Concentration analysis of multivariate elliptic diffusions

Stochastics Seminar – Aarhus

Lukas Trottner
joint work with Cathrine Aeckerle-Willems and Claudia Strauch

20 April 2023

Aarhus University University of Mannheim

AARHUS UNIVERSITY

Some known concentration results for Markov processes

Let X be a nice ergodic Markov processes on $\mathbb{R}^d$ with semigroup $(P_t)_{t\geqslant 0}$, generator L and invariant distribution μ. We are interested in

$$\mathbb{C}_\nu(f, T, x) \coloneqq \mathbb{P}^\nu\left(\left|\frac{1}{T}\int_0^T f(X_t)\,\mathrm{d}t - \mu(f)\right| > x\right), \quad f \in \mathbb{L}^2(\mu), x, T > 0.$$

Some known concentration results for Markov processes

Let X be a nice ergodic Markov processes on $\mathbb{R}^d$ with semigroup $(P_t)_{t\geqslant 0}$, generator L and invariant distribution μ. We are interested in

$$\mathbb{C}_\nu(f,T,x) \coloneqq \mathbb{P}^\nu\left(\left|\frac{1}{T}\int_0^T f(X_t)\,\mathrm{d}t - \mu(f)\right| > x\right), \quad f \in \mathbb{L}^2(\mu), x, T > 0.$$

Bounds have been mostly studied with two approaches (*Lyapunov vs. Poincaré* [BCG08]):

1. Functional inequalities:
   - Poincaré inequality:

$$\mathrm{Var}_\mu(g) \coloneqq \mu(g^2) - \mu(g)^2 \leqslant C_\mathrm{P}\langle -Lg, g\rangle_\mu \coloneqq C_\mathrm{P}\int g(x)(-Lg(x))\,\mu(\mathrm{d}x), \quad g \in D(L).$$

[Lez01] For $\|f\|_\infty < \infty$ and $\nu \ll \mu$, $\mathrm{d}\nu/\mathrm{d}\mu \in \mathbb{L}^2(\mu)$,

$$\mathbb{C}_\nu(f,T,x) \leqslant 2\left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{\mathbb{L}^2(\mu)} \exp\left(-\frac{Tx^2}{2(\sigma^2(f) + 2C_\mathrm{P}\|f\|_\infty x)}\right),$$

Some known concentration results for Markov processes

Let X be a nice ergodic Markov processes on $\mathbb{R}^d$ with semigroup $(P_t)_{t\geqslant 0}$, generator L and invariant distribution μ. We are interested in

$$\mathbb{C}_\nu(f,T,x) \coloneqq \mathbb{P}^\nu\Big(\Big|\frac{1}{T}\int_0^T f(X_t)\,\mathrm{d}t - \mu(f)\Big| > x\Big), \quad f\in\mathbb{L}^2(\mu), x, T>0.$$

Bounds have been mostly studied with two approaches (*Lyapunov vs. Poincaré* [BCG08]):

1. Functional inequalities:
   - Poincaré inequality:
     $$\mathrm{Var}_\mu(g) \coloneqq \mu(g^2)-\mu(g)^2 \leqslant C_{\mathrm{P}}\langle -Lg, g\rangle_\mu \coloneqq C_{\mathrm{P}}\int g(x)(-Lg(x))\,\mu(\mathrm{d}x), \quad g\in D(L).$$
     [Lez01] For $\|f\|_\infty < \infty$ and $\nu \ll \mu$, $\mathrm{d}\nu/\mathrm{d}\mu \in \mathbb{L}^2(\mu)$,
     $$\mathbb{C}_\nu(f,T,x) \leqslant 2\Big\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\Big\|_{\mathbb{L}^2(\mu)}\exp\Big(-\frac{Tx^2}{2(\sigma^2(f)+2C_{\mathrm{P}}\|f\|_\infty x)}\Big),$$
   - log-Sobolev inequality: $(P_t)_{t\geqslant 0}$ symmetric and
     $$\mathrm{Ent}_\mu(g^2) \coloneqq \mu(g^2\log g^2)-\mu(g^2)\log\mu(g^2) \leqslant 2C_{\mathrm{LS}}\langle -Lg, g\rangle_\mu, \quad g\in D(L).$$
     [GGW14] For $|f(x)| \lesssim 1+\|x\|^2$,
     $$\mathbb{C}_\nu(f,T,x) \leqslant 2\Big\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\Big\|_{\mathbb{L}^2(\mu)}\exp\Big(-\frac{Tx^2}{2(\sigma^2(f)+C_{\mathrm{P}}(\Lambda^*)^{-1}(2C_{\mathrm{LS}}/C_{\mathrm{P}})x)}\Big)$$

