

Convergence of Noise-Free Sampling Algorithms with Regularized Wasserstein Proximals

Fuqun Han

FQHAN@MATH.UCLA.EDU

*Department of Mathematics
University of California, Los Angeles
Los Angeles, CA, USA*

Stanley Osher

SJO@MATH.UCLA.EDU

*Department of Mathematics
University of California, Los Angeles
Los Angeles, CA, USA*

Wuchen Li

WUCHEN@MAILBOX.SC.EDU

*Department of Mathematics
University of South Carolina
Columbia, SC, USA*

Editor: Krishnakumar Balasubramanian

Abstract

In this work, we investigate the convergence properties of the backward regularized Wasserstein proximal (BRWP) method for sampling a target distribution. The BRWP approach can be shown as a semi-implicit time discretization for a probability flow ODE with the score function whose density satisfies the Fokker-Planck equation of the overdamped Langevin dynamics. Specifically, the evolution of the density—hence the score function—is approximated via a kernel representation derived from the regularized Wasserstein proximal operator. By applying the dual formulation and a localized Taylor series to obtain the asymptotic expansion of this kernel formula, we establish guaranteed convergence in terms of the Kullback–Leibler divergence for the BRWP method towards a strongly log-concave target distribution. Our analysis also identifies the optimal and maximum step sizes for convergence. Furthermore, we demonstrate that the deterministic and semi-implicit BRWP scheme outperforms many classical Langevin Monte Carlo methods, such as the Unadjusted Langevin Algorithm (ULA), by offering faster convergence and reduced bias. Numerical experiments further validate the convergence analysis of the BRWP method.

Keywords: Entropy dissipation; Regularized Wasserstein proximal; Score functions; Optimal time stepsize.

1. Introduction

Sampling from complex and potentially high-dimensional distributions is increasingly crucial in data science (Andrieu et al., 2003), computational mathematics (Durmus and Moulines, 2018), and engineering (Leimkuhler and Matthews, 2015). Efficient sampling algorithms are central to numerous real-world applications, including identifying global optimizers for high-dimensional functions (Ma et al., 2019), generating samples from latent spaces in generative modeling (Song and Ermon, 2019), and solving Bayesian inverse problems to estimate posterior distributions (Stuart, 2010; Garbuno-Inigo et al., 2020).

The success of these applications heavily relies on the efficiency, reliability, and theoretical convergence guarantees of the employed sampling algorithms.

Given the importance of sampling from known distributions, various algorithms have been developed and analyzed. Markov Chain Monte Carlo (MCMC) methods are widely used (Ma et al., 2021; Mengersen and Tweedie, 1996; Bélisle et al., 1993; Carrillo et al., 2022; Betancourt, 2017). A well-known example is the overdamped Langevin dynamics, which relies on a gradient drift vector field and a simple diffusion process introduced by Gaussian noise. Theoretical results show that the continuous-time Langevin Monte Carlo (LMC) method can converge to a stationary distribution under appropriate assumptions (Otto and Villani, 2000). In practice, however, challenges arise with LMC when discretizing the Langevin dynamics in time, particularly due to the dimension dependence caused by the random walk component and the bias introduced by the finite time step-size. Notable discretizations include the Unadjusted Langevin Algorithm (ULA), which employs an explicit Euler update (Parisi, 1981), the Metropolis-Adjusted Langevin Algorithm (MALA), which incorporates an acceptance–rejection step (Xifara et al., 2014), the Proximal Langevin Algorithm, which uses implicit updates in the drift to handle nonsmooth potentials (Liang and Chen, 2022), and forward–backward splitting schemes formulated directly in the Wasserstein space (Salim et al., 2020; Diao et al., 2023). While these implicit or semi-implicit methods can partially mitigate discretization-induced bias, Wasserstein splitting approaches typically require solving a variational subproblem at each iteration, leading to substantial computational overhead except in special cases where closed-form updates are available.

Motivated by these limitations of LMC-based approaches, an alternative perspective is to reformulate the diffusion process in terms of the score function and consider the corresponding probability flow ODE. In recent years, studies such as (Song and Ermon, 2019; Lu et al., 2022; Chen et al., 2024; Huang et al., 2024) have focused on this probability flow ODE derived from Langevin dynamics. This approach expresses the diffusion process, originally driven by Brownian motion, in terms of the score function—the gradient of the logarithm of the density function. The density itself satisfies the Fokker–Planck equation associated with overdamped Langevin dynamics. This reformulation also connects to the gradient flow of the Kullback–Leibler (KL) divergence in the Wasserstein space, the space of probability densities endowed with the Wasserstein-2 metric (Otto and Villani, 2000). In this context, the score function represents the Wasserstein gradient of the negative Boltzmann–Shannon entropy. From a numerical standpoint, algorithms that explicitly utilize the score often achieve improved stability and accuracy (Lu et al., 2022; Zhang and Chen, 2022; Zhao et al., 2024; Wang and Li, 2022).

Despite this advantage, applying the probability flow ODE to sampling introduces several significant challenges. First, the computation or approximation of the score function can be both computationally intensive and inaccurate when only finite samples are available. Second, the choice of time discretization is crucial: while a stable implicit scheme is desirable, it is often difficult to implement. The time-implicit Jordan–Kinderlehrer–Otto (JKO) (Jordan et al., 1998) scheme addresses this issue by iteratively solving the proximal operator of the KL divergence in Wasserstein space. However, the JKO update is computationally demanding due to the absence of a closed-form formula for the Wasserstein proximal operator, which typically requires solving an additional optimization problem at each time step.

To mitigate these difficulties, a regularization term can be added to the Wasserstein proximal operator, as demonstrated in (Li et al., 2023), yielding a system that admits a closed-form solution via the Hopf–Cole transformation. This solution, known as the kernel formula, allows a semi-implicit time discretization of the evolution of the score function and forms the basis of the Backward Regularized Wasserstein Proximal (BRWP) scheme.

The BRWP scheme, initially introduced in (Tan et al., 2024) using empirical particle approximations for the density, was later enhanced in (Han et al., 2025) through tensor-train representations to accommodate high-dimensional settings. Analysis in (Han et al., 2025; Tan et al., 2024) demonstrates that, for Gaussian target distributions, the BRWP scheme achieves faster convergence and exhibits reduced sensitivity to dimensionality in terms of mixing time. These results suggest its potential applicability to a broader class of non-Gaussian distributions, motivating the more general convergence analysis.

In this work, we explore the convergence properties of the BRWP scheme for sampling from strongly log-concave distributions. By leveraging the dual expansion on the kernel formula, we show that the BRWP algorithm is a semi-implicit discretization of the probability flow ODE, where the score function is evaluated at the next time point and free from Brownian motion. This implicit approach enhances the robustness and stability of the sampling process. Moreover, since the probability flow ODE in the BRWP scheme is deterministic, we can conduct weak second-order numerical analysis, which allows us to obtain an optimal step size to achieve faster convergence, as demonstrated in our upcoming analysis.

We informally demonstrate the main results as follows. We aim to sample a target distribution $\rho^* := \frac{1}{Z} \exp(-\beta V)$ with a given potential function $V : \mathbb{R}^d \rightarrow \mathbb{R}$, a normalization constant $Z = \int_{\mathbb{R}^d} \exp(-\beta V(x)) dx < \infty$, a constant inverse temperature β , and an initial density ρ_0 . In particular, in Section 3, we show that the following kernel formula from the regularized Wasserstein proximal operator (RWPO)

$$\mathcal{K}_V^h \rho(x) = \int_{\mathbb{R}^d} \frac{\exp \left[-\frac{\beta}{2} \left(V(x) + \frac{\|x-y\|_2^2}{2h} \right) \right]}{\int_{\mathbb{R}^d} \exp \left[-\frac{\beta}{2} \left(V(z) + \frac{\|z-y\|_2^2}{2h} \right) \right] dz} \rho_0(y) dy, \quad (1)$$

provides a one-step time approximation of the Fokker-Planck equation:

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\nabla V \rho) + \beta^{-1} \Delta \rho.$$

Here, the stepsize is given as $h > 0$.

Theorem 1 (*Informal, see Theorem 4*) *For fixed test function $\varphi \in C^{2,1}(U)$ and stepsize $h > 0$, suppose ρ_0 satisfies the Fokker-Planck equation at time t_0 . Then we have*

$$\left| \left\langle \varphi, \mathcal{K}_V^h \rho_0 - \rho_0 - h \frac{\partial \rho_0}{\partial t} \Big|_{t=t_0} \right\rangle \right| \leq C h^2 \|\varphi\|_{C^{2,1}(U)}, \quad (2)$$

where the constant depends on V through its derivative up to order 3 and the local domain U .

With the help of the kernel formula (1) and assuming that it can be computed accurately, we implement the following semi-implicit discretization with the stepsize $h > 0$ of the probability flow ODE to sample from ρ^* :

$$x_{k+1} = x_k - h(\nabla V(x_k) + \beta^{-1} \nabla \log \mathcal{K}_V^h \rho_k(x_k)), \quad (3)$$

where $x_k \sim \rho_k$ and $\mathcal{K}_V^h \rho_k$ provides $\mathcal{O}(h^2)$ weak approximation to ρ_{k+1} . Considering the evolution of the density function associated with x_k , denoted as ρ_k , in Section 4, we show the convergence guarantees of the proposed sampling method in terms of the KL divergence where $D_{\text{KL}}(\rho_k \|\rho^*) = \int \rho_k \log \frac{\rho_k}{\rho^*} dx$.

Theorem 2 (*Informal, see Theorem 12 and Assumptions 1 to 4*) Assume ρ^* is strongly log-concave with constant α . For a sufficient small constant $\delta > 0$, the BRWP algorithm will achieve an error of $D_{\text{KL}}(\rho_k \|\rho^*) \leq \delta$, with

$$k = \mathcal{O}\left(\frac{|\ln \delta|}{2\alpha\sqrt{\delta}}\right),$$

where the stepsize is chosen as $h = \sqrt{\delta}$ when $\sqrt{\delta} < 2/(3\alpha)$.

The convergence of Langevin–dynamics–based sampling methods has been extensively studied under various assumptions; see, for example, Balasubramanian et al. (2022); Durmus and Moulines (2018); Dwivedi et al. (2019); Chewi et al. (2025). In contrast to methods such as ULA, proximal Langevin, and other stochastic Langevin-type algorithms, our scheme—being entirely noise-free—achieves an improved sampling complexity under a suitable score oracle assumption, while incurring only an $\mathcal{O}(\sqrt{\delta})$ bias. We further outline in Section 5 several practical scenarios in which this score oracle assumption can be guaranteed.

The deterministic structure of our method also allows for an explicit characterization of the KL decay, which in turn leads to a principled choice of an optimal stepsize. This differs from sampling algorithms derived from the probability flow ODE, whose convergence properties have been analyzed in (Gao and Zhu, 2024; Chen et al., 2023, 2022). Those approaches may suffer from instability, with errors potentially increasing over time, and typically assume accurate estimation of the score function at each iteration (Yang and Wibisono, 2022). In our setting, the score information is obtained through the RWPO kernel, and theoretical analysis in Section 3 with practical consideration in Section 5 provide conditions under which the error in the score approximation can be controlled. Moreover, the potential growth of the approximation error at infinity is handled by establishing uniform $C^{3,1}$ bounds on the discrete densities, leveraging the regularizing effect of the kernel formula. As a result, the optimal stepsize arises from balancing the KL contraction rate, available under strong convexity, with the weak truncation error of the RWPO kernel, as detailed in Section 4.

Finally, in Section 6, we present numerical experiments comparing our method against ULA and proximal-type sampling algorithms, thereby validating our theoretical findings. We conclude with a brief discussion of several promising directions for future research.

2. Review on Probability Flow ODE with Score Function, Regularized Wasserstein Proximal Operator, and BRWP Algorithm

In this section, we begin by outlining the sampling problem for a given target distribution and reviewing classical sampling algorithms derived from overdamped Langevin dynamics, along with their convergence analysis in Wasserstein space. We then discuss various sampling algorithms based on discrete-time approximations of the Fokker-Planck equation, including the ULA, the JKO scheme, and BRWP. To provide a comprehensive comparison, we summarize the assumptions required in the upcoming analysis, sampling complexity, and optimal stepsize for BRWP and several popular existing methods in Section 2.4 and Table 1.

2.1 Sampling problem

We aim to generate samples $x_{k,j} \in \mathbb{R}^d$, where $k \in \mathbb{N}$ denotes the iteration index and $j = 1, \dots, N$ indexes the particles. The target distribution is known as a Gibbs measure

$$\rho^*(x) = \frac{1}{Z} \exp(-\beta V(x)),$$

where $Z = \int_{\mathbb{R}^d} \exp(-\beta V(x)) dx < +\infty$ is the normalization constant, $\beta > 0$ is the inverse temperature, and V represents the potential function. The collection $\{x_{0,j}\}_{j=1}^N$ is an arbitrary set of initial particles, and the objective of the sampling algorithm is to ensure that, as k increases, the distribution of $\{x_{k,j}\}_{j=1}^N$ approximates the density function ρ^* .

We first recall the classical overdamped Langevin dynamics to highlight the difficulties that the BRWP method resolves. The dynamic is described by a stochastic differential equation

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dB_t, \quad (4)$$

where B_t denotes the standard Brownian motion in \mathbb{R}^d . The density function of X_t evolves according to the Kolmogorov forward equation, also named Fokker-Planck equation:

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \nabla \cdot (\rho \nabla V) + \beta^{-1} \Delta \rho \\ &= \nabla \cdot (\rho \nabla V) + \beta^{-1} \nabla \cdot (\rho \nabla \log \rho) \\ &= \beta^{-1} \nabla \cdot \left(\rho \nabla \log \frac{\rho}{\rho^*} \right), \end{aligned} \quad (5)$$

where the second equality is based on the fact that $\nabla \log \rho = \nabla \rho / \rho$. Clearly, ρ^* is an equilibrium of the Fokker-Planck equation (5). Thus, ρ^* is the unique invariant distribution of the Langevin dynamics.

A known fact is that the Fokker-Planck equation corresponds to the Wasserstein-2 gradient flow of the KL divergence

$$D_{\text{KL}}(\rho \parallel \rho^*) := \int \rho \log \frac{\rho}{\rho^*} dx.$$

This means that the Fokker-Planck equation (5) can be formulated as

$$\frac{\partial \rho}{\partial t} = \beta^{-1} \nabla \cdot \left(\rho \nabla \left(\frac{\delta}{\delta \rho} D_{\text{KL}}(\rho \parallel \rho^*) \right) \right),$$

where $\frac{\delta}{\delta \rho}$ is the L^2 first variation with respect to the density function ρ . Furthermore, in the Wasserstein space, the squared norm of the gradient of the KL divergence, also known as the relative Fisher information, is written as

$$\mathcal{I}(\rho \parallel \rho^*) = \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \rho dx.$$

From any initial distribution, the probability density of Langevin dynamics converges to the target distribution ρ^* under mild conditions. Specifically, we have

$$\frac{d}{dt} D_{\text{KL}}(\rho \parallel \rho^*) = -\beta^{-1} \mathcal{I}(\rho \parallel \rho^*) < 0.$$

Moreover, when ρ^* satisfies the log-Sobolev inequality (LSI) with constant α , we can establish the relationship between the KL divergence and Fisher information

$$D_{\text{KL}}(\rho \parallel \rho^*) \leq \frac{1}{2\beta\alpha} \mathcal{I}(\rho \parallel \rho^*). \quad (6)$$

This inequality can be viewed as the gradient-dominated condition, commonly known as the Polyak-Lojasiewicz (PL) inequality in the Wasserstein space. From now on, we assume that $\nabla^2 V \succeq \alpha I$, or ρ^* is strongly log-concave, such that the LSI holds (Gross, 1975; Bakry and Émery, 2006).

Utilizing the PL inequality, one can show that the KL divergence converges exponentially fast along the Wasserstein gradient flow (see Lemma 24 in the appendix for more details)

$$D_{\text{KL}}(\rho_t \|\rho^*) \leq \exp(-2\alpha t) D_{\text{KL}}(\rho_0 \|\rho^*),$$

where $\rho_t(x) = \rho(t, x)$.

Although the continuous-time convergence is well understood, a numerical analysis of the particle trajectories arising from the discretized Langevin scheme shows that the method induces a systematic bias and has a dependence on the dimensionality of the problem. This issue arises due to the evolution of Brownian motions in approximating Langevin dynamics, which can slow down convergence in high-dimensional sampling problems, as its variance is proportional to the dimension. More details will be provided in the next subsection.

Given these difficulties, in this work, we instead consider a probability flow ODE, where the diffusion is generated by the score function:

$$dX_t = -\nabla V(X_t)dt - \beta^{-1} \nabla \log \rho_t(X_t)dt. \tag{7}$$

The trajectory of (7) differs from the Langevin dynamics (4). However, since the drift field satisfies $\nabla \log \rho = \nabla \rho / \rho$, the Liouville equation coincides with the Fokker-Planck equation (5). Consequently, the corresponding probability flow continues to be the Wasserstein gradient flow for the KL divergence. In continuous time, the KL divergence still converges exponentially fast along the flow given by (7).

Moreover, since the ODE (7) is deterministic, it is feasible to perform higher-order numerical analysis under finite time stepsizes to explore the higher-order error terms introduced by discretization and the optimal stepsize associated with the discretization. This analysis will be the main focus of Section 4, where we demonstrate the improved convergence rate and accuracy of our scheme.

2.2 Discrete time approximation

Although the KL divergence converges exponentially fast in continuous time along the flow given by the Fokker-Planck equation, discretizing Langevin dynamics slows down convergence. It also introduces a bias term due to the discretization error. For example, applying explicit Euler discretization with stepsize h to (4) results in the ULA

$$x_{k+1} = x_k - h\nabla V(x_k) + \sqrt{2\beta^{-1}h}z_k, \tag{8}$$

where $z_k \sim \mathcal{N}(0, I)$. The convergence of ULA is well studied, and (Vempala and Wibisono, 2019) shows

$$D_{\text{KL}}(\rho_k \|\rho^*) \leq \exp(-\alpha hk) D_{\text{KL}}(\rho_0 \|\rho^*) + \frac{8hdL^2}{\alpha},$$

where ρ_k represents the density at time $t_k = kh$ and $-LI \preceq \nabla^2 V \preceq LI$. The discretization induces an error term of order h , which depends on the dimension d that arises from the Brownian motion term. This makes high-dimensional cases particularly challenging and requires a small stepsize for the convergence.

Then we turn to study the discretization of the probability flow ODE (7) and note that one of the primary challenges is the approximation of the score function. It becomes increasingly difficult and inaccurate in high-dimensional spaces. As a result, an explicit forward Euler discretization can substantially misestimate the transport field induced by $\nabla \log \rho$. This amplification of discretization error leads to numerical instability and may cause the particle system to collapse or diverge. Hence, an implicit discretization of (7) becomes highly demanding.

Implicit methods for solving Langevin dynamics have also been extensively studied, including proximal Langevin (Wibisono, 2019), proximal sampler with restricted Gaussian oracle (Liang and Chen, 2022), and implicit Langevin (Hodgkinson et al., 2021). In particular, from the density level, the implicit method usually corresponds to a certain splitting (Bernton, 2018) of the classical JKO scheme for the Fokker-Planck equation

$$\rho_{k+1} = \arg \min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \beta^{-1} \text{D}_{\text{KL}}(\rho \| \rho^*) + \frac{1}{2h} W_2(\rho, \rho_k)^2, \quad (9)$$

where $W_2(\rho, \rho_k)$ is the Wasserstein-2 distance between densities ρ and ρ_k .

However, the aforementioned implicit implementation requires solving certain proximal operators or a system of nonlinear equations, which could be time-consuming and difficult for general families of density functions. In light of this, a semi-implicit Euler discretization method was proposed in (Tan et al., 2024; Han et al., 2025) for the particle evolution equation (7) at time t_k :

$$x_{k+1} = x_k - h (\nabla V(x_k) + \beta^{-1} \nabla \log \rho_{k+1}(x_k)). \quad (10)$$

In other words, we only evaluate the score function in the next time step.

2.3 Regularized Wasserstein proximal operator and kernel formula

Concerning (10), one may note that approximating the terminal density $\rho_{k+1} := \rho(t_k + h, \cdot)$ is challenging due to its nonlinear dependence on the initial density and the high-dimensional nature of ρ . A natural approach for discretizing the Fokker-Planck equation is the classical JKO scheme (9), but each iteration requires solving a Wasserstein proximal operator, which in general involves an expensive optimization problem.

To address this difficulty, we recall that the JKO update can be interpreted as a Wasserstein proximal operator associated with the KL divergence. This perspective suggests that, if a tractable surrogate for this proximal operator were available, one could design a stable and semi-implicit discretization of the density evolution. Motivated by this idea, we consider a regularized Wasserstein proximal operator (Li et al., 2023), which admits a closed-form solution and can therefore be used to approximate the Fokker-Planck dynamics with a small stepsize.

Let ρ_0 be the initial density, we first recall the Wasserstein proximal operator associated with a linear energy functional:

$$\arg \min_q \left[\frac{1}{2h} W_2(\rho_0, q)^2 + \int_{\mathbb{R}^d} V(x) q(x) dx \right]. \quad (11)$$

Note that this differs from the Wasserstein proximal operator with KL divergence used in the JKO scheme; this connection will become clear once a Laplacian regularization is added.

By the Benamou-Brenier formula (Benamou and Brenier, 2000), the Wasserstein-2 distance can be written as an optimal transport problem:

$$\frac{W_2(\rho_0, q)^2}{2h} = \inf_{\rho, v} \int_0^h \int_{\mathbb{R}^d} \frac{1}{2} \|v(t, x)\|_2^2 \rho(t, x) dx dt,$$

where the minimization is taken over vector fields $v : [0, h] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and densities $\rho : [0, h] \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \quad \rho(0, x) = \rho_0(x), \quad \rho(h, x) = q(x).$$

Solving (11) directly is typically a challenging variational problem. In (Li et al., 2023), motivated by Schrödinger bridge systems, a regularized Wasserstein proximal operator is introduced by adding a Laplacian term to the dynamics, leading to

$$\mathcal{K}_V^h \rho_0 := \arg \min_q \inf_{v, \rho} \int_0^h \int_{\mathbb{R}^d} \frac{1}{2} \|v(t, x)\|_2^2 \rho(t, x) dx dt + \int_{\mathbb{R}^d} V(x) q(x) dx, \quad (12)$$

subject to the modified continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = \beta^{-1} \Delta \rho, \quad \rho(0, x) = \rho_0(x), \quad \rho(h, x) = q(x). \quad (13)$$

We remark that this regularization is the key step: it enables a closed-form solution via the Hopf-Cole transformation, leading to the kernel formula that forms the basis of the BRWP scheme.

Introducing a Lagrange multiplier function Φ , we find that solving $\mathcal{K}_V^h \rho_0$ is equivalent to computing the solution of the coupled PDEs

$$\begin{cases} \partial_t \rho + \nabla_x \cdot (\rho \nabla_x \Phi) = \beta^{-1} \Delta_x \rho, \\ \partial_t \Phi + \frac{1}{2} \|\nabla_x \Phi\|_2^2 = -\beta^{-1} \Delta_x \Phi, \\ \rho(0, x) = \rho_0(x), \quad \Phi(h, x) = -V(x). \end{cases} \quad (14)$$

Comparing the first equation in (14) with the Fokker-Planck equation defined in (5), we observe that the solution to the regularized Wasserstein proximal operator (12) approximates the terminal density ρ when h is small. In other words, by solving the coupled PDE (14), we can approximate the evolution of the Fokker-Planck equation. This approximation will be justified rigorously in Section 3.

The primary motivation for considering the regularized Wasserstein proximal operator as an approximate solution lies in the fact that the coupled PDEs (14) can be solved using a Hopf-Cole type transformation (Evans, 2022, Sec. 4.4) by taking $\eta = \exp(\beta/2\Phi)$ and $\hat{\eta} = \rho/\eta$. Then (14) is equivalent to a system of backward-forward heat equations

$$\rho(t, x) = \eta(t, x) \hat{\eta}(t, x), \quad (15)$$

where $\hat{\eta}$ and η satisfy

$$\begin{aligned} \partial_t \hat{\eta}(t, x) &= \beta^{-1} \Delta_x \hat{\eta}(t, x), \\ \partial_t \eta(t, x) &= -\beta^{-1} \Delta_x \eta(t, x), \\ \hat{\eta}(0, x) \hat{\eta}(0, x) &= \rho_0(x), \quad \eta(h, x) = \exp(-\beta/2V(x)). \end{aligned} \quad (16)$$

The system (15) can be explicitly solved using the heat kernel

$$G_h(x - y) = \frac{1}{(4\pi h/\beta)^{d/2}} \exp\left(-\beta \frac{\|x - y\|_2^2}{4h}\right),$$

corresponding to a diffusion process with thermal diffusivity β . For instance, the forward component can be written as

$$\eta(t, x) = \int_{\mathbb{R}^d} G_h(x - y) e^{-\frac{\beta}{2}V(y)} dy,$$

and an analogous expression holds for $\hat{\eta}$. Through the coupling relation in (15), the final solution to the regularized Wasserstein proximal operator (12) can then be obtained in

closed form:

$$\begin{cases} \rho(t, x) = \left(G_{h-t} * \exp\left(-\frac{V}{2\beta}\right) \right) (x) \cdot \left(G_h * \frac{\rho_0}{G_h * \exp\left(-\frac{V}{2\beta}\right)} \right) (x), \\ v(t, x) = 2\beta^{-1} \nabla \log \left(G_{h-t} * \exp\left(-\frac{V}{2\beta}\right) \right) (x), \\ \rho(h, x) = \exp\left(-\frac{V(x)}{2\beta}\right) \cdot \left(G_h * \frac{\rho_0}{G_h * \exp\left(-\frac{V}{2\beta}\right)} \right) (x). \end{cases}$$

In particular, we focus on the terminal density $\rho(h, x)$, which serves as an approximation to the solution of the Fokker-Planck equation at time h with initial density ρ_0 . For this purpose, we write down the explicit kernel formula of $\rho(h, x)$, in which the normalization constant of the heat kernel is omitted, as it cancels out and plays no role in the subsequent analysis

$$\mathcal{K}_V^h(\rho_0)(x) := \rho(h, x) = \int_{\mathbb{R}^d} \frac{\exp\left[-\frac{\beta}{2}\left(V(x) + \frac{\|x-y\|_2^2}{2h}\right)\right]}{\int_{\mathbb{R}^d} \exp\left[-\frac{\beta}{2}\left(V(z) + \frac{\|z-y\|_2^2}{2h}\right)\right] dz} \rho_0(y) dy. \quad (17)$$

2.4 BRWP algorithm

With the aid of (17), we approximate ρ_{k+1} in (10) by $\mathcal{K}_V^h \rho_k$. This approximation leads to the BRWP sampling algorithm, which is derived from a semi-implicit discretization of the probability flow ODE and whose key advantage is that it admits a closed-form update, in contrast to other proximal-based Langevin algorithms.

Algorithm 1 BRWP for Sampling from $\rho^* = \frac{1}{Z} \exp(-\beta V)$

- 1: **Input:** Initial particle set $\{x_{0,j}\}_{j=1}^N$
- 2: **for** $k = 1, 2, 3, \dots$ **do**
- 3: Given ρ_k as the density associated with $\{x_{k,j}\}^1$
- 4: **for** each particle $j = 1, 2, \dots, N$ **do**
- 5: Update particle:

$$x_{k+1,j} = x_{k,j} - h \left(\nabla V(x_{k,j}) + \beta^{-1} \nabla \log \tilde{\rho}_{k+1}(x_{k,j}) \right)$$

with $\tilde{\rho}_{k+1} = \mathcal{K}_V^h(\rho_k)$ defined in (17) that approximate ρ_{k+1} .

- 6: **end for**
 - 7: **end for**
-

The goal of this work is to establish results on the sampling complexity and maximum stepsize h required for the convergence of the interacting particle system generated by Algorithm 1 under the following assumptions.

Assumption 1 (Regularity) *The potential $V \in C_{\text{loc}}^{4,1}(\mathbb{R}^d)$ is α -strongly convex and L -smooth; that is,*

$$\alpha I_d \preceq \nabla^2 V(x) \preceq L I_d \quad \text{for some } \alpha, L > 0, \quad \forall x \in \mathbb{R}^d.$$

Assumption 2 (Polynomial growth) *There exist constants $a_1, a_2 > 0$ and $m \geq 1$ such that*

$$V(x) \geq a_1 \|x\|^m - a_2.$$

Moreover, there exist constants $C_V, q_V \geq 0$ such that, for all multi-indices $|\lambda| \leq 3$,

$$|\partial^\lambda V(x)| \leq C_V(1 + \|x\|)^{q_V}.$$

Assumption 3 (Finite moment and regularity of initial density) *The initial density $\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$ has a finite p -th moment for some $p > d + q_V + 3$, belongs to $C^{3,1}(U)$, and is strictly positive on a neighborhood of a bounded set U (see Section 3 for the precise specification of U).*

Assumption 1 provides the smoothness of V required to justify the second- and third-order Taylor expansions that underlie our local weak error analysis. We remark that, to establish the second-order weak accuracy of the kernel formula in Lemma 3, it would suffice to assume $V \in C^{3,1}(U)$. However, computing the score function $\nabla \log \mathcal{K}_V^h \rho_k$ requires differentiating $\mathcal{K}_V^h \rho_k$ once more, which in turn takes an additional derivative of V . This is the reason for the stronger regularity requirement adopted in the present work. The convexity condition further guarantees the validity of the log-Sobolev inequality (6), which yields exponential convergence of the corresponding Fokker-Planck flow toward equilibrium.

