# FINITE-PARTICLE CONVERGENCE RATES FOR CONSERVATIVE AND NON-CONSERVATIVE DRIFTING MODELS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

We analyze finite-particle drifting models for one-step generative modeling. For a conservative velocity given by the difference of the kernel-smoothed data and model scores, a joint-entropy identity yields continuous-time bounds for the smoothed Fisher discrepancy and the squared particle velocity. The finite-particle correction involves reciprocal kernel density estimates. We give local-occupancy conditions and exact expectation bounds, and show that these expectations diverge for full-support initial densities in a shrinking-bandwidth regime. Keeping the bandwidth dependence of the quadrature constants explicit yields a conditional root residual-velocity rate of $N^{-1/(d+4)}$ under uniform regularity, with a corresponding rate under weaker quadrature growth conditions. We also analyze the original displacement field with the exact Laplace kernel. A companion kernel gives a weighted coercivity estimate that accommodates the unbounded empirical scale and the kernel's nondifferentiability. Localization controls the particle velocity, while relative scale alignment absorbs the mismatch into dissipation. Both analyses quantify the size of an additional drift correction under explicit regularity and stability assumptions.

## 1 INTRODUCTION

Drifting models (Deng et al., 2026) learn one-step generators by moving the model distribution during training, avoiding a long iterative sampler at inference time. We study their finite-particle behavior. Given the data law $\nu$, a model law $\mu$, and a positive kernel $K_h$, we define

$$M_{\alpha,h}(z) := \frac{\int (y-z)K_h(z-y)\,\alpha(\mathrm{d}y)}{\int K_h(z-y)\,\alpha(\mathrm{d}y)}, \qquad u^{\mathrm{disp}}_{\nu,\mu,h} := M_{\nu,h} - M_{\mu,h}. \tag{1}$$

The data term attracts a point toward the data, whereas the model term repels it from regions represented by the model. Training regresses a generated point $z$ toward the frozen target $z + \eta u^{\mathrm{disp}}_{\nu,\mu,h}(z)$. We show that exact regression on a fixed latent batch gives a particle Euler update (see Appendix H for more details), and analyze its small-step continuous-time limit. Specifically, we analyze how batch size and kernel bandwidth control the residual drift.

The displacement field is generally non-conservative because its normalization varies with position (Franz et al., 2026). We analyze it alongside a conservative alternative based on kernel density estimator (KDE) scores. With $\rho_{\alpha,h} = K_h * \alpha$ and $s_{\alpha,h} = \nabla \log \rho_{\alpha,h}$, we define

$$b_{\nu,\mu,h} := s_{\nu,h} - s_{\mu,h} = \nabla\{\log \rho_{\nu,h} - \log \rho_{\mu,h}\}. \tag{2}$$

For the Gaussian kernel with covariance $h^2 I_d$, $M_{\alpha,h} = h^2 s_{\alpha,h}$, so the dynamics agree after rescaling time. For other kernels, displacement averaging and KDE differentiation produce distinct fields. We write $b$ for a conservative velocity and $u$ for a displacement velocity.

Figure 1 illustrates the fields from a common initial configuration. We show that the comparison is more nuanced than "conservative is better." Although the conservative field has zero spatial circulation, its potential evolves with the particle law, so conservativeness alone does not imply monotonicity along trajectories. We also identify an exact reciprocal self-interaction correction generated



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

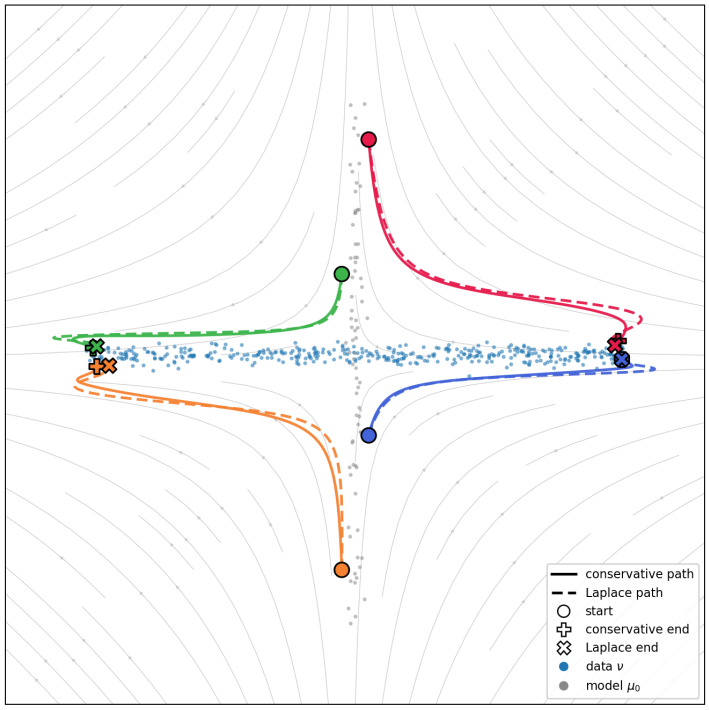



Figure 1: Conservative Gaussian drift (solid) and displacement-based Laplace drift (dashed) from the same initial cloud. The data occupy a thin horizontal cluster (blue), and the initial model occupies a thin vertical cluster (grey). Four trajectories are shown from $N = 80$ particles at bandwidth $h = 0.55$. The conservative field is a spatial gradient at each time; the Laplace field can have nonzero circulation. Curvature of the paths alone does not distinguish these properties. Appendix A gives the accompanying discussion and a Helmholtz–Hodge decomposition of the two fields.

by the finite-particle KDE. The self-masked Laplace field avoids this correction, but its position-dependent normalization introduces circulation and scale mismatch. We derive $u_x = a_{x,h} b_x^{\#} + e_x$, expressing the field as a positively preconditioned conservative score mismatch plus an error. We show that scale derivatives govern circulation, whereas data–model scale differences determine $e_x$; these effects are related but distinct. Thus, the fields face different finite-particle obstructions, and neither uniformly dominates.

For conservative drifting, we prove a joint-entropy identity and quadrature comparisons controlling the smoothed Fisher discrepancy and particle velocity. Under uniform quadrature control, we obtain the conditional root rate $N^{-1/(d+4)}$ and, for growing quadrature constants, the modified rate in Section 3. We control reciprocal KDEs via local occupancy and an exact binomial identity, and show that full-support initialization makes their expectation diverge in the shrinking-bandwidth regime. For the exact Laplace field, we construct a companion kernel yielding weighted coercivity despite the cusp and unbounded scale. We combine localization with relative scale alignment to control particle velocity and absorb the mismatch with coefficient $1 - \sqrt{\theta_h}$.

