From stability of Langevin diffusion to convergence of proximal MCMC for non-log-concave sampling

Marien Renaud [1,3] Valentin de Bortoli [2] Arthur Leclaire [3] Nicolas Papadakis [1]

[1]Univ. Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251, F-33400 Talence, France [2]ENS, CNRS, PSL University, Paris, 75005, FRANCE [3]LTCI, Télécom Paris, IP Paris, France

Overview

Objective: Sampling a distribution $\pi \propto e^{-V}$, with V non convex.

Langevin. We ise the Unajusted Langevin Algorithm (ULA) MCMC to sample π, defined by

$$X_{k+1} = X_k - \gamma \nabla V(X_k) + \sqrt{2\gamma} Z_{k+1} \tag{1}$$

with $\gamma > 0$ the stepsize and $Z_{k+1} \sim \mathcal{N}(0, I_d)$. In practice, ∇V is approximated by $b \approx \nabla V$, leading to the inexact ULA defined by

$$X_{k+1} = X_k - \gamma b(X_k) + \sqrt{2\gamma} Z_{k+1}$$

Composite potential $V = f + g$. Many problem, e.g. inverse problem in imaging, involves a composite potential $V = f + g$ with two non convex functions f, g and a non smooth function g.
ULA cannot be applied. Instead we rely on the Proximal Stochastic Gradient Langevin Algorithm (PSGLA)

$$X_{k+1} = \mathsf{Prox}_{\gamma g}\left(X_k - \gamma \nabla f(X_k) + \sqrt{2\gamma} Z_{k+1}\right), \tag{2}$$

with the proximal operator defined by $\mathsf{Prox}_{\gamma g}(x) = \arg\min_{y \in \mathbb{R}^d} \frac{1}{2\gamma}\|x - y\|^2 + g(y)$.

Contributions

- New stability result in b for iULA algorithm and bound on the discretization error
- First non convex proof of convergence for PSGLA
- Practical gain of PSGLA in realistic scenario of inverse problem in imaging

Definitions

- The p-Wasserstein distance between μ and ν is defined, for $p \geq 1$, by

$$\mathbf{W}_p(\mu, \nu) = \left(\min_{\beta \in \Pi} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\beta(x, y)\right)^{\frac{1}{p}}, \tag{3}$$

with Π the set of probability law β on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν.

- g is ρ-weakly convex, with $\rho > 0$, if and only if $g + \frac{\rho}{2}\|\cdot\|^2$ is convex.
- $(X_k)_{k \geq 0}$ has an *invariant law* if there exists μ such that if $X_0 \sim \mu$, then $\forall k \geq 0$, $X_k \sim \mu$.
- $(X_k)_{k \geq 0}$ is *geometrically ergodic*, if the law of X_k converges geometrically to the invariant law, i.e. there exist $A \geq 0$, $\rho \in (0, 1)$, an invariant law μ, such that $\mathbf{W}_1(p_{X_k}, \mu) \leq A\rho^k$ with p_{X_k} the law of X_k.

Stability of Langevin diffusion

For two iULA with two drifts defined for $i \in \{1, 2\}$ by

$$X^i_{k+1} = X^i_k - \gamma b^i(X^i_k) + \sqrt{2\gamma} Z^i_{k+1}, \tag{4}$$

Assumption 1: There exist $L, R \geq 0$ and $m > 0$ such that the drift b verifies (i) b is L-Lipschitz, i.e. $\forall x, y \in \mathbb{R}^d$, $\|b(x) - b(y)\| \leq L\|x - y\|$ (ii) $\forall x, y \in \mathbb{R}^d$ such that $\|x - y\| \geq R$, we have $\langle b(x) - b(y), x - y\rangle \geq m\|x - y\|^2$.

Theorem: If b^1, b^2 satisfy Assumption 1. X^1_k, X^2_k are two geometrically ergodic Markov Chains with invariant laws $\pi^1_\gamma, \pi^2_\gamma$. Then for $\gamma_0 = \frac{m}{L^2}$ and $p \in \mathbb{N}^\star$ there exist $C_p, C \geq 0$ such that $\forall \gamma \in (0, \gamma_0]$, we have

$$\mathbf{W}_p(\pi^1_\gamma, \pi^2_\gamma) \leq C_p \|b^1 - b^2\|^{\frac{1}{p}}_{\ell_2(\pi^1_\gamma)}, \tag{5}$$

$$\|\pi^1_\gamma - \pi^2_\gamma\|_{TV} \leq C\|b^1 - b^2\|_{\ell_2(\pi^1_\gamma)}. \tag{6}$$

Corollary: If $b, \nabla V$ verify Assumption 1. Then for $\gamma_0 = \frac{m}{L^2}$, and $\gamma \in (0, \gamma_0]$, the Markov Chain (1) is geometrically ergodic with an invariant law $\hat{\pi}_\gamma$. Moreover for $p \in \mathbb{N}^\star$, there exist $C_p, D_p, C, D \geq 0$ such that $\forall \gamma \in (0, \gamma_0]$, we have

$$\mathbf{W}_p(\hat{\pi}_\gamma, \pi) \leq C_p\|b - \nabla V\|^{\frac{1}{p}}_{\ell_2(\hat{\pi}_\gamma)} + D_p \gamma^{\frac{1}{2p}}$$

$$\|\hat{\pi}_\gamma - \pi\|_{TV} \leq C\|b - \nabla V\|_{\ell_2(\hat{\pi}_\gamma)} + D\gamma^{\frac{1}{2}}.$$

Convergence of PSGLA

Question: Does PSGLA converge in non-convex setting?

