# High dimensional adaptive Monte Carlo algorithm

Arthur VILHELM[1] – Marcelo PEREYRA[2]
(1) *Ecole Normale Supérieure de Cachan, Fr* – (2) *University of Bristol, UK*

## Abstract

• We are interested in the recovering of unobserved or **partially observed image from noisy data**. This problem is a cornerstone in imaging science and some canonical examples are : *image denoising, deconvolution, super-resolution, inpainting or tomographic reconstruction, etc.*

• We use **Markov Chain Monte Carlo methods** to compute an estimator of the original image and a statistical summary of the posterior distribution. We develop **an adaptive framework** to increase performance and autonomy.

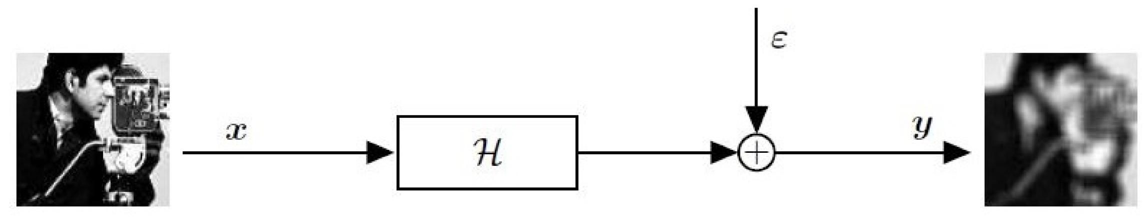
## Introduction

• Markov Chain Monte Carlo (MCMC) methods are well known as a key tool for statistical data analysis, from signal processing to biology or economics.

• In image processing, the objects we deal with are likely **high-dimensional** and classical samplers may fail to **scale up**.

• **Langevin-based Monte Carlo** samplers (LMC) are very effective in high-dimensional setting because Langevin dynamics consist in the trade-off between **random walk** and **gradient mapping**.

## Model




$$\mathbf{y} = H\mathbf{x} + \mathbf{n}$$

$\mathbf{y}$ is the noisy observation, $\mathbf{x}$ is the image to recover, $\mathbf{n}$ is the noise and $H$ is a linear operator.

• The **posterior distribution** is given by Bayes formula $\pi(\mathbf{x}) = p(\mathbf{y}|\mathbf{x})p(\mathbf{x})/p(\mathbf{y})$ to add prior information to the likelihood

$$\pi(\mathbf{x}) = \exp\{-g(\mathbf{x})\}/\kappa_0$$

• $g(\mathbf{x}) = l(\mathbf{x}) + h(\mathbf{x})$ is the convex potential :

(i) $\lim_{||x||\to\infty} g(\mathbf{x}) = +\infty$

(ii) $l : \mathbb{R}^n \to \mathbb{R}$ is **convex**, continuously differentiable and gradient Lipschitz with Lipschitz constant $L_l$

(iii) $h : \mathbb{R}^n \to (-\infty, +\infty]$ is strictly convex, **lower semi-continuous** (l.s.c) and Lipschitz ($\Gamma_0$)

• **Moreau approximation** :

$$\pi^\lambda(\mathbf{x}) = \sup_{\mathbf{u}\in\mathbb{R}^n} e^{-\frac{||\mathbf{u}-\mathbf{x}||_2^2}{2\lambda} - h(\mathbf{u})} e^{-l(\mathbf{x})}/\kappa_\lambda$$

(sub)–gradient mapping (a.k.a proximal mapping) :

$$\text{prox}_h^\lambda(\mathbf{x}) = \underset{\mathbf{x}\in\mathbb{R}^n}{\text{argmin}} \left\{ f(\mathbf{x}) + \frac{||\mathbf{u}-\mathbf{x}||^2}{2\lambda} \right\}$$

## Fully adaptive Proximal Langevin algorithm

**Metropolis-Hastings scheme** [1] [3]

– Drift :

$$\mathring{D}^{(k)} = -\nabla\mathring{l}(\mathbf{x}^{(k)}) + \frac{1}{\lambda}\left(U\text{prox}_h^\lambda\left(U^T\mathring{\mathbf{x}}^{(k)}\right) - \mathring{\mathbf{x}}^{(k)}\right)$$