Assumption 2 ensures that V and its derivatives exhibit at most polynomial growth. This condition, along with uniform moment control in Lemma 15, justifies the reduction from an unbounded to a bounded domain, as employed in Theorem 4. Under these assumptions, the kernel operator \mathcal{K}_V^h is well defined and yields the desired approximation properties. Finally, Assumption 3 guarantees the well-posedness of the iterative scheme and the associated score approximation, starting from a properly defined initial state in the proof of Section 3.2.

Numerical experiments (see Section 6) further demonstrate that the proposed algorithms remain robust beyond these idealized conditions, performing effectively even for nonconvex or multimodal target distributions. Our proof in Section 3 will rely primarily on Assumptions 1-3, where we establish that $\mathcal{K}_V^h(\rho_k) \approx \rho_{k+1}$ in the weak sense.

We first summarize the theoretical results we will prove in Section 4 in Table 1, where we compare BRWP with other popular methods. Sampling complexity refers to the number of iterations needed to achieve a given level of accuracy.

Algorithm	Number of iterations required to achieve $D_{\text{KL}}(\rho_k \ \rho^*) \leq \delta$	Maximum stepsize
BRWP (this paper)	$\mathcal{O}\left(\frac{ \ln \delta }{2\alpha\delta^{1/2}}\right)$	$\frac{2}{3\alpha}$
ULA ((Vempala and Wibisono, 2019))	$\mathcal{O}\left(\frac{dL^2}{\alpha^2\delta}\right)$	$\frac{\alpha}{4L^2}$
Underdamped Langevin ((Ma et al., 2021))	$\mathcal{O}\left(\frac{d^{1/2}(L^{3/2} + d^{1/2}K)}{\alpha^2\delta^{1/2}}\right)$	$\mathcal{O}\left(\frac{\alpha}{L^{3/2}}\right)$
Proximal Langevin ((Wibisono, 2019))	$\mathcal{O}\left(\frac{d^{1/2}(L^{3/2} + dK)}{\alpha^{3/2}\delta^{1/2}}\right)$	$\min\left\{\frac{1}{8L}, \frac{1}{K}, \frac{3\alpha}{32L^2}\right\}$

Table 1: Iteration complexities for sampling algorithms under Assumptions 1-3 in Section 2.4 and the score-oracle Assumption 4 in Section 4.

In the next section, we shall focus on investigating the approximate density given by the kernel formula (17) and the regularity property of the discrete density from the BRWP iteration.

1. For BRWP, the stepsize is additionally restricted by a positive number h_0 ensuring the regularity of ρ_k (Section 3); thus, the effective maximal stepsize is $\min\{h_0, 2/(3\alpha)\}$.

3. Kernel Approximation and Regularity of the BRWP Update

In this section, we mainly study and utilize the properties of the kernel formula \mathcal{K}_V^h underlying the BRWP update. We show that \mathcal{K}_V^h yields a weak $\mathcal{O}(h^2)$ approximation of the Fokker-Planck evolution and that the densities generated by the BRWP iteration remain uniformly locally regular and integrable. All approximation results are understood in a weak, local sense; global estimates follow from suitable moment bounds and localization arguments.

The motivation is the following. The semi-implicit BRWP scheme requires, at each step, an approximation of the next-step score function $\nabla \log \rho_{k+1}$ via the kernel-based quantity $\nabla \log(\mathcal{K}_V^h \rho_k)$. To control the global error of the BRWP algorithm, it is therefore essential to establish two properties: (i) the kernel formula \mathcal{K}_V^h approximates the Fokker-Planck flow with second-order weak accuracy, and (ii) the resulting densities retain sufficient regularity so that the score evaluation remains well posed throughout the iteration.

The key observation is that \mathcal{K}_V^h admits a short-time expansion whose leading-order terms coincide with the generator of the Fokker-Planck equation. Under Assumption 1, this yields a second-order weak Taylor expansion: when tested against smooth functions, the discrepancy between $\mathcal{K}_V^h \rho$ and the exact Fokker-Planck evolution is of order $\mathcal{O}(h^2)$. This expansion forms the basis of our weak error analysis.

At the same time, Assumptions 2–3 guarantee stability of the BRWP update. The Gaussian structure of \mathcal{K}_V^h induces a localized smoothing effect that propagates uniform $C^{3,1}$ regularity on bounded sets, while the polynomial growth conditions on V ensure that densities remain integrable along the iteration. As a consequence, the BRWP densities stay uniformly locally regular and bounded at every step.

Taken together, these results demonstrate that \mathcal{K}_V^h serves as both an accurate local approximation of the Fokker-Planck flow and a stable update mechanism for the approximate score function in the BRWP scheme, thereby providing the analytical foundation for the convergence analysis presented in the subsequent sections.

We begin by introducing the notation and operators used throughout this section. For a smooth test function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, the Langevin generator and its adjoint are

$$(\mathcal{L}\varphi)(x) = \beta^{-1} \Delta \varphi(x) - \nabla V(x) \cdot \nabla \varphi(x), \quad (\mathcal{L}^* \rho)(x) = \beta^{-1} \Delta \rho(x) + \nabla \cdot (\rho \nabla V(x)). \quad (18)$$

Define

$$\psi(x) := e^{-\frac{\beta}{2} V(x)}, \quad \rho^*(x) = \frac{1}{Z} \psi(x)^2.$$

The introduction of ψ is motivated by the weight that appears in \mathcal{K}_V^h in (17). Let \mathcal{G}_h denote the Gaussian heat semigroup,

$$(\mathcal{G}_h f)(y) = \int_{\mathbb{R}^d} G_h(x-y) f(x) dx,$$

and write $Z := \mathcal{G}_h \psi$. Then, for test functions φ and densities ρ , the dual kernel operator takes the form

$$P_V^h \varphi(y) := \frac{\mathcal{G}_h(\psi \varphi)(y)}{Z(y)}, \quad \langle \varphi, \mathcal{K}_V^h \rho \rangle = \langle P_V^h \varphi, \rho \rangle, \quad (19)$$

where $\langle u, v \rangle = \int_{\mathbb{R}^d} u(x) v(x) dx$. With the above notation, the kernel formula can be written as

$$\mathcal{K}_V^h f(x) = \psi(x) \int_{\mathbb{R}^d} \frac{G_h(x-y) f(y)}{Z(y)} dy. \quad (20)$$

Equivalently, P_V^h admits the Gibbs-type representation

$$P_V^h \varphi(y) = \mathbb{E}_{\mu_y}[\varphi(X)], \quad (21)$$

where

$$U_y(x) := \frac{\beta}{2}V(x) + \frac{\beta}{4h}\|x - y\|^2, \quad \mu_y(dx) = \frac{e^{-U_y(x)}}{\int_{\mathbb{R}^d} e^{-U_y(x)} dx} dx.$$

We will compare three families of densities, all starting from the same initial data ρ_0 :

- $\rho(t_k)$: the exact Fokker-Planck solution at time t_k ;
- $\widehat{\rho}_k$: the kernel-propagated densities $\widehat{\rho}_{k+1} = \mathcal{K}_V^h \widehat{\rho}_k$;
- ρ_k : the BRWP iterates in the Algorithm 1 which can be written equivalently as

$$\rho_{k+1} = (F_k)_\# \rho_k, \quad F_k(x) = x - h(\nabla V(x) + \beta^{-1} \nabla \log(\mathcal{K}_V^h \rho_k)(x)).$$

For a finite signed measure μ supported in a bounded set $U \subset \mathbb{R}^d$, define

$$\|\mu\|_{(C^{1,1}(U))^*} := \sup_{\|\varphi\|_{C^{1,1}(U)} \leq 1} |\langle \varphi, \mu \rangle|, \quad \|\varphi\|_{C^{1,1}(U)} := \|\varphi\|_{L^\infty(U)} + \|\nabla \varphi\|_{L^\infty(U)} + [\nabla \varphi]_{\text{Lip}(U)}, \quad (22)$$

$$\|\mu\|_{(C_0^{1,1}(U))^*} := \sup_{\|\varphi\|_{C_0^{1,1}(U)} \leq 1} |\langle \varphi, \mu \rangle|, \quad \|\varphi\|_{C_0^{1,1}(U)} := \|\nabla \varphi\|_{L^\infty(U)} + [\nabla \varphi]_{\text{Lip}(U)}. \quad (23)$$

Higher-order norms $\|\cdot\|_{C^{m,1}(U)}$ are defined analogously. The dual norm above appears naturally in the weak formulation that we will work with in this section. Moreover, we write the weighted p moment as

$$M_p(\mu) := \int_{\mathbb{R}^d} (1 + \|y\|)^p |\mu(y)| dy.$$

For the following analysis, we fix bounded open sets with smooth boundaries and define

$$U \Subset U_r \subset \mathbb{R}^d, \quad r_0 := \text{dist}(U, \partial U_r) > 0.$$

The set U is the region on which all local expansions and regularity estimates are performed, while U_r serves as a slightly enlarged localization domain that contains all near-field interactions generated by the kernel for sufficiently small step sizes. Unless otherwise specified, all $C^{m,1}$ -norms are taken on U . All constants appearing below may depend on d, β, α , the domains $U \Subset U_r$, and $\|V\|_{C^{3,1}(U_r)}$, but are uniform in the step size h and the iteration index k . The choice of U depends on β and the maximum admissible stepsize. Further details on the specification of U are provided in the remark following Theorem 4.

This localization allows us to cleanly separate near-field and far-field contributions in the kernel expansions. Near-field interactions, corresponding to $\|x - y\| \leq r_0$, remain inside U_r and are handled by local Taylor expansions and regularity estimates. Far-field contributions are controlled using the Gaussian decay of the kernel, together with the polynomial growth condition on V (Assumption 2) and uniform moment bounds. This structure is essential for establishing uniform local $C^{3,1}$ bounds on the discrete densities and for proving the $\mathcal{O}(h^2)$ weak error estimates. Since U_r is a fixed localization domain satisfying $U \Subset U_r$, we will often state estimates only on U , with the understanding that all constants and expansions implicitly depend on the ambient domain U_r .

The detailed proof for results in this section can be found in the Appendix A.1.

3.1 Weak second-order kernel approximation of the Fokker-Planck flow

Firstly, we establish that the kernel formula provides a weak approximation of the Fokker-Planck evolution on a compact set. The proof is based on successive Taylor expansions of smooth functions, combined with integration by parts and moment estimates for the Gaussian kernel.

Lemma 3 (Local weak $O(h^2)$ expansion) For $\varphi \in C^{2,1}(U)$ supported in U , there exists $h_0 > 0$ such that for all $h \in (0, h_0]$,

$$|\langle \varphi, \mathcal{K}_V^h u - u - h\mathcal{L}^* u \rangle| = |\langle P_V^h \varphi - \varphi - h\mathcal{L}\varphi, u \rangle| \leq C h^2 \|\varphi\|_{C^{2,1}(U)} \|u\|_{C^{0,1}(U)}, \quad (24)$$

for some constant $C = C(d, \beta, U_r, \|V\|_{C^{3,1}(U_r)})$.

Proof [Sketch of proof] We only indicate the main ideas and refer to the appendix for full details. For $x \in \mathbb{R}^d$, write $\mathcal{G}_h[f(x, y)] = \int_{\mathbb{R}^d} G_h(y) f(x, y) dy$ where we integrate with respect to y , then

$$P_V^h \varphi(x) = \varphi(x) + \frac{N(x)}{Z(x)}, \quad N(x) := \mathcal{G}_h[\psi(x+y)(\varphi(x+y) - \varphi(x))], \quad Z(x) := \mathcal{G}_h[\psi(x+y)]. \quad (25)$$

Choose

$$r_h := \min\left\{\frac{r_0}{2}, \sqrt{\frac{12}{\beta} h \log(1/h)}\right\}.$$

Standard Gaussian tail estimates imply $\mathcal{G}_h[\mathbf{1}_{\{\|y\| > r_h\}}] \leq Ch^2$ uniformly in $x \in U$. Consequently, for any bounded integrand and any $u \in C^{0,1}(U)$, the contribution of the region $\{\|y\| > r_h\}$ to the weak pairing is of order $\mathcal{O}(h^2)$. Since $r_h \leq r_0/2$, we have $x+B(0, r_h) \subset U_r$ for all $x \in U$. Therefore, up to an $\mathcal{O}(h^2)$ error in the weak sense, we may restrict the y -integration to the ball $B(0, r_h)$, on which all local Taylor expansions are valid.

For $\|y\| \leq r_h$, expand

$$\psi(x+y) = \psi(x) + \nabla\psi(x) \cdot y + \frac{1}{2} y^\top \nabla^2 \psi(x) y + \frac{1}{6} D^3 \psi(x)[y, y, y] + R_\psi(x, y),$$

with $|R_\psi(x, y)| \leq C\|y\|^4$.

Using

$$\mathcal{G}_h[y] = 0, \quad \mathcal{G}_h[yy^\top] = 2h\beta^{-1}I_d, \quad \mathcal{G}_h[y_i y_j y_k] = 0, \quad \mathcal{G}_h[\|y\|^4] = \mathcal{O}(h^2),$$

together with Stein's identity $\mathcal{G}_h[y \cdot f(y)] = 2h\beta^{-1} \mathcal{G}_h[\nabla_y \cdot f(y)]$, all cubic terms vanish and only the even-order terms survive. This yields

$$Z(x) = \psi(x) + h\beta^{-1} \Delta\psi(x) + \mathcal{O}(h^2 \|\psi\|_{C^{3,1}(U_r)}),$$

uniformly on U . Insert the Taylor expansions of $\psi(x+y)$ and $\varphi(x+y) - \varphi(x)$ into $N(x)$ and keep only terms contributing up to order h . Using the Gaussian identities above and $\psi = e^{-\beta V/2}$, one obtains

$$\left\langle \frac{N(x)}{Z(x)}, u \right\rangle = h \langle \beta^{-1} \Delta\varphi - \nabla V \cdot \nabla\varphi, u \rangle + \mathcal{O}(h^2 \|\varphi\|_{C^{2,1}(U)} \|u\|_{C^{0,1}(U)}).$$

To obtain the above $\mathcal{O}(h^2)$ weak remainder, one needs four total derivatives on the integrand: the kernel expansion produces up to third-order derivatives of φ via Stein's identity, while one derivative comes from integrating by parts against u . The intermediate orders $h^{1/2}$ and $h^{3/2}$ vanish because all odd Gaussian moments are zero.

This proves the claimed local weak $\mathcal{O}(h^2)$ expansion. ■

Subsequently, we extend the approximation in Lemma 3 to globally supported test functions by imposing suitable polynomial moment assumptions. These moment bounds, together with Gaussian tail estimates, enable a truncation of the far-field contribution and control of the resulting error.

To carry this out, we employ a standard localization argument: a smooth cutoff χ is introduced to decompose φ and u into compactly supported inner parts, on which the local expansion on U_r applies directly, and outer parts supported in U^c . The near-field interactions between inner and outer regions remain inside U_r when h is small, and thus also satisfy the local $O(h^2)$ estimate. The remaining far-field interactions, with $\|x - y\| > r_h$, are then controlled using the moment bound M_p and the rapid Gaussian decay of the kernel, which together imply that the far region contributes only $O(h^2)$. Combining these bounds yields the global weak expansion in the following theorem.

Theorem 4 (Global weak $O(h^2)$ expansion with unbounded supports) *Let $\varphi \in C_{\text{loc}}^{2,1}(\mathbb{R}^d)$ and $u \in C_{\text{loc}}^{0,1}(\mathbb{R}^d)$ have unbounded support and satisfy the weighted moment bound*

$$M_p := \int_{\mathbb{R}^d} (1 + \|x\|)^p (|u(x)| + |\varphi(x)|) dx < \infty, \quad p > 2 + q_V.$$

Then there exists $h_0 > 0$ and a constant

$$C = C(d, \beta, C_V, q_V, p, M_p) < \infty$$

such that, for any fixed bounded open set $U \subset \mathbb{R}^d$ and all $h \in (0, h_0]$,

$$|\langle \varphi, \mathcal{K}_V^h u - u - h\mathcal{L}^* u \rangle| = \left| \langle P_V^h \varphi - \varphi - h\mathcal{L}\varphi, u \rangle \right| \leq C h^2 \left(\|\varphi\|_{C^{2,1}(U)} \|u\|_{C^{0,1}(U)} + 1 \right). \quad (26)$$

The dependence of the step-size restriction h_0 on the bounded set U enters only through the Gaussian tail estimates used in the proofs of Lemma 3 and Theorem 4. We illustrate this dependence explicitly. Fix $R > 0$ and assume that

$$U \subset B(0, R) \subset B(0, 2R) \subset U_r.$$

For any $x \in U$, we have $\|x\| \leq R$, and therefore $\|x - y\| \geq R$ for all $y \in U_r^c$. It follows that

$$\int_{U_r^c} G_h(x - y) dy \leq \int_{\{\|z\| \geq R\}} G_h(z) dz \leq C \exp\left(-\frac{\beta R^2}{4h}\right),$$

where the last inequality is the standard Gaussian tail bound. The estimate holds uniformly for all $x \in U$. Consequently, there exists $h_0(R) \in (0, 1]$ such that

$$C \exp\left(-\frac{\beta R^2}{4h}\right) \leq h^2, \quad 0 < h \leq h_0(R),$$

which ensures that $\int_{U_r^c} G_h(x - y) dy \leq h^2$ uniformly for $x \in U$. This shows that the contribution from U_r^c is of order $\mathcal{O}(h^2)$ and justifies the truncation used in the weak expansion.

Moreover, through one more step of integration by parts, it is straightforward to show

$$|\langle \varphi, \mathcal{K}_V^h u - u - h\mathcal{L}^* u \rangle| \leq C h^2 (\|\varphi\|_{C^{1,1}(U)} \|u\|_{C^{1,1}(U)} + 1), \quad (27)$$

or

$$|\langle \varphi, \mathcal{K}_V^h u - u - h\mathcal{L}^* u \rangle| \leq C h^2 (\|\varphi\|_{C^{0,1}(U)} \|u\|_{C^{2,1}(U)} + 1), \quad (28)$$

following the proof in Lemma 3 and Theorem 4.

We now quantify how the weak local error of the kernel formula \mathcal{K}_V^h accumulates along the discrete density propagation. The next result provides control over both the local truncation error and the error inherited from previous steps.

Theorem 5 (Weak score oracle with propagated base error) *Consider the kernel formula iteration $\widehat{\rho}_{k+1} = \mathcal{K}_V^h \widehat{\rho}_k$. There exists $h_0 \in (0, 1]$ such that, for every $h \in (0, h_0]$, every test function $\varphi \in C_{\text{loc}}^{2,1}(\mathbb{R}^d)$, and densities $\rho(t_k), \widehat{\rho}_k \in C_{\text{loc}}^{0,1}(\mathbb{R}^d)$ satisfying the weighted moment condition of Theorem 4, one has*

$$\begin{aligned} \left| \langle \varphi, \widehat{\rho}_{k+1} - \rho(t_k) - h\mathcal{L}^* \rho(t_k) \rangle \right| &\leq C h^2 (\|\varphi\|_{C^{2,1}(U)} \|\rho(t_k)\|_{C^{0,1}(U)} + 1) \\ &\quad + \frac{1}{1 + \alpha h} \|\widehat{\rho}_k - \rho(t_k)\|_{(C^{0,1})^*(U)} \|\varphi\|_{C^{0,1}(U)}. \end{aligned} \quad (29)$$

Moreover, if at each step the oracle is refreshed so that the past error vanishes, then the propagation term in (29) disappears, and

$$\left| \langle \varphi, \mathcal{K}_V^h(\rho(t_k)) - \rho(t_k) - h\mathcal{L}^* \rho(t_k) \rangle \right| \leq C h^2 (\|\varphi\|_{C^{2,1}(U)} \|\rho(t_k)\|_{C^{0,1}(U)} + 1).$$

Proof Add and subtract $\mathcal{K}_V^h \rho(t_k)$ and apply duality:

$$\langle \varphi, \widehat{\rho}_{k+1} - \rho(t_k) - h\mathcal{L}^* \rho(t_k) \rangle = \underbrace{\langle \varphi, \mathcal{K}_V^h \rho(t_k) - \rho(t_k) - h\mathcal{L}^* \rho(t_k) \rangle}_{\text{local truncation}} + \underbrace{\langle P_V^h \varphi, \widehat{\rho}_k - \rho(t_k) \rangle}_{\text{propagated base error}}.$$

By Theorem 4, the local truncation term is bounded by $C h^2 (\|\varphi\|_{C^{2,1}} \|\rho(t_k)\|_{C^{0,1}} + 1)$. For the propagated error, Lemma 18 yields the $C^{0,1}$ -stability estimate

$$\|P_V^h \varphi\|_{C^{0,1}(U)} \leq \frac{1}{1 + \alpha h} \|\varphi\|_{C^{0,1}(U)}.$$

Substituting this into the dual pairing gives (29). ■

The weak expansion established in Theorem 5 shows that the kernel operator \mathcal{K}_V^h approximates the Fokker–Planck flow with second-order accuracy in the weak sense, up to a contraction of previously accumulated errors in the dual Lipschitz norm. This estimate holds provided the test functions admit a uniform $C^{2,1}$ bound and the evolving densities are uniformly controlled in $C^{0,1}$ on bounded sets.

In the convergence analysis of the BRWP scheme, however, test functions are no longer arbitrary: they will be replaced by expressions involving the densities themselves, including the KL divergence and the Fisher information. To justify such substitutions and the associated integration-by-parts identities, it is therefore necessary to establish higher regularity and uniform positivity of the discrete densities generated by the algorithm.

Accordingly, in the next subsection we study the densities produced by the BRWP iteration and prove that the sequence $\{\rho_k\}$ enjoys uniform local $C^{3,1}$ bounds under Assumptions 1–3. This regularity ensures that the weak second-order accuracy of \mathcal{K}_V^h remains valid when test functions are replaced by density-dependent quantities, thereby allowing the weak expansion to propagate through the full BRWP algorithm and into the KL-based convergence analysis.

3.2 Uniform local regularity of the BRWP iterates

We now show that the iterates ρ_k remain uniformly bounded in $C^{3,1}(U)$. The argument has two main components.

First, we compare the BRWP sequence ρ_k with the smooth reference sequence $\widehat{\rho}_k$ generated by repeated applications of the kernel formula. The construction of $\widehat{\rho}_k$ makes it easier to obtain uniform $C^{3,1}$ estimates, and the comparison $\rho_k - \widehat{\rho}_k$ allows us to transfer these bounds back to the original BRWP iterates.

Second, we rely on a lifting argument based on the localized resolvent equation $(I - \mathcal{K}_V^h)u = \mu$ where μ represents the weak error accumulated at each iteration. This equation exploits the smoothing property of \mathcal{K}_V^h to upgrade moment bounds on a measure μ into $C^{3,1}$ control of the corresponding solution u . We begin by stating the key lemma that encodes this lifting property.

Lemma 6 (Lifting lemma for localized resolvent in $C^{3,1}$) *Let $\eta \in C^\infty(U_r)$ satisfy $\eta \equiv 1$ on U . Let μ be a finite signed measure on U_r with $M_p(\mu) < \infty$, for $p > d + 2$. For $h \in (0, h_0]$, consider the localized resolvent equation*

$$u - \mathcal{K}_V^h u = \eta\mu, \quad u|_{\partial U_r} = 0. \quad (30)$$

Then (30) admits a unique solution $u \in C^{3,1}(U)$ satisfying

$$\|u\|_{C^{3,1}(U)} \leq C M_p(\mu), \quad (31)$$

where $C = C(d, U, U_r, p, \|V\|_{C^{4,1}(U_r)})$ is independent of $h \in (0, h_0]$.

Proof [Sketch of proof] Writing $u = \psi v$ and using the kernel formula $\mathcal{K}_V^h f = \psi \mathcal{G}_h(Z^{-1}f)$ with the heat-kernel expansion $Z = \psi + h\Delta\psi + O(h^2)$, we obtain

$$v - \mathcal{G}_h v = \mathcal{G}_h((w_h - 1)v) + \psi^{-1}\eta\mu, \quad w_h := \psi/Z = 1 + O(h).$$

Next, decompose $\mathcal{G}_h = \mathcal{G}_h^D + R_h$ using the Dirichlet heat kernel on U_r . The boundary remainder satisfies the Gaussian smallness estimate

$$\|R_h\|_{L^1(U_r) \rightarrow C^{3,1}(U)} \leq C e^{-cr_0^2/h}, \quad r_0 := \text{dist}(U, \partial U_r),$$

so that applying the Dirichlet resolvent $\mathcal{R}_h^D = (I - \mathcal{G}_h^D)^{-1} = \sum_{n \geq 0} (\mathcal{G}_h^D)^n$ gives

$$v = (\mathcal{A}_h^{(0)} + \mathcal{A}_h^{(1)} + \mathcal{A}_h^{(2)})v + \mathcal{R}_h^D(\psi^{-1}\eta\mu),$$

where $\mathcal{A}_h^{(0)}$, $\mathcal{A}_h^{(1)}$, and $\mathcal{A}_h^{(2)}$ collect the terms involving R_h and $(w_h - 1)$.

Since the condition $p > d + 2$ ensures that the Dirichlet resolvent maps measures with finite p -moment into $C^{3,1}(U)$, the Heat-kernel bounds in Lemma 16 imply

$$\|\mathcal{A}_h^{(0)}\| \leq C e^{-cr_0^2/h}, \quad \|\mathcal{A}_h^{(1)}\| \leq Ch, \quad \|\mathcal{A}_h^{(2)}\| \leq Ch e^{-cr_0^2/h}.$$

Choosing $h_0 > 0$ small ensures $\|\mathcal{A}_h^{(0)} + \mathcal{A}_h^{(1)} + \mathcal{A}_h^{(2)}\| \leq \frac{1}{2}$, for $h \leq h_0$, so the equation for v can be inverted by a Neumann series:

$$v = \left(I - \mathcal{A}_h^{(0)} - \mathcal{A}_h^{(1)} - \mathcal{A}_h^{(2)} \right)^{-1} \mathcal{R}_h^D(\psi^{-1}\eta\mu).$$

By Lemma 17, the Dirichlet resolvent satisfies the interior estimate

$$\|\mathcal{R}_h^D \nu\|_{C^{3,1}(U)} \leq C M_p(\nu).$$

Since $\psi^{-1}\eta$ is smooth and bounded, $\|v\|_{C^{3,1}(U)} \leq C M_p(\mu)$.

Finally, $\psi \in C^{3,1}(U_r)$ implies

$$\|u\|_{C^{3,1}(U)} = \|\psi v\|_{C^{3,1}(U)} \leq C \|v\|_{C^{3,1}(U)} \leq C M_p(\mu). \quad \blacksquare$$

Finally, we establish uniform local $C^{3,1}$ regularity for the BRWP iterates. The argument is based on a lifting mechanism for weak errors through a localized resolvent equation.

Theorem 7 (Uniform local $C^{3,1}$ bound for the BRWP iterates) *Let $0 < h \leq h_0$, where $h_0 \in (0, 1]$ is the constant in Lemma 6. Then there exists a constant*

$$C_* = C_*(d, \beta, \alpha, \|V\|_{C^{4,1}(U_r)}, U, U_r, p, M_p(\rho_0), h_0) < \infty$$

such that, for all $k \geq 0$,

$$\|\rho_k\|_{C^{3,1}(U)} \leq C_*. \quad (32)$$

In particular, C_* is independent of the iteration index k and the step size $h \in (0, h_0]$.

Proof [Sketch of proof] Let $\epsilon_k := \rho_k - \widehat{\rho}_k$ denote the difference between the BRWP iterate and the kernel-propagated reference density. On the localization domain U , this error satisfies a resolvent relation of the form

$$(I - \mathcal{K}_V^h)u = \epsilon_k.$$

Applying Lemma 6 with $\mu = \epsilon_k$ yields a function $u_k \in C^{3,1}(U)$ such that

$$\epsilon_k = u_k - \mathcal{K}_V^h u_k, \quad \|u_k\|_{C^{3,1}(U)} \leq C M_p(\epsilon_k).$$

Since the kernel operator \mathcal{K}_V^h is bounded on $C^{3,1}(U)$ (Lemma 19), this representation lifts the weak control of ϵ_k into a strong bound,

$$\|\epsilon_k\|_{C^{3,1}(U)} \leq C M_p(\epsilon_k).$$

The same argument applies to the reference sequence $(\widehat{\rho}_k)$. Combined with the uniform moment bounds, this yields

$$\|\widehat{\rho}_k\|_{C^{3,1}(U)} \leq C,$$

uniformly in k . Since $\rho_k = \widehat{\rho}_k + \epsilon_k$ and both components are uniformly controlled, the BRWP iterates inherit a uniform local $C^{3,1}$ bound. \blacksquare

The $C^{3,1}$ bounds established above are local in space and depend on the choice of the bounded set $U \Subset \mathbb{R}^d$. We do not claim global uniform $C^{3,1}$ regularity on \mathbb{R}^d . Instead, the behavior at infinity is controlled by Assumption 2 together with the uniform p -moment bounds from Lemma 15. These conditions ensure the required weighted integrability of the densities (and of the kernel expressions appearing in the analysis), so that all global integrals and Gaussian integration-by-parts identities used in Section 4 are well defined, with constants uniform in k .

3.3 Weak error of the kernel-based score oracle

We now establish the key error estimate for the score oracle produced by the kernel formula. Our goal is to show that $\nabla \log(\mathcal{K}_V^h \rho_k)$ provides an $\mathcal{O}(h^2)$ approximation to the score function at the next time point $\nabla \log \rho_{k+1}$, thereby ensuring that the BRWP iteration acts as a semi-implicit discretization in the score function. We emphasize that the next theorem is phrased in terms of the Langevin generator \mathcal{L}^* : indeed, the quantity

$$\nabla \log \rho_k + h \nabla \left(\frac{\mathcal{L}^* \rho_k}{\rho_k} \right)$$

also serves as the approximation to $\nabla \log \rho_{k+1}$ that will be verified in Section 4.

Theorem 8 (Weak score-oracle) *Let ρ_k be the BRWP iterates, and let $\tilde{\rho}_k$ be any density satisfying the same regularity bounds as ρ_k . Then, for every vector field $w \in C^{1,1}(U; \mathbb{R}^d)$ and every $k \geq 0$,*

$$\left| \int_U w \cdot \left[\nabla \log(\mathcal{K}_V^h \tilde{\rho}_k) - \nabla \log \rho_k - h \nabla \left(\frac{\mathcal{L}^* \rho_k}{\rho_k} \right) \right] \rho_k dx \right| \leq C \left(h^2 + \|\tilde{\rho}_k - \rho_k\|_{C^{2,1}(U)} \right) \|w\|_{C^{1,1}(U)}. \quad (33)$$

The constant C depends only on the parameters and the uniform $C^{3,1}$ bounds for $(\rho_k)_{k \geq 0}$.

Proof [Sketch of proof.] The proof follows the idea of the weak $O(h^2)$ expansion for test functions. Fix $k \geq 0$ and write $r := \rho_k$, $r_K := \mathcal{K}_V^h \rho_k$, and $r_* := r + h \mathcal{L}^* r$. We decompose the error in approximating the score function into an ideal kernel contribution and an oracle perturbation:

$$\begin{aligned} & \int w \cdot \left[\nabla \log(\mathcal{K}_V^h \tilde{\rho}_k) - \nabla \log r - h \nabla \left(\frac{\mathcal{L}^* r}{r} \right) \right] r dx \\ &= \int w \cdot \left[\nabla \log r_K - \nabla \log r - h \nabla \left(\frac{\mathcal{L}^* r}{r} \right) \right] r dx + \int w \cdot \left[\nabla \log(\mathcal{K}_V^h \tilde{\rho}_k) - \nabla \log(\mathcal{K}_V^h \rho_k) \right] \rho_k dx \end{aligned}$$

where the score is well defined due to the uniform positivity of ρ_k by Lemma 20. We denote the first and second terms above as I^0 and I^1 .

For I^0 , we insert and subtract $\nabla \log r_*$ and exploit positivity and $C^{3,1}$ regularity of all densities to rewrite I^0 as integrals against $r_K - r_*$. The coefficients appearing in these integrals lie in $C^{0,1}$ with norms controlled by $\|w\|_{C^{1,1}(U)}$. Applying the weak $O(h^2)$ expansion for the kernel formula (Theorem 4) yields $I^0 = O(h^2 \|w\|_{C^{1,1}(U)})$.

For I^1 , the $C^{3,1}$ -smoothing property of \mathcal{K}_V^h implies $\|\mathcal{K}_V^h \tilde{\rho}_k - \mathcal{K}_V^h \rho_k\|_{C^1(U)} \leq C \|\tilde{\rho}_k - \rho_k\|_{C^{2,1}(U)}$. Since the map $\rho \mapsto \nabla \log \rho$ is Lipschitz from $C^{2,1}(U)$ to $C^1(U)$ on positive densities, we obtain

$$|I^1| \leq C \|w\|_{C^{1,1}(U)} \|\tilde{\rho}_k - \rho_k\|_{C^{2,1}(U)}.$$

Combining the bounds on I^0 and I^1 gives the claimed weak score oracle approximation with propagated base error. \blacksquare

We remark that the above theorem applies if we replace ρ_k by the exact solution to the Fokker-Planck equation $\rho(t_k)$. The local regularity of $\rho(t_k)$ follows from the global hypo-coercive regularity for the Fokker-Planck semigroup with strongly convex potential V (see, e.g., (Villani, 2009)).

4. Convergence Analysis of the BRWP Update in KL Divergence

After establishing the approximation properties of the kernel operator \mathcal{K}_V^h in Section 3, we now turn to the convergence analysis of the BRWP update. Our objective is to quantify the decay of the KL divergence along the iteration by deriving a weak one-step expansion of the density evolution.

Recall that the BRWP iteration is given by

$$x_{k+1} = x_k - h \left(\nabla V(x_k) + \beta^{-1} \nabla \log \tilde{\rho}_{k+1}(x_k) \right), \quad (34)$$

where the next-step density is approximated by $\tilde{\rho}_{k+1} = \mathcal{K}_V^h \rho_k$.

To streamline the convergence analysis, we isolate the numerical property of the score approximation required in this section, namely a weak second-order consistency condition.

This is not an additional analytical assumption for the BRWP scheme: for the kernel-based update $\tilde{\rho}_{k+1} = \mathcal{K}_V^h \rho_k$, the condition has already been verified constructively in Theorem 8. We state it explicitly in order to decouple the analytical convergence argument from the numerical realization of the kernel formula, which is discussed separately in Section 5.

Viewed in this way, the following assumption serves as a numerical interface assumption: it captures the minimal weak accuracy required of the score approximation without committing to a specific implementation strategy. The subsequent analysis will still rely on the regularity and moment assumptions already established for the BRWP scheme, and is not intended to cover alternative sampling schemes. Moreover, the condition is understood to hold only for stepsizes $0 < h \leq h_0$, where $h_0 > 0$ is the maximal range on which the $O(h^2)$ weak expansion and the regularity estimates established in Section 3 remain valid; all stepsize conditions below are implicitly subject to this restriction.

Assumption 4 (Weak score oracle) *For any fixed bounded open set $U \subset \mathbb{R}^d$,*

$$\left| \int_{\mathbb{R}^d} w \cdot \left[\nabla \log \tilde{\rho}_{k+1} - \nabla \log \rho_k - h \nabla \left(\frac{\mathcal{L}^* \rho_k}{\rho_k} \right) \right] \rho_k dx \right| \leq C h^2 \|w\|_{C^{1,1}(U)}, \quad (35)$$

for all $w \in C^{1,1}(U; \mathbb{R}^d)$, uniformly in $k \geq 0$.

We refer to $\nabla \log \tilde{\rho}_{k+1}$ as a weak score oracle. The above $O(h^2)$ bound means that, when tested against smooth vector fields, the approximate score reproduces the first two terms of the short-time Fokker–Planck expansion in a weak sense; see Lemma 9.

Throughout this section, all results are derived under Assumptions 1–3 stated in Section 2.4. All $\mathcal{O}(\cdot)$ terms are understood to be uniform on compact subsets of \mathbb{R}^d . Unless stated otherwise, constants may depend on the potential V , the dimension d , the inverse temperature β , and the local domains U , but are independent of the stepsize h and the iteration index k .

By Theorem 7, the BRWP iterates ρ_k enjoy uniform local $C^{3,1}$ regularity on U , and Lemma 20 ensures uniform positivity on U . Consequently, $\log(\rho_k/\rho^*)$ is well defined in $C^{3,1}(U)$, and all differential expressions appearing below—such as $\nabla \log \rho_k$ and $\nabla \cdot (\rho_k \nabla u)$ —are well defined pointwise.

Moreover, the uniform p -moment bounds from Lemma 15 guarantee sufficient decay of ρ_k at infinity. As a result, integration by parts over \mathbb{R}^d is justified, and the identity

$$\int_{\mathbb{R}^d} f \nabla \cdot (\rho_k \nabla g) dx = - \int_{\mathbb{R}^d} \langle \nabla f, \nabla g \rangle \rho_k dx$$

holds rigorously.

Finally, the localization domain U is chosen sufficiently large so that the cutoff radius condition in the proof of Theorem 4 is satisfied for all admissible stepsizes considered in this section.

4.1 Weak one-step expansion of the BRWP density

In this subsection, we derive the weak one-step expansion of the BRWP update under an $O(h^2)$ -accurate score oracle. This expansion provides the discrete analogue of the infinitesimal generator identity for the Fokker–Planck flow. The calculation is carried out in the weak sense, and all derivatives fall on smooth test functions.

Lemma 9 (Weak expansion of the BRWP density update) *Let $(\rho_k)_{k \geq 0}$ be the BRWP iterates. Let $u \in C_{loc}^{3,1}(\mathbb{R}^d)$ satisfy $M_p(u) < \infty$ for $p > d + 3$. Then*

$$\begin{aligned} \langle u, \rho_{k+1} - \rho_k \rangle &= h \left\langle u, \nabla \cdot \left(\rho_k \nabla \left(\beta^{-1} \log \frac{\rho_k}{\rho^*} \right) \right) \right\rangle \\ &\quad - h^2 \beta^{-1} \left\langle u, \nabla \cdot \left(\rho_k \nabla \left(\frac{\mathcal{L}^* \rho_k}{\rho_k} \right) \right) \right\rangle + \frac{h^2}{2} \langle u, \nabla \cdot (\rho_k \zeta_k) \rangle + \mathcal{O}(h^3 \|u\|_{C^{3,1}(U)}), \end{aligned} \quad (36)$$

where $U \Subset \mathbb{R}^d$ and

$$\zeta_k := \nabla \cdot ((\nabla \Phi_k)(\nabla \Phi_k)^\top) + (\nabla \Phi_k \cdot \nabla \log \rho_k) \nabla \Phi_k, \quad \text{where } \Phi_k = \beta^{-1} \log \frac{\rho_k}{\rho^*},$$

Here $\nabla \cdot [A]$ denotes the vector field whose i th component is $\sum_j \partial_{x_j} A_{ij}$.

Proof [Proof sketch] The full proof is given in Appendix B. We outline the main steps.

A cutoff argument reduces the proof to a compact domain. The BRWP update is expanded using a second-order Taylor expansion of the transport map around x_k , with all derivatives falling on the test function u . The weak score oracle property (Assumption 4), verified for the kernel-based BRWP update in Section 3, is used precisely to replace $\nabla \log \tilde{\rho}_{k+1}$ by $\nabla \log \rho_k + h \nabla (\mathcal{L}^* \rho_k / \rho_k)$ in the weak sense. Collecting all terms yields the stated expansion. \blacksquare

We remark that the moment condition on u is needed only to justify passing to the limit in the cutoff argument: dominated convergence requires $(1 + \|x\|)^p u(x) \in L^1$ which already suffices since the BRWP iterates have uniformly bounded p -moments (Lemma 15). The stronger assumption $p > d + 3$ is imposed only for consistency with the following discussion, where repeated integration by parts and the polynomial growth of the derivatives of V introduce higher-order polynomial factors that must remain integrable.

To express the expansion more compactly, introduce the weighted elliptic operator

$$\mathcal{D}_k^\beta(u) := \frac{\beta^{-1}}{\rho_k} \nabla \cdot (\rho_k \nabla u), \quad \mathcal{D}_k^\beta(u) \rho_k = \beta^{-1} \nabla \cdot (\rho_k \nabla u). \quad (37)$$

Using the identity

$$\frac{\mathcal{L}^* \rho_k}{\rho_k} = \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right),$$

we can rewrite the second-order weak expansion (36) as

$$\rho_{k+1} = \rho_k + h \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \rho_k + h^2 (\mathcal{D}_k^\beta \circ \mathcal{D}_k^\beta) \left(\log \frac{\rho_k}{\rho^*} \right) \rho_k + \frac{h^2}{2} \nabla \cdot (\zeta_k \rho_k) + \mathcal{O}(h^3). \quad (38)$$

In applications, the test function u in (36) is chosen according to the functional we wish to expand. For example, $u = \rho^*$ for $\int \log \rho_{k+1} \rho^* dx$, and $u = \rho_{k+1}$ for $\int \rho_{k+1} \log \rho_{k+1} dx$. These choices generate the appropriate couplings in the weak formulation and allow all occurrences of ρ_{k+1} to be replaced by (38).

4.2 One-step decay of the KL divergence

With the one-step weak expansion established, we now analyze how the KL divergence evolves along the BRWP iteration by quantifying the decrease of $D_{\text{KL}}(\rho \|\rho^*)$ produced by (34). Let $\mathcal{I}(\rho \|\rho^*) := \int \|\nabla \log(\rho/\rho^*)\|^2 \rho dx$ be the relative Fisher information, then we have the following.

Lemma 10 (KL contraction per step) *Along the update (34) from t_k to $t_{k+1} = t_k + h$,*

$$\begin{aligned} \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) - \mathrm{D}_{\mathrm{KL}}(\rho_{k+1} \|\rho^*) &= h \beta^{-1} \mathcal{I}(\rho_k \|\rho^*) - \frac{3h^2}{2} \int \left| \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \right|^2 \rho_k dx \\ &\quad - \frac{h^2}{2} \beta^{-2} \int \left\langle \nabla \log \frac{\rho_k}{\rho^*}, \nabla^2 \log \frac{\rho_k}{\rho^*} \nabla \log \frac{\rho_k}{\rho^*} \right\rangle \rho_k dx + O(h^3 \|\rho_k\|_{C^{3,1}(U)}). \end{aligned} \quad (39)$$

All derivatives are understood in the weak sense, and the $O(h^3)$ is uniform for k by Theorem 7.

Proof Write $\mathrm{D}_{\mathrm{KL}}(\rho \|\rho^*) = \int \log(\rho/\rho^*) \rho dx$. Expand ρ_{k+1} by (38) and $\log(\rho_{k+1}/\rho^*)$ by a Taylor expansion at ρ_k ; keep terms up to $O(h^2)$, and test all expressions against smooth compactly supported functions to justify integrations by parts. The $O(h)$ contribution equals

$$h \int \nabla \log \frac{\rho_k}{\rho^*} \cdot \left(\beta^{-1} \rho_k \nabla \log \frac{\rho_k}{\rho^*} \right) dx = h \beta^{-1} \mathcal{I}(\rho_k \|\rho^*).$$

The quadratic h^2 terms combine into the negative second term in (39) after using the identity $\int \nabla u \cdot \nabla v \rho_k dx = \int \mathcal{D}_k^\beta(u) v \rho_k dx$ and symmetrization. The cubic form in the last line of (39) comes from the $O(h^2)$ part of the pushforward (the $\tilde{\phi}$ term in (38)) when paired with $\log(\rho_{k+1}/\rho^*)$. Remainders are $O(h^3)$ uniformly thanks to Lemma 15. \blacksquare

We remark that the appearance of $-3h^2/2$ term reflects the fact that using the implicit score term improves the dissipation compared to the plain explicit Euler step.

4.3 Convergence of KL divergence and the mixing time

In the expansion of Lemma 10, the leading term corresponds to the Fisher information and, via the PL inequality (6), yields the exponential decay. Thus, the remaining second and third-order contributions, which arise purely from the discretization, become decisive when comparing the convergence behavior of explicit schemes with that of our semi-implicit update.

Since the discrete update admits a second-order weak expansion with an $O(h^3)$ remainder, it is natural to invoke identities for the second time derivative of the KL divergence along the Fokker-Planck flow. Let $(\rho_t)_{t \geq 0}$ solve the Fokker-Planck equation with initial condition $\rho_0 = \rho_k$. By Lemma 25 and Lemma 26 in the appendix, we obtain

$$\begin{aligned} \frac{1}{2} \frac{d^2}{dt^2} \mathrm{D}_{\mathrm{KL}}(\rho_t \|\rho^*) \Big|_{t=0} &= \beta^{-1} \int \left\langle \nabla \log \frac{\rho_k}{\rho^*}, \nabla^2 V \nabla \log \frac{\rho_k}{\rho^*} \right\rangle \rho_k dx + \beta^{-2} \int \left\| \nabla^2 \log \frac{\rho_k}{\rho^*} \right\|_F^2 \rho_k dx \\ &= \int \left| \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \right|^2 \rho_k dx + \beta^{-2} \int \left\langle \nabla \log \frac{\rho_k}{\rho^*}, \nabla^2 \log \frac{\rho_k}{\rho^*} \nabla \log \frac{\rho_k}{\rho^*} \right\rangle \rho_k dx. \end{aligned} \quad (40)$$

Since ρ_k enjoys uniform $C^{3,1}$ regularity (Section 3), all terms above are well defined, and the identity can be evaluated directly at the discrete state ρ_k .

Combining the above second-order time derivative identity with the PL inequality and the contraction of the fourth-order information term established in Lemma 28, we obtain a Grönwall-type inequality for the decay of the KL divergence between t_k and t_{k+1} .

Lemma 11 *Let $M_0 := \beta^{-2} \int \|\nabla \log \frac{\rho_0}{\rho^*}\|^4 \rho_0 dx$ and $t_k = kh$. Then the one-step decay of the KL divergence satisfies*

$$\mathrm{D}_{\mathrm{KL}}(\rho_{k+1} \|\rho^*) \leq [1 - 2\alpha h + 3\alpha^2 h^2] \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) + \frac{h^2}{2} M_0 e^{-4\alpha h k} + O(h^3). \quad (41)$$

With this estimate in hand, we are ready to derive the main convergence guarantee for the BRWP algorithm in terms of KL divergence.

Theorem 12 *There exists $h_0 > 0$ such that for any stepsize*

$$0 < h \leq h^* := \min\left\{h_0, \frac{2}{3\alpha}\right\},$$

the KL divergence of the BRWP iteration (34) at step k satisfies

$$\begin{aligned} \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) &\leq \exp[-\alpha kh(2 - 3\alpha h)] \mathrm{D}_{\mathrm{KL}}(\rho_0 \|\rho^*) \\ &\quad + \frac{h^2}{2} M_0 (k + 1) \max\left\{(1 - 2\alpha h + 3\alpha^2 h^2)^k, e^{-4\alpha kh}\right\} + \mathcal{O}(h^2). \end{aligned} \tag{42}$$

Moreover, the algorithm achieves an error $\mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^) \leq \delta$ with*

$$k = \mathcal{O}\left(\frac{|\ln \delta|}{2\alpha\sqrt{\delta}}\right), \quad h = \sqrt{\delta},$$

whenever $\sqrt{\delta} \leq h^$ and $\delta \ll 1$.*

Proof Apply Lemma 29 to the one-step decay from Lemma 11:

$$\mathrm{D}_{\mathrm{KL}}(\rho_{k+1} \|\rho^*) \leq (1 - 2\alpha h + 3\alpha^2 h^2) \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) + \frac{h^2}{2} M_0 e^{-4\alpha kh} + \mathcal{O}(h^3).$$

With $a_k = \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*)$, $c_1 = \alpha(2 - 3\alpha h)$ (so that $1 - c_1 h = 1 - 2\alpha h + 3\alpha^2 h^2$), $c_2 = M_0/2$, and $c_3 = 4\alpha$, Lemma 29 yields (42).

For the complexity statement, set $h^2 = \delta$ and require the first term in (42) to be $\mathcal{O}(\delta)$, which gives

$$k \geq \frac{\ln(\mathrm{D}_{\mathrm{KL}}(\rho_0 \|\rho^*)) - \ln \delta}{\alpha h(2 - 3\alpha h)}.$$

Substituting $h = \sqrt{\delta}$ yields the claimed scaling. ■

From (Han et al., 2025; Tan et al., 2024), the BRWP scheme exhibits a bias of order $\mathcal{O}(h^2)$ for Gaussian targets, which is consistent with our analysis.

We remark that the existing $\mathcal{O}(\log d)$ lower bounds for sampling complexity (Chewi et al., 2025) are derived for algorithms that interact with the target distribution through noisy or finite-sample gradient oracles. In contrast, our iteration-complexity bound is obtained under a weak score oracle assumption and analyzes the deterministic BRWP update map itself, rather than a stochastic oracle model. As a result, these information-theoretic lower bounds do not apply directly to the deterministic model considered here, and no explicit $\log d$ factor appears in the convergence rate.

In practice, one often chooses h small to reduce the bias, and it is natural to ask for explicit guidance on the largest permissible or optimal stepsize. The next corollary addresses this question. It follows directly from the explicit decay estimate in Theorem 12 and is especially useful when the strong-convexity constant α is small or when adaptive stepsizes are employed—for instance, using a larger stepsize during the initial iterations.

Corollary 13 *Let $h_0 > 0$ be the maximal stepsize for which the weak $\mathcal{O}(h^2)$ expansion and regularity estimates established in Section 3 remain valid. The BRWP iteration enjoys KL contraction whenever $0 < h < \frac{2}{3\alpha}$ and $h \leq h_0$. Thus, the maximal admissible stepsize is $\min\{h_0, 2/(3\alpha)\}$.*

Some other quantities are also often useful for measuring the efficiency of a sampling algorithm, including the Wasserstein-2 distance between ρ_k and ρ^* and the mixing time. In particular, the mixing time is defined as

$$t_{mix}(\delta, \rho_0) = \min\{k \mid d_{TV}(\rho_k, \rho^*) \leq \delta\}, \quad (43)$$

where ρ_k is the density function at time t_k starting from ρ_0 and

$$d_{TV}(\rho_k, \rho^*) = \frac{1}{2} \int_{\mathbb{R}^d} |\rho_k(x) - \rho^*(x)| dx. \quad (44)$$

Recalling Pinsker's inequality

$$d_{TV}(\rho_k, \rho^*)^2 \leq \frac{1}{2} \text{D}_{\text{KL}}(\rho_k \parallel \rho^*),$$

and the Talagrand's inequality

$$\frac{\alpha}{2} W_2(\rho_k, \rho^*)^2 \leq \text{D}_{\text{KL}}(\rho_k \parallel \rho^*),$$

we obtain the following direct corollary of Theorem 12.

Corollary 14 *Let $\{\rho_k\}$ be the BRWP iterates with stepsize $h \in (0, h_0]$. Then, using the discrete KL contraction of Theorem 12, the Wasserstein-2 error satisfies*

$$W_2(\rho_k, \rho^*)^2 \leq \frac{2e^{-\alpha kh(2-3\alpha h)}}{\alpha} \text{D}_{\text{KL}}(\rho_0 \parallel \rho^*) + \frac{2h^2 M_0 e^{-4\alpha kh}}{\alpha(e^{-4\alpha h} - (1 - 2\alpha h + 3\alpha^2 h^2))} + \mathcal{O}(h^2).$$

In particular, to reach accuracy $W_2(\rho_k, \rho^*)^2 \leq \delta$, choose any stepsize $h = \delta$ such that

$$\delta \leq h^* := \min\left\{h_0, \frac{2}{3\alpha}\right\}.$$

Then the mixing time satisfies

$$t_{\text{mix}}(\delta, \rho_0) = kh = \mathcal{O}\left(\frac{1}{\alpha} \left| \ln \frac{\text{D}_{\text{KL}}(\rho_0 \parallel \rho^*)}{\delta} \right| \right).$$

5. Estimation of the Score Function and Practical Considerations

The numerical accuracy of the score approximation used in the BRWP update is a critical factor in the performance of the algorithm. Any error in the evaluation of the score enters the iteration directly and may propagate across steps as an additional perturbation. In this section, we discuss practical score approximation schemes through the lens of Assumption 4 and outline several numerical strategies that can meet this weak accuracy requirement in practice.

A natural approach, adopted in (Tan et al., 2024; Han et al., 2025), is to estimate the score using kernel density estimation based on the particle ensemble $\{x_{k,j}\}_{j=1}^N$. Classical results on score estimation (e.g. Wibisono et al., 2024; Jiang and Zhang, 2009) can ensure high accuracy in low to moderate dimensions or when sufficiently many particles are available, though the estimation error typically deteriorates rapidly as the dimension increases. When the resulting score approximation satisfies an $O(h^2)$ weak accuracy bound in the sense of Assumption 4, the convergence guarantees of Section 4 apply without modification. By contrast, if the numerical score is less accurate, the induced weak error may

dominate the intrinsic discretization error of the BRWP scheme and thereby degrade the predicted convergence behavior.

As an alternative to kernel density estimation, one may propagate an auxiliary density using the kernel formula (17) and evaluate the score from this density. This avoids constructing an empirical density at each iteration and instead produces a sequence of deterministic densities

$$\widehat{\rho}_{k+1} = \mathcal{K}_V^h(\widehat{\rho}_k), \quad \widehat{\rho}_0 = \rho_0, \quad (45)$$

which can be evaluated without particle sampling, for example, via tensor-based approximations. Because the Gaussian factors in the kernel formula factorize across dimensions, the associated integrals admit efficient structured quadrature schemes. As a result, computation shifts to the evaluation of high-dimensional, tensorized integrals, whose complexity depends on the structural properties of the target distribution and the achievable tensor ranks. This strategy avoids variance due to density estimation and is often numerically more stable than KDE; however, when iterated with a fixed stepsize h , it is generally only weakly first-order accurate over long time horizons, as established in Corollary 22.

To increase the order of accuracy, we can improve the oracle using a Richardson-type combination of the kernel evaluations at two internal step sizes. Given the kernel formula \mathcal{K}_V^h , Corollary 22 shows that $\mathcal{K}_V^h \rho_k = \rho_{k+1} + O(h)$ in the weak sense. We therefore define the corrected density

$$\widehat{\rho}_{k+1}^{\text{rich}} = 2 \mathcal{K}_V^{h/2}(\mathcal{K}_V^{h/2}(\widehat{\rho}_k)) - \mathcal{K}_V^h(\widehat{\rho}_k),$$

which cancels the leading weak truncation error and yields a second-order weak approximation.

Algorithm 2 BRWP with Richardson-corrected regularized Wasserstein proximal operator

- 1: Initialize $\widehat{\rho}_0 = \rho_0$.
- 2: **for** $k = 0, 1, 2, \dots$ **do**
- 3: Compute $\mathcal{K}_V^{h/2}(\widehat{\rho}_k)$, $\mathcal{K}_V^{h/2}(\mathcal{K}_V^{h/2}(\widehat{\rho}_k))$, and $\mathcal{K}_V^h(\widehat{\rho}_k)$ by tensor computation.
- 4: Form the Richardson-corrected next-step density

$$\widehat{\rho}_{k+1} = 2 \mathcal{K}_V^{h/2}(\mathcal{K}_V^{h/2}(\widehat{\rho}_k)) - \mathcal{K}_V^h(\widehat{\rho}_k),$$

and evaluate $\nabla \log \widehat{\rho}_{k+1}$.

- 5: **for** each particle $j = 1, \dots, N$ **do**
- 6:

$$x_{k+1,j} = x_{k,j} - h \left(\nabla V(x_{k,j}) + \beta^{-1} \nabla \log \widehat{\rho}_{k+1}(x_{k,j}) \right).$$

- 7: **end for**
 - 8: **end for**
-

The Richardson extrapolation step is not positivity preserving in general. In practice, this can be mitigated by choosing the stepsize h sufficiently small, restricting the evaluation to a bounded domain, or applying a mild positivity-preserving regularization before computing $\nabla \log \widehat{\rho}_{k+1}$. A complete quantitative analysis of this construction needs an $O(h^3)$ weak expansion of \mathcal{K}_V^h , which would require refining the analysis in Section 3. Since our theoretical analysis focuses on the BRWP iterations under weak accuracy assumptions, we do not pursue these numerical safeguards or higher-order error estimates here.

We summarize several settings in which the weak $O(h^2)$ accuracy required in Assumption 4 can be achieved:

1. (*Many particles*). A KDE-based approximation of the score computed from the particle ensemble $\{x_{k,j}\}_{j=1}^N$ requires $N \gtrsim h^{-(d+4)}$ particles to attain weak $O(h^2)$ accuracy (Wibisono et al., 2024). Although this scaling is unfavorable in high dimensions, the BRWP update is fully parallelizable, since the score can be evaluated independently at each particle.
2. (*Richardson-type correction*). Instead of approximating the score from particles, one may propagate an auxiliary density $\hat{\rho}_k$ using the kernel formula and form a Richardson-extrapolated update, as in Algorithm 2. The resulting density $\hat{\rho}_{k+1}$ provides a weakly second-order accurate approximation of the true next-step density ρ_{k+1} , and the score $\nabla \log \hat{\rho}_{k+1}$ is used to advance the particles.
3. (*Laplace approximation*). Since the kernel formula involves a heat kernel with a small parameter h , the integral is dominated, as $h \rightarrow 0$, by a neighborhood of the minimizer of the exponent. This observation can be made rigorous via a Laplace approximation. For example,

$$\int_{\mathbb{R}^d} \exp\left[-\frac{\beta}{2}\left(V(z) + \frac{\|z - y\|_2^2}{2h}\right)\right] dz \approx C_d \exp\left[-\frac{\beta}{2}\left(V(y^*) + \frac{\|y^* - y\|_2^2}{2h}\right)\right],$$

where

$$y^* = \arg \min_{z \in \mathbb{R}^d} \left(V(z) + \frac{\|z - y\|_2^2}{2h} \right) = \text{prox}_V^h(y).$$

The approximation error admits an explicit expansion, and higher-order corrections are available within the same framework. This Laplace approximation reduces the original high-dimensional integral to a Gaussian integral around y^* , which can be evaluated efficiently in closed form.

4. (*Low-dimensional or parametric structure*). When the evolving density admits a low-dimensional or parametric representation (e.g., tensor trains, Gaussian mixtures, or normalizing flows), the score or the kernel operator \mathcal{K}_V^h can often be evaluated directly from the representation, without relying on kernel density estimation. In such cases, the approximation error is governed primarily by the accuracy of the representation rather than by the ambient dimension, and weak $O(h^2)$ accuracy may be achievable at a computational cost polynomial in the effective dimension. This makes Assumption 4 feasible in practice.

6. Numerical Experiments

In this section, we present several numerical experiments to illustrate the theoretical results derived in the previous sections. All below numerical experiments are conducted in a 10-dimensional sample space, i.e., \mathbb{R}^d with $d = 10$ and we only plot the result in the first dimension for the purpose of presentation.

Example 1. In this example, we explore the evolution of density functions over several iterations using the kernel formula (17). Consistent with Theorem 4 and Corollary 22, we numerically demonstrate that the computed density converges to the target density under an appropriately chosen step size. To efficiently manage the high-dimensional integrations involved, we employ tensor train approximation for the density functions. Further details on this approach can be found in our previous work (Han et al., 2025).

