Bayesian inference and mathematical imaging. Part II: Markov chain Monte Carlo.

Dr. Marcelo Pereyra
http://www.macs.hw.ac.uk/~mp71/

Maxwell Institute for Mathematical Sciences, Heriot-Watt University

Outline

1. Bayesian inference in imaging inverse problems
2. Proximal Markov chain Monte Carlo
3. Uncertainty quantification in astronomical and medical imaging
4. Image model selection and model calibration
5. Conclusion
We are interested in an unknown image $x \in \mathbb{R}^d$.  

We measure $y$, related to $x$ by a statistical model $p(y|x)$.  

The recovery of $x$ from $y$ is ill-posed or ill-conditioned, resulting in significant uncertainty about $x$.  

For example, in many imaging problems 

$$y = Ax + w,$$

for some operator $A$ that is rank-deficient, and additive noise $w$.  

---  

M. Pereyra (MI — HWU)  
Bayesian mathematical imaging
The Bayesian framework

- We use priors to reduce uncertainty and deliver accurate results.

- Given the prior $p(x)$, the posterior distribution of $x$ given $y$
  \[ p(x|y) = \frac{p(y|x)p(x)}{p(y)} \]
  models our knowledge about $x$ after observing $y$.

- In this talk we consider that $p(x|y)$ is log-concave; i.e.,
  \[ p(x|y) = \exp\{ -\phi(x) \}/Z, \]
  where $\phi(x)$ is a convex function and $Z = \int \exp\{ -\phi(x) \}dx$. 

M. Pereyra (MI — HWU)  
Bayesian mathematical imaging  
3 / 44
Maximum-a-posteriori (MAP) estimation

The predominant Bayesian approach in imaging is MAP estimation

\[ \hat{x}_{\text{MAP}} = \arg\max_{x \in \mathbb{R}^d} p(x | y), \]
\[ = \arg\min_{x \in \mathbb{R}^d} \phi(x), \quad (1) \]

computed efficiently, even in very high dimensions, by (proximal) convex optimisation (Chambolle and Pock, 2016).
Illustrative example: astronomical image reconstruction

Recover $x \in \mathbb{R}^d$ from low-dimensional degraded observation

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

where $\mathcal{F}$ is the continuous Fourier transform, $M \in \mathbb{C}^{m \times d}$ is a measurement operator and $w$ is Gaussian noise. We use the model

$$p(x|y) \propto \exp \left(-\|y - M\mathcal{F}x\|^2/2\sigma^2 - \theta \|\psi x\|_1\right)1_{\mathbb{R}_+^d}(x).$$

(2)

Figure: Radio-interferometric image reconstruction of the W28 supernova.
MAP estimation by proximal optimisation

To compute $\hat{x}_{MAP}$ we use a proximal splitting algorithm. Let

$$f(x) = \|y - MFX\|^2/2\sigma^2,$$
and

$$g(x) = \theta \|\Psi x\|_1 - \log 1_{\mathbb{R}_+^n}(x),$$

where $f$ and $g$ are l.s.c. convex on $\mathbb{R}^d$, and $f$ is $L_f$-Lipschitz differentiable.

For example, we could use a \textbf{proximal gradient} iteration

$$x^{m+1} = prox_{g}^{L_f^{-1}} \{x^m + L_f^{-1} \nabla f(x^m)\},$$

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

\textbf{Definition} For $\lambda > 0$, the $\lambda$-proximal operator of a convex l.s.c. function $g$ is defined as (Moreau, 1962)

$$\text{prox}_g^\lambda(x) \triangleq \arg\min_{u \in \mathbb{R}^N} g(u) + \frac{1}{2\lambda} \|u - x\|^2.$$
MAP estimation by proximal optimisation

The **alternating direction method of multipliers (ADMM)** algorithm

\[
\begin{align*}
x^{m+1} &= \text{prox}_{\lambda f}\{z^m - u^m\}, \\
z^{m+1} &= \text{prox}_{\lambda g}\{x^{m+1} + u^m\}, \\
u^{m+1} &= u^m + x^{m+1} - z^{m+1},
\end{align*}
\]

also converges to \(\hat{x}_{\text{MAP}}\) very quickly, and does not require \(f\) to be smooth.