– Sample a perturbation : $\mathring{W}^{(k)} \sim \mathcal{N}(0, \Gamma^{(k)})$

– Compute a candidate :

$$\mathring{\mathbf{x}}^* = \mathring{\mathbf{x}}^{(k)} + \frac{\delta^{(k)}}{2}\Gamma^{(k)}D(\mathring{\mathbf{x}}^{(k)}) + \sqrt{\delta}\mathring{W}^{(k)}$$

– Compute the acceptance probability :

$$\alpha_\lambda^{(k)} = \min\left\{1; \frac{\pi^\lambda(\mathbf{x}^*)q(\mathbf{x}^{(k)}|\mathbf{x}^*)}{\pi^\lambda(\mathbf{x}^{(k)})q(\mathbf{x}^*|\mathbf{x}^{(k)})}\right\}$$

– Set $\mathbf{x}^{(k+1)} = \mathbf{x}^*$ with probability $\alpha_\lambda^{(k)}$

**Update parameters** [2]

– Adaptive step-size :

$$\delta^{(k)} = P_\delta\left\{\delta^{(k-1)}\left(1 + \iota^{(k)}(\alpha_\lambda^{(k)} - \alpha_{opt})\right)\right\}$$

– Update mean and covariance estimates :

$$\mathring{\mu}^{(k)} = \mathring{\mu}^{(k-1)} + \iota^{(k)}\left(\mathring{\mathbf{x}}^{(k)} - \mathring{\mu}^{(k-1)}\right)$$

$$\Gamma_{0\,i,i}^{(k)} = \Gamma_{0\,i,i}^{(k-1)} + \iota^{(k)}\left(\left(\mathring{\mathbf{x}}_i^{(k)} - \mathring{\mu}_i^{(k)}\right)^2 - \Gamma_{0\,i,i}^{(k-1)}\right)$$

**Stopping criterion**

– Effective Sample Size [4]

## Theoretical and experimental results

### Convergence guarantees

The fully adaptive Proximal Langevin algorithm is a special case of Metropolis Adjusted Langevin Algorithm (MALA) which target $\pi^\lambda$ a $\lambda$–approximamation of $\pi$. It is guaranteed to converge ergodically if :

**Condition 1.** The density $\pi^\lambda$ is positive with continuous first derivative such that

$$\lim_{||\mathbf{x}||\to\infty} -\frac{\mathbf{x}}{||\mathbf{x}||}\nabla g(\mathbf{x}) = -\infty \quad \text{(Coercivity)}$$

$$\limsup_{||\mathbf{x}||\to\infty} -\frac{\mathbf{x}}{||\mathbf{x}||}\frac{\nabla g(\mathbf{x})}{||\nabla g(\mathbf{x})||} < 0 \quad \text{(Curvature)}$$

If $\pi$ has a polynomial form $\mathcal{P}_\beta$ such as $g(\mathbf{x}) = \gamma||\mathbf{x}||^\beta$ :

