arXiv:submit/7627190 [math.PR] 22 May 2026

ERROR ESTIMATES FOR TAMED EULER AND RANDOMIZED EULER SCHEMES FOR SDES WITH LOCALLY LIPSCHITZ DRIFT WITH APPLICATIONS TO NON-LOGCONCAVE SAMPLING AND OPTIMIZATION

IOSIF LYTRAS[†] AND ANGELOS NTOUSIS[*]

ABSTRACT. In this paper, we study the numerical discretization of stochastic differential equations with locally Lipschitz, super-linearly growing drift, and the resulting implications for sampling from non-log-concave distributions satisfying a logarithmic Sobolev inequality. In this regime, the classical Euler–Maruyama scheme underlying the unadjusted Langevin algorithm (ULA) is known to be unstable. We analyze the KL-accelerated tamed unadjusted Langevin algorithm (kTULA) and introduce a new tamed randomized midpoint scheme, termed tRLMC. Building on the shifted-composition approach of [5], we develop two new local-error frameworks that yield finite-time, non-asymptotic error estimates against the underlying SDE—in KL divergence for kTULA, and in total variation for tRLMC —valid for general locally Lipschitz drift. Specializing these frameworks to the sampling problem under a logarithmic Sobolev inequality, we obtain a near-optimal $\widetilde{O}(\varepsilon^{-1/2})$ iteration complexity for kTULA in KL divergence, with corresponding guarantees in total variation and Wasserstein distance. We further establish, for the first time, a non-asymptotic guarantee in total variation for a tamed randomized Langevin scheme under super-linear drift growth, together with the corresponding Wasserstein-distance bound, both with $\widetilde{O}(\varepsilon^{-1})$ complexity for tRLMC. As a consequence, both schemes yield non-asymptotic bounds for a non-convex excess-risk optimization problem.

1. INTRODUCTION

The numerical discretization of stochastic differential equations (SDEs) with super-linearly growing drift has attracted considerable attention in the numerical analysis of stochastic dynamics. Beyond its intrinsic interest, this analysis underpins a wide range of sampling algorithms used in computational statistics, Bayesian inference, and large-scale optimization. The general SDE of interest in this work takes the form

$$dX_t = -h(X_t)\,dt + \sigma\,dB_t, \tag{SDE}$$

where $h : \mathbb{R}^d \to \mathbb{R}^d$ is a locally Lipschitz drift that may grow super-linearly at infinity, $\sigma > 0$ is a constant volatility, and $(B_t)_{t\geq 0}$ is a canonical Wiener process in $\mathbb{R}^d$ with unit covariance. A canonical application consists of sampling from a target distribution π on $\mathbb{R}^d$, typically expressed in Gibbs form as $\pi(x) \propto e^{-\beta u(x)}$ for some potential function $u : \mathbb{R}^d \to \mathbb{R}$. Under suitable assumptions on u, this distribution arises as the invariant measure of the Langevin SDE

$$dX_t = -\nabla u(X_t)\,dt + \sqrt{\tfrac{2}{\beta}}\,dB_t, \tag{LSDE}$$

which is precisely (SDE) specialized to $h := \nabla u$ and $\sigma := \sqrt{2/\beta}$. Motivated by this fundamental property of (LSDE), one of the most widely used algorithmic schemes for sampling from π is the *unadjusted Langevin algorithm* (ULA), given recursively by

$$\theta_0^{\mathrm{ULA}} := \theta_0, \qquad \theta_{n+1}^{\mathrm{ULA}} = \theta_n^{\mathrm{ULA}} - \lambda\, h\big(\theta_n^{\mathrm{ULA}}\big) + \sqrt{\tfrac{2\lambda}{\beta}}\,\xi_{n+1}, \qquad n \in \mathbb{N}_0, \tag{ULA}$$

where $\theta_n \in \mathbb{R}^d$, $n = 1, 2, \ldots$, denotes the state variable of the algorithm, $(\xi_n)_{n\geq 1}$ is an independent and identically distributed (i.i.d.) sequence of standard d-dimensional random vectors with unit covariance, $\lambda > 0$ is a step-size parameter, and $h := \nabla u$ denotes the gradient of the potential u. The guiding idea behind (ULA) is that it may be viewed as an Euler–Maruyama discretization of (LSDE), so that, for sufficiently large n and sufficiently small λ, θ_n^{ULA} is distributed approximately according to the invariant measure of (LSDE), which is precisely the target distribution π.

This connection has motivated a substantial literature devoted to establishing non-asymptotic convergence guarantees for (ULA) in various probability metrics, most notably Wasserstein distances, total variation, and information-theoretic divergences such as the Kullback–Leibler (KL) divergence. Much of this literature focuses on the case where π is log-concave and the gradient ∇u is globally Lipschitz,

corresponding respectively to convexity and smoothness of the potential u; see, e.g., [8, 9, 10] and references therein.

Beyond the globally Lipschitz log-concave setting, a substantial body of work has sought to relax convexity and smoothness assumptions on u, notably through dissipativity, convexity-at-infinity, and related structural conditions; see, for instance, [7, 24, 6, 27]. A complementary line of research, initiated by the insights of Vempala and Wibisono [31], establishes convergence under isoperimetric inequalities, such as Poincaré and logarithmic Sobolev inequalities; see [11, 12, 6, 25].

This work examines the numerical discretization of (SDE) in a regime that lies at the intersection of two well-studied directions: relaxing global Lipschitz assumptions on the drift, and—in the sampling specialization—weakening convexity via isoperimetric inequalities. While each of these settings has been analyzed extensively in isolation, rigorous non-asymptotic guarantees in their combination—particularly in the presence of super-linear drift growth—remain limited in the literature. Crucially, the numerical-analytic obstacles arising from super-linear drift are not specific to the gradient case $h = \nabla u$: explosion of moments, divergence of Euler–Maruyama in mean-square sense, and breakdown of standard local-error frameworks all occur at the level of (SDE), prompting a treatment that is general at the level of the SDE and only later specialized to sampling. This motivates the following two-fold question:

How does one derive finite-time error estimates for numerical discretizations of SDEs with super-linearly growing locally Lipschitz drift, and what guarantees can such estimates yield for sampling and optimization in the absence of log-concavity and global smoothness?

This scenario presents substantial challenges. As noted in several prior studies, both (ULA) and its stochastic variants, such as SGLD, may become unstable in such regimes. In particular, when h grows super-linearly, the associated Euler–Maruyama scheme—which forms the backbone of (ULA)—can fail dramatically. A pivotal result in [12] showed that, in such cases, the numerical approximation may diverge from the true SDE solution in mean-square sense, even over finite time horizons. This phenomenon is directly linked to the explosion of moments in the discretized process, highlighting why standard schemes may break down when applied to approximate (SDE) with super-linearly growing drift h.

Addressing this issue requires a fundamentally different approach—one that revisits the way in which (ULA) is designed as a numerical discretization of (LSDE), and leverages insights from the theory of numerical methods for SDEs. In this vein, a promising class of techniques—known as tamed Euler schemes—emerged in [13, 28, 29]. These methods modify the drift h in order to ensure the stability of (ULA) even under super-linear growth. Specifically, they replace the original coefficient h with a modified version h_λ, depending on the stepsize λ, which is constructed to satisfy two essential properties:

(P1) The tamed coefficient h_λ grows at most linearly, i.e., $h_\lambda(\theta) = O(\|\theta\|)$ as $\|\theta\| \to \infty$.
(P2) h_λ converges pointwise to the original coefficient h as $\lambda \to 0$.

These properties ensure both the stability of the numerical scheme and consistency with the original dynamics as the stepsize vanishes.

Tamed Euler algorithms. In the context of Langevin-based sampling, taming techniques have been applied under strong dissipativity or convexity assumptions [3, 15, 21], in stochastic-gradient settings [19, 17], under a "convexity at infinity" assumption [26], and more recently under functional inequalities [22, 20, 23]. In parallel, truncated or projected schemes have been examined under logarithmic Sobolev assumptions [34].

In this work, we build on the KL-accelerated tamed unadjusted Langevin algorithm (kTULA) introduced in [23], which can be viewed as a modification of TULA designed to ensure stability under super-linear drift growth. The key idea underlying our analysis is that taming induces state-dependent perturbations in the dynamics, which fall outside the scope of existing KL frameworks. By explicitly controlling these perturbations, we establish non-asymptotic finite-time error estimates for kTULA in KL divergence under polynomial smoothness and dissipativity of the drift; when specialized to the sampling setting $h = \nabla u$ under a logarithmic Sobolev inequality, this yields improved convergence rates toward π_β even in the presence of super-linear gradient growth.

Tamed randomized Euler algorithms. At the same time, an alternative line of work has focused on randomized discretization schemes for Langevin dynamics, where randomness is introduced directly at the level of the numerical integrator, for instance through randomized time steps or intermediate evaluations of the drift. Such methods have been shown to improve stability properties, reduce discretization bias, and, in certain regimes, achieve improved convergence rates compared to their deterministic counterparts.

A representative example, in the Langevin case $h = \nabla u$, is the randomized midpoint Langevin algorithm, given by

$$Y_{n+1}^{\tau} = Y_n - \tau_{n+1}\nabla u(Y_n)\,\lambda + \sqrt{2}\,\Delta W_{n+1}^{\tau}, \qquad Y_0 = x_0,$$
$$Y_{n+1} = Y_n - \nabla u(Y_{n+1}^{\tau})\,\lambda + \sqrt{2}\,\Delta W_{n+1}, \qquad n \in \mathbb{N}_0,$$

where $(\tau_n)_{n\in\mathbb{N}}$ are i.i.d. uniform random variables on $[0, 1]$.

Originally introduced as numerical methods for SDEs with irregular coefficients, such randomized schemes were later adapted to sampling settings in [30] and further developed in [4, 36, 5, 16]. In the context of sampling, randomized discretizations are particularly appealing, as they can better capture the underlying stochastic dynamics while mitigating the accumulation of discretization error.

However, existing analyses of these methods rely crucially on global Lipschitz assumptions on the gradient. In particular, it remains unclear whether the stability and bias-reduction properties of randomized schemes persist when the drift exhibits super-linear growth, as is typical in non-log-concave and high-dimensional learning problems.

Motivated by the developments in both the randomized Euler and tamed Euler literature, it is natural to investigate whether the stabilizing effect of taming can be combined with the robustness properties of randomized discretizations. This raises the question of whether one can construct discretization schemes that remain stable under super-linear drift growth while retaining the favorable bias properties of randomized methods.

In this work, we address this question by introducing and analyzing a tamed randomized Langevin Monte Carlo algorithm (tRLMC). We establish non-asymptotic finite-time error estimates for this method in total variation against the underlying SDE, and, in the sampling specialization, derive non-asymptotic convergence guarantees in non-log-concave regimes, including explicit bounds in Wasserstein distance and total variation, with further implications for optimization.

Our contributions. Our results are organized in two layers: finite-time error estimates against the underlying (SDE), valid for general locally Lipschitz drift, and their specialization to the sampling problem under $h = \nabla u$ and a logarithmic Sobolev inequality. The two finite-time frameworks address complementary settings: the KL framework applies to the deterministic tamed Euler scheme, where a sharp cross-regularity estimate is available, whereas the TV framework applies to the randomized scheme, for which such a cross-regularity estimate is currently out of reach due to the randomization of the integration time.

(i) **Finite-time KL error estimates for SDEs with locally Lipschitz drift.** Building on [5], we develop a new local-error framework for controlling the KL divergence between a discretization scheme and the (SDE) when the drift grows super-linearly. Applied to (kTULA), the framework yields a finite-time estimate of the form

$$\mathrm{KL}\big(\mu\widehat{P}^N \,\big\|\, \nu P^N\big) \;\lesssim\; \tfrac{1}{\lambda}\,\mathcal{W}_2^2(\mu,\nu) \;+\; \big((N\lambda)\vee\log N\big)\,\lambda^2,$$

valid for general locally Lipschitz drift; see Theorem 1.

(ii) **Finite-time TV error estimates for randomized schemes under super-linear drift.** By combining the taming methodology with a randomized midpoint method, we introduce a new discretization, termed tRLMC, and develop a corresponding total variation local-error framework. This yields the first finite-time error estimate in total variation for a randomized scheme applied to SDEs with super-linearly growing locally Lipschitz drift; see Theorem 2.

(iii) **Near-optimal sampling guarantees for tamed Euler schemes in** KL **divergence.** Specializing the KL framework to the Langevin case $h = \nabla u$ under a logarithmic Sobolev inequality, we derive an explicit $\tilde{O}(\epsilon^{-\frac{1}{2}})$ complexity bound for kTULA to reach accuracy ϵ in KL divergence toward π_β. This further yields improved guarantees in Wasserstein and total variation distances.

(iv) **Sampling guarantees for tamed randomized schemes.** Specializing the TV framework to the Langevin case, we show that tRLMC achieves $\tilde{O}(\epsilon^{-1})$ complexity for sampling toward π_β in W_2 and total variation distance under a logarithmic Sobolev inequality.

(v) **Connection with optimization.** We apply our Wasserstein convergence results to obtain explicit non-asymptotic excess-risk bounds for tamed Langevin schemes in the non-log-concave setting.

More precisely, we decompose the excess risk as:

$$\underbrace{\mathbb{E}\big[u(\theta_n)\big] - \mathbb{E}_{\pi_\beta}\big[u(x)\big]}_{\text{sampling error}} + \underbrace{\mathbb{E}_{\pi_\beta}\big[u(x)\big] - u^*}_{\text{concentration for large } \beta},$$

where θ_n denotes the n-th iterate of the algorithm. Thus, a solution to the sampling problem yields corresponding optimization guarantees for sufficiently large β.

Organization of the paper. The remainder of the paper is organized as follows. Section 2 introduces the necessary notation, the structural assumptions on the drift of (SDE), and the functional inequalities used in the sampling specialization. Section 3 introduces the two tamed schemes (kTULA) and (tRLMC) and presents the main finite-time error estimates against the underlying SDE in KL divergence and total variation, together with their specialization to sampling in KL, total variation, and Wasserstein distance, and the resulting implications for optimization. Section 4 reports numerical experiments illustrating the stabilizing effect of taming both for sampling and for a nonlinear optimization problem. Section 5 presents the new local-error frameworks underlying our analysis, including the KL framework for kTULA, the corresponding TV framework for the randomized scheme, and the Wasserstein convergence argument for tRLMC. Detailed proofs of all auxiliary results and of the main theorems are deferred to Appendices A-D.

2. Preliminaries and Blanket Assumptions

In this section, we provide the necessary groundwork for stating the proposed algorithmic schemes and our main results. We adopt throughout a two-layer convention: structural assumptions and notation are formulated at the level of the general (SDE) with locally Lipschitz drift h, while functional inequalities and the target measure π_β are introduced only in the sampling specialization $h = \nabla u$.

2.1. **Notation.** Throughout, $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$ and $\langle\cdot,\cdot\rangle$ its associated inner product. For a sufficiently smooth function $f : \mathbb{R}^d \to \mathbb{R}$, we write ∇f, $\nabla^2 f$ and Δf for its gradient, Hessian matrix, and Laplacian respectively. For a continuously differentiable map $F : \mathbb{R}^d \to \mathbb{R}^d$, we write $J(F)$ for its Jacobian. In the gradient case $F = \nabla u$, one has $J(\nabla u) = \nabla^2 u$. For a measurable map T, we write $\mathrm{Law}(T)$ for its induced distribution. For a probability measure μ and Markov kernel P, the pushforward measure μP is

$$(\mu P)(A) := \int_{\mathbb{R}^d} P(x, A)\, \mu(dx), \qquad \mu P^n := \mu \underbrace{P \cdots P}_{n \text{ times}}.$$

SDE semigroup and one-step kernel. Let $(P_t)_{t\geq 0}$ denote the Markov semigroup of (SDE), i.e., $P_t(x,\cdot) = \mathrm{Law}(X_t^x)$ where $X_0^x = x$. Fix a step size $\lambda > 0$ and define the one-step diffusion kernel

$$P := P_\lambda.$$

In the sampling specialization $h = \nabla u$, $(P_t)_{t\geq 0}$ coincides with the Langevin semigroup of (LSDE).

2-Wasserstein distance. Let $\mathcal{P}_2(\mathbb{R}^d)$ denote the space of probability measures with finite second moment. For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the 2-Wasserstein distance is defined by

$$\mathcal{W}_2(\mu,\nu) := \left(\inf_{\gamma \in \mathcal{C}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^2 \, \gamma(dx,dy) \right)^{1/2}, \tag{1}$$

where $\mathcal{C}(\mu,\nu)$ denotes the set of couplings of μ and ν, that is, probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. Equivalently, (X,Y) is a coupling of μ and ν if $\mathrm{Law}(X) = \mu$ and $\mathrm{Law}(Y) = \nu$.

Kullback–Leibler divergence. Let $\mathcal{P}(\mathbb{R}^d)$ denote the space of probability measures on $\mathbb{R}^d$. For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, the Kullback–Leibler divergence is defined by

$$\mathrm{KL}(\mu\|\nu) := \int_{\mathbb{R}^d} \log\left(\frac{d\mu}{d\nu}\right) d\mu, \tag{2}$$

whenever μ is absolutely continuous with respect to ν, and $\mathrm{KL}(\mu\|\nu) := +\infty$ otherwise.

We rely on the standard chain rule and data processing inequality for relative entropy, which yield the marginal bound $\mathrm{KL}(\mu^Y\|\nu^Y) \le \mathrm{KL}(\mu^{X,Y}\|\nu^{X,Y})$. A key refinement for our purposes is the *shifted chain rule* [5]. By introducing an auxiliary variable X', it provides the flexible bound

$$\mathrm{KL}(\mu^Y\|\nu^Y) \le \mathrm{KL}(\mu^{X'}\|\nu^X) + \inf_{\gamma\in\mathcal{C}(\mu^X,\mu^{X'})} \int \mathrm{KL}\big(\mu^{Y|X=x} \,\big\|\, \nu^{Y|X=x'}\big)\,\gamma(dx,dx'). \tag{3}$$

This freedom to couple X and X' is central to the shifted comparison arguments developed in Section 5.

Rényi divergence. The Rényi divergence of order $q > 1$ between two probability measures μ, ν is defined by

$$\mathsf{R}_q(\mu\|\nu) := \frac{1}{q-1}\log\int\left(\frac{d\mu}{d\nu}\right)^q d\nu, \tag{4}$$

whenever $\mu \ll \nu$, and $+\infty$ otherwise. In the limit as $q \to 1$, this recovers the Kullback–Leibler divergence, while the case $q = 2$ yields $\mathsf{R}_2 = \log(1+\chi^2)$, where χ^2 is the chi-squared divergence.

Infinitesimal generator. We denote by $\mathcal{L}$ the infinitesimal generator of (SDE),

$$\mathcal{L}f \;:=\; \tfrac{\sigma^2}{2}\,\Delta f \;-\; \langle h, \nabla f\rangle,$$

which, in the sampling specialization $h = \nabla u$ and $\sigma = \sqrt{\frac{2}{\beta}}$, reduces to the Langevin generator $\mathcal{L}f = \frac{1}{\beta}\Delta f - \langle \nabla u, \nabla f\rangle$.

2.2. **Assumptions on the drift.** We impose the following structural conditions on the drift h of (SDE).

(A1) **Polynomial Jacobian growth.** There exist constants $L > 0$ and $\ell \ge 0$ such that

$$\max\big\{\|h(x)\|,\ \|J(h)(x)\|\big\} \le L\big(1+\|x\|^{2\ell}\big), \qquad \forall x\in\mathbb{R}^d. \tag{PJG}$$

(A2) **Polynomial Lipschitz continuity.** There exist constants $L' > 0$ and $\ell' \ge 0$ such that

$$\|h(x)-h(y)\| \le L'\big(1+\|x\|+\|y\|\big)^{\ell'}\|x-y\|, \qquad \forall x,y\in\mathbb{R}^d. \tag{PLC}$$

(A3) **One-sided Lipschitz continuity.** There exists a constant $K' \in \mathbb{R}$ such that

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

(A4) **Dissipativity.** There exist constants $a, b > 0$ such that

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

(A5) **Polynomial Jacobian Lipschitz continuity.** There exist constants $L'' > 0$ and $\ell'' \ge 0$ such that

$$\|J(h)(x)-J(h)(y)\| \le L''(1+\|x\|+\|y\|)^{\ell''}\|x-y\|, \qquad \forall x,y\in\mathbb{R}^d. \tag{JLC}$$

Assumption (A1) allows the drift and its Jacobian to grow polynomially at infinity, thereby encompassing a wide class of drift coefficients with super-linear growth that fall outside the scope of the standard global Lipschitz framework. In the sampling specialization $h = \nabla u$, it covers non-convex potentials for which classical convergence analyses of (ULA) are not directly applicable.

Assumption (A2) ensures that the drift is locally Lipschitz continuous, with a Lipschitz modulus that may itself grow polynomially. This condition is substantially weaker than global Lipschitz continuity and is compatible with super-linear drifts, while still being sufficient to establish well-posedness of (SDE) and stability of the numerical schemes considered in this work.

Assumption (A3) is a one-sided Lipschitz condition, which plays a key role in controlling the propagation of errors between coupled trajectories of (SDE). It is strictly weaker than monotonicity or strong convexity, and is particularly well suited to the analysis of non-convex dynamics.

Assumption (A4) is a dissipativity condition that ensures coercive behavior of the drift at infinity. This assumption is central to our analysis, as it allows us to derive uniform-in-time moment bounds for both (SDE) and its discretizations, and is a natural coercivity-type condition in applications.

Assumption (A5) is imposed only in the analysis of the (kTULA) scheme. It provides the additional Jacobian regularity—namely, local Lipschitz continuity of $J(h)$ with polynomial modulus—required to obtain the improved weak error rate underlying our KL convergence bounds. In the gradient case $h = \nabla u$, since $J(h) = \nabla^2 u$, this reduces to a polynomial-Lipschitz Hessian condition on the potential. The assumption is not needed for the randomized scheme (tRLMC), whose bias-reduction mechanism does not rely on second-order smoothness of the drift.

2.3. **Sampling and functional inequalities.** We now turn to the specialization of (SDE) that underlies the sampling application. Throughout this subsection and whenever sampling guarantees are stated, we assume that $h = \nabla u$ for some $u : \mathbb{R}^d \to \mathbb{R}$ satisfying $\int_{\mathbb{R}^d} e^{-\beta u(x)}\, dx < \infty$ for any $\beta > 0$, and we set $\sigma = \sqrt{\frac{2}{\beta}}$. Under this specialization, (SDE) reduces to (LSDE), which admits as invariant measure the Gibbs distribution

$$\pi_\beta(A) \;:=\; \frac{\int_A e^{-\beta u(x)}\, dx}{\int_{\mathbb{R}^d} e^{-\beta u(x)}\, dx}, \qquad A \in \mathcal{B}(\mathbb{R}^d).$$

In addition to the structural assumptions (A1)–(A5) on $h = \nabla u$, we impose a functional inequality on π_β.

(A6) **Logarithmic Sobolev inequality (LSI).** The target distribution π_β satisfies a logarithmic Sobolev inequality with constant $C_{\mathrm{LSI}} > 0$, meaning that for every probability measure $\nu \ll \pi_\beta$ with density $f := \frac{d\nu}{d\pi_\beta}$,

$$\mathrm{KL}(\nu \| \pi_\beta) = \int_{\mathbb{R}^d} f \log f \, d\pi_\beta \le \frac{1}{2C_{\mathrm{LSI}}}\, I_{\pi_\beta}(\nu), \tag{LSI}$$

where $I_{\pi_\beta}(\nu)$ denotes the *relative Fisher information* of ν with respect to π_β.

Assumption (A6) is widely used in the analysis of Langevin dynamics since it implies exponential ergodicity in relative entropy. In particular, by the Bakry–Émery theorem [1], (LSI) holds for strongly convex potentials and is stable under bounded perturbations, Lipschitz mappings, and convolutions.

Finally, (LSI) implies Talagrand's transportation-cost inequality: for every $\nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\mathcal{W}_2(\nu, \pi_\beta) \;\le\; \sqrt{\frac{2}{C_{\mathrm{LSI}}}\, \mathrm{KL}(\nu \| \pi_\beta)}. \tag{TI}$$

3. Tamed schemes and main results

This section introduces the two tamed discretization schemes analyzed in this work—(kTULA) and (tRLMC)—and presents the corresponding main results. The results in this section are organized into two distinct layers. First, we establish finite-time error estimates against the underlying (SDE), which rely solely on the structural assumptions (A1)–(A5) for the general drift h. Second, we derive sampling and optimization guarantees toward π_β, which additionally invoke the gradient specialization $h = \nabla u$ and Assumption (A6).

3.1. **Tamed discretizations.** Throughout this subsection, we fix constants $a > 0$ and $\ell \ge 0$, and define, for each $\lambda > 0$, the tamed drift $h_\lambda : \mathbb{R}^d \to \mathbb{R}^d$ associated with h by

$$h_\lambda(x) := ax + \frac{h(x) - ax}{\left(1 + \lambda \|x\|^{2(\ell+1)}\right)^{1/2}}. \tag{5}$$

The construction (5) is well defined for any locally Lipschitz drift h and does not require h to be a gradient field.

The kTULA scheme. The KL-accelerated tamed unadjusted Langevin algorithm is given by

$$\widehat{X}_0^\lambda := X_0, \qquad \widehat{X}_{n+1}^\lambda = \widehat{X}_n^\lambda - \lambda\, h_\lambda(\widehat{X}_n^\lambda) + \sqrt{\frac{2\lambda}{\beta}}\, \xi_{n+1}, \qquad n \in \mathbb{N}_0, \tag{kTULA}$$

where X_0 is an $\mathbb{R}^d$-valued random variable, $\lambda > 0$ is the stepsize, $\beta > 0$ is the inverse temperature, and $(\xi_n)_{n \in \mathbb{N}_0}$ are i.i.d. standard d-dimensional Gaussian random vectors. We denote by π_n^λ the density of $\mathrm{Law}(\widehat{X}_n^\lambda)$ for all $n \in \mathbb{N}_0$.

Remark 1. The design of the tamed coefficient h_λ in (5) follows from that of mTULA [26] and sTULA [20]. It allows us to derive several properties of h_λ, which are crucial to establish moment estimates and convergence results of (kTULA). More precisely, by adopting the splitting trick originally used in [20], the tamed coefficient h_λ satisfies a dissipativity condition, which enables contraction of the algorithm (kTULA) and therefore an easier computation of the associated moment bounds. Moreover, dividing by the term $\left(1 + \lambda \|x\|^{2(\ell+1)}\right)^{1/2}$ in (5) yields an improved upper bound for the difference between h_λ and h, compared to [20].

Main results. Denote by

$$\lambda_{\max}^{\mathrm{kTULA}} := \min\left\{1,\ \frac{1}{8a},\ \frac{1}{(6L_0)^2}\right\}, \tag{6}$$

where $L_0 := 2a + 4L + (\ell+1)(2L+a)$.

Throughout this subsection, for a fixed step size $\lambda > 0$, we write $P := P_\lambda$ for the one-step kernel of (SDE) run for time λ, and $\widehat{P}$ for the one-step kernel associated with (kTULA).

3.1.1. *Finite-time KL error estimate against the underlying SDE.* Our first result is a finite-time KL error estimate between (kTULA) and (SDE), valid for general locally Lipschitz drift.

Theorem 1 (KL local-error bound for kTULA)**.** *Assume* (A1)–(A5)*. Let* $\lambda \in (0, \lambda_{\max}^{\mathrm{kTULA}}]$*. Then there exist constants* $C_1, C > 0$*, depending at most polynomially on the dimension and independent of* N *and* λ*, such that for all* $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ *and all* $N \in \mathbb{N}$ *with* $N \geq 3$*,*

$$\mathrm{KL}\big(\mu\,\widehat{P}^N \,\big\|\, \nu\,P^N\big) \ \leq\ C_1\,\frac{1}{\lambda}\,\mathcal{W}_2^2(\mu,\nu) \ +\ C\,\big((N\lambda)\vee\log N\big)\lambda^2. \tag{7}$$

Remark 2. In particular, for $\mu = \nu$ and $N \geq 3$, the Wasserstein term vanishes, and

$$\mathrm{KL}\big(\mu\,\widehat{P}^N \,\big\|\, \mu\,P^N\big) \ \leq\ C\,\big((N\lambda)\vee\log N\big)\lambda^2.$$

3.1.2. *Sampling specialization.* We now specialize Theorem 1 to the sampling setting introduced in Section 2.3. From here on through the end of this subsection, we assume $h = \nabla u$, so that (LSDE) admits π_β as its invariant measure, and we additionally impose the logarithmic Sobolev inequality (A6). Under this specialization, the kernel P becomes the one-step Langevin diffusion kernel of (LSDE), and Theorem 1 combined with the exponential ergodicity of $(P_t)_{t\geq 0}$ yields sampling guarantees toward π_β.

Corollary 1 (KL sampling guarantee for kTULA under LSI)**.** *Assume* (A1)–(A6)*, and let* $\lambda \in (0, \lambda_{\max}^{\mathrm{kTULA}}]$*. Then there exist constants* $C_1, C > 0$ *such that, for all* $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ *and all* $N \geq 3$*,*

$$\mathrm{KL}\big(\mu\widehat{P}^N \,\big\|\, \pi_\beta\big) \leq \mathrm{KL}\big(\mu\widehat{P}^N \,\big\|\, \mu P^N\big) + R_2\big(\mu P^N \,\big\|\, \pi_\beta\big), \tag{8}$$

where $R_2(\cdot\|\cdot)$ *denotes the Rényi divergence of order* 2*. Consequently,*

$$\mathrm{KL}\big(\mu\widehat{P}^N \,\big\|\, \pi_\beta\big) \leq C\,\big((N\lambda)\vee\log N\big)\lambda^2 + e^{-2C_{\mathrm{LSI}}N\lambda}\,R_2\big(\mu \,\big\|\, \pi_\beta\big). \tag{9}$$

We next deduce explicit mixing-time guarantees from this bound.

Proposition 1 (Mixing time in KL for kTULA under LSI)**.** *Assume* (A1)–(A6)*, and let* $\lambda \in (0, \lambda_{\max}^{\mathrm{kTULA}}]$*. Fix* $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ *and assume* $R_0 := R_2(\mu\|\pi_\beta) < \infty$*. Then*

$$\mathrm{KL}\big(\mu\,\widehat{P}^N \,\big\|\, \pi_\beta\big) \ \leq\ C\,\big((N\lambda)\vee\log N\big)\lambda^2 \ +\ e^{-2C_{\mathrm{LSI}}N\lambda}\,R_0.$$

In particular, for $\lambda \leq \widetilde{O}(\sqrt{\varepsilon})$*, it suffices to take* $N \geq \widetilde{O}\Big(\frac{1}{\sqrt{\varepsilon}}\Big)$ *to ensure*

$$\mathrm{KL}\big(\mu\,\widehat{P}^N \,\big\|\, \pi_\beta\big) \leq \varepsilon.$$

The KL bound further implies convergence in stronger metrics.

Corollary 2 (TV mixing time via Pinsker)**.** *Assume the conditions of Corollary 1 and let* $\lambda \in (0, \lambda_{\max}^{\mathrm{kTULA}}]$*. Fix* $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ *and assume* $R_0 := R_2(\mu\|\pi_\beta) < \infty$*. Then*

$$TV(\mu\widehat{P}^N, \pi_\beta) \ \leq\ C\,\lambda\sqrt{(N\lambda)\vee\log N} \ +\ e^{-C_{\mathrm{LSI}}N\lambda}\sqrt{R_0}.$$

As a consequence, for $\lambda \leq \widetilde{O}(\varepsilon)$*, it suffices to take* $N \geq \widetilde{O}(1/\varepsilon)$ *to ensure*

$$TV(\mu\widehat{P}^N, \pi_\beta) \leq \varepsilon.$$

Corollary 3 (W_2 mixing time via Talagrand)**.** *Assume the conditions of Corollary 1 and let* $\lambda \in (0, \lambda_{\max}^{\mathrm{kTULA}}]$*. Fix* $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ *and assume* $R_0 := R_2(\mu\|\pi_\beta) < \infty$*. Then there exist constants* $C_{W_2}, c_0 > 0$*, where* C_{W_2} *depends at most polynomially on the dimension and* c_0 *depends on* C_{LSI}*, such that*

$$\mathcal{W}_2\big(\mu\widehat{P}^N, \pi_\beta\big) \ \leq\ C_{W_2}\Big(\lambda\sqrt{(N\lambda)\vee\log N} + e^{-c_0 N\lambda}\Big).$$

As a consequence, for $\lambda \leq \widetilde{O}(\varepsilon)$*, it suffices to take* $N \geq \widetilde{O}(1/\varepsilon)$ *to guarantee*

$$\mathcal{W}_2\big(\mu\widehat{P}^N, \pi_\beta\big) \leq \varepsilon.$$

Implications for optimization. When β is sufficiently large, the Gibbs measure π_β concentrates around the minimizers of u [14]. As a result, (kTULA) may also be used to address the optimization problem

$$\text{minimize} \qquad \mathbb{R}^d \ni \theta \mapsto u(\theta).$$

Corollary 4 (Expected excess risk bound for kTULA)**.** *Assume* (A1)–(A6)*, and let* $\lambda \in (0, \lambda_{\max}^{\mathrm{kTULA}}]$*. Let* $(\widehat{X}_n)_{n\geq 0}$ *be the* (kTULA) *iterates. Fix* $\mu := \mathrm{Law}(\widehat{X}_0)$ *such that* $R_0 := R_2(\mu\|\pi_\beta) < \infty$*. Then there exist constants* $C_0, C_1, C_2, C_3 > 0$*, independent of* N *and* λ*, such that for all* $N \in \mathbb{N}$*,*

$$\mathbb{E}\big[u(\widehat{X}_N)\big] - \inf_{\theta\in\mathbb{R}^d} u(\theta) \;\leq\; C_1\, e^{-C_0 N\lambda} + C_2\,\lambda\,\sqrt{(N\lambda)\vee \log N} + C_3 \frac{\log\beta}{\beta}. \tag{10}$$

3.2. **Tamed randomized discretization.** In addition to the fully explicit (kTULA), we also consider a randomized midpoint discretization equipped with our taming scheme. This construction is designed to handle super-linearly growing drifts within the randomized framework, while preserving stability and consistency with the underlying (SDE).

Using this randomized midpoint construction, we define the *tamed randomized Langevin Monte Carlo* algorithm as follows:

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

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

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

where $t_n = n\lambda$ for $n\in\mathbb{N}_0$, and the tamed drift h_λ is defined by (5).

Main results. Denote by

$$\lambda_{\max}^{\mathrm{tRLMC}} := \min\Big\{1,\ \tfrac{1}{8a},\ \tfrac{a}{2CL_0^2},\ \big(\tfrac{pa}{4C_p}\big)^2\Big\}, \tag{11}$$

where the constants are made explicit in the proof of Lemma 12.

Throughout this subsection, for a fixed step size $\lambda > 0$, we write $P := P_\lambda$ for the one-step kernel of (SDE) run for time λ, and $\widetilde{P}$ for the one-step kernel associated with (tRLMC).

3.2.1. *Finite-time TV error estimate against the underlying SDE.* Our main finite-time result for (tRLMC) is a total-variation error estimate against (SDE), valid for general locally Lipschitz drift.

Theorem 2 (TV local-error bound for tRLMC)**.** *Assume* (A1)–(A4) *and let* $\lambda \in (0, \lambda_{\max}^{\mathrm{tRLMC}}]$*. Then there exist constants* $C_1, C > 0$*, independent of* N *and* λ*, such that for all* $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$ *and all* $N\in\mathbb{N}$ *with* $N\geq 3$*,*

$$TV\big(\mu\,\widetilde{P}^N \,\|\, \nu\,P^N\big) \;\leq\; C_1\,\frac{1}{\sqrt{\lambda}}\,\mathcal{W}_2(\mu,\nu) \;+\; C\lambda\sqrt{\big((N\lambda)\vee\log N\big)}. \tag{12}$$

Theorem 2 requires only (A1)–(A4)—in particular, it does not require the Jacobian regularity (A5) used in the KL framework—and is valid for general locally Lipschitz drift, with no gradient structure imposed on h. To the best of our knowledge, this is the first finite-time total-variation error estimate for a randomized discretization scheme applied to SDEs with super-linearly growing drift.

3.2.2. *Sampling specialization.* We now specialize Theorem 2 to the sampling setting. From here on through the end of this subsection, we assume $h = \nabla u$, so that (LSDE) admits π_β as its invariant measure, and we additionally impose Assumption (A6).

Corollary 5 (TV mixing time for tRLMC)**.** *Assume* (A1)–(A6)*, and let* $\lambda \in (0, \lambda_{\max}^{\mathrm{tRLMC}}]$*. There holds*

$$TV(\mu_0\widetilde{P}^N, \pi_\beta) \;\leq\; C\lambda\sqrt{\big((N\lambda)\vee\log N\big)} \;+\; \sqrt{e^{-C_{\mathrm{LSI}}\lambda N}\mathrm{KL}(\mu_0\,\|\,\pi_\beta)}.$$

As a result, for $\lambda \leq \widetilde{O}(\varepsilon)$ *one needs* $N \geq \widetilde{O}\big(\frac{1}{\varepsilon}\big)$ *to guarantee*

$$TV(\mu_0\widetilde{P}^N, \pi_\beta) \leq \varepsilon.$$

We also obtain convergence bounds in Wasserstein distance.

Theorem 3 (Wasserstein convergence for tRLMC)**.** *Assume* (A1)–(A6)*, and let* $\lambda \in (0, \lambda_{\max}^{\mathrm{tRLMC}}]$*. There holds*

$$W_2(\mu_0 \widetilde{P}^n, \pi_\beta) \;\le\; C_{W_2}\big(\lambda + e^{-c_0 \lambda n}\big),$$

where C_{W_2} *depends polynomially on the dimension and* c_0 *depends on* C_{LSI}*. As a result, for* $\lambda \le \widetilde{O}(\varepsilon)$ *one needs* $N \ge \widetilde{O}\big(\frac{1}{\varepsilon}\big)$ *to guarantee*

$$W_2(\mu_0 \widetilde{P}^N, \pi_\beta) \le \varepsilon.$$

Implications for optimization. As for (kTULA), we connect the Wasserstein convergence bound to the optimization problem $\min_{\theta \in \mathbb{R}^d} u(\theta)$ via the concentration of π_β around the minimizers of u.

Corollary 6 (Expected excess risk bound for tRLMC)**.** *Assume* (A1)–(A6)*, and let* x_n *be the* n*-th iterate of the tamed randomized Langevin Monte Carlo scheme* (tRLMC)*. There exist constants* $C, C', c_0 > 0$*, depending at most polynomially on the dimension, such that*

$$\mathbb{E}[u(x_n)] - u^* \;\le\; C\big(\lambda + e^{-c_0 \lambda n}\big) + C' \frac{\log \beta}{\beta}.$$

3.3. **Comparison with related literature.** We compare our results with the existing literature on Langevin-based sampling under non-convexity and local Lipschitz continuity of the gradient. We emphasize that Theorems 1 and 2 are valid for general locally Lipschitz drift on (SDE) and are, to the best of our knowledge, the first finite-time error estimates of their kind under super-linear drift growth.

Comparison in Wasserstein distance. Several recent works on taming and projection schemes provide convergence rates of $\mathcal{O}(\varepsilon^{-1})$ in Wasserstein distance under super-linear drift growth [34, 2]. Our analysis of (kTULA) matches this rate under the same or weaker assumptions, and extends the analysis to KL divergence and total variation, both of which are stronger metrics.

Comparison in KL divergence. For tamed schemes in KL divergence, the closest comparison is with [22, 20, 23], of which [23] provides the strongest known guarantees prior to the present work. The current article provides the best known complexity for convergence to the invariant measure under these assumptions. More specifically, ignoring logarithmic factors in ε, our algorithm converges in KL divergence after

$$n \;\ge\; \widetilde{\mathcal{O}}\big(d^{3(\ell+1)/2}\, \varepsilon^{-1/2}\big).$$

On the other hand, [23] achieves a general rate of $\widetilde{\mathcal{O}}\big(d^{\frac{\ell+1}{\epsilon_h}}\, \varepsilon^{-\frac{1}{2-\epsilon_h}}\big)$ for $\epsilon_h \in \big(0, \frac{1}{2}\big]$, which explodes in the dimension as $\epsilon_h \to 0$. Specializing their framework to $\epsilon_h = \frac{1}{2}$ corresponding to our setting, that work achieves a complexity of $\widetilde{\mathcal{O}}\big(d^{2(\ell+1)}\, \varepsilon^{-\frac{2}{3}}\big)$ for KL accuracy ε.

Comparison for randomized schemes. To the best of our knowledge, the only existing analysis of randomized Langevin schemes under super-linear drift growth is [33], which establishes Wasserstein convergence. Our analysis of (tRLMC) matches the Wasserstein rate of [33] and additionally provides, for the first time in this setting, a non-asymptotic convergence guarantee in total variation, obtained through a new total variation local-error framework.

4. Numerical Experiments

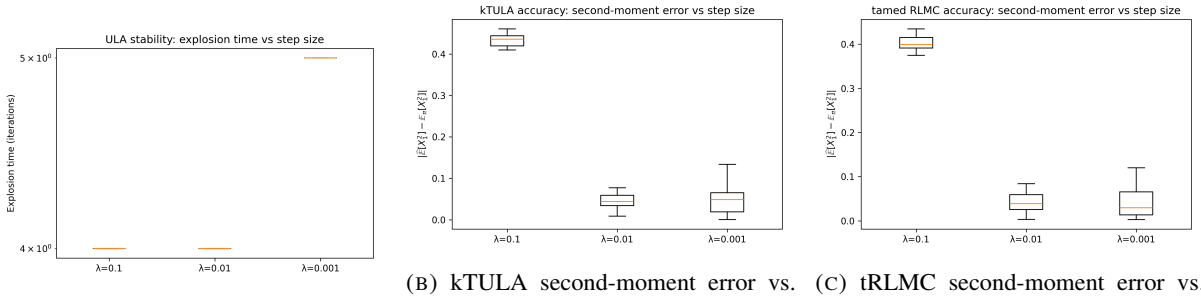
4.1. **Sampling experiments: stability and accuracy.** We first study the stability and accuracy of the proposed kTULA scheme (kTULA) and the tamed randomized Langevin method (tRLMC), and compare them with the classical Unadjusted Langevin Algorithm (ULA). As a benchmark we consider the double-well potential

$$U(x) = \tfrac{1}{4}x^4 - \tfrac{1}{2}x^2, \tag{13}$$

with gradient $\nabla U(x) = x^3 - x$. This example satisfies the dissipativity assumptions used in our analysis while exhibiting superlinear drift growth at infinity.

All experiments are performed in dimension $d = 100$ with inverse temperature $\beta = 1$. For each stepsize $\lambda \in \{0.1, 0.01, 0.001\}$, ULA, kTULA, and tRLMC are run for $N = 2 \times 10^5$ iterations. The initial condition is chosen deterministically as $X_0 = (200, 0, \ldots, 0) \in \mathbb{R}^{100}$, placing the system far from equilibrium and in a regime where the superlinear drift dominates. For kTULA and tRLMC, samples are collected after a burn-in period of 5×10^4 iterations. Each experiment is repeated independently 30 times, with all randomness generated from fixed seeds.

To assess stability, we record the *explosion time* of ULA, defined as the first iteration at which numerical overflow occurs. To assess accuracy for kTULA and tRLMC, we estimate the second moment of the first coordinate and report the absolute error $\left|\widehat{\mathbb{E}}[X_1^2] - \mathbb{E}_{\pi_\beta}[X_1^2]\right|$. Here π_β denotes the invariant law with density proportional to $\exp\bigl(-\beta \sum_{i=1}^d U(x_i)\bigr)$. Since the coordinates are independent under the target distribution, $\mathbb{E}_{\pi_\beta}[X_1^2]$ is the second moment of the associated one-dimensional marginal. In the experiments this reference value is computed numerically by trapezoidal quadrature on a dense grid over $[-4, 4]$.



(A) ULA explosion time vs. stepsize.

(B) kTULA second-moment error vs. stepsize.

(C) tRLMC second-moment error vs. stepsize.

FIGURE 1. Stability and accuracy in the aggressive regime ($d = 100$). Left: ULA explosion time across the tested stepsizes. Center–right: second-moment error for kTULA and tRLMC.

Figure 1 shows a clear separation between the untamed and tamed schemes. For every tested stepsize, ULA becomes unstable after only a few iterations, which is consistent with the known failure of Euler-type Langevin discretizations under superlinear drifts. By contrast, both kTULA and tRLMC remain numerically stable throughout the simulation horizon. Their empirical second-moment error also decreases substantially when the stepsize is reduced from $\lambda = 0.1$ to $\lambda = 0.01$, while stability is preserved across all tested values of λ.

TABLE 1. Summary statistics over 30 independent repetitions. For ULA we report explosion time; for kTULA and tRLMC we report the absolute second-moment error. Entries are empirical mean $\pm$ standard deviation.

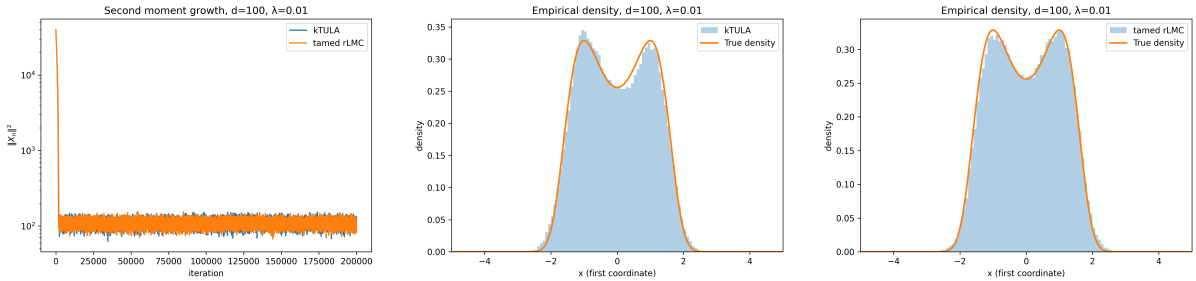
Method	$\lambda = 0.1$	$\lambda = 0.01$	$\lambda = 0.001$
ULA explosion time	4.00 ± 0.00	4.00 ± 0.00	5.00 ± 0.00
kTULA: $\left\lvert\widehat{\mathbb{E}}[X_1^2] - \mathbb{E}_{\pi_\beta}[X_1^2]\right\rvert$	0.4336 ± 0.0154	0.0453 ± 0.0170	0.0455 ± 0.0325
tRLMC: $\left\lvert\widehat{\mathbb{E}}[X_1^2] - \mathbb{E}_{\pi_\beta}[X_1^2]\right\rvert$	0.4025 ± 0.0159	0.0428 ± 0.0211	0.0437 ± 0.0360

Table 1 makes the separation quantitative. ULA explodes after roughly four to five iterations for all tested stepsizes. The zero empirical standard deviation for the ULA explosion times indicates that, for each tested stepsize, overflow occurred at the same iteration in all 30 repetitions from the chosen deterministic initialization. In contrast, both tamed schemes remain stable in all runs.

Figure 2 complements this quantitative comparison by displaying the trajectory behavior of the two tamed schemes at the representative stepsize $\lambda = 0.01$. We plot the evolution of $\|X_n\|^2$ and the empirical marginal density of the first coordinate against the corresponding one-dimensional marginal of the target law.