However, MAP estimation has some limitations, e.g.,

1. it provides little information about \(p(x|y)\),
2. it struggles with unknown/partially unknown models,
Outline

1. Bayesian inference in imaging inverse problems
2. Proximal Markov chain Monte Carlo
3. Uncertainty quantification in astronomical and medical imaging
4. Image model selection and model calibration
5. Conclusion
Monte Carlo integration
Given a set of samples $X_1,\ldots,X_M$ distributed according to $p(x|y)$, we approximate posterior expectations and probabilities

$$\frac{1}{M} \sum_{m=1}^{M} h(X_m) \rightarrow \mathbb{E}\{h(x)|y\}, \quad \text{as } M \rightarrow \infty$$

Markov chain Monte Carlo:
Construct a Markov kernel $X_{m+1}|X_m \sim K(\cdot|X_m)$ such that the Markov chain $X_1,\ldots,X_M$ has $p(x|y)$ as stationary distribution.

MCMC simulation in high-dimensional spaces is very challenging.
Suppose for now that $p(x|y) \in C^1$. Then, we can generate samples by mimicking a Langevin diffusion process that converges to $p(x|y)$ as $t \to \infty$,

$$
X : \quad \text{d}X_t = \frac{1}{2} \nabla \log p(X_t|y) \, \text{d}t + \text{d}W_t, \quad 0 \leq t \leq T, \quad X(0) = x_0.
$$

where $W$ is the $n$-dimensional Brownian motion.

Because solving $X_t$ exactly is generally not possible, we use an Euler Maruyama approximation and obtain the “unadjusted Langevin algorithm”

$$
\text{ULA} : \quad X_{m+1} = X_m + \delta \nabla \log p(X_m|y) + \sqrt{2\delta} Z_{m+1}, \quad Z_{m+1} \sim \mathcal{N}(0, \mathbb{I}_n)
$$

ULA is remarkably efficient when $p(x|y)$ is sufficiently regular.
Non-smooth models

However, imaging models are often not smooth. Suppose that

$$p(x|y) \propto \exp \{-f(x) - g(x)\}$$  \hspace{1cm} (3)

where \(f(x)\) and \(g(x)\) are l.s.c. convex functions from \(\mathbb{R}^d \rightarrow (-\infty, +\infty]\), \(f\) is \(L_f\)-Lipschitz differentiable, and \(g \notin C^1\).

For example,

$$f(x) = \frac{1}{2\sigma^2} \|y - Ax\|_2^2, \quad g(x) = \alpha \|Bx\|_\dagger + 1_S(x),$$

for some linear operators \(A, B\), norm \(\| \cdot \|_\dagger\), and convex set \(S\).

Unfortunately, such non-models are beyond the scope of ULA.

**Idea:** Regularise \(p(x|y)\) to enable efficiently Langevin sampling.
Approximation of \( p(x|y) \)

**Moreau-Yoshida approximation of \( p(x|y) \) (Pereyra, 2015):**

Let \( \lambda > 0 \). We propose to approximate \( p(x|y) \) with the density

\[
p_{\lambda}(x|y) = \frac{\exp[-f(x) - g_{\lambda}(x)]}{\int_{\mathbb{R}^d} \exp[-f(x) - g_{\lambda}(x)] dx},
\]

where \( g_{\lambda} \) is the Moreau-Yoshida envelope of \( g \) given by

\[
g_{\lambda}(x) = \inf_{u \in \mathbb{R}^d} \{ g(u) + (2\lambda)^{-1} \| u - x \|^2_2 \},
\]

and where \( \lambda \) controls the approximation error involved.
Moreau-Yoshida approximations

Key properties (Pereyra, 2015; Durmus et al., 2018):

1. \( \forall \lambda > 0, \ p_\lambda \) defines a proper density of a probability measure on \( \mathbb{R}^d \).

2. Convexity and differentiability:
   - \( p_\lambda \) is log-concave on \( \mathbb{R}^d \).
   - \( p_\lambda \in C^1 \) even if \( p \) not differentiable, with
     \[
     \nabla \log p_\lambda(x|y) = -\nabla f(x) + \{\text{prox}_{\lambda g}(x) - x\}/\lambda,
     \]
     and \( \text{prox}_{\lambda g}(x) = \arg\min_{u\in\mathbb{R}^N} g(u) + \frac{1}{2\lambda} ||u - x||^2 \).
   - \( \nabla \log p_\lambda \) is Lipchitz continuous with constant \( L \leq L_f + \lambda^{-1} \).

3. Approximation error between \( p_\lambda(x|y) \) and \( p(x|y) \):
   - \( \lim_{\lambda \to 0} \|p_\lambda - p\|_{TV} = 0 \).
   - If \( g \) is \( L_g \)-Lipchitz, then \( \|p_\lambda - p\|_{TV} \leq \lambda L_g^2 \).
Examples of Moreau-Yoshida approximations:

\[ p(x) \propto \exp(-|x|) \]
\[ p(x) \propto \exp(-x^4) \]
\[ p(x) \propto 1_{[-0.5,0.5]}(x) \]

**Figure:** True densities (solid blue) and approximations (dashed red).
Proximal ULA

We approximate $\mathbf{X}$ with the “regularised” auxiliary Langevin diffusion

$$
\mathbf{X}^\lambda : \quad d\mathbf{X}^\lambda_t = \frac{1}{2} \nabla \log p_\lambda (\mathbf{X}^\lambda_t | y) \, dt + dW_t, \quad 0 \leq t \leq T, \quad \mathbf{X}^\lambda(0) = x_0,
$$

which targets $p_\lambda (x | y)$. Remark: we can make $\mathbf{X}^\lambda$ arbitrarily close to $\mathbf{X}$.

Finally, an Euler Maruyama discretisation of $\mathbf{X}^\lambda$ leads to the (Moreau-Yoshida regularised) proximal ULA

$$
\text{MYULA} : \quad X_{m+1} = (1 - \frac{\delta}{\lambda})X_m - \delta \nabla f\{X_m\} + \frac{\delta}{\lambda} \text{prox}_g^\lambda \{X_m\} + \sqrt{2\delta} Z_{m+1},
$$

where we used that $\nabla g_\lambda(x) = \{x - \text{prox}_g^\lambda(x)\}/\lambda$. 

Convergence results

Non-asymptotic estimation error bound

**Theorem 2.1 (Durmus et al. (2018))**

Let $\delta_{\lambda}^{\text{max}} = (L_1 + 1/\lambda)^{-1}$. Assume that $g$ is Lipchitz continuous. Then, there exist $\delta_\epsilon \in (0, \delta_{\lambda}^{\text{max}}]$ and $M_\epsilon \in \mathbb{N}$ such that $\forall \delta < \delta_\epsilon$ and $\forall M \geq M_\epsilon$,

$$\|\delta_{x_0} Q_\delta^M - p\|_{TV} < \epsilon + \lambda L_g^2,$$

where $Q_\delta^M$ is the kernel assoc. with $M$ iterations of MYULA with step $\delta$.

Note: $\delta_\epsilon$ and $M_\epsilon$ are explicit and tractable. If $f + g$ is strongly convex outside some ball, then $M_\epsilon$ scales with order $O(d \log(d))$. See Durmus et al. (2018) for other convergence results.
**Illustrative examples:**

- $p(x) \propto \exp(-|x|)$
- $p(x) \propto \exp(-x^4)$
- $p(x) \propto 1_{[-0.5,0.5]}(x)$

**Figure:** True densities (blue) and MC approximations (red histogram).
Recent surveys on Bayesian computation...

25th anniversary special issue on Bayesian computation


Special issue on “Stochastic simulation and optimisation in signal processing”

1. Bayesian inference in imaging inverse problems
2. Proximal Markov chain Monte Carlo
3. Uncertainty quantification in astronomical and medical imaging
4. Image model selection and model calibration
5. Conclusion
Where does the posterior probability mass of $x$ lie?