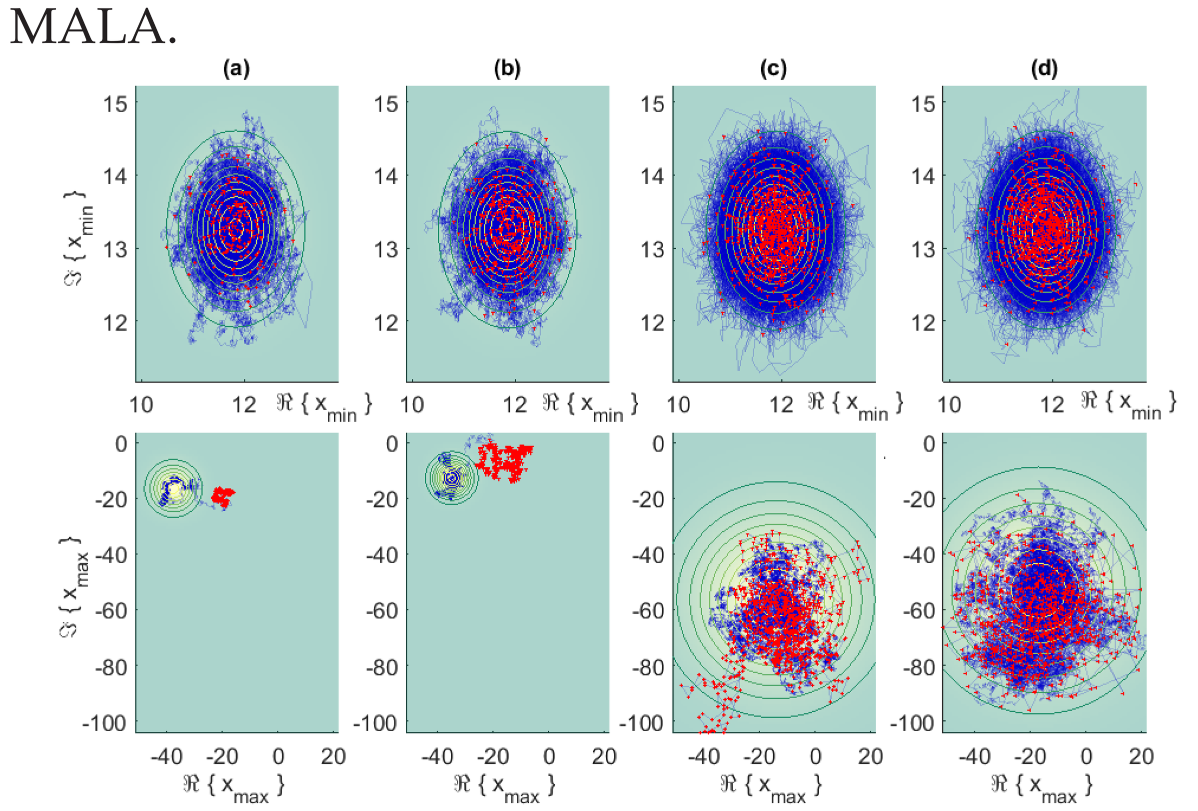
**Proposition 1.** Condition (1) is satisfied if one of the following hold :

1. $\exp(-l(\mathbf{x})) \in \mathcal{P}_{\beta_1}$ and $\exp(-h(\mathbf{x})) \in \mathcal{P}_{\beta_2}$, such as $\beta_2 \geq 2 \geq \beta_1 > 0$ or $\max(\beta_1, \beta_2) \geq 2$ is even
2. $\pi$ is strongly log-concave

It is also possible to target $\pi$ directly by substituting $\pi$ to $\pi^\lambda$ in the acceptance probability. The algorithm converges if $\pi \in \mathcal{C}^2$ and we conjecture that it also converges ergodically if prop 1. is satisfied.

### Experimental mixing improvement

Comparison of the improvement in the autocorrelation (after $10^6$ iterations) : (a) RWM, (b) MYMALA, (c) adaptive step-size (only) MYMALA, (d) fully adaptive MYMALA.