2. Mixing assumptions: for $q \in [0,1)$,

$$\alpha_\nu(t) \coloneqq \sup_{s\geqslant 0} \sup_{A\in\sigma(X_u, u\leqslant s), B\in\sigma(X_u, u\geqslant s+t)} |\mathbb{P}^\nu(A\cap B) - \mathbb{P}^\nu(A)\mathbb{P}^\nu(B)| \lesssim \exp(-t^{\frac{1-q}{1+q}}).$$

For reasonable ν guaranteed given (sub)exponential ergodicity of (P_t), i.e.,

$$\|P_t(x,\cdot) - \mu\|_{\mathrm{TV}} \lesssim V(x)\exp(-t^{\frac{1-q}{1+q}}).$$

[CG08] For $\|f\|_\infty < \infty$,

$$\mathbb{C}_\mu(f,T,x) \leqslant 2\exp\left(-c(q)\left(\frac{x\sqrt{T}}{\|f\|_\infty}\right)^{1-q}\right), \quad x \geqslant C(c,q)/\sqrt{T}.$$

Martingale approximation for diffusions

- Let X be a (weak) solution to 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, $a \coloneqq \sigma\sigma^\top$ s.t. $\lambda_-\mathbb{I} \leqslant a(x) \leqslant \lambda_+\mathbb{I}$, $\forall x$

Martingale approximation for diffusions

- Let X be a (weak) solution to the SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t,$$

$b \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$, $\sigma \in \mathrm{Lip}(\mathbb{R}^d;\mathbb{R}^{d\times d})$ and bounded, $a \coloneqq \sigma\sigma^\top$ s.t. $\lambda_- \mathbb{I} \leqslant a(x) \leqslant \lambda_+ \mathbb{I}$, $\forall x$

- Let $L = b^\top \nabla + \sum_{i,j} a_{i,j} \partial_{x_i} \partial_{x_j}$ and suppose that for given $f : \mathbb{R}^d \to \mathbb{R}$ the Poisson equation $Lg = f$ has some sufficiently regular solution $L^{-1}[f]$

Martingale approximation for diffusions

- Let X be a (weak) solution to 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, $a := \sigma\sigma^\top$ s.t. $\lambda_-\mathbb{I} \leqslant a(x) \leqslant \lambda_+\mathbb{I}$, $\forall x$
- Let $L = b^\top\nabla + \sum_{i,j} a_{i,j}\partial_{x_i}\partial_{x_j}$ and suppose that for given $f\colon \mathbb{R}^d \to \mathbb{R}$ the Poisson equation $Lg = f$ has some sufficiently regular solution $L^{-1}[f]$
- By Itō's formula: $L^{-1}[f](X_t) - L^{-1}[f](X_0) = \int_0^t LL^{-1}[f](X_s)\,\mathrm{d}s + \int_0^t (\nabla L^{-1}[f](X_s))^\top \sigma(X_s)\,\mathrm{d}W_s$ and hence
$$\int_0^t f(X_s)\,\mathrm{d}s = \underbrace{\int_0^t (-\nabla L^{-1}[f](X_s))^\top \sigma(X_s)\,\mathrm{d}W_s}_{\text{(loc.) martingale}} + \underbrace{L^{-1}[f](X_t) - L^{-1}[f](X_0)}_{\text{remainder}}$$