- A set $C_\alpha$ is a posterior credible region of confidence level $(1 - \alpha)\%$ if
  \[ P \left[ x \in C_\alpha | y \right] = 1 - \alpha. \]

- The highest posterior density (HPD) region is decision-theoretically optimal (Robert, 2001)
  \[ C_{\alpha}^* = \{ x : \phi(x) \leq \gamma_\alpha \} \]
  with $\gamma_\alpha \in \mathbb{R}$ chosen such that $\int_{C_{\alpha}^*} p(x|y)dx = 1 - \alpha$ holds.
Visualising uncertainty in radio-interferometric imaging

Astro-imaging experiment with redundant wavelet frame (Cai et al., 2017).

\[ \hat{x}_{\text{penMLE}}(y) \]

\[ \hat{x}_{\text{MAP}} \text{(by optimisation)} \]

credible intervals (scale $10 \times 10$)

3C2888 and M31 radio galaxies (size $256 \times 256$ pixels). Estimation error w.r.t. MH implementation 3%.
Hypothesis testing

Bayesian hypothesis test for specific image structures (e.g., lesions)

\(H_0: \) The structure of interest is ABSENT in the true image
\(H_1: \) The structure of interest is PRESENT in the true image

The null hypothesis \(H_0\) is rejected with significance \(\alpha\) if

\[ P(H_0 | y) \leq \alpha. \]

**Theorem** (Repetti et al., 2018)

Let \(S\) denote the region of \(\mathbb{R}^d\) associated with \(H_0\), containing all images *without the structure* of interest. Then

\[ S \cap C_\alpha^* = \emptyset \implies P(H_0 | y) \leq \alpha. \]

If in addition \(S\) is convex, then checking \(S \cap C_\alpha^* = \emptyset\) is a convex problem

\[ \min_{\bar{x}, x \in \mathbb{R}^d} \| \bar{x} - x \|^2_2 \text{ s.t. } \bar{x} \in C_\alpha^*, \quad x \in S. \]
Uncertainty quantification in MRI imaging

MRI experiment: test images $\bar{x} = x$, hence we fail to reject $H_0$ and conclude that there is little evidence to support the observed structure.
Uncertainty quantification in MRI imaging

MRI experiment: test images $\tilde{x} \neq x$, hence we reject $H_0$ and conclude that there is significant evidence in favour of the observed structure.
Uncertainty quantification in radio-interferometric imaging

Quantification of minimum energy of different energy structures, at level $(1 - \alpha) = 0.99$, as the number of measurements $T = \dim(y)/2$ increases.

$\hat{x}_{\text{MAP}}(T = 200)$

$\rho_\alpha$, energy ratio preserved at $\alpha = 0.01$

Note: energy ratio calculated as

$$\rho_\alpha = \frac{\|\bar{x} - \bar{x}\|_2}{\|x_{\text{MAP}} - \tilde{x}_{\text{MAP}}\|_2}$$

where $\bar{x}, \bar{x}$ are computed with $\alpha = 0.01$, and $\tilde{x}_{\text{MAP}}$ is a modified version of $x_{\text{MAP}}$ where the structure of interest has been carefully removed from the image.

Figure: Analysis of 3 structures in the W28 supernova RI image.
Outline

1. Bayesian inference in imaging inverse problems
2. Proximal Markov chain Monte Carlo
3. Uncertainty quantification in astronomical and medical imaging
4. Image model selection and model calibration
5. Conclusion
The Bayesian framework provides theory for comparing models objectively.

Given $K$ alternative models $\{M_j\}_{j=1}^K$ with posterior densities

$$M_j : \quad p_j(x|y) = p_j(y|x)p_j(x)/p_j(y),$$

we compute the (marginal) posterior probability of each model, i.e.,

$$p(M_j|y) \propto p(y|M_j)p(M_j) \quad (4)$$

where $p(y|M_j) = p_j(y) = \int p_j(y|x)p_j(x)dx$ measures model-fit-to-data.