The first distribution we consider is a mixed Gaussian distribution defined as

$$\rho^*(x) = \frac{1}{2(2\pi\sigma^2)^{d/2}} \left[\exp\left(-\frac{\|x - a\|_2^2}{2\sigma^2}\right) + \exp\left(-\frac{\|x + a\|_2^2}{2\sigma^2}\right) \right], \quad (46)$$

where $a = (2, \dots, 2)$.

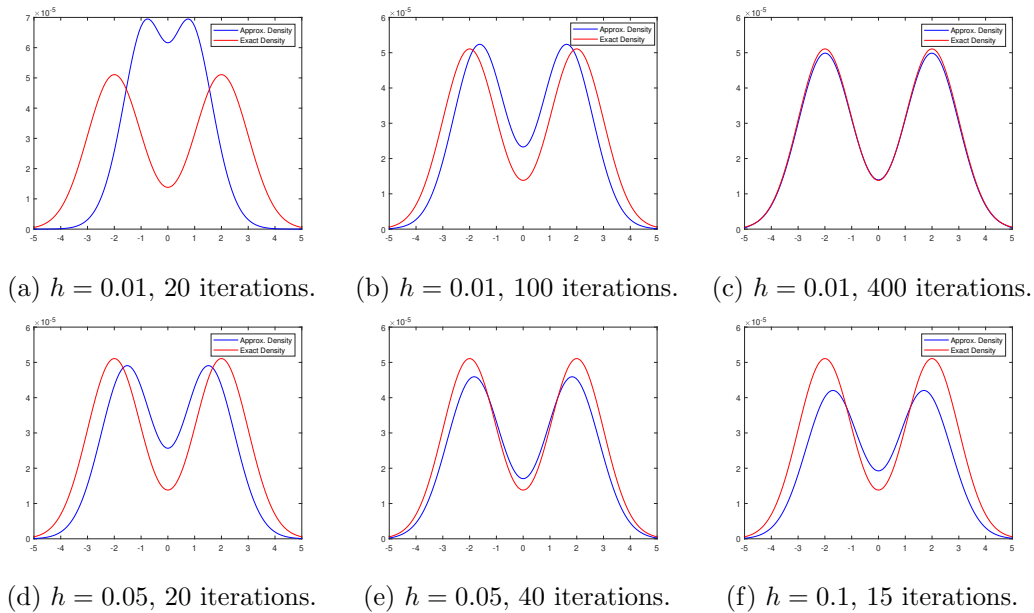


Figure 1: Evolution of the density function (blue) with (17) for different stepsizes h for the first dimension. The initial density is $\mathcal{N}(0, 2)$. The target density is a mixed Gaussian (red).

From Fig. 1, we observe that for sufficiently small stepsizes, the generated density converges to the target density very satisfactorily. Furthermore, comparing the first and last plots, we note that a larger time stepsize results in faster convergence while having a larger approximation error.

Next, we consider a mixture of L_1 and $L_{1/2}$ norms, where

$$\rho^*(x) = \frac{1}{Z} \left[\exp(-\|x + 2\vec{e}_1\|_1) + \frac{1}{2} \exp(-\|x - 2\vec{e}_1\|_{1/2}^{1/2}) \right],$$

\vec{e}_1 is the vector with the first entry equal to 1 and all other entries equal to 0, and Z is the normalization constant.

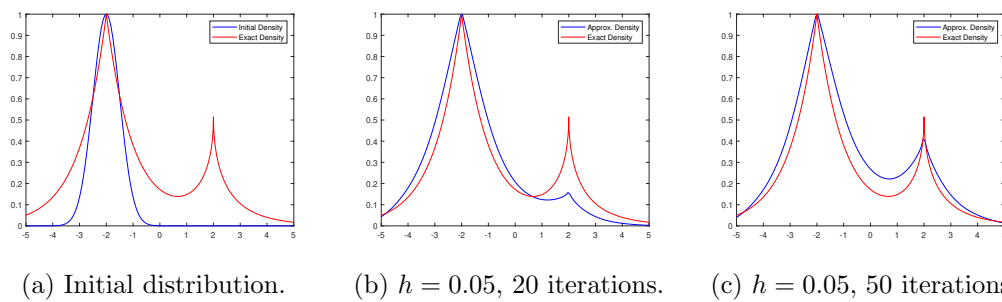


Figure 2: Evolution of the density function (blue) for the first dimension with (17) as the target density with a mixture of L_1 and $L_{1/2}$ norms (red).

From Fig. 2, we observe that even for the non-smooth potential function, which exceeds the assumptions made in Section 3, the density still converges to the target distribution satisfactorily using the kernel formula.

Example 2: In the second example, we assess the convergence of the BRWP algorithm by computing the score function based on the density function evolved using the regularized Wasserstein proximal operator, as outlined in Algorithm 2. We compare the performance of the BRWP method with that of the explicit Euler discretization of the probability flow ODE, where the score function at time t_k is approximated by a Gaussian kernel KDE from particles, the proximal Langevin algorithm in (Liang and Chen, 2022), and the ULA described in (8).

The figure below shows the distribution of particles after 50 iterations for the mixed Gaussian distribution defined in (46).

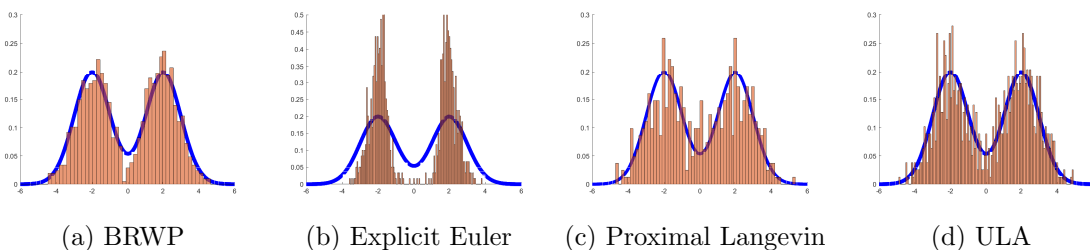


Figure 3: Histogram of 500 particles after 50 iterations in the first dimension for a Gaussian mixture distribution with $h = 0.02$.

Comparing the first and second graphs in Fig. 3, it is evident that the semi-implicit discretization (BRWP) improves the robustness of sampling and mitigates the variance collapse artifacts commonly observed in the explicit Euler discretization. Additionally, comparing the BRWP and proximal Langevin and ULA, we note that the BRWP algorithm provides a more accurate and structured approximation to the target distribution due to its noise-free nature.

In the second experiment, we consider a mixture of Gaussian and Laplace distributions defined as

$$\rho^*(x) = \frac{1}{Z} \left[\exp\left(-\frac{\|x - 2\|_2^2}{2\sigma^2}\right) + \exp\left(-\frac{\|x + 2\|_1}{2b}\right) \right],$$

where Z is the normalization constant.

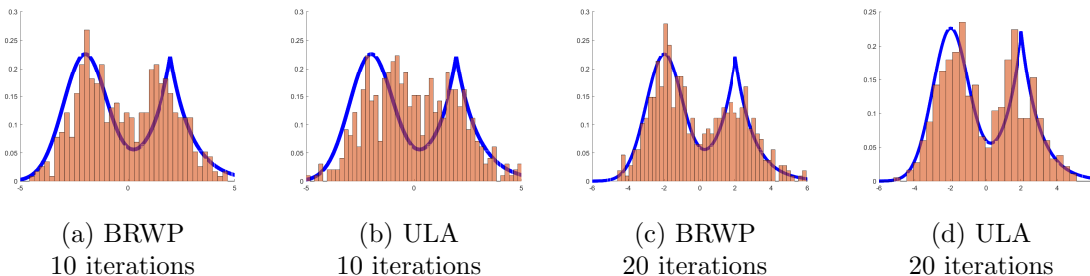


Figure 4: Histogram of 500 particles in the first dimension for the mixture of Gaussian and Laplace distributions with $h = 0.02$.

In Fig. 4, the TT-BRWP algorithm exhibits faster convergence to the target distribution compared to ULA, which is consistent with our previous theoretical results.

7. Conclusion and discussion

In this work, we present the convergence analysis of the BRWP algorithm, which is designed to sample from a known distribution up to a normalization constant. The algorithm relies on a semi-implicit time discretization of a deterministic probability flow ODE whose associated Liouville equation coincides with the Fokker–Planck equation of overdamped Langevin dynamics. To address the challenge of approximating the evolution of the density function, which usually involves intensive optimization, we apply the regularized Wasserstein proximal operator, whose solution is represented by a closed-form kernel formula. Using the weak formulation and Taylor expansion, we first demonstrate that the kernel formula serves as an approximation to the evolution of the Fokker–Planck equation. Given the noise-free nature of the new sampling algorithm, we then conduct a weak second-order numerical analysis to study the KL divergence convergence guarantee of the BRWP algorithm. The analysis shows that the semi-implicit structure of BRWP leads to enhanced dissipation in KL divergence and reduced discretization bias compared with fully explicit schemes, particularly when the kernel formula can be evaluated accurately.

Our BRWP sampling method can be viewed as an interacting particle system based on a kernel derived from the regularized Wasserstein proximal operator. This concept shares similar motivations with sampling algorithms from interacting particle systems proposed in (Garbuno-Inigo et al., 2020; Carrillo et al., 2022; Reich and Weissmann, 2021; Carrillo et al., 2019; Liu and Wang, 2016). However, our use of these kernels is fundamentally different: in BRWP the kernel serves as a linear approximation of the Fokker–Planck semigroup, whereas classical interacting particle samplers employ nonlinear kernel interactions to define new particle dynamics. It is essential to highlight that a major practical challenge lies in accurately approximating the evolving density and efficiently evaluating the kernel formula, especially in high dimensions.

There are several interesting future directions to explore from this work. First, it would be valuable to extend the discussion to more general non-log-concave distributions and investigate the convergence of the algorithm. Second, the current analysis and numerical framework can be extended and generalized to different schemes, addressing challenges such as constrained sampling problems, sampling from group symmetric distributions, and sampling under different metrics. Lastly, we are also applying this approach to broader fields, including global optimization, time-reversible diffusion, and solving high-dimensional Hamilton–Jacobi equations.

Appendix A. Postponed Proofs and Lemmas for Section 3

A.1 Postponed Proofs

In this Appendix, we list proofs of the main Lemmas and theorems for Section 3.

Proof [Proof of Lemma 3] *Step 1. Notation and setup.* For $x \in \mathbb{R}^d$, write $\mathcal{G}_h[f(x, y)] = \int_{\mathbb{R}^d} G_h(y) f(x, y) dy$, then

$$P_V^h \varphi(x) = \varphi(x) + \frac{N(x)}{Z(x)}, \quad N(x) := \mathcal{G}_h[\psi(x+y) (\varphi(x+y) - \varphi(x))], \quad Z(x) = \mathcal{G}_h[\psi(x+y)], \quad (47)$$

where the Gaussian convolution is in the y -variable.

For all $x, y \in \mathbb{R}^d$,

$$\varphi(x+y) - \varphi(x) = \int_0^1 \nabla \varphi(x + \theta y) \cdot y \, d\theta,$$

hence

$$N(x) = \int_0^1 \mathcal{G}_h[\psi(x+y) (\nabla \varphi(x + \theta y) \cdot y)] \, d\theta.$$

Let

$$r_h := \min \left\{ \frac{r_0}{2}, \sqrt{\frac{12}{\beta} h \log \frac{1}{h}} \right\}.$$

A standard change of variables gives

$$\int_{\{\|y\| > r_h\}} h^{-d/2} e^{-c_0 \|y\|^2/h} \, dy \leq C_d \left(\log \frac{1}{h} \right)^{\frac{d}{2}-1} h^3 \leq C_{d, h_0} h^2. \quad (48)$$

Since $h(\log \frac{1}{h})^{d/2-1}$ is bounded for $h \in (0, h_0]$,

$$\mathcal{G}_h[\mathbf{1}_{\{\|y\| > r_h\}}] \leq C h^2.$$

Thus, tail contributions are $\mathcal{O}(h^2)$ and can be absorbed in the remainder term. For $x \in U$ we also have

$$x + B(0, r_h) \subset U_r,$$

so Gaussian integrals can be truncated to $B(0, r_h)$ with an $\mathcal{O}(h^2)$ error.

Step 2. Expansion and integral identities. For $x \in U$ and $y \in B(0, r_h)$,

$$\psi(x+y) = \psi(x) + \nabla \psi(x) \cdot y + \frac{1}{2} y^\top \nabla^2 \psi(x) y + \frac{1}{6} \nabla^3 \psi(x)[y, y, y] + R_\psi(x, y), \quad (49)$$

with

$$|R_\psi(x, y)| \leq C \|\psi\|_{C^{3,1}(U_r)} \|y\|^4.$$

Furthermore,

$$Z(x) = \mathcal{G}_h[\psi(x+y)] = \psi(x) + h\beta^{-1} \Delta \psi(x) + \mathcal{O}(h^2 \|\psi\|_{C^{3,1}(U_r)}). \quad (50)$$

Since V is bounded on U , the quantity $Z(x)$ is bounded below uniformly in h .

Step 3. Splitting by order in ψ . Recall the Gaussian identities for $y \sim \mathcal{N}(0, 2h\beta^{-1}I_d)$:

$$\mathcal{G}_h[y] = 0, \quad \mathcal{G}_h[yy^\top] = 2h\beta^{-1}I_d, \quad \mathcal{G}_h[y_i y_j y_k] = 0.$$

For smooth f ,

$$\mathcal{G}_h[y \cdot f(y)] = 2h\beta^{-1} \mathcal{G}_h[\nabla_y \cdot f(y)]. \quad (51)$$

Substituting (49) into $N(x)$ gives

$$N(x) = I_0(x) + I_1(x) + I_2(x) + I_3(x) + I_{\text{rem}}(x),$$

where

$$\begin{aligned} I_0(x) &:= \psi(x) \int_0^1 \mathcal{G}_h[\nabla \varphi(x + \theta y) \cdot y] \, d\theta, \\ I_1(x) &:= \int_0^1 \mathcal{G}_h[(\nabla \psi(x) \cdot y)(\nabla \varphi(x + \theta y) \cdot y)] \, d\theta, \\ I_2(x) &:= \frac{1}{2} \int_0^1 \mathcal{G}_h[(y^\top \nabla^2 \psi(x) y)(\nabla \varphi(x + \theta y) \cdot y)] \, d\theta, \\ I_3(x) &:= \frac{1}{6} \int_0^1 \mathcal{G}_h[\nabla^3 \psi(x)[y, y, y](\nabla \varphi(x + \theta y) \cdot y)] \, d\theta, \\ I_{\text{rem}}(x) &:= \int_0^1 \mathcal{G}_h[R_\psi(x, y)(\nabla \varphi(x + \theta y) \cdot y)] \, d\theta. \end{aligned}$$

Since $\mathcal{G}_h[\|y\|^4] = \mathcal{O}(h^2)$ and $\mathcal{G}_h[\|y\|^5] = \mathcal{O}(h^{5/2})$,

$$|I_3(x)| + |I_{\text{rem}}(x)| \leq C h^2 \|\psi\|_{C^{3,1}(U_r)} \|\nabla\varphi\|_{L^\infty(U)}.$$

Set $p(x) := \frac{u(x)}{Z(x)}$, then

$$\langle I_2, p \rangle = \frac{1}{2} \int_0^1 \int_U \int_{\mathbb{R}^d} G_h(y) (y^\top \nabla^2 \psi(x) y) (\nabla\varphi(x + \theta y) \cdot y) dy p(x) dx d\theta. \quad (52)$$

Integration by parts in x gives

$$\langle I_2, p \rangle = -\frac{1}{2} \int_0^1 \int_U \int_{\mathbb{R}^d} G_h(y) (y \otimes y \otimes y) : (\varphi(x + \theta y) \nabla(\nabla^2 \psi(x) p(x))) dy dx d\theta.$$

Using the evenness of G_h and $\int G_h \|y\|^4 dy = \mathcal{O}(h^2)$,

$$|\langle I_2, p \rangle| \leq C h^2 \|\psi\|_{C^{3,1}(U_r)} \|\nabla\varphi\|_{L^\infty(U)} \|u\|_{C^{0,1}(U)}.$$

Step 4. Main terms I_0 and I_1 . Define

$$v(x) := \frac{\psi(x)u(x)}{Z(x)}, \quad w(x) := \frac{\nabla\psi(x)u(x)}{Z(x)}.$$

I_0 . Using (51) with $F_\theta(y) = \nabla\varphi(x + \theta y)$,

$$\mathcal{G}_h[\nabla\varphi(x + \theta y) \cdot y] = 2h\beta^{-1}\theta \mathcal{G}_h[\Delta\varphi(x + \theta y)].$$

Hence

$$\begin{aligned} \langle I_0, p \rangle &= 2h\beta^{-1} \int_0^1 \theta \int_U \mathcal{G}_h[\Delta\varphi(x + \theta y)] v(x) dx d\theta \\ &= -2h\beta^{-1} \int_0^1 \theta \mathcal{G}_h \left[\int_U \nabla\varphi(x + \theta y) \cdot \nabla v(x) dx \right] d\theta. \end{aligned} \quad (53)$$

I_1 . Similarly, we have

$$\begin{aligned} \langle I_1, p \rangle &= \int_0^1 \int_U \mathcal{G}_h[(y \cdot \nabla\varphi(x + \theta y))y] \cdot w(x) dx d\theta \\ &= 2h\beta^{-1} \int_0^1 \int_U \mathcal{G}_h[(\nabla_y \cdot (\nabla\varphi(x + \theta y))y)] \cdot w(x) dx d\theta. \end{aligned} \quad (54)$$

Summing up and noting that the odd power in h will vanish, we have

$$\begin{aligned} \langle I_0 + I_1, p \rangle &= -2h\beta^{-1} \int_0^1 \left(\theta \langle \mathcal{G}_h[\nabla\varphi(x + \theta y)], \nabla v \rangle - \langle \mathcal{G}_h[\nabla \cdot (\nabla\varphi(x + \theta y))y], w \rangle \right) d\theta \\ &= -h\beta^{-1} (\langle \nabla\varphi, \nabla v \rangle - 2\langle \nabla\varphi, w \rangle) + \mathcal{O}(h^2 \|\varphi\|_{C^{2,1}(U)} \|u\|_{C^{0,1}(U)}), \end{aligned}$$

using $\mathcal{G}_h[\nabla\varphi(x + \theta y)] = \nabla\varphi(x) + \mathcal{O}(h)$ and $\int_0^1 \theta d\theta = 1/2$.

Substituting $v = \psi u/Z$ and $w = \nabla\psi u/Z$, and using $\nabla \log \psi = -\frac{\beta}{2} \nabla V$ together with (50), gives

$$\langle I_0 + I_1, p \rangle = h\beta^{-1} \langle \Delta\varphi - \beta \nabla V \cdot \nabla\varphi, u \rangle + \mathcal{O}(h^2 \|\varphi\|_{C^{2,1}(U)} \|u\|_{C^{0,1}(U)}).$$

Therefore

$$|\langle P_V^h \varphi - \varphi - h\mathcal{L}\varphi, u \rangle| \leq C h^2 \|\varphi\|_{C^{2,1}(U)} \|u\|_{C^{0,1}(U)},$$

which completes the proof of (24). \blacksquare

Proof [Proof of Theorem 4] Fix a bounded open set $U \subset \mathbb{R}^d$. Choose $\chi \in C_c^\infty(\mathbb{R}^d)$ with

$$\chi \equiv 1 \text{ on } U, \quad U \Subset \{\chi = 1\} \Subset \text{supp } \chi \Subset U_r,$$

where U_r is the fixed bounded domain used in the local weak expansion. Define the cutoff decompositions

$$\varphi_{\text{in}} := \chi\varphi, \quad \varphi_{\text{out}} := (1 - \chi)\varphi, \quad u_{\text{in}} := \chi u, \quad u_{\text{out}} := (1 - \chi)u.$$

Then

$$\langle P_V^h \varphi - \varphi - h\mathcal{L}\varphi, u \rangle = \langle P_V^h \varphi_{\text{in}} - \varphi_{\text{in}} - h\mathcal{L}\varphi_{\text{in}}, u_{\text{in}} \rangle + E_{\text{out}},$$

where E_{out} consists of all contributions in which at least one factor is supported in U^c .

Step 1. Local part and Gaussian tail. The local weak expansion on U_r yields

$$|\langle P_V^h \varphi_{\text{in}} - \varphi_{\text{in}} - h\mathcal{L}\varphi_{\text{in}}, u_{\text{in}} \rangle| \leq C h^2 \|\varphi\|_{C^{2,1}(U_r)} \|u\|_{C^{0,1}(U_r)}. \quad (55)$$

By assumption 2, the weight ratio satisfies

$$|\psi(x)\psi(y)^{-1}| \leq C(1 + \|x\|^{q_V} + \|y\|^{q_V}).$$

Using the kernel representation with $c_0 \leq Z(x) \leq C_0$

$$\begin{aligned} |P_V^h \varphi(x)| &= \left| Z(x)^{-1} \psi(x) \int_{\mathbb{R}^d} G_h(x-y) \psi(y)^{-1} \varphi(y) dy \right| \\ &\leq C \int_{\mathbb{R}^d} G_h(x-y) (1 + \|x\|^{q_V} + \|y\|^{q_V}) |\varphi(y)| dy. \end{aligned} \quad (56)$$

Fix $p > 2 + q_V$, define $c_p = p - q_V > 2$, and choose a slightly larger radius than the one in the Proof of Lemma 3

$$r_h := \sqrt{\frac{8 + 2d + 2c_p}{\beta} h \log \frac{1}{h}}, \quad 0 < h \leq h_0 \leq e^{-1}.$$

A direct computation gives the Gaussian tail estimate

$$\sup_{\|z\| \geq r_h} G_h(z) (1 + \|z\|)^{-c_p} \leq C h^2. \quad (57)$$

Step 2. Cross terms. We will show that all near-field contributions with either φ_{in} or u_{in} are supported in a fixed bounded set. Since $\text{dist}(\text{supp}\{\chi\}, U_r^c) \geq r_1 > 0$ and $r_h \leq r_1/2$, if $y \in \text{supp}\{\chi\}$ and $\|x - y\| \leq r_h$, then $x \in U_r$ as well. Thus, any such near-field term is a pairing of the form

$$\langle P_V^h f - f - h\mathcal{L}f, g \rangle, \quad \text{supp}\{f\} \cap \text{supp}\{g\} \subset U_r,$$

with $f \in C^{2,1}(U_r)$ and $g \in C^{0,1}(U_r)$. By the local weak expansion,

$$|\text{near region with one inner factor}| \leq C h^2 \|\varphi\|_{C^{2,1}(U_r)} \|u\|_{C^{0,1}(U_r)}.$$

Step 3. Far region $\{\|x - y\| > r_h\}$. Introduce the weighted functions

$$f(x) := (1 + \|x\|)^p |u(x)|, \quad g(y) := (1 + \|y\|)^p |\varphi(y)|,$$

so that, by assumption, $\int f \leq M_p$ and $\int g \leq M_p$. Using (56) and Fubini,

$$\begin{aligned} J_{\text{far}} &:= \iint_{\{\|x-y\|>r_h\}} G_h(x-y)(1 + \|x\|^{q_V} + \|y\|^{q_V})|\varphi(y)| |u(x)| dx dy \\ &\leq M_p^2 \sup_{\|x-y\|>r_h} G_h(x-y)W(x,y), \end{aligned}$$

where

$$W(x,y) := (1 + \|x\|^{q_V} + \|y\|^{q_V})(1 + \|x\|)^{-p}(1 + \|y\|)^{-p}.$$

Weight inequality. Let $a := \|x\|$, $b := \|y\|$, $r := \|x - y\|$, and $R := \max\{a, b\}$. A simple two-case estimate (depending on whether $\min\{a, b\} \geq R/2$ or not) yields

$$W(x,y) \leq C(1+r)^{-(p-q_V)} = C(1+r)^{-c_p}.$$

Combining the above with (57),

$$J_{\text{far}} \leq C M_p^2 \sup_{\|z\|>r_h} G_h(z)(1 + \|z\|)^{-c_p} \leq C M_p^2 h^2.$$

All purely outer terms are treated identically, since the kernel bound (56) and the weighted integrability of u and φ apply without using any local regularity.

Hence

$$|E_{\text{out}}| \leq C h^2 (\|\varphi\|_{C^{2,1}(U_r)} \|u\|_{C^{0,1}(U_r)} + M_p^2 + 1).$$

Combining (55) with the bound for E_{out} proves (26) where we replace U_r by U as the set is arbitrary. \blacksquare

Proof [Proof of Lemma 6] *Step 0. Heat kernel expansions.* Since $V \in C^{4,1}(U_r)$, the weight $\psi = e^{-\beta V/2}$ satisfies

$$\|\psi^{\pm 1}\|_{C^{3,1}(U_r)} \leq C.$$

Recall $Z = \mathcal{G}_h(\psi)$, as in (50), the classical heat kernel expansion gives

$$\|Z - \psi - h\beta^{-1}\Delta\psi\|_{L^\infty(U_r)} \leq C h^2.$$

Since $\psi > 0$, we also have $0 < c_0 \leq Z \leq C_0$.

Define the multiplier

$$w_h(x) := \psi(x) Z^{-1}(x).$$

Then

$$w_h(x) = 1 - h\beta^{-1} \frac{\Delta\psi(x)}{\psi(x)} + O(h^2), \quad \|w_h - 1\|_{(C_0^{0,1}(U_r))^*} \leq C h. \quad (58)$$

Step 1. Reduction to a perturbed Dirichlet problem.

Write $u = \psi v$. Using $\mathcal{K}_V^h f = \psi \mathcal{G}_h(Z^{-1}f)$, (30) becomes

$$v - \mathcal{G}_h v = \mathcal{G}_h((w_h - 1)v) + \psi^{-1}\eta\mu. \quad (59)$$

Let G_h^D be the Dirichlet heat kernel on U_r and let \mathcal{G}_h^D be the convolution operator with kernel G_h^D . Write the boundary remainder as $R_h := \mathcal{G}_h - \mathcal{G}_h^D$. By Lemma 16,

$$\|R_h\|_{L^1(U_r) \rightarrow C^{3,1}(U)} \leq C e^{-c_0^2/h}, \quad r_0 = \text{dist}(U, \partial U_r). \quad (60)$$

Rewrite (59) as

$$v - \mathcal{G}_h^D v = \mathcal{G}_h^D ((w_h - 1)v) + R_h v + R_h ((w_h - 1)v) + \psi^{-1} \eta \mu. \quad (61)$$

Define the Dirichlet resolvent

$$\mathcal{R}_h^D := (I - \mathcal{G}_h^D)^{-1} = \sum_{n \geq 0} (\mathcal{G}_h^D)^n.$$

By Lemma 17,

$$\|\mathcal{R}_h^D \nu\|_{C^{3,1}(U)} \leq C M_p(\nu). \quad (62)$$

Step 2. Perturbative inversion. Apply \mathcal{R}_h^D to (61):

$$v = \mathcal{A}_h^{(0)} v + \mathcal{A}_h^{(1)} v + \mathcal{A}_h^{(2)} v + \mathcal{R}_h^D (\psi^{-1} \eta \mu),$$

where

$$\mathcal{A}_h^{(0)} := \mathcal{R}_h^D \circ R_h, \quad \mathcal{A}_h^{(1)} := \mathcal{R}_h^D (\mathcal{G}_h^D ((w_h - 1) \cdot)), \quad \mathcal{A}_h^{(2)} := \mathcal{R}_h^D (R_h ((w_h - 1) \cdot)).$$

Since

$$\int_{U_r} (1 + \|x\|)^p |\mathcal{G}_h^D ((w_h - 1)v)(x)| dx \leq \|(w_h - 1)v\|_\infty \int_{U_r} \int_{U_r} (1 + \|x\|)^p G_h^D(x, y) dx dy \leq Ch \|v\|_{C^{3,1}(U_r)}.$$

Applying (62),

$$\|\mathcal{A}_h^{(1)}\|_{C^{3,1}(U_r) \rightarrow C^{3,1}(U)} \leq Ch. \quad (63)$$

Next, from (60),

$$\|R_h((w_h - 1)v)\|_{C^{3,1}(U)} \leq Ch e^{-cr_0^2/h} \|v\|_{C^{3,1}(U_r)}.$$

Thus

$$\|\mathcal{A}_h^{(0)}\| \leq C e^{-cr_0^2/h}, \quad \|\mathcal{A}_h^{(2)}\| \leq Ch e^{-cr_0^2/h}, \quad (64)$$

as operators $C^{3,1}(U_r) \rightarrow C^{3,1}(U)$.