Martingale approximation for diffusions

- Let X be a (weak) solution to 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, $a := \sigma\sigma^\top$ s.t. $\lambda_-\mathbb{I} \leqslant a(x) \leqslant \lambda_+\mathbb{I}$, $\forall x$

- Let $L = b^\top\nabla + \sum_{i,j} a_{i,j}\partial_{x_i}\partial_{x_j}$ and suppose that for given $f\colon \mathbb{R}^d \to \mathbb{R}$ the Poisson equation $Lg = f$ has some sufficiently regular solution $L^{-1}[f]$
- By Itō's formula: $L^{-1}[f](X_t) - L^{-1}[f](X_0) = \int_0^t LL^{-1}[f](X_s)\,\mathrm{d}s + \int_0^t (\nabla L^{-1}[f](X_s))^\top\sigma(X_s)\,\mathrm{d}W_s$ and hence

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

$\rightsquigarrow$ If we have some control on $L^{-1}[f], \nabla L^{-1}[f]$ we can use martingale approximation for derivation of concentration bounds

Martingale approximation for diffusions

- Let X be a (weak) solution to 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, $a \coloneqq \sigma\sigma^\top$ s.t. $\lambda_- \mathbb{I} \leqslant a(x) \leqslant \lambda_+ \mathbb{I}$, $\forall x$

- Let $L = b^\top \nabla + \sum_{i,j} a_{i,j}\partial_{x_i}\partial_{x_j}$ and suppose that for given $f\colon \mathbb{R}^d \to \mathbb{R}$ the Poisson equation $Lg = f$ has some sufficiently regular solution $L^{-1}[f]$
- By Itō's formula: $L^{-1}[f](X_t) - L^{-1}[f](X_0) = \int_0^t LL^{-1}[f](X_s)\,\mathrm{d}s + \int_0^t (\nabla L^{-1}[f](X_s))^\top \sigma(X_s)\,\mathrm{d}W_s$ and hence

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

$\rightsquigarrow$ If we have some control on $L^{-1}[f], \nabla L^{-1}[f]$ we can use martingale approximation for derivation of concentration bounds

- employed in case $d = 1$ for exponentially ergodic diffusions in [AWS21; GP07] and for $d \geqslant 1$ and periodic drift [NR20] in the context of drift estimation

Poisson equation under subexponential drift assumptions

Assume $\|b(x)\| \lesssim 1 + \|x\|^{\kappa}$ and for some $q \in (-1, 1)$, $\mathfrak{r}, A > 0$,

$$\langle b(x), x/\|x\| \rangle \leqslant -\mathfrak{r}\|x\|^{-q}, \quad \|x\| > A. \tag{$\mathcal{D}(q)$}$$

[DFG09] implies

$$\|P_t(x,\cdot) - \mu\|_{\mathrm{TV}} \lesssim \exp\left(\iota\|x\|^{1-q_+}\right) \exp\left(-\iota' t^{\frac{1-q_+}{1+q_+}}\right) \quad \text{and} \quad \int_{\mathbb{R}^d} \exp\left(\iota\|x\|^{1-q_+}\right) \mu(\mathrm{d}x) < \infty.$$

[PV01; BRS18] If $\mu(f) = 0$ and $|f(x)| \lesssim 1 + \|x\|^{\eta}$, then for $L^{-1}[f](x) \coloneqq -\int_0^{\infty} P_t f(x)\,\mathrm{d}t$ we have $L^{-1}[f] \in \mathcal{W}^{2,p}_{\mathrm{loc}}(\mathbb{R}^d)$ for any $p > 1$, $L^{-1}[f]$ solves the Poisson equation and

$$|L^{-1}[f](x)| \lesssim 1 + \|x\|^{\eta+1+q}, \quad \|\nabla L^{-1}[f](x)\| \lesssim 1 + \|x\|^{\eta+\kappa+1+q}.$$

Poisson equation under subexponential drift assumptions

Assume $\|b(x)\| \lesssim 1 + \|x\|^\kappa$ and for some $q \in (-1,1)$, $\mathfrak{r}, A > 0$,

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

[DFG09] implies

$$\|P_t(x,\cdot) - \mu\|_{\mathrm{TV}} \lesssim \exp\left(\iota\|x\|^{1-q_+}\right)\exp\left(-\iota' t^{\frac{1-q_+}{1+q_+}}\right) \quad \text{and} \quad \int_{\mathbb{R}^d} \exp\left(\iota\|x\|^{1-q_+}\right)\mu(\mathrm{d}x) < \infty.$$

[PV01; BRS18] If $\mu(f) = 0$ and $|f(x)| \lesssim 1 + \|x\|^\eta$, then for $L^{-1}[f](x) := -\int_0^\infty P_t f(x)\,\mathrm{d}t$ we have $L^{-1}[f] \in \mathcal{W}^{2,p}_{\mathrm{loc}}(\mathbb{R}^d)$ for any $p > 1$, $L^{-1}[f]$ solves the Poisson equation and

$$|L^{-1}[f](x)| \lesssim 1 + \|x\|^{\eta+1+q}, \quad \|\nabla L^{-1}[f](x)\| \lesssim 1 + \|x\|^{\eta+\kappa+1+q}.$$

Proposition [DFG09], [TAWS23]

Given $(\mathcal{D}(q))$ we have for $\gamma \geqslant 1+q$, $r_{\gamma,q}(t) \sim (1+t)^{(\gamma-(1+q))/(1+q)}$, $f_{\gamma,q}(x) \sim 1 + \|x\|^{\gamma-(1+q)}$,

$$(\Psi_1(r_{\gamma,q}(t)) \vee 1)\|P_t(x,\cdot) - \mu\|_{1\vee\Psi_2\circ f_{\gamma,q}} \leqslant C(\Psi)(1 + \|x\|^\gamma),$$

where $\|\nu\|_f := \sup_{|g|\leqslant f}|\nu(g)|$ and (Ψ_1, Ψ_2) is a pair of inverse Young functions (i.e., $xy \leqslant \Psi_1^{-1}(x) + \Psi_2^{-1}(y)$)

Continuous-time concentration result

Theorem [TAWS23]

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

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

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

$$\mathbb{C}_{\mu}(f, T, x) \coloneqq \mathbb{P}^{\mu}\Big(\Big|\frac{1}{T}\int_0^T f(X_t)\,\mathrm{d}t - \mu(f)\Big| > x\Big) \leqslant \exp\Big(-\mathfrak{c}\Big(\frac{x\sqrt{T}}{\mathfrak{L}}\Big)^{1/\rho(\eta,\kappa,q)}\Big).$$

Continuous-time concentration result

Theorem [TAWS23]

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

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

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

$$\mathbb{C}_\mu(f, T, x) \coloneqq \mathbb{P}^\mu\left(\left|\frac{1}{T}\int_0^T f(X_t)\,\mathrm{d}t - \mu(f)\right| > x\right) \leqslant \exp\left(-\mathfrak{c}\left(\frac{x\sqrt{T}}{\mathfrak{L}}\right)^{1/\rho(\eta,\kappa,q)}\right).$$

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

Table 1: Order of sufficient sample length $\Psi(\varepsilon, \delta)$ s.t. (ε, δ)-PAC-bound $\mathbb{P}^\mu(|\mu_T(f) - \mu(f)| \leqslant \varepsilon) \geqslant 1 - \delta$ holds for $T \geqslant \Psi(\varepsilon, \delta)$

Discrete-time concentration result

Let observations $(X_{k\Delta})_{k=1,\dots,n}$ be given for some $\Delta \leqslant 1$. Discrete MC-estimator:

$$\mathbb{H}_n^{\Delta}(f) \coloneqq \frac{1}{n\Delta} \sum_{k=1}^{n} f(X_{k\Delta})\Delta.$$

Then for $\mathbb{H}_t(f) \coloneqq t^{-1} \int_0^T f(X_t)\,\mathrm{d}t$, $f = \widetilde{f} - \mu(\widetilde{f})$, $\Phi_k(t) \coloneqq \int_t^{k\Delta} (L\widetilde{f}(X_s) - \mu(L\widetilde{f}))\,\mathrm{d}s$, $\omega_k(t) \coloneqq \int_t^{k\Delta} \nabla \widetilde{f}(X_s)^{\top} \sigma(X_s)\,\mathrm{d}W_s$,

$$n\Delta(\mathbb{H}_n^{\Delta}(f) - \mathbb{H}_{n\Delta}(f)) = \mu(L\widetilde{f})\frac{n\Delta^2}{2} + \sum_{k=1}^{n} \int_{(k-1)\Delta}^{k\Delta} \Phi_k(t)\,\mathrm{d}t + \sum_{k=1}^{n} \int_{(k-1)\Delta}^{k\Delta} \omega_k(t)\,\mathrm{d}t.$$

Let observations $(X_{k\Delta})_{k=1,\dots,n}$ be given for some $\Delta \leqslant 1$. Define

$$\mathbb{H}_n^{\Delta}(f) := \frac{1}{n\Delta}\sum_{k=1}^{n} f(X_{k\Delta})\Delta.$$

Then for $\mathbb{H}_t(f) := t^{-1}\int_0^T f(X_t)\,\mathrm{d}t$, $f = \widetilde{f} - \mu(\widetilde{f})$, $\Phi_k(t) := \int_t^{k\Delta}(L\widetilde{f}(X_s) - \mu(L\widetilde{f}))\,\mathrm{d}s$, $\omega_k(t) := \int_t^{k\Delta}\nabla\widetilde{f}(X_s)^\top\sigma(X_s)\,\mathrm{d}W_s$,

$$n\Delta(\mathbb{H}_n^{\Delta}(f) - \mathbb{H}_{n\Delta}(f)) = \mu(L\widetilde{f})\frac{n\Delta^2}{2} + \sum_{k=1}^{n}\int_{(k-1)\Delta}^{k\Delta}\Phi_k(t)\,\mathrm{d}t + \sum_{k=1}^{n}\int_{(k-1)\Delta}^{k\Delta}\omega_k(t)\,\mathrm{d}t.$$

Theorem [TAWS23]

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 $p \geqslant 2$,

$$\|\mathbb{H}_n^{\Delta}(f) - \mu(f)\|_{L^p(\mathbb{P}^{\mu})} \leqslant \mathfrak{D}\left(\Delta + \sqrt{\frac{\Delta}{n}}\,p^{\frac{\max\{(\widetilde{\gamma}+2\alpha+1-q_+)/2,\,\eta_1+1-q_+\}}{1-q_+}} + \frac{1}{\sqrt{n\Delta}}p^{\frac{1}{2}+\frac{\eta+\kappa+1+q}{1-q_+}}\right) := \Phi(n, \Delta, p),$$

and

$$\mathbb{P}^{\mu}\left(|\mathbb{H}_n^{\Delta}(f) - \mu(f)| > \mathrm{e}\Phi(n, \Delta, x)\right) \leqslant \mathrm{e}^{-x}, \quad x \geqslant 2.$$

Application

MCMC for moderately heavy tailed targets

- Langevin diffusion
$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t,$$
has invariant density $\pi(x) \propto \exp(-U(x)) \rightsquigarrow$ sampling from π by numerical approximation of X, e.g., Euler scheme
$$\vartheta_{n+1}^{(\Delta)} = \vartheta_n^{(\Delta)} - \Delta\nabla U(\vartheta_n^{(\Delta)}) + \sqrt{2\Delta}\xi_{n+1}, \quad \vartheta_0^{(\Delta)} \sim X_0, \quad (\xi_n) \underset{\text{iid}}{\sim} \mathcal{N}(0, \mathbb{I}_d)$$
- abundant literature on sampling precision in TV or Wasserstein distance for U strongly convex or modifications thereof [Dal17; DK19; DM17; DMM19] $\rightsquigarrow \pi(x)\,\mathrm{d}x$ sub-Gaussian

MCMC for moderately heavy tailed targets

- Langevin diffusion

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

has invariant density $\pi(x) \propto \exp(-U(x)) \rightsquigarrow$ sampling from π by numerical approximation of X, e.g., Euler scheme

$$\vartheta_{n+1}^{(\Delta)} = \vartheta_n^{(\Delta)} - \Delta \nabla U(\vartheta_n^{(\Delta)}) + \sqrt{2\Delta}\xi_{n+1}, \quad \vartheta_0^{(\Delta)} \sim X_0, \quad (\xi_n) \underset{\text{iid}}{\sim} \mathcal{N}(0, \mathbb{I}_d)$$

- abundant literature on sampling precision in TV or Wasserstein distance for U strongly convex or modifications thereof [Dal17; DK19; DM17; DMM19] $\rightsquigarrow \pi(x)\,\mathrm{d}x$ sub-Gaussian
- Assume instead that for some $q \in (0, 1)$

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

$\rightsquigarrow \exists \lambda > 0 : \int_{\mathbb{R}^d} \exp\left(\lambda \|x\|^{\widetilde{q}}\right)\pi(x)\,\mathrm{d}x < \infty \iff \widetilde{q} \leqslant 1 - q$

$\rightsquigarrow$ prototypical example: $\pi(x) \propto \exp(-\beta\|x\|^{1-q})$ outside some ball around the origin

Proposition [TAWS23]

Assume $(\mathcal{U}(q))$ and that ∇U is bounded. Let $f \in C^2(\mathbb{R}^d)$ s.t. $\|D^k f(x)\| \lesssim 1 + \|x\|^{\eta_k}$, $k = 0, 1, 2$, and consider the burn-in estimator

$$\mathbb{H}_{n,m,\Delta}(f) \coloneqq \mathbb{H}_{n,\Delta}(f) \circ \theta_m = \frac{1}{n} \sum_{k=m+1}^{n+m} f(X_{k\Delta}).$$

Then we have the following approximation guarantees:

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

Table 2: Order of sufficient sample size n and burn-in m for (ε, δ)-PAC bounds and sampling within ε-TV margin

Lasso for parametrized drifts

For a given dictionary $\{\psi_1, \ldots, \psi_N\}$ of Lipschitz functions $\psi_i \colon \mathbb{R}^d \to \mathbb{R}^d$, let X be the strong solution to

$$\mathrm{d}X_t = b_{\theta^0}(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad \text{where} \quad b_{\theta^0}(x) = \sum_{i=1}^N \theta_i^0 \psi_i(x).$$

Let $\boldsymbol{\psi}(x) = (\psi_1(x), \ldots, \psi_N(x))$, $\boldsymbol{\Psi}(x) \coloneqq (\sigma^{-1}(x)\boldsymbol{\psi}(x))^\top \sigma^{-1}(x)\boldsymbol{\psi}(x)$ and $\overline{\boldsymbol{\Psi}}_T \coloneqq T^{-1}\int_0^T \boldsymbol{\Psi}(X_t)\,\mathrm{d}t$.
Then for $b_\theta \coloneqq \boldsymbol{\psi}\theta$, negative log-likelihood given by

$$\mathcal{L}_T(\theta) = \mathcal{L}_T(b_\theta) = \theta^\top \overline{\boldsymbol{\Psi}}_T \theta - 2\theta^T \frac{1}{T}\int_0^T \boldsymbol{\psi}(X_t)^\top a^{-1}(X_t)\,\mathrm{d}X_t.$$

Lasso for parametrized drifts

For a given dictionary $\{\psi_1, \dots, \psi_N\}$ of Lipschitz functions $\psi_i \colon \mathbb{R}^d \to \mathbb{R}^d$, let X be the strong solution to

$$\mathrm{d}X_t = b_{\theta^0}(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \quad \text{where} \quad b_{\theta^0}(x) = \sum_{i=1}^N \theta_i^0 \psi_i(x).$$

Let $\boldsymbol{\psi}(x) = (\psi_1(x), \dots, \psi_N(x))$, $\boldsymbol{\Psi}(x) \coloneqq (\sigma^{-1}(x)\boldsymbol{\psi}(x))^\top \sigma^{-1}(x)\boldsymbol{\psi}(x)$ and $\overline{\boldsymbol{\Psi}}_T \coloneqq T^{-1}\int_0^T \boldsymbol{\Psi}(X_t)\,\mathrm{d}t$.
Then for $b_\theta \coloneqq \boldsymbol{\psi}\theta$, negative log-likelihood given by

$$\mathcal{L}_T(\theta) = \mathcal{L}_T(b_\theta) = \theta^\top \overline{\boldsymbol{\Psi}}_T \theta - 2\theta^T \frac{1}{T}\int_0^T \boldsymbol{\psi}(X_t)^\top a^{-1}(X_t)\,\mathrm{d}X_t.$$

Goal

Study convergence guarantees of Lasso estimator

$$\widehat{\theta}_T \coloneqq \operatorname*{arg\,min}_{\theta \in \mathbb{R}^N} \left\{ \mathcal{L}_T(\theta) + \lambda \|\theta\|_1 \right\},$$

under sparsity assumptions on θ^0, i.e., $\|\theta^0\|_0 \leqslant s_0$.

Assumptions and examples

We assume

1. $\exists A, \mathfrak{r} > 0, q \in [-1, 1) : \quad \langle b_{\theta^0}(x), x/\|x\| \rangle \leqslant -\mathfrak{r}\|x\|^{-q}, \quad \|x\| > A;$
2. $\lambda_{\max}(\mathbf{\Psi}(x)) \lesssim 1 + \|x\|^{2\eta};$
3. $\overline{\mathbf{\Psi}}_T$ is positive definite $\mathbb{P}_{\theta^0}$-a.s.

Example 1: [GM19; CMP20]

Ornstein–Uhlenbeck process: $N = d^2$,

$$b_{\theta^0}(x) = A_{\theta^0} x.$$

If A_{θ^0} is symmetric, negative definite $\rightsquigarrow q = -1, \eta = 1$.

Example 2:

$N = 2d^2$,

$$b_{\theta^0}(x) = A_{\theta^0} x + B_{\theta^0} x(\alpha + \|x\|)^{-(1+\widetilde{q})}.$$

If A_{θ^0} is singular and negative semi-definite and B_{θ^0} is negative definite $\rightsquigarrow q = \widetilde{q}, \eta = 1$

Restricted eigenvalue property

- Proof of high probability bounds relies on having good control over the spectrum of the empirical Gram matrix $\overline{\boldsymbol{\Psi}}_T = \frac{1}{T}\int_0^T \boldsymbol{\Psi}(X_t)\,\mathrm{d}t$

Restricted eigenvalue property

- Proof of high probability bounds relies on having good control over the spectrum of the empirical Gram matrix $\overline{\boldsymbol{\Psi}}_T = \frac{1}{T}\int_0^T \boldsymbol{\Psi}(X_t)\,\mathrm{d}t$

$\rightsquigarrow$ control

$$\inf_{\theta\in\mathcal{S}} \theta^\top \overline{\boldsymbol{\Psi}}_T \theta = \inf_{\theta\in\mathcal{S}} \frac{1}{T}\int_0^T \|\sigma^{-1}(X_t) b_\theta(X_t)\|^2\,\mathrm{d}t,$$

for appropriate $\mathcal{S} \subset \mathbb{R}^N$ in terms of $\lambda_{\min}(\mathbb{E}[\overline{\boldsymbol{\Psi}}_T]) =: \lambda_{\min}^\infty$ via concentration inequality for (unbounded) b_θ and covering arguments

Restricted eigenvalue property

- Proof of high probability bounds relies on having good control over the spectrum of the empirical Gram matrix $\overline{\boldsymbol{\Psi}}_T = \frac{1}{T}\int_0^T \boldsymbol{\Psi}(X_t)\,\mathrm{d}t$

$\rightsquigarrow$ control

$$\inf_{\theta\in\mathcal{S}} \theta^\top \overline{\boldsymbol{\Psi}}_T \theta = \inf_{\theta\in\mathcal{S}} \frac{1}{T}\int_0^T \|\sigma^{-1}(X_t) b_\theta(X_t)\|^2\,\mathrm{d}t,$$

for appropriate $\mathcal{S} \subset \mathbb{R}^N$ in terms of $\lambda_{\min}(\mathbb{E}[\overline{\boldsymbol{\Psi}}_T]) =: \lambda_{\min}^\infty$ via concentration inequality for (unbounded) b_θ and covering arguments

- for some sparsity dependent $\mathcal{S}(s)$, we obtain

$$\mathbb{P}\Big(\inf_{\theta\in\mathcal{S}(s)} \theta^\top \overline{\boldsymbol{\Psi}}_T \theta \geqslant \frac{\lambda_{\min}^\infty}{2}\Big) \geqslant 1-\varepsilon,$$

for

$$T \geqslant T_0(\varepsilon, s, d, q, \eta) \sim \Big\{\log\Big(21^{2s}\Big(d \wedge \big(\tfrac{ed}{2s}\big)^{2s}\Big)\Big) + \log(1/\varepsilon)\Big\}^{\frac{6\eta+2q+3-q_+}{1-q_+}} \cdot \frac{1}{(\lambda_{\min}^\infty)^2}.$$

High probability bound

Theorem [TAWS23]

Suppose $\|\theta^0\|_0 \leqslant s_0$ and fix $\varepsilon \in (0,1)$. If $T \geqslant T_0(\varepsilon/3, s_0, d, q, \eta)$, then for the choice $\lambda \asymp \sqrt{\log(N/\varepsilon)/T}$ with probability at least $1-\varepsilon$,

$$\|\widehat{\theta}_T - \theta_0\|_{L^2}^2 \coloneqq (\widehat{\theta}_T - \theta_0)^\top \overline{\boldsymbol{\Psi}}_T (\widehat{\theta}_T - \theta_0) \lesssim \frac{\log(N/\varepsilon)s_0}{T}.$$

Summary

- we provide concentration inequalities for subexponentially ergodic diffusions and polynomially bounded functions given continuous observations
- Concentration inequalities for sampled chains are derived from the continuous observation result
- we demonstrate implications on sufficient sample sizes for MCMC for moderately heavy tailed targets as well as sparse estimation of parametrized diffusion models

Paper available as

"Trottner, L., Aeckerle-Willems, C., and C. Strauch. Concentration analysis of multivariate elliptic diffusions. *Journal of Machine Learning Research* 24 (2023), paper no. 106, pp. 1–38."

Summary

- we provide concentration inequalities for subexponentially ergodic diffusions and polynomially bounded functions given continuous observations
- Concentration inequalities for sampled chains are derived from the continuous observation result
- we demonstrate implications on sufficient sample sizes for MCMC for moderately heavy tailed targets as well as sparse estimation of parametrized diffusion models

Paper available as
"Trottner, L., Aeckerle-Willems, C., and C. Strauch. Concentration analysis of multivariate elliptic diffusions. *Journal of Machine Learning Research* 24 (2023), paper no. 106, pp. 1–38."

Thank you for your attention!