In Section 6, we verify that these terms are active. We show that omitting the conservative correction accounts for the full entropy-identity defect, and identify rare low-density particles as the source of full-support growth. For Laplace drifting, we show that $\theta = \mathcal{D}_h^{\mathrm{Lap}} / \mathcal{W}_h^{\mathrm{Lap}}$ organizes a clear phase diagram. The relative bound certifies a median 93.3% of observed dissipation, compared with 83.2% for the absolute bound, and remains informative along simulated flows. Full proofs appear in the appendices.

**Related work.** Drifting (Deng et al., 2026) is related to score-difference flows (Weber, 2023) and to population gradient-flow interpretations of kernel-smoothed divergences (Cao et al., 2026; Lai et al., 2026; Turan & Ovsjanikov, 2026; Gretton et al., 2026). Concurrent work also intro-

duces kernel-gradient drifting (Esteban-Casadevall et al., 2026). Our contribution is a finite-particle entropy-rate analysis, including reciprocal-KDE corrections and a separate treatment of the Laplace displacement field. The sharp-normalization construction of Franz et al. (2026) produces a conservative field by scoring a companion KDE. Here we use that construction to analyze the original Laplace normalization. Identifiability and stability of companion kernels are studied by Lee & Chun (2026).

Our joint-entropy argument builds on finite-particle analyses of Stein variational gradient descent (SVGD) (Shi & Mackey, 2023; Banerjee et al., 2025; He et al., 2026b). SVGD uses a kernelized Stein velocity (Liu & Wang, 2016; Liu, 2017), whereas conservative drifting uses the KDE-score residual itself. Amortized SVGD (Feng et al., 2017) likewise regresses a generator toward transported samples. We discuss further connections to generative transport, constrained drifting, and Sinkhorn and lookahead methods in Appendix B.

## 2 Conservative Drifting: Set-up and Assumptions

Let $K_h(u) = h^{-d}K(u/h)$, $\rho = K_h * \nu$, and $s_\rho = \nabla \log \rho$. For $x = (x_1, \dots, x_N)$, define

$$\mu_x^N = \frac{1}{N}\sum_i \delta_{x_i}, \qquad q_x(z) = \frac{1}{N}\sum_i K_h(z - x_i), \qquad b_x = s_\rho - \nabla \log q_x. \tag{3}$$

The interacting-particle system is $\dot{X}_i(t) = b_{X(t)}(X_i(t))$. For its joint density $p_t^N$, set $H_N(t) = \mathrm{KL}(p_t^N \| \rho^{\otimes N})$. Expectations $\mathbb{E}_t$ are with respect to this joint law. The Stein divergence is $\mathcal{A}_\rho f = \nabla \cdot f + s_\rho \cdot f$, and the relevant quantities are

$$\mathsf{S}_N(x) = \frac{1}{N}\sum_i \mathcal{A}_\rho b_x(x_i), \qquad \mathsf{R}_N(x) = \frac{1}{N}\sum_i \frac{1}{q_x(x_i)},$$
$$\mathsf{I}_N(x) = \int \|b_x(z)\|^2 \, q_x(z)\,\mathrm{d}z, \qquad \mathsf{V}_N(x) = \frac{1}{N}\sum_i \|b_x(x_i)\|^2. \tag{4}$$

Thus $\mathsf{I}_N$ is the smoothed Fisher discrepancy, and $\mathsf{V}_N$ is the mean squared particle speed. These quantities need not coincide with the center-evaluated Stein drift $\mathsf{S}_N$.

**Assumption 2.1** (Conservative regularity and stability)**.** *The kernel $K$ is positive, even, normalized, and $C^3$, with $m_2(K) := \int \|u\|^2 K(u)\,\mathrm{d}u < \infty$ and finite $\Delta K(0)$. The target score is $C^2$. The particle ODE has a unique solution on $[0, T]$ with a differentiable joint density solving its Liouville equation. The entropy is absolutely continuous, differentiation under its integral is justified, and the spatial and joint-space integrations by parts used below have vanishing boundary terms. All displayed integrals are finite. There is a deterministic $\Lambda_T < \infty$ such that*

$$\sup_{0 \le t \le T} \mathbb{E}_t[\mathsf{R}_N(X(t))] \le \Lambda_T.$$

*For every visited configuration, the scalar functions $\mathcal{A}_\rho b_x$ and $\|b_x\|^2$ are twice continuously differentiable and satisfy*

$$\sup_z \left\| D^2(\mathcal{A}_\rho b_x)(z) \right\|_{\mathrm{op}} \le B_{A,T}(h), \qquad \sup_z \left\| D^2 \|b_x(z)\|^2 \right\|_{\mathrm{op}} \le B_{V,T}(h).$$

The constants may depend on $N, h, T$ and the initial law. Uniformity is required only when an asymptotic rate invokes it. These are assumptions on the evolving system, including sufficient smoothness for the scalar Hessians; $K \in C^3$ alone need not provide that higher regularity. Appendix C gives the separate assumptions and sufficient derivative conditions. No log-concavity, log-Sobolev, or Poincaré inequality is imposed for the residual bounds.

## 3 Finite-particle Rates for Conservative Drifting

**Theorem 3.1** (Conservative entropy identity)**.** *Under the kernel and regular-flow conditions of Assumption 2.1,*

$$\frac{\mathrm{d}}{\mathrm{d}t} H_N(t) = -N\mathbb{E}_t[\mathsf{S}_N(X(t))] - \Delta K_h(0)\mathbb{E}_t[\mathsf{R}_N(X(t))]. \tag{5}$$

The correction in (5) comes from the dependence of $q_x$ on the center $x_i$ as well as on the evaluation point. For kernels with $\Delta K_h(0) < 0$, it contributes positively to the entropy derivative. A spatial integration by parts gives $\int \mathcal{A}_\rho b_x \, q_x = \mathsf{I}_N(x)$. Symmetric-kernel quadrature then compares this integral with $\mathsf{S}_N$ and compares $\mathsf{I}_N$ with $\mathsf{V}_N$. The complete entropy calculation and both quadrature comparisons are in Appendices D–E.

**Theorem 3.2** (Conservative Fisher and velocity bounds)**.** *Under Assumption 2.1, let* $a_1 = \{-\Delta K(0)\}_+$. *Then*

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{I}_N(X(t))]\,\mathrm{d}t \le \frac{H_N(0)}{NT} + \frac{a_1\Lambda_T}{Nh^{d+2}} + \frac{m_2(K)}{2}B_{A,T}(h)h^2, \tag{6}$$

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t \le \frac{H_N(0)}{NT} + \frac{a_1\Lambda_T}{Nh^{d+2}} + C_T(h)h^2, \tag{7}$$

*where* $C_T(h) = \frac{1}{2}m_2(K)\{B_{A,T}(h) + B_{V,T}(h)\}$.

The proofs are Theorem E.5 and Corollary E.7. If $H_N(0) \le \kappa_0 N$ and $T = N$, the three terms in (7) become

$$\frac{\kappa_0}{N} + \frac{a_1\Lambda_N(h)}{Nh^{d+2}} + C_N(h)h^2. \tag{8}$$

The first is the initial entropy per particle divided by the horizon; the second is the finite-particle self-interaction correction; the third is the quadrature error.

