High-dimensional MCMC: Sampling from log-concave distrutions.

Alain Durmus, Eric Moulines, Marcelo Pereyra

Telecom ParisTech, LTCI, Ecole Polytechnique, CMAP, University of Bristol, Stat. Lab.

July 8, 2016

Introduction

- Sampling distribution over high-dimensional state-space has recently attracted a lot of research efforts in computational statistics and machine learning...
- Applications (non-exhaustive)
  1. Bayesian inference for high-dimensional models
  2. Bayesian non parametrics
  3. Aggregation of estimators and experts
- Most of the sampling techniques known so far do not scale to high-dimension... Challenges are numerous in this area...

Logistic regression

- Likelihood: Binary regression set-up in which the binary observations (responses) $(Y_1, \ldots, Y_n)$ are conditionally independent Bernoulli random variables with success probability $F(\boldsymbol{\beta}^T X_i)$, where
  1. X_i is a d dimensional vector of known covariates,
  2. $\boldsymbol{\beta}$ is a d dimensional vector of unknown regression coefficient
  3. F is a distribution function.
- logistic regression: F is the standard logistic distribution function,

$$F(t) = \mathrm{e}^t / (1 + \mathrm{e}^t)$$

.

- Problem: the number of predictor variables d can be large.

Bayes 101

- Bayesian analysis requires a prior distribution for the unknown regression parameter

$$\pi(\boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}'\Sigma_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}\right) \quad \text{or} \quad \pi(\boldsymbol{\beta}) = \exp\left(-\sum_{i=1}^{d}\alpha_i|\beta_i|\right)$$

.

- The posterior of $\boldsymbol{\beta}$ is up to a proportionality constant given by

$$\pi(\boldsymbol{\beta}|(Y,X)) \propto \prod_{i=1}^{n} F^{Y_i}(\beta' X_i)(1-F(\beta' X_i))^{1-Y_i}\pi(\boldsymbol{\beta})$$

A daunting problem ?

- The posterior density distribution of β is given by Bayes' rule, up to a proportionality constant by

$$\pi(\beta|(Y,X)) \propto \exp(-U(\beta)) .$$

where the potential $U(\beta)$ is given by

$$U(\beta) = -\sum_{i=1}^{p} \{Y_i \log \frac{F(\beta^T X_i)}{1 - F(\beta^T X_i)} + \log(1 - F(\beta^T X_i))\} + \|B\beta\|^{1,2}$$

- Classical composite objective function... The prior plays the role of regularization penalty.

A daunting problem ?

- In the case of the ridge regression, the potential U is smooth strongly convex.
- In the case of the lasso regression, the potential U is non-smooth but still convex...
- A wealth of reasonably fast optimisation algorithms are available to solve this problem in high-dimension...

Framework

- Denote by π a target density w.r.t. the Lebesgue measure on $\mathbb{R}^d$, known up to a normalisation factor

$$x \mapsto \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y \ ,$$

  Implicitly, $d \gg 1$.

- Assumption: U is L-smooth : continuously differentiable and there exists a constant L such that for all $x, y \in \mathbb{R}^d$,

$$\|\nabla U(x) - \nabla U(y)\| \leq L\|x - y\| \ .$$

Langevin diffusion

- Langevin SDE:

$$\mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \ ,$$

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian Motion.

- Denote for all $x \in \mathbb{R}^d$, $\delta_x P_t$ the law Y_t started at x.
- $\pi \propto \mathrm{e}^{-U}$ is reversible $\rightsquigarrow$ the unique invariant probability measure.
- The convergence to the stationary distribution takes place at geometrical rate.
- Precise estimates of the convergence rate (TV, relative entropy) can be obtained using:
  - Functional inequalities: Poincaré or Log-Sobolev inequalities
  - Coupling techniques: synchronous or reflection coupling, depending upon the assumptions (Eberle, 2015)

Discretized Langevin diffusion

- Idea: Sample the diffusion paths, using for example the Euler-Maruyama (EM) scheme:

$$X_{k+1} = X_k - \gamma_{k+1} \nabla U(X_k) + \sqrt{2\gamma_{k+1}} Z_{k+1}$$

  where
  - $(Z_k)_{k \geq 1}$ is i.i.d. $\mathcal{N}(0, \mathrm{I}_d)$
  - $(\gamma_k)_{k \geq 1}$ is a sequence of stepsizes, which can either be held constant or be chosen to decrease to 0 at a certain rate.
- Euler discretization = gradient algorithm + noise.