We then select for our inferences the “best” model, i.e.,

$$M^* = \arg\max_{j \in \{1,\ldots,K\}} p(M_j|y).$$
Experiment setup

We degrade the Boat image of size $256 \times 256$ pixels with a $5 \times 5$ uniform blur operator $A^*$ and Gaussian noise $w \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_N)$ with $\sigma = 0.5$.

$$y = A^* x + w$$

We consider four alternative models to estimate $x$, given by

$$\mathcal{M}_j : \quad p_j(x|y) \propto \exp \left[ -\left( \|y - A_j x\|^2 / 2\sigma^2 \right) - \beta_j \phi_j(x) \right]$$

(5)

with fixed hyper-parameters $\sigma$ and $\beta$, and where:

- $\mathcal{M}_1$: $A_1$ is the correct blur operator and $\phi_j(x) = TV(x)$.
- $\mathcal{M}_2$: $A_2$ is a mildly misspecified blur operator and $\phi_j(x) = TV(x)$.
- $\mathcal{M}_3$: $A_3$ is the correct blur operator and $\phi_j(x) = \|\Psi x\|_1$.
- $\mathcal{M}_4$: $A_4$ is a mildly misspecified blur operator and $\phi_j(x) = \|\Psi x\|_1$.

where $\Psi$ is a wavelet frame and $TV(x) = \|\nabla_d x\|_{1-2}$ is the total-variation pseudo-norm. The $\beta_j$ are adjusted automatically (see model calibration).
To perform model selection we use MYULA to approximate the posterior probabilities $p(M_j|y)$ for $j = 1, 2, 3, 4$ by Monte Carlo integration.

For each model we generate $n = 10^5$ samples $\{X^j_k\}_{k=1}^n \sim p(x|y, M_j)$ and use the truncated harmonic mean estimator

$$p(y|M_j) \approx \left( \sum_{k=1}^n \frac{1_{S^*}(X^M_k)}{p(X^M_k, y|M_j)} \right)^{-1} \text{vol}(S^*), \quad j = \{1, 2, 3, 4\} \quad (6)$$

where $S^*$ is a union of highest posterior density sets of $p(x|y, M_j)$, also estimated from $\{X^j_k\}_{k=1}^n$.

Computing time approx. 30 minutes per model.
Numerical results

We obtain that $p(M_1|y) \approx 0.68$ and $p(M_3|y) \approx 0.27$ with the correct blur are the best models, $p(M_2|y) < 0.05$ and $p(M_4|y) < 0.01$ perform poorly.

Figure: MAP estimation results for the Boat image deblurring experiment. (Note: error w.r.t. “exact” probabilities from Px-MALA approx. 0.5%).
Numerical results

MYULA and Px-MALA efficiency comparison:

Figure: (a) Convergence of the chains to the typical set of $x | y$ under model $\mathcal{M}_1$, (b) chain autocorrelation function (ACF).
Empirical Bayesian model calibration

For illustration, consider the class of Bayesian models

\[ p(x|y, \theta) = \frac{p(y|x)p(x|\theta)}{p(y|\theta)}, \]

parametrised by a regularisation parameter \( \theta \in \Theta \). For example,

\[ p(x|\theta) = \frac{1}{C(\theta)} \exp\{-\theta \varphi(x)\}, \quad p(y|x) \propto \exp\{-f_y(x)\}, \]

with \( f_y \) and \( \varphi \) convex l.s.c. functions, and \( f_y \) \( L \)-Lipschitz differentiable.

We assume that \( p(x|\theta) \) is proper, i.e.,

\[ C(\theta) = \int_{\mathbb{R}^d} \exp\{-\theta \varphi(x)\} \, dx < \infty, \]

with \( C(\theta) \) unknown and generally intractable.
Maximum-a-posteriori estimation

If $\theta$ is fixed, the posterior $p(x|y, \theta)$ is log-concave and

$$
\hat{x}_{MAP} = \arg\min_{x \in \mathbb{R}^d} f_y(x) + \theta \varphi(x)
$$

is a convex optimisation problem that can be often solved efficiently.

For example, the proximal gradient algorithm

$$
x^{m+1} = \text{prox}_{\varphi}^{L^{-1}} \left\{ x^m + L^{-1} \nabla f_y(x^m) \right\},
$$

converges to $\hat{x}_{MAP}$ as $m \to \infty$.

