mathbf{z}^{(m)}.$$

Stability condition: step-size $\delta \leq \delta_\lambda^{max} = (L_1 + 1/\lambda)^{-1}$.

Rule of thumb: set $\lambda = L_1^{-1}$ and $\delta \in [L_1^{-1}/10, L_1^{-1}/2]$.

Some fundamental questions

Starting from some arbitrary initial condition $\mathbf{x}_0 \in \mathbb{R}^d$, we perform $M \in \mathbb{N}$ iterations of MYULA targeting $\pi^*_\lambda \approx \pi$...

Some fundamental questions:

1. Does MYULA converge to a stationary distribution as $M \to \infty$?
2. Is this stationary distribution close to π in some sense?
3. Are there any accuracy guarantees for finite M?
4. How do these guarantees scale with M and with the dimension d?

Asymptotic results

1. Does MYULA converge to a stationary distribution as $M \to \infty$?
2. Is this stationary distribution close to π in some sense?

Assumption 2.1

Let $\Gamma_0(\mathbb{R}^d)$ be the class of lower semi-continous convex functions from $\mathbb{R}^d \to (-\infty, +\infty]$. Assume that $\pi(\mathbf{x}) \propto \exp\{-g_1(\mathbf{x}) - g_2(\mathbf{x})\}$, with $g_1, g_2 \in \Gamma_0(\mathbb{R}^d)$, and ∇g_1 Lipschitz continuous with constant L_1.

Theorem 2.1 (Durmus et al. (2016))

Suppose that Assumption 2.1 holds. Then, $\forall \mathbf{x}_0 \in \mathbb{R}^d$ and $\forall \delta < \delta_\lambda^{max}$, MYULA converges geometrically fast to an invariant measure $\tilde{\pi}_\lambda^\delta$ satisfying

$$\|\tilde{\pi}_\lambda^\delta - \pi_\lambda^*\|_{TV} = \mathcal{O}(\delta^{1/2}),$$

as $M \to \infty$.

Non-asymptotic results

3. Are there any accuracy guarantees for finite M?

Theorem 2.2 (Durmus et al. (2016))

Suppose that Assumption 1 holds. Then, there exist $\delta_\epsilon \in (0, \delta_\lambda^{max}]$ and $M_\epsilon \in \mathbb{N}$ such that $\forall \delta < \delta_\epsilon$ and $\forall M \geq M_\epsilon$

$$\|\delta_{\mathbf{x}_0} Q_\delta^M - \pi^*_{\delta/2}\|_{TV} < \epsilon,$$

where Q_δ^M is the kernel associated with M MYULA iterations with step δ.

If in addition g_2 is Lipchitz continuous with constant L_2, then

$$\|\delta_{\mathbf{x}_0} Q_\delta^M - \pi\|_{TV} < \epsilon + \frac{\delta}{2} L_2^2.$$

Scaling with dimension

4. How do these bounds scale with M and with the dimension d?

- Dependence of δ_ϵ and M_ϵ on dimension d and ϵ (Durmus et al., 2016):

	n	ε		n	ε
δ	$\mathcal{O}(d^{-5})$	$\mathcal{O}(\varepsilon^2/\log(\varepsilon^{-1}))$	δ	$\mathcal{O}(d^{-1})$	$\mathcal{O}(\varepsilon^2/\log(\varepsilon^{-1}))$
M	$\mathcal{O}(d^9)$	$\mathcal{O}(\varepsilon^{-2}\log^2(\varepsilon^{-1}))$	M	$\mathcal{O}(d)$	$\mathcal{O}(\varepsilon^{-2}\log^2(\varepsilon^{-1}))$

general bounds for Assumption 2.1 — bounds if the drift is strongly convex outside some ball

- The bound $\|\pi_\lambda^* - \pi\|_{TV} \leq \frac{\delta}{2} L_2^2$ is typically $\mathcal{O}(d)$.

Conclusion: MYULA delivers **reliable and computationally efficient approximations**, with good control of accuracy vs. computing-time.

Outline

Sparse image deblurring with an ℓ_1 prior

Recover a sparse high-resolution image $\mathbf{x} \in \mathbb{R}^n$ from a blurred and noisy observation

$$\mathbf{y} = H\mathbf{x} + \mathbf{w},$$

where H is a linear blur operator and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_d)$.

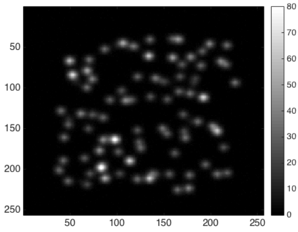
We use the Bayesian model

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\|\mathbf{y} - H\mathbf{x}\|^2/2\sigma^2 - \beta\|\mathbf{x}\|_1\right). \tag{6}$$

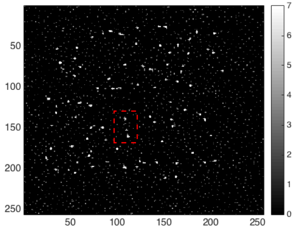
with $\beta = 0.01$.

Microscopy experiment

MAP estimation - live cell microscopy dataset (Zhu et al., 2012):



Microscopic image $\mathbf{y}$



$\hat{\mathbf{x}}MAP$ (log-scale)

Computing $\hat{\mathbf{x}}_{MAP}$ by convex optimisation Afonso et al. (2011) required 2.3 seconds.

Consider the 3-molecule structure in the highlighted region, how confident are we about this structure (its presence, position, etc.)?

Bayesian analysis - Posterior credible regions

Where does the posterior probability mass of $\mathbf{x}$ lie?

- A set C_α is a posterior credible region of confidence level $(1-\alpha)\%$ if

$$\mathrm{P}\left[\mathbf{x} \in C_\alpha | \mathbf{y}\right] = 1 - \alpha.$$

- The *highest posterior density* (HPD) region is decision-theoretically optimal (Robert, 2001)

$$C^*_\alpha = \{\mathbf{x} : g_1(\mathbf{x}) + g_2(\mathbf{x}) \leq \gamma_\alpha\}$$

  with $\gamma_\alpha \in \mathbb{R}$ chosen such that $\int_{C^*_\alpha} p(\mathbf{x}|\mathbf{y})\mathrm{d}\mathbf{x} = 1 - \alpha$ holds.

Bayesian analysis - Knockout hypothesis testing

"Knockout" test: double negation approach - assume that the structure is NOT present in the image and seek to REJECT the hypothesis.

Test procedure:

1. Generate a surrogate test image $\mathbf{x}_\dagger$ by modifying $\hat{\mathbf{x}}_{MAP}$ to remove the structure of interest.
2. If $\mathbf{x}_\dagger \notin \tilde{C}_\alpha$ the model rejects $\mathbf{x}_\dagger$ with probability $(1-\alpha)$, suggesting that the structure is present in the true image with high probability.
3. Otherwise, if $\mathbf{x}_\dagger \in \tilde{C}_\alpha$ the posterior uncertainty about the structure is too high to draw conclusions → increase measurements / reduce noise.

Bayesian computation

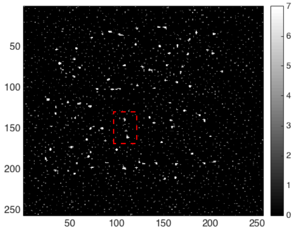
Estimation of C^*_α:

- We use MYULA to generate $n = 10^5$ samples $\{X_k^M\}_{k=1}^n$ and compute the HPD threshold γ_α by solving the quantile estimation problem

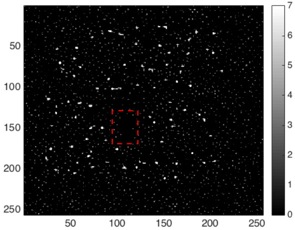
$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{(-\infty,\gamma_\alpha]}\left[g_1(X_k^M) + g_2(X_k^M)\right] = 1 - \alpha.$$

- We implement MYULA with:
  - $g_1(\mathbf{x}) = \|\mathbf{y} - H\mathbf{x}\|^2/2\sigma^2$
  - $g_2(\mathbf{x}) = \beta\|\mathbf{x}\|_1$.
  - $\text{prox}^\lambda_{g_2}(\mathbf{x})$ is the soft-thresholding operator with parameter $\beta\lambda$.
  - Algorithm parameters $\lambda = L_f^{-1} = 1.2$ and $\delta = \delta_\lambda^{max} = 0.6$.
- Computing time 4 minutes.

Knockout test:



$\hat{\mathbf{x}}MAP$ (log-scale)

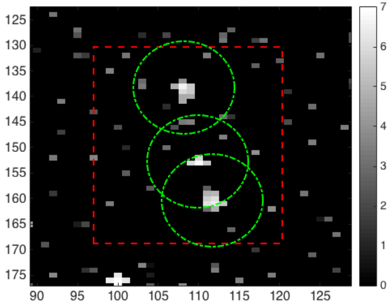


Test image $\mathbf{x}_\dagger$ (log-scale)

1. Score $g_1(\mathbf{x}_\dagger) + g_2(\mathbf{x}_\dagger) = 1.19 \times 10^5$.
2. The 99% threshold $\gamma_{0.01} = 9.69 \times 10^4$.
3. Therefore $\mathbf{x}_\dagger \notin \tilde{C}_\alpha$, rejecting the knockout hypothesis and providing evidence in favour of the structure considered.

Position uncertainty quantification

Find maximum molecule displacement within $\tilde{C}_\alpha$:

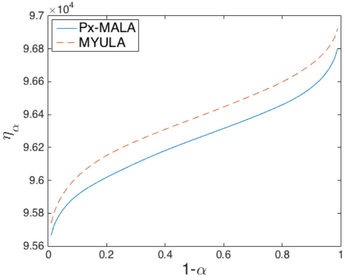


Mocule position uncertainty ($\pm 5 \times \pm 8$ pixels)

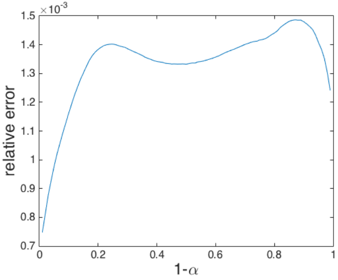
Note: Uncertainty analysis ($\pm 78nm \times \pm 125nm$) in close agreement with the experimental results (average precision $80nm$) of Zhu et al. (2012).

Microscopy experiment - Approximation error analysis

To assess the approximation error we benchmark estimations against proximal MALA (Px-MALA), which targets $p(\mathbf{x}|\mathbf{y})$ exactly (Pereyra, 2015). We use $n = 10^7$ iterations of Px-MALA (computing time 24 hours).



(a)



(b)

Figure : Microscopy experiment: (a) HDP region thresholds η_α for MYULA and Px-MALA, (b) relative approximation error of MYULA.

Outline

Conclusion

- The challenges facing modern image processing require a paradigm shift, and a new wave of analysis and computation methodologies.
- Great potential for synergy between Bayesian and variational approaches at algorithmic, methodological, and theoretical levels.
- MYULA delivers reliable and computationally efficient approximate inferences, with good control of accuracy vs. computing-time.

Thank you!

Fancy a Postdoc?

Bibliography I

Afonso, M., Bioucas-Dias, J., and Figueiredo, M. (2011). An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE. Trans. on Image Process.*, 20(3):681–695.

Combettes, P. L. and Pesquet, J.-C. (2011). Proximal splitting methods in signal processing. In Bauschke, H. H., Burachik, R. S., Combettes, P. L., Elser, V., Luke, D. R., and Wolkowicz, H., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer New York.

Durmus, A., Moulines, E., and Pereyra, M. (2016). Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau. *SIAM J. Imaging Sci.* in preparation.

Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. Roy. Stat. Soc. Ser. B*, 73(2):123–214.

Moreau, J.-J. (1962). Fonctions convexes duales et points proximaux dans un espace Hilbertien. *C. R. Acad. Sci. Paris Sér. A Math.*, 255:2897–2899.

Neal, R. (2012). MCMC using Hamiltonian dynamics. *ArXiv e-prints*.

Pereyra, M. (2015). Proximal Markov chain Monte Carlo algorithms. *Statistics and Computing*. open access paper, http://dx.doi.org/10.1007/s11222-015-9567-4.

Bibliography II

Robert, C. P. (2001). *The Bayesian Choice (second edition)*. Springer Verlag, New-York.

Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernulli*, 2(4):341–363.

Zhu, L., Zhang, W., Elnatan, D., and Huang, B. (2012). Faster STORM using compressed sensing. *Nat. Meth.*, 9(7):721–723.