**Bandwidth dependence.** Suppose $\kappa_0$ and $a_1\Lambda_N(h)$ are uniformly bounded and $C_N(h) \le C_* h^{-\beta}$ for $0 \le \beta < 2$ throughout the admissible bandwidth range. Choosing $h \asymp N^{-1/(d+4-\beta)}$ gives

$$\left(\frac{1}{N}\int_0^N \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t\right)^{1/2} = O\left(N^{-\frac{2-\beta}{2(d+4-\beta)}}\right). \tag{9}$$

The case $\beta = 0$ yields $N^{-1/(d+4)}$. Similar to classical minimax bounds in nonparametric estimation (Tsybakov, 2009), our bounds also deteriorate with the ambient dimension, so the curse of dimensionality is naturally expected here rather than merely an artifact of this upper-bound argument. The bandwidth calculation is given in Appendix E.3.

The factor $h^2$ from the kernel second moment does not by itself make the quadrature error small. For example, derivative bounds at the kernel scale can give $B_{A,N}(h) + B_{V,N}(h) = O(h^{-4})$, producing an $O(h^{-2})$ term. The scalar Hessian assumptions must therefore be checked along the trajectory, including any cancellations between the target and model scores. If $\beta \ge 2$, this bound does not yield a vanishing quadrature error by shrinking $h$.

## 4 Reciprocal KDE Control and Initialization

Suppose $K(u) \ge \kappa_K > 0$ for $\|u\| \le r_K$. Let $M_i(x;h)$ count the centers within $r_K h$ of $x_i$, including $i$. Proposition F.2 shows that

$$M_i(x;h) \ge \alpha N h^d \text{ for all } i \quad \Longrightarrow \quad q_x(x_i) \ge \kappa_K\alpha, \qquad \mathsf{R}_N(x) \le (\kappa_K\alpha)^{-1}.$$

This controls denominators at particle centers without requiring a uniform density lower bound on $\mathbb{R}^d$.

**Proposition 4.1** (Independent samples)**.** *Let* $Y_1, \dots, Y_N$ *be i.i.d. from* $\mu$, *with* $N \ge 2$, *and write* $p_h(y) = \mu(B(y, r_K h))$. *Under the kernel lower bound above,*

$$\mathbb{E}[\mathsf{R}_N(Y)] \le \frac{h^d}{\kappa_K}\int \frac{1-(1-p_h(y))^N}{p_h(y)}\,\mu(\mathrm{d}y). \tag{10}$$

*The ratio is defined by its limit* $N$ *at* $p_h(y) = 0$. *If* $p_h(y) \ge p_0 h^d$ *for* $\mu$*-almost every* $y$, *then*

$$\mathbb{E}[\mathsf{R}_N(Y)] \le \frac{1}{\kappa_K p_0}, \qquad \mathbb{P}\left(\mathsf{R}_N(Y) > \frac{4}{\kappa_K p_0}\right) \le N\exp\left(-\frac{p_0 N h^d}{16}\right).$$

The proof uses the exact conditional binomial identity for $1/M_i$, together with a lower-tail bound; see Proposition F.5 and (62). The expectation estimate does not require $Nh^d \gtrsim \log N$. That condition arises when simultaneous occupancy is required with high probability. A density bounded below on a sufficiently thick bounded support supplies the lower local mass condition; the corresponding example is in Appendix F.

**Propagation along the flow.** If $\left\|b_{X(t)}(z) - b_{X(t)}(w)\right\| \le L(t)\,\|z-w\|$ with deterministic $\Gamma_T = \int_0^T L(t)\,\mathrm{d}t$, initial neighbors within $r_K h e^{-\Gamma_T}$ remain within $r_K h$. For i.i.d. initialization satisfying the lower local mass condition at bandwidth $he^{-\Gamma_T}$, Corollary F.10 gives

$$\mathbb{E}\left[\sup_{0\le t\le T} \mathsf{R}_N(X(t))\right] \le \frac{e^{d\Gamma_T}}{\kappa_K p_0}. \tag{11}$$

Uniform distortion on a growing horizon is a separate stability condition. In particular, a Lipschitz bound of order $h^{-2}$ integrated over $[0,N]$ does not give a uniform right-hand side in (11).

**Proposition 4.2** (Full-support initialization)**.** *Let $Y_1,\dots,Y_N$ be i.i.d. from a bounded continuous density $f$ positive on $\mathbb{R}^d$. Suppose $K$ is nonnegative and normalized, $0 < K(0) < \infty$, $h_N \to 0$, and $Nh_N^d \to \infty$. Then $\mathbb{E}[\mathsf{R}_N(Y)] \to \infty$.*

The proof is Proposition F.12. This includes Gaussian initialization. Since the time supremum includes time zero, $\Lambda_N(h_N)$ cannot be uniformly bounded in this regime. The finite-particle inequality remains valid with its actual reciprocal coefficient, whose growth must enter any bandwidth optimization.

## 5 Non-conservative Drifting with the Laplace Kernel

Let $K_h(u) = c_d h^{-d} e^{-\|u\|/h}$, normalized on $\mathbb{R}^d$. For $N \ge 2$, set $\mu^N_{x,-i} = (N-1)^{-1}\sum_{j\ne i}\delta_{x_j}$. We analyze the self-masked dynamics

$$\dot X_i(t) = u_{X(t),-i}(X_i(t)), \qquad u_{x,-i} = M_{\nu,h} - M_{\mu^N_{x,-i},h}, \qquad u_x = M_{\nu,h} - M_{\mu^N_x,h}. \tag{12}$$

Self-masking agrees with the diagonal masking in the original drifting implementation (Deng et al., 2026, Algorithm 2). It also removes the model's dependence on $x_i$ when differentiating $u_{x,-i}(z)$ at $z = x_i$.

### 5.1 Sharp scores and the local scale

Define the companion kernel and its convolutions by

$$L_h(u) = h(\|u\| + h)K_h(u), \qquad Q_{\alpha,h} = K_h * \alpha, \qquad R_{\alpha,h} = L_h * \alpha. \tag{13}$$

Write $\sigma_{\alpha,h} = \nabla \log R_{\alpha,h}$ and $a_{\alpha,h} = R_{\alpha,h}/Q_{\alpha,h}$. Lemmas G.1 and G.17 establish

$$\nabla L_h(u) = -uK_h(u), \quad M_{\alpha,h} = a_{\alpha,h}\sigma_{\alpha,h}, \quad Z_{\#,h} := \int L_h = (d+1)h^2, \quad \|\sigma_{\alpha,h}\|_\infty \le h^{-1}. \tag{14}$$

Let $\rho^{\#}_{\alpha,h} = R_{\alpha,h}/Z_{\#,h}$, and use the subscript $x$ for $\alpha = \mu^N_x$. The full field decomposes as

$$u_x = a_{x,h} b^{\#}_x + e_x, \qquad b^{\#}_x = \sigma_{\nu,h} - \sigma_{x,h}, \qquad e_x = (a_{\nu,h} - a_{x,h})\sigma_{\nu,h}. \tag{15}$$

The scale is $a_{\alpha,h} = h(\bar r_{\alpha,h} + h)$, where $\bar r_{\alpha,h}$ is the kernel-weighted mean distance from $z$ to $y$. It is at least $h^2$. For a finite configuration with $R_x = \max_j \|x_j\|$,

$$h(\|z\| - R_x + h) \le a_{x,h}(z) \le h(\|z\| + R_x + h).$$

Thus no finite global upper bound is available. We retain the scale inside the energies

$$\mathcal{W}^{\mathrm{Lap}}_h(x) = \int \frac{\|u_x\|^2}{a_{x,h}}\rho^{\#}_{x,h}, \qquad \mathcal{D}^{\mathrm{Lap}}_h(x) = \int \frac{\|e_x\|^2}{a_{x,h}}\rho^{\#}_{x,h}. \tag{16}$$

### 5.2 ENTROPY, COERCIVITY, AND PARTICLE RATES

Set $\mathcal{A}_{\#} f = \nabla \cdot f + \sigma_{\nu,h} \cdot f$, and define

$$\mathsf{S}_N^{\mathrm{Lap}}(x) = \frac{1}{N} \sum_i \mathcal{A}_{\#} u_{x,-i}(x_i), \qquad \mathcal{J}_h^{\mathrm{Lap}}(x) = \int \mathcal{A}_{\#} u_x \, \rho_{x,h}^{\#},$$
$$\mathsf{V}_N^{\mathrm{Lap}}(x) = \frac{1}{N} \sum_i \left\| u_{x,-i}(x_i) \right\|^2, \qquad \mathcal{V}_h^{\mathrm{Lap}}(x) = \int \|u_x\|^2 \, \rho_{x,h}^{\#}. \tag{17}$$

The full field is locally Lipschitz but may be nondifferentiable at particle centers. Its divergence in $\mathcal{J}_h^{\mathrm{Lap}}$ is a weak derivative, represented almost everywhere. Classical point evaluations are used only for the leave-one-out fields.

**Assumption 5.1** (Laplace regularity and approximation)**.** *The regular-flow and integration conditions in Assumption G.2 hold on $[0,T]$. In particular, entropy differentiation is justified, the joint entropy is absolutely continuous, spatial integrations by parts are valid for the weak derivatives, and each $u_{x,-i}$ is classically differentiable at $x_i$. Uniformly over visited configurations, all energies are finite and*

$$\mathcal{D}_h^{\mathrm{Lap}}(x) \le \Delta_h^2,$$
$$\left| \mathsf{S}_N^{\mathrm{Lap}}(x) - \mathcal{J}_h^{\mathrm{Lap}}(x) \right| \le \varepsilon_{S,h,N}, \qquad \left| \mathsf{V}_N^{\mathrm{Lap}}(x) - \mathcal{V}_h^{\mathrm{Lap}}(x) \right| \le \varepsilon_{V,h,N}.$$

*For the unweighted bound, also assume there is a measurable $\Omega_x$ with*

$$a_{x,h}(z) \le L_h \quad (z \in \Omega_x), \qquad \int_{\Omega_x^c} \|u_x\|^2 \, \rho_{x,h}^{\#} \le \tau_{h,N}.$$

*Here the scalar $L_h > 0$ and the other constants are independent of configuration and time on $[0,T]$; they may depend on $N, h, T$. The notation $L_h$ without a spatial argument denotes this scalar bound.*

Write $H_N^{\#}(t) = \mathrm{KL}(p_t^N \| (\rho_{\nu,h}^{\#})^{\otimes N})$. Theorems and lemmas in Appendix G give the two identities and the coercivity bound

$$\frac{\mathrm{d}}{\mathrm{d}t} H_N^{\#}(t) = -N \mathbb{E}_t[\mathsf{S}_N^{\mathrm{Lap}}(X(t))],$$
$$\mathcal{J}_h^{\mathrm{Lap}}(x) = \int b_x^{\#} \cdot u_x \, \rho_{x,h}^{\#} \ge \frac{1}{2} \mathcal{W}_h^{\mathrm{Lap}}(x) - \frac{1}{2} \mathcal{D}_h^{\mathrm{Lap}}(x). \tag{18}$$

There is no reciprocal self-interaction correction in this entropy identity. The scale mismatch enters the coercivity comparison instead.

**Theorem 5.2** (Laplace residual bounds)**.** *Under Assumption 5.1,*

$$\frac{1}{T} \int_0^T \mathbb{E}_t[\mathcal{W}_h^{\mathrm{Lap}}(X(t))] \, \mathrm{d}t \le \frac{2 H_N^{\#}(0)}{NT} + \Delta_h^2 + 2\varepsilon_{S,h,N}, \tag{19}$$
$$\frac{1}{T} \int_0^T \mathbb{E}_t[\mathsf{V}_N^{\mathrm{Lap}}(X(t))] \, \mathrm{d}t \le \frac{2 L_h H_N^{\#}(0)}{NT} + L_h \Delta_h^2 + 2 L_h \varepsilon_{S,h,N} + \tau_{h,N} + \varepsilon_{V,h,N}. \tag{20}$$

*The first bound does not require localization. If the absolute mismatch bound is replaced by $\mathcal{D}_h^{\mathrm{Lap}} \le \theta_h \mathcal{W}_h^{\mathrm{Lap}}$ with $0 \le \theta_h < 1$, then*

$$\frac{1}{T} \int_0^T \mathbb{E}_t[\mathsf{V}_N^{\mathrm{Lap}}(X(t))] \, \mathrm{d}t \le \frac{L_h}{1 - \sqrt{\theta_h}} \left\{ \frac{H_N^{\#}(0)}{NT} + \varepsilon_{S,h,N} \right\} + \tau_{h,N} + \varepsilon_{V,h,N}. \tag{21}$$

The proofs are Theorem G.13 and Corollary G.14. Absolute alignment gives an additive upper bound on the residual; it does not imply a positive limiting error floor. Relative alignment absorbs the mismatch into the weighted dissipation with coefficient $1 - \sqrt{\theta_h}$.

### 5.3 Checking the Laplace assumptions

The supporting estimates separate alignment, tails, and approximation. If $\|\sigma_{\nu,h}\|_\infty \le G_h$ and $\int |\bar{r}_{\nu,h} - \bar{r}_{x,h}|^2 \rho^{\#}_{x,h} \le \delta^2_{r,h}$, then $\Delta_h^2 = G_h^2 \delta^2_{r,h}$ is admissible. The universal choice is $G_h = h^{-1}$; additional target regularity may give a smaller value.