Assumption 2: The potential V is composite, i.e. $V = f + g$ where (i) f is L_f-smooth, i.e. ∇f is L_f-Lipschitz. (ii) g is ρ-weakly convex. PSGLA (2) can be reformulated as a two points algorithm

$$Y_{k+1} = X_k - \gamma \nabla f(X_k) + \sqrt{2\gamma} Z_{k+1}$$
$$X_{k+1} = \mathsf{Prox}_{\gamma g}(Y_{k+1}).$$

Under Assumption 2, we get that the iterates Y_k verify the equation

$$Y_{k+1} = \mathsf{Prox}_{\gamma g}(Y_k) - \gamma \nabla f(\mathsf{Prox}_{\gamma g}(Y_k)) + \sqrt{2\gamma} Z_{k+1}$$
$$= Y_k - \gamma b^\gamma(Y_k) + \sqrt{2\gamma} Z_{k+1},$$

where the drift b^γ is defined for $y \in \mathbb{R}^d$ as

$$b^\gamma(y) = \nabla f(y - \gamma \nabla g^\gamma(y)) + \nabla g^\gamma(y).$$

Assumption 3: (i) $\forall \gamma \in (0, \frac{1}{\rho})$, g is L_g-smooth on $\mathsf{Prox}_{\gamma g}\left(\mathbb{R}^d\right)$. (ii) g^γ is μ-strongly convex at infinity with $\mu \geq 8L_f + 4L_g$, i.e. there exists $\gamma_1 > 0$ and $R_0 \geq 0$ such that $\forall \gamma \in (0, \gamma_1]$, $\nabla^2 g^\gamma \succeq \mu I_d$, on $\mathbb{R}^d \setminus B(0, R_0)$.

Theorem of convergence of PSGLA

Under Assumptions 2-3, there exist $r \in (0, 1)$, $C_1, C_2 \in \mathbb{R}_+$ such that $\forall \gamma \in (0, \bar{\gamma}]$, with $\bar{\gamma} = \min\left(\frac{1}{2L_g}, \frac{\mu}{32(L_f + L_g)^2}, \frac{1}{2\rho}, \gamma_1\right)$, where L_g, L_f, ρ, γ_1 are defined in Assumptions 2-3, and $\forall k \in \mathbb{N}$, we have

$$\mathbf{W}_p(p_{X_k}, \nu_\gamma) \leq C_1 r^{k\gamma} + C_2 \gamma^{\frac{1}{2p}}, \tag{7}$$

with p_{X_k} the distribution of Y_k and $\nu_\gamma \propto \mathsf{Prox}_{\gamma g}\#e^{-f-g^\gamma}$.

Experiments

PnP-PSGLA: approximation of the Proximal operator by a pretrained Gaussian denoiser

$$X_{k+1} = D_{\sqrt{\gamma}}\left(X_k - \frac{\gamma}{\lambda}\nabla f(X_k) + \sqrt{2\gamma} Z_{k+1}\right).$$

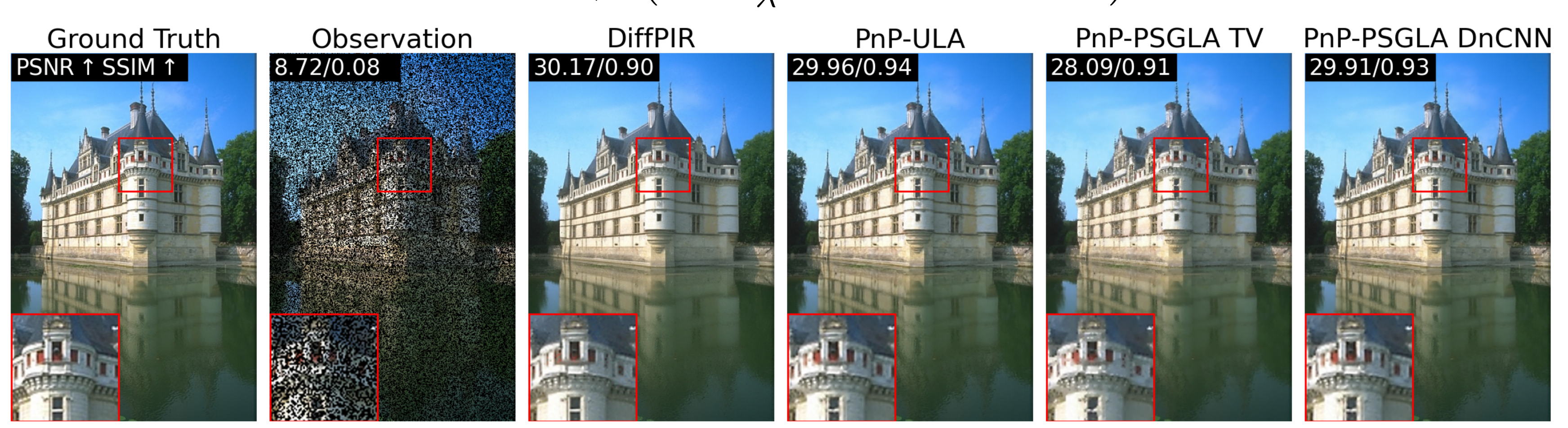

Figure 1. Qualitative result for image inpainting with 50% masked pixels and a noise level of $\sigma = 1/255$. PnP-ULA is run with 1,000,000 iterations and PSGLA with 10,000 iterations.

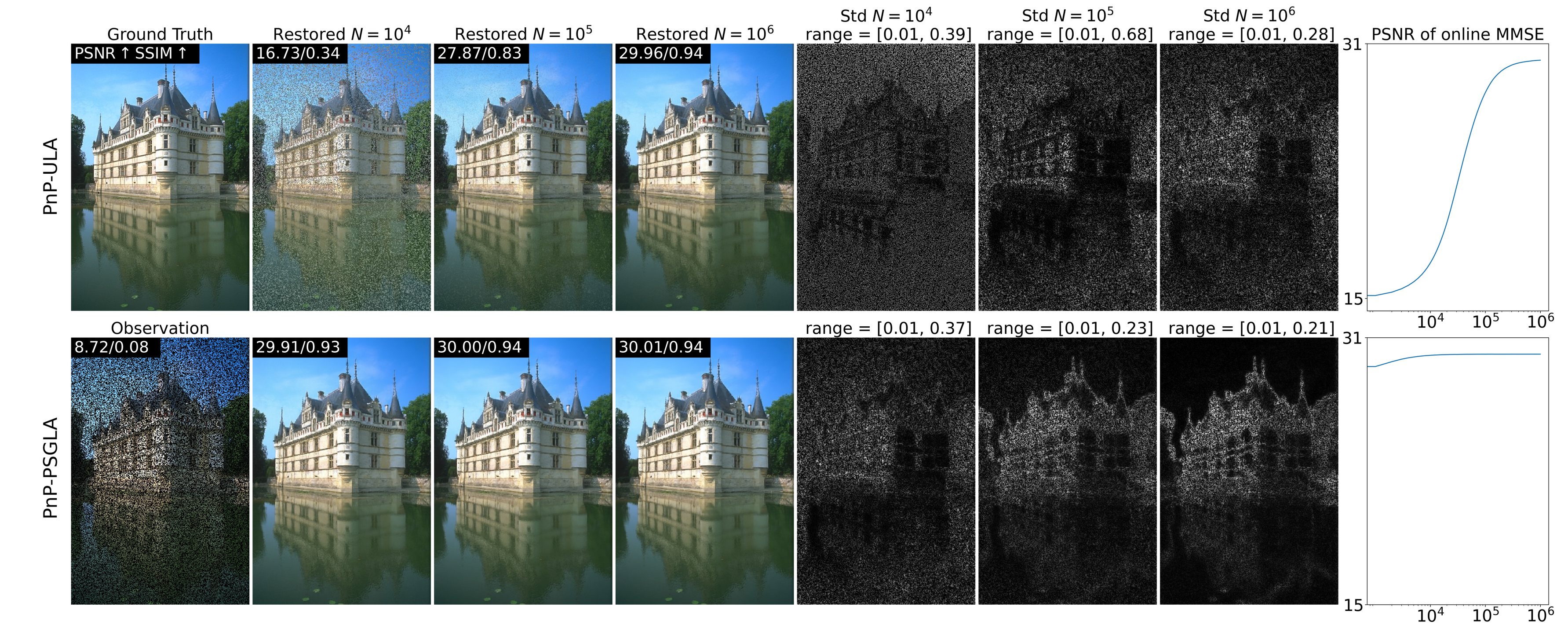

Figure 2. PnP-ULA and PnP-PSGLA with various N.

Algorithm	Denoiser	PSNR ↑	SSIM ↑	LPIPS ↓	N ↓	time (s) ↓	convergent
DiffPIR	GSDRUNet	29.99	0.88	0.06	**20**	**1**	✗
RED	DnCNN	30.49	0.89	0.06	500	6	✓
RED	GSDRUNet	29.26	0.88	0.12	500	20	✓
PnP	DnCNN	30.50	0.91	0.06	500	6	✓
PnP	GSDRUNet	30.52	**0.92**	0.07	500	20	✓
PnP-ULA	DnCNN	27.89	0.82	0.12	100,000	1,200	✓
PnP-PSGLA	TV	29.24	0.89	0.08	1,000	25	✓
PnP-PSGLA	DnCNN	**30.81**	**0.92**	**0.05**	10,000	120	✓

Table 1. Quantitative results for image inpainting with 50% masked pixels on CBSD68 dataset.