Choose h_0 so small that

$$C e^{-cr_0^2/h_0} \leq \frac{1}{6}, \quad Ch_0 \leq \frac{1}{6}, \quad Ch_0 e^{-cr_0^2/h_0} \leq \frac{1}{6}.$$

Then, for all $h \in (0, h_0]$,

$$\|\mathcal{A}_h^{(0)} + \mathcal{A}_h^{(1)} + \mathcal{A}_h^{(2)}\| \leq \frac{1}{2}.$$

Thus, the Neumann expansion converges:

$$v = \left(I - \mathcal{A}_h^{(0)} - \mathcal{A}_h^{(1)} - \mathcal{A}_h^{(2)} \right)^{-1} \mathcal{R}_h^D (\psi^{-1} \eta \mu),$$

so that

$$\|v\|_{C^{3,1}(U)} \leq 2 \|\mathcal{R}_h^D (\psi^{-1} \eta \mu)\|_{C^{3,1}(U)}.$$

Using (62) and boundedness of $\psi^{-1} \eta$,

$$\|v\|_{C^{3,1}(U)} \leq C M_p(\mu). \quad (65)$$

Step 3. Return to $u = \psi v$. Since $\psi \in C^{3,1}(U_r)$,

$$\|u\|_{C^{3,1}(U)} = \|\psi v\|_{C^{3,1}(U)} \leq C \|v\|_{C^{3,1}(U)} \leq C M_p(\mu).$$

This proves (31). ■

Proof [Proof of Theorem 7] *Step 1. Strong control of ϵ_k* . We apply Lemma 6 with $\mu = \epsilon_k$ and $\eta \equiv 1$ on U . This gives a solution $u_k \in C^{3,1}(U)$ to

$$u_k - \mathcal{K}_V^h u_k = \epsilon_k \quad \text{on } U,$$

satisfying

$$\|u_k\|_{C^{3,1}(U)} \leq CM_p(\epsilon_k). \quad (66)$$

By Lemma 19, we have

$$\|\mathcal{K}_V^h f\|_{C^{3,1}(U)} \leq C \|f\|_{C^{3,1}(U)}. \quad (67)$$

Because $\epsilon_k = u_k - \mathcal{K}_V^h u_k$ on U , combining (66)–(67) gives

$$\|\epsilon_k\|_{C^{3,1}(U)} \leq \|u_k\|_{C^{3,1}(U)} + \|\mathcal{K}_V^h u_k\|_{C^{3,1}(U)} \leq C \|u_k\|_{C^{3,1}(U)} \leq CM_p(\epsilon_k). \quad (68)$$

Step 2. Uniform $C^{3,1}$ bound for the kernel chain $\hat{\rho}_k$. From Lemma 15, we have $\sup_{k \geq 0} M_p(\hat{\rho}_k) \leq C$. Applying Lemma 6 to $\mu = \eta \hat{\rho}_k$ yields $v_k \in C^{3,1}(U)$ with

$$v_k - \mathcal{K}_V^h v_k = \eta \hat{\rho}_k \quad \text{on } U,$$

and

$$\|v_k\|_{C^{3,1}(U)} \leq C M_p(\hat{\rho}_k) \leq C'. \quad (69)$$

Using again (67),

$$\|\hat{\rho}_k\|_{C^{3,1}(U)} = \|v_k - \mathcal{K}_V^h v_k\|_{C^{3,1}(U)} \leq C \|v_k\|_{C^{3,1}(U)} \leq C'. \quad (70)$$

Step 3. Closing the recursion and uniform bound. By Lemma 15, we have

$$\sup_{k \geq 0} M_p(\rho_k) + M_p(\hat{\rho}_k) \leq C,$$

hence

$$M_p(\epsilon_k) = M_p(\rho_k - \hat{\rho}_k) \leq M_p(\rho_k) + M_p(\hat{\rho}_k) \leq C$$

for all $k \geq 0$. Combining this with (68) gives

$$\sup_{k \geq 0} \|\epsilon_k\|_{C^{3,1}(U)} \leq C. \quad (71)$$

Together with (70), we obtain a uniform $C^{3,1}$ bound for ρ_k :

$$\|\rho_k\|_{C^{3,1}(U)} \leq \|\hat{\rho}_k\|_{C^{3,1}(U)} + \|\epsilon_k\|_{C^{3,1}(U)} \leq C, \quad k \geq 0. \quad (72)$$

■

Proof [Proof of Theorem 8] Fix $k \geq 0$ and assume first that $\text{supp}\{w\} \subset U$. The general case follows from a cutoff argument exactly as in Theorem 4.

From Theorem 7, Lemma 18, Lemma 19, and Lemma 20

$$0 < c \leq \rho_k(x), \mathcal{K}_V^h \rho_k(x), \mathcal{K}_V^h \tilde{\rho}_k(x) \leq C \quad \text{and} \quad \rho_k, \mathcal{K}_V^h \rho_k, \mathcal{K}_V^h \tilde{\rho}_k \in C^{3,1}(U).$$

For notational sake, set

$$r := \rho_k, \quad r_K := \mathcal{K}_V^h \rho_k, \quad r_* := r + h \mathcal{L}^* r.$$

Then we have

$$\begin{aligned} & \int w \cdot \left[\nabla \log(\mathcal{K}_V^h \tilde{\rho}_k) - \nabla \log \rho_k - h \nabla \left(\frac{\mathcal{L}^* \rho_k}{\rho_k} \right) \right] \rho_k dx \\ &= \int w \cdot \left[\nabla \log r_K - \nabla \log r - h \nabla \left(\frac{\mathcal{L}^* \rho_k}{\rho_k} \right) \right] \rho_k dx + \int w \cdot \left[\nabla \log(\mathcal{K}_V^h \tilde{\rho}_k) - \nabla \log(\mathcal{K}_V^h \rho_k) \right] \rho_k dx \end{aligned} \quad (73)$$

and we denote the first and second term as I^0 and I^1 .

Step 1. Ideal kernel part I^0 . Insert and subtract $\nabla \log r_*$:

$$\begin{aligned} I^0 &= \int w \cdot (\nabla \log r_K - \nabla \log r_*) r dx + \int w \cdot \left(\nabla \log r_* - \nabla \log r - h \nabla \left(\frac{\mathcal{L}^* r}{r} \right) \right) r dx \\ &=: I_a^0 + I_b^0. \end{aligned} \quad (74)$$

Using

$$\nabla \log r_K - \nabla \log r_* = \frac{\nabla(r_K - r_*)}{r_*} + \nabla r_K \frac{r_* - r_K}{r_K r_*},$$

we write

$$\begin{aligned} I_a^0 &= \int w \cdot \frac{\nabla(r_K - r_*)}{r_*} r dx + \int w \cdot \nabla r_K \frac{r_* - r_K}{r_K r_*} r dx \\ &= - \int \nabla \cdot \left(\frac{wr}{r_*} \right) (r_K - r_*) dx - \int \left(w \cdot \nabla r_K \frac{r}{r_K r_*} \right) (r_K - r_*) dx. \end{aligned}$$

Since $r, r_K, r_* \in C^{1,1}(U)$ and $r_*, r_K \geq c > 0$, both coefficients

$$\nabla \cdot \left(\frac{wr}{r_*} \right), \quad w \cdot \nabla r_K \frac{r}{r_K r_*}$$

lie in $C^{0,1}(U)$ with norms

$$\left\| \nabla \cdot \left(\frac{wr}{r_*} \right) \right\|_{C^{0,1}(U)} + \left\| w \cdot \nabla r_K \frac{r}{r_K r_*} \right\|_{C^{0,1}(U)} \leq C \|w\|_{C^{1,1}(U)}.$$

Combining the above with the weak approximation in Theorem 4 implies

$$|I_a^0| \leq Ch^2 \|w\|_{C^{1,1}(U)}. \quad (75)$$

For I_b^0 , Taylor expansion of $\log(r + \theta h \mathcal{L}^* r)$ in θ yields

$$\nabla \log r_* - \nabla \log r = h \nabla \left(\frac{\mathcal{L}^* r}{r} \right) + h^2 R(x), \quad \|R\|_{C^0(U)} \leq C.$$

Thus

$$|I_b^0| \leq Ch^2 \|w\|_{L^\infty(U)} \leq Ch^2 \|w\|_{C^{1,1}(U)}. \quad (76)$$

From (75)–(76),

$$|I^0| \leq Ch^2 \|w\|_{C^{1,1}(U)}. \quad (77)$$

Step 2. Oracle perturbation I^1 . Since \mathcal{K}_V^h is $C^{3,1}$ -smoothing,

$$\|\mathcal{K}_V^h \tilde{\rho}_k - \mathcal{K}_V^h \rho_k\|_{C^1(U)} \leq C \|\tilde{\rho}_k - \rho_k\|_{C^{2,1}(U)}. \quad (78)$$

The score map $\rho \mapsto \nabla \log \rho$ is C^1 on the set $\{f \in C^{2,1}(U) : c \leq f \leq C\}$, hence

$$\|\nabla \log(\mathcal{K}_V^h \tilde{\rho}_k) - \nabla \log(\mathcal{K}_V^h \rho_k)\|_{C^0(U)} \leq C' \|\mathcal{K}_V^h \tilde{\rho}_k - \mathcal{K}_V^h \rho_k\|_{C^1(U)} \leq C \|\tilde{\rho}_k - \rho_k\|_{C^{2,1}(U)}. \quad (79)$$

Thus

$$|I^1| \leq C \|w\|_{L^\infty(U)} \|\tilde{\rho}_k - \rho_k\|_{C^{2,1}(U)} \leq C \|w\|_{C^{1,1}(U)} \|\tilde{\rho}_k - \rho_k\|_{C^{2,1}(U)}. \quad (80)$$

Combining (77) and (80) gives the desired result.

If $\text{supp}\{w\} \not\subseteq U$, choose $\chi \in C_c^\infty(U_r)$ with $\chi \equiv 1$ on U and replace w by χw ; this leaves the integral unchanged and only modifies C by a universal factor. \blacksquare

A.2 Postponed Lemmas

In this Appendix, we verify the corresponding assumptions and establish uniform bounds for the quantities appearing in Theorem 4 and 7, including the uniform p -moment bounds in Section A.2.1, estimates involve a Gaussian kernel in Section A.2.2, boundedness of the kernel formula in Section A.2.3, and local positivity of discrete iterates in Section A.2.4.

A.2.1 FINITE p -MOMENT BOUND

We will first establish a uniform p -moment estimate for the discrete propagation $\rho_{k+1} = (F_k)_\# \rho_k$ which guarantees that all iterates of the discrete kernel remain integrable, and we can truncate the integration to a bounded domain which is required in the proof of Theorem 4.

Lemma 15 (Uniform discrete p -moment bound for the BRWP iteration) *Let $p \geq 2$. Then there exist constants $h_0 \in (0, 1]$, $a_p, c_p \in (0, 1)$, and $b_p, d_p < \infty$, depending only on $(\alpha, \beta, d, p, \|V\|_{C^{3,1}})$, such that for all $h \in (0, h_0]$ and all $k \geq 0$,*

$$\int_{\mathbb{R}^d} \|x\|^p \rho_{k+1}(x) dx \leq (1 - a_p h) \int_{\mathbb{R}^d} \|y\|^p \rho_k(y) dy + b_p h, \quad (81)$$

and

$$\int_{\mathbb{R}^d} \|x\|^p (\mathcal{K}_V^h \hat{\rho}_k)(x) dx \leq (1 - c_p h) \int_{\mathbb{R}^d} \|y\|^p \hat{\rho}_k(y) dy + d_p h. \quad (82)$$

Consequently,

$$\sup_{k \geq 0} \int_{\mathbb{R}^d} \|x\|^p \rho_k(x) dx \leq \max \left\{ \int_{\mathbb{R}^d} \|x\|^p \rho_0(x) dx, \frac{b_p}{a_p} \right\} < \infty, \quad (83)$$

and

$$\sup_{k \geq 0} \int_{\mathbb{R}^d} \|x\|^p \hat{\rho}_k(x) dx \leq \max \left\{ \int_{\mathbb{R}^d} \|x\|^p \rho_0(x) dx, \frac{d_p}{c_p} \right\} < \infty. \quad (84)$$

Proof Fix $p \geq 2$. Constants denoted by C may vary from line to line and depend only on $(\alpha, \beta, d, p, \|V\|_{C^{3,1}})$.

Step 1. Score decomposition and identities. We begin with the identity

$$\log(\mathcal{K}_V^h \rho_k) = \log \psi + \log(\mathcal{G}_h(Z^{-1} \rho_k)).$$

Since $\nabla \log \psi = -(\beta/2) \nabla V$, it follows that

$$\nabla V + \beta^{-1} \nabla \log(\mathcal{K}_V^h \rho_k) = \frac{1}{2} \nabla V + \beta^{-1} \nabla \log(\mathcal{G}_h(Z^{-1} \rho_k)),$$

and therefore

$$F_k(x) = x - \frac{h}{2} \nabla V(x) - h \beta^{-1} \nabla \log(\mathcal{G}_h(Z^{-1} \rho_k))(x). \quad (85)$$

For any $f \geq 0$,

$$\nabla \log(\mathcal{G}_h f)(x) = \frac{-\beta}{2h} \frac{\int_{\mathbb{R}^d} (x-y) G_h(x-y) f(y) dy}{\int_{\mathbb{R}^d} G_h(x-y) f(y) dy} = -\beta \frac{x - m_f(x)}{2h},$$

where

$$m_f(x) := \frac{\int_{\mathbb{R}^d} y G_h(x-y) f(y) dy}{\int_{\mathbb{R}^d} G_h(x-y) f(y) dy}$$

is the local Gaussian barycenter. Applying this with $f = Z^{-1} \rho_k$ gives

$$F_k(x) = x - \frac{h}{2} \nabla V(x) + \frac{1}{2} (x - m_{\rho_k/Z}(x)). \quad (86)$$

Using Cauchy–Schwarz under the normalized weight proportional to $G_h(x-y) Z^{-1}(y) \rho_k(y)$,

$$\|x - m_{\rho_k/Z}(x)\|^2 \leq \frac{\int_{\mathbb{R}^d} \|x-y\|^2 G_h(x-y) Z^{-1}(y) \rho_k(y) dy}{\int_{\mathbb{R}^d} G_h(x-y) Z^{-1}(y) \rho_k(y) dy}.$$

It implies the following averaged estimate:

$$\int_{\mathbb{R}^d} \|x - m_{\rho_k/Z}(x)\|^p \rho_k(x) dx \leq C h^{p/2}, \quad p \geq 2, \quad (87)$$

where C depends only on $(d, p, \beta, \|V\|_{C^{3,1}})$.

Step 2. Drift contraction. Combining (86) and (87), let $\Lambda_k(x) := \beta(x - m_{\rho_k/Z}(x))$. Then

$$F_k(x) = x - \frac{h}{2} \nabla V(x) + \frac{1}{2\beta} \Lambda_k(x), \quad \int \|\Lambda_k(x)\|^p \rho_k(x) dx \leq C h^{p/2}.$$

Let \bar{x} denote a minimizer of V , so $\nabla V(\bar{x}) = 0$. By the mean-value theorem,

$$\nabla V(x) - \nabla V(\bar{x}) = H_\xi(x - \bar{x}), \quad H_\xi := \int_0^1 \nabla^2 V(\bar{x} + t(x - \bar{x})) dt.$$

Thus,

$$x - \frac{h}{2} \nabla V(x) - \bar{x} = (I - \frac{h}{2} H_\xi)(x - \bar{x}).$$

Since $\alpha I_d \preceq H_\xi \preceq L I_d$ on bounded sets, choose $h_0 \leq \min\{1, 2/L\}$ so that

$$\|I - \frac{h}{2} H_\xi\| \leq 1 - \frac{\alpha h}{2}.$$

Therefore,

$$\|x - \frac{h}{2} \nabla V(x) - \bar{x}\| \leq (1 - \frac{\alpha h}{2}) \|x - \bar{x}\|. \quad (88)$$

Raising the inequality to the p -th power and using $(1 - \theta)^p \leq 1 - \frac{(p-1)\theta}{2}$ gives there exists $a_p > 0$, such that

$$\|x - \frac{h}{2} \nabla V(x) - \bar{x}\|^p \leq (1 - a_p h) \|x - \bar{x}\|^p. \quad (89)$$

Step 3. Adding the fluctuation. Write

$$F_k(x) - \bar{x} = U(x) + W(x), \quad U(x) := x - \frac{h}{2} \nabla V(x) - \bar{x}, \quad W(x) := \frac{1}{2\beta} \Lambda_k(x).$$

For $p \geq 2$ and $\varepsilon \in (0, 1]$, Young's inequality gives

$$\|U + W\|^p \leq (1 + \varepsilon) \|U\|^p + C_p \varepsilon^{1-p} \|W\|^p.$$

Choose $\varepsilon = 1$, integrate against ρ_k , and apply (89) together with (87) to obtain

$$\|F_k(x) - \bar{x}\|^p \leq (1 - a_p h) \|x - \bar{x}\|^p + C_p h. \quad (90)$$

Since $\|x\|^p \leq 2^{p-1} (\|x - \bar{x}\|^p + \|\bar{x}\|^p)$, integrating (90) against ρ_k (and noting the $\|\bar{x}\|^p$ term is constant) yields

$$\int \|F_k(x)\|^p \rho_k(x) dx \leq (1 - a_p h) \int \|x\|^p \rho_k(x) dx + b_p h, \quad b_p := C_p (1 + \|\bar{x}\|^p).$$

Because $\rho_{k+1} = (F_k)_\# \rho_k$, the left-hand side equals $\int \|x\|^p \rho_{k+1}(x) dx$, giving (81). Iterating proves (83).

Step 4. p -moment for the kernel iteration. Let $p \geq 2$ and set $r := \max\{p, q_V + 2\}$. Let $\Phi \in C^{3,1}(\mathbb{R}^d)$ be a smooth radial Lyapunov function defined as $\Phi(x) = (1 + \|x\|^2)^{r/2}$ when $\|x\| \geq 2$ and $\Phi \geq 1$ everywhere. A direct calculation shows

$$\mathcal{L}\Phi := \beta^{-1} \Delta \Phi - \nabla V \cdot \nabla \Phi \leq -c_1 \Phi + C_1 \quad (\|x\| \geq R),$$

for some suitable $R, c_1, C_1 > 0$, using $\nabla^2 V \succeq \alpha I_d$ and polynomial bounds inside the ball $\{\|x\| \leq R\}$.

By the local weak expansion (Lemma 3),

$$P_V^h \Phi(y) = \Phi(y) + h \mathcal{L}\Phi(y) + R_h(y), \quad |R_h(y)| \leq C h^2 (1 + \|y\|^{q_V}),$$

for some $q_V > 0$ in Assumption 2. Since $r \geq q_V$, we have $1 + \|y\|^{q_V} \leq C(1 + \|y\|^r) \leq C\Phi(y)$ for all y , and hence, for $h \leq h_0$ sufficiently small,

$$|R_h(y)| \leq \frac{c_1}{4} h \Phi(y) + C_3 h.$$

Thus, for $y \in \mathbb{R}^d$

$$P_V^h \Phi(y) \leq (1 - c_0 h) \Phi(y) + C_0 h$$

for suitable constants $c_0 \in (0, c_1/2]$ and $C_0 > 0$.

Since $c_1(1 + \|x\|^r) \leq \Phi(x) \leq c_2(1 + \|x\|^r)$ and $r \geq p$, we have $1 + \|x\|^p \leq C(1 + \|x\|^r) \leq C\Phi(x)$, so the same inequality holds for $\phi(x) = \|x\|^p$ after adjusting constants. Finally, by duality,

$$\int \|x\|^p (\mathcal{K}_V^h \rho)(x) dx = \langle P_V^h \phi, \rho \rangle \leq (1 - c_0 h) \int \|y\|^p \rho(y) dy + C_0 h,$$

which proves (82). ■

A.2.2 AUXILIARY LEMMAS IN THE PROOF OF LEMMA 6

In this section, we use the interior regularity of the integral equation associated with \mathcal{K}_V^h to lift the moment or $(C_0^{0,1})^*$ estimate to the desired $C^{3,1}$ regularity. Firstly, we will show two auxiliary lemmas. The following heat kernel estimate is classical and can be found in, for example, (Davies, 1989), while we still give an elementary proof for completeness.

Lemma 16 (Derivative and boundary-decay estimates for the Dirichlet heat kernel) *Let G_h^D be the Dirichlet heat kernel on U_r . There exist constants $C, c > 0$, depending only on (d, U, U_r) , such that for all $h \in (0, 1]$ and all $x \in U$, $y \in U_r$:*

(a) (**Boundary decay**). *For every integer $\ell = 0, 1, 2, 3, 4$,*

$$|\partial_x^\ell G_h^D(x, y) - \partial_x^\ell G_h(x - y)| \leq C e^{-cr_0^2/h} h^{-\ell/2} G_{ch}(x - y). \quad (91)$$

(b) (**Interior derivative bound**). *For every finite signed measure ν on U_r and every $\ell = 0, 1, 2, 3, 4$,*

$$\int_{U_r} |\partial_x^\ell G_h^D(x, y)| |\nu(y)| dy \leq C h^{-(d+\ell)/2} \int_{U_r} e^{-c|x-y|^2/h} |\nu(y)| dy. \quad (92)$$

Proof (a) Fix $y \in U_r$ and define

$$H_h(x) := G_h(x - y) - G_h^D(x, y).$$

Then

$$(\partial_h - \Delta_x)H_h = 0 \quad \text{in } (0, h] \times U_r, \quad H_h(x) = G_h(x - y) \text{ on } \partial U_r, \quad \lim_{t \rightarrow 0} H_t \equiv 0.$$

By the parabolic maximum principle,

$$0 \leq H_h(x) \leq \sup_{x_\partial \in \partial U_r} G_h(x_\partial - y). \quad (93)$$

If $x \in U$, then $\text{dist}(x, \partial U_r) \geq r_0$, hence $|x_\partial - y| \geq ||x - y| - r_0|$. Using $(a - b)^2 \geq \frac{a^2}{2} - b^2$ with $a = |x - y|$ and $b = r_0$,

$$e^{-|x_\partial - y|^2/(4h)} \leq e^{-(|x - y| - r_0)^2/(4h)} \leq e^{-|x - y|^2/(8h)} e^{-r_0^2/(8h)}.$$

Thus,

$$|H_h(x)| \leq C e^{-cr_0^2/h} G_{ch}(x - y), \quad x \in U.$$

Fix $|\lambda| = \ell \in \{1, 2, 3, 4\}$. Then $u := \partial_x^\lambda H_h$ solves the heat equation on $(0, h] \times U_r$. Let $s := \min\{r_0/2, \sqrt{h}\}$ and $Q_s(x, h) := B_s(x) \times [h - s^2, h]$. By interior parabolic estimates,

$$\|\partial_x^\lambda H_h\|_{L^\infty(B_{s/2}(x))} \leq C s^{-\ell} \|H\|_{L^\infty(Q_s(x, h))}.$$

Since $x \in U$ and $s \leq r_0/2$, the bound on H_h from above holds in $Q_s(x, h)$. Moreover, for all $h \in (0, 1]$ we have $s^{-\ell} \leq C h^{-\ell/2}$, with a constant depending only on r_0 and the finite range $\ell \leq 4$. Therefore,

$$|\partial_x^\ell H_h(x)| \leq C h^{-\ell/2} e^{-cr_0^2/h} G_{ch}(x - y).$$

Since $H_h = G_h - G_h^D$, this is exactly (91) for $\ell \geq 1$, and for $\ell = 0$ we already obtained the same bound on H_h above.

(b) For $\ell \leq 4$,

$$\partial_x^\ell G_h^D(x, y) = \partial_x^\ell G_h(x - y) + \partial_x^\ell (G_h^D(x, y) - G_h(x - y)).$$

The free term satisfies the standard Gaussian derivative bound

$$|\partial_x^\ell G_h(x - y)| \leq C h^{-(d+\ell)/2} e^{-c|x-y|^2/h}. \quad (94)$$

The difference term is bounded by part (a):

$$|\partial_x^\ell (G_h^D - G_h)(x, y)| \leq C e^{-cr_0^2/h} h^{-\ell/2} G_{ch}(x - y) \leq C h^{-(d+\ell)/2} e^{-c'|x-y|^2/h},$$

where we absorbed the prefactor $e^{-cr_0^2/h}$ into the Gaussian with a possibly different constant $c' > 0$. Hence,

$$|\partial_x^\ell G_h^D(x, y)| \leq C h^{-(d+\ell)/2} e^{-c|x-y|^2/h}.$$

Integrating against $|\nu|$ yields (92). This completes the proof. \blacksquare

Lemma 17 (Kernel resolvent bound in $C^{3,1}$) *Let G_h^D be the Dirichlet heat kernel on U_r . Fix $p > d + 2$ and for $h \in (0, 1]$, set*

$$\mathcal{G}_h^D \nu(x) := \int_{U_r} G_h^D(x, y) \nu(y) dy, \quad \mathcal{R}_h^D \nu := \sum_{n \geq 1} (\mathcal{G}_h^D)^n \nu.$$

Then there exists a constant $C = C(d, U, U_r, p)$ such that, for all $h \in (0, 1]$ and all finite signed measures ν on U_r ,

$$\|\mathcal{R}_h^D \nu\|_{C^{3,1}(U)} \leq C M_p(\nu).$$

Proof *Step 1. One-step $C^{3,1}$ bounds.* From Lemma 16 (valid for $\ell = 0, 1, 2, 3, 4$), we have

$$|\partial_x^\ell G_h^D(x, y)| \leq C h^{-(d+\ell)/2} e^{-c\|x-y\|^2/h},$$

and hence

$$\|\partial_x^\ell \mathcal{G}_h^D \nu\|_{L^\infty(U)} \leq C h^{-(d+\ell)/2} \int_{U_r} e^{-c\|x-y\|^2/h} |\nu(y)| dy.$$

Since U_r is bounded and $p > d$, there is $C = C(d, U_r, p)$ such that

$$\int_{U_r} e^{-c\|x-y\|^2/h} |\nu(y)| dy \leq C M_p(\nu), \quad x \in U, h \in (0, 1].$$

Thus

$$\|\partial_x^\ell \mathcal{G}_h^D \nu\|_{L^\infty(U)} \leq C h^{-(d+\ell)/2} M_p(\nu), \quad \ell = 0, 1, 2, 3. \quad (95)$$

For the Lipschitz seminorm of $\nabla_x^3 f$, note that on the bounded set U ,

$$[\nabla_x^3 f]_{\text{Lip}(U)} \leq C(U) \|\nabla_x^4 f\|_{L^\infty(U)}.$$

Since Lemma 16 holds for $\ell = 4$,

$$\|\nabla_x^4 \mathcal{G}_h^D \nu\|_{L^\infty(U)} \leq C h^{-(d+4)/2} M_p(\nu).$$

Combining (95) for $\ell \leq 3$ with this Lipschitz control,

$$\|\mathcal{G}_h^D \nu\|_{C^{3,1}(U)} \leq C \left(h^{-d/2} + h^{-(d+1)/2} + h^{-(d+2)/2} + h^{-(d+3)/2} + h^{-(d+4)/2} \right) M_p(\nu). \quad (96)$$

Step 2. Telescoping identity and decomposition. For $N \geq 1$, since $s \mapsto \mathcal{G}_s^D \nu$ is smooth in $s > 0$, there exists a constant C independent of h, N such that

$$\left\| \sum_{n=1}^N \mathcal{G}_{nh}^D \nu - \left(\mathcal{G}_h^D \nu + \int_h^{Nh} \partial_s (\mathcal{G}_s^D \nu) ds \right) \right\|_{C^{3,1}(U)} \leq C M_p(\nu). \quad (97)$$

Decompose

$$\Delta_x(G_s^D * \nu) = \Delta_x(G_s * \nu) + \Delta_x((G_s^D - G_s) * \nu).$$

Since $\partial_s G_s = \Delta_x G_s$,

$$\int_h^{Nh} \Delta_x(G_s * \nu) ds = G_{Nh} * \nu - G_h * \nu.$$

Hence, combining with (97), we obtain

$$\left\| \sum_{n=1}^N \mathcal{G}_{nh}^D \nu - \left(G_{Nh} * \nu + (G_h^D - G_h) * \nu + \int_h^{Nh} \Delta_x((G_s^D - G_s) * \nu) ds \right) \right\|_{C^{3,1}(U)} \leq C M_p(\nu). \quad (98)$$

For the boundary correction term, Lemma 16 yields, for $\ell = 0, 1, 2, 3$,

$$|\partial_x^\ell \Delta_x(G_s^D - G_s)(x, y)| \leq C e^{-cr_0^2/s} s^{-(d+\ell+2)/2} e^{-c\|x-y\|^2/s}.$$

Integrating against $|\nu|$ gives

$$\|\partial_x^\ell \Delta_x((G_s^D - G_s) * \nu)\|_{L^\infty(U)} \leq C e^{-cr_0^2/s} s^{-(d+\ell+2)/2} M_p(\nu).$$

Using again the Lipschitz equivalence $\text{Lip}(\nabla^3 f) \leq C \|\nabla^4 f\|_{L^\infty}$,

$$\|\Delta_x((G_s^D - G_s) * \nu)\|_{C^{3,1}(U)} \leq C e^{-cr_0^2/s} s^{-2-d/2} M_p(\nu).$$

Therefore,

$$\left\| \int_h^{Nh} \Delta_x((G_s^D - G_s) * \nu) ds \right\|_{C^{3,1}(U)} \leq C M_p(\nu) \int_h^{Nh} e^{-cr_0^2/s} s^{-2-d/2} ds. \quad (99)$$

The integral in (99) is uniformly bounded when $Nh \geq r_0^2$. Furthermore,

$$\|(G_h^D - G_h) * \nu\|_{C^{3,1}(U)} \leq C e^{-cr_0^2/h} h^{-(d+4)/2} M_p(\nu).$$

Combining the above gives

$$\left\| \sum_{n=1}^N \mathcal{G}_{nh}^D \nu \right\|_{C^{3,1}(U)} \leq \|G_{Nh} * \nu\|_{C^{3,1}(U)} + C M_p(\nu), \quad Nh \geq r_0^2. \quad (100)$$

Step 3. Choice of block length and spectral gap. Choose

$$N_0 := \lceil r_0^2/h \rceil, \quad N_0 h \in [r_0^2, r_0^2 + 1].$$

Then from (96),

$$\|G_{N_0 h} * \nu\|_{C^{3,1}(U)} \leq C M_p(\nu).$$

Inserting this into (100),

$$\left\| \sum_{n=1}^{N_0} \mathcal{G}_{nh}^D \nu \right\|_{C^{3,1}(U)} \leq C M_p(\nu).$$

The tail satisfies (using the Dirichlet spectral gap)

$$\|\mathcal{G}_s^D \nu\|_{C^{3,1}(U)} \leq C e^{-\lambda_1^D s} M_p(\nu), \quad s \geq r_0^2,$$

hence

$$\sum_{n \geq N_0+1} \|\mathcal{G}_{nh}^D \nu\|_{C^{3,1}(U)} \leq C M_p(\nu).$$

Combining block and tail gives

$$\sum_{n \geq 1} \|\mathcal{G}_{nh}^D \nu\|_{C^{3,1}(U)} \leq C M_p(\nu).$$

Since

$$\mathcal{R}_h^D \nu = \sum_{n \geq 1} (\mathcal{G}_h^D)^n \nu = \sum_{n \geq 1} \mathcal{G}_{nh}^D \nu,$$

the proof is complete. \blacksquare

A.2.3 STABILITY AND CONTRACTION OF THE KERNEL FORMULA

The results in this section establish the regularity and boundedness properties of the kernel sequence $\hat{\rho}_k$, which serve as intermediate steps for the proof in Section 3.

Lemma 18 (Regularization and contraction of P_V^h) *Then there exists $h_0 \in (0, 1]$ and a constant $C = C(d, \beta, \alpha, \|V\|_{C^{3,1}(U_r)}, U, U_r)$ such that for all $h \in (0, h_0]$ and $\varphi \in C^{2,1}(U_r)$,*

$$\|P_V^h \varphi\|_{L^\infty(U_r)} \leq \|\varphi\|_{L^\infty(U_r)}, \quad (101)$$

$$\|\nabla P_V^h \varphi\|_{L^\infty(U)} \leq \frac{1}{1 + \alpha h} \|\nabla \varphi\|_{L^\infty(U_r)}, \quad (102)$$

$$\|P_V^h \varphi\|_{C^{3,1}(U)} \leq C \|\varphi\|_{C^{3,1}(U_r)}. \quad (103)$$

Proof *Step 1. L^∞ contraction and Strong log-concavity.* For each fixed $y \in U_r$, $P_V^h \varphi(y) = \mathbb{E}_{\mu_y}[\varphi(X)]$ is an average of φ against a probability measure. Hence

$$|P_V^h \varphi(y)| \leq \|\varphi\|_{L^\infty(U_r)}, \quad y \in U_r.$$

The density of μ_y is proportional to $e^{-U_y(x)}$, where

$$\nabla_x^2 U_y(x) = \frac{\beta}{2} (\nabla^2 V(x) + h^{-1} I_d) \succeq \kappa I_d, \quad \kappa := \frac{\beta}{2} \left(\alpha + \frac{1}{h} \right).$$

Thus μ_y is κ -strongly log-concave. By Poincaré inequality,

$$\text{Var}_{\mu_y}(f) \leq \frac{1}{\kappa} \mathbb{E}_{\mu_y} \|\nabla f(X)\|^2, \quad f \in C_{\text{loc}}^1(\mathbb{R}^d). \quad (104)$$

In particular, for any unit $u \in \mathbb{R}^d$,

$$\text{Var}_{\mu_y}(\langle X - y, u \rangle) \leq \frac{1}{\kappa}, \quad \kappa = \frac{\beta}{2} \left(\alpha + \frac{1}{h} \right).$$

Step 2. Score identity and gradient contraction. Recall the score identity (differentiate under the integral):

$$\nabla_y \mathbb{E}_{\mu_y}[g(X)] = \frac{\beta}{2h} \text{Cov}_{\mu_y}(g(X), X - y), \quad g \in C_{\text{loc}}^1(\mathbb{R}^d). \quad (105)$$

For any unit u ,

$$D_u \mathbb{E}_{\mu_y}[g(X)] = \frac{\beta}{2h} \text{Cov}_{\mu_y}(g(X), \langle X - y, u \rangle).$$

Cauchy-Schwarz and the Poincaré inequality (104) give

$$|\text{Cov}_{\mu_y}(g, \langle X - y, u \rangle)| \leq \sqrt{\text{Var}_{\mu_y}(g) \text{Var}_{\mu_y}(\langle X - y, u \rangle)} \leq \kappa^{-1} \sqrt{\mathbb{E}_{\mu_y} \|\nabla g(X)\|^2} \leq \kappa^{-1} \|\nabla g\|_{L^\infty(U_r)}.$$

Thus

$$|D_u \mathbb{E}_{\mu_y}[g(X)]| \leq \frac{\beta}{2h} \kappa^{-1} \|\nabla g\|_{L^\infty(U_r)} = \frac{1}{1 + \alpha h} \|\nabla g\|_{L^\infty(U_r)}.$$

Applying this to $g = \varphi$ and taking the supremum over all unit u gives precisely

$$\|\nabla P_V^h \varphi\|_{L^\infty(U)} = \|\nabla_y \mathbb{E}_{\mu_y}[\varphi(X)]\|_{L^\infty(U)} \leq \frac{1}{1 + \alpha h} \|\nabla \varphi\|_{L^\infty(U_r)},$$

which is (102).

Step 3. Second derivatives and $C^{2,1}$ bound. By definition, $\nabla P_V^h \varphi$ can be written as

$$\nabla P_V^h \varphi(y) = \underbrace{\mathbb{E}_{\mu_y}[\nabla \varphi(X)]}_{=:A(y)} - \frac{\beta}{2} \underbrace{\text{Cov}_{\mu_y}(\nabla V(X), \varphi(X))}_{=:B(y)}. \quad (106)$$

Thus

$$\nabla^2 P_V^h \varphi(y) = \nabla A(y) - \frac{\beta}{2} \nabla B(y).$$

Each component of A is of the form $A_i(y) = \mathbb{E}_{\mu_y}[\partial_{x_i} \varphi(X)]$. Applying the score identity (105) to $g = \partial_{x_i} \varphi$, and repeating the argument in Step 2, we obtain

$$\|\nabla A_i\|_{L^\infty(U)} \leq \frac{1}{1 + \alpha h} \|\nabla(\partial_{x_i} \varphi)\|_{L^\infty(U_r)} \leq \frac{1}{1 + \alpha h} \|\nabla^2 \varphi\|_{L^\infty(U_r)}.$$

Hence

$$\|\nabla A\|_{L^\infty(U)} \leq \frac{1}{1 + \alpha h} \|\nabla^2 \varphi\|_{L^\infty(U_r)}. \quad (107)$$

Write

$$B(y) = \mathbb{E}_{\mu_y}[\nabla V(X) \varphi(X)] - \mathbb{E}_{\mu_y}[\nabla V(X)] \mathbb{E}_{\mu_y}[\varphi(X)].$$

Differentiate along a unit direction u . Using the score identity (105) for each expectation, we obtain

$$D_u B(y) = \frac{\beta}{2h} \left\{ \text{Cov}_{\mu_y}(\nabla V \varphi, \langle X - y, u \rangle) - \mathbb{E}_{\mu_y}[\varphi] \text{Cov}_{\mu_y}(\nabla V, \langle X - y, u \rangle) - \mathbb{E}_{\mu_y}[\nabla V] \text{Cov}_{\mu_y}(\varphi, \langle X - y, u \rangle) \right\}.$$

Each covariance is controlled by Poincaré (104). For example, since $\|\nabla(\nabla V \varphi)\| \leq \|\nabla^2 V\| \|\varphi\| + \|\nabla V\| \|\nabla \varphi\|$, we have

$$\|\text{Cov}_{\mu_y}(\nabla V \varphi, \langle X - y, u \rangle)\| \leq \kappa^{-1} \|\nabla(\nabla V \varphi)\|_{L^\infty(U_r)} \leq Ch \|\varphi\|_{C^{1,1}(U_r)}.$$

Similarly,

$$\|\text{Cov}_{\mu_y}(\nabla V, \langle X - y, u \rangle)\| \leq Ch, \quad |\text{Cov}_{\mu_y}(\varphi, \langle X - y, u \rangle)| \leq Ch \|\varphi\|_{C^{0,1}(U_r)}.$$

Combining these bounds, we obtain

$$\|\nabla B\|_{L^\infty(U)} = [B]_{\text{Lip}(U)} \leq Ch \|\varphi\|_{C^{1,1}(U_r)} \leq Ch \|\varphi\|_{C^{2,1}(U_r)}. \quad (108)$$

From (107) and (108),

$$\|\nabla^2 P_V^h \varphi\|_{L^\infty(U)} \leq \|\nabla A\|_{L^\infty(U)} + \frac{\beta}{2} \|\nabla B\|_{L^\infty(U)} \leq \frac{1}{1 + \alpha h} \|\nabla^2 \varphi\|_{L^\infty(U_r)} + Ch \|\varphi\|_{C^{2,1}(U_r)}.$$

Since $h \leq h_0 \leq 1$, the right-hand side is bounded by

$$\|\nabla^2 P_V^h \varphi\|_{L^\infty(U)} \leq C \|\varphi\|_{C^{2,1}(U_r)}.$$

A similar differentiation of (106) one more time (using the score identity and Poincaré, and the fact that $V \in C^{4,1}$) shows that the Lipschitz seminorm of $\nabla^3 P_V^h \varphi$ is also bounded by $C \|\varphi\|_{C^{3,1}(U_r)}$. We omit the repetitive details and summarize

$$\|P_V^h \varphi\|_{C^{3,1}(U)} \leq C \|\varphi\|_{C^{3,1}(U_r)},$$

for all $h \in (0, h_0]$, which is (103). ■

Lemma 19 (Local $C^{3,1}$ stability of \mathcal{K}_V^h) *There exist $h_0 \in (0, 1]$ and $C < \infty$, depending only on*

$$(d, \beta, \alpha, \|V\|_{C^{4,1}(U_r)}, U, U_r),$$

such that for all $h \in (0, h_0]$ and all $f \in C^{3,1}(U_r)$,

$$\|\mathcal{K}_V^h f\|_{C^{3,1}(U)} \leq C \|f\|_{C^{3,1}(U_r)}. \quad (109)$$

Proof By the kernel representation (20),

$$\mathcal{K}_V^h f(x) = \psi(x) \int_{\mathbb{R}^d} G_h(x-y) \frac{f(y)}{Z(y)} dy, \quad x \in U_r.$$

Assumption 1 gives $\psi^{\pm 1} \in C^{4,1}(U_r)$. Moreover, for $h \leq h_0$ sufficiently small, the normalizing factor Z satisfies $0 < c_0 \leq Z \leq C_0$ on U_r and $Z^{\pm 1} \in C^{3,1}(U_r)$ with bounds depending only on the listed parameters.

Let $|\lambda| \leq 3$. Using $\partial_{x_i} G_h(x-y) = -\partial_{y_i} G_h(x-y)$, we repeatedly integrate by parts in y to transfer all x -derivatives from G_h onto f/Z . This yields the representation

$$\partial_x^\lambda \mathcal{K}_V^h f(x) = \sum_{\mu \leq \lambda} a_{\lambda, \mu}(x) \int_{\mathbb{R}^d} G_h(x-y) D_y^{\lambda-\mu} \left(\frac{f(y)}{Z(y)} \right) dy,$$

where each coefficient $a_{\lambda, \mu}$ is a linear combination of $\partial_x^\nu \psi(x)$ for $|\nu| \leq |\lambda|$. Hence $\|a_{\lambda, \mu}\|_{C^{0,1}(U)} \leq C$ since $\psi \in C^{4,1}(U_r)$.

Because $\int_{\mathbb{R}^d} G_h(x-y) dy = 1$ and $x \in U \Subset U_r$, we obtain

$$\sup_{x \in U} |\partial_x^\lambda \mathcal{K}_V^h f(x)| \leq C \sup_{y \in U_r} \max_{|\gamma| \leq |\lambda|} |D_y^\gamma (f(y)/Z(y))| \leq C \|f\|_{C^{3,1}(U_r)}.$$

The same formula, applied to difference quotients in x , yields uniform Lipschitz bounds for all derivatives up to order 3 on U , again bounded by $C \|f\|_{C^{3,1}(U_r)}$. Thus (109) follows. ■

A.2.4 LOCAL UNIFORM POSITIVITY FOR DISCRETE ITERATES

The next lemma establishes the uniform positivity of the discrete iterates. This property will be used in the proof of Theorem 8, where the score term appears in the form $\nabla \rho / \rho$ and it is therefore essential to ensure that the denominator remains uniformly bounded away from zero.

Lemma 20 *There exist $h_0 \in (0, 1]$ and $c_{U, h_0} > 0$ such that, for all $0 < h \leq h_0$ and all $k \geq 0$,*

$$\inf_{x \in U} \widehat{\rho}_k(x) \geq c_{U, h_0}, \quad \inf_{x \in U} \rho_k(x) \geq c_{U, h_0}. \quad (110)$$

Proof We first construct a lower bound for the density in the continuous Fokker-Planck equation and then use the weak convergence estimates to derive a lower bound for the discrete sequence.

(i) *FP flow.* By standard parabolic regularity and strict ellipticity of $\beta^{-1}\Delta - \nabla V \cdot \nabla$, the Fokker-Planck solution $\rho(t, \cdot)$ is $C^{2,1}$ on U_r for each $t > 0$, and the map $(t, x) \mapsto \rho(t, x)$ is continuous on $[0, T] \times \overline{U}$ for every $T > 0$.

By Assumption 3 on the initial density, there exists $c_0 > 0$ such that

$$\inf_{x \in U_r} \rho_0(x) \geq c_0.$$

Since $\rho(t, x) > 0$ for all $t > 0$ and $x \in U$ (Evans, 2022), and $(t, x) \mapsto \rho(t, x)$ is continuous on $[0, h_0] \times \overline{U}$ for any fixed $h_0 > 0$, the infimum value on this compact set is strictly positive:

$$c_1 := \inf_{0 \leq t \leq h_0} \inf_{x \in U} \rho(t, x) > 0.$$

On the other hand, the invariant density

$$\rho^*(x) = Z^{-1} e^{-\beta V(x)}$$

is continuous and strictly positive on U_r , hence

$$c_* := \inf_{x \in U} \rho^*(x) > 0.$$

Moreover, by the uniform convexity assumption on V , the Fokker-Planck semigroup converges exponentially fast to ρ^* in $C^0(U_r)$; namely, there exist constants $C \geq 1$ and $\lambda > 0$ such that

$$\|\rho(t) - \rho^*\|_{C^0(U_r)} \leq C e^{-\lambda t} \|\rho_0 - \rho^*\|_{C^0(U_r)}, \quad t \geq 0. \quad (111)$$

Since $U \Subset U_r$ and $\rho^* \propto e^{-\beta V}$ is continuous and strictly positive on U_r , there exists $c_* > 0$ with $\inf_{x \in U} \rho^*(x) \geq c_*$. Choose $h_1 \geq h_0$ large enough so that

$$C e^{-\lambda h_1} \|\rho_0 - \rho^*\|_{C^0(U_r)} \leq \frac{1}{2} c_*.$$

Then for all $t \geq h_1$ and $x \in U$,

$$\rho(t, x) \geq \rho^*(x) - \|\rho(t) - \rho^*\|_{C^0(U_r)} \geq c_* - \frac{1}{2} c_* = \frac{1}{2} c_*.$$

Combining this with the definition of c_1 (and enlarging h_0 to h_1 if necessary) gives

$$\inf_{t \geq 0} \inf_{x \in U} \rho(t, x) \geq \min\{c_1, \frac{1}{2} c_*\} =: c_U > 0. \quad (112)$$

(ii) *Discrete iterates.* We now transfer the bound (112) to the discrete chains. By Theorem 4, combined with a discrete Grönwall argument as in Theorem 23, there exists $C_1 > 0$ such that, for all $0 < h \leq 1$, all $k \geq 0$ and all test functions $\phi \in C_C^{1,1}(U_r)$,

$$|\langle \phi, \rho_k - \rho(kh) \rangle| + |\langle \phi, \widehat{\rho}_k - \rho(kh) \rangle| \leq C_1 h \|\phi\|_{C^{1,1}(U_r)}. \quad (113)$$

Moreover, by Theorem 7 the families $(\rho(t))_{t \geq 0}$, $(\rho_k)_{k \geq 0}$ and $(\widehat{\rho}_k)_{k \geq 0}$ are uniformly bounded in $C^{2,1}(U_r)$. In particular, all three families are uniformly Lipschitz on U :

$$\sup_{t \geq 0} \|\rho(t)\|_{C^{0,1}(U)} + \sup_{k \geq 0} \|\rho_k\|_{C^{0,1}(U)} + \sup_{k \geq 0} \|\widehat{\rho}_k\|_{C^{0,1}(U)} \leq C_2,$$

for some constant C_2 independent of h .

Fix $x \in U$ and let $\eta \in C^\infty(B_1(0))$ be a standard bump with $\int \eta = 1$. For $r > 0$ small enough so that $B_r(x) \subset U$, define

$$\phi_{x,r}(y) := \frac{1}{r^d} \eta\left(\frac{y-x}{r}\right), \quad y \in \mathbb{R}^d.$$

Then $\phi_{x,r} \in C_C^{1,1}(U_r)$, and by scaling we have

$$\|\phi_{x,r}\|_{C^{1,1}(U_r)} \leq C r^{-d-2},$$

for some constant C depending only on η and d . Moreover, for any Lipschitz density f on U ,

$$\left| \int f(y) \phi_{x,r}(y) dy - f(x) \right| \leq C r \|f\|_{C^{0,1}(U)}.$$

Applying this first to $f = \rho(kh, \cdot)$ and then to $f = \rho_k(\cdot)$ and $f = \widehat{\rho}_k(\cdot)$, and using (113), we obtain

$$\begin{aligned} |\rho_k(x) - \rho(kh, x)| &\leq |\langle \phi_{x,r}, \rho_k - \rho(kh) \rangle| + C r (\|\rho_k\|_{C^{0,1}(U)} + \|\rho(kh)\|_{C^{0,1}(U)}), \\ |\widehat{\rho}_k(x) - \rho(kh, x)| &\leq |\langle \phi_{x,r}, \widehat{\rho}_k - \rho(kh) \rangle| + C r (\|\widehat{\rho}_k\|_{C^{0,1}(U)} + \|\rho(kh)\|_{C^{0,1}(U)}). \end{aligned}$$

By (113) and the uniform Lipschitz bounds, this yields

$$|\rho_k(x) - \rho(kh, x)| + |\widehat{\rho}_k(x) - \rho(kh, x)| \leq C_3 (h r^{-d-2} + r),$$

for some constant C_3 independent of k , $x \in U$ and $0 < h \leq 1$.

Choosing, for example, $r = h^{1/(d+3)}$ gives

$$\sup_{k \geq 0} \|\rho_k - \rho(kh)\|_{L^\infty(U)} + \sup_{k \geq 0} \|\widehat{\rho}_k - \rho(kh)\|_{L^\infty(U)} \leq C_4 h^{1/(d+3)},$$

for some $C_4 > 0$ independent of h .

Combining this with the lower bound (112), we obtain, for all $x \in U$, $k \geq 0$ and $0 < h \leq 1$,

$$\rho_k(x) \geq \rho(kh, x) - C_4 h^{1/(d+3)} \geq c_U - C_4 h^{1/(d+3)},$$

and the same inequality with ρ_k replaced by $\widehat{\rho}_k$. Finally, choose $h_0 \in (0, 1]$ so small that $C_4 h_0^{1/(d+3)} \leq c_U/2$, and set

$$c_{U, h_0} := \frac{1}{2} c_U > 0.$$

Then for all $0 < h \leq h_0$, all $k \geq 0$ and all $x \in U$,

$$\rho_k(x) \geq c_{U, h_0}, \quad \widehat{\rho}_k(x) \geq c_{U, h_0},$$

which proves (110). ■

A.2.5 COROLLARIES FOR KERNEL APPROXIMATION

In this section, we record several approximation properties of $\mathcal{K}_V^h \rho_k$. Although our algorithm ultimately requires only the score approximation, the corresponding density estimates are not the main focus. Nevertheless, the following corollaries remain of independent interest for applications in which one wishes to use the kernel formula to approximate the full density evolution.

Lemma 21 (One-step weak error via kernel comparison in the dual $C^{0,1}$ norm) *Define the local error and its weak $C^{0,1}$ -dual norm as*

$$\epsilon_k := \rho_k - \widehat{\rho}_k, \quad e_k^{0,1} := \|\epsilon_k\|_{(C^{0,1}(U_r))^*}.$$

Then there exist constants $h_0 \in (0, 1]$, $c_1 \in (0, \alpha/2]$, and $C_2, C_3 < \infty$, depending only on

$$(d, \beta, \alpha, \|V\|_{C^{3,1}(U_r)}, U, U_r, M_p(\rho_0)),$$

such that for all $k \geq 0$ and $h \in (0, h_0]$,

$$e_{k+1}^{0,1} \leq (1 - c_1 h + C_2 h^2) e_k^{0,1} + C_3 h^2 \|\rho_k\|_{C^{2,1}(U_r)}. \quad (114)$$

Proof Fix $\varphi \in C^{0,1}(U)$ with $\|\varphi\|_{C^{0,1}(U)} \leq 1$, and extend it to $\tilde{\varphi} \in C^{0,1}(U_r)$ with

$$\tilde{\varphi}|_U = \varphi, \quad \|\tilde{\varphi}\|_{C^{0,1}(U_r)} \leq C_{\text{ext}}.$$

Then

$$\langle \varphi, \epsilon_{k+1} \rangle = \langle \varphi, \rho_{k+1} - \widehat{\rho}_{k+1} \rangle = \underbrace{\langle \varphi, (F_k)_\# \rho_k - \mathcal{K}_V^h \rho_k \rangle}_{\mathcal{E}_1} + \underbrace{\langle \varphi, \mathcal{K}_V^h(\rho_k - \widehat{\rho}_k) \rangle}_{\mathcal{E}_2}. \quad (115)$$

By duality and the gradient contraction in Lemma 18, for some $c_1 \in (0, \alpha/2]$ and $C_2 < \infty$ depending only on $(d, \beta, \alpha, \|V\|_{C^{3,1}}, U, U_r)$, we have

$$|\mathcal{E}_2| = |\langle P_V^h \varphi, \epsilon_k \rangle| \leq \|P_V^h \varphi\|_{C^{0,1}(U)} e_k^{0,1} \leq (1 - c_1 h + C_2 h^2) e_k^{0,1}. \quad (116)$$

By the weak expansion of P_V^h in Theorem 4 and the expansion of $(F_k)_\# \rho_k$ through the change of variable formula, there exists a constant $C_3 < \infty$ depending only on $(d, \beta, \alpha, \|V\|_{C^{3,1}}, U, U_r, M_p(\rho_0))$ such that for all $\varphi \in C^{0,1}(U_r)$,

$$|\langle \varphi, (F_k)_\# \rho_k - \mathcal{K}_V^h \rho_k \rangle| \leq C_3 h^2 \|\varphi\|_{C^{0,1}(U_r)} \|\rho_k\|_{C^{2,1}(U_r)}. \quad (117)$$

Combining (116) and (117) in (115), we obtain

$$|\langle \varphi, \epsilon_{k+1} \rangle| \leq (1 - c_1 h + C_2 h^2) e_k^{0,1} + C_3 h^2 \|\rho_k\|_{C^{2,1}(U_r)}.$$

Taking the supremum over all $\varphi \in C^{0,1}(U)$ with $\|\varphi\|_{C^{0,1}(U)} \leq 1$ yields

$$e_{k+1}^{0,1} \leq (1 - c_1 h + C_2 h^2) e_k^{0,1} + C_3 h^2 \|\rho_k\|_{C^{2,1}(U_r)},$$

which is exactly (114). ■

We remark that the above Lemma implies $e_k^{0,1}$ is a contraction when $1 - c_1 h + C_2 h^2 < 1$, hence, in the following, we assume $h_0 \leq c_1/(2C_2)$ which ensures that

$$e_{k+1}^{0,1} \leq (1 - c_1 h/2) e_k^{0,1} + C_3 h^2 \|\rho_k\|_{C^{2,1}(U_r)}.$$

We stated the following approximation results for non-refreshed chains as a direct corollary for Lemma 21.

Corollary 22 For any $\varphi \in C_{\text{loc}}^{2,1}(\mathbb{R}^d)$ and any bounded open set $U \subset \mathbb{R}^d$, there exist constants $C > 0$ and $h_0 > 0$ such that for all $h \in (0, h_0]$ and all $k \geq 0$,

$$|\langle \varphi, \widehat{\rho}_k - \rho_k \rangle| \leq Ch.$$

Proof By Lemma 21, the local weak error between one kernel step and one BRWP step satisfies

$$|\langle \varphi, \widehat{\rho}_{k+1} - \rho_{k+1} \rangle| \leq (1 - ch + Ch^2) |\langle \varphi, \widehat{\rho}_k - \rho_k \rangle| + C\|\varphi\|_{C^{2,1}(U)}h^2.$$

Applying the discrete Grönwall inequality to this recursion yields the uniform bound

$$|\langle \varphi, \widehat{\rho}_k - \rho_k \rangle| \leq Ch, \quad k \geq 0,$$

where C depends on φ and U , but is independent of k and of h for $h \leq h_0$. ■

The next result states the approximation we have if we apply the kernel formula to the exact solution of the Fokker–Planck equation.

Corollary 23 (Weak one–step density approximation) There exist $h_0 \in (0, 1]$ and

$$C = C(d, \beta, a_1, a_2, C_V, q_V, p) < \infty$$

such that, for any $t \geq 0$, and all $h \in (0, h_0]$,

$$\left| \langle \rho(t), \mathcal{K}_V^h \rho(t) - \rho(t+h) \rangle \right| \leq Ch^2 \left(\|\rho(t)\|_{C^{2,1}(U)} \|\rho(t)\|_{C^{0,1}(U)} + 1 \right). \quad (118)$$

Proof Fix $t \geq 0$ and write $\rho := \rho(t)$. We add and subtract the first–order term generated by the Fokker–Planck operator:

$$\langle \rho, \mathcal{K}_V^h \rho - \rho(t+h) \rangle = \underbrace{\langle \rho, \mathcal{K}_V^h \rho - \rho - h \mathcal{L}^* \rho \rangle}_{\text{(I)}} + \underbrace{\langle \rho, \rho + h \mathcal{L}^* \rho - \rho(t+h) \rangle}_{\text{(II)}}.$$

For (I), Theorem 4 and the duality $\langle P_V^h f, g \rangle = \langle f, \mathcal{K}_V^h g \rangle$ give

$$|\text{(I)}| = |\langle P_V^h \rho - \rho - h \mathcal{L} \rho, \rho \rangle| \leq Ch^2 \left(\|\rho\|_{C^{2,1}(U)} \|\rho\|_{C^{0,1}(U)} + 1 \right).$$

For (II), we use the Fokker–Planck equation. A Taylor expansion in time yields

$$\rho(t+h) = \rho + h \mathcal{L}^* \rho + \mathcal{O}(h^2) \quad \text{in the weak sense against } C^{0,1}(U) \text{ test functions,}$$

with the $\mathcal{O}(h^2)$ term bounded by $Ch^2(\|\rho\|_{C^{2,1}(U)} + 1)$. Testing this expansion against ρ gives

$$|\text{(II)}| \leq Ch^2 \left(\|\rho\|_{C^{2,1}(U)} \|\rho\|_{C^{0,1}(U)} + 1 \right).$$

Combining the bounds for (I) and (II) yields (118). ■

Appendix B. Postponed Proofs and Lemmas for Section 4

In this section, we provide proofs for lemmas and theorems that are used in our analysis. Some of the results in Section B.1 are well-known. We include them in the appendix for the sake of completeness.

B.1 Identities and inequalities along the flow of the Fokker-Planck equation

Let $\rho^*(x) = \frac{1}{Z} \exp(-\beta V(x))$. When $\nabla^2 V \succeq \alpha I$, it is known that ρ^* satisfies the log-Sobolev inequality. Specifically, for any smooth function g with $E_{\rho^*}(g^2) \leq \infty$, we have

$$\int g^2 \log g^2 \rho^* dx - \int g^2 \rho^* dx \log \int g^2 \rho^* dx \leq \frac{2}{\beta \alpha} \int \|\nabla g\|^2 \rho^* dx. \quad (119)$$

Note that the factor of β arises from the definition $\rho^* = \frac{1}{Z} \exp(-\beta V)$, and our assumption applies to the function V .

Using the log-Sobolev inequality, we can derive the following dissipation result for the Fokker-Planck equation (5) with solution ρ .

Lemma 24 *If ρ^* satisfies the log-Sobolev inequality (119) and ρ is the solution to the Fokker-Planck equation (5), then*

$$\frac{d}{dt} \mathcal{D}_{\text{KL}}(\rho \| \rho^*) = -\beta^{-1} \mathcal{I}(\rho \| \rho^*) \leq -2\alpha \mathcal{D}_{\text{KL}}(\rho \| \rho^*).$$

Proof By the definition of KL divergence, since ρ decays sufficiently fast and all derivatives are integrable, boundary terms vanish, we have

$$\frac{d}{dt} \mathcal{D}_{\text{KL}}(\rho \| \rho^*) = \int \frac{\partial}{\partial t} \rho \log \frac{\rho}{\rho^*} dx = \beta^{-1} \int \nabla \cdot \left(\nabla \log \frac{\rho}{\rho^*} \rho \right) \log \frac{\rho}{\rho^*} dx = -\beta^{-1} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \rho dx.$$

Substituting $g^2 = \frac{\rho}{\rho^*}$ to (119), we further obtain the following inequality relating KL divergence and Fisher information

$$\mathcal{D}_{\text{KL}}(\rho \| \rho^*) = \int \log \frac{\rho}{\rho^*} \rho dx \leq \frac{1}{2\beta\alpha} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \rho dx = \frac{1}{2\beta\alpha} \mathcal{I}(\rho \| \rho^*).$$

■

Then, for the second-order time derivative of the KL divergence, we can derive the following two relations using the definition and integration by parts. These relations are crucial for the proof of Lemma 11.

Lemma 25 *When ρ^* satisfies the log-Sobolev inequality and ρ is the solution to the Fokker-Planck equation, the second-order time derivative of KL divergence satisfies*

$$(i) \quad \frac{d^2}{dt^2} \mathcal{D}_{\text{KL}}(\rho \| \rho^*) = 2\beta^{-2} \int \left\| \nabla^2 \log \frac{\rho}{\rho^*} \right\|_{\text{F}}^2 \rho dx + 2\beta^{-1} \int \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla^2 V \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx.$$

$$(ii) \quad \frac{d^2}{dt^2} \mathcal{D}_{\text{KL}}(\rho \| \rho^*) = -\beta^{-1} \frac{d}{dt} \mathcal{I}(\rho \| \rho^*) \geq 2\beta^{-1} \alpha \mathcal{I}(\rho \| \rho^*) \geq 4\alpha^2 \mathcal{D}_{\text{KL}}(\rho \| \rho^*).$$

Proof Using the fact that ρ satisfies the Fokker-Planck equation, and recall our definitions of \mathcal{D}^β and the generator \mathcal{L} as used in Section 4 which are

$$\mathcal{D}^\beta(u) := \frac{\beta^{-1}}{\rho} \nabla \cdot (\nabla u \rho), \quad \mathcal{L}(v) = \beta^{-1} \Delta v - \nabla V \cdot \nabla v. \quad (120)$$

Using integration by parts, we then get

$$\begin{aligned}
 \frac{d^2}{dt^2} \text{D}_{\text{KL}}(\rho \|\rho^*) &= \frac{d}{dt} \int \frac{\partial}{\partial t} \rho \log \frac{\rho}{\rho^*} dx = \int \frac{\partial^2 \rho}{\partial t^2} \log \frac{\rho}{\rho^*} dx + \int \left| \frac{\partial \rho}{\partial t} \right|^2 \frac{1}{\rho} dx \\
 &= \int \left[\beta^{-1} \Delta \frac{\partial \rho}{\partial t} + \nabla \cdot \left(\nabla V \frac{\partial \rho}{\partial t} \right) \right] \log \frac{\rho}{\rho^*} dx + \int \left| \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \right|^2 \rho dx \\
 &= \int \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \mathcal{L} \left(\log \frac{\rho}{\rho^*} \right) \rho dx + \int \left| \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \right|^2 \rho dx \\
 &= 2 \int \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \mathcal{L} \left(\log \frac{\rho}{\rho^*} \right) \rho dx + \int \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \left[\mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) - \mathcal{L} \left(\log \frac{\rho}{\rho^*} \right) \right] \rho dx \\
 &= -2\beta^{-1} \int \nabla \log \frac{\rho}{\rho^*} \cdot \nabla \mathcal{L} \left(\log \frac{\rho}{\rho^*} \right) \rho dx + \beta^{-1} \int \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \rho dx \\
 &= -2\beta^{-1} \int \nabla \log \frac{\rho}{\rho^*} \cdot \nabla \mathcal{L} \left(\log \frac{\rho}{\rho^*} \right) \rho dx + \beta^{-1} \int \mathcal{L} \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \rho dx,
 \end{aligned} \tag{121}$$

where the last inequality comes from the fact that $\mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \rho = \beta^{-1} \Delta \rho + \nabla \cdot (\nabla V \rho)$.

The commutator between \mathcal{L} and ∇ for a smooth function f can be written as

$$\nabla \mathcal{L} f - \mathcal{L} \nabla f = \nabla (\beta^{-1} \Delta f - \nabla V \cdot \nabla f) - \beta^{-1} \Delta (\nabla f) + \nabla V \cdot \nabla^2 f = -\nabla^2 V \nabla f.$$

Using Bochner's formula

$$\frac{1}{2} \Delta (\|\nabla f\|_2^2) = \Delta \nabla f \cdot \nabla f + \|\nabla^2 f\|_{\text{F}}^2,$$

we note that

$$\begin{aligned}
 &\frac{1}{2} \mathcal{L} \|\nabla f\|_2^2 - \nabla f \cdot \nabla \mathcal{L}(f) \\
 &= \frac{1}{2} \mathcal{L} \|\nabla f\|_2^2 - \nabla f \cdot \mathcal{L} \nabla f + \langle \nabla f, \nabla^2 V \nabla f \rangle \\
 &= \beta^{-1} \left(\frac{1}{2} \Delta \|\nabla f\|_2^2 - \nabla f \cdot \Delta \nabla f \right) - \frac{1}{2} \nabla V \cdot \nabla \|\nabla f\|_2^2 + \langle \nabla f, \nabla^2 f \nabla V \rangle + \langle \nabla f, \nabla^2 V \nabla f \rangle \\
 &= \beta^{-1} \|\nabla^2 f\|_{\text{F}}^2 + \langle \nabla f, \nabla^2 V \nabla f \rangle,
 \end{aligned} \tag{122}$$

Combining the final equality in (122) with (121), we obtain the first statement.

Substituting $f = \log \frac{\rho}{\rho^*}$ into (122) and combined with (121), we get

$$\begin{aligned}
 \frac{d^2}{dt^2} \text{D}_{\text{KL}}(\rho \|\rho^*) &= 2\beta^{-2} \int \left\| \nabla^2 \log \frac{\rho}{\rho^*} \right\|_{\text{F}}^2 \rho dx + 2\beta^{-1} \int \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla^2 V \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx \\
 &\geq 2\beta^{-1} \alpha \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \rho dx = 2\beta^{-1} \alpha \mathcal{I}(\rho \|\rho^*) \geq 4\alpha^2 \text{D}_{\text{KL}}(\rho \|\rho^*),
 \end{aligned}$$

where we used $\nabla^2 V \succeq \alpha I_d$ and the final inequality follows from Lemma 24. \blacksquare

Lemma 26 *Let ρ satisfy the Fokker-Planck equation (5). Then, the second-order time derivative of the KL divergence satisfies*

$$\frac{1}{2} \frac{d^2}{dt^2} \text{D}_{\text{KL}}(\rho \|\rho^*) = \int \left| \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \right|^2 \rho dx + \beta^{-2} \int \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla^2 \log \frac{\rho}{\rho^*} \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx,$$

where \mathcal{D}^β is defined in (120).

Proof Firstly, by the definition of KL divergence and computation in (121), we have

$$\frac{d^2}{dt^2} \text{D}_{\text{KL}}(\rho \|\rho^*) = \int \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \mathcal{L} \left(\log \frac{\rho}{\rho^*} \right) \rho dx + \int \left| \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \right|^2 \rho dx.$$

Moreover, we note the following relationship

$$\mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) - \mathcal{L} \left(\log \frac{\rho}{\rho^*} \right) = \beta^{-1} \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2.$$

Thus, we can express

$$\begin{aligned} \int \left| \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \right|^2 \rho dx &= \frac{1}{2} \left[\int \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \mathcal{L} \left(\log \frac{\rho}{\rho^*} \right) \rho dx + \int \left| \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \right|^2 \rho dx \right] \\ &\quad + \frac{1}{2} \int \left(\mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) - \mathcal{L} \left(\log \frac{\rho}{\rho^*} \right) \right) \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \rho dx \\ &= \frac{1}{2} \frac{d^2}{dt^2} \text{D}_{\text{KL}}(\rho \|\rho^*) + \frac{\beta^{-1}}{2} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \mathcal{D}^\beta \left(\log \frac{\rho}{\rho^*} \right) \rho dx \\ &= \frac{1}{2} \frac{d^2}{dt^2} \text{D}_{\text{KL}}(\rho \|\rho^*) - \beta^{-2} \int \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla^2 \log \frac{\rho}{\rho^*} \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx, \end{aligned}$$

where the last equality uses the identity for any smooth function u

$$\begin{aligned} \int \|\nabla u\|_2^2 \mathcal{D}^\beta(u) \rho dx &= \beta^{-1} \int \|\nabla u\|_2^2 \nabla \cdot (\nabla u \rho) dx \\ &= -\beta^{-1} \int \nabla \|\nabla u\|_2^2 \cdot \nabla u \rho dx = -2\beta^{-1} \int \langle \nabla u, \nabla^2 u \nabla u \rangle \rho dx. \end{aligned}$$

■

B.1.1 CONVERGENCE OF FOURTH ORDER TERM

We show the fourth power of $\|\nabla \log \frac{\rho}{\rho^*}\|$ also converges exponentially in Wasserstein space. The below result is used in the proof of Theorem 12.

Lemma 27 *When ρ^* is strongly log-concave with $\nabla^2 V \succeq \alpha I$ and ρ is the solution to the Fokker-Planck equation, we have*

$$\frac{\partial}{\partial t} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^4 \rho dx \leq -4\alpha \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^4 \rho dx.$$

Proof Taking the time derivative directly, we have

$$\frac{\partial}{\partial t} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^4 \rho dx = 4 \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla \frac{\partial \rho / \partial t}{\rho} \right\rangle \rho dx + \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^4 \frac{\partial \rho}{\partial t} dx. \quad (123)$$

Using the identity $\frac{\partial \rho}{\partial t} = \beta^{-1} \nabla \cdot \left(\nabla \log \frac{\rho}{\rho^*} \right)$ and integration by parts, the first term above equals to

$$\begin{aligned}
 & 4 \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla \frac{\partial \rho / \partial t}{\rho} \right\rangle \rho dx \tag{124} \\
 &= -4 \int \nabla \cdot \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right) \frac{\partial \rho / \partial t}{\rho} \rho dx - 4 \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\langle \nabla \log \frac{\rho}{\rho^*}, \frac{\nabla \rho}{\rho} \right\rangle \frac{\partial \rho}{\partial t} dx \\
 &= -4\beta^{-1} \int \nabla \cdot \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right) \nabla \cdot \left(\nabla \log \frac{\rho}{\rho^*} \rho \right) dx \\
 &\quad - 4\beta^{-1} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\langle \nabla \log \frac{\rho}{\rho^*}, \frac{\nabla \rho}{\rho} \right\rangle \nabla \cdot \left(\nabla \log \frac{\rho}{\rho^*} \rho \right) dx \\
 &= 4\beta^{-1} \int \left\langle \nabla \nabla \cdot \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right), \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx \\
 &\quad + 4\beta^{-1} \int \left\langle \left[\nabla \left\langle \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*}, \frac{\nabla \rho}{\rho} \right\rangle \right], \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx.
 \end{aligned}$$

For the last term in (124), it can be simplified as

$$\begin{aligned}
 & 4\beta^{-1} \int \left\langle \left[\nabla \left\langle \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*}, \frac{\nabla \rho}{\rho} \right\rangle \right], \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx \tag{125} \\
 &= 4\beta^{-1} \int \left\langle \left\langle \nabla \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right), \frac{\nabla \rho}{\rho} \right\rangle, \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx \\
 &\quad + 4\beta^{-1} \int \left\langle \left\langle \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*}, \nabla \left(\frac{\nabla \rho}{\rho} \right) \right\rangle, \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx \\
 &= 4\beta^{-1} \int \left\langle \left(\nabla \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right) \right)^T \nabla \log \frac{\rho}{\rho^*}, \nabla \rho \right\rangle dx \\
 &\quad + 4\beta^{-1} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla \left(\frac{\nabla \rho}{\rho} \right) \nabla \log \frac{\rho}{\rho^*} \right\rangle dx,
 \end{aligned}$$

where we have used for $u = \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*}$, $v = \frac{\nabla \rho}{\rho}$

$$(\nabla(u \cdot v))_i = \sum_j \left[\frac{\partial u_j}{\partial x_i} v_j + u_j \frac{\partial v_j}{\partial x_i} \right] = ((\nabla u) \cdot v)_i + (u \cdot (\nabla v))_i \quad \text{where } (\nabla u)_{ij} = \frac{\partial u_j}{\partial x_i},$$

in first equality. We also used

$$(A \cdot u) \cdot v = \sum_{ij} a_{ij} u_j v_i = \sum_j \left(\sum_i a_{ij} v_i \right) u_j = (A^T v) \cdot u, \quad (u \cdot A) \cdot v = (u^T A) v = u \cdot (Av),$$

in the second equality.

Noting that $\nabla(\nabla \rho / \rho) = \nabla^2 \log \rho$, the last term in (125) equals to

$$\begin{aligned}
 & 4\beta^{-1} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla \left(\frac{\nabla \rho}{\rho} \right) \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx \tag{126} \\
 &= 4\beta^{-1} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla^2 \log \rho \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx.
 \end{aligned}$$

Moreover, applying integration by parts with respect to $\nabla\rho$ to the first term in the last line of (125), we get

$$\begin{aligned}
 & 4\beta^{-1} \int \left\langle \left(\nabla \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right) \right)^T \nabla \log \frac{\rho}{\rho^*}, \nabla \rho \right\rangle dx \\
 &= -4\beta^{-1} \int \nabla \cdot \left[\left(\nabla \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right) \right)^T \nabla \log \frac{\rho}{\rho^*} \right] \rho dx \\
 &= -4\beta^{-1} \int \left\langle \nabla \cdot \left(\nabla \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right)^T \right), \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx \\
 &\quad - 4\beta^{-1} \int \text{Tr} \left\{ \nabla \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right) \nabla^2 \log \frac{\rho}{\rho^*} \right\} \rho dx,
 \end{aligned} \tag{127}$$

where we used for $A = \nabla \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right)$, $b = \nabla \log \frac{\rho}{\rho^*}$, and

$$\begin{aligned}
 \nabla \cdot (A^T b) &= \sum_j \frac{\partial \sum_i a_{ij} b_i}{\partial x_j} = \sum_{ij} \left(\frac{\partial a_{ij}}{\partial x_j} b_i + a_{ij} \frac{\partial b_i}{\partial x_j} \right) = (\nabla \cdot A^T) \cdot b + \sum_{ij} a_{ij} (\nabla b)_i^T \\
 &= (\nabla \cdot A^T) \cdot b + \text{Tr} \{ A(\nabla b) \}, \quad (A_{ij}) = a_{ij}, \quad \nabla \cdot A^T = \sum_j \frac{\partial a_{ij}}{\partial x_j}.
 \end{aligned}$$

Next noting that when $u = \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*}$

$$(\nabla \cdot (\nabla u)^T)_i = \sum_j \frac{\partial (\nabla u)_{ij}}{\partial x_j} = \sum_j \frac{\partial}{\partial x_j} \frac{\partial u_j}{\partial x_i} = \frac{\partial}{\partial x_i} \sum_j \frac{\partial u_j}{\partial x_j} = (\nabla(\nabla \cdot u))_i,$$

hence the first term in both the last line of (127) and (124) cancels. Now, it only remains to look at the last term in (127). Firstly, for a scalar function $\zeta = \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2$ and a vector function $u = \nabla \log \frac{\rho}{\rho^*}$, we have

$$(\nabla(\zeta u))_{ij} = \frac{\partial \zeta u_j}{\partial x_i} = \frac{\partial \zeta}{\partial x_i} u_j + \zeta \frac{\partial u_j}{\partial x_i} = (\nabla \zeta u^T)_{ij} + \zeta (\nabla u)_{ij}.$$

This implies

$$\begin{aligned}
 & -4\beta^{-1} \int \text{Tr} \left\{ \nabla \left(\left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla \log \frac{\rho}{\rho^*} \right) \nabla^2 \log \frac{\rho}{\rho^*} \right\} \rho dx \\
 &= -4\beta^{-1} \int \text{Tr} \left\{ \left(\nabla \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left(\nabla \log \frac{\rho}{\rho^*} \right)^T + \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \nabla^2 \log \frac{\rho}{\rho^*} \right) \nabla^2 \log \frac{\rho}{\rho^*} \right\} \rho dx \\
 &= -8\beta^{-1} \int \text{Tr} \left\{ \left(\nabla^2 \log \frac{\rho}{\rho^*} \cdot \nabla \log \frac{\rho}{\rho^*} \right) \left(\nabla^2 \log \frac{\rho}{\rho^*} \cdot \nabla \log \frac{\rho}{\rho^*} \right)^T \right\} \rho dx \\
 &\quad - 4\beta^{-1} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\| \nabla^2 \log \frac{\rho}{\rho^*} \right\|_F^2 \rho dx \\
 &= -8\beta^{-1} \int \left\| \nabla^2 \log \frac{\rho}{\rho^*} \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \rho dx - 4\beta^{-1} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\| \nabla^2 \log \frac{\rho}{\rho^*} \right\|_F^2 \rho dx,
 \end{aligned} \tag{128}$$

where we also used the fact that $\nabla^2 \log \frac{\rho}{\rho^*}$ is symmetric and $Abb^T A^T = Ab(Ab)^T$.

Finally, the second term in (123) will be

$$\int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^4 \frac{\partial \rho}{\partial t} dx = -4\beta^{-1} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla^2 \log \frac{\rho}{\rho^*} \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx. \quad (129)$$

Combing equations (123) to (129), we arrive

$$\begin{aligned} \frac{\partial}{\partial t} \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^4 \rho dx &= -4\beta^{-1} \int \left\| \nabla^2 \log \frac{\rho}{\rho^*} \right\|_F^2 \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \rho dx \\ &\quad - 4 \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \left\langle \nabla \log \frac{\rho}{\rho^*}, \nabla^2 V \nabla \log \frac{\rho}{\rho^*} \right\rangle \rho dx - 8\beta^{-1} \int \left\| \nabla^2 \log \frac{\rho}{\rho^*} \nabla \log \frac{\rho}{\rho^*} \right\|_2^2 \rho dx. \end{aligned}$$

The desired result can now be achieved by noting the condition that $\nabla^2 V \succeq \alpha I$. \blacksquare

Hence, for the density function evolving according to (38), we have the following result.

Lemma 28 *There exists $C < \infty$, independent of k and h , such that for all sufficiently small $h > 0$,*

$$\int \left\| \nabla \log \frac{\rho_k}{\rho^*} \right\|_2^4 \rho_k dx \leq e^{-4\alpha hk} \int \left\| \nabla \log \frac{\rho_0}{\rho^*} \right\|_2^4 \rho_0 dx + Ch. \quad (130)$$

In particular, the fourth-order moment is uniformly bounded in k .

Proof Let

$$F(\rho) := \int \left\| \nabla \log \frac{\rho}{\rho^*} \right\|_2^4 \rho dx.$$

By Lemma 27, for the exact Fokker–Planck flow ρ_t we have

$$\frac{d}{dt} F(\rho_t) \leq -4\alpha F(\rho_t).$$

Hence,

$$F(\rho_{t+h}) \leq e^{-4\alpha h} F(\rho_t). \quad (131)$$

By the weak one-step expansion in Lemma 9 and the uniform $C^{3,1}$ bounds on ρ_k (using ρ_k as the test function), we have

$$F(\rho_{k+1}) = F(\rho_{t_k+h}) + O(h^2),$$

where ρ_{t_k+h} is the exact Fokker–Planck solution starting from ρ_k . Combining with (131) yields

$$F(\rho_{k+1}) \leq e^{-4\alpha h} F(\rho_k) + Ch^2.$$

Iterating the above inequality gives

$$F(\rho_k) \leq e^{-4\alpha hk} F(\rho_0) + Ch^2 \sum_{j=0}^{k-1} e^{-4\alpha hj} \leq e^{-4\alpha hk} F(\rho_0) + Ch,$$

since $\sum_{j \geq 0} e^{-4\alpha hj} \leq (4\alpha h)^{-1}$. \blacksquare

B.2 Postponed proof and additional Lemma used in Section 4

Proof [Proof of Lemma 9] *Reduction by moment condition.* Choose $\chi_R \in C_c^\infty(\mathbb{R}^d)$ with $\chi_R \equiv 1$ on B_R and $\chi_R \equiv 0$ on B_{2R}^c , and set $u_R = \chi_R u$. The moment condition in Lemma 15 imply

$$\langle u_R, \rho_{k+1} - \rho_k \rangle \rightarrow \langle u, \rho_{k+1} - \rho_k \rangle, \quad R \rightarrow \infty.$$

Hence, it suffices to work with $u \in C^{3,1}(U)$.

Step 1. Taylor expansion. From the BRWP update,

$$\langle u, \rho_{k+1} \rangle = \int u \left(x - h \nabla \left(\beta^{-1} \log \frac{\tilde{\rho}_{k+1}}{\rho^*} \right) (x) \right) \rho_k(x) dx.$$

Taylor expansion gives

$$u(x - h \tilde{\Phi}'_k) = u - h \tilde{\Phi}'_k \cdot \nabla u + \frac{h^2}{2} \tilde{\Phi}'_k{}^\top \nabla^2 u \tilde{\Phi}'_k + R_3, \quad |R_3| \leq Ch^3 \|u\|_{C^{3,1}},$$

where

$$\tilde{\Phi}'_k := \nabla \left(\beta^{-1} \log \frac{\tilde{\rho}_{k+1}}{\rho^*} \right).$$

Thus

$$\langle u, \rho_{k+1} - \rho_k \rangle = -h \langle \tilde{\Phi}'_k \cdot \nabla u, \rho_k \rangle + \frac{h^2}{2} \langle \tilde{\Phi}'_k{}^\top \nabla^2 u \tilde{\Phi}'_k, \rho_k \rangle + O(h^3 \|u\|_{C^{3,1}}). \quad (132)$$

Step 2. Drift decomposition via the weak oracle. Write

$$\nabla \left(\beta^{-1} \log \frac{\tilde{\rho}_{k+1}}{\rho^*} \right) = \nabla \left(\beta^{-1} \log \frac{\rho_k}{\rho^*} \right) + h \beta^{-1} \nabla \left(\frac{\mathcal{L}^* \rho_k}{\rho_k} \right) + \varepsilon_{k+1},$$

with the score oracle Assumption 4

$$|\langle w \cdot \varepsilon_{k+1}, \rho_k \rangle| \leq Ch^2 \|w\|_{C^{1,1}(U)}.$$

Applying this with $w = \nabla u$ and $w = (\nabla^2 u) \nabla(\beta^{-1} \log(\rho_k/\rho^*))$ shows that ε_{k+1} contributes only $O(h^3)$.

Step 3. The ideal drift. Substituting into (132) yields

$$\begin{aligned} \langle u, \rho_{k+1} - \rho_k \rangle &= -h \left\langle \nabla \left(\beta^{-1} \log \frac{\rho_k}{\rho^*} \right) \cdot \nabla u, \rho_k \right\rangle \\ &\quad - h^2 \beta^{-1} \left\langle \nabla \left(\frac{\mathcal{L}^* \rho_k}{\rho_k} \right) \cdot \nabla u, \rho_k \right\rangle + \frac{h^2}{2} \left\langle \Phi'_k{}^\top \nabla^2 u \Phi'_k, \rho_k \right\rangle + O(h^3), \end{aligned} \quad (133)$$

where $\Phi'_k = \nabla(\beta^{-1} \log(\rho_k/\rho^*))$.

Hessian term. As usual,

$$\Phi'_k{}^\top \nabla^2 u \Phi'_k = \sum_{i,j} \partial_{ij} u \Phi'_{k,i} \Phi'_{k,j},$$

and integrating by parts gives

$$\langle \Phi'_k{}^\top \nabla^2 u \Phi'_k, \rho_k \rangle = - \langle \nabla u, \rho_k \zeta_k \rangle,$$

where ζ_k is defined in the lemma. Substituting completes the proof. ■

Proof [Proof of Lemma 10.] Using the Taylor expansion of the log function and the expansion of ρ_{k+1} , we obtain

$$\begin{aligned} \log \rho_{k+1} &= \log \rho_k + h \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) - \frac{h^2}{2} \left(\mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \right)^2 \\ &\quad + h^2 \mathcal{D}_k^\beta \circ \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) + \frac{h^2}{2\rho_k} \nabla \cdot (\zeta \rho_k) + \mathcal{O}(h^3). \end{aligned}$$

Multiplying by ρ_{k+1} and expanding, we obtain

$$\begin{aligned} \rho_{k+1} \log \frac{\rho_{k+1}}{\rho^*} &= \rho_k \log \frac{\rho_k}{\rho^*} + \frac{h^2}{2} \rho_k \left(\mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \right)^2 \\ &\quad + \left[h \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) + h^2 \mathcal{D}_k^\beta \circ \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) + \frac{h^2}{2\rho_k} \nabla \cdot (\zeta \rho_k) \right] \rho_k \left(1 + \log \frac{\rho_k}{\rho^*} \right) + \mathcal{O}(h^3). \end{aligned}$$

Substituting into the KL formula gives

$$\begin{aligned} &D_{\text{KL}}(\rho_k \|\rho^*) - D_{\text{KL}}(\rho_{k+1} \|\rho^*) \\ &= -h \left[\int \left(\mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \rho_k + h \mathcal{D}_k^\beta \circ \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \rho_k - \frac{h}{2} \nabla \cdot (\zeta \rho_k) \right) \left(1 + \log \frac{\rho_k}{\rho^*} \right) dx \right] \\ &\quad - \frac{h^2}{2} \int \left(\mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \right)^2 \rho_k dx + \mathcal{O}(h^3). \end{aligned} \tag{134}$$

(1) *The $\mathcal{O}(h)$ term.* Using integration by parts and the definition of \mathcal{D}_k^β ,

$$- \int \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \left(1 + \log \frac{\rho_k}{\rho^*} \right) \rho_k dx = \beta^{-1} \int \|\nabla \log \frac{\rho_k}{\rho^*}\|^2 \rho_k dx = \beta^{-1} I(\rho_k \|\rho^*).$$

(2) *The ζ -term.* Recall $\tilde{\phi} = \beta^{-1} \log(\rho_k/\rho^*) + \mathcal{O}(h)$ and $\zeta = \beta^{-2} \nabla \cdot (\Psi \rho_k)$, where $\Psi = \nabla \log \frac{\rho_k}{\rho^*} (\nabla \log \frac{\rho_k}{\rho^*})^T$. We compute

$$\begin{aligned} \int \langle \zeta, \nabla \log \frac{\rho_k}{\rho^*} \rangle \rho_k dx &= \beta^{-2} \int \left\langle \nabla \log \frac{\rho_k}{\rho^*}, \nabla \cdot (\Psi \rho_k) \right\rangle dx \\ &= -\beta^{-2} \int \left\langle \nabla \log \frac{\rho_k}{\rho^*}, \nabla^2 \log \frac{\rho_k}{\rho^*} \nabla \log \frac{\rho_k}{\rho^*} \right\rangle \rho_k dx, \end{aligned}$$

where the last equality uses integration by parts.

(3) *The h^2 term from $\mathcal{D}_k^\beta \circ \mathcal{D}_k^\beta$.* Using integration by parts:

$$\begin{aligned} &\int \mathcal{D}_k^\beta \circ \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \left(1 + \log \frac{\rho_k}{\rho^*} \right) \rho_k dx \\ &= \int \beta^{-1} \mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \nabla \cdot \left(\nabla \log \frac{\rho_k}{\rho^*} \rho_k \right) dx = \int \left(\mathcal{D}_k^\beta \left(\log \frac{\rho_k}{\rho^*} \right) \right)^2 \rho_k dx. \end{aligned}$$

Putting (1)-(3) into (134) yields exactly the one-step KL expansion stated in (39). ■

Proof [Proof of Lemma 11] In the following proof, all time derivatives with respect to ρ_k are taken along the Fokker-Planck flow starting from the initial density ρ_k . Substituting the expression (40) into (39), we obtain

$$\begin{aligned}
 & \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) - \mathrm{D}_{\mathrm{KL}}(\rho_{k+1} \|\rho^*) \tag{135} \\
 & \geq h\beta^{-1} \mathcal{I}(\rho_k \|\rho^*) - \frac{3h^2}{4} \frac{d^2}{dt^2} \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) \\
 & \quad - \frac{h^2}{2} \beta^{-2} \left[\int \left\| \nabla^2 \log \frac{\rho_k}{\rho^*} \right\|_{\mathrm{F}}^2 \rho_k dx + \int \left\| \nabla \log \frac{\rho_k}{\rho^*} \right\|_2^4 \rho_k dx \right] + \mathcal{O}(h^3) \\
 & = h\beta^{-1} \mathcal{I}(\rho_k \|\rho^*) - h^2 \frac{d^2}{dt^2} \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) \\
 & \quad + \frac{h^2 \beta^{-1}}{2} \int \left\langle \nabla \log \frac{\rho_k}{\rho^*}, \nabla^2 V \nabla \log \frac{\rho_k}{\rho^*} \right\rangle \rho_k dx - \frac{h^2 \beta^{-2}}{2} \int \left\| \nabla \log \frac{\rho_k}{\rho^*} \right\|_2^4 \rho_k dx + \mathcal{O}(h^3),
 \end{aligned}$$

where we used the inequality

$$\|A\|_{\mathrm{F}}^2 + \|x\|_2^4 \geq -2 \sum_{ij} x_i a_{ij} x_j = -2x^T A x.$$

For the fourth-order term, Lemma 27 shows that, along the continuous Fokker-Planck flow,

$$\frac{\partial}{\partial t} \int \left\| \nabla \log \frac{\rho_t}{\rho^*} \right\|_2^4 \rho_t dx \leq -4\alpha \int \left\| \nabla \log \frac{\rho_t}{\rho^*} \right\|_2^4 \rho_t dx,$$

which implies exponential decay.

For the second-order time derivative, since ρ_k is uniformly $C^{3,1}$, the KL functional along the flow admits a locally uniform third derivative. Hence,

$$\begin{aligned}
 \frac{d^2}{dt^2} \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) &= \frac{1}{h} \left[\frac{d}{dt} \mathrm{D}_{\mathrm{KL}}(\rho_{t_k+h} \|\rho^*) - \frac{d}{dt} \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) \right] + \mathcal{O}(h) \\
 &= \frac{\beta^{-1}}{h} [\mathcal{I}(\rho_k \|\rho^*) - \mathcal{I}(\rho_{k+1} \|\rho^*)] + \mathcal{O}(h), \tag{136}
 \end{aligned}$$

where we used

$$\frac{d}{dt} \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) = -\beta^{-1} \mathcal{I}(\rho_k \|\rho^*), \quad \mathcal{I}(\rho_{k+1} \|\rho^*) = \mathcal{I}(\rho_{t_k+h} \|\rho^*) + \mathcal{O}(h^2).$$

Substituting (130) and (136) into (135), we obtain

$$\begin{aligned}
 & \mathrm{D}_{\mathrm{KL}}(\rho_{k+1} \|\rho^*) + \frac{h}{\beta} \mathcal{I}(\rho_{k+1} \|\rho^*) \tag{137} \\
 & \leq \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) - \frac{h}{\beta} \left(1 - \alpha \frac{h}{2} \right) \mathcal{I}(\rho_k \|\rho^*) + \frac{h^2}{2} M_0 \exp(-4\alpha t_k) + \mathcal{O}(h^3),
 \end{aligned}$$

where

$$M_0 := \beta^{-2} \int \left\| \nabla \log \frac{\rho_0}{\rho^*} \right\|_2^4 \rho_0 dx,$$

and we used $\nabla^2 V \succeq \alpha I$.

Using the PL inequality (6), (137) implies

$$(1 + 2\alpha h) \mathrm{D}_{\mathrm{KL}}(\rho_{k+1} \|\rho^*) \leq (1 - \alpha^2 h^2) \mathrm{D}_{\mathrm{KL}}(\rho_k \|\rho^*) + \frac{h^2}{2} M_0 \exp(-4\alpha h k) + \mathcal{O}(h^3),$$

which is equivalent to

$$\begin{aligned}
 D_{\text{KL}}(\rho_{k+1} \|\rho^*) &\leq \frac{1 - \alpha^2 h^2}{1 + 2\alpha h} D_{\text{KL}}(\rho_k \|\rho^*) + \frac{h^2}{2(1 + 2\alpha h)} M_0 \exp(-4\alpha h k) + \mathcal{O}(h^3) \\
 &= \left(1 - h\alpha \frac{2 + \alpha h}{1 + 2\alpha h}\right) D_{\text{KL}}(\rho_k \|\rho^*) + \frac{h^2}{2} M_0 \exp(-4\alpha h k) + \mathcal{O}(h^3) \\
 &= [1 - 2\alpha h + 3\alpha^2 h^2] D_{\text{KL}}(\rho_k \|\rho^*) + \frac{h^2}{2} M_0 \exp(-4\alpha h k) + \mathcal{O}(h^3).
 \end{aligned}$$

■

The next lemma is used in the proof of Theorem 12 to derive the exponential decay of KL divergence with a bias term.

Lemma 29 *For a sequence that satisfies*

$$a_{k+1} \leq (1 - c_1 h) a_k + h^2 c_2 e^{-c_3 k h} + \mathcal{O}(h^3),$$

where $h, c_i > 0$ and $c_1 h < 1$, we have for all $k \geq 1$

$$a_k \leq (1 - c_1 h)^k a_0 + h^2 c_2 k \max\{(1 - c_1 h)^{k-1}, e^{-c_3(k-1)h}\} + \mathcal{O}(h^3).$$

In particular, for all $k \geq 1$,

$$a_k \leq e^{-c_1 k h} a_0 + h^2 c_2 k e^{-\min\{c_1, c_3\}(k-1)h} + \mathcal{O}(h^3).$$

Proof We first show the sequence satisfies the following inductive relationship

$$a_{k+1} \leq (1 - c_1 h)^{k+1} a_0 + h^2 c_2 \sum_{j=0}^k (1 - c_1 h)^j \exp(-c_3(k-j)h) + \mathcal{O}(h^3).$$

It is true for $k = 0$. And for the case $k \geq 1$, we have

$$\begin{aligned}
 a_{k+1} &\leq (1 - c_1 h) a_k + h^2 c_2 \exp(-c_3 k h) + \mathcal{O}(h^3) \\
 &\leq (1 - c_1 h)^{k+1} a_0 + h^2 c_2 \left[\sum_{j=0}^{k-1} (1 - c_1 h)^j \exp(-c_3(k-1-j)h) + \exp(-c_3 k h) \right] + \mathcal{O}(h^3) \\
 &= (1 - c_1 h)^{k+1} a_0 + h^2 c_2 \sum_{j=0}^k (1 - c_1 h)^j \exp(-c_3(k-j)h) + \mathcal{O}(h^3).
 \end{aligned}$$

For the term under summation, set

$$b_j := (1 - c_1 h)^j e^{-c_3(k-j)h} = e^{-c_3 k h} \left((1 - c_1 h) e^{c_3 h} \right)^j.$$

Since $b_j \geq 0$, we may bound

$$\sum_{j=0}^k b_j \leq (k+1) \max_{0 \leq j \leq k} b_j.$$

Because b_j is geometric in j , its maximum on $\{0, \dots, k\}$ is attained at an endpoint, hence

$$\max_{0 \leq j \leq k} b_j = \max\{b_0, b_k\} = \max\{e^{-c_3 k h}, (1 - c_1 h)^k\}.$$

Therefore

$$\sum_{j=0}^k (1 - c_1 h)^j e^{-c_3(k-j)h} \leq (k+1) \max\{(1 - c_1 h)^k, e^{-c_3 k h}\}.$$

Substituting this into the inductive estimate yields

$$a_{k+1} \leq (1 - c_1 h)^{k+1} a_0 + h^2 c_2 (k+1) \max\{(1 - c_1 h)^k, e^{-c_3 k h}\} + \mathcal{O}(h^3).$$

Replacing $k+1$ by k gives the claim. ■

Appendix C. Laplace Approximation to the Kernel Formula

In this section, we show supplementary results on the pointwise approximation provided by the regularized Wasserstein proximal operator (17) to the evolution of the Fokker-Planck equation (5) under the convexity condition when h is small by utilizing the Laplace method. Although this analysis is not required for our main convergence result, it offers additional insight into how the kernel formula approximates the Fokker-Planck dynamics at the pointwise level. In particular, it clarifies how the scheme behaves in settings where pointwise accuracy is essential—such as optimization procedures involving a decaying inverse-temperature schedule.

Our strategy is to employ the **Laplace method** up to two terms (Bleistein and Handelsman, 1986)

$$\int_{\mathbb{R}^d} g(x) \exp\left(-\frac{f(x)}{h}\right) dx = (2\pi h)^{d/2} \exp\left(-\frac{f(x^*)}{h}\right) \left[\frac{g(x^*)}{|\nabla^2 f(x^*)|^{1/2}} + \frac{h}{2} H_1(x^*) + \mathcal{O}(h^2) \right], \quad (138)$$

where $x^* = \arg \min_x f(x)$ is the unique minimizer of f , $|\nabla^2 f|$ is the determinant of the Hessian matrix of f , and H_1 is the first-order term of the expansion. The explicit form for H_1 is given below, with its detailed derivation available in Section 8.3 of (Bleistein and Handelsman, 1986). Writing $f_p = \frac{\partial}{\partial x_p} f$ for $p = 1, \dots, d$, we have

$$H_1(x^*) = -|\nabla^2 f(x^*)|^{-1/2} \left\{ -f_{srq} B_{sq} B_{rp} g_p + \text{Tr}(CB) \right. \\ \left. + g \left[f_{pqr} f_{stu} \left(\frac{1}{4} B_{ps} B_{qr} B_{tu} + \frac{1}{6} B_{ps} B_{qt} B_{ru} \right) - \frac{1}{4} f_{pqr} B_{pr} B_{qs} \right] \right\}_{x=x^*}, \quad (139)$$

where we use the summation convention to sum over all indices from 1 to d . The matrices in the expression are defined as

$$C = \{g_{pq}\}, \quad B = \{B_{pq}\}, \quad \sum B_{pq} f_{qr}(x^*) = \delta_{pr}.$$

We can then apply this approximation to the kernel formula of the regularized Wasserstein proximal operator in (17) to obtain the asymptotic expansion when h is small. In Theorem 31, we will show that $\mathcal{K}_V^h \rho_0$, computed from the kernel formula satisfies

$$\mathcal{K}_V^h \rho_0(x) = \rho_0(x) + h \left. \frac{\partial \rho_0}{\partial t} \right|_{t=t_0}(x) + \mathcal{O}(h^2),$$

where ρ_0 satisfies the Fokker-Planck equation at time t_0 . The proof relies on exploring the explicit representations of the first two terms in the approximation (138), the Taylor expansion of the potential function V (by the bounded derivative assumption), and the uniqueness of the minimizer x^* in (139), which follows from the log-concavity assumption.

C.1 Approximation to the normalization term

Firstly, we derive the approximation result for the denominator in the kernel formula (17), as shown in the following lemma.

Lemma 30 *For any y , assuming $h\Delta V(s_y) \leq 1$, we have*

$$\begin{aligned} & \left(\frac{\beta}{4\pi h}\right)^{d/2} \int_{\mathbb{R}^d} \exp\left(-\frac{\beta}{2}\left(V(z) + \frac{\|z-y\|_2^2}{2h}\right)\right) dz \\ &= \frac{1}{1+h/2\Delta V(\tilde{s}_y)} \exp\left(-\frac{\beta}{2}\left(V(\tilde{s}_y) + \frac{\|\tilde{s}_y-y\|_2^2}{2h}\right)\right) + \mathcal{O}(h^2), \end{aligned} \quad (140)$$

where

$$s_y = y - h\nabla V(s_y), \quad \tilde{s}_y = y - h\nabla V(y).$$

Proof For the normalization term in (17) and fixed y , we have

$$\begin{aligned} & \left(\frac{\beta}{4\pi h}\right)^{d/2} \int_{\mathbb{R}^d} \exp\left(-\frac{\beta}{2}\left(V(z) + \frac{\|z-y\|_2^2}{2h}\right)\right) dz \\ &= \left(\frac{\beta}{4\pi h}\right)^{d/2} \int_{\mathbb{R}^d} \exp\left(-\frac{\beta}{2h}\left(hV(z) + \frac{\|z-y\|_2^2}{2}\right)\right) dz \\ &= \exp\left(-\frac{\beta}{2h}\left(hV(s_y) + \frac{\|s_y-y\|_2^2}{2}\right)\right) \left[\frac{1}{|1+h\nabla^2 V(s_y)|^{1/2}} + \frac{h}{\beta} H_1(s_y) + \mathcal{O}\left(\frac{2h}{\beta}\right)^2 \right], \end{aligned} \quad (141)$$

where s_y is defined as the arg min of the exponent

$$s_y := \arg \min_z \left\{ hV(z) + \frac{\|y-z\|_2^2}{2} \right\}.$$

Next, we verify that $H_1(s_y)$ is of order $\mathcal{O}(h)$. Concerning the general form of Laplace method (138), we note

$$g(z) = 1, \quad f(z) = hV(z) + \frac{\|z-y\|_2^2}{2}, \quad f_{pq}(z) = h \frac{\partial^2 V}{\partial z_p \partial z_q}(z) + \delta_{pq}, \quad f_{pqr}(z) = h \frac{\partial^3 V}{\partial z_p \partial z_q \partial z_r}(z),$$

where the quadratic term becomes a constant after taking the second-order partial derivative and vanishes after taking the third-order partial derivative. Given that the B matrix in (139) is the inverse of a diagonal matrix plus $h\nabla^2 V$, the magnitude of the diagonal entries of B is $\mathcal{O}(1)$.

Thus, looking at the expression of $H_1(s_y)$, we confirm that

$$H_1(s_y) = \mathcal{O}(h),$$

as all terms in (139) are of the order $\mathcal{O}(h)$. Then the approximation in (141) is simplified to

$$\begin{aligned} & \left(\frac{\beta}{4\pi h}\right)^{d/2} \int_{\mathbb{R}^d} \exp\left(-\frac{\beta}{2}\left(V(z) + \frac{\|z-y\|_2^2}{2h}\right)\right) dz \\ &= \frac{1}{|1+h\nabla^2 V(s_y)|^{1/2}} \exp\left(-\frac{\beta}{2h}\left(hV(s_y) + \frac{\|s_y-y\|_2^2}{2}\right)\right) + \mathcal{O}(h^2). \end{aligned} \quad (142)$$

Next, we compute s_y more explicitly. Since $V(z)$ is convex, so is $hV(z) + \|z - y\|_2^2/2$ for any y and h with respect to z . This ensures s_y is unique. By the first-order optimality condition, s_y satisfies

$$h\nabla V(s_y) + s_y = y.$$

To further simplify the expression and compute the integral explicitly, we define an approximation for s_y as

$$\tilde{s}_y := y - h\nabla V(y). \quad (143)$$

By definition of \tilde{s}_y , we have

$$\|\tilde{s}_y - s_y\|_2 = h\|\nabla V(s_y) - \nabla V(y)\|_2 \leq Lh\|s_y - y\|_2 = Lh^2\|\nabla V(s_y)\|_2 = \mathcal{O}(h^2).$$

The inequality holds when ∇V is Lipschitz continuous with constant L which can be implied by $\nabla^2 V$ is bounded by L along the line segment connecting y and s_y .

Finally, for the determinant of the Hessian matrix, we apply the Taylor expansion of the determinant operator and the square root function to obtain

$$|I + h\nabla^2 V(s_y)| = 1 + h\Delta V(s_y) + \mathcal{O}(h^2), \quad (144)$$

$$|I + h\nabla^2 V(s_y)|^{1/2} = 1 + \frac{h}{2}\Delta V(s_y) + \mathcal{O}(h^2),$$

which converges when $h\Delta V(s_y) \leq 1$.

We arrive at the desired result by combining equations (141) to (144). \blacksquare

C.2 Approximation to the kernel formula

Letting

$$V_0(x) = -\beta^{-1} \log \rho_0(x),$$

we follow a similar approach as in the proof of Lemma 30 to derive the asymptotic approximation of $\mathcal{K}_V^h \rho_0$.

Theorem 31 *For fixed x , assume that ΔV_0 is bounded above on the line segment connecting x and r_x , where r_x satisfies $r_x = x + h\nabla(V - 2V_0)(r_x)$. Moreover, when $h\Delta V(r_x) \leq 1$ and $h\Delta V_0(r_x) \leq 1$, we have*

$$\mathcal{K}_V^h \rho_0(x) = \rho_0(x) [1 - \beta h \nabla(V - V_0) \cdot \nabla V_0(x) + h\Delta(V - V_0)(x)] + \mathcal{O}(h^2). \quad (145)$$

If ρ_0 satisfies the Fokker-Planck equation at time t_0 , we have

$$\mathcal{K}_V^h \rho_0(x) = \rho_0 + h\beta^{-1} \nabla \cdot \left(\nabla \log \frac{\rho_0}{\rho_0^*} \rho_0 \right) + \mathcal{O}(h^2) = \rho_0 + h \frac{\partial \rho}{\partial t} \Big|_{t=t_0} + \mathcal{O}(h^2). \quad (146)$$

Proof Substituting (140) into the expression for $\mathcal{K}_V^h \rho_0(x)$ in (17), we arrive

$$\begin{aligned} \mathcal{K}_V^h \rho_0(x) &= \exp\left(-\frac{\beta}{2}V(x)\right) \int_{\mathbb{R}^d} \frac{\exp\left(-\beta \frac{\|x-y\|_2^2}{4h}\right)}{\int_{\mathbb{R}^d} \exp\left[-\frac{\beta}{2}\left(V(z) + \frac{\|z-y\|_2^2}{2h}\right)\right] dz} \rho_0(y) dy \\ &= \exp\left(-\frac{\beta}{2}V(x)\right) \int_{\mathbb{R}^d} \frac{1 + \frac{h}{2}\Delta V(\tilde{s}_y)}{(4\pi h/\beta)^{d/2}} \exp\left[-\frac{\beta}{2}\left(\frac{\|x-y\|_2^2}{2h} + 2V_0(y) - V(\tilde{s}_y) - \frac{\|y-\tilde{s}_y\|_2^2}{2h}\right)\right] dy + \mathcal{O}(h^2). \end{aligned} \quad (147)$$

For fixed x , to apply the approximation with the Laplace method as in (138), we let

$$f(y) = \frac{\|x-y\|_2^2}{2} + 2hV_0(y) - hV(\tilde{s}_y) - \frac{\|y-\tilde{s}_y\|_2^2}{2}, \quad g(y) = 1. \quad (148)$$

Then we may write the minimizer for $f(y)$ which is a function of x as

$$r_x = \arg \min_y f(y) = \arg \min_y \left\{ \frac{\|x - y\|_2^2}{2} + 2hV_0(y) - hV(\tilde{s}_y) - \frac{\|y - \tilde{s}_y\|_2^2}{2} \right\}.$$

Similar to the derivation for Lemma 30, the first-order optimality condition leads to

$$-(r_x - x) - 2h\nabla V_0(r_x) + \left[h \frac{\partial \tilde{s}_{r_x}}{\partial r_x} \nabla V(\tilde{s}_{r_x}) + \left(1 - \frac{\partial \tilde{s}_{r_x}}{\partial r_x} \right) (r_x - \tilde{s}_{r_x}) \right] = 0,$$

where $\frac{\partial \tilde{s}_{r_x}}{\partial r_x}$ is the Jacobian of \tilde{s}_{r_x} .

To simplify the expression for r_x , by definition of \tilde{s}_{r_x} in (143) and replacing y with r_x , we have

$$r_x - \tilde{s}_{r_x} = h\nabla V(r_x).$$

This leads to

$$-(r_x - x) - 2h\nabla V_0(r_x) + h\nabla V(r_x) = 0 \quad \Rightarrow \quad r_x = x + h\nabla(V - 2V_0)(r_x),$$

as the term involves $\frac{\partial \tilde{s}_{r_x}}{\partial r_x}$ cancel out.

Then as a similar argument as in the proof of Lemma 30, we can define a linearized approximate solution to r_x as

$$\tilde{r}_x = x + h\nabla(V - 2V_0)(x), \quad (149)$$

where $|\tilde{r}_x - r_x| = \mathcal{O}(h^2)$ under the assumption $\nabla^2(V_0 - V)$ is bounded on the line segment connection x and r_x .

Now, note that the Hessian of $f(r_x)$ in (148) will be

$$\nabla^2 f(r_x) = 1 + h\nabla^2(V_0 - V)(r_x) - \frac{h^2}{2}\nabla^3 V(r_x) = 1 + h\nabla^2(V_0 - V)(r_x) + \mathcal{O}(h^2).$$

Using the Taylor expansion for the determinant function in (144) and the definition of \tilde{r}_x , we are ready to apply the Laplace method to $\mathcal{K}_V^h \rho_0$ in (147) to get

$$\begin{aligned} \mathcal{K}_V^h \rho_0(x)(x) &= \exp\left(-\frac{\beta}{2}V(x)\right) \int_{\mathbb{R}^d} \frac{\exp\left(-\beta\frac{\|x-y\|_2^2}{4h}\right)}{\int_{\mathbb{R}^d} \exp\left[-\frac{\beta}{2}\left(V(z) + \frac{\|z-y\|_2^2}{2h}\right)\right] dz} \rho_0(y) dy \\ &= \frac{\exp\left(-\frac{\beta}{2}V(x)\right) [1 + \frac{h}{2}\Delta V(\tilde{s}_{r_x})]}{|1 + h\nabla^2(V_0 - V)(r_x) + \mathcal{O}(h^2)|^{1/2}} \exp\left[-\frac{\beta}{2}\left(\frac{\|x-r_x\|_2^2}{2h} - \frac{\|r_x - \tilde{s}_{r_x}\|_2^2}{2h} + 2V_0(r_x) - V(\tilde{s}_{r_x})\right)\right] + \mathcal{O}(h^2) \\ &= \frac{\exp\left(-\frac{\beta}{2}V(x)\right) [1 + \frac{h}{2}\Delta V(\tilde{s}_{r_x})]}{1 + \frac{h}{2}\Delta(2V_0 - V)(\tilde{r}_x)} \exp\left[-\frac{\beta}{2}\left(\frac{\|x-\tilde{r}_x\|_2^2}{2h} - \frac{\|\tilde{r}_x - \tilde{s}_{\tilde{r}_x}\|_2^2}{2h} + 2V_0(\tilde{r}_x) - V(\tilde{s}_{\tilde{r}_x})\right)\right] + \mathcal{O}(h^2). \end{aligned} \quad (150)$$

For the exponent in the second exponential function in the last line of (150), using the definition of \tilde{r}_x and \tilde{s}_x and omitting the factor $-\beta/2$ for clarity, it can be simplified as

$$\begin{aligned} &2V_0(\tilde{r}_x) - V(\tilde{s}_{\tilde{r}_x}) + \frac{h}{2}\|\nabla(V - 2V_0)(x)\|_2^2 - \frac{h}{2}\|\nabla V(\tilde{r}_x)\|_2^2 + \mathcal{O}(h^2) \\ &= 2V_0(x + h\nabla(V - 2V_0)(x)) - V(x - 2h\nabla V_0(x)) + \frac{h}{2}\|\nabla(V - 2V_0)(x)\|_2^2 - \frac{h}{2}\|\nabla V(x)\|_2^2 + \mathcal{O}(h^2) \\ &= 2V_0(x) + 2h\nabla V_0 \cdot \nabla(V - 2V_0)(x) - V(x) + 2h\nabla V \cdot \nabla V_0(x) \\ &\quad + \frac{h}{2}\|\nabla(V - 2V_0)(x)\|_2^2 - \frac{h}{2}\|\nabla V(x)\|_2^2 + \mathcal{O}(h^2) \\ &= 2V_0(x) - V(x) + 2h\nabla(V - V_0) \cdot \nabla V_0(x) + \mathcal{O}(h^2), \end{aligned}$$

where we have used the Taylor expansion on V_0 and V , and the relation

$$\tilde{s}_{\tilde{r}_x} = \tilde{r}_x - h\nabla V(\tilde{r}_x) = x + h\nabla(V - 2V_0)(x) - h\nabla V(x) + \mathcal{O}(h^2) = x - 2h\nabla V_0(x) + \mathcal{O}(h^2).$$

Lastly, for the coefficient before the exponential term in (150), with the help of the Neumann series, we derive

$$\frac{1 + \frac{h}{2}\Delta V(\tilde{s}_{\tilde{r}_x})}{1 + \frac{h}{2}\Delta(2V_0 - V)(\tilde{r}_x)} = \left[1 + \frac{h}{2}\Delta V(x)\right] \left[1 - \frac{h}{2}\Delta(2V_0 - V)(x)\right] + \mathcal{O}(h^2) = 1 + h\Delta(V - V_0)(x) + \mathcal{O}(h^2),$$

under the assumption that $|h\Delta(2V_0 - V)(x)| \leq 2$. Combining the above expression, we arrive at the desired result in Theorem 31. \blacksquare

We observe that it is sufficient for the Hessian of V_0 to be bounded only within a h -neighborhood around the sampling point x , rather than requiring global boundedness, which can be challenging to verify in practice. Additionally, since most sampling points are typically situated near the high-density regions of ρ^* , it is adequate for V_0 to exhibit reasonable smoothness specifically within these high-density areas.

References

- Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50:5–43, 2003.
- Dominique Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX 1983/84: Proceedings*, pages 177–206. Springer, 2006.
- Krishna Balasubramanian, Sinho Chewi, Murat A Erdogdu, Adil Salim, and Shunshi Zhang. Towards a theory of non-log-concave sampling: first-order stationarity guarantees for Langevin Monte Carlo. In *Conference on Learning Theory*, pages 2896–2923. PMLR, 2022.
- Claude J. P. Bélisle, H. Edwin Romeijn, and Robert L. Smith. Hit-and-run algorithms for generating multivariate distributions. *Mathematics of Operations Research*, 18(2):255–266, 1993.
- Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- Espen Bernton. Langevin Monte Carlo and JKO splitting. In *Conference on Learning Theory*, pages 1777–1798. PMLR, 2018.
- Michael Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*, 2017.
- N. Bleistein and R.A. Handelsman. *Asymptotic Expansions of Integrals*. Dover Books on Mathematics Series. Dover Publications, 1986.
- José A Carrillo, Franca Hoffmann, Andrew M Stuart, and Urbain Vaes. Consensus-based sampling. *Studies in Applied Mathematics*, 148(3):1069–1140, 2022.
- José Antonio Carrillo, Katy Craig, and Francesco S Patacchini. A blob method for diffusion. *Calculus of Variations and Partial Differential Equations*, 58:1–53, 2019.
- Hongrui Chen, Holden Lee, and Jianfeng Lu. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *International Conference on Machine Learning*, pages 4735–4763. PMLR, 2023.

- Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R. Zhang. Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. *arXiv preprint arXiv:2209.11215*, 2022.
- Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ODE is provably fast. In *Advances in Neural Information Processing Systems*, volume 36, 2024.
- Sinho Chewi, Murat A. Erdogdu, Mufan Li, Ruoqi Shen, and Matthew S. Zhang. Analysis of Langevin Monte Carlo from Poincaré to log-Sobolev. *Foundations of Computational Mathematics*, 25(4):1–51, 2025.
- Edward Brian Davies. *Heat kernels and spectral theory*. Cambridge University Press, 1989.
- Michael Ziyang Diao, Krishna Balasubramanian, Sinho Chewi, and Adil Salim. Forward-backward Gaussian variational inference via JKO in the Bures-Wasserstein space. In *International Conference on Machine Learning*, pages 7960–7991. PMLR, 2023.
- Alain Durmus and Eric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *arXiv preprint arXiv:1605.01559*, 2018.
- Raaz Dwivedi, Yuansi Chen, Martin J Wainwright, and Bin Yu. Log-concave sampling: Metropolis-Hastings algorithms are fast. *Journal of Machine Learning Research*, 20(183):1–42, 2019.
- Lawrence C Evans. *Partial Differential Equations*, volume 19. American Mathematical Society, 2022.
- Xuefeng Gao and Lingjiong Zhu. Convergence analysis for general probability flow ODEs of diffusion models in Wasserstein distances. *arXiv preprint arXiv:2401.17958*, 2024.
- Alfredo Garbuno-Inigo, Franca Hoffmann, Wuchen Li, and Andrew M. Stuart. Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler. *SIAM Journal on Applied Dynamical Systems*, 2020. doi: 10.1137/19M1251655.
- Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.
- Fuqun Han, Stanley Osher, and Wuchen Li. Tensor train based sampling algorithms for approximating regularized Wasserstein proximal operators. *SIAM/ASA Journal on Uncertainty Quantification*, 13(2):775–804, 2025.
- Liam Hodgkinson, Robert Salomone, and Fred Roosta. Implicit Langevin algorithms for sampling from log-concave densities. *Journal of Machine Learning Research*, 22(136):1–30, 2021.
- Daniel Zhengyu Huang, Jiaoyang Huang, and Zhengjiang Lin. Convergence analysis of probability flow ODE for score-based generative models. *arXiv preprint arXiv:2404.09730*, 2024.
- Wenhua Jiang and Cun-Hui Zhang. General maximum likelihood empirical Bayes estimation of normal means. *The Annals of Statistics*, 37(4):1647–1684, 2009.
- Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.
- Ben Leimkuhler and Charles Matthews. *Molecular Dynamics*, volume 39 of *Interdisciplinary Applied Mathematics*. Springer, 2015.

- Wuchen Li, Siting Liu, and Stanley Osher. A kernel formula for regularized Wasserstein proximal operators. *Research in Mathematical Science*, 10:43, 2023.
- Jiaming Liang and Yongxin Chen. A proximal algorithm for sampling from non-smooth potentials. In *2022 Winter Simulation Conference (WSC)*. IEEE, 2022.
- Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose Bayesian inference algorithm. In *Advances in Neural Information Processing Systems*, volume 29, 2016.
- Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, volume 35, pages 5775–5787, 2022.
- Yi-An Ma, Yuansi Chen, Chi Jin, Nicolas Flammarion, and Michael I. Jordan. Sampling can be faster than optimization. *Proceedings of the National Academy of Sciences of the United States of America*, 116(42):20881–20885, 2019.
- Yi-An Ma, Niladri S. Chatterji, Xiang Cheng, Nicolas Flammarion, Peter L. Bartlett, and Michael I. Jordan. Is there an analog of Nesterov acceleration for gradient-based MCMC? *Bernoulli*, 27(3):1942–1992, 2021.
- K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24(1):101–121, 1996.
- Felix Otto and Cédric Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400, 2000.
- Giorgio Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180(3):378–384, 1981.
- Sebastian Reich and Simon Weissmann. Fokker–Planck particle systems for Bayesian inference: Computational approaches. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):446–482, 2021.
- Adil Salim, Anna Korba, and Giulia Luise. The Wasserstein proximal gradient algorithm. In *Advances in Neural Information Processing Systems*, volume 33, pages 12356–12366, 2020.
- Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32, page 11895–11907, 2019.
- Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.
- Hong Ye Tan, Stanley Osher, and Wuchen Li. Noise-free sampling algorithms via regularized Wasserstein proximals. *Research in Mathematical Science*, 11:65, 2024.
- Santosh Vempala and Andre Wibisono. Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- Cédric Villani. *Hypocoercivity*, volume 202. American Mathematical Society, 2009.
- Yifei Wang and Wuchen Li. Accelerated information gradient flow. *Journal of Scientific Computing*, 90:1–47, 2022.
- Andre Wibisono. Proximal Langevin algorithm: Rapid convergence under isoperimetry. *arXiv preprint arXiv:1911.01469*, 2019.

- Andre Wibisono, Yihong Wu, and Kaylee Yingxi Yang. Optimal score estimation via empirical Bayes smoothing. *arXiv preprint arXiv:2402.07747*, 2024.
- Tatiana Xifara, Chris Sherlock, Samuel Livingstone, Simon Byrne, and Mark Girolami. Langevin diffusions and the Metropolis-adjusted Langevin algorithm. *Statistics & Probability Letters*, 91: 14–19, 2014.
- Kaylee Yingxi Yang and Andre Wibisono. Convergence in KL and Rényi divergence of the unadjusted Langevin algorithm using estimated score. In *NeurIPS 2022 Workshop on Score-Based Methods*, 2022.
- Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. *arXiv preprint arXiv:2204.13902*, 2022.
- Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. UniPC: A unified predictor-corrector framework for fast sampling of diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, 2024.