Discretized Langevin diffusion: constant stepzize

- When $\gamma_k = \gamma$, then $(X_k)_{k \geq 1}$ is an homogeneous Markov chain with Markov kernel R_γ with density

$$r_\gamma(x, y) = (4\pi\gamma)^{-d/2} \exp\left(-(4\gamma)^{-1} \left\|y - x + \gamma \nabla U(x)\right\|^2\right) .$$

- Under some appropriate conditions (a bit of positive curvature at infinity), this Markov chain is irreducible, positive recurrent $\rightsquigarrow$ unique invariant distribution π_γ.
- Problem: $\pi_\gamma \neq \pi$.

The EM Markov chain

- When $(\gamma_k)_{k\geq 1}$ is nonincreasing and non constant, $(X_k)_{k\geq 1}$ is an inhomogeneous Markov chain associated with the sequence of Markov kernel $(R_{\gamma_k})_{k\geq 1}$.
- Denote by $\delta_x Q_\gamma^p$ the law of X_p stated at x.
- Reminder: the diffusion converges to the target distribution
- Question: since the EM disretization approximates the diffusion, can it be used to sample from π ?
  - Is $\delta_x Q_\gamma^p$ close to π for which p ?
  - Can we have some theoretical guarantees ? Particular attention to the dimension d, see also Dalalyan 2014.

Metric on probability spaces

Definition

For μ, ν two probabilities measure on $\mathbb{R}^d$, define

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{|f| \leq 1} |\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]| \ .$$

$$W_2^2(\mu, \nu) = \inf_{\xi \in \mathcal{C}(\mu,\nu)} \int_{\mathbb{R}^{2d}} \|x - y\|^2 \, \xi(\mathrm{d}x, \mathrm{d}y) \ .$$

Geometric convergence of the Langevin diffusion

- If there exists a Lyapunov function for the generator of the diffusion then there exists $\kappa \in [0, 1)$ such that for any initial distribution μ_0 and $t > 0$,

$$\|\mu_0 P_t - \pi\|_{\mathrm{TV}} \leq C(\mu_0)\kappa^t ,$$

  for some explicit function of the initial probability $C(\mu_0)$.
- Explicit expressions of the constant (the way dimension impacts theses constants) critically depends on
  - the assumptions on the potential U
  - the technique of proofs (functional inequalities, coupling constructions, etc...)

Foster-Lyapunov condition

- A function $V \in C^2(\mathbb{R}^d)$ is a Lyapunov function if $V \geq 1$ and if there exists $\theta > 0$, $b \geq 0$ such that,

$$\mathscr{A} V \leq -\theta V + b \, ,$$

where $\mathscr{A} f = - \langle \nabla U, \nabla f \rangle + \Delta f$ is the generator of the diffusion

- Example: If there exist $\alpha > 1$, $\rho > 0$ and $M_\rho \geq 0$ such that for all $y \in \mathbb{R}^d$, $\|y\| \geq M_\rho$:

$$\langle \nabla U(y), y \rangle \geq \rho \|y\|^\alpha \, .$$

then $V(x) = \exp(U(x)/2)$ is a Lyapunov function.

Geometric convergence of the Euler discretization

- Let $(\gamma_k)_{k\geq 1}$ be a sequence of positive and non-increasing step sizes
- Euler discretization:

$$X_{k+1} = X_k - \gamma_{k+1}\nabla U(X_k) + \sqrt{2\gamma_{k+1}} Z_{k+1} ,$$

where $(Z_k)_{k\geq 1}$ is i.i.d. $\mathcal{N}(0, \mathrm{I}_d)$, independent of X_0.

- Markov kernel R_γ and $x \in \mathbb{R}^d$ by

$$R_\gamma(x, A) = \int_A \frac{1}{(4\pi\gamma)^{d/2}} \exp\left(-(4\gamma)^{-1} \|y - x + \gamma\nabla U(x)\|^2\right) \mathrm{d}y .$$

- The sequence $(X_n)_{n\geq 0}$ is a (possibly) time-nonhomogeneous Markov chain whose distribution is specified by the Markov kernels $(R_{\gamma_n})_{n\geq 1}$.

Level-0 results

- The Markov kernel R_γ is strongly Feller, irreducible, and hence all the compact sets are therefore small.
- Typically, the R_γ satisfies a Foster-Lyapunov drift condition of a particular form, *i.e.* there exists $\kappa \in [0,1)$, $b > 0$ such that for all $\gamma > 0$

$$R_\gamma V \leq \kappa^\gamma V + \gamma b \; .$$

- R_γ admits a unique stationary distribution π_γ and is V-uniformly geometrically ergodic, in the sense that, for some constant $C < \infty$ and $\kappa \in [0,1)$, such that for all $x \in \mathbb{R}^d$,

$$\left\| R_\gamma^k(x,\cdot) - \pi_\gamma \right\|_V \leq C(\gamma) \kappa^{\gamma k} V(x) \; .$$

Example: A drift condition for R_γ

Theorem

Assume U is L-smooth and there exist $\rho > 0$, $\alpha > 1$ and $M_\rho \geq 0$ such that :

$$\langle \nabla U(y), y \rangle \geq \rho \|y\|^\alpha \;, \quad \text{for all } y \in \mathbb{R}^d,\ \|y\| \geq M_\rho$$

Then for all $\bar{\gamma} \in \left(0, L^{-1}\right)$, there exists $b \geq 0$ and $s > 0$ such that

$$R_\gamma V(x) \leq \kappa^\gamma V(x) + \gamma b \;, \quad \text{for all } \gamma \in (0, \bar{\gamma}] \text{ and } x \in \mathbb{R}^d,$$

where

$$V(x) = \exp(U(x)/2).$$

Control of moments

- By a straightforward induction, we get for all $n \geq 0$ and $x \in \mathbb{R}^d$,

$$Q_\gamma^n V \leq \kappa^{\Gamma_{1,n}} V + b \sum_{i=1}^{n} \gamma_i \kappa^{\Gamma_{i+1,n}} \ ,$$

  where $\Gamma_{n,p} = \sum_{k=n}^{p} \gamma_k$.
- Note that we have

$$\sup_{n \geq 1} \sum_{i=1}^{n} \gamma_i \kappa^{\Gamma_{i+1,n}} < \infty \ .$$

Wasserstein distance convergence

- We assume in this part that U is strongly convex: there exist and $m > 0$, such that for all $x, y \in \mathbb{R}^d$,

$$\langle \nabla U(x) - \nabla U(y), x - y \rangle \geq m \|x - y\|^2 \ .$$

Wasserstein distance convergence

Theorem

Assume that U is strongly convex. Then,

(i) *For all probability measures $x, y \in \mathbb{R}^d$ and $t \geq 0$,*

$$W_2(\delta_x P_t, \delta_y P_t) \leq \mathrm{e}^{-mt} \|x - y\|$$

(ii) *In addition, for any $x \in \mathbb{R}^d$,*

$$W_2(\delta_x P_t, \pi) \leq C(d)\mathrm{e}^{-mt} .$$

Elements of proof

$$\begin{cases} \mathrm{d}Y_t &= -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \ , \\ \mathrm{d}\tilde{Y}_t &= -\nabla U(\tilde{Y}_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \ , \end{cases}$$

where $(Y_0, \tilde{Y}_0) = (x, y)$. This SDE has a unique strong solution $(Y_t, \tilde{Y}_t)_{t \geq 0}$ associated to $(B_t)_{t \geq 0}$. Moreover

$$\left\| Y_t - \tilde{Y}_t \right\|^2 = \left\| Y_0 - \tilde{Y}_0 \right\|^2 - 2 \int_0^t \left\langle (\nabla U(Y_s) - \nabla U(\tilde{Y}_s)), Y_s - \tilde{Y}_s \right\rangle \mathrm{d}s \ ,$$

which implies using Grönwall's inequality that

$$\left\| Y_t - \tilde{Y}_t \right\|^2 \leq \left\| Y_0 - \tilde{Y}_0 \right\|^2 - 2m \int_0^t \left\| Y_s - \tilde{Y}_s \right\|^2 \mathrm{d}s \leq \left\| Y_0 - \tilde{Y}_0 \right\|^2 \mathrm{e}^{-2mt} \ .$$

The proof follows since for all $t \geq 0$, the law of $(Y_t, \tilde{Y}_t)$ is a coupling between $\delta_x P_t$ and $\delta_y P_t$.

Coupling between the Langevin and the EM discretization

Consider the synchronuous coupling between $(Y_t)_{t\geq 0}$ and $(X_k)_{k\geq 0}$ defined by

$$\begin{cases} \mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \\ X_{k+1} = X_k - \gamma_{k+1}\nabla U(X_k) + \sqrt{2\gamma_{k+1}}(B_{\Gamma_{k+1}} - B_{\Gamma_k}) \ , \end{cases}$$

started at $(y, x) \in \mathbb{R}^d \times \mathbb{R}^d$.

Bound in Wasserstein distance of order 2

- We make an additional regularity assumption on U: The potential U is three times continuously differentiable and there exists $\tilde{L}$ such that for all $x, y \in \mathbb{R}^d$:

$$\left\|\nabla^2 U(x) - \nabla^2 U(y)\right\| \leq \tilde{L} \left\|x - y\right\| \ .$$

- Then for all $x \in \mathbb{R}^d$ and $n \geq 1$,

$$W_2^2(\delta_x Q_\gamma^n, \pi) \leq u_n^{(1)}(\gamma) \int_{\mathbb{R}^d} \left\|y - x\right\|^2 \pi(\mathrm{d}y) + u_n^{(2)}(\gamma) \ ,$$

where $u_n^{(1)}(\gamma)$ and $u_n^{(2)}(\gamma)$ are explicit.

Decreasing step sizes

- If $\lim_{k\to+\infty} \gamma_k = 0$ and $\lim_{k\to+\infty} \Gamma_k = +\infty$, then

$$\lim_{p\to+\infty} W_2^2(\delta_x Q_\gamma^p, \pi) = 0 ,$$

  with explicit convergence.
- Order of convergence of $W_2^2(\delta_x Q_\gamma^n, \pi)$ for $\gamma_k = \gamma_1 k^{-\alpha}$: $\mathcal{O}(n^{-2\alpha})$.

Fixed step sizes

When $(\gamma_k)_{k\geq 1}$ is held constant:

- We optimize γ and p to get $W_1(\delta_x Q_\gamma^p, \pi) \leq \epsilon$. In particular, we find:

$$p = \mathcal{O}(\sqrt{d}\epsilon^{-1}) \ .$$

- At fixed number of iteration p, we can choose γ such that $W_1(\delta_x Q_\gamma^p, \pi) \leq Cn^{-1}$.
- Letting $p \to +\infty$, we get:

$$W_2(\pi_\gamma, \pi) \leq C\gamma \ .$$

Convergence of the Euler discretization in total variation

- If $\lim_{\gamma_k \to +\infty} \gamma_k = 0$, and $\sum_k \gamma_k = +\infty$ then

$$\lim_{p \to +\infty} \|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0 \; .$$

- If the step sizes are constant $\gamma_k = \gamma$ for all $k \in \mathbb{N}$, we can optimize γ and p to get

$$\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \epsilon \; ,$$

for a target precision $\epsilon > 0$.

	d	ε	L	m
γ	$\mathcal{O}(d^{-1})$	$\mathcal{O}(\varepsilon^2 / \log(\varepsilon^{-1}))$	$\mathcal{O}(L^{-2})$	$\mathcal{O}(m)$
p	$\mathcal{O}(d \log(d))$	$\mathcal{O}(\varepsilon^{-2} \log^2(\varepsilon^{-1}))$	$\mathcal{O}(L^2)$	$\mathcal{O}(m^{-2})$

Non-smooth potentials

The target distribution has a density π with respect to the Lebesgue measure on $\mathbb{R}^d$ of the form $x \mapsto \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y$ where $U = f + g$, with $f : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R}^d \to (-\infty, +\infty]$ are two lower bounded, convex functions satisfying:

1. f is continuously differentiable and gradient Lipschitz with Lipschitz constant L_f, *i.e.* for all $x, y \in \mathbb{R}^d$

$$\|\nabla f(x) - \nabla f(y)\| \leq L_f \|x - y\| \ .$$

2. g is lower semi-continuous and $\int_{\mathbb{R}^d} \mathrm{e}^{-g(y)} \mathrm{d}y \in (0, +\infty)$.

Moreau-Yosida regularization

- Let $\mathrm{h} : \mathbb{R}^d \to (-\infty, +\infty]$ be a l.s.c convex function and $\lambda > 0$. The λ-Moreau-Yosida envelope $\mathrm{h}^\lambda : \mathbb{R}^d \to \mathbb{R}$ and the proximal operator $\mathrm{prox}_\mathrm{h}^\lambda : \mathbb{R}^d \to \mathbb{R}^d$ associated with h are defined for all $x \in \mathbb{R}^d$ by

$$\mathrm{h}^\lambda(x) = \inf_{y \in \mathbb{R}^d} \left\{ \mathrm{h}(y) + (2\lambda)^{-1} \|x - y\|^2 \right\} \leq \mathrm{h}(x) \; .$$

- For every $x \in \mathbb{R}^d$, the minimum is achieved at a unique point, $\mathrm{prox}_\mathrm{h}^\lambda(x)$, which is characterized by the inclusion

$$x - \mathrm{prox}_\mathrm{h}^\lambda(x) \in \gamma \partial \mathrm{h}(\mathrm{prox}_\mathrm{h}^\lambda(x)) \; .$$

- The Moreau-Yosida envelope is a regularized version of g, which approximates g from below.

Properties of proximal operators

- As $\lambda \downarrow 0$, converges h^λ converges pointwise h, *i.e.* for all $x \in \mathbb{R}^d$,

$$\mathrm{h}^\lambda(x) \uparrow \mathrm{h}(x) \;, \quad \text{as } \lambda \downarrow 0 \;.$$

- The function h^λ is convex and continuously differentiable

$$\nabla \mathrm{h}^\lambda(x) = \lambda^{-1}(x - \mathrm{prox}_{\mathrm{h}}^\lambda(x)) \;.$$

- The Moreau-Yosida envelope is L-smooth: $\left\|\nabla \mathrm{h}^\lambda(x) - \nabla \mathrm{h}^\lambda(y)\right\| \leq \lambda^{-1} \left\|x - y\right\|$, for all $x, y \in \mathbb{R}^d$.

MY regularized potential

- If g is not differentiable, but the proximal operator associated with g is available, its λ-Moreau Yosida envelope g^λ can be considered.
- This leads to the approximation of the potential $U^\lambda : \mathbb{R}^d \to \mathbb{R}$ defined for all $x \in \mathbb{R}^d$ by

$$U^\lambda(x) = f(x) + g^\lambda(x) \; .$$

Theorem

Under (H), for all $\lambda > 0$, $0 < \int_{\mathbb{R}^d} \mathrm{e}^{-U^\lambda(y)} \mathrm{d}y < +\infty$.

Some approximation results

Theorem

Assume (H).

1. *Then,* $\lim_{\lambda \to 0} \|\pi^\lambda - \pi\|_{\mathrm{TV}} = 0$.
2. *Assume in addition that* g *is Lipschitz. Then for all* $\lambda > 0$,

$$\|\pi^\lambda - \pi\|_{\mathrm{TV}} \leq \lambda \|g\|_{\mathrm{Lip}}^2 \ .$$

3. *If* $g = \iota_{\mathcal{K}}$ *where* $\mathcal{K}$ *is a convex body of* $\mathbb{R}^d$. *Then for all* $\lambda > 0$ *we have*

$$\|\pi^\lambda - \pi\|_{\mathrm{TV}} \leq 2\left(1 + \mathsf{D}(\mathcal{K}, \lambda)\right)^{-1} \ ,$$

where $\mathsf{D}(\mathcal{K}, \lambda)$ *is explicit in the proof, and is of order* $\mathcal{O}(\lambda^{-1})$ *as* λ *goes to* 0.

The MYULA algorithm-I

Given a regularization parameter $\lambda > 0$ and a sequence of stepsizes $\{\gamma_k,\ k \in \mathbb{N}^*\}$, the algorithm produces the Markov chain $\{X_k^{\mathrm{M}},\ k \in \mathbb{N}\}$: for all $k \geq 0$,

$$X_{k+1}^{\mathrm{M}} = X_k^{\mathrm{M}} - \gamma_{k+1} \left\{ \nabla f(X_k^{\mathrm{M}}) + \lambda^{-1}(X_k^{\mathrm{M}} - \mathrm{prox}_g^{\lambda}(X_k^{\mathrm{M}})) \right\} + \sqrt{2\gamma_{k+1}} Z_{k+1} ,$$

where $\{Z_k,\ k \in \mathbb{N}^*\}$ is a sequence of i.i.d. d-dimensional standard Gaussian random variables.

The MYULA algorithm-II

- The ULA target the smoothed distribution π^λ.
- To compute the expectation of a function $h : \mathbb{R}^d \to \mathbb{R}$ under π from $\{X_k^{\mathrm{M}} \,;\, 0 \le k \le n\}$, an importance sampling step is used to correct the regularization.
- This step amounts to approximate $\int_{\mathbb{R}^d} h(x)\pi(x)\mathrm{d}x$ by the weighted sum

$$\mathrm{S}_n^h = \sum_{k=0}^{n} \omega_{k,n} h(X_k) \; , \; \text{with } \omega_{k,n} = \left\{ \sum_{k=0}^{n} \gamma_k \mathrm{e}^{\bar{g}^\lambda(X_k^{\mathrm{M}})} \right\}^{-1} \gamma_k \mathrm{e}^{\bar{g}^\lambda(X_k^{\mathrm{M}})} \; ,$$

where for all $x \in \mathbb{R}^d$

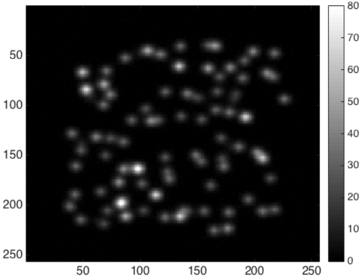
$$\bar{g}^\lambda(x) = g^\lambda(x) - g(x) = g(\mathrm{prox}_g^\lambda(x)) - g(x) + (2\lambda)^{-1} \left\| x - \mathrm{prox}_g^\lambda(x) \right\|^2 \; .$$

Image deconvolution

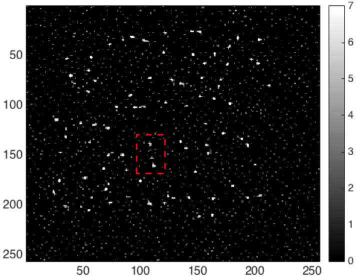
- Objective recover an original image $\boldsymbol{x} \in \mathbb{R}^n$ from a blurred and noisy observed image $\boldsymbol{y} \in \mathbb{R}^n$ related to $\boldsymbol{x}$ by the linear observation model $\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{w}$, where H is a linear operator representing the blur point spread function and $\boldsymbol{w}$ is a Gaussian vector with zero-mean and covariance matrix $\sigma^2 \boldsymbol{I}_n$.
- This inverse problem is usually ill-posed or ill-conditioned: exploits prior knowledge about $\boldsymbol{x}$.
- Consider the ℓ_1 norm prior, $\pi(\boldsymbol{x}) \propto \exp\left(-\alpha \|\boldsymbol{x}\|_1\right)$, then

$$\pi(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left[-\|\boldsymbol{y} - H\boldsymbol{x}\|^2 / 2\sigma^2 - \alpha \|\boldsymbol{x}\|_1\right].$$

Microscopy dataset



(a)



(b)

HPD credible region

- We want to test if the molecules are indeed present in true image (as opposed to being noise artefacts for example),
- Uncertainty about their position.
- For this, it can be relevent to compute the HPD credible region

$$C^*_\alpha = \{x : U(x) \leq \eta_\alpha\}$$

with $\eta_\alpha \in \mathbb{R}$ chosen such that $\mathbb{P}(\boldsymbol{x} \in C^*_\alpha | \boldsymbol{y}) = 1 - \alpha$ holds. Here we use $\alpha = 0.01$ related to the 99% confidence level.

Convergence of MYULA

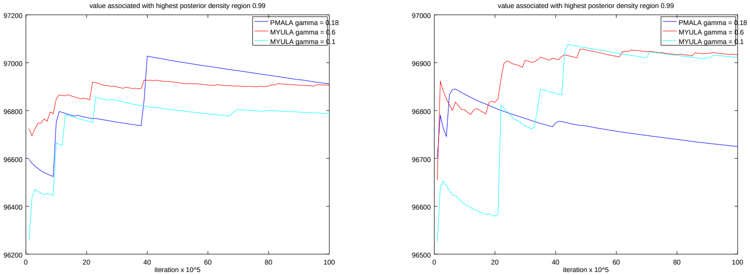


Figure : Convergence of MYULA and PMALA after 5×10^6 iterations

Comparison with PMALA

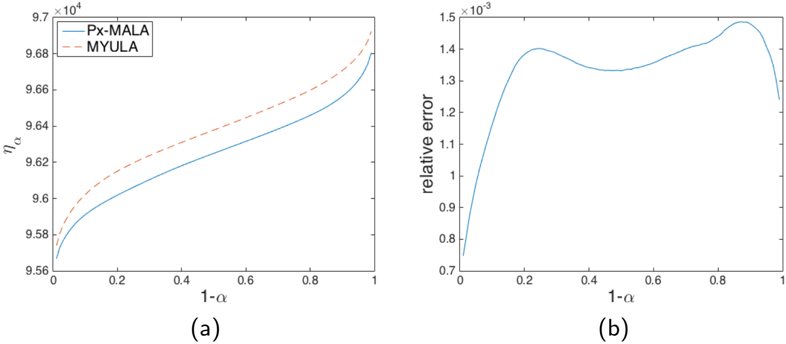


Figure : Microscopy experiment: (a) HDP region thresholds η_α for MYULA (2×10^6 iterations $\lambda = 1, \gamma = 0.6$) and PMALA ($2 \times 10^7$ iterations), (b) relative approximation error of MYULA.

Application of MYULA

- We check that a group of three molecules are in the image by a test.
- MYULA (2×10^6 iterations $\lambda = 1, \gamma = 0.6$) and PMALA ($2 \times 10^7$ iterations) give the same answer.

Thank you for your attention.