However, when $\theta$ is unknown this significantly complicates the problem.
We adopt an empirical Bayes approach and calibrate the model maximising the evidence or marginal likelihood, i.e.,

\[ \hat{\theta} = \arg\max_{\theta \in \Theta} p(y|\theta), \]

\[ = \arg\max_{\theta \in \Theta} \int_{\mathbb{R}^d} p(y, x|\theta) dx, \]

which we solve efficiently by using a stochastic gradient algorithm driven by two proximal MCMC kernels (see Fernandez-Vidal and Pereyra (2018)).

Given \( \hat{\theta} \), we then straightforwardly compute

\[ \hat{x}_{MAP} = \arg\min_{x \in \mathbb{R}^d} f_y(x) + \hat{\theta} \varphi(x). \] (7)
Projected gradient algorithm

Assume that $\Theta$ is convex, and that $\hat{\theta}$ is the only root of $\nabla_\theta \log p(y|\theta)$ in $\Theta$. Then $\hat{\theta}$ is also the unique solution of the fixed-point equation

$$\theta = P_\Theta \left[ \theta + \delta \nabla_\theta \log p(y|\theta) \right].$$

where $P_\Theta$ is the projection operator on $\Theta$ and $\delta > 0$.

If $\nabla \log p(y|\theta)$ was tractable, we could compute $\hat{\theta}$ iteratively by using

$$\theta^{(t+1)} = P_\Theta \left[ \theta^{(t)} + \delta_t \nabla_\theta \log p(y|\theta^{(t)}) \right],$$

with sequence $\delta_t = \alpha t^{-\beta}$, $\alpha > 0$, $\beta \in [1/2, 1]$.

However, $\nabla \log p(y|\theta)$ is “doubly” intractable...
Stochastic projected gradient algorithm

To circumvent the intractability of $\nabla_{\theta} \log p(y|\theta)$ we use Fisher's identity

$$\nabla_{\theta} \log p(y|\theta) = \mathbb{E}_{x|y,\theta} \{ \nabla_{\theta} \log p(x, y|\theta) \},$$

$$= -\mathbb{E}_{x|y,\theta} \{ \varphi + \nabla_{\theta} \log C(\theta) \},$$

together with the identity

$$\nabla_{\theta} \log C(\theta) = -\mathbb{E}_{x|\theta} \{ \varphi(x) \},$$

to obtain $\nabla_{\theta} \log p(y|\theta) = \mathbb{E}_{x|\theta} \{ \varphi(x) \} - \mathbb{E}_{x|y,\theta} \{ \varphi(x) \}$.

This leads to the equivalent fixed-point equation

$$\theta = P_{\Theta} \left( \theta + \delta \mathbb{E}_{x|\theta} \{ \varphi(x) \} - \delta \mathbb{E}_{x|y,\theta} \{ \varphi(x) \} \right), \quad (8)$$

which we solve by using a stochastic approximation algorithm.
Stochastic Approximation algorithm to compute \( \hat{\theta} \)

We use the following MCMC-driven stochastic gradient algorithm:
Initialisation \( x^{(0)}, u^{(0)} \in \mathbb{R}^d, \theta^{(0)} \in \Theta, \delta_t = \delta_0 t^{-0.8}. \)

for \( t = 0 \) to \( n \)

1. MCMC update \( x^{(t+1)} \sim M_{x|y,\theta^{(t)}}(\cdot|x^{(t)}) \) targeting \( p(x|y, \theta^{(t)}) \)
2. MCMC update \( u^{(t+1)} \sim K_{x|\theta^{(t)}}(\cdot|u^{(t)}) \) targeting \( p(x|\theta^{(t)}) \)
3. Stoch. grad. update

\[
\theta^{(t+1)} = P_\Theta \left[ \theta^{(t)} + \delta_t \varphi(u^{(t+1)}) - \delta_t \varphi(x^{(t+1)}) \right].
\]

end for

Output The iterates \( \theta^{(t)} \to \hat{\theta} \) as \( n \to \infty. \)
SAPG algorithm driven MCMC kernels

Initialisation \( x^{(0)}, u^{(0)} \in \mathbb{R}^d, \theta^{(0)} \in \Theta, \delta_t = \delta_0 t^{-0.8}, \lambda = 1/L, \gamma = 1/4L. \)

\textbf{for} \( t = 0 \) to \( n \)