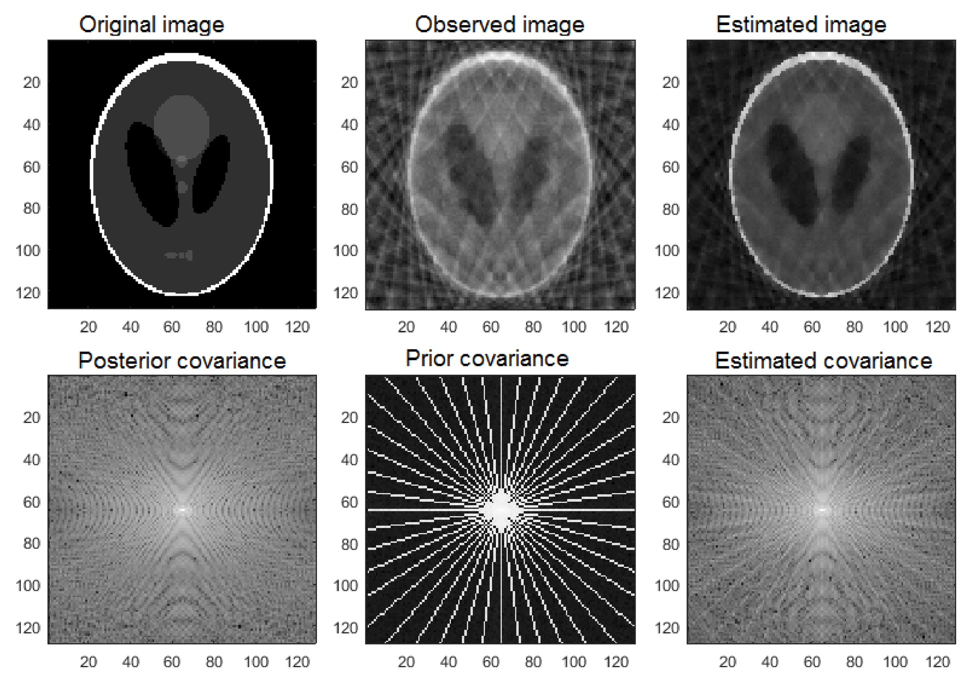

*Representation of the distribution smallest and the biggest eigenvectors of a 16 384–dimensional target*

The **Effective Sample Size** $N^{\text{eff}}$ is the equivalent number of independent sample, and the ratio $N^{\text{eff}}/N^{\text{tot}}$ is, for each method : $1.97 \times 10^{-5}$, $2.32 \times 10^{-5}$, $5.46 \times 10^{-5}$ and for the proposed algorithm $2.32 \times 10^{-4}$.

## Conclusion

• We have proposed here a fully adaptive Langevin scheme which **drastically improves the performances** reached by more classical algorithms such as Random Walk Metropolis, MALA or proximal-MALA, in terms of sample correlation and computing time. This let **scale up** Monte Carlo methods to image processing problems to perform a descriptive statistical summary.

• Application for **medical image processing**




*MRI deconvolution and denoising*

• Moreau approximation and proximal mapping broaden the class of density to **non-smooth targets** with convergences and **ergodicity guaranties**.

• These targets are more and more present due to regularisation methods, such as in the **total variation** (TV) prior, where the likelihood is Gaussian and the prior on $\mathbf{x}$ is of the form $g(\mathbf{x}) = ||\nabla\mathbf{x}||_1$, which is **piecewise differentiable**.

## Future Research

• Further the canonical application of deconvolution, we have been investigating a few applications such as **target detection** in MRI (of the round gray areas) with confidence interval tests in highest posterior density regions.

• It is also of interest to perform **hypothesis testing** and **prior selection**, that we have experimented through Bayes Factor and **harmonic mean estimators** (HME). HME is not a robust estimator and the convergence and approximation bias are still open questions.

• **Convergence** of a Monte Carlo algorithm is never an easy question, and we could not give sufficient proofs to ensure the exact and adaptive proximal-MALA to be ergodic. Even though, [5] gives **strong intuition** that it would converge to the exact posterior and we already have encouraging results on **Langevin diffusion convergence**.

## References

[1] C. Robert and G. Casella. *Monte Carlo statistical methods*. 2013.
[2] Y. F. Atchade. An adaptive version for the metropolis adjusted langevin algorithm. 2006.
[3] A. Durmus, E. Moulines, and M. Pereyra. Sampling from convex non continuously differentiable functions. 2016.
[4] L. Gong and J. M. Flegal. A practical sequential stopping rule for high-dimensional mcmc. 2014.
[5] G. O. Roberts and R. L. Tweedie. Exponential convergence of langevin distributions. 1996.

## Contact Information

**Email**[1] arthur.vilhelm@ens-cachan.fr
**Email**[2] marcelo.pereyra@bristol.ac.uk