The left panel of Figure 2 shows that both tamed schemes enter a stable regime after a short transient and maintain controlled second moments. The center and right panels indicate that the empirical marginals recover the expected bimodal structure of the target distribution, providing further evidence that the tamed schemes remain stable while capturing the qualitative shape of the invariant law.

Taken together, these experiments illustrate the stabilizing effect of taming in the sampling regime. In this aggressive setting, ULA fails almost immediately, whereas kTULA and tRLMC remain stable over long horizons and attain substantially smaller second-moment error once the stepsize is reduced from the coarsest regime.



(A) Second-moment growth. (B) kTULA empirical density. (C) tRLMC empirical density.

FIGURE 2. Representative behavior at $\lambda = 0.01$ in the aggressive regime ($d = 100$). Left: evolution of $\|X_n\|^2$ (log scale) for kTULA and tRLMC. Center–right: empirical marginal densities of the first coordinate compared with the corresponding target marginal.

4.2. **Optimization involving a neural network in the fixed-feature setting.** We finally consider the optimization problem

$$\min_{\theta \in \mathbb{R}^d} u_n(\theta) := \frac{1}{n} \sum_{j=1}^{n} \big(Y_j - \mathcal{N}(\theta, Z_j)\big)^2 + \frac{\eta}{6} \sum_{k=1}^{d} |\theta_k|^6, \tag{14}$$

where $X_j = (Z_j, Y_j) \in \mathbb{R}^m$, with $Y_j \in \mathbb{R}$ the target variable and $Z_j \in \mathbb{R}^{m-1}$ the input variable, $\eta > 0$ is a regularization constant, and $\mathcal{N} : \mathbb{R}^d \times \mathbb{R}^{m-1} \to \mathbb{R}$ is a single-hidden-layer feed-forward neural network with fixed random features, given by $\mathcal{N}(\theta, z) := \sum_{i=1}^{d_1} W_i \, \sigma(\langle c_i, z\rangle + b_i)$, where $c_i \in \mathbb{R}^{m-1}$ are fixed random feature vectors, $W_i \in \mathbb{R}$ are trainable output weights, and $b_i \in \mathbb{R}$ are trainable bias parameters. The activation is the sigmoid linear unit $\sigma(x) = x/(1 + e^{-x})$, and the trainable parameter is $\theta = ((W_1, \dots, W_{d_1}), (b_1, \dots, b_{d_1})) \in \mathbb{R}^{2d_1}$, so that $d = 2d_1$.

We ask whether the provable stability of the tamed schemes manifests in a simple nonlinear problem of the form (14) when the learning rate is taken into an aggressive regime, and whether standard first-order optimizers — which carry no comparable guarantees on (14) — display the deterioration one would expect.

We generate a synthetic regression dataset from a teacher model of the same fixed-feature form. The inputs are sampled from a standard Gaussian distribution in dimension $m - 1 = 20$; the teacher network has hidden width 80; independent Gaussian noise with standard deviation 0.05 is added to the teacher output; and the student network has hidden width $d_1 = 100$. The training set size is $n = 4000$, the test set size is 1000, and training is performed with mini-batches of size 128.

We compare SGD, Adam, AMSGrad, kTULA, and tRLMC on the empirical objective (14). For SGD we use momentum 0.9. For the two tamed schemes we set $\beta = 10^6$, $a = 10^{-2}$, $\ell = 4$, and $\eta = 0.05$. Here a is the linear stabilizing component in the taming map; we fix $a = 10^{-2}$ throughout all runs as a small baseline dissipative term. Each method is run for 20 epochs, repeated over the five seeds $1, 2, 3, 4, 5$, and tested at learning rates $\lambda \in \{0.1, 0.2, 0.3\}$.

Figure 3 reports the final test MSE as a function of the learning rate. The main qualitative trend is that the two tamed methods are empirically less sensitive to increasing learning rate than SGD, Adam, and AMSGrad in this synthetic setting. In particular, the three standard optimizers deteriorate substantially as λ increases, whereas kTULA varies only moderately across the tested range. tRLMC performs comparably to kTULA at $\lambda = 0.1$ and $\lambda = 0.2$, but degrades more noticeably at $\lambda = 0.3$. Since this experiment is small and synthetic, we interpret these results as evidence of a favorable stability trend rather than as a broad benchmark conclusion.

Table 2 makes the final-error comparison quantitative. In particular, kTULA has the smallest mean final test MSE at $\lambda = 0.2$ and $\lambda = 0.3$, while tRLMC is slightly better at $\lambda = 0.1$ but becomes more variable at $\lambda = 0.3$.

Figure 4 complements this comparison by displaying the epoch-wise evolution of the parameter norm, test MSE, and training objective at the representative learning rate $\lambda = 0.1$. The parameter-norm panel provides the clearest visual separation between the standard and tamed methods. SGD exhibits steadily increasing norm, while Adam and AMSGrad operate at substantially larger parameter norms throughout training. By contrast, kTULA and tRLMC remain in a much smaller norm regime over the full horizon.

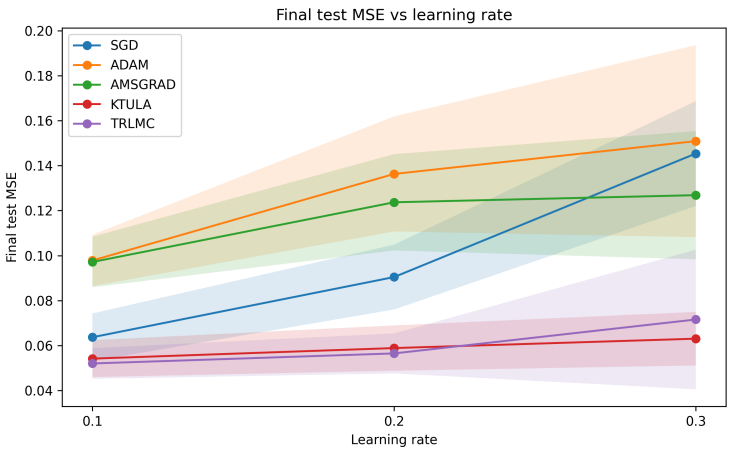
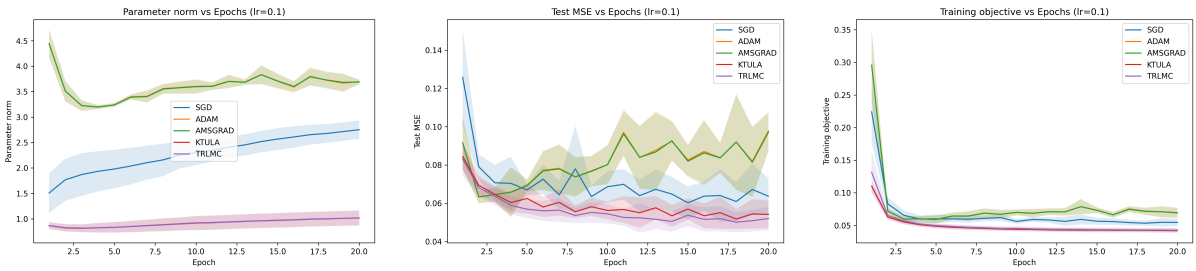


FIGURE 3. Final test MSE versus learning rate in the fixed-feature neural-network experiment. The tamed methods are empirically less sensitive to increasing learning rate than SGD, Adam, and AMSGrad on this synthetic task.

TABLE 2. Final test MSE over 5 seeds in the fixed-feature neural-network experiment. Entries are empirical mean $\pm$ standard deviation across seeds.

Method	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$
SGD	0.0637 ± 0.0106	0.0904 ± 0.0144	0.1454 ± 0.0233
Adam	0.0978 ± 0.0114	0.1363 ± 0.0256	0.1509 ± 0.0427
AMSGrad	0.0972 ± 0.0112	0.1237 ± 0.0214	0.1269 ± 0.0285
kTULA	0.0542 ± 0.0082	0.0589 ± 0.0101	0.0630 ± 0.0119
tRLMC	0.0520 ± 0.0068	0.0565 ± 0.0089	0.0716 ± 0.0310

The test-MSE panel suggests that, after the initial transient, the two tamed methods maintain lower average test error than the standard baselines at $\lambda = 0.1$. Likewise, the training-objective panel indicates that kTULA and tRLMC settle at lower objective values, while Adam and AMSGrad level off at visibly higher levels. These plots do not show instability in the strict sense of numerical explosion, but they do indicate that taming leads to more controlled iterates and more stable empirical performance in this regime.



(A) Parameter norm. (B) Test MSE. (C) Training objective.

FIGURE 4. Epoch-wise behavior at learning rate $\lambda = 0.1$. The tamed methods remain in a smaller parameter-norm regime and achieve lower average test error and training objective than the standard baselines on this synthetic task. Shaded regions indicate one empirical standard deviation across the five seeds.

Taken together, these experiments support the same qualitative message as the preceding study. In this nonlinear regression problem with high-order regularization and relatively aggressive learning rates, the tamed Langevin-type schemes display more controlled parameter growth and more robust empirical behavior than SGD, Adam, and AMSGrad. We emphasize, however, that this experiment is intended as a small-scale diagnostic study.

5. Proof outline and technical innovations

The convergence analyses underlying our main results share a common technical core, which we develop in detail in this section. The goal is to bound the divergence between two stochastic processes driven by *different* Markov kernels: the discrete scheme kTULA or tRLMC on the one hand, and the (SDE) on the other. Classical coupling arguments are tailored to the Wasserstein metric and do not transfer verbatim to information-theoretic divergences such as Kullback–Leibler or total variation. The shifted composition methodology of [5] provides a principled remedy by introducing an auxiliary, third process that interpolates between the two processes of interest. We adapt this methodology to the present setting, where the super-linear growth of the drift induces a state-dependent coupling increment that does not appear in [5]. This is the only point at which the abstract machinery genuinely departs from [5].

Following the two-layer structure of the paper, the local-error frameworks developed in this section—both the KL framework of §5.1–§5.5 and the TV framework of §5.6—are formulated at the level of the (SDE) with general locally Lipschitz drift h, requiring no gradient structure. The Wasserstein convergence argument of §5.7 additionally invokes the sampling specialization $h = \nabla u$ and Assumption (A6), as it relies on the exponential contraction of $(P_t)_{t\geq 0}$ toward π_β.

5.1. **Setup, kernels, and one-step local errors.** Fix a step size $\lambda \in (0, 1]$. Let $P = P_\lambda$ denote the Markov kernel of (SDE) run for time λ, and let $\widehat{P}$ denote the one-step transition kernel of (kTULA). For any $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and $n \in \mathbb{N}_0$, set $\widehat{\mu}_n := \mu\,\widehat{P}^{\,n}$ and $\nu_n := \nu\,P^n$. The objective of the framework is to control $\mathrm{KL}(\widehat{\mu}_N \,\|\, \nu_N)$ in terms of *one-step*, i.e. local, quantities computed at a single iteration.

We work with three measurable functions encoding the one-step discrepancy between $\widehat{P}$ and P:

$$E_{\mathrm{strong}}, E_{\mathrm{weak}} \colon \mathbb{R}^d \to [0, \infty), \qquad \Gamma \colon \mathbb{R}^d \times \mathbb{R}^d \times (0, \infty) \to [0, \infty).$$

Here $E_{\mathrm{strong}}(x)$ controls the L^2-distance between one step of $\widehat{P}$ and one step of P from the same point x; $E_{\mathrm{weak}}(x)$ is the corresponding bias, which exploits cancellations and is therefore typically smaller than $E_{\mathrm{strong}}(x)$; and $\Gamma(x, y, \lambda)$ is a *coupling increment* controlling the synchronous-coupling distortion between two trajectories of (SDE) started at x and y. For convenience we set

$$a_0(x) := E_{\mathrm{strong}}(x), \qquad a_1(x, y) := E_{\mathrm{weak}}(x) + \Gamma(x, y, \lambda)\, E_{\mathrm{strong}}(x). \tag{15}$$

The framework is fed by the following one-step Wasserstein bound, verified for kTULA in Lemma 9: there exists $L > 0$ such that, for all $x, y \in \mathbb{R}^d$,

$$\mathcal{W}_2^2\big(\delta_x \widehat{P}, \delta_y P\big) \;\leq\; L^2\, \|x - y\|^2 + 2\, a_1(x, y)\, \|x - y\| + a_0(x)^2. \tag{16}$$

Remark 3 (State-dependence of a_1). In [5], the coupling increment γ is a constant. In the present super-linear setting, the polynomial Lipschitz condition (A2) forces Γ to depend on both endpoints, namely $\Gamma(x, y, \lambda) = C_{\mathrm{cpl}}\lambda(1 + \|x\|^{\ell'} + \|y\|^{\ell'})$. This is the only structural difference between our framework and [5]; as we shall see, it is benign provided uniform moments along the auxiliary process introduced in §5.2 are established.

5.2. **The auxiliary interpolating process.** The principal obstacle in bounding $\mathrm{KL}(\widehat{\mu}_N \,\|\, \nu_N)$ directly is that $\widehat{\mu}_N$ and ν_N are produced by *different* Markov kernels: any one-step KL bound therefore encodes both the discretization bias and the contraction of the SDE semigroup, and the two effects are difficult to separate. The shifted composition strategy circumvents this obstacle by introducing a third process $(Y'_n)_{n=0}^N$ that (i) starts at ν_0, (ii) terminates at $\widehat{\mu}_N$, and (iii) evolves at intermediate steps via the SDE kernel P, so that intermediate one-step KL terms only involve P and are amenable to standard regularity estimates of the semigroup $(P_t)_{t\geq 0}$.

Construction. Let $(\widehat{X}_n)_{n\geq 0}$, $(Y_n)_{n\geq 0}$ denote the kTULA chain 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

$$\widetilde{Y}_n := (1 - \eta_n)\, Y'_n + \eta_n\, \widehat{X}_n, \qquad n = 0, 1, \ldots, N-1. \tag{17}$$

Geometrically, $\widetilde{Y}_n$ lies on the segment between Y'_n and $\widehat{X}_n$; the parameter η_n measures how far the auxiliary process is pushed toward the kTULA chain at step n. Define

$$Q_n := \begin{cases} P, & 0 \leq n \leq N-2, \\ \widehat{P}, & n = N-1, \end{cases} \tag{18}$$

and update the auxiliary process by

$$Y'_{n+1} \sim Q_n(\widetilde{Y}_n, \cdot), \qquad n = 0, 1, \ldots, N-1, \qquad Y'_0 = Y_0. \tag{19}$$

We write $\nu'_n := \mathrm{Law}(Y'_n)$.

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

$$\nu'_0 \;=\; \nu_0, \qquad \nu'_N \;=\; \widehat{\mu}_N, \tag{20}$$

since $\widetilde{Y}_{N-1} = \widehat{X}_{N-1}$ and the last update applies $\widehat{P}$ to $\widehat{X}_{N-1}$. Figure 5 illustrates the construction. The freedom to choose $\eta_n \in [0, 1]$ is what enables the framework to exploit *both* the contraction of P and the local discretization errors of $\widehat{P}$: the bounds, determined in Proposition 2, dictate the final iteration complexity.

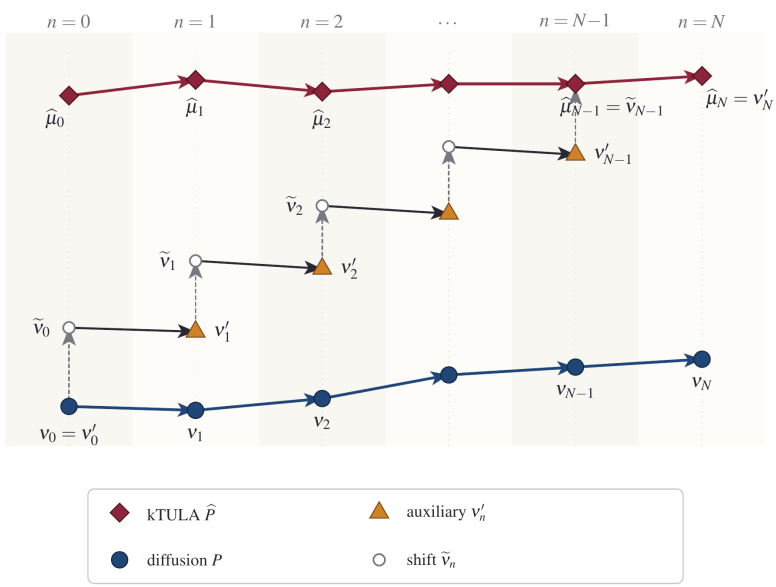


FIGURE 5. Auxiliary interpolating process underlying the shifted-composition argument. The top trajectory $\{\widehat{\mu}_n\}$ (red diamonds) is generated by the kTULA kernel $\widehat{P}$ from initial law μ, while the bottom trajectory $\{\nu_n\}$ (blue circles) is generated by the SDE kernel P from initial law ν. The auxiliary process $\{\nu'_n\}$ (orange triangles) matches the bottom at $n = 0$ and the top at $n = N$. At each step $n < N - 1$, the law ν'_n is shifted along the Wasserstein geodesic toward $\widehat{\mu}_n$ (gray dashed arrow) to produce the shifted state $\widetilde{\nu}_n$ (open circle), after which the SDE kernel P is applied. Only at $n = N - 1$, where the constraint $\eta_{N-1} = 1$ forces $\widetilde{\nu}_{N-1} = \widehat{\mu}_{N-1}$, the kTULA kernel $\widehat{P}$ is applied instead to enforce (20).

5.3. **KL telescoping via the shifted chain rule.** The bridge between the auxiliary process and the KL divergence is the shifted chain rule (3) introduced in Section 2. Recall that the auxiliary variable X' allows comparison of Markov updates from *different* starting points, which is precisely what is needed when the underlying kernels disagree.

We assume the following two one-step KL bounds:

$$\mathrm{KL}\big(\delta_x P \,\|\, \delta_y P\big) \;\le\; c\,\|x - y\|^2, \tag{21}$$

$$\mathrm{KL}\big(\delta_x \widehat{P} \,\|\, \delta_y P\big) \;\le\; c'\,\|x - y\|^2 + b(x)^2, \tag{22}$$

for constants $c, c' \ge 0$ and a measurable function $b : \mathbb{R}^d \to [0, \infty)$. We refer to (21) as the *regularity* of the SDE kernel and to (22) as the *cross-regularity* of the scheme. For kTULA, both are established in Propositions 5–8.

Lemma 1 (KL telescoping bound)**.** *Let* $(Y'_n)_{n=0}^N$ *and* $(\widetilde{Y}_n)_{n=0}^{N-1}$ *be defined by* (17)–(19)*. Then*

$$\mathrm{KL}\big(\widehat{\mu}_N \,\|\, \nu_N\big) \;\le\; \sum_{n=0}^{N-1} \mathbb{E}\Big[\mathrm{KL}\big(Q_n(\widetilde{Y}_n,\cdot) \,\big\|\, P(Y'_n,\cdot)\big)\Big]\,. \tag{23}$$

If (21)–(22) *hold, then*

$$\mathrm{KL}\big(\widehat{\mu}_N \,\|\, \nu_N\big) \;\le\; c\sum_{n=0}^{N-2} \eta_n^2\,\mathbb{E}\|\widehat{X}_n - Y'_n\|^2 + c'\,\mathbb{E}\|\widehat{X}_{N-1} - Y'_{N-1}\|^2 + \bar{b}^{\,2}, \tag{24}$$

where $\bar{b}^{\,2} := \|b\|^2_{L^2(\widehat{\mu}_{N-1})}$.

5.4. **Shifted Wasserstein recursion.** The right-hand side of (24) is controlled entirely by the squared coupling distances

$$d_n^2 \;:=\; \mathbb{E}\|\widehat{X}_n - Y'_n\|^2, \qquad n = 0,1,\dots,N-1, \tag{25}$$

which propagate forward in time according to the next lemma. This is the only place in the analysis where the state-dependent coefficient a_1 enters.

Lemma 2 (Shifted Wasserstein recursion)**.** *Assume* (16)*. For all* $0 \le n \le N-2$*,*

$$d_{n+1}^2 \;\le\; L^2(1-\eta_n)^2\, d_n^2 + 2\,(1-\eta_n)\,\overline{a}_1\, d_n + \overline{a}_0^2, \tag{26}$$

where $\overline{a}_0 := \sup_{0\le k<N} \|a_0(\widehat{X}_k)\|_{L^2}$ *and* $\overline{a}_1 := \sup_{0\le k<N} \|a_1(\widehat{X}_k,\widetilde{Y}_k)\|_{L^2}$.

Proof. By the Wasserstein lift of the coupling and (16),

$$\mathcal{W}_2^2(\widehat{\mu}_{n+1},\nu'_{n+1}) \;\le\; L^2\,\mathbb{E}\|\widehat{X}_n - \widetilde{Y}_n\|^2 + 2\,\mathbb{E}\Big[a_1(\widehat{X}_n,\widetilde{Y}_n)\,\|\widehat{X}_n - \widetilde{Y}_n\|\Big] + \mathbb{E}\Big[a_0(\widehat{X}_n)^2\Big]\,.$$

Since $\widehat{X}_n - \widetilde{Y}_n = (1-\eta_n)(\widehat{X}_n - Y'_n)$, the first term equals $L^2(1-\eta_n)^2 d_n^2$, Cauchy–Schwarz bounds the cross term by $2(1-\eta_n)\overline{a}_1 d_n$, and the third term by $\overline{a}_0^2$. Choosing the synchronous Wasserstein coupling on the next step yields (26). □

Remark 4 (Effect of state-dependence on $\overline{a}_1$). The supremum defining $\overline{a}_1$ involves the joint law of $(\widehat{X}_k,\widetilde{Y}_k)$, hence requires uniform moment bounds on both $(\widehat{X}_k)$ and the shifted process $(\widetilde{Y}_k)$. Since $\widetilde{Y}_k$ is a convex combination of Y'_k and $\widehat{X}_k$, controlling its moments reduces to controlling those of Y'_k, which we establish in Lemma 10 via the dissipativity condition (A4) on the drift of (SDE). This is the only place where the polynomial-Lipschitz nature of the drift plays a non-trivial role at the level of the framework.

The recursion (26) is purely deterministic once $\overline{a}_0, \overline{a}_1, L$ have been fixed. Its analysis is identical to that of the corresponding recursion in [5, Lemma B.5], and yields the following optimal control of the shifted distances.

Proposition 2 (Deterministic control of the shifted distances)**.** *Let* $L \in [\frac{1}{2}, 2]$*,* $\overline{a}_0, \overline{a}_1 \ge 0$*, and* $\bar{N} := N \wedge 1/(1-L)_+$*. There exists a sequence* $(\eta_n)_{n=0}^{N-2} \subset [0,1]$ *with* $\eta_{N-1} = 1$ *such that any nonnegative sequence* $(d_n)_{n=0}^{N-1}$ *satisfying* (26) *and* $d_0 = \mathcal{W}_2(\mu,\nu)$ *obeys*

$$d_{N-1}^2 \;\lesssim\; \tfrac{L^{-1}-1}{L^{-N}-1}\, d_0^2 + \big(((L-1)N) \vee \log \bar{N}\big)\,\overline{a}_0^2 + \bar{N}\,\overline{a}_1^2, \tag{27}$$

$$\sum_{n=0}^{N-2} \eta_n^2\, d_n^2 \;\lesssim\; \tfrac{L^{-1}-1}{L^{-N}-1}\, d_0^2 + \big(((L-1)N) \vee \log \bar{N}\big)\,\overline{a}_0^2 + \bar{N}\,\overline{a}_1^2. \tag{28}$$

Proof. The recursion (26) has the same form as the recursion analyzed in [5, Appendix B.2], with deterministic coefficients $\overline{a}_0, \overline{a}_1$ in place of the local errors used there. The optimal shift sequence and the dynamic programming argument of [5, Lemmas B.4–B.6] apply verbatim and yield (27)–(28). □

5.5. **The KL local-error bound for kTULA.** Combining Lemma 1 with Proposition 2 yields the main KL bound of the framework.

Theorem 4 (KL local-error framework)**.** *Assume* (16)*,* (21)*, and* (22)*, and let* $L \in [\frac{1}{2}, 2]$*. With* $\overline{a}_0, \overline{a}_1$ *as in Lemma 2,* $\bar{b} := \max_{0\le n<N} \|b\|_{L^2(\widehat{\mu}_n)}$*, and* $\bar{N} = N \wedge 1/(1-L)_+$*,*

$$\mathrm{KL}\big(\mu\widehat{P}^N \,\|\, \nu P^N\big) \;\lesssim\; (c+c')\Big[\tfrac{L^{-1}-1}{L^{-N}-1}\,\mathcal{W}_2^2(\mu,\nu) + \big(((L-1)N) \vee \log \bar{N}\big)\,\overline{a}_0^2 + \bar{N}\,\overline{a}_1^2\Big] + \bar{b}^{\,2}. \tag{29}$$

We emphasize that the bound (29) compares the discretization $\mu\widehat{P}^N$ against the law νP^N of the (SDE); no invariant-measure or gradient structure is invoked at this stage. The specialization to sampling toward π_β proceeds via the additional ergodicity input provided by Assumption (A6), as carried out in Section 3.

5.6. **A TV bound for the randomized scheme tRLMC.** For the tamed randomized midpoint scheme tRLMC, the KL framework above is not directly applicable: the randomization of the integration time prevents the natural construction of an adapted continuous-time interpolation, and consequently a sharp cross-regularity estimate of the form (22) is not currently available. We therefore work in the weaker total variation metric, where a direct one-step estimate is at hand.

Let $\widetilde{P}$ denote the one-step kernel of tRLMC, and write $\widetilde{\mu}_n := \mu\widetilde{P}^n$. We assume, in addition to the regularity (21), the existence of a one-step TV local error: there exists a measurable function $C_{\mathrm{TV}}\colon \mathbb{R}^d \to [0,\infty)$ such that, for all $x \in \mathbb{R}^d$,

$$TV\big(\delta_x\widetilde{P}, \delta_x P\big) \;\le\; C_{\mathrm{TV}}(x)\,\lambda. \tag{30}$$

For tRLMC, this estimate is established in Proposition 11.

The auxiliary process construction of §5.2 remains valid with $\widehat{P}$ replaced by $\widetilde{P}$ throughout.

Lemma 3 (TV bound via the shifted construction)**.** *Assume* (21) *and* (30)*, and let* $d_n^2 := \mathbb{E}\|\widetilde{X}_n - Y_n'\|^2$. *Writing* $\overline{C}_{\mathrm{TV}} := \sup_{0\le n<N} \|C_{\mathrm{TV}}(x)\|_{L^2(\widetilde{\mu}_n)}$,

$$TV\big(\widetilde{\mu}_N, \nu_N\big) \;\le\; \overline{C}_{\mathrm{TV}}\,\lambda + \sqrt{\tfrac{c}{2}}\,d_{N-1} + \sqrt{\tfrac{c}{2}\sum_{n=0}^{N-2}\eta_n^2\,d_n^2}. \tag{31}$$

Theorem 5 (TV local-error framework)**.** *Assume* (16)*,* (21)*, and* (30)*, and let* $L \in [\frac{1}{2}, 2]$. *With* $\overline{a}_0, \overline{a}_1$ *as in Lemma 2 (with* $\widehat{P}$ *replaced by* $\widetilde{P}$*) and* $\bar{N} = N \wedge 1/(1-L)_+$,

$$TV\big(\mu\widetilde{P}^N, \nu P^N\big) \;\lesssim\; \overline{C}_{\mathrm{TV}}\,\lambda + \sqrt{\tfrac{L^{-1}-1}{L^{-N}-1}\,\mathcal{W}_2^2(\mu,\nu) + \big(((L-1)N)\vee\log\bar{N}\big)\,\overline{a}_0^2 + \bar{N}\,\overline{a}_1^2}. \tag{32}$$

The proof combines Lemma 3 with Proposition 2. The KL and TV frameworks rest on the same shifted Wasserstein recursion and differ only at the terminal step: the KL framework requires the cross-regularity estimate (22), whereas the TV framework only requires (30), making Theorem 5 applicable to schemes such as tRLMC for which a sharp cross-regularity estimate is currently out of reach. As with Theorem 4, the bound (32) is stated against the law νP^N of (SDE) and requires no gradient structure on the drift.

5.7. **Wasserstein convergence of tRLMC.** Wasserstein convergence for tRLMC avoids the auxiliary process construction and proceeds via a more direct route based on the contraction of the underlying semigroup toward π_β. Unlike the local-error frameworks of §5.1–§5.6, this argument is genuinely sampling-specific: throughout the present subsection we therefore assume the specialization $h = \nabla u$, $\sigma = \sqrt{2/\beta}$ of (SDE) to the (LSDE), and we additionally impose Assumption (A6).

Under this specialization, the local Lipschitz property of the drift, combined with the one-sided Lipschitz condition (A3), yields a one-step Wasserstein bound of the form (16) with $\Gamma(x,y,\lambda) = C_{\mathrm{cpl}}\lambda(1 + \|x\|^{2\ell} + \|y\|^{2\ell})$. Iterating this estimate over n_0 steps and exploiting the uniform moment bounds of Lemma 12 furnishes the long-time discretization-error estimate

$$\mathcal{W}_2\Big(\mu_0\widetilde{P}^n,\ \big(\mu_0\widetilde{P}^{n-n_0}\big)P^{n_0}\Big) \;\le\; C\,e^{c\lambda n_0}\,\lambda, \qquad n \ge n_0 \ge 0, \tag{33}$$

where C, c depend polynomially on the dimension. In parallel, the combination of the logarithmic Sobolev inequality (A6) and the one-sided Lipschitz condition (A3) implies the contraction

$$\mathcal{W}_2\big(\nu P^t, \pi_\beta\big) \;\le\; C_{\mathrm{contr}}\,e^{-\dot{c}\,t}\,\mathcal{W}_2\big(\nu, \pi_\beta\big), \qquad t \ge 0, \tag{34}$$

for some $\dot{c} > 0$ depending on C_{LSI}. Combining (33) and (34) via the triangle inequality

$$\mathcal{W}_2\big(\mu_0\widetilde{P}^n, \pi_\beta\big) \;\le\; \mathcal{W}_2\Big(\mu_0\widetilde{P}^n,\ \big(\mu_0\widetilde{P}^{n-n_0}\big)P^{n_0}\Big) + \mathcal{W}_2\Big(\big(\mu_0\widetilde{P}^{n-n_0}\big)P^{n_0}, \pi_\beta\Big),$$

and balancing the two contributions by an appropriate choice of n_0 yields the announced Wasserstein convergence estimate of Theorem 3.

APPENDIX A. DRIFT PROPERTIES AND MOMENT ESTIMATES

This appendix collects technical estimates underlying the analysis. We record structural properties of the tamed drift, including dissipativity, growth, and global Lipschitz bounds, and establish uniform moment bounds for both the underlying LSDE and the kTULA iterates. We also include a local Taylor remainder estimate under polynomially growing Jacobian regularity, which is used in the control of one-step discretization errors. Proofs are deferred to Section D unless otherwise stated.

A.1. **Properties of the tamed drift.**

Lemma 4 (Key properties of the tamed drift)**.** *Assume that* (A1)*,* (A2)*, and* (A4) *hold. Then for any* $\theta, \bar{\theta} \in \mathbb{R}^d$ *and any* $0 < \lambda < 1$*, the following statements hold.*

(i) Dissipativity. *The tamed drift satisfies*

$$\langle h_\lambda(\theta), \theta \rangle \geq a\|\theta\|^2 - b.$$

(ii) Growth control. *The magnitude of* h_λ *satisfies*

$$\|h_\lambda(\theta)\| \leq 2a\|\theta\| + 2L\,\lambda^{-1/2}, \qquad \|h_\lambda(\theta)\| \leq (2a + L)\big(1 + \|\theta\|^{\ell+1}\big).$$

(iii) Lipschitz continuity. *There exists a constant* $L_0 > 0$*, depending only on* a, L, ℓ*, such that*

$$\|h_\lambda(\theta) - h_\lambda(\bar{\theta})\| \leq L_0\,\lambda^{-1/2}\,\|\theta - \bar{\theta}\|,$$

where

$$L_0 := 2a + 4L + (\ell + 1)(2L + a).$$

(iv) Taming error. *The deviation between the original and tamed drifts satisfies*

$$\|h(\theta) - h_\lambda(\theta)\|^2 \leq 4\lambda^2(L + a)^2\big(1 + \|\theta\|^{6(\ell+1)}\big).$$

Proof. The result follows from [23, Lemma 4.1] specialized to $\varepsilon_h = \frac{1}{2}$. □

A.2. **Moment bounds for the underlying SDE.**

Lemma 5 (Uniform $2p$-moment bounds for LSDE)**.** *Let Assumptions* (A1)–(A4) *hold, and let* $(X_t)_{t\geq 0}$ *solve LSDE with* $X_0 = x$*. Then for any* $p \in \mathbb{N}$ *there exist constants* $c_p > 0$ *and* $C_p > 0$ *such that for all* $t \geq 0$*,*

$$\mathbb{E}\|X_t\|^{2p} \;\leq\; e^{-c_p t}\,\|x\|^{2p} \;+\; C_p,$$

where one can take

$$c_p = pa, \qquad C_p = C\Big(b + \tfrac{1}{\beta}(d + 2(p-1))\Big)^p$$

for a constant $C > 0$ *depending only on* p*.*

Proof. The proof follows from [18, Lemma A.1]. □

Lemma 6 (One-step p-moment drift for LSDE)**.** *Fix* $p \geq 2$*. Under Assumption* (A4)*, there exist constants* $q > 0$ *and* $C_p > 0$*, depending only on* p*,* d*,* a*,* b*, and* β*, such that for every* $\lambda \in (0, 1]$ *and every* $x \in \mathbb{R}^d$*,*

$$\mathbb{E}\|X(t, x; t + \lambda)\|^p \leq e^{-q\lambda}\|x\|^p + C_p\lambda,$$

where $X(t, x; \cdot)$ *denotes the solution of* (LSDE) *started from* x *at time* t*.*

Proof. The proof is given in D.1.1. □

A.3. **Moment bounds for the kTULA scheme.**

We now establish uniform-in-time moment bounds for the kTULA chain.

Lemma 7 (Uniform moments for the (kTULA) scheme)**.** *Let Assumptions (A1)–(A4) hold. There exists* $\lambda_{\max} > 0$ *such that for all* $0 < \lambda \leq \lambda_{\max}$ *the following holds. Let* $(\widehat{X}_n)_{n\geq 0}$ *denote the kTULA iterates and let* $\widehat{X}_t$ *be any continuous interpolation on* $[n\lambda, (n+1)\lambda]$ *satisfying* $\widehat{X}_{n\lambda} = \widehat{X}_n$*. Then for every* $p \geq 2$ *there exists a constant* $c_p < \infty$ *such that for all* $t \in [n\lambda, (n+1)\lambda]$*,*

$$\mathbb{E}\|\widehat{X}_t\|^{2p} \;\leq\; (1 - a(t - n\lambda))(1 - a\lambda)^n\,\mathbb{E}\|\widehat{X}_0\|^{2p} + c_p\Big(1 + \frac{1}{a}\Big).$$

In particular,

$$\sup_{t\geq 0}\mathbb{E}\|\widehat{X}_t\|^{2p} \;\leq\; \mathbb{E}\|\widehat{X}_0\|^{2p} + c_p\Big(1 + \frac{1}{a}\Big) < \infty.$$

Proof. The result follows from [23, Lemma 4.2]. □

A.4. **A local smoothness remainder bound.** We close this appendix with an elementary Taylor remainder estimate that will be invoked repeatedly when controlling one-step discretization errors under polynomially growing local smoothness of the Jacobian.

Lemma 8 (Taylor remainder under locally Lipschitz Jacobian)**.** *Let $h : \mathbb{R}^d \to \mathbb{R}^d$ be continuously differentiable. Under Assumption* (A5)*, for all $x, y \in \mathbb{R}^d$,*

$$\big\|h(x) - h(y) - J(h)(x)(x-y)\big\| \le \frac{L''}{2}\,(1 + \|x\| + \|y\|)^{\ell''}\,\|x-y\|^2.$$

Proof. A Taylor expansion of $h\big(y + t(x-y)\big)$ for $t \in [0,1]$, combined with (A5) to bound the increment of the Jacobian $J(h)$, yields the claim. □

APPENDIX B. LOCAL ERROR TOOLS AND REGULARITY

This appendix collects auxiliary estimates for the underlying (SDE) that are used to verify the local-error framework. In particular, we establish W_2-Lipschitz continuity, derive a coupling increment bound, and record the resulting one-step Wasserstein estimate together with a KL regularity bound.

Proposition 3 (W_2–Lipschitz continuity)**.** *Assume* (A3)*. Let $(X_t)_{t\ge0}$ and $(Y_t)_{t\ge0}$ be solutions of* (SDE) *driven by the same Brownian motion, with initial conditions $X_0 = x$ and $Y_0 = y$. Then for all $t \ge 0$,*

$$\|X_t - Y_t\|_{L^2} \le e^{K't}\,\|x-y\|.$$

In particular, for $t \in [0,\lambda]$, setting $L := e^{K'\lambda}$ yields

$$\|X_t - Y_t\|_{L^2} \le L\,\|x-y\|.$$

For $\lambda \le \frac{\ln 2}{|K'|}$, then $L \in [\frac{1}{2}, 2]$.

Proof. The proof is given in Subsection D.2.1. □

Proposition 4 (Coupling increment)**.** *Assume* (A1)–(A3)*. There exists a constant $C_{\mathrm{cpl}} > 0$, depending on L', ℓ', and β, such that for all $t \in [0,\lambda]$, with $\lambda \le \frac{\ln 2}{|K'|}$,*

$$\|X_t - x - (Y_t - y)\|_{L^2} \;\le\; \Gamma(x,y,t)\,\|x-y\|,$$

where $(X_t)_{t\ge0}$, $(Y_t)_{t\ge0}$ are solutions of (LSDE) *driven by the same Brownian motion with $X_0 = x$, $Y_0 = y$, and*

$$\Gamma(x,y,t) := C_{\mathrm{cpl}}\,t\,\big(1 + \|x\|^{\ell'} + \|y\|^{\ell'}\big).$$

In particular, taking $t = \lambda$ yields

$$\Gamma(x,y,\lambda) := C_{\mathrm{cpl}}\,\lambda\,\big(1 + \|x\|^{\ell'} + \|y\|^{\ell'}\big).$$

Proof. The proof is given in Subsection D.2.2. □

Lemma 9 (One-step Wasserstein inequality)**.** *Let $L > 0$ be the W_2-Lipschitz constant from Proposition 3, let $\Gamma(x,y,\lambda)$ be the coupling increment from Proposition 4, and let $E_{\mathrm{weak}}, E_{\mathrm{strong}} : \mathbb{R}^d \to [0,\infty)$ denote the weak and strong one-step local errors of* (kTULA) *(Propositions 6–7). Then, for all $x, y \in \mathbb{R}^d$,*

$$\mathcal{W}_2^2(\delta_x\widehat{P}, \delta_y P) \le L^2\|x-y\|^2 + 2\Big(E_{\mathrm{weak}}(x) + \Gamma(x,y,\lambda)\,E_{\mathrm{strong}}(x)\Big)\|x-y\| + E_{\mathrm{strong}}(x)^2.$$

Proof. The result follows by the same argument as in [5], with the only difference that here the coupling increment $\Gamma(x,y,\lambda)$ depends on both endpoints rather than only on x, leading to the state-dependent coefficient. □

Remark 5. With the notation

$$a_0(x) := E_{\mathrm{strong}}(x), \qquad a_1(x,y) := E_{\mathrm{weak}}(x) + \Gamma(x,y,\lambda)E_{\mathrm{strong}}(x),$$

the one-step Wasserstein estimate takes the form

$$\mathcal{W}_2^2(\delta_x\widehat{P}, \delta_y P) \le L^2\|x-y\|^2 + 2a_1(x,y)\|x-y\| + a_0(x)^2.$$

Proposition 5 (Regularity [32, Theorem 1.1])**.** *Let $P = P_\lambda$ denote the one-step kernel of* (SDE)*. There exists a constant $C_{\mathrm{reg}} > 0$, depending on β, such that for all $x, y \in \mathbb{R}^d$ and all $t \in (0, \lambda]$,*

$$\mathrm{KL}\big(\delta_x P_t \,\big\|\, \delta_y P_t\big) \;\le\; R_2(\delta_x P_t, \delta_y P_t) \;\le\; \frac{C_{\mathrm{reg}}}{1 - e^{-K't}}\, \|x - y\|^2.$$

In particular, if $K'\lambda \le 2$, then at $t = \lambda$ one may take

$$c = \frac{3C_{\mathrm{reg}}}{\lambda}.$$

Appendix C. Auxiliary Results for the Main Theorems

This appendix collects auxiliary results used in the proofs of the main convergence theorems. We record the strong and weak local error bounds, cross-regularity estimates, and shifted moment controls needed for kTULA, together with the corresponding moment, local-error, total variation, and Wasserstein tools for the tamed randomized scheme.

C.1. Strong and weak local errors of kTULA.

Proposition 6 (Weak local error)**.** *Let Assumptions* (A1)–(A5) *hold. Let X_t denote the solution of* (LSDE) *started from $X_0 = x$, and $\widehat{X}_t$ the one-step interpolation associated with* (kTULA) *started from $\widehat{X}_0 = x$. Set $M := \max\{3(\ell+1),\, 4\ell + \ell''\}$. Then there exists a constant $C > 0$, independent of $x \in \mathbb{R}^d$, d, and $\lambda \in (0, 1]$, such that for every $x \in \mathbb{R}^d$,*

$$\big\|\mathbb{E}[\widehat{X}_\lambda] - \mathbb{E}[X_\lambda]\big\| \;\le\; C\big(1 + d^{M/2} + \|x\|^M\big)\, \lambda^2.$$

In particular,

$$E_{\mathrm{weak}}(x) := C\big(1 + d^{M/2} + \|x\|^M\big)\, \lambda^2.$$

Proof. The proof is given in Subsection D.3.1. □

Proposition 7 (Strong local error)**.** *Let Assumptions* (A1)–(A4) *hold. Let X_t denote the solution of* (LSDE) *started from $X_0 = x$, and $\widehat{X}_t$ the one-step interpolation associated with* (kTULA) *started from $\widehat{X}_0 = x$. Set $M := \max\{3(\ell+1),\, 2\ell' + 4\ell\}$. Then there exists a constant $C > 0$, independent of $x \in \mathbb{R}^d$, d, and $\lambda \in (0, 1]$, such that for every $x \in \mathbb{R}^d$,*

$$\|\widehat{X}_\lambda - X_\lambda\|_{L^2} \;\le\; C\big(1 + d^{M/2} + \|x\|^M\big)\, \lambda^{3/2}.$$

In particular,

$$E_{\mathrm{strong}}(x) := C\big(1 + d^{M/2} + \|x\|^M\big)\, \lambda^{3/2}.$$

Proof. The proof is given in Subsection D.3.2. □

C.2. Cross-regularity for kTULA.

Proposition 8 (Cross-regularity)**.** *Assume* (A1)–(A4)*. There exist a constant $C_{\mathrm{cross}} > 0$ and a measurable function $b : \mathbb{R}^d \to [0, \infty)$ such that, for all $x, y \in \mathbb{R}^d$ and all $t \in (0, \lambda]$,*

$$\mathrm{KL}\big(\delta_x \widehat{P} \,\big\|\, \delta_y P_t\big) \;\le\; \frac{C_{\mathrm{cross}}}{1 - e^{-K't}}\, \|x - y\|^2 \;+\; b(x)^2,$$

where

$$b(x)^2 \le C\, \lambda^2 \big(1 + \|x\|^{2\ell+2}\big).$$

In particular, taking $t = \lambda$, one may take

$$c' = \frac{C_{\mathrm{cross}}}{\lambda}.$$

Proof. The proof is given in Subsection D.3.3. □

C.3. **Shifted recursion and auxiliary moment control for kTULA.**

Lemma 10 (Uniform moment bounds for the shifted process)**.** *Let $p \geq 1$ and define the shifted process $\widetilde{Y}_n := (1-\eta_n)Y'_n + \eta_n \widehat{X}_n$ for $n \geq 0$. Under the hypotheses of Lemma 6 (with constants $q, C > 0$), and assuming $\widehat{M}_p := \sup_{n\geq 0} \mathbb{E}\|\widehat{X}_n\|^p < \infty$, we have*

$$\mathbb{E}\|\widetilde{Y}_n\|^p \leq e^{-q\lambda n}\mathbb{E}\|\widetilde{Y}_0\|^p + \frac{C\lambda + \widehat{M}_p}{1-e^{-q\lambda}},$$

yielding the uniform bound $\sup_{n\geq 0}\mathbb{E}\|\widetilde{Y}_n\|^p \leq \mathbb{E}\|\widetilde{Y}_0\|^p + \frac{C\lambda+\widehat{M}_p}{1-e^{-q\lambda}} < \infty$.

Proof. The proof is given in Subsection D.3.4. □

Lemma 11 (Shifted distance recursion)**.** *Let E_{weak}, E_{strong}, and Γ denote the weak-error, strong-error, and coupling-increment functions of Propositions 6, 7, and 4, and set*

$$a_0(x) := E_{\text{strong}}(x), \qquad a_1(x,y) := E_{\text{weak}}(x) + \Gamma(x,y,\lambda)\, E_{\text{strong}}(x).$$

For $0 \leq n \leq N-1$, let $\eta_n \in [0,1]$, $\widetilde{Y}_n := (1-\eta_n)Y'_n + \eta_n\widehat{X}_n$, and $d_n^2 := \mathbb{E}\|\widehat{X}_n - Y'_n\|^2$, and put

$$\overline{a}_0 := \sup_{0\leq k<N} \|a_0(\widehat{X}_k)\|_{L^2}, \qquad \overline{a}_1 := \sup_{0\leq k<N} \|a_1(\widehat{X}_k, \widetilde{Y}_k)\|_{L^2}.$$

Then, with $L := e^{K'\lambda}$, for every $0 \leq n \leq N-2$,

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

Proof. The proof is given in Subsection D.3.5. □

C.4. **Auxiliary moment bounds for tRLMC.**

Lemma 12 (Uniform moments for tRLMC)**.** *Assume* (A1)*,* (A2)*, and* (A4)*. Let $(\bar{Y}_n)_{n\geq 0}$ denote the tamed randomized Langevin Monte Carlo scheme* (tRLMC)*, initialized at $\bar{Y}_0 = x_0 \in \mathbb{R}^d$. Then the following statements hold.*

(i) *There exist constants $\lambda_{2,\max}^{tRLMC} > 0$, $\mu > 0$, and $M_2 > 0$, depending only on a, b, L, L_0, β, such that for every $\lambda \in (0, \lambda_{2,\max}^{tRLMC}]$,*

$$\mathbb{E}\|\bar{Y}_n\|^2 \leq e^{-\mu n\lambda}\,\mathbb{E}\|x_0\|^2 + M_2\, d, \qquad n \in \mathbb{N}_0.$$

(ii) *For every $p \in [2, 4\ell] \cap \mathbb{N}$, there exist constants $\lambda_{p,\max}^{tRLMC} > 0$, $c_p > 0$, and $\widetilde{C}_p > 0$, depending only on p, a, b, L, L_0, β, such that for every $\lambda \in (0, \lambda_{p,\max}^{tRLMC}]$,*

$$\mathbb{E}\|\bar{Y}_n\|^{2p} \leq (1-c_p\lambda)^n\,\mathbb{E}\|x_0\|^{2p} + \widetilde{C}_p(1+d^p), \qquad n \in \mathbb{N}_0.$$

In particular,

$$\sup_{n\geq 0}\mathbb{E}\|\bar{Y}_n\|^{2p} < \infty.$$

Proof. The proof of (i) is given in Subsection D.4.1, and the proof of (ii) is given in Subsection D.4.2. □

Lemma 13 (One-step p-moment drift for RLMC)**.** *Fix $p \geq 2$. Under Assumption* (A4)*, there exist constants $q > 0$ and $C_p > 0$, depending only on p, d, a, b, and β, such that for all $\lambda \in (0,1]$,*

$$\mathbb{E}\|Y'_{n+1}\|^p \leq e^{-q\lambda}\,\mathbb{E}\|\widetilde{Y}_n\|^p + C_p\lambda,$$

where

$$Y'_{n+1} \sim P_\lambda(\widetilde{Y}_n, \cdot), \qquad \widetilde{Y}_n = (1-\eta_n)Y'_n + \eta_n\bar{Y}_n, \qquad \eta_n \in [0,1].$$

Proof. The proof is given in Subsection D.4.3. □

Lemma 14 (Uniform moment bounds for the shifted process under tRLMC)**.** *Let $p \geq 2$ and define the shifted process $\widetilde{Y}_n := (1-\eta_n)Y'_n + \eta_n\bar{Y}_n$ for $n \geq 0$. Under the hypotheses of Lemma 13 (with constants $q, C_p > 0$), and assuming $\bar{M}_p := \sup_{n\geq 0}\mathbb{E}\|\bar{Y}_n\|^p < \infty$, we have*

$$\mathbb{E}\|\widetilde{Y}_n\|^p \leq e^{-q\lambda n}\mathbb{E}\|\widetilde{Y}_0\|^p + \frac{C_p\lambda + \bar{M}_p}{1-e^{-q\lambda}},$$

yielding the uniform bound $\sup_{n\geq 0}\mathbb{E}\|\widetilde{Y}_n\|^p \leq \mathbb{E}\|\widetilde{Y}_0\|^p + \frac{C_p\lambda+\bar{M}_p}{1-e^{-q\lambda}} < \infty$.

Proof. The proof is given in Subsection D.4.4. □

C.5. **Weak and strong local errors for tRLMC.**

Proposition 9 (Weak error for tRLMC)**.** *Assume* (A1)*,* (A2)*, and* (A4)*. Let* $X(t,x;t+\lambda)$ *denote the solution of* (LSDE) *at time* $t+\lambda$*, starting from* $X_t = x$*, and let* $\lambda \in (0,1)$*. Let* $\bar{Y}(t,x;t+\lambda)$ *be the one-step* tamed randomized midpoint *approximation defined in* (tRLMC)*. Set* $R := 3(\ell+1)$*. Then, for all* $t \geq 0$ *and* $x \in \mathbb{R}^d$*, there exists a constant* $C > 0$*, independent of* d, t, x, λ*, such that*

$$\left\|\mathbb{E}\big[X(t,x;t+\lambda) - \bar{Y}(t,x;t+\lambda)\big]\right\| \;\leq\; C\left(1 + d^{R/2} + \|x\|^R\right)\lambda^2.$$

Proof. The proof is given in Subsection D.4.5. □

Proposition 10 (Strong error for tRLMC)**.** *Assume* (A1)*,* (A2)*, and* (A4)*. Let* $X(t,x;t+\lambda)$ *denote the solution of* (SDE) *at time* $t+\lambda$*, starting from* $X_t = x$*, and let* $\lambda \in (0,1)$*. Let* $\bar{Y}(t,x;t+\lambda)$ *be the one-step* tamed randomized midpoint *approximation defined in* (tRLMC)*. Set* $\widetilde{R} := \max\{2\ell' + 4\ell,\, 6(\ell+1)\}$*. Then, for all* $t \geq 0$ *and* $x \in \mathbb{R}^d$*, there exists a constant* $C > 0$*, independent of* d, t, x, λ*, such that*

$$\mathbb{E}\big[\|X(t,x;t+\lambda) - \bar{Y}(t,x;t+\lambda)\|^2\big] \;\leq\; C\left(1 + d^{\widetilde{R}} + \|x\|^{2\widetilde{R}}\right)\lambda^3.$$

Proof. The proof is given in Subsection D.4.6. □

C.6. **Total variation estimates for tRLMC.**

Proposition 11 (Cross-TV regularity for tRLMC)**.** *There exist a constant* $C > 0$ *and a measurable function* $C_{\mathrm{TV}} : \mathbb{R}^d \to [0,\infty)$ *such that, for all* $x, y \in \mathbb{R}^d$*,*

$$\mathrm{TV}(\delta_x \widetilde{P}, \delta_y P) \;\leq\; C_{\mathrm{TV}}(x)\,\lambda \;+\; C\,\frac{\|x-y\|}{\sqrt{\lambda}}.$$

Proof. The proof is given in Section D.4.7. □

Lemma 15 (TV bound via the shifted construction)**.** *Assume* (21) *and* (30)*. Let*

$$d_n^2 := \mathbb{E}\|\bar{Y}_n - Y_n'\|^2.$$

Then

$$\mathrm{TV}(\widetilde{\mu}_N, \nu_N) \leq \overline{C}_{\mathrm{TV}}\,\lambda + \sqrt{\tfrac{c}{2}}\,d_{N-1} + \sqrt{\tfrac{c}{2}\sum_{n=0}^{N-2}\eta_n^2\,d_n^2}, \tag{35}$$

where $c = \mathcal{O}(\lambda^{-1})$*.*

Proof. The proof is given in Section D.4.8. □

C.7. **Wasserstein convergence for tRLMC.** Throughout this subsection, we work in the sampling specialization $h = \nabla u$, $\sigma = \sqrt{\frac{2}{\beta}}$, under Assumption (A6). We denote by P the one-step Langevin kernel of (LSDE), and by $\widetilde{P}$ the one-step kernel of (tRLMC).

Lemma 16. *For any* $n, n_0 \in \mathbb{N}$*,*

$$W_2\big(\mu_0\widetilde{P}^n,\; \big(\mu_0\widetilde{P}^{\,n-n_0}\big)P^{n_0}\big) \;\leq\; C_d\, e^{\frac{3}{2}K'\lambda n_0}\,\lambda,$$

where $C_d = \mathcal{O}(d^{2\ell+2})$*.*

Proof. Recall the one-step Wasserstein estimate (16):

$$W_2^2(\delta_x\widetilde{P}, \delta_y P) \leq L^2\|x-y\|^2 + 2a_1(x,y)\|x-y\| + a_0(x)^2.$$

Applying Young's inequality yields

$$W_2^2(\delta_x\widetilde{P}, \delta_y P) \leq L^3\|x-y\|^2 + \frac{a_1(x,y)^2}{(L-1)L^2} + a_0(x)^2.$$

By convexity of the Wasserstein distance,

$$W_2^2(\mu\widetilde{P}, \nu P) \leq L^3 W_2^2(\mu,\nu) + \frac{1}{(L-1)L^2}\int a_1(x,y)^2\,d\mu(x)\,d\nu(y) + \int a_0(x)^2\,d\mu(x).$$

Define

$$\bar{a}_0^2 := \sup_{0\leq k<n_0}\mathbb{E}\big[a_0(\bar{Y}_k)^2\big], \qquad \bar{a}_1^2 := \sup_{0\leq k<n_0}\mathbb{E}\big[a_1(\bar{Y}_k, Y_k)^2\big],$$

where $(\bar{Y}_k)$ are the (tRLMC) iterates and (Y_k) the corresponding SDE updates. Iterating the above inequality yields

$$W_2^2\big(\mu_0\widetilde{P}^n,\ \big(\mu_0\widetilde{P}^{\,n-n_0}\big)P^{n_0}\big) \le L^{3n_0}\Big(\frac{\bar{a}_1^2}{(L-1)^2}+\frac{\bar{a}_0^2}{L-1}\Big).$$

Since $L = e^{K'\lambda}$, one has $(L-1)^{-1} = \mathcal{O}(\lambda^{-1})$. Using the moment bounds from Lemma 12, together with the local error estimates from Propositions 9–10, we obtain

$$\bar{a}_0^2 = \mathcal{O}(d^{2\ell+2}\lambda^3), \qquad \bar{a}_1^2 = \mathcal{O}(d^{4\ell}\lambda^2).$$

Substituting these bounds yields the result. □

Lemma 17 (Contraction)**.** *For all $n, n_0 \in \mathbb{N}$,*

$$W_2\big(\big(\mu_0\widetilde{P}^{\,n-n_0}\big)P^{n_0},\pi_\beta\big) \le C_{\mathrm{contr}}\, e^{-\dot{c}\,\lambda n_0}\, W_2\big(\mu_0\widetilde{P}^{\,n-n_0},\pi_\beta\big).$$

Proof. Since π_β satisfies a logarithmic Sobolev inequality (Assumption (A6)) and ∇u is one-sided Lipschitz (Assumption (A3)), the result follows from [34, Proposition 2.5] applied at time $t = \lambda n_0$. □

Proposition 12 (Wasserstein convergence)**.** *For all $n \in \mathbb{N}$,*

$$W_2(\mu_0\widetilde{P}^n,\pi_\beta)\ \le\ \frac{C^*C_d}{1-e^{-1}}\,\lambda\ +\ C'e^{-c_0\lambda n},$$

where $C_d, C^*, C', c_0 > 0$ are given explicitly in the proof. This estimate yields Theorem 3.

Proof. Let $n_0 = \left\lceil \frac{\log C_{\mathrm{contr}}+1}{\dot{c}\,\lambda}\right\rceil$, so that $C_{\mathrm{contr}}e^{-\dot{c}\lambda n_0} \le e^{-1}$, and set $C^* := e^{\frac{3}{2}K'\lambda n_0}$.

By Lemma 16 and Lemma 17, for $n \ge n_0$,

$$\begin{aligned}
W_2(\mu_0\widetilde{P}^n,\pi_\beta) &= W_2\big(\big(\mu_0\widetilde{P}^{n-n_0}\big)\widetilde{P}^{n_0},\pi_\beta\big)\\
&\le W_2\big(\big(\mu_0\widetilde{P}^{n-n_0}\big)\widetilde{P}^{n_0},\ \big(\mu_0\widetilde{P}^{n-n_0}\big)P^{n_0}\big) + W_2\big(\big(\mu_0\widetilde{P}^{n-n_0}\big)P^{n_0},\pi_\beta\big)\\
&\le C_d e^{\frac{3}{2}K'\lambda n_0}\lambda + C_{\mathrm{contr}}e^{-\dot{c}\lambda n_0}\, W_2(\mu_0\widetilde{P}^{n-n_0},\pi_\beta)\\
&\le C^*C_d\lambda + \frac{1}{e}\, W_2(\mu_0\widetilde{P}^{n-n_0},\pi_\beta).
\end{aligned}$$

By [35, Lemma 3.18],

$$W_2(\mu_0\widetilde{P}^n,\pi_\beta) \le \frac{C^*C_d}{1-e^{-1}}\,\lambda + \sup_{N\ge 0} W_2(\mu_0\widetilde{P}^N,\pi_\beta)\, e^{-\frac{n}{n_0}+1}.$$

From the uniform moment bounds in Lemma 12,

$$W_2(\mu_0\widetilde{P}^N,\pi_\beta) \le \sqrt{\mathbb{E}_{\mu_0\widetilde{P}^N}\|x\|^2 + \mathbb{E}_{\pi_\beta}\|x\|^2} \le C_{\mathrm{mom}},$$

where $C_{\mathrm{mom}} = \mathcal{O}(d)$. Hence, defining $C' := C_{\mathrm{mom}}\, e$ and $c_0 = \frac{\dot{c}}{\log C_{\mathrm{contr}}+1}$,

$$W_2(\mu_0\widetilde{P}^n,\pi_\beta) \le \frac{C^*C_d}{1-e^{-1}}\,\lambda + C'\, e^{-c_0\lambda n},$$

as claimed. □

Appendix D. Detailed Proofs

This appendix gathers the detailed proofs of the auxiliary lemmas and main results stated throughout the paper.

D.1. **Proofs for Appendix B: Drift Properties and Moment Estimates.**

D.1.1. *Proof of Lemma 6.*

Proof. Set $X_s := X(t, x; t+s)$ for $s \in [0, \lambda]$ and consider the Lyapunov function $V(z) := \|z\|^p$. Evaluating the infinitesimal generator $\mathcal{L}$ of the SDE on V, and invoking the dissipativity of h alongside Young's inequality, yields

$$\mathcal{L}V(z) = -p\|z\|^{p-2}\langle h(z), z\rangle + \frac{p(d+p-2)}{\beta}\|z\|^{p-2} \le -qV(z) + C_p,$$

for some constants $q, C_p > 0$. Applying Itô's formula and taking expectations yields the differential inequality $\frac{d}{ds}\mathbb{E}\|X_s\|^p \le -q\mathbb{E}\|X_s\|^p + C_p$. Integrating this over $[0, \lambda]$ with the initial condition $X_0 = x$, and applying the elementary bound $1 - e^{-q\lambda} \le q\lambda$, establishes

$$\mathbb{E}\|X(t, x; t+\lambda)\|^p \le e^{-q\lambda}\|x\|^p + \frac{C_p}{q}(1 - e^{-q\lambda}) \le e^{-q\lambda}\|x\|^p + C_p\lambda,$$

which completes the proof. □

D.2. **Proofs for Appendix C: Local-Errors and Regularity.**

D.2.1. *Proof of Proposition 3.*

Proof. Let (X_t) and (Y_t) solve (LSDE) from $X_0 = x$ and $Y_0 = y$ respectively, driven by the same Brownian motion, and set $\Delta_t := X_t - Y_t$. Under this synchronous coupling the diffusion terms cancel, so

$$d\Delta_t = -\big(h(X_t) - h(Y_t)\big)\, dt.$$

Consequently by Itô's lemma,

$$\frac{d}{dt}\|\Delta_t\|^2 = -2\big\langle \Delta_t,\, h(X_t) - h(Y_t)\big\rangle.$$

By the one-sided Lipschitz condition (A3), $\langle h(X_t) - h(Y_t),\, \Delta_t\rangle \ge -K'\|\Delta_t\|^2$, hence

$$\frac{d}{dt}\|\Delta_t\|^2 \le 2K'\|\Delta_t\|^2.$$

Grönwall's lemma yields $\|\Delta_t\|^2 \le e^{2K't}\|x - y\|^2$ pointwise, and taking expectations and square roots gives

$$\|X_t - Y_t\|_{L^2} \le e^{K't}\|x - y\|, \qquad t \ge 0.$$

In particular, for $t \in [0, \lambda]$,

$$\|X_t - Y_t\|_{L^2} \le L\,\|x - y\|, \qquad L := e^{K'\lambda}.$$

Finally, if $\lambda \le \frac{\ln 2}{|K'|}$ then $|K'\lambda| \le \ln 2$, so $e^{K'\lambda} \in [\frac{1}{2}, 2]$. □

D.2.2. *Proof of Proposition 4.*

Proof. Let (X_t) and (Y_t) solve (LSDE) from $X_0 = x$ and $Y_0 = y$ respectively, driven by the same Brownian motion. Under this synchronous coupling the diffusion terms cancel, so

$$X_t - x - (Y_t - y) = -\int_0^t \big(h(X_s) - h(Y_s)\big)\, ds.$$

As in the proof of Proposition 3, the same coupling gives the pathwise contraction estimate

$$\|X_s - Y_s\| \le e^{K's}\|x - y\|, \qquad s \ge 0.$$

By the Cauchy–Schwarz inequality applied to the time integral,

$$\Big\|\int_0^t \big(h(X_s) - h(Y_s)\big)\, ds\Big\|^2 \le t\int_0^t \|h(X_s) - h(Y_s)\|^2\, ds,$$

and the polynomial Lipschitz condition (A2) gives,

$$\|h(X_s) - h(Y_s)\|^2 \le L'^2\big(1 + \|X_s\| + \|Y_s\|\big)^{2\ell'}\|X_s - Y_s\|^2.$$

Taking expectations and applying Cauchy–Schwarz,

$$\mathbb{E}\|h(X_s) - h(Y_s)\|^2 \le L'^2\big(\mathbb{E}(1 + \|X_s\| + \|Y_s\|)^{4\ell'}\big)^{1/2}\big(\mathbb{E}\|X_s - Y_s\|^4\big)^{1/2}.$$

Substituting the pathwise contraction estimate, which yields $\big(\mathbb{E}\|X_s - Y_s\|^4\big)^{1/2} \le e^{2K's}\|x-y\|^2$, and writing

$$M_s := \big(\mathbb{E}(1 + \|X_s\| + \|Y_s\|)^{4\ell'}\big)^{1/2},$$

we obtain

$$\|X_t - x - (Y_t - y)\|_{L^2}^2 \le t\, L'^2\, \|x-y\|^2 \int_0^t e^{2K's}\, M_s\, ds \le t\, L'^2\, \|x-y\|^2 \Big(\sup_{0\le s\le t} M_s\Big) \int_0^t e^{2K's}\, ds.$$

By Lemma 5 with $p = 2\ell'$,

$$\sup_{0\le s\le t} M_s \le \widetilde{C}_{\ell'}\big(1 + \|x\|^{2\ell'} + \|y\|^{2\ell'}\big).$$

For the remaining integral, if $t \le \frac{\ln 2}{|K'|}$, then $e^{2K's} \le 4$ on $[0,t]$, yielding the bound $t\int_0^t e^{2K's}\, ds \le 4t^2 \le C_{K'}^2\, t^2$.

Combining these two bounds and taking square roots, gives $\big(1+\|x\|^{2\ell'}+\|y\|^{2\ell'}\big)^{1/2} \le 1+\|x\|^{\ell'}+\|y\|^{\ell'}$, hence

$$\|X_t - x - (Y_t - y)\|_{L^2} \le C_{\mathrm{cpl}}\, t\, \big(1 + \|x\|^{\ell'} + \|y\|^{\ell'}\big)\, \|x-y\|,$$

for a constant $C_{\mathrm{cpl}} > 0$ depending only on L', ℓ', K', and β. □

D.3. **Proofs for Appendix D: Auxiliary Results for kTULA.**

D.3.1. *Proof of Proposition 6.*

Proof. Let $\widehat{X}_t := x - t\, h_\lambda(x) + \sqrt{\frac{2}{\beta}}\, B_t$ for $t \in [0,\lambda]$ denote the one-step interpolation started at x, and let $(X_t)_{t\in[0,\lambda]}$ solve (LSDE) from $X_0 = x$ driven by the same Brownian motion. Subtracting the two mild forms, the Brownian terms cancel and

$$\widehat{X}_\lambda - X_\lambda = \int_0^\lambda \big(h(X_s) - h_\lambda(x)\big)\, ds.$$

Adding and subtracting $h(x)$ inside the integral and taking expectations,

$$\big\|\mathbb{E}[\widehat{X}_\lambda - X_\lambda]\big\| \le \lambda\, \|h_\lambda(x) - h(x)\| + \int_0^\lambda \|\mathbb{E}[h(X_s)] - h(x)\|\, ds. \tag{36}$$

By Lemma 4(iv), the taming error satisfies $\|h(x) - h_\lambda(x)\| \le C(1+\|x\|^{3(\ell+1)})\lambda$, so the first term in (36) is bounded by $C(1 + \|x\|^{3(\ell+1)})\lambda^2$.

It remains to bound the integrand $\|\mathbb{E}[h(X_s)] - h(x)\|$. By Lemma 8 under Assumption (A5), h admits the first-order expansion

$$h(X_s) - h(x) = J(h)(x)\,(X_s - x) + R_s, \qquad \|R_s\| \le \frac{L''}{2}\big(1 + \|x\| + \|X_s\|\big)^{\ell''}\|X_s - x\|^2,$$

so that, taking expectations,

$$\|\mathbb{E}[h(X_s)] - h(x)\| \le \|J(h)(x)\|\, \|\mathbb{E}[X_s - x]\| + \mathbb{E}\|R_s\|. \tag{37}$$

We estimate the two contributions separately.

For the first, the mild form gives $\mathbb{E}[X_s - x] = -\int_0^s \mathbb{E}[h(X_r)]\, dr$, hence $\|\mathbb{E}[X_s - x]\| \le \int_0^s \mathbb{E}\|h(X_r)\|\, dr$. Assumption (A1) bounds $\mathbb{E}\|h(X_r)\| \le L(1 + \mathbb{E}\|X_r\|^{2\ell})$, and Lemma 5 controls the moment, yielding

$$\|\mathbb{E}[X_s - x]\| \le C\big(1 + d^\ell + \|x\|^{2\ell}\big)\, s.$$

Assumption (A1) also gives the Jacobian growth bound $\|J(h)(x)\| \le L(1 + \|x\|^{2\ell})$, and combining the two,

$$\|J(h)(x)\|\, \|\mathbb{E}[X_s - x]\| \le C\big(1 + d^{2\ell} + \|x\|^{4\ell}\big)\, s. \tag{38}$$

For the remainder $\mathbb{E}\|R_s\|$, Cauchy–Schwarz gives

$$\mathbb{E}\|R_s\| \le \frac{L''}{2}\big(\mathbb{E}(1 + \|x\| + \|X_s\|)^{2\ell''}\big)^{1/2}\big(\mathbb{E}\|X_s - x\|^4\big)^{1/2}.$$

The first factor is bounded by $C(1 + d^{\ell''/2} + \|x\|^{\ell''})$ via Lemma 5 with $2p = 2\ell''$. For the second, the mild form $X_s - x = -\int_0^s h(X_r)\,dr + \sqrt{\frac{2}{\beta}}\,B_s$, Jensen's inequality, Assumption (A1) together with Lemma 5 (with $2p = 8\ell$), and the Gaussian moment $\mathbb{E}\|B_s\|^4 \le Cd^2s^2$, give, using $s \le \lambda \le 1$,

$$\mathbb{E}\|X_s - x\|^4 \le C\big(1 + d^{4\ell} + \|x\|^{8\ell}\big)\, s^2.$$

Combining the two factors,

$$\mathbb{E}\|R_s\| \le C\big(1 + d^{\ell''/2} + \|x\|^{\ell''}\big)\big(1 + d^{2\ell} + \|x\|^{4\ell}\big)\, s. \tag{39}$$

Substituting (38) and (39) into (37) and retaining the dominant powers,

$$\|\mathbb{E}[h(X_s)] - h(x)\| \le C\big(1 + d^{M/2} + \|x\|^M\big)\, s, \qquad s \in [0, \lambda], \qquad M := \max\{3(\ell + 1),\, 4\ell + \ell''\}.$$

Integrating over $[0, \lambda]$ contributes a factor λ^2, and combining with the taming term in (36) yields

$$\big\|\mathbb{E}[\widehat{X}_\lambda - X_\lambda]\big\| \le C\big(1 + d^{M/2} + \|x\|^M\big)\lambda^2,$$

for a constant $C > 0$ independent of x, d, and λ. □

D.3.2. *Proof of Proposition 7.*

Proof. Let $\widehat{X}_s := x - s\,h_\lambda(x) + \sqrt{\frac{2}{\beta}}\,B_s$ and $E_s := \widehat{X}_s - X_s$ for $s \in [0, \lambda]$, where (X_s) solves (LSDE) from $X_0 = x$ under the same Brownian motion. Subtracting the two mild forms, the Brownian terms cancel and $dE_s = -(h_\lambda(x) - h(X_s))\,ds$, so by Itô's lemma

$$\frac{d}{ds}\|E_s\|^2 = -2\big\langle E_s,\, h_\lambda(x) - h(X_s)\big\rangle.$$

By Young's inequality with parameter $\varepsilon = \lambda^{-1}$,

$$\frac{d}{ds}\|E_s\|^2 \le \frac{1}{\lambda}\|E_s\|^2 + \lambda\,\|h(X_s) - h_\lambda(x)\|^2,$$

and integrating over $[0, t]$ and taking expectations,

$$\mathbb{E}\|E_t\|^2 \le \frac{1}{\lambda}\int_0^t \mathbb{E}\|E_s\|^2\,ds + \lambda\int_0^t \mathbb{E}\|h(X_s) - h_\lambda(x)\|^2\,ds. \tag{40}$$

We bound the second integrand. Splitting

$$\|h(X_s) - h_\lambda(x)\|^2 \le 2\|h(X_s) - h(x)\|^2 + 2\|h(x) - h_\lambda(x)\|^2,$$

we treat the two terms separately.

For the drift increment, Assumption (A2) and the Cauchy–Schwarz inequality give

$$\mathbb{E}\|h(X_s) - h(x)\|^2 \le C\big(\mathbb{E}(1 + \|x\| + \|X_s\|)^{4\ell'}\big)^{1/2}\big(\mathbb{E}\|X_s - x\|^4\big)^{1/2}.$$

The first factor is bounded by $C(1 + d^{\ell'} + \|x\|^{2\ell'})$ via Lemma 5 with $2p = 4\ell'$. For the second, the mild form $X_s - x = -\int_0^s h(X_r)\,dr + \sqrt{\frac{2}{\beta}}\,B_s$, Jensen's inequality, Assumption (A1) together with Lemma 5 (with $2p = 8\ell$), and the Gaussian moment $\mathbb{E}\|B_s\|^4 \le Cd^2s^2$ yield, using $s \le \lambda \le 1$,

$$\mathbb{E}\|X_s - x\|^4 \le C\big(1 + d^{4\ell} + \|x\|^{8\ell}\big)\, s^2.$$

Combining the two factors and retaining the dominant powers,

$$\mathbb{E}\|h(X_s) - h(x)\|^2 \le C\big(1 + d^{(2\ell' + 4\ell)/2} + \|x\|^{2\ell' + 4\ell}\big)\, s. \tag{41}$$

For the taming term, Lemma 4(iv) gives

$$\|h(x) - h_\lambda(x)\|^2 \le C\big(1 + \|x\|^{6(\ell+1)}\big)\lambda^2. \tag{42}$$

Writing $M := \max\{3(\ell + 1),\, 2\ell' + 4\ell\}$, the exponent in (42) is $6(\ell + 1) = 2 \cdot 3(\ell + 1) \le 2M$ and that in (41) is $2\ell' + 4\ell \le M$ with dimension exponent $(2\ell' + 4\ell)/2 \le M/2$. Integrating both over $[0, t]$, which contributes a factor λ^2, using $s \le \lambda \le 1$,

$$\int_0^t \mathbb{E}\|h(X_s) - h_\lambda(x)\|^2\,ds \le C\big(1 + d^{M/2} + \|x\|^M\big)\lambda^2. \tag{43}$$

Substituting (43) into (40) gives

$$\mathbb{E}\|E_t\|^2 \le \frac{1}{\lambda}\int_0^t \mathbb{E}\|E_s\|^2\,ds + C\big(1 + d^{M/2} + \|x\|^M\big)\lambda^3, \qquad t \in [0,\lambda],$$

and Grönwall's inequality on $[0,\lambda]$, where the prefactor of the integral term is λ^{-1} and the interval has length at most λ, yields

$$\mathbb{E}\|E_t\|^2 \le C\big(1 + d^{M/2} + \|x\|^M\big)\lambda^3\, e^{t/\lambda}, \qquad t \in [0,\lambda].$$

Evaluating at $t = \lambda$, so that $e^{t/\lambda} \le e$, and taking square roots,

$$\|\widehat{X}_\lambda - X_\lambda\|_{L^2} \le C\big(1 + d^{M/2} + \|x\|^M\big)\lambda^{3/2},$$

for a constant $C > 0$ independent of x, d, and λ. □

D.3.3. *Proof of Proposition 8.*

Proof. We first estimate the divergence $\mathrm{KL}(\delta_x\widehat{P}\|\delta_x P)$. By Girsanov's theorem, (see similar calculations in [3]) one obtains

$$\mathrm{KL}(\delta_x\widehat{P}\|\delta_x P) \le \mathbb{E}\int_0^\lambda |h_\lambda(x) - h(x_t)|^2\,dt$$

where x_t is the law of continuous time interpolation of the algorithm at time t. Bounding

$$\mathbb{E}\int_0^\lambda |h_\lambda(x) - h(x_t)|^2\,dt \le 2\mathbb{E}\int_0^\lambda |h(x) - h(x_t)|^2\,dt + 2\lambda|h(x) - h_\lambda(x)|^2$$

By the local Lipschitz continuity, by standard calculations done before, the first term is $\mathcal{O}(\lambda^2)$ and the second term is the taming error which is again $\mathcal{O}(\lambda^2)$. Using the weak triangle inequality for KL divergence

$$KL(\delta_x\hat{P}||\delta_y\hat{P}) \le KL(\delta_x\hat{P}||\delta_x P) + R_2(\delta_x P||\delta_y P)$$

using the Renyi regularity of the Langevin kernel in Proposition 5 yields the result. □

D.3.4. *Proof of Lemma 10.*

Proof. By convexity of $\|\cdot\|^p$ and $\widetilde{Y}_n = (1-\eta_n)Y_n' + \eta_n\widehat{X}_n$,

$$\mathbb{E}\|\widetilde{Y}_n\|^p \le (1-\eta_n)\,\mathbb{E}\|Y_n'\|^p + \eta_n\widehat{M}_p, \qquad \widehat{M}_p := \sup_{n\ge 0}\mathbb{E}\|\widehat{X}_n\|^p.$$

Since $Y_n' \sim P_\lambda(\widetilde{Y}_{n-1},\cdot)$, Lemma 6 gives $\mathbb{E}\|Y_n'\|^p \le e^{-q\lambda}\mathbb{E}\|\widetilde{Y}_{n-1}\|^p + C\lambda$, and since $1-\eta_n \le 1$,

$$\mathbb{E}\|\widetilde{Y}_n\|^p \le e^{-q\lambda}\,\mathbb{E}\|\widetilde{Y}_{n-1}\|^p + C\lambda + \widehat{M}_p.$$

Iterating this recursion and summing the geometric series $\sum_{k=0}^{n-1} e^{-q\lambda k} \le \frac{1}{1-e^{-q\lambda}}$ yields

$$\sup_{n\ge 0}\mathbb{E}\|\widetilde{Y}_n\|^p \le \mathbb{E}\|\widetilde{Y}_0\|^p + \frac{C\lambda + \widehat{M}_p}{1-e^{-q\lambda}} < \infty,$$

which is finite by Lemma 7. □

D.3.5. *Proof of Lemma 11.*

Proof. By the W_2 coupling lift, $d_{n+1}^2 \le \mathbb{E}[W_2^2(\delta_{\widehat{X}_n}\widehat{P}, \delta_{\widetilde{Y}_n}P)]$, and Lemma 9 bounds the right-hand side by

$$\mathbb{E}\big[L^2\|\widehat{X}_n - \widetilde{Y}_n\|^2 + 2\,a_1(\widehat{X}_n,\widetilde{Y}_n)\,\|\widehat{X}_n - \widetilde{Y}_n\| + a_0(\widehat{X}_n)^2\big].$$

Since $\widetilde{Y}_n = (1-\eta_n)Y_n' + \eta_n\widehat{X}_n$, we have $\widehat{X}_n - \widetilde{Y}_n = (1-\eta_n)(\widehat{X}_n - Y_n')$, hence $\mathbb{E}\|\widehat{X}_n - \widetilde{Y}_n\|^2 = (1-\eta_n)^2 d_n^2$. Applying Cauchy–Schwarz to the cross term and the definitions

$$\overline{a}_1 := \sup_{0\le k<N}\|a_1(\widehat{X}_k,\widetilde{Y}_k)\|_{L^2}, \qquad \overline{a}_0 := \sup_{0\le k<N}\|a_0(\widehat{X}_k)\|_{L^2},$$

gives $\mathbb{E}[a_1(\widehat{X}_n,\widetilde{Y}_n)\|\widehat{X}_n - \widetilde{Y}_n\|] \le \overline{a}_1(1-\eta_n)d_n$ and $\mathbb{E}[a_0(\widehat{X}_n)^2] \le \overline{a}_0^2$. Substituting these three bounds yields

$$d_{n+1}^2 \le L^2(1-\eta_n)^2 d_n^2 + 2(1-\eta_n)\overline{a}_1 d_n + \overline{a}_0^2,$$

as claimed. □

D.4. **Proofs for Appendix D: Auxiliary Results for tRLMC.**

D.4.1. *Proof of Lemma 12(i).*

Proof. Rewrite (tRLMC) as

$$\bar{Y}_{n+1} = \bar{Y}_n - \lambda h_\lambda(\bar{Y}_n) + \sqrt{\frac{2}{\beta}}\,\Delta W_{n+1} - \lambda\,\Xi_{n+1}, \qquad \Xi_{n+1} := h_\lambda(\bar{Y}^\tau_{n+1}) - h_\lambda(\bar{Y}_n). \tag{44}$$

Expanding $\|\bar{Y}_{n+1}\|^2$, taking expectations, and using that ΔW_{n+1} is centred and independent of $\bar{Y}_n$, so the three Brownian cross terms vanish, the surviving inner products are estimated by Cauchy–Schwarz, giving

$$\begin{aligned}\mathbb{E}\|\bar{Y}_{n+1}\|^2 \le{}& \mathbb{E}\|\bar{Y}_n\|^2 - 2\lambda\,\mathbb{E}\langle \bar{Y}_n, h_\lambda(\bar{Y}_n)\rangle + \lambda^2\mathbb{E}\|h_\lambda(\bar{Y}_n)\|^2 + \frac{2}{\beta}\mathbb{E}\|\Delta W_{n+1}\|^2 \\ &+ 2\lambda\,\mathbb{E}\big[\|\bar{Y}_n\|\,\|\Xi_{n+1}\|\big] + 2\lambda^2\,\mathbb{E}\big[\|h_\lambda(\bar{Y}_n)\|\,\|\Xi_{n+1}\|\big] \\ &+ 2\lambda\sqrt{\frac{2}{\beta}}\,\mathbb{E}\big[\|\Delta W_{n+1}\|\,\|\Xi_{n+1}\|\big] + \lambda^2\mathbb{E}\|\Xi_{n+1}\|^2.\end{aligned} \tag{45}$$

By the dissipativity bound of Lemma 4(i), $-2\lambda\,\mathbb{E}\langle \bar{Y}_n, h_\lambda(\bar{Y}_n)\rangle \le -2a\lambda\,\mathbb{E}\|\bar{Y}_n\|^2 + 2b\lambda$. Applying Young's inequality to the cross term $2\lambda\,\mathbb{E}[\|\bar{Y}_n\|\,\|\Xi_{n+1}\|] \le \frac{a}{2}\lambda\,\mathbb{E}\|\bar{Y}_n\|^2 + \frac{2\lambda}{a}\mathbb{E}\|\Xi_{n+1}\|^2$, and to the two remaining cross terms with parameter $\varepsilon = 1$, we collect

$$\mathbb{E}\|\bar{Y}_{n+1}\|^2 \le \Big(1 - \tfrac{3a}{2}\lambda\Big)\mathbb{E}\|\bar{Y}_n\|^2 + 2\lambda^2\mathbb{E}\|h_\lambda(\bar{Y}_n)\|^2 + \frac{4}{\beta}\mathbb{E}\|\Delta W_{n+1}\|^2 + 2b\lambda + \Big(\tfrac{2\lambda}{a} + 3\lambda^2\Big)\mathbb{E}\|\Xi_{n+1}\|^2 \tag{46}$$

It remains to bound the two h_λ-dependent terms. By Lemma 4(ii),

$$\lambda^2\mathbb{E}\|h_\lambda(\bar{Y}_n)\|^2 \le 8a^2\lambda^2\,\mathbb{E}\|\bar{Y}_n\|^2 + 8L^2\lambda. \tag{47}$$

For Ξ_{n+1}, the Lipschitz bound of Lemma 4(iii) gives $\|\Xi_{n+1}\|^2 \le L_0^2\lambda^{-1}\|\bar{Y}^\tau_{n+1} - \bar{Y}_n\|^2$, and from $\bar{Y}^\tau_{n+1} - \bar{Y}_n = -\lambda\tau_{n+1}h_\lambda(\bar{Y}_n) + \sqrt{\frac{2}{\beta}}\,\Delta W^\tau_{n+1}$ with $\tau_{n+1} \in (0,1)$,

$$\mathbb{E}\|\Xi_{n+1}\|^2 \le 2L_0^2\lambda\,\mathbb{E}\|h_\lambda(\bar{Y}_n)\|^2 + \frac{4L_0^2}{\beta}\lambda^{-1}\mathbb{E}\|\Delta W^\tau_{n+1}\|^2.$$

Conditionally on τ_{n+1}, ΔW^τ_{n+1} is a Gaussian increment over an interval of length $\tau_{n+1}\lambda$, so $\mathbb{E}[\|\Delta W^\tau_{n+1}\|^2 \mid \tau_{n+1}] = \tau_{n+1}\lambda d$, and taking expectation over $\tau_{n+1} \sim \mathcal{U}(0,1)$ gives $\mathbb{E}\|\Delta W^\tau_{n+1}\|^2 = \frac{1}{2}\lambda d \le \lambda d$. With $\mathbb{E}\|\Delta W_{n+1}\|^2 = \lambda d$ and $0 < \lambda \le 1$, combining the last two displays with (47) yields

$$\Big(\tfrac{2\lambda}{a} + 3\lambda^2\Big)\mathbb{E}\|\Xi_{n+1}\|^2 \le C\lambda^2\,\mathbb{E}\|\bar{Y}_n\|^2 + C\lambda d, \tag{48}$$

for a constant $C > 0$ depending only on a, L, L_0, β.

Substituting (47) and (48) into (46) and using $\mathbb{E}\|\Delta W_{n+1}\|^2 = \lambda d$,

$$\mathbb{E}\|\bar{Y}_{n+1}\|^2 \le \Big(1 - \tfrac{3a}{2}\lambda + C\lambda^2\Big)\mathbb{E}\|\bar{Y}_n\|^2 + C\lambda d.$$

Set

$$\lambda^{tRLMC}_{2,\max} := \min\Big\{1,\ \tfrac{1}{8a},\ \tfrac{a}{2CL_0^2}\Big\}.$$

For $\lambda \in (0, \lambda^{tRLMC}_{2,\max}]$ one has $C\lambda^2 \le C\lambda \le \frac{a}{2}$, so $1 - \frac{3a}{2}\lambda + C\lambda^2 \le 1 - a\lambda$, hence, with $\mu := a$,

$$\mathbb{E}\|\bar{Y}_{n+1}\|^2 \le (1 - \mu\lambda)\,\mathbb{E}\|\bar{Y}_n\|^2 + C\lambda d.$$

Iterating, and using $\sum_{k=0}^{n-1}(1-\mu\lambda)^k \le (\mu\lambda)^{-1}$ and $(1-\mu\lambda)^n \le e^{-\mu n\lambda}$,

$$\mathbb{E}\|\bar{Y}_n\|^2 \le e^{-\mu n\lambda}\,\mathbb{E}\|x_0\|^2 + \frac{C}{\mu}\,d,$$

so the claim holds with $M_2 := C/\mu$. □

D.4.2. *Proof of Lemma 12(ii).*

Proof. Fix $p \in [2, 4\ell] \cap \mathbb{N}$ and $0 < \lambda \le \lambda_{p,\max}^{tRLMC}$, with the threshold $\lambda_{p,\max}^{tRLMC} \le 1$ to be specified. Let $C_p > 0$ denote a generic constant depending only on p, a, b, L, L_0, β. Defining the residual $\Xi_{n+1} := h_\lambda(\bar{Y}_{n+1}^\tau) - h_\lambda(\bar{Y}_n)$ yields the increment decomposition

$$\bar{Y}_{n+1} - \bar{Y}_n = -\lambda h_\lambda(\bar{Y}_n) + \sqrt{\frac{2}{\beta}}\,\Delta W_{n+1} - \lambda\,\Xi_{n+1}.$$

By Lemma 4(ii)(iii) and the inner increment $\bar{Y}_{n+1}^\tau - \bar{Y}_n = -\lambda\tau_{n+1}h_\lambda(\bar{Y}_n) + \sqrt{\frac{2}{\beta}}\,\Delta W_{n+1}^\tau$, the residual satisfies $\lambda\|\Xi_{n+1}\| \le C_p\lambda^{3/2}\|\bar{Y}_n\| + C_p\lambda + C_p\lambda^{1/2}\|\Delta W_{n+1}^\tau\|$. Since $\Delta W_{n+1} \sim \mathcal{N}(0, \lambda I_d)$ and $\Delta W_{n+1}^\tau \sim \mathcal{N}(0, \tau_{n+1}\lambda I_d)$, taking conditional Gaussian moments yields

$$\mathbb{E}\big[\|\bar{Y}_{n+1}-\bar{Y}_n\|^2 \mid \bar{Y}_n\big] \le C_p\lambda^2\|\bar{Y}_n\|^2+C_p\lambda(1+d), \qquad \mathbb{E}\big[\|\bar{Y}_{n+1}-\bar{Y}_n\|^{2p} \mid \bar{Y}_n\big] \le C_p\lambda^{2p}\|\bar{Y}_n\|^{2p}+C_p\lambda^p(1+d^p).$$

The polynomial expansion $\|x+z\|^{2p} \le \|x\|^{2p} + 2p\|x\|^{2p-2}\langle x, z\rangle + C_p(\|x\|^{2p-2}\|z\|^2 + \|z\|^{2p})$ furnishes the conditional bound

$$\begin{aligned}\mathbb{E}\big[\|\bar{Y}_{n+1}\|^{2p} \mid \bar{Y}_n\big] &\le \|\bar{Y}_n\|^{2p} + 2p\|\bar{Y}_n\|^{2p-2}\mathbb{E}\big[\langle \bar{Y}_n, \bar{Y}_{n+1} - \bar{Y}_n\rangle \mid \bar{Y}_n\big] \\ &\quad + C_p\|\bar{Y}_n\|^{2p-2}\mathbb{E}\big[\|\bar{Y}_{n+1} - \bar{Y}_n\|^2 \mid \bar{Y}_n\big] + C_p\mathbb{E}\big[\|\bar{Y}_{n+1} - \bar{Y}_n\|^{2p} \mid \bar{Y}_n\big].\end{aligned}$$

Since the Brownian increment is zero-mean, the conditional drift is bounded via the dissipativity condition of Lemma 4(i) and Young's inequality, yielding

$$2p\|\bar{Y}_n\|^{2p-2}\mathbb{E}\big[\langle \bar{Y}_n, \bar{Y}_{n+1} - \bar{Y}_n\rangle \mid \bar{Y}_n\big] \le -\tfrac{3pa}{2}\lambda\|\bar{Y}_n\|^{2p} + C_p\lambda^{3/2}\|\bar{Y}_n\|^{2p} + C_p\lambda(1 + d^p).$$

Inserting the drift and increment moment bounds directly into the Taylor expansion yields

$$\mathbb{E}\big[\|\bar{Y}_{n+1}\|^{2p} \mid \bar{Y}_n\big] \le \Big(1 - \tfrac{3pa}{2}\lambda + C_p\lambda^{3/2} + C_p\lambda^2\Big)\|\bar{Y}_n\|^{2p} + C_p\lambda(1 + d)\|\bar{Y}_n\|^{2p-2} + C_p\lambda(1 + d^p).$$

Defining the step-size threshold

$$\lambda_{p,\max}^{tRLMC} := \min\left\{\lambda_{2,\max}^{tRLMC},\ 1,\ \left(\frac{pa}{4C_p}\right)^2,\ \frac{pa}{4C_p}\right\}$$

ensures the leading bracket is strictly bounded by $1 - pa\lambda$, giving

$$\mathbb{E}\big[\|\bar{Y}_{n+1}\|^{2p} \mid \bar{Y}_n\big] \le (1 - pa\lambda)\|\bar{Y}_n\|^{2p} + C_p\lambda(1 + d)\|\bar{Y}_n\|^{2p-2} + C_p\lambda(1 + d^p).$$

Setting $M_p^2 := \frac{2C_p(1+d)}{pa}$, the strict polynomial decay $-pa\,r^{2p} + C_p(1 + d)\,r^{2p-2} \le -\frac{pa}{2}r^{2p}$ holds for $r \ge M_p$. Applying this for $\|\bar{Y}_n\| > M_p$ and uniformly bounding the terms for $\|\bar{Y}_n\| \le M_p$, it follows that for $c_p := \frac{pa}{2}$,

$$\mathbb{E}\big[\|\bar{Y}_{n+1}\|^{2p} \mid \bar{Y}_n\big] \le (1 - c_p\lambda)\|\bar{Y}_n\|^{2p} + C_p\lambda(1 + d^p).$$

Taking unconditional expectations and iterating the recurrence bounds the geometric sum, yielding

$$\mathbb{E}\|\bar{Y}_n\|^{2p} \le e^{-c_p n\lambda}\,\mathbb{E}\|x_0\|^{2p} + \frac{C_p}{c_p}(1 + d^p),$$

ensuring $\sup_{n\ge0}\mathbb{E}\|\bar{Y}_n\|^{2p} < \infty$ as claimed. □

D.4.3. *Proof of Lemma 13.*

Proof. Let

$$Y'_{n+1} = X(t_n, \widetilde{Y}_n; t_{n+1}), \qquad t_{n+1} - t_n = \lambda,$$

so that, by definition,

$$Y'_{n+1} \sim P_\lambda(\widetilde{Y}_n, \cdot).$$

By Lemma 6 for the Langevin diffusion over a time interval of length λ, there exist $q > 0$ and $C_p > 0$, depending on p, d, a, and b, such that for every $x \in \mathbb{R}^d$ and every $\lambda \in (0, 1]$,

$$\mathbb{E}\big[\|X(t_n, x; t_{n+1})\|^p\big] \le e^{-q\lambda}\|x\|^p + C_p\lambda.$$

Applying this with $x = \widetilde{Y}_n$, conditionally on $\widetilde{Y}_n$, yields

$$\mathbb{E}\Big[\|Y'_{n+1}\|^p \mid \widetilde{Y}_n\Big] = \mathbb{E}\Big[\|X(t_n, \widetilde{Y}_n; t_{n+1})\|^p \mid \widetilde{Y}_n\Big] \le e^{-q\lambda}\|\widetilde{Y}_n\|^p + C_p\lambda.$$

Taking expectations gives

$$\mathbb{E}\|Y'_{n+1}\|^p \le e^{-q\lambda}\mathbb{E}\|\widetilde{Y}_n\|^p + C_p\lambda.$$

This proves the claim. □

D.4.4. *Proof of Lemma 14.*

Proof. The argument is identical to that of Lemma 10, with $\widehat{X}_n$ replaced by the (tRLMC) iterate $\bar{Y}_n$ and Lemma 6 replaced by Lemma 13. By convexity of $\|\cdot\|^p$ and $\widetilde{Y}_n = (1-\eta_n)Y'_n + \eta_n\bar{Y}_n$,

$$\mathbb{E}\|\widetilde{Y}_n\|^p \le (1-\eta_n)\,\mathbb{E}\|Y'_n\|^p + \eta_n\bar{M}_p, \qquad \bar{M}_p := \sup_{n\ge 0}\mathbb{E}\|\bar{Y}_n\|^p,$$

which is finite by Lemma 12. Since $Y'_n \sim P_\lambda(\widetilde{Y}_{n-1},\cdot)$, Lemma 13 gives $\mathbb{E}\|Y'_n\|^p \le e^{-q\lambda}\mathbb{E}\|\widetilde{Y}_{n-1}\|^p + C_p\lambda$, and since $1-\eta_n \le 1$,

$$\mathbb{E}\|\widetilde{Y}_n\|^p \le e^{-q\lambda}\,\mathbb{E}\|\widetilde{Y}_{n-1}\|^p + C_p\lambda + \bar{M}_p.$$

Iterating and summing $\sum_{k=0}^{n-1} e^{-q\lambda k} \le (1-e^{-q\lambda})^{-1}$ yields

$$\sup_{n\ge 0}\mathbb{E}\|\widetilde{Y}_n\|^p \le \mathbb{E}\|\widetilde{Y}_0\|^p + \frac{C_p\lambda + \bar{M}_p}{1-e^{-q\lambda}} < \infty.$$

□

D.4.5. *Proof of Proposition 9.*

Proof. Let $\tau \sim \mathcal{U}(0,1)$ be independent of the driving Brownian motion. Writing the exact solution and the one-step numerical approximation (tRLMC) in their integral forms yields

$$X(t,x;t+\lambda) = x - \int_t^{t+\lambda} h(X(t,x;s))\,ds + \sqrt{\frac{2}{\beta}}\,(B_{t+\lambda} - B_t),$$

$$\bar{Y}_m(t,x;t+\tau\lambda) = x - \tau\lambda\, h_\lambda(x) + \sqrt{\frac{2}{\beta}}\,(B_{t+\tau\lambda} - B_t),$$

and

$$\bar{Y}(t,x;t+\lambda) = x - \lambda\, h_\lambda\big(\bar{Y}_m(t,x;t+\tau\lambda)\big) + \sqrt{\frac{2}{\beta}}\,(B_{t+\lambda} - B_t). \tag{49}$$

Subtracting the approximation from the exact solution, the Brownian increments cancel. Adding and subtracting the intermediate values $h(X(t,x;t+\tau\lambda))$ and $h_\lambda(X(t,x;t+\tau\lambda))$ inside the integral yields the decomposition

$$\begin{aligned} X(t,x;t+\lambda) - \bar{Y}(t,x;t+\lambda) = &-\int_t^{t+\lambda}\big(h(X(t,x;s)) - h(X(t,x;t+\tau\lambda))\big)\,ds \\ &+ \lambda\big(h(X(t,x;t+\tau\lambda)) - h_\lambda(X(t,x;t+\tau\lambda))\big) \\ &+ \lambda\big(h_\lambda(X(t,x;t+\tau\lambda)) - h_\lambda(\bar{Y}_m(t,x;t+\tau\lambda))\big). \end{aligned} \tag{50}$$

Taking the norm of the expectation and applying the triangle inequality yields the bound

$$\big\|\mathbb{E}[X(t,x;t+\lambda) - \bar{Y}(t,x;t+\lambda)]\big\| \le I_1 + I_2 + I_3,$$

where I_1, I_2, I_3 are the norms of the expectations of the respective terms above.

For the first term, the change of variables $s = t+u\lambda$ and Fubini's theorem, exploiting the independence of τ, imply

$$\mathbb{E}\Big[\int_t^{t+\lambda} h(X(t,x;s))\,ds\Big] = \lambda\,\mathbb{E}\big[h(X(t,x;t+\tau\lambda))\big].$$

Since the subtracted term $h(X(t,x;t+\tau\lambda))$ is independent of s, its integral over $[t,t+\lambda]$ equals to the same quantity. Consequently, the expectations cancel entirely, yielding $I_1 = 0$.

For the second term, setting $Z := X(t,x;t+\tau\lambda)$ and $R := 3(\ell+1)$, the taming error bound of Lemma 4(iv) yields $\|h(Z) - h_\lambda(Z)\| \le 2(L+a)\lambda\,(1+\|Z\|^{2R})^{1/2}$. Taking expectations and applying Jensen's inequality yields

$$I_2 \le 2(L+a)\lambda^2\big(1 + (\mathbb{E}\|Z\|^{2R})^{1/2}\big).$$

Applying the uniform moment bound of Lemma 5 ensures $\mathbb{E}\|Z\|^{2R} \le C(\|x\|^{2R} + d^R)$, which, upon taking the square root, simplifies the estimate to $I_2 \le C\lambda^2\big(1 + d^{R/2} + \|x\|^R\big)$.

For the third term, the local Lipschitz property of Lemma 4(iii) gives

$$I_3 \le L_0\,\lambda\,\mathbb{E}\big\|X(t,x;t+\tau\lambda) - \bar{Y}_m(t,x;t+\tau\lambda)\big\|. \tag{51}$$

Expanding this difference via the mild form on $[t, t+\tau\lambda]$ and canceling the Brownian increments provides

$$X(t,x;t+\tau\lambda) - \bar{Y}_m(t,x;t+\tau\lambda) = -\int_t^{t+\tau\lambda} \big(h(X(t,x;s)) - h(x)\big)\,ds + \tau\lambda\big(h_\lambda(x) - h(x)\big).$$

Taking norms and expectations, and noting $\mathbb{E}[\tau] = \frac{1}{2}$ for the deterministic second term, yields

$$\mathbb{E}\big\|X(t,x;t+\tau\lambda) - \bar{Y}_m(t,x;t+\tau\lambda)\big\| \le A_\lambda + B_\lambda, \tag{52}$$

where A_λ bounds the integral term and B_λ bounds the taming term. By Lemma 4(iv), the taming term satisfies

$$B_\lambda \le C\lambda^2(1 + \|x\|^R). \tag{53}$$

For A_λ, applying the Cauchy–Schwarz and Jensen's inequalities, gives

$$A_\lambda \le \Big(\lambda \int_t^{t+\lambda} \mathbb{E}\|h(X(t,x;s)) - h(x)\|^2\,ds\Big)^{1/2}.$$

By Assumption (A2) and Cauchy–Schwarz, the integrand is bounded by the product of the moment estimates $\big(\mathbb{E}(1 + \|X(t,x;s)\| + \|x\|)^{4\ell'}\big)^{1/2}$ and $\big(\mathbb{E}\|X(t,x;s) - x\|^4\big)^{1/2}$. Evaluating these via Assumption (A1) and Lemma 5 yields $\mathbb{E}\|h(X(t,x;s)) - h(x)\|^2 \le C(1 + d^R + \|x\|^{2R})(s-t)$. Integrating this over $[t, t+\lambda]$ and taking the square root establishes

$$A_\lambda \le C\big(1 + d^{R/2} + \|x\|^R\big)\lambda^{3/2}. \tag{54}$$

Substituting (52), (53), and (54) into (51) and using $\lambda^{5/2} \le \lambda^2$ for $\lambda \in (0,1)$ establishes

$$I_3 \le C\big(1 + d^{R/2} + \|x\|^R\big)\lambda^2.$$

Combining the bounds for I_1, I_2, and I_3 yields the final one-step estimate

$$\big\|\mathbb{E}[X(t,x;t+\lambda) - \bar{Y}(t,x;t+\lambda)]\big\| \le C\,\big(1 + d^{R/2} + \|x\|^R\big)\lambda^2,$$

which proves the claim. □

Lemma 18. *Under Assumptions (A1) and (A2), there exists a constant $C > 0$, independent of d, t, x, λ, such that for every $0 < \lambda \le 1$ and every $s_1, s_2 \in [t, t+\lambda]$,*

$$\mathbb{E}\|h(X(t,x;s_1)) - h(X(t,x;s_2))\|^2 \le C\big(1 + d^{2\ell'+4\ell} + \|x\|^{2\ell'+4\ell}\big)|s_1 - s_2|.$$

Proof. Let $s_1 < s_2$. Applying Assumption (A2) and the Cauchy–Schwarz inequality yields

$$\mathbb{E}\|h(X_{s_1}) - h(X_{s_2})\|^2 \le L'^2\big(\mathbb{E}(1 + \|X_{s_1}\| + \|X_{s_2}\|)^{4\ell'}\big)^{1/2}\big(\mathbb{E}\|X_{s_2} - X_{s_1}\|^4\big)^{1/2}.$$

By Lemma 5, the first factor is bounded by $C(1 + d^{2\ell'} + \|x\|^{2\ell'})$. For the second factor, the integral form $X_{s_2} - X_{s_1} = -\int_{s_1}^{s_2} h(X_r)\,dr + \sqrt{\frac{2}{\beta}}(B_{s_2} - B_{s_1})$, Jensen's inequality, Assumption (A1), and the Gaussian fourth moment yield the bound

$$\mathbb{E}\|X_{s_2} - X_{s_1}\|^4 \le C\big(1 + d^{8\ell} + \|x\|^{8\ell}\big)(s_2 - s_1)^2.$$

Taking the square root and multiplying the bounds yields the claim. □

D.4.6. *Proof of Proposition 10.*

Proof. The exact decomposition (50) of the error $X(t,x;t+\lambda) - \bar{Y}(t,x;t+\lambda)$ from the weak-error proof applies verbatim. Applying the elementary bound $\|a+b+c\|^2 \le 3(\|a\|^2 + \|b\|^2 + \|c\|^2)$ yields

$$\mathbb{E}\|X(t,x;t+\lambda) - \bar{Y}(t,x;t+\lambda)\|^2 \le 3(J_1 + J_2 + J_3),$$

where J_1, J_2, J_3 denote the squared L^2 norms of the corresponding three terms in (50).

By the Cauchy–Schwarz inequality applied to the time integral, the first term satisfies

$$J_1 \le \lambda \int_t^{t+\lambda} \mathbb{E}\|h(X(t,x;s)) - h(X(t,x;t+\tau\lambda))\|^2\,ds.$$

Applying the time-regularity bound of Lemma 18 to the integrand provides the upper bound $C(1 + d^{2\ell'+4\ell} + \|x\|^{2\ell'+4\ell})|s-(t+\tau\lambda)|$. Since $|s-(t+\tau\lambda)| \le \lambda$, integrating over the interval of length λ furnishes

$$J_1 \le C\big(1 + d^{2\ell'+4\ell} + \|x\|^{2\ell'+4\ell}\big)\lambda^3.$$

For the second term, applying Lemma 4(iv) and the uniform SDE moment bound of Lemma 5 yields

$$J_2 = \lambda^2\mathbb{E}\|h(X(t,x;t+\tau\lambda)) - h_\lambda(X(t,x;t+\tau\lambda))\|^2 \le C\lambda^4\,\mathbb{E}\big[1 + \|X(t,x;t+\tau\lambda)\|^{6(\ell+1)}\big].$$

Evaluating the moment bound directly yields $J_2 \le C\big(1 + d^{6(\ell+1)} + \|x\|^{6(\ell+1)}\big)\lambda^4$.

For the third term, the local Lipschitz property of Lemma 4(iii) gives

$$J_3 \le L_0^2\,\lambda\,\mathbb{E}\|X(t,x;t+\tau\lambda) - \bar{Y}_m(t,x;t+\tau\lambda)\|^2.$$

Expanding the difference via the identity (52), applying $\|u+v\|^2 \le 2(\|u\|^2 + \|v\|^2)$, and integrating via Cauchy–Schwarz yields

$$\mathbb{E}\|X(t,x;t+\tau\lambda) - \bar{Y}_m(t,x;t+\tau\lambda)\|^2 \le 2\lambda\int_t^{t+\lambda}\mathbb{E}\|h(X(t,x;s)) - h(x)\|^2\,ds + 2\lambda^2\|h_\lambda(x) - h(x)\|^2.$$

The time-regularity of Lemma 18 bounds the first integrand by $C(1 + d^{2\ell'+4\ell} + \|x\|^{2\ell'+4\ell})(s-t)$, which integrates to an order-λ^2 term. The taming error is bounded by $C(1 + \|x\|^{6(\ell+1)})\lambda^2$. Substituting these estimates into the J_3 inequality and using $\lambda^5 \le \lambda^4$ for $\lambda \in (0,1)$ establishes

$$J_3 \le C\big(1 + +d^{\widetilde{R}} + \|x\|^{\widetilde{R}}\big)\lambda^4.$$

Combining the bounds for J_1, J_2, and J_3, and absorbing the higher-order λ^4 terms into λ^3, yields the final one-step estimate

$$\mathbb{E}\|X(t,x;t+\lambda) - \bar{Y}(t,x;t+\lambda)\|^2 \le C\big(1 + d^{\widetilde{R}} + \|x\|^{\widetilde{R}}\big)\lambda^3,$$

where $\widetilde{R} := \max\{2\ell' + 4\ell,\ 6(\ell+1)\}$. This proves the claim. □

D.4.7. *Cross-regularity bound for tamed RLMC.*

Lemma 19. *Let $\hat{P}^u$ be the Markov kernel of* (tRLMC) *run with fixed u and $x \in \mathbb{R}^d$. For $u \le 1 - \lambda^2$ there holds*

$$TV(\delta_x\hat{P}^u \| \delta_x P) \le \sqrt{C(x)}(1-u)^{-\frac{1}{2}}\lambda$$

Proof. Conditioned on the Brownian motion $B_{\lambda u}$ the continuous time interpolation the tamed RLMC can be described by the following SDE:

$$\hat{X}_t = X_0 - \int_0^t h_\lambda(\hat{X}^+_{\lambda u})ds + \sqrt{\frac{2}{\beta}}\int_0^t d\tilde{B}_s$$

where $X_0 = x - \lambda u h_\lambda(\hat{X}^+_{\lambda u}) + \sqrt{\frac{2}{\beta}}B_{\lambda u}$. Let

$$Y_t = X_0 - \int_0^t h(Y_s)ds + \sqrt{\frac{2}{\beta}}\int_0^t d\tilde{B}_s.$$

By applying Girsanov's theorem one obtains

$$\begin{aligned} KL(\mathcal{L}(\hat{X}_{\lambda(1-u)})|\mathcal{L}(Y_{\lambda(1-u)})) &\le \int_0^{\lambda(1-u)}\mathbb{E}|h_\lambda(\hat{X}^+_{\lambda u}) - h(\hat{X}_s)|^2 ds \\ &\le 2\int_0^{\lambda(1-u)}\mathbb{E}|h_\lambda(\hat{X}^+_{\lambda u}) - h(\hat{X}^+_{\lambda u})|^2 ds + 2\int_0^{\lambda(1-u)}\mathbb{E}|h(\hat{X}^+_{\lambda u}) - h(\hat{X}_s)|^2 ds \\ &\le C\lambda^2 d^{2l+2} + \int_0^{\lambda(1-u)}\sqrt{\mathbb{E}|\hat{X}^+_{\lambda u} - \hat{X}_s|^4}\sqrt{\mathbb{E}(|\hat{X}^+_{\lambda u}| + |\hat{X}_s| + 1)^{4l}}ds \end{aligned} \tag{55}$$

where the first term is bounded by the taming approximation and the second by the local Lipschitz property of h. Noticing that

$$\mathbb{E}|\hat{X}_s - \hat{X}^+_{\lambda u}|^4 \le C(\mathbb{E}|\hat{X}_s - X_0|^4 + \mathbb{E}|\hat{X}^+_{\lambda u} - X_0|^4) \le Cd^{2l+2}\lambda^2 + \lambda^4\mathbb{E}|h_\lambda(\hat{X}^+_{\lambda u}) - h_\lambda(x)|^4 \le C(x)\lambda^2$$

and substituting in (55) one deduces that

$$TV(\mathcal{L}(\hat{X}_{\lambda(1-u)})|\mathcal{L}(Y_{\lambda(1-u)})) \leq \sqrt{2KL(\mathcal{L}(\hat{X}_{\lambda(1-u)})|\mathcal{L}(Y_{\lambda(1-u)}))} \leq \sqrt{C(x)}\lambda.$$

In addition, by the Renyi regularity of Langevin Kernel, there holds that if $\mu_0 = \mathcal{L}(X_0)$

$$TV((\delta_x P^{\lambda u})P^{\lambda(1-u)}, \mu_0 P^{\lambda(1-u)}) \leq \frac{C}{\sqrt{\lambda(1-u)}} W_2(\mu_0, \delta_x P^{\lambda u}) \leq \frac{C}{\lambda} W_2(\mu_0, \delta_x P^{\lambda u}) \tag{56}$$

where the last step is due to the assumption $u \leq 1-\lambda$.
It now remains to bound the Wasserstein distance. Taking

$$Y_{\lambda u} = x - \int_0^{\lambda u} h(Y_s)ds + \int_0^{\lambda u} \sqrt{2\beta^{-1}} dB_s$$

with same Brownian motion as X_0 one obtains that

$$\begin{aligned}
\mathbb{E}|X_0 - Y_{\lambda u}|^2 &\leq \mathbb{E}|\int_0^{\lambda u} h_\lambda(\hat{X}^+_{\lambda u}) - h(Y_s)ds|^2 \\
&\leq 2\lambda u \mathbb{E}\int_0^{\lambda u} |h_\lambda(\hat{X}^+_{\lambda u}) - h(\hat{X}^+_{\lambda u})|^2 ds + 2\lambda u \mathbb{E}\int_0^{\lambda u} |h(\hat{X}^+_{\lambda u}) - h(Y_s)|^2 ds \\
&\leq \lambda^4 \mathbb{E}|\hat{X}^+_{\lambda u}|^{2l+2} + 2\lambda u \int_0^{\lambda u} \sqrt{\mathbb{E}|\hat{X}^+_{\lambda u} - Y_s|^4}\sqrt{\mathbb{E}(1+|\hat{X}^+_{\lambda u}| + |Y_s|)^{2l}} ds
\end{aligned}$$

where the first term was bounded by taming error while the second is controlled by the local Lipschitz continuity and Cauchy Swartz inequality. Since

$$\mathbb{E}|Y_s - \hat{X}^+_{\lambda u}|^4 \leq 8\mathbb{E}|\int_0^s h(Y_r)dr|^4 + 8\mathbb{E}|\int_0^{\lambda u} h_\lambda(x)dr|^4 \leq (\lambda u)^3 \int_0^{\lambda u} \mathbb{E}|h(Y_r)|^4 dr + 8(\lambda u)^4 |x|^{4l+4}$$

Since the moment bounds of Y_r can be controlled by the moment bounds of the initial condition x while the same is true for the moment bounds of $\hat{X}^+_{\lambda u}$, one deduces that

$$\mathbb{E}|X_0 - Y_{\lambda u}|^2 \leq \lambda^4 C(x)$$

where $C(x)$ depends polynomially on x. Plugging this into (56) yields

$$TV((\delta_x P^{\lambda u})P^{\lambda(1-u)}||\mu_0 P^{\lambda(1-u)}) \leq C_{TV}(x)\lambda.$$

□

Lemma 20. *There holds*

$$TV(\delta_x P^{RLMC}, \delta_x P) \leq C_{TV}(x)\lambda$$

Proof. Conditioning on the uniform random variable u of the randomized scheme,

$$\begin{aligned}
TV(\delta_x P^{RLMC}, \delta_x P) &\leq \mathbb{E}_u TV(\delta_x \hat{P}^u, \delta_x P) \\
&= \int_0^{1-\lambda} TV(\delta_x \hat{P}^u, \delta_x P)du + \int_{1-\lambda}^1 TV(\delta_x \hat{P}^u, \delta_x P)du \\
&\leq \int_0^{1-\lambda} C(x)\lambda du + \lambda \\
&\leq C_{TV}(x)\lambda
\end{aligned}$$

where the first term is given by Lemma 19 □

D.4.8. *TV-shifted bound.*

Proof of Lemma 3. Since $\widehat{\mu}_N = \nu'_N$, $Y'_N \sim \widehat{P}(\widetilde{Y}_{N-1}, \cdot)$, and $Y_N \sim P(Y_{N-1}, \cdot)$, applying the triangle inequality yields the decomposition

$$\mathrm{TV}(\widehat{\mu}_N, \nu_N) = \mathrm{TV}\big(\mathrm{Law}(Y'_N), \mathrm{Law}(Y_N)\big) \leq I_1 + I_2 + I_3,$$

where

$$I_1 := \mathbb{E}\big[\mathrm{TV}\big(\widehat{P}(\widetilde{Y}_{N-1},\cdot), P(\widetilde{Y}_{N-1},\cdot)\big)\big],$$
$$I_2 := \mathbb{E}\big[\mathrm{TV}\big(P(\widetilde{Y}_{N-1},\cdot), P(Y'_{N-1},\cdot)\big)\big],$$
$$I_3 := \mathrm{TV}(\nu'_{N-1}P, \nu_{N-1}P).$$

By the local error estimate (30), the first term satisfies $I_1 \le C_{\mathrm{TV}}(x)\lambda$.

For the second term, applying Pinsker's inequality and the regularity condition (21) almost surely yields $\mathrm{TV}\big(P(\widetilde{Y}_{N-1},\cdot), P(Y'_{N-1},\cdot)\big) \le \sqrt{\frac{c}{2}\|\widetilde{Y}_{N-1} - Y'_{N-1}\|^2}$. Taking the expectation and applying Jensen's inequality provides

$$I_2 \le \sqrt{\frac{c}{2}\,\mathbb{E}\|\widetilde{Y}_{N-1} - Y'_{N-1}\|^2}.$$

Since $\eta_{N-1} = 1$, the identity $\widetilde{Y}_{N-1} = \widehat{X}_{N-1}$ holds exactly, bounding the second term by $I_2 \le \sqrt{\frac{c}{2}}\, d_{N-1}$.

For the third term, the contraction of total variation under Markov kernels implies $I_3 \le \mathrm{TV}(\nu'_{N-1}, \nu_{N-1})$. Applying Pinsker's inequality and telescoping the shifted chain rule up to time $N-1$ provides the relative entropy bound

$$\mathrm{KL}(\nu'_{N-1}\|\nu_{N-1}) \le \sum_{n=0}^{N-2} \mathbb{E}\Big[\mathrm{KL}\big(P(\widetilde{Y}_n,\cdot)\,\big\|\,P(Y'_n,\cdot)\big)\Big].$$

Invoking (21) and the interpolation identity $\widetilde{Y}_n - Y'_n = \eta_n(\widehat{X}_n - Y'_n)$ evaluates the sum as

$$\mathrm{KL}(\nu'_{N-1}\|\nu_{N-1}) \le c\sum_{n=0}^{N-2} \mathbb{E}\|\widetilde{Y}_n - Y'_n\|^2 = c\sum_{n=0}^{N-2} \eta_n^2\, d_n^2.$$

Substituting this into the Pinsker bound yields $I_3 \le \sqrt{\frac{c}{2}\sum_{n=0}^{N-2}\eta_n^2\, d_n^2}$.

Combining the bounds for I_1, I_2, and I_3 establishes (35). □

D.5. **Proofs of the Main Results for kTULA.**

D.5.1. *Proof of Theorem 1.*

Proof. By Propositions 3, 5, and 8, together with the one-step weak and strong error bounds, the quantities of the KL framework 4 satisfy, uniformly over $0 \le n < N$,

$$\overline{a}_0^2 = O(\lambda^3), \qquad \overline{a}_1^2 = O(\lambda^4), \qquad c = O(\lambda^{-1}), \qquad c' = O(\lambda^{-1}).$$

Set $L := e^{K'\lambda}$, $\bar{N} := N \wedge (1-L)_+^{-1}$, and $\bar{b} := \sup_{0\le n<N}\|b\|_{L^2(\widehat{\mu}_n)}$. By hypothesis $L \in [\frac{1}{2}, 2]$, and $\bar{N} \le N$ in both regimes: $\bar{N} = N$ when $L \ge 1$, and $\bar{N} \le (1-L)^{-1} \le N$ when $L < 1$. For $L \ne 1$, the Wasserstein coefficient satisfies

$$\frac{L^{-1}-1}{L^{-N}-1} = \frac{L-1}{L^N-1} \in [0,1],$$

the bound holding when $L \ge 1$, since then $L^N - 1 \ge L - 1 \ge 0$, and when $L \le 1$, then $L^N - 1 \le L - 1 < 0$, and the case $L = 1$ following by continuity.

Applying the estimate (29) and bounding the Wasserstein coefficient by 1,

$$\mathrm{KL}\big(\mu\widehat{P}^N \,\big\|\, \nu P^N\big) \;\lesssim\; (c+c')\Big[W_2^2(\mu,\nu) + \big((L-1)N \vee \log\bar{N}\big)\overline{a}_0^2 + \bar{N}\,\overline{a}_1^2\Big] + \bar{b}^{\,2}.$$

We bound the bracketed terms. Since $L \in [\frac{1}{2}, 2]$, the mean value theorem gives

$$|L-1| = |e^{K'\lambda} - 1| \le e^{|K'|\lambda}|K'|\lambda \le 2|K'|\lambda = O(\lambda),$$

so, together with $\log\bar{N} \le \log N$ and $\overline{a}_0^2 = O(\lambda^3)$, the second bracketed term is $O\big(((N\lambda)\vee\log N)\lambda^3\big)$, while $\bar{N}\,\overline{a}_1^2 \le N\,\overline{a}_1^2 = O(N\lambda^4)$. Multiplying the bracket by $(c+c') = O(\lambda^{-1})$ therefore contributes

$$O(\lambda^{-1})\,W_2^2(\mu,\nu) + O\big(((N\lambda)\vee\log N)\lambda^2\big) + O(N\lambda^3).$$

For the remaining term, Proposition 8 yields $b(x)^2 \le C\lambda^2(1 + \|x\|^{2\ell+2})$, so

$$\bar{b}^{\,2} \le C\lambda^2\Big(1 + \sup_{n\ge 0}\mathbb{E}\|\widehat{X}_n\|^{2\ell+2}\Big) \le C\lambda^2,$$

the last bound by Lemma 7 with $p = \ell + 1$.

Collecting these contributions,

$$\mathrm{KL}\big(\mu\widehat{P}^N \,\|\, \nu P^N\big) \le \frac{C_1}{\lambda}\, W_2^2(\mu,\nu) + C\big((N\lambda)\vee \log N\big)\lambda^2 + C\,N\lambda^3 + C\,\lambda^2.$$

Finally, since $0<\lambda\le 1$ and $N\ge 3$ (so $\log N\ge 1$), both λ^2 and $N\lambda^3=(N\lambda)\lambda^2$ are dominated by $((N\lambda)\vee\log N)\lambda^2$; absorbing them yields

$$\mathrm{KL}\big(\mu\widehat{P}^N \,\|\, \nu P^N\big) \le \frac{C_1}{\lambda}\, W_2^2(\mu,\nu) + C\big((N\lambda)\vee \log N\big)\lambda^2,$$

for a constant $C>0$ independent of N and λ. □

D.5.2. *Proof of Corollary 1.*

Proof. Applying the divergence bound $\mathrm{KL}(\rho\|\pi)\le \mathrm{KL}(\rho\|\eta)+R_2(\eta\|\pi)$ to $\rho=\mu\widehat{P}^N$, $\eta=\nu P^N$, and $\pi=\pi_\beta$ directly yields (8). To deduce (9), we bound the components of (8) using Theorem 1 and the Rényi-2 contraction of the diffusion semigroup under (A6), obtaining

$$\mathrm{KL}\big(\mu\widehat{P}^N\|\pi_\beta\big) \le \frac{C_1}{\lambda}\mathcal{W}_2^2(\mu,\nu) + C\big((N\lambda)\vee\log N\big)\lambda^2 + e^{-2C_{\mathrm{LSI}}N\lambda}R_2\big(\nu\|\pi_\beta\big).$$

Evaluating this at $\nu=\mu$ forces $\mathcal{W}_2^2(\mu,\mu)=0$, yielding (9) as claimed. □

D.5.3. *Proof of Proposition 1.*

Proof. By (9),

$$\mathrm{KL}\big(\mu\widehat{P}^N \,\|\, \pi_\beta\big) \le C\,\big((N\lambda)\vee\log N\big)\lambda^2 + e^{-2C_{\mathrm{LSI}}N\lambda}R_0,$$

so it suffices to make each term at most $\varepsilon/2$. Set

$$T := \frac{1}{2C_{\mathrm{LSI}}}\Big[\log\Big(\tfrac{2R_0}{\varepsilon}\Big)\Big]_+, \qquad \lambda := \min\Big\{\lambda_{\max}^{kTULA},\ \sqrt{\tfrac{\varepsilon}{2C\,(T\vee\log(3+T/\lambda))}}\Big\}, \qquad N := \max\big\{3,\lceil T/\lambda\rceil\big\}.$$

Since $N\ge\lceil T/\lambda\rceil$ we have $N\lambda\ge T$, hence $e^{-2C_{\mathrm{LSI}}N\lambda}R_0\le e^{-2C_{\mathrm{LSI}}T}R_0$. If $T>0$ then, by the definition of T, $e^{-2C_{\mathrm{LSI}}T}R_0=e^{-\log(2R_0/\varepsilon)}R_0=\varepsilon/2$. If $T=0$ then $R_0\le\varepsilon/2$ by the same definition. In either case

$$e^{-2C_{\mathrm{LSI}}N\lambda}R_0\le\frac{\varepsilon}{2}.$$

From $N\le 3+T/\lambda$ we get $N\lambda\le T+3\lambda$ and $\log N\le\log(3+T/\lambda)$, so, using $0<\lambda\le\lambda_{\max}^{kTULA}$,

$$(N\lambda)\vee\log N\le (T+3\lambda)\vee\log\Big(3+\tfrac{T}{\lambda}\Big)\le C'\Big(T\vee\log\Big(3+\tfrac{T}{\lambda}\Big)\Big)$$

for an absolute constant $C'>0$. By the choice of λ and absorbing C' into C,

$$C\,\big((N\lambda)\vee\log N\big)\lambda^2\le CC'\Big(T\vee\log\big(3+\tfrac{T}{\lambda}\big)\Big)\lambda^2\le\frac{\varepsilon}{2}.$$

Combining the two estimates gives $\mathrm{KL}(\mu\widehat{P}^N\|\pi_\beta)\le\varepsilon$.

There $\lambda\asymp\sqrt{\varepsilon/T}$ up to logarithmic factors, so

$$N\asymp\frac{T}{\lambda}=O\Big(\frac{T^{3/2}}{\sqrt{\varepsilon}}\Big)=O\Big(\frac{1}{\sqrt{\varepsilon}}\Big[\log\Big(\tfrac{2R_0}{\varepsilon}\Big)\Big]_+^{3/2}\Big)=\widetilde{O}\big(\varepsilon^{-1/2}\big),$$

since $T=\frac{1}{2C_{\mathrm{LSI}}}[\log(\frac{2R_0}{\varepsilon})]_+$. □

D.5.4. *Proof of Corollary 2.*

Proof. By Pinsker's inequality and (9),

$$\mathrm{TV}(\mu\widehat{P}^N,\pi_\beta)\le\sqrt{\tfrac12\,\mathrm{KL}\big(\mu\widehat{P}^N\,\|\,\pi_\beta\big)}\le\sqrt{\tfrac12\Big(C\,((N\lambda)\vee\log N)\lambda^2+e^{-2C_{\mathrm{LSI}}N\lambda}R_0\Big)},$$

so it suffices to make each term inside the square root at most $\varepsilon^2/2$. This is achieved exactly as in the proof of Proposition 1, with ε replaced by ε^2 in the definition of T: setting

$$T := \frac{1}{2C_{\mathrm{LSI}}}\Big[\log\Big(\tfrac{2R_0}{\varepsilon^2}\Big)\Big]_+, \qquad \lambda := \min\Big\{\lambda_{\max}^{kTULA},\ \sqrt{\tfrac{\varepsilon^2}{2C\,(T\vee\log(3+T/\lambda))}}\Big\}, \qquad N := \max\big\{3,\lceil T/\lambda\rceil\big\},$$

the same two estimates as there give $e^{-2C_{\mathrm{LSI}}N\lambda}R_0 \le \varepsilon^2/2$ and $C((N\lambda)\vee \log N)\lambda^2 \le \varepsilon^2/2$. Hence

$$\mathrm{TV}(\mu\widehat{P}^N, \pi_\beta) \le \sqrt{\tfrac{1}{2}\big(\tfrac{\varepsilon^2}{2}+\tfrac{\varepsilon^2}{2}\big)} = \frac{\varepsilon}{\sqrt{2}} \le \varepsilon.$$

For the complexity, $\lambda \asymp \varepsilon/\sqrt{T}$ up to logarithmic factors, so

$$N \asymp \frac{T}{\lambda} = O\Big(\frac{T^{3/2}}{\varepsilon}\Big) = \widetilde{O}\big(\varepsilon^{-1}\big),$$

since $T = \frac{1}{2C_{\mathrm{LSI}}}[\log(2R_0/\varepsilon^2)]_+$, where $\widetilde{O}(\cdot)$ hides polylogarithmic factors in ε^{-1} and R_0. □

D.5.5. *Proof of Corollary 3.*

Proof. By Talagrand's inequality implied by Assumption (A6) and (9),

$$W_2\big(\mu\widehat{P}^N, \pi_\beta\big) \le \sqrt{\tfrac{2}{C_{\mathrm{LSI}}}\,\mathrm{KL}\big(\mu\widehat{P}^N \,\|\, \pi_\beta\big)} \le \sqrt{\tfrac{2}{C_{\mathrm{LSI}}}\Big(C\,((N\lambda)\vee\log N)\lambda^2 + e^{-2C_{\mathrm{LSI}}N\lambda}R_0\Big)},$$

so it suffices to make each term inside the square root at most $\frac{C_{\mathrm{LSI}}}{4}\varepsilon^2$. This is achieved exactly as in the proof of Proposition 1, with the target accuracy $\frac{C_{\mathrm{LSI}}}{4}\varepsilon^2$ in place of $\varepsilon/2$: setting

$$T := \frac{1}{2C_{\mathrm{LSI}}}\Big[\log\Big(\tfrac{4R_0}{C_{\mathrm{LSI}}\varepsilon^2}\Big)\Big]_+, \qquad \lambda := \min\Big\{\lambda_{\max}^{kTULA},\ \sqrt{\tfrac{C_{\mathrm{LSI}}\varepsilon^2}{4C\,(T\vee\log(3+T/\lambda))}}\Big\}, \qquad N := \max\big\{3, \lceil T/\lambda\rceil\big\},$$

the same two estimates as there give $e^{-2C_{\mathrm{LSI}}N\lambda}R_0 \le \frac{C_{\mathrm{LSI}}}{4}\varepsilon^2$ and $C((N\lambda)\vee\log N)\lambda^2 \le \frac{C_{\mathrm{LSI}}}{4}\varepsilon^2$. Hence

$$W_2\big(\mu\widehat{P}^N, \pi_\beta\big) \le \sqrt{\tfrac{2}{C_{\mathrm{LSI}}}\cdot\tfrac{C_{\mathrm{LSI}}}{2}\,\varepsilon^2} = \varepsilon.$$

For the complexity, $\lambda \asymp \varepsilon/\sqrt{T}$ up to logarithmic factors, so

$$N \asymp \frac{T}{\lambda} = O\Big(\frac{T^{3/2}}{\varepsilon}\Big) = \widetilde{O}\big(\varepsilon^{-1}\big),$$

since $T = \frac{1}{2C_{\mathrm{LSI}}}[\log\big(\frac{4R_0}{C_{\mathrm{LSI}}\varepsilon^2}\big)]_+$, where $\widetilde{O}(\cdot)$ hides polylogarithmic factors in ε^{-1} and R_0. □

D.5.6. *Proof of Corollary 4.*

Proof. Recall that $h = \nabla u$, and let Z_∞ be an $\mathbb{R}^d$-valued random variable with $\mathrm{Law}(Z_\infty) = \pi_\beta$. Then

$$\mathbb{E}\big[u(\widehat{X}_N)\big] - \inf_{\theta\in\mathbb{R}^d} u(\theta) = \Big(\mathbb{E}[u(\widehat{X}_N)] - \mathbb{E}[u(Z_\infty)]\Big) + \Big(\mathbb{E}[u(Z_\infty)] - \inf_{\theta\in\mathbb{R}^d} u(\theta)\Big). \tag{57}$$

We first estimate the difference $\mathbb{E}[u(\widehat{X}_N)] - \mathbb{E}[u(Z_\infty)]$. Since $h = \nabla u$, the fundamental theorem of calculus gives, for all $x, y \in \mathbb{R}^d$,

$$u(x) - u(y) = \int_0^1 \langle h\big(tx + (1-t)y\big),\, x - y\rangle\, dt.$$

Using Assumption (A1), we obtain

$$\begin{aligned}
|u(x) - u(y)| &\le \int_0^1 \big\|h\big(tx + (1-t)y\big)\big\|\,\|x-y\|\,dt \\
&\le L\int_0^1 \Big(1 + \big\|tx + (1-t)y\big\|^{2\ell}\Big)\,dt\,\|x-y\| \\
&\le C_h\big(1 + \|x\|^{2\ell} + \|y\|^{2\ell}\big)\|x-y\|,
\end{aligned}$$

for some constant $C_h > 0$.

Let $P^\star \in \mathcal{C}(\mathrm{Law}(\widehat{X}_N), \pi_\beta)$ be an optimal coupling, so that

$$W_2^2\big(\mathrm{Law}(\widehat{X}_N), \pi_\beta\big) = \mathbb{E}_{P^\star}\|\widehat{X}_N - Z_\infty\|^2.$$

Then

$$\begin{aligned}
\mathbb{E}[u(\widehat{X}_N)] - \mathbb{E}[u(Z_\infty)] &= \mathbb{E}_{P^\star}\big[u(\widehat{X}_N) - u(Z_\infty)\big] \\
&\le C_h\,\mathbb{E}_{P^\star}\Big[\big(1 + \|\widehat{X}_N\|^{2\ell} + \|Z_\infty\|^{2\ell}\big)\|\widehat{X}_N - Z_\infty\|\Big].
\end{aligned}$$

By Cauchy–Schwarz,

$$\mathbb{E}[u(\widehat{X}_N)] - \mathbb{E}[u(Z_\infty)] \le C_h\Big(1 + \big(\mathbb{E}\|\widehat{X}_N\|^{4\ell}\big)^{1/2} + \big(\mathbb{E}\|Z_\infty\|^{4\ell}\big)^{1/2}\Big) \times W_2\big(\mathrm{Law}(\widehat{X}_N), \pi_\beta\big). \tag{58}$$

By Lemma 7,

$$\sup_{N\ge 0} \mathbb{E}\|\widehat{X}_N\|^{4\ell} < \infty,$$

and since $Z_\infty \sim \pi_\beta$, Lemma 5 yields

$$\mathbb{E}\|Z_\infty\|^{4\ell} < \infty.$$

Hence there exists a constant $M > 0$, independent of N and λ, such that

$$1 + \big(\mathbb{E}\|\widehat{X}_N\|^{4\ell}\big)^{1/2} + \big(\mathbb{E}\|Z_\infty\|^{4\ell}\big)^{1/2} \le M.$$

Substituting this into (58) yields

$$\mathbb{E}[u(\widehat{X}_N)] - \mathbb{E}[u(Z_\infty)] \le C_h M\, W_2\big(\mathrm{Law}(\widehat{X}_N), \pi_\beta\big).$$

By Corollary 3 (Talagrand's inequality combined with (9)),

$$W_2\big(\mathrm{Law}(\widehat{X}_N), \pi_\beta\big) \le C_1\, e^{-C_0 N\lambda} + C_2\, \lambda\, \sqrt{(N\lambda) \vee \log N},$$

for constants $C_0, C_1, C_2 > 0$ depending only on C_{LSI}, R_0, and the constant of (9). Substituting into $\mathbb{E}[u(\widehat{X}_N)] - \mathbb{E}[u(Z_\infty)] \le C_h M\, W_2(\mathrm{Law}(\widehat{X}_N), \pi_\beta)$ and absorbing $C_h M$ into the constants,

$$\mathbb{E}[u(\widehat{X}_N)] - \mathbb{E}[u(Z_\infty)] \le C_1\, e^{-C_0 N\lambda} + C_2\, \lambda\, \sqrt{(N\lambda) \vee \log N}. \tag{59}$$

For the second term in (57), a bound is obtained by following exactly the same argument as in [18, Lemma 4.9]. In the present setting, this yields a constant $C_3 > 0$ such that

$$\mathbb{E}[u(Z_\infty)] - \inf_{\theta\in\mathbb{R}^d} u(\theta) \le \frac{C_3 \log \beta}{\beta}. \tag{60}$$

Combining (57), (59), and (60) yields

$$\mathbb{E}\big[u(\widehat{X}_N)\big] - \inf_{\theta\in\mathbb{R}^d} u(\theta) \le C_1\, e^{-C_0 N\lambda} + C_2\, \lambda\, \sqrt{(N\lambda) \vee \log N} + \frac{C_3 \log \beta}{\beta},$$

which proves (10). □

D.6. **Proofs of the Main Results for tRLMC.**

Proof of Theorem 5. Applying Lemma 15, it remains to bound the shifted-distance term. For this, we use the deterministic estimates (27) and (28), where the quantities $\overline{a}_0$ and $\overline{a}_1$ are controlled through the strong local error from Proposition 10, the weak local error from Proposition 9, and the uniform moment bounds from Lemmas 12 and 14. Substituting these bounds into Lemma 15 yields the claimed estimate. □

REFERENCES

[1] Dominique Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX 1983/84: Proceedings*, pages 177–206. Springer, 2006.

[2] Jianhai Bao, Mateusz B Majka, and Jian Wang. Geometric ergodicity of modified euler schemes for sdes with super-linearity. *arXiv preprint arXiv:2412.19377*, 2024.

[3] Nicolas Brosse, Alain Durmus, Éric Moulines, and Sotirios Sabanis. The tamed unadjusted Langevin algorithm. *Stochastic Processes and their Applications*, 129(10):3638–3663, 2019.

[4] Yu Cao, Jianfeng Lu, and Lihan Wang. Complexity of randomized algorithms for underdamped langevin dynamics. *arXiv preprint arXiv:2003.09906*, 2020.

[5] Sinho Chewi, Jason Altschuler, and Holden Lu. Shifted composition: A local error framework for kl analysis of sampling algorithms. *Journal of Machine Learning Research*, 25(123):1–42, 2024.

[6] Sinho Chewi, Murat A. Erdogdu, Mufan Bill Li, Ruoqi Shen, and Matthew Zhang. Analysis of langevin monte carlo from poincaré to log-sobolev, 2024.

[7] Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities, 2016.

[8] Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3): 651–676, 2017.

[9] Alain Durmus, Eric Moulines, et al. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability*, 27(3):1551–1587, 2017.

[10] Alain Durmus, Eric Moulines, et al. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 2019.

[11] Murat A Erdogdu and Rasa Hosseinzadeh. On the convergence of langevin monte carlo: The interplay between tail growth and smoothness. In *Conference on Learning Theory*, pages 1776–1822. PMLR, 2021.

[12] Murat A. Erdogdu, Rasa Hosseinzadeh, and Shunshi Zhang. Convergence of langevin monte carlo in chi-squared and rényi divergence. In *International Conference on Artificial Intelligence and Statistics*, pages 8151–8175. PMLR, 2022.

[13] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong convergence of an explicit numerical method for sdes with nonglobally lipschitz continuous coefficients. *Ann. Appl. Probab.*, 22(4):1611–1641, 08 2012.

[14] Chii-Ruey Hwang. Laplace's method revisited: Weak convergence of probability measures. *The Annals of Probability*, 8(6):1177–1182, 1980.

[15] T. Johnston, I. Lytras, and S. Sabanis. Kinetic langevin mcmc sampling without gradient lipschitz continuity—the strongly convex case. *J. Complex.*, 85, 2024.

[16] Ruinan Li, Tian Shen, and Zhonggen Su. Convergence rate of randomized midpoint langevin monte carlo. *arXiv preprint arXiv:2511.13093*, 2025.

[17] Dong-Young Lim, Ariel Neufeld, Sotirios Sabanis, and Ying Zhang. Non-asymptotic estimates for TUSLA algorithm for non-convex learning with applications to neural networks with ReLU activation function. *IMA Journal of Numerical Analysis*, page drad038, 2023.

[18] Dong-Young Lim, Ariel Neufeld, Sotirios Sabanis, and Ying Zhang. Non-asymptotic estimates for tusla algorithm for non-convex learning with applications to neural networks with relu activation function, 2023. URL `https://arxiv.org/abs/2107.08649`.

[19] Attila Lovas, Iosif Lytras, Miklós Rásonyi, and Sotirios Sabanis. Taming neural networks with tusla: Nonconvex learning via adaptive stochastic gradient langevin algorithms. *SIAM Journal on Mathematics of Data Science*, 5(2):323–345, 2023.

[20] I. Lytras and S. Sabanis. Taming under isoperimetry. *Stochastic Process. Appl.*, 188, 2025.

[21] Iosif Lytras and Panagiotis Mertikopoulos. Contractive kinetic langevin samplers beyond global lipschitz continuity. *arXiv preprint arXiv: 2509.12031*, 2025.

[22] Iosif Lytras and Panayotis Mertikopoulos. Tamed langevin sampling under weaker conditions. *arXiv preprint arXiv:2405.17693*, 2024.

[23] Iosif Lytras, 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.

[24] Mateusz B Majka, Aleksandar Mijatović, and Lukasz Szpruch. Non-asymptotic bounds for sampling algorithms without log-concavity. *arXiv preprint arXiv:1808.07105*, 2018.

[25] Alireza Mousavi-Hosseini, Tyler K Farghly, Ye He, Krishna Balasubramanian, and Murat A Erdogdu. Towards a complete analysis of langevin monte carlo: Beyond poincaré inequality. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 1–35. PMLR, 2023.

[26] Ariel Neufeld, Matthew Ng Cheng En, and Ying Zhang. Non-asymptotic convergence bounds for modified tamed unadjusted langevin algorithm in non-convex setting, 2024. URL `https://arxiv.org/abs/2207.02600`.

[27] Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via stochastic gradient langevin dynamics: a nonasymptotic analysis. In *Conference on Learning Theory*, pages 1674–1703, 2017.

[28] Sotirios Sabanis. A note on tamed euler approximations. *Electron. Commun. Probab.*, 18(47):1–10, 2013.

[29] Sotirios Sabanis. Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients. *The Annals of Applied Probability*, 26(4):2083–2105, 2016.

[30] Ruoqi Shen and Yin Tat Lee. The randomized midpoint method for log-concave sampling. *Advances in Neural Information Processing Systems*, 32, 2019.

[31] Santosh Vempala and Andre Wibisono. Rapid convergence of the unadjusted langevin algorithm: Isoperimetry suffices. *Advances in neural information processing systems*, 32, 2019.
[32] Feng-Yu Wang. Harnack inequality for sde with multiplicative noise and extension to neumann semigroup on nonconvex manifolds. *The Annals of Probability*, 39(4), 2011.
[33] Xiaojie Wang and Bin Yang. When langevin monte carlo meets randomization: Non-asymptotic error bounds beyond log-concavity and gradient lipschitzness. *arXiv preprint arXiv:2509.25630*, 2025.
[34] Bin Yang and Xiaojie Wang. Non-asymptotic error bounds in w2-distance with sqrt(d) dimension dependence and first order convergence for langevin monte carlo beyond log-concavity. In *Proceedings of the 42nd International Conference on Machine Learning*, pages 71358–71382, 2025.
[35] Xuda Ye and Zhennan Zhou. Error analysis of time-discrete random batch method for interacting particle systems and associated mean-field limits. *IMA Journal of Numerical Analysis*, 44(3): 1660–1698, 2024.
[36] Lu Yu, Avetik Karagulyan, and Arnak Dalalyan. Langevin monte carlo for strongly log-concave distributions: Randomized midpoint revisited. *arXiv preprint arXiv:2306.08494*, 2023.

ARCHIMEDES, ATHENA RESEARCH CENTER, GREECE AND NATIONAL TECHNICAL UNIVERSITY OF ATHENS, ATHENS,GREECE.
Email address: i.lytras@athenarc.gr

* ARCHIMEDES, ATHENA RESEARCH CENTER, GREECE
Email address: aggelosntousis02@gmail.com