For data supported in $B(0, R_\nu)$ and particles confined to $B(0, R_T)$, Proposition G.19 gives, for every $r > 0$,

$$\Omega_x = B(0, R_T + r), \qquad L_h = h(2R_T + r + h), \qquad \tau_{h,N} = (R_\nu + R_T)^2 p_d(r/h), \tag{22}$$

where the exact companion-kernel tail is

$$p_d(s) = e^{-s} \left\{ \frac{d}{d+1} \sum_{k=0}^{d} \frac{s^k}{k!} + \frac{1}{d+1} \sum_{k=0}^{d-1} \frac{s^k}{k!} \right\}.$$

Taking $r = ch \log N$ yields a tail proportional to $(1 + (\log N)^d) N^{-c}$, with the confinement factor retained. A bounded initial cloud gives $R_T \le (R_0 + R_\nu) e^T - R_\nu$ on a regular finite horizon. Uniform control on $T = N$ requires a stronger confinement estimate.

For quadrature, let $k_i(z) = L_h(z - x_i)/Z_{\#,h}$ and $d_i = u_{x,-i} - u_x$. Let $q_S$ and $q_V$ be the average absolute errors between the value at $x_i$ and the $k_i$-integral of $\mathcal{A}_\# u_{x,-i}$ and $\|u_{x,-i}\|^2$, respectively. Lemma G.21 gives

$$\begin{aligned} \left| \mathsf{S}_N^{\mathrm{Lap}} - \mathcal{J}_h^{\mathrm{Lap}} \right| &\le q_S + \frac{2}{Nh} \sum_i \int k_i \, \|d_i\| \,, \\ \left| \mathsf{V}_N^{\mathrm{Lap}} - \mathcal{V}_h^{\mathrm{Lap}} \right| &\le q_V + \frac{1}{N} \sum_i \int k_i (\|u_{x,-i}\| + \|u_x\|) \, \|d_i\| \,. \end{aligned} \tag{23}$$

The comparison uses integration by parts against each $k_i$ and differences of field values. It never differentiates the full field at a particle center. Local Taylor estimates for $q_S, q_V$ apply on balls where the leave-one-out integrands are smooth; the complementary integrals are bounded explicitly. With finite data support, these balls must avoid the other centers and all data atoms. Their radii, local Hessian bounds, and integrated leave-one-out errors all enter the final constants. Center-denominator control alone does not ensure a uniform quadrature rate.

For $T = N$ and $H_N^{\#}(0) \le \kappa_0 N$, the right side of (20) is

$$\mathcal{R}_{N,h} := \frac{2\kappa_0 L_h}{N} + L_h \Delta_h^2 + 2 L_h \varepsilon_{S,h,N} + \tau_{h,N} + \varepsilon_{V,h,N}. \tag{24}$$

A bandwidth sequence gives a vanishing root residual if $\kappa_0$ stays bounded and $L_h/N$, $L_h \Delta_h^2$, $L_h \varepsilon_{S,h,N}$, and $\tau_{h,N} + \varepsilon_{V,h,N}$ tend to zero, with the assumptions valid throughout $[0, N]$. The condition $Nh^d \gtrsim \log N$ can support local occupancy, but does not by itself establish these requirements.

## 6 Numerical Experiments

We use low-dimensional simulations to isolate two mechanisms that are specific to the theory: the reciprocal self-interaction correction in the conservative entropy identity, and the scale-alignment coefficient in the Laplace coercivity estimate. The experiments are diagnostic rather than generative benchmarks. Target scores and kernel sums are evaluated analytically or as finite sums, and uncertainty is computed across independent particle clouds; particles within one interacting cloud are not treated as independent replicates.

### 6.1 Conservative correction and reciprocal tails

**Aim and setup.** The first experiment asks whether the correction in (5) is numerically necessary and whether the reciprocal-KDE assumptions distinguish bounded- from full-support initialization. We use the Gaussian conservative field with analytic one- and two-dimensional Gaussian-mixture targets. For the identity check, we take $d \in \{1, 2\}$, $N \in \{8, 16, 32\}$, $h \in \{0.15, 0.30, 0.60\}$, and

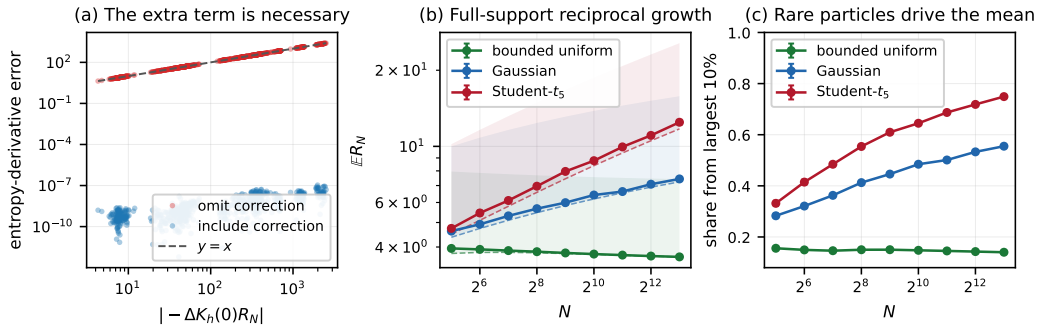



Figure 2: **The conservative correction and reciprocal-KDE obstruction.** (a) Error in the independently evaluated pathwise entropy derivative, plotted against the magnitude of the reciprocal correction. Omitting the correction follows the identity line; including it leaves only finite-difference error. (b) In $d = 1$ with $h_N = N^{-1/5}$, the mean reciprocal KDE remains bounded for uniform initialization but grows for the two full-support laws. Error bars are 95% Monte Carlo intervals; shaded regions run from the Jensen lower bound to the exact expectation upper bound, with the lower bound dashed. (c) Fraction of $\mathsf{R}_N$ contributed by the largest 10% of particle reciprocals.

five independent clouds from each of a variance-matched uniform, Gaussian, and Student-$t_5$ law. At times $0, 0.05, 0.25$ along the conservative ODE, coordinatewise central differences give an independent configuration divergence and hence the pathwise derivative $\widehat{\dot{G}} = -\widehat{\nabla_x \cdot F} - \sum_i s_{\rho,h}(X_i) \cdot F_i$. We compare its error under the naive prediction $-N\mathsf{S}_N$ and the corrected prediction $-N\mathsf{S}_N - \Delta K_h(0)\mathsf{R}_N$. This gives 810 derivative comparisons. Separately, at $h_N = N^{-1/(d+4)}$, we estimate $\mathbb{E}\mathsf{R}_N$ up to $N = 8192$ in $d = 1$ and $N = 4096$ in $d = 2$. Each cell uses 250 clouds through $N = 1024$, 100 through $N = 4096$, and 40 at $N = 8192$; additional runs use bandwidth constants $c \in \{0.5, 2\}$.

**Results and interpretation.** Figure 2(a) shows that the naive error equals the omitted correction to relative error below $4 \times 10^{-10}$ in every comparison. With the correction included, the median error relative to the correction is $2.3 \times 10^{-11}$ and the maximum is $3.8 \times 10^{-10}$. The correction is not a negligible numerical detail: its median magnitude is $0.72$ times that of the Stein term $-N\mathsf{S}_N$, with 10th–90th percentile range $0.25$–$2.67$.

In $d = 1$, increasing $N$ from 32 to 8192 changes $\mathbb{E}\mathsf{R}_N$ from $3.95$ to $3.66$ for bounded uniform initialization, from $4.62$ to $7.43$ for Gaussian initialization, and from $4.74$ to $12.45$ for Student-$t_5$. At the largest $N$, the most isolated 10% of particles contribute respectively 14%, 56%, and 75% of the sum. The same separation occurs in $d = 2$, and every measured mean lies between the evaluated lower and upper bounds. Thus the self-interaction term is required even at moderate $N$, while the full-support obstruction is carried by increasingly rare low-density particles, precisely the behavior that a median diagnostic would conceal.

### 6.2 Laplace scale-alignment phase diagram

**Aim and setup.** The second experiment asks whether the observable ratio $\theta = \mathcal{D}_h^{\mathrm{Lap}}/\mathcal{W}_h^{\mathrm{Lap}}$ predicts the strength of the Laplace coercivity certificate. The target is eight fixed atoms in $\mathbb{R}^2$, so all target-side Laplace quantities are exact. We draw $N = 128$ particles from $\mathcal{N}(m, s^2 I_2)$, with five shifts $m \in \{(0,0), (1.5,0), (3,3), (-4.5,1), (7,0)\}$, five spreads $s \in \{0.2, 0.4, 1, 2.5, 5\}$, three bandwidths $h \in \{0.15, 0.30, 0.60\}$, and six seeds, giving 450 configurations. For each, approximately 24,000 stratified exact draws from $\rho^{\#}_{x,h}$ estimate $\mathcal{W}_h^{\mathrm{Lap}}$, $\mathcal{D}_h^{\mathrm{Lap}}$, and $\mathcal{J}_h^{\mathrm{Lap}}$. We also evolve four representative initialization families under the ODE induced by the exact leave-one-out Laplace field, using five seeds at each bandwidth. The confinement radius $r = ch \log N$ used in the accompanying velocity diagnostics is optimized over $c \in [0.15, 6]$ rather than fixed in advance.

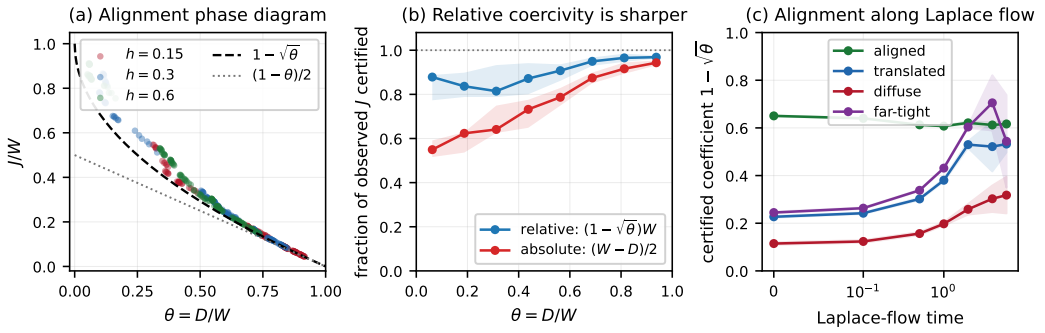



Figure 3: **Weighted scale alignment for exact Laplace drifting.** (a) The measured ratio $\mathcal{J}_h^{\mathrm{Lap}}/\mathcal{W}_h^{\mathrm{Lap}}$ against $\theta = \mathcal{D}_h^{\mathrm{Lap}}/\mathcal{W}_h^{\mathrm{Lap}}$. The dashed curve is the relative certificate $1 - \sqrt{\theta}$ and the dotted curve is the absolute certificate $(1 - \theta)/2$. (b) Fraction of the observed $\mathcal{J}_h^{\mathrm{Lap}}$ certified by each lower bound. Lines are binwise medians and bands are 10th–90th percentiles. (c) The relative coefficient along selected Laplace flows at $h = 0.30$; lines and bands are means and 95% intervals across five independent clouds.

**Results and interpretation.** The configurations span $\theta = 0.051$ to $0.917$. Every measured point in Figure 3(a) lies above $1 - \sqrt{\theta}$; the smallest observed normalized slack is $8.6 \times 10^{-4}$. More importantly, the relative lower bound certifies a median 93.3% of the observed $\mathcal{J}_h^{\mathrm{Lap}}$ (10th–90th percentiles 81.4%–97.1%), compared with 83.2% (55.6%–93.2%) for $(\mathcal{W} - \mathcal{D})/2$.

The trajectory experiment shows how the diagnostic behaves on states actually visited by the particle system. At $h = 0.30$, the coefficient $1 - \sqrt{\theta}$ rises from $0.23$ to $0.53$ for translated clouds, from $0.12$ to $0.32$ for diffuse clouds, and from $0.25$ to $0.54$ for far but tight clouds; an initially well-aligned centered cloud remains near $0.6$. This is not a claim that alignment is monotone for every trajectory. Rather, it shows that the scale-mismatch term is quantitatively active and that $\theta$ separates strongly and weakly coercive regimes along the exact-field dynamics. Dropping $\mathcal{D}_h^{\mathrm{Lap}}$ would miss this distinction, whereas the relative theory gives a nearly sharp pointwise certificate over the tested range.

## 7 Conclusion and Limitations

We developed finite-particle convergence guarantees for conservative kernel-score drifting and the original non-conservative Laplace displacement field, showing that conservativeness alone does not determine performance. For conservative dynamics, we control the smoothed score discrepancy and particle velocity and identify an exact self-interaction correction involving reciprocal kernel-density estimates: local particle occupancy keeps it finite, but full-support initializations can make its expectation grow as bandwidth shrinks. Although circulation-free, this field has a distribution-dependent potential that need not improve monotonically along individual trajectories. The leave-one-out Laplace field avoids the correction, but its position-dependent normalization introduces circulation and scale mismatch; a companion-kernel representation nevertheless controls velocity despite these scale issues, the kernel cusp, and unbounded empirical scales. Simulations confirm that the conservative correction explains the observed entropy-identity defect and that relative scale alignment sharpens the Laplace certificate. Together, these results identify the quantities governing each field's distinct failure modes and clarify when each guarantee is informative.

Our guarantees are time-averaged and conditional on density, quadrature, and stability assumptions; they do not yet establish global distributional convergence or a terminal-time guarantee. The analysis also idealizes training as an exact continuous-time particle system, while practical neural network training adds discretization, approximation, optimization, and minibatch errors. Future work should connect the entropy bounds to terminal convergence, extend the analysis to finite-step learned generators and broader kernels, develop adaptive methods for low-density particles and scale mismatch, and validate these mechanisms on realistic high-dimensional data.

### AI USE STATEMENT

Generative AI tools (ChatGPT) were used to check mathematical arguments, identify proof gaps, assist in formulating and writing revisions to the Laplace analysis and reciprocal-KDE bounds, and edit and format the manuscript. They also assisted with literature searches and preparation of the LaTeX source. Responsibility for the mathematical claims, citations, and final manuscript rests with the authors.

### REPRODUCIBILITY STATEMENT

The appendices provide the full assumptions and proofs for the entropy identities, conservative rates, reciprocal-KDE estimates, and Laplace bounds. They also give the kernel constants, quadrature and tail estimates, bandwidth calculations, and the connection between stop-gradient regression and the particle ODE. The hypotheses required for uniform asymptotic rates are distinguished from the finite-horizon inequalities. Simulation code, fixed parameter grids, random seeds, numerical checks, and figure-generation scripts are included in the supplementary file.

## REFERENCES

Michael Samuel Albergo and Eric Vanden-Eijnden. Building Normalizing Flows with Stochastic Interpolants. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=li7qeBbCR1t.

Sayan Banerjee, Krishnakumar Balasubramanian, and Promit Ghosal. Improved Finite-Particle Convergence Rates for Stein Variational Gradient Descent. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=sbG8qhMjkZ.

Andrew D Barbour. Stein's method and Poisson process convergence. *Journal of Applied Probability*, 25(A):175–184, 1988.

Nicholas Matthew Boffi, Michael Samuel Albergo, and Eric Vanden-Eijnden. Flow map matching with stochastic interpolants: A mathematical framework for consistency models. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=cqDH0e6ak2.

Jiarui Cao, Zixuan Wei, and Yuxin Liu. Gradient flow drifting: Generative modeling via Wasserstein gradient flows of KDE-approximated divergences. *arXiv preprint arXiv:2603.10592*, 2026.

Mingyang Deng, He Li, Tianhong Li, Yilun Du, and Kaiming He. Generative modeling via drifting. *arXiv preprint arXiv:2602.04770*, 2026.

Théo Dumont, Théo Lacombe, and François-Xavier Vialard. Learning Monge maps with constrained drifting models. *arXiv preprint arXiv:2603.25182*, 2026.

Maria Esteban-Casadevall, Jorge Carrasco-Pollo, Max Welling, Jan-Willem van de Meent, Erik J. Bekkers, and Floor Eijkelboom. Kernel-gradient drifting models. *arXiv preprint arXiv:2605.10727*, 2026.

Yihao Feng, Dilin Wang, and Qiang Liu. Learning to draw samples with amortized Stein variational gradient descent. In *The Conference on Uncertainty in Artificial Intelligence (UAI)*, 2017.

Leonard Franz, Sebastian Hoffmann, and Georg Martius. Drifting fields are not conservative. *arXiv preprint arXiv:2604.06333*, 2026.

Jackson Gorham and Lester Mackey. Measuring sample quality with Stein's method. *Advances in neural information processing systems*, 28, 2015.

Arthur Gretton, Li Kevin Wenliang, Alexandre Galashov, James Thornton, Valentin De Bortoli, and Arnaud Doucet. On the Wasserstein Gradient Flow Interpretation of Drifting Models. *arXiv preprint arXiv:2605.05118*, 2026.

Ping He, Om Khangaonkar, Hamed Pirsiavash, Yikun Bai, and Soheil Kolouri. Sinkhorn-Drifting Generative Models. *arXiv preprint arXiv:2603.12366*, 2026a.

Ye He, Krishnakumar Balasubramanian, Sayan Banerjee, and Promit Ghosal. Finite-Particle Rates for Regularized Stein Variational Gradient Descent. *arXiv preprint arXiv:2602.05172*, 2026b.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.

Chieh-Hsin Lai, Bac Nguyen, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yuki Mitsufuji, Stefano Ermon, and Molei Tao. A unified view of drifting and score-based models. *arXiv preprint arXiv:2603.07514*, 2026.

Hak Geun Lee and Hyonho Chun. Identifiability and Stability of Generative Drifting with Companion-Elliptic Kernel Families. *arXiv preprint arXiv:2604.24196*, 2026.

Zhiqi Li and Bo Zhu. A Long-Short Flow-Map Perspective for Drifting Models. *arXiv preprint arXiv:2602.20463*, 2026.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow Matching for Generative Modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Qiang Liu. Stein variational gradient descent as gradient flow. *Advances in neural information processing systems*, 30, 2017.

Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. *Advances in neural information processing systems*, 29, 2016.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Mathew Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=TIdIXIpzhoI.

Jiaxin Shi and Lester Mackey. A finite-particle convergence rate for Stein variational gradient descent. *Advances in Neural Information Processing Systems*, 36:26831–26844, 2023.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning*, 2023.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009.

Erkan Turan and Maks Ovsjanikov. Generative drifting is secretly score matching: A spectral and variational perspective. *arXiv preprint arXiv:2603.09936*, 2026.

Romann M. Weber. The Score-Difference Flow for Implicit Generative Modeling. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=dpGSNLUCzu.

Guoqiang Zhang, Kenta Niwa, and W. Bastiaan Kleijn. Lookahead drifting model. *arXiv preprint arXiv:2605.04060*, 2026.

# A Conservative versus Non-conservative Drift: A Visual Comparison

To build intuition for the distinction between the conservative and non-conservative drifting fields, consider a simple two-dimensional setup in which the data distribution $\nu$ is supported on a thin horizontal cluster along the $x$-axis and the model distribution $\mu$ is initialized on a thin vertical cluster along the $y$-axis. We run the finite-particle dynamics from the same initial configuration under each of the two velocity fields, $b_{\nu,\mu,h}(z)$ with a Gaussian kernel for the conservative case, and $u^{\mathrm{Lap}}_{\nu,\mu,h}(z) := M_{\nu,h}(z) - M_{\mu,h}(z)$ with a Laplace kernel for the non-conservative case, as in Deng et al. (2026). We track four model particles whose initial positions are identical across the two simulations.

Figure 1 shows the resulting trajectories. Both fields transport the displayed model cloud toward the data cluster, and the four pairs of paths share their starting points. At time $t$, the conservative field is the spatial gradient of

$$\Phi_t(z) := \log \rho_{\nu,h}(z) - \log \rho_{\mu_t,h}(z), \qquad b_{\nu,\mu_t,h}(z) = \nabla_z \Phi_t(z).$$

Thus its circulation around a closed curve is zero whenever the spatial derivatives are defined. The potential itself evolves with $\mu_t$. Along a trajectory $\dot{Z}(t) = \nabla_z \Phi_t(Z(t))$, the chain rule gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_t(Z(t)) = \partial_t \Phi_t(Z(t)) + \|\nabla_z \Phi_t(Z(t))\|^2 .$$

The sign of the full time derivative therefore depends on the changing model density as well as on the instantaneous drift.

For the Laplace field, the position-dependent normalization can produce nonzero circulation. Curvature of a trajectory by itself does not distinguish the two fields: a spatial gradient field can also have curved integral curves. The algebraic distinction is captured by the local scale $a_{\alpha,h} = R_{\alpha,h}/Q_{\alpha,h}$ in Lemma G.1. Its spatial variation enters the curl, while its difference between the data and model enters the scale-mismatch term. The weighted and localized estimates in Theorem G.13 keep this scale dependence explicit; see Remark G.16 for the circulation identity.

## A.1 Helmholtz–Hodge decomposition of the initial fields

We test the distinction directly in a deterministic replicate of the Figure 1 setting. The replicate uses 400 data atoms in a thin horizontal cloud and the same 80 model particles in a thin vertical cloud for both fields. The long data and model coordinates are sampled uniformly on $[-2.85, 2.85]$ and $[-2.65, 2.65]$, respectively; the transverse noise is Gaussian with standard deviation $0.065$, and each cloud is recentered. We use seed 1701, bandwidth $h = 0.55$, and a $101 \times 101$ node grid over $[-3.4, 3.4]^2$.

Because a finite-domain Helmholtz decomposition depends on its boundary condition, we avoid a periodic Fourier projection. Instead, we sample the component tangent to each grid edge and use the orthogonal graph-Hodge decomposition

$$f = G\phi + r, \qquad G^\top r = 0,$$

where $G$ is the node-to-edge gradient. The first term is a discrete curl-free field, and the residual is discrete divergence-free under the natural grid boundary condition. This construction exactly reconstructs the sampled field and separates circulation from the curvature of individual trajectories.

Figure 4 makes two points. First, the conservative field is recovered almost entirely by the gradient projection, and its remaining rotational energy decreases rapidly under grid refinement. Second, the Laplace field is still predominantly gradient-like, but it has a small, spatially organized rotational component that persists under refinement. Non-conservative therefore does not mean "mostly rotational"; it means that a nonzero circulation component remains after the best curl-free projection. The figure and all diagnostics are reproduced by `supplement/exp7_helmholtz_decomposition.py`.

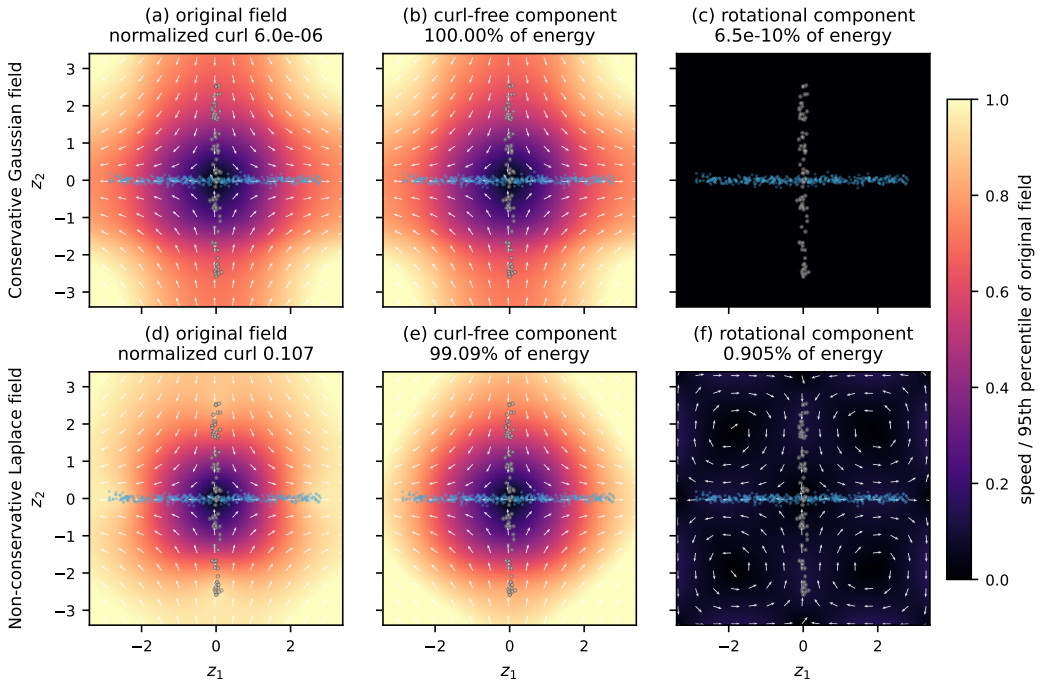



Figure 4: **Helmholtz–Hodge decomposition in the Figure 1 setting.** Rows show the conservative Gaussian field and non-conservative Laplace field at initialization; columns show the original field, its curl-free projection, and its rotational residual. Color is component speed divided by the 95th percentile of the corresponding original field, so the magnitude scale is shared within each row. Arrow lengths are normalized wherever the component exceeds 0.5% of that reference and therefore show direction rather than relative magnitude. The conservative residual contains $6.5 \times 10^{-10}\%$ of the edge energy and has bandwidth-normalized root-mean-square curl $6.0 \times 10^{-6}$, consistent with discretizing an exact gradient. The Laplace residual contains $0.905\%$ of the energy, has normalized curl $0.107$, and forms four coherent circulation cells. On a $51 \times 51$ grid, the corresponding fractions are $1.1 \times 10^{-8}\%$ and $0.937\%$, respectively.

## B Further Related Work

Drifting models were introduced by Deng et al. (2026) as a training-time transport paradigm for one-step generation: generated samples are repeatedly moved by a data-attractive and model-repulsive drift field, and the generator is trained to reproduce the drifted samples through a stop-gradient target. This viewpoint is also related to score-difference flows for implicit generative modeling, where the velocity is written as the difference between target and source scores (Weber, 2023). Subsequent work has connected drifting to score matching and Wasserstein gradient flows of smoothed divergences (in the infinite-particle setting), especially in the Gaussian-kernel case where the displacement field can be written as a difference of KDE log-density gradients (Cao et al., 2026; Lai et al., 2026; Turan & Ovsjanikov, 2026; Gretton et al., 2026).

A complementary line of work emphasizes the non-conservative nature of the original displacement-based field. Franz et al. (2026) show that the position-dependent normalization in the original field generally destroys conservatism, with the Gaussian kernel being a special case where the field remains a gradient field; they also introduce sharp normalizations that restore conservatism for broader radial kernels. Lee & Chun (2026) study identifiability and stability for companion-elliptic kernel families, including the Laplace kernel, showing that vanishing drift can identify the target under appropriate kernel structure while also highlighting possible instability modes. These works motivate our distinction between the conservative drifting method, based on KDE-score gradients, and the non-conservative drifting method by Deng et al. (2026), based on the original displacement velocity.

For the latter, our analysis uses weