Concentration analysis of multivariate elliptic diffusions

Lukas Trottner (Aarhus University)

joint work with: Cathrine Aeckerle-Willems (University of Mannheim) and Claudia Strauch (Aarhus University)

Setup and aims of the paper

We consider a (weak) solution X of the SDE

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,$$

$b \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$, $\sigma \in \mathrm{Lip}(\mathbb{R}^d;\mathbb{R}^{d\times d})$ and bounded, uniformly elliptic diffusion matrix $a := \sigma\sigma^\top$, given the drift condition

$$\langle b(x), x/\|x\|\rangle \leq -\mathfrak{r}\|x\|^{-q}, \quad \|x\| > A, q \in [-1,1). \qquad (\mathcal{D}(q))$$

The smaller the parameter q, the faster the (sub)exponential speed of ergodicity towards the invariant distribution π of X.

We quantify the concentration around the ergodic mean $\pi(f) := \int f(x)\,\pi(\mathrm{d}x)$ of both the continuous-time MC estimator and its discrete version

$$\widehat{\pi}_T(f) := \frac{1}{T}\int_0^T f(X_t)\,\mathrm{d}t, \quad \widehat{\pi}_{n,\Delta}(f) := \frac{1}{n}\sum_{k=1}^n f(X_{k\Delta}),$$

for polynomially growing functions f.

Strategy: Martingale approximation

For the generator $L = b^\top\nabla + \frac{1}{2}\sum_{i,j} a_{i,j}\partial_{x_i}\partial_{x_j}$ the Poisson equation

$$Lg = f$$

has a nice solution $L^{-1}[f]$ if f is π-centered and grows polynomially. The Itō–Krylov formula then gives

$$\int_0^t f(X_s)\,\mathrm{d}s = \underbrace{\int_0^t (-\nabla L^{-1}[f](X_s))^\top \sigma(X_s)\,\mathrm{d}W_s}_{\text{(loc.) martingale}} + \underbrace{L^{-1}[f](X_t) - L^{-1}[f](X_0)}_{\text{remainder}}$$

☞ use polynomial growth bounds on on $L^{-1}[f], \nabla L^{-1}[f]$ from [4]

☞ precise understanding of the tails of π given ($\mathcal{D}(q)$) then allows to control all L^p-norms of the centered statistic $\widehat{\pi}_T(f) - \pi(f)$

Main results

Theorem 1. Assume ($\mathcal{D}(q)$), $\|b(x)\| \lesssim 1 + \|x\|^\kappa$ and $|f(x)| \leq \mathfrak{L}(1+\|x\|^\eta)$. Let

$$\rho(\eta,\kappa,q) := \begin{cases} 1/(1-q_+), & \eta = 0 \\ \frac{1}{2} + \frac{\eta+\kappa+1+q}{1-q_+}, & \eta > 0. \end{cases}$$

Then, there exists a constant $\mathfrak{c} > 0$ s.t. for any $x \geq 2/\sqrt{T}$,

$$\mathbb{P}^\pi\Big(\big|\widehat{\pi}_T(f) - \pi(f)\big| > x\Big) \leq \exp\Big(-\mathfrak{c}\Big(\frac{x\sqrt{T}}{\mathfrak{L}}\Big)^{1/\rho(\eta,\kappa,q)}\Big).$$

This translates into the following PAC-bounds

Poincaré, $\eta = 0$	log-Sobolev, $\eta \leq 2$	subexponential, $\eta > 0$
$\frac{\log(1/\delta)}{\varepsilon}$	$\frac{\log(1/\delta)}{\varepsilon}$	$\frac{\log(1/\delta)^{2\rho(\eta,\kappa,q)}}{\varepsilon^2}$

sample complexity s.t. (ε,δ)-PAC-bound $\mathbb{P}^\pi(|\widehat{\pi}_T(f) - \pi(f)| \leq \varepsilon) \geq 1-\delta$ holds

By relating the discrete estimator with sampling frequency Δ to the continuous estimator, we obtain concentration bounds for $\widehat{\pi}_{n,\Delta}(f)$:

Theorem 2. Assume ($\mathcal{D}(q)$), $\|b(x)\| \lesssim 1 + \|x\|^\kappa$ and $\|D^k f(x)\| \lesssim 1 + \|x\|^{\eta_k}$, $k = 0,1,2$. Define $\alpha := (\kappa+\eta_1)\vee\eta_2$, and let $\widetilde{\gamma} > 1+q$, $r > 1$, s.t. $\widetilde{\gamma} - (1+q) > r(\alpha\vee(1+q)/(r-1))$. Then, for $p \geq 2$,

$$\begin{aligned}
&\|\widehat{\pi}_{n,\Delta}(f) - \pi(f)\|_{L^p(\mathbb{P}^\pi)} \\
&\leq \mathfrak{D}\Big(\underbrace{\Delta + \sqrt{\frac{\Delta}{n}}p^{\frac{\max\{(\widetilde{\gamma}+2\alpha+1-q_+)/2,\eta_1+1-q_+\}}{1-q_+}}}_{\text{sampling error}} + \frac{1}{\sqrt{n\Delta}}p^{\frac{1}{2}+\frac{\eta+\kappa+1+q}{1-q_+}}\Big) \\
&=: \Phi(n,\Delta,p),
\end{aligned}$$

and

$$\mathbb{P}^\pi\Big(\big|\widehat{\pi}_{n,\Delta}(f) - \pi(f)\big| > \mathrm{e}\Phi(n,\Delta,x)\Big) \leq \mathrm{e}^{-x}, \quad x \geq 2.$$

Sample complexities for heavy-tailed Langevin MCMC

Under mild assumptions on the potential U, the Langevin diffusion

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

has invariant density $\pi(x) \propto \exp(-U(x))$

☞ approximate sampling from π by numerical approximation of X, e.g., Euler scheme with $(\xi_n) \underset{\text{iid}}{\sim} \mathcal{N}(0,\mathbb{I}_d)$:

$$\vartheta_{n+1}^{(\Delta)} = \vartheta_n^{(\Delta)} - \Delta\nabla U(\vartheta_n^{(\Delta)}) + \sqrt{2\Delta}\xi_{n+1}, \quad \vartheta_0^{(\Delta)} \sim X_0.$$

☞ abundant literature on sampling precision in TV or Wasserstein distance for U strongly convex or modifications thereof [1, 2, 3] ⇝ $\pi(x)\,\mathrm{d}x$ is sub-Gaussian

Assume instead that for $q \in (0,1)$

$$\langle\nabla U(x), x/\|x\|\rangle \geq \mathfrak{r}\|x\|^{-q}, \quad \|x\| > A. \qquad (\mathcal{U}(q))$$

⇝ prototypical example: $\pi(x) \propto \exp(-\beta\|x\|^{1-q})$ is heavy-tailed and U is non-convex

Proposition 3.

	step length Δ	sample size n	burn-in m
ε-prec. sampling	$\frac{\varepsilon^2}{d(\log(\mathfrak{C}/\varepsilon))^{(1-q)/(1+q)}}$	$\frac{d(\log(\mathfrak{C}/\varepsilon))^{2(1-q)/(1+q)}}{\varepsilon^2}$	–
(ε,δ)-PAC bound	$\frac{(\delta\varepsilon)^2}{d(\log(1/\delta))^{2(\eta_0+(q+3)/2)/(1-q)}}$	$\frac{d\mathfrak{D}^2(\log(1/\delta))^{(4(\eta_0+(q+3)/2))/(1-q)}}{\delta^2\varepsilon^4}$	$\frac{d(\log(1/\delta))^{2(\eta_0+q+2)/(1-q)}}{(\delta\varepsilon)^2}$

Order of sufficient sampling frequency Δ, sample size n and burn-in m for (ε,δ)-PAC bounds and sampling within ε-TV margin

References

[1] A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676, 2017.

[2] A. Durmus, S. Majewski, and B. a. Miasojedow. Analysis of Langevin Monte Carlo via convex optimization. *J. Mach. Learn. Res.*, 20:Paper No. 73, 46, 2019.

[3] A. Durmus and E. Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 2017.

[4] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085, 2001.