1. Coupled Proximal MCMC updates: generate \( z^{(t+1)} \sim \mathcal{N}(0, \mathbb{I}_d) \)

\[
x^{(t+1)} = (1 - \frac{\gamma}{\lambda}) x^{(t)} - \gamma \nabla f_y \left( x^{(t)} \right) + \frac{\gamma}{\lambda} \text{prox}_{\varphi}^{\theta \lambda} \left( x^{(t)} \right) + \sqrt{2\gamma} z^{(t+1)},
\]

\[
u^{(t+1)} = (1 - \frac{\gamma}{\lambda}) u^{(t)} + \frac{\gamma}{\lambda} \text{prox}_{\varphi}^{\theta \lambda} \left( u^{(t)} \right) + \sqrt{2\gamma} z^{(t+1)},
\]

2. Stochastic gradient update

\[
\theta^{(t+1)} = P_{\Theta} \left[ \theta^{(t)} + \delta_t \varphi(u^{(t+1)}) - \delta_t \varphi(x^{(t+1)}) \right].
\]

\textbf{end for}

\textbf{Output} Averaged estimator \( \bar{\theta} = n^{-1} \sum_{t=1}^{n} \theta^{(t+1)} \) converges approx. to \( \hat{\theta} \).
Illustrative example - Image deblurring with $\ell_1$ prior

We consider again the live-cell microscopy setup

$$p(x|y, \theta) \propto \exp \left( -\|y - Ax\|_2^2/2\sigma^2 - \theta \|x\|_1 \right),$$

and compute $\hat{\theta} = \arg\max_{\theta \in \mathbb{R}^+} p(y|\theta)$.

Figure: Molecules image deconvolution experiment, computing time 0.75 secs.
Illustrative example - Image deblurring with TV-$\ell_2$ prior

Similarly, for the Bayesian image deblurring model

$$p(x|y, \theta) \propto \exp\left(-\frac{1}{2}\|y - Ax\|^2/\sigma^2 - \alpha\|x\|_2 - \theta \|\nabla_d x\|_{1-2}\right),$$

we compute $\hat{\theta} = \text{argmax}_{\theta \in \mathbb{R}^+} p(y|\theta)$.

Figure: Boat image deconvolution experiment.
Image deblurring with TV-$\ell_2$ prior

Comparison with the (non-Bayesian) SUGAR method (Deledalle et al., 2014), and an oracle that knows the optimal value of $\theta$. Average values over 6 test images of size 512 $\times$ 512 pixels.

<table>
<thead>
<tr>
<th>Method</th>
<th>SNR=20 dB</th>
<th></th>
<th>SNR=30 dB</th>
<th></th>
<th>SNR=40 dB</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg. MSE</td>
<td>Avg. Time</td>
<td>Avg. MSE</td>
<td>Avg. Time</td>
<td>Avg. MSE</td>
<td>Avg. Time</td>
</tr>
<tr>
<td>$\theta^*$ (Oracle)</td>
<td>22.95 ± 3.10</td>
<td>-</td>
<td>21.05 ± 3.19</td>
<td>-</td>
<td>18.76 ± 3.19</td>
<td>-</td>
</tr>
<tr>
<td>Empirical Bayes</td>
<td>23.24 ± 3.23</td>
<td>43.01</td>
<td>21.16 ± 3.24</td>
<td>41.50</td>
<td>18.90 ± 3.39</td>
<td>42.85</td>
</tr>
<tr>
<td>SUGAR</td>
<td>24.14 ± 3.19</td>
<td>15.74</td>
<td>23.96 ± 3.26</td>
<td>20.87</td>
<td>23.94 ± 3.27</td>
<td>20.59</td>
</tr>
</tbody>
</table>
1. Bayesian inference in imaging inverse problems
2. Proximal Markov chain Monte Carlo
3. Uncertainty quantification in astronomical and medical imaging
4. Image model selection and model calibration
5. Conclusion
The challenges facing modern imaging sciences require a methodological paradigm shift to go beyond point estimation.

The Bayesian framework can support this paradigm shift, but this requires significantly accelerating computation methods.

We explored improving efficiency by integrating modern stochastic and variational approaches.

Thank you!
Bibliography:


