

Error estimates for Euler schemes and Randomized Euler schemes beyond Lipschitz continuity

Iosif Lytras

The University of Edinburgh

Joint work with A.Ntoussis



Stochastic Differential Equations

Consider the SDE

$$dX_t = f(t, X_t) dt + g(t, X_t) dW_t, \quad t \in [0, T].$$

where

$$f : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad g : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}.$$

Goal:

- Existence and uniqueness of solutions
- Moment estimates
- Numerical approximation schemes

Linear Growth

$$|f(t, x)|^2 \vee |g(t, x)|^2 \leq K(1 + |x|^2).$$

Global Lipschitz

$$|f(t, x) - f(t, y)|^2 \vee |g(t, x) - g(t, y)|^2 \leq K|x - y|^2.$$

Main Result

- Unique strong solution exists on $[0, T]$.

Moment Bounds

If $p \geq 2$ and $\mathbb{E}|\xi|^p < \infty$, then

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} |X_t|^p \right] \leq (1 + 3^{p-1} \mathbb{E}|\xi|^p) e^{cT}.$$

Consequences:

- Finite moments of all orders.
- Stability of the solution.
- Foundation for convergence analysis.

Monotonicity Condition

$$\langle x, f(t, x) \rangle + \frac{1}{2}|g(t, x)|^2 \leq K(1 + |x|^2).$$

Local Lipschitz Condition

For every $R > 0$:

$$|f(t, x) - f(t, y)|^2 + |g(t, x) - g(t, y)|^2 \leq L_R |x - y|^2.$$

for $|x|, |y| \leq R$.

Existence Beyond Lipschitz

Under monotonicity and local Lipschitz assumptions:

- A unique solution exists.
- Explosion is avoided.
- Finite moments are preserved.

Moreover,

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} |X_t|^p \right] \leq C.$$

For

$$dY_t = b(Y_t) dt + \sigma(Y_t) dW_t,$$

the Euler-Maruyama scheme is

$$X_{k+1} = X_k + b(X_k) \Delta t + \sigma(X_k) \Delta W_k,$$

where

$$\Delta W_k \sim \mathcal{N}(0, \Delta t).$$

Euler Convergence

Under linear growth:

$$\sup_{0 \leq t \leq T} \mathbb{E} |Y_t^n|^2 \leq C.$$

Under linear growth and Lipschitz assumptions:

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} |X_t - Y_t|^2 \right] \leq C\lambda.$$

Strong convergence order:

$$\mathcal{O}(\lambda^{1/2}).$$

For constant diffusion this increases to rate 1 (Euler coincides with Milstein scheme)

$$dZ_t = -h(Z_t)dt + \sqrt{\beta^{-1}}dB_t$$

Unadjusted Langevin Algorithm (ULA)

Definition (ULA)

$$\theta_0^\lambda := \theta_0, \quad \theta_{n+1}^\lambda := \theta_n^\lambda - \lambda h(\theta_n^\lambda) + \sqrt{\frac{2\lambda}{\beta}} \xi_{n+1}, \quad n \in \mathbb{N}$$

where $h := \nabla u$ and ξ_n is a sequence of d -dimensional Gaussian random variables.

Simple negative example for superlinear drift

Setting $h(x) = x^3/3$ and applying vanilla ULA with step-size λ and initial condition $\mathcal{L}(X_0) = \mathcal{N}(0, \frac{4}{\lambda})$, we get

$$X_{n+1} = X_n - \lambda X_n^2 + \sqrt{2\lambda} \xi_{n+1}$$

After some calculations,

$$\mathbb{E}X_{n+1}^2 \geq \mathbb{E}X_0^2 + 2\lambda n. \quad (1)$$

Moment diverges as $n \rightarrow \infty$ even in simple example!

Divergence of Euler scheme

For superlinear drift, the Euler scheme diverges in the strong L^p sense, i.e.,

$$\lim_{\lambda \rightarrow 0} \mathbb{E} \left[\left| Y_T - Y_T^{\lambda, \text{EM}} \right|^p \right] = \infty,$$

$\forall p \geq 1$, see (Hutzenthaler et al., 2011, Theorem 2.1). The theorem also implies

$$\lim_{\lambda \rightarrow 0} \mathbb{E} \left[\left| Y_T^{\lambda, \text{EM}} \right|^p \right] = \infty.$$

We need to make the coefficients have linear growth!!

Convergence of taming schemes

- To approximate SDEs with non-globally Lipschitz continuous drift coefficient, Hutzenthaler et al. (2012) proposed a tamed Euler scheme, and obtained a convergence result of the tamed scheme in the strong L^p -sense with the supremum over the time interval $[0, T]$.
- Sabanis (2013) extends the tamed scheme to approximate (time-inhomogeneous) SDEs with locally one-sided Lipschitz continuous drift coefficient.

$$\bar{\theta}_0^\lambda := \theta_0, \quad \bar{\theta}_{n+1}^\lambda := \bar{\theta}_n^\lambda - \lambda h_\lambda \left(\bar{\theta}_n^\lambda \right) + \sqrt{2\lambda\beta^{-1}}\xi_{n+1}, \quad n \in \mathbb{N}_0, \quad (2)$$

We use a modified drift h_λ

- h_λ should have at most linear growth to provide stability
- $h_\lambda \rightarrow h$ as $\lambda \rightarrow 0$ to provide a good approximation to the SDE

$$\pi_\beta(A) = \int_A e^{-\beta u(\theta)} d\theta / \int_{\mathbb{R}^d} e^{-\beta u(\theta)} d\theta, \quad A \in \mathcal{B}(\mathbb{R}^d)$$

Under suitable assumptions on ∇u , π_β is the unique invariant distribution of the **Langevin** SDE:

$$dZ_t = -\nabla u(Z_t) dt + \sqrt{2\beta^{-1}} dB_t, \quad t > 0 \quad (3)$$

Introduction to Langevin-based sampling

Idea: Use an Euler-Maruyama (EM) scheme with starting (possibly random) value L_0 :

$$L_{k+1}^\gamma = L_k^\gamma - \gamma_{k+1} h(L_k^\gamma) + \sqrt{2\beta^{-1}\gamma_{k+1}} \xi_{k+1}$$

,

- $\{\xi_k\}_{k \geq 1}$ is i.i.d. $N(0, I_d)$.
- $\{\gamma_k\}_{k \geq 1}$ is a sequence of stepsizes; either constant (fixed gain) or decreasing to 0 at a certain rate.

Unadjusted Langevin Algorithm (ULA)

Definition (ULA)

$$\theta_0^\lambda := \theta_0, \quad \theta_{n+1}^\lambda := \theta_n^\lambda - \lambda h(\theta_n^\lambda) + \sqrt{\frac{2\lambda}{\beta}} \xi_{n+1}, \quad n \in \mathbb{N}$$

where $h := \nabla u$ and ξ_n is a sequence of d -dimensional Gaussian random variables.

Many important non-asymptotic convergence results for ULA under:

Assumption (Strong convexity)

$$\langle \nabla u(x) - \nabla u(y), x - y \rangle \geq a|x - y|^2 \quad \forall x, y \in \mathbb{R}^d, \quad a > 0.$$

Assumption (Global Lipschitzness)

$$|\nabla u(x) - \nabla u(y)| \leq L|x - y| \quad \forall x, y \in \mathbb{R}^d$$

Can we go non-convex and Local Lipschitz and obtain results in KL, TV and W_2 ?

Assumptions on the drift

- **Polynomial Jacobian growth.** There exist constants $L > 0$ and $\ell \geq 0$ such that

$$\max\{\|h(x)\|, \|J(h)(x)\|\} \leq L(1 + \|x\|^{\ell+1}), \quad \forall x \in \mathbb{R}^d. \quad (\text{PJG})$$

- **Polynomial Lipschitz continuity.** There exist constants $L' > 0$ and $\ell' \geq 0$ such that

$$\|h(x) - h(y)\| \leq L'(1 + \|x\| + \|y\|)^{\ell'} \|x - y\|, \quad \forall x, y \in \mathbb{R}^d. \quad (\text{PLC})$$

- **One-sided Lipschitz continuity.** There exists a constant $K' \in \mathbb{R}$ such that

$$\langle h(x) - h(y), x - y \rangle \geq -K' \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^d. \quad (\text{OSL})$$

- **Dissipativity.** There exist constants $a, b > 0$ such that

$$\langle h(x), x \rangle \geq a\|x\|^2 - b, \quad \forall x \in \mathbb{R}^d. \quad (\text{D})$$

We propose the splitted tamed unadjusted Langevin algorithm (**sTULA**)

$$\bar{\theta}_0^\lambda := \theta_0, \quad \bar{\theta}_{n+1}^\lambda := \bar{\theta}_n^\lambda - \lambda h_\lambda \left(\bar{\theta}_n^\lambda \right) + \sqrt{2\lambda\beta^{-1}}\xi_{n+1}, \quad n \in \mathbb{N}_0, \quad (4)$$

where $\lambda > 0$ is the step size, $(\xi_n)_{n \in \mathbb{N}}$ is a sequence of independent standard d -dimensional random variables independent of the $\mathbb{R}^d$ -valued random variable θ_0 , and where for all $\theta \in \mathbb{R}^d$, the tamed coefficient of our iterative scheme shall be given as

$$h_\lambda(x) := ax + \frac{h(x) - ax}{\sqrt{1 + g(\lambda\|x\|^{2(\ell+1)})}}, \quad (5)$$

where

$$g(t) = \begin{cases} 0, & 0 \leq t < 1, \\ 5(t-1)^2 - 3(t-1)^3, & 1 \leq t < 2, \\ t, & t \geq 2. \end{cases}$$

Tamed Sampling Literature

-  Brosse, Nicolas, et al. "The tamed unadjusted Langevin algorithm." *Stochastic Processes and their Applications* 129.10 (2019): 3638-3663.
-  Lovas, A., Lytras, I., Rásonyi, M., Sabanis, S. (2023). Taming neural networks with tusla: Nonconvex learning via adaptive stochastic gradient langevin algorithms. *SIAM Journal on Mathematics of Data Science*, 5(2), 323-345.
-  Neufeld, A., En, M. N. C., Zhang, Y. (2022). Non-asymptotic convergence bounds for modified tamed unadjusted Langevin algorithm in non-convex setting. *arXiv preprint arXiv:2207.02600*.
-  Lytras, Iosif, and Sotirios Sabanis. "Taming under isoperimetry." *Stochastic Processes and their Applications* (2025): 104684.
-  Lytras, Iosif, Sotirios Sabanis, and Ying Zhang. "kTULA: A Langevin sampling algorithm with improved KL bounds under super-linear log-gradients." *arXiv preprint arXiv:2506.04878* (2025).

Goal

Best rate achieved in KL divergence is $\tilde{O}(\epsilon^{-\frac{3}{4}})$.

Goal: Achieve optimal rate $\tilde{O}(\epsilon^{-\frac{1}{2}})$

Basic Properties of the tamed scheme

- There exist c_1, c_2 such that

$$|h_\lambda(x)| \leq c_1|x| + \frac{c_2}{\sqrt{\lambda}}$$

- There exists $p > 0$ such that

$$|h_\lambda(x) - h(x)|^2 \leq \lambda^2|x|^p$$

- h_λ is $\mathcal{O}\left(\frac{1}{\sqrt{\lambda}}\right)$ -Lipschitz

Tamed scheme inherits dissipativity condition

Lemma

Let the 2-dissipativity condition hold for h . Then, $\langle h_\lambda(x), x \rangle \geq \frac{a}{2}|x|^2 - b$.

Moment bounds

For $\lambda < \min\{1, \frac{1}{4(2a+4L)^2}\}$, there holds for any $p \geq 1$

$$\sup_n \mathbb{E}|\hat{X}_n|^{2p} \leq C_p$$

where C_p has at most polynomial dependence on the dimension.

Key properties of the Langevin SDE

Let X, Y two Langevin SDEs started at $x, y \in \mathbb{R}^d$ with the same driving Brownian motion. For small time $\lambda > 0$, there holds

1

$$\|X_\lambda - Y_\lambda\|_{L^2} \leq e^{K\lambda} \|x - y\|$$

2

$$\|X_\lambda - x - (Y_\lambda - y)\|_{L^2} \leq \lambda(1 + \|x\| + \|y\|)^{2l} \|x - y\|$$

3

$$KL(\mathcal{L}(X_\lambda) \parallel \mathcal{L}(Y_\lambda)) \leq C \frac{\|x - y\|^2}{\lambda} \quad (\text{regularity})$$

KL local framework- properties of the algorithm



$$\sup_n \mathbb{E}|X_n|^p < \infty \quad \forall p$$



$$KL(\delta_x \hat{P} || \delta_y P) \leq C \frac{|x - y|^2}{\lambda} + b(x) \quad (\text{cross-regularity})$$

where $b(x) = C(x)\lambda^2$

Construction

Let $(\widehat{X}_n)_{n \geq 0}$, $(Y_n)_{n \geq 0}$ denote the chain of the algorithm and the chain of one-step SDE updates with $\widehat{X}_0 \sim \mu$ and $Y_0 \sim \nu$. Choose shift coefficients $(\eta_n)_{n=0}^{N-1} \subset [0, 1]$ with $\eta_{N-1} = 1$, and define

$$\widetilde{Y}_n := (1 - \eta_n) Y'_n + \eta_n \widehat{X}_n, \quad n = 0, 1, \dots, N-1. \quad (6)$$

Define

$$Q_n := \begin{cases} P, & 0 \leq n \leq N-2, \\ \widehat{P}, & n = N-1, \end{cases} \quad (7)$$

and update the auxiliary process by

$$Y'_{n+1} \sim Q_n(\widetilde{Y}_n, \cdot), \quad n = 0, 1, \dots, N-1, \quad Y'_0 = Y_0. \quad (8)$$

We write $\nu'_n := (Y'_n)$.

The terminal condition $\eta_{N-1} = 1$ together with the choice $Q_{N-1} = \widehat{P}$ guarantees the interpolation property

$$\nu'_0 = \nu_0, \quad \nu'_N = \widehat{\mu}_N, \quad (9)$$

$$KL(\hat{\mu}_N \parallel \nu_N) \leq \sum_{n=0}^{N-1} \mathbb{E} \left[KL(Q_n(\tilde{Y}_n, \cdot) \parallel P(Y'_n, \cdot)) \right]. \quad (10)$$

$$KL(\hat{\mu}_N \parallel \nu_N) \leq c \sum_{n=0}^{N-2} \eta_n^2 \mathbb{E} \|\hat{X}_n - Y'_n\|^2 + c' \mathbb{E} \|\hat{X}_{N-1} - Y'_{N-1}\|^2 + \bar{b}^2, \quad (11)$$

where $\bar{b}^2 := \|b\|_{L^2(\hat{\mu}_{N-1})}^2$ and $c, c' \leq \mathcal{O}(\lambda^{-1})$.

Strong and weak errors

Let $\hat{P}$ the one-step of an algorithm with stepsize kernel and P the Markov kernel of the SDE at time λ . Let mutually defined random variables $X, \hat{X}$ $\mathcal{L}(\hat{X}) = \delta_x \hat{P}$ and $\mathcal{L}(X) = \delta_x P$. We define

$$E_{\text{strong}}(x) := \sqrt{\mathbb{E}|X - \hat{X}|^2}$$

and

$$E_{\text{weak}}(x) := ||\mathbb{E}(X) - \mathbb{E}(\hat{X})||$$

$$a_0(x) := E_{\text{strong}}(x), \quad a_1(x, y) := E_{\text{weak}}(x) + \lambda(1 + |x|'' + |y|'') E_{\text{strong}}(x). \quad (12)$$

Lemma (Shifted Wasserstein recursion)

The right-hand side of (11) is controlled entirely by the squared coupling distances

$$d_n^2 := \mathbb{E} \|\hat{X}_n - Y'_n\|^2, \quad n = 0, 1, \dots, N-1, \quad (13)$$

which propagate forward in time according to the next lemma. For all $0 \leq n \leq N-2$,

$$d_{n+1}^2 \leq L^2(1 - \eta_n)^2 d_n^2 + 2(1 - \eta_n) \bar{a}_1 d_n + \bar{a}_0^2, \quad (14)$$

where $\bar{a}_0 := \sup_{0 \leq k < N} \|a_0(\hat{X}_k)\|_{L^2}$ and $\bar{a}_1 := \sup_{0 \leq k < N} \|a_1(\hat{X}_k, \tilde{Y}_k)\|_{L^2}$.

Theorem (Lytras and Ntoussis (2026))

Let the previous assumptions on the algorithm and the SDE hold. Then,

$$\begin{aligned} \text{KL} \left(\mu \hat{P}^N \| \nu P^N \right) &\lesssim \frac{1}{\lambda} \frac{e^{-K\lambda} - 1}{e^{-K\lambda N} - 1} W_2^2(\mu, \nu) \\ &+ \frac{1}{\lambda} \left[(\lambda N \vee \log \bar{N}) \bar{\mathcal{E}}_{strong}^2 + \bar{N} (\bar{E}_{weak} + \lambda C_m \bar{E}_{strong})^2 \right] \\ &+ \bar{b}^2 \end{aligned}$$

$$\bar{N} := N \wedge \frac{1}{\lambda}, \quad \bar{b} := \max_{0 \leq k < N} \mathbb{E} |\hat{X}_k|^{2l+2}, \quad C_m := \max_{0 \leq k < N} \sqrt{\mathbb{E} |\hat{X}_k|^{2l'}}$$

Goal: Recall that $\bar{b} \leq \mathcal{O}(\lambda^2)$. Also $\bar{E}_{strong} \leq \mathcal{O}(\lambda^{\frac{3}{2}})$. We need weak error to be of order $\mathcal{O}(\lambda^2)$

Key additional assumption to increase weak error

Assume that

$$\|Dh(x) - Dh(y)\|_2 \leq L(1 + |x| + |y|)' |x - y| \quad \forall x, y \in \mathbb{R}^d$$

Then,

$$E_{weak} \leq C_w \lambda^2$$

Theorem

Then there exist constants $C_1, C > 0$, depending at most polynomially on the dimension and independent of N and λ ,

$$KL(\mu \hat{P}^N \parallel \nu P^N) \leq C_1 \frac{1}{\lambda} W_2^2(\mu, \nu) + C ((N\lambda) \vee \log N) \lambda^2.$$

Definition (LSI)

A Gibbs probability measure π_β satisfies the logarithmic Sobolev inequality with constant $\alpha > 0$, denoted $\text{LSI}(\alpha)$, if for all probability measures ν such that $\nu \ll \pi_\beta$

$$(\text{LSI}(\alpha)) \quad KL(\nu|\pi_\beta) \leq \frac{1}{2\alpha} \frac{1}{\beta} \int_{\mathbb{R}^d} \frac{|\nabla f|^2}{f} d\pi_\beta := \frac{1}{2\alpha} FI(\nu|\pi_\beta),$$

LSI-Optimization in the space of measures

In the Riemannian structure induced by the W_2 metric for the functional $E(\mu) = H_\pi(\mu)$ there holds $|\text{grad } E|_\mu|^2 = I_\pi(\mu)$

$$\underbrace{E(\mu) \leq c |\text{grad } E|_\mu|^2}_{\text{P-L condition}} \implies \underbrace{H_\pi(\mu) \leq c I_\pi(\mu)}_{\text{LSI}}$$

Examples of LSI

Theorem (Bakry-Emery theorem)

$$\text{Hess}(u) \geq m \implies \text{LSI}(m)$$

Theorem (Hooley-Strook perturbation theorem)

Let $f : |u - f|_\infty \leq \xi$ and $\nu = e^{-f}$.

$$\nu \text{ satisfies } \text{LSI}(a) \implies \pi \text{ satisfies } \text{LSI}(e^{2\xi} a)$$

Theorem (Cattiaux et al)

dissipativity + one-sided Lipschitzness + (Poincaré inequality) $\implies$ LSI

LSI implies contraction in Renyi

Under LSI, there holds

$$R_2(\mu P_t | \pi_\beta) \leq e^{-ct} R_2(\mu | \pi)$$

Sampling under LSI

Writing

$$KL(\mu^{\hat{P}^N}|\pi) \leq KL(\mu^{\hat{P}^N}|\mu^{P^N}) + 2R_2(\mu^{P^N}|\pi)$$

leads to

$$KL(\mu^{\hat{P}^N}|\pi) \leq C ((N\lambda) \vee \log N) \lambda^2 + 2e^{-c\lambda N} R_2(\mu|\pi)$$

- Due to LSI

$$W_2(\mu\hat{P}^N, \pi) \leq \sqrt{CLSIKL(\mu\hat{P}^N, \pi)}$$

- Due to Pinsker inequality

$$TV(\mu\hat{P}^N, \pi) \leq \sqrt{2KL(\mu\hat{P}^N, \pi)}$$

To reach ϵ accuracy in KL divergence, one needs $\tilde{O}(\epsilon^{-\frac{1}{2}})$ iterations.

To reach ϵ accuracy in TV, one needs $\tilde{O}(\epsilon^{-1})$ iterations.

To reach ϵ accuracy in W_2 , one needs $\tilde{O}(\epsilon^{-1})$ iterations.

Global optimization guarantee

Theorem

Let u^* the global minimum of u . Then,

$$\mathbb{E} \left[u \left(\bar{\theta}_n^\lambda \right) \right] - u_* \leq CW_2 \left(\mathcal{L} \left(\bar{\theta}_n^\lambda \right), \pi_\beta \right) + \mathcal{O} \left(\frac{\log \beta}{\beta} \right).$$

This is achieved by decomposing as

$$\mathbb{E} \left[u \left(\bar{\theta}_n^\lambda \right) \right] - u_* = \underbrace{\mathbb{E} \left[u \left(\bar{\theta}_n^\lambda \right) \right] - \mathbb{E}_{\pi_\beta} [u(x)]}_{\text{sampling problem}} + \underbrace{\mathbb{E}_{\pi_\beta} [u(x)] - u_*}_{\text{concentration around minimum}}$$

What if no Hessian Lipschitz

We need to find a method that has increased weak error. Randomized schemes have been used as an alternative to Euler schemes, to solve ODEs and SDEs to cater for loss of regularity.

Tamed Randomized scheme

We define the tamed randomized Langevin Monte Carlo algorithm as follows:

$$\begin{aligned}\bar{Y}_{n+1}^\tau &= \bar{Y}_n - \lambda h_\lambda(\bar{Y}_n) \tau_{n+1} + \sqrt{\frac{2}{\beta}} \Delta W_{n+1}^\tau, & \bar{Y}_0 &= x_0, \\ \bar{Y}_{n+1} &= \bar{Y}_n - \lambda h_\lambda(\bar{Y}_{n+1}^\tau) + \sqrt{\frac{2}{\beta}} \Delta W_{n+1}, & n &\in \mathbb{N}_0.\end{aligned}\tag{tRLMC}$$

Here, $(\tau_n)_{n \geq 1}$ is an i.i.d. sequence of random variables uniformly distributed on $(0, 1)$, independent of the Brownian motion W , and

$$\Delta W_{n+1}^\tau := W_{t_n + \tau_{n+1}\lambda} - W_{t_n}, \quad \Delta W_{n+1} := W_{t_{n+1}} - W_{t_n},$$

where $t_n = n\lambda$ for $n \in \mathbb{N}_0$.

Properties of the algorithm



$$\sup_n \mathbb{E} |\tilde{Y}_n|^p \leq C_p$$



$$E_{strong}(x) \leq \lambda^{\frac{3}{2}} C(x)$$



$$E_{weak}(x) \leq \lambda^2 C'(x)$$



$$TV(\delta_x \tilde{P}, \delta_y P) \leq \lambda b(x) + \frac{C}{\sqrt{\lambda}} |x - y|$$

Increased weak error without Hessian Lipschitzness!

Unfortunately, we could prove cross-regularity only in Total variation..

Theorem

$$TV(\mu \tilde{P}^N \parallel \nu P^N) \leq C_1 \frac{1}{\sqrt{\lambda}} W_2(\mu, \nu) + C\lambda \sqrt{((N\lambda) \vee \log N)}. \quad (15)$$

Total variation guarantee under LSI

Corollary

There holds

$$TV(\mu\tilde{P}^N, \pi_\beta) \leq C\lambda\sqrt{((N\lambda) \vee \log N)} + \sqrt{e^{-C_{LSI}\lambda N} KL(\mu \parallel \pi_\beta)}.$$

As a result, for $\lambda \leq \tilde{O}(\epsilon)$ one needs $N \geq \tilde{O}(\frac{1}{\epsilon})$ to guarantee

$$TV(\mu\tilde{P}^N, \pi_\beta) \leq \epsilon$$

Let $n, n_0 \in \mathbb{N}$. There holds

$$\begin{aligned} & W_2^2((\mu \tilde{P}^{n-n_0}) \tilde{P}, \mu \tilde{P}^{n-n_0} P^{n_0}) \\ & \leq C e^{3K\lambda n_0} \left(\frac{1}{\lambda^2} \left(\tilde{E}_w + \lambda \tilde{\Gamma} \tilde{E}_{strong} \right)^2 + \frac{1}{\lambda} \tilde{E}_{strong}^2 \right) \end{aligned}$$

Finite time bound

There holds

$$W_2(\mu \tilde{P}^{n-n_0} \tilde{P}^{n_0}, \mu \tilde{P}^{n-n_0} P^{n_0}) \leq C e^{c\lambda n_0} \lambda$$

LSI+one-sided Lipschitzness implies

$$W_2(\mu \hat{P}^{n-n_0} P^{n_0}, \pi_\beta) \leq C e^{-\dot{c} \lambda n_0} W_2(\mu \hat{P}^{n-n_0}, \pi_\beta)$$

$$W_2(\mu\tilde{P}^n, \pi_\beta) \leq Ce^{c\lambda n_0}\lambda + Ce^{-\dot{c}\lambda n_0}W_2(\mu\tilde{P}^{n-n_0}, \pi_\beta)$$

Picking the right n_0 one obtains

$$W_2(\mu\tilde{P}^n, \pi_\beta) \leq C^*\lambda + \frac{1}{e}W_2(\mu\tilde{P}^{n-n_0}, \pi_\beta)$$

Theorem

There holds

$$W_2(\mu\tilde{P}^n, \pi_\beta) \leq C(\lambda + e^{-c\lambda n})$$

As a result, for $\lambda \leq \tilde{O}(\epsilon)$ one needs $N \geq \tilde{O}(\frac{1}{\epsilon})$ to guarantee

$$W_2(\mu\tilde{P}^N, \pi_\beta) \leq \epsilon$$

Global optimization guarantee

Theorem

Let u^* the global minimum of u . Then,

$$\mathbb{E} \left[u \left(\bar{\theta}_n^\lambda \right) \right] - u_* \leq CW_2 \left(\mathcal{L} \left(\bar{\theta}_n^\lambda \right), \pi_\beta \right) + \mathcal{O} \left(\frac{\log \beta}{\beta} \right).$$

This is achieved by decomposing as

$$\mathbb{E} \left[u \left(\bar{\theta}_n^\lambda \right) \right] - u_* = \underbrace{\mathbb{E} \left[u \left(\bar{\theta}_n^\lambda \right) \right] - \mathbb{E}_{\pi_\beta} [u(x)]}_{\text{sampling problem}} + \underbrace{\mathbb{E}_{\pi_\beta} [u(x)] - u_*}_{\text{concentration around minimum}}$$

LSI may be large in β

Note that the convergence in the sampling problem relies on the LSI constant! In many cases (for example the double-well potential) this constant is exponential in β which cannot be avoided which slows the convergence with respect to the optimization problem.

Two ways to tackle this problem :

- Use a Polyak-Lojacevic condition and treat the algorithm using only optimization algorithm
- Find conditions where the LSI constant is polynomial in β .

Optimization result under P-L(no LSI assumed)

Theorem

Assume that u satisfies a Polyak–Łojasiewicz condition.

$$u(x) - u^* \leq C_{\text{PL}} \|\nabla u(x)\|^2 \quad \forall x \in \mathbb{R}^d. \quad (16)$$

Let x_n be the n – th iterate of adTULA or adTRLMC . Then,

$$\mathbb{E}u(x_{n+1}) - u^* \leq (1 - \lambda C_{\text{PL}}^{-1})^{n+1} (\mathbb{E}u(x_0) - u^*) + C_{\text{PL}} \left(C^{\text{opt}} \lambda + \frac{d}{\beta} \right),$$

where C^{opt} depends on the moment bounds of the algorithm and is specified in the proof.

Extra Conditions for LSI polynomial in β

- There exists $C' > 0$ such that,

$$\|\nabla^2 u(x)\| \leq C'(1 + |h(x)|) \quad \forall x \in \mathbb{R}^d$$

- u has a unique local minimum.
- There exists a constant $I^* > 0$ such that

$$I^* \leq \inf \{ |\bar{\lambda}_i(\nabla^2 u(y))| \mid h(y) = 0, i \in \{1, 2, \dots, d\} \}.$$

Denoting by S the set containing all saddle points and local maxima of u ,

$$\sup_{x \in S} \bar{\lambda}_{\min}(\nabla^2 u(x)) \leq -I^* < 0.$$

Theorem (Lytras and Sabanis, Stochastic Processes and Appl. 2025)

Under the above conditions, for β large enough,

$$(C_{LSI})^{-1} \leq \mathcal{O}(\beta^2)$$

Summary of contributions

- ① Finite time KL guarantees for tamed Euler and TV guarantees for Tamed Randomized for SDEs with Locally Lipschitz drift
- ② Optimal rates for sampling in KL, TV, W_2 for tamed Euler and in TV, W_2 for tamed Randomized
- ③ Optimization guarantees

From theory to practice: the boosting factor

Bridge to the optimizer

The practical optimizer (TheoPouLA-style, Lim–Sabanis) augments the coordinate-wise taming map with a **boosting factor** that scales small gradient coordinates up:

$$\mathbf{T}(h_i) = \left(1 + \frac{\sqrt{\lambda}}{\varepsilon + |h_i|}\right) \frac{h_i}{\sqrt{1 + g(\lambda |h_i|)}}.$$

- Left factor **boosts** small coordinates (helps vanishing gradients); right factor **tames** large ones (prevents gradient explosions).
- **Theory transfers.** The boosting factor is a bounded multiplicative perturbation: the growth, closeness, and dissipativity lemmas all go through with only constant changes.

Other directions of research

- Sampling using subgradient Langevin dynamics from non-smooth potentials e.g (Johnston et al 2025, TMLR), (Lytras, Makras, Sabanis 2026++)
- Sampling using tamed SGLD e.g (Lovas et al 2023)
- Sampling using tamed kinetic Langevin e.g (Lytras and Mertikopoulos 2025, Johnston et al 2024, Journal of Complexity)
- Tamed SGD (Apidopoulos, Lytras and Mertikopoulos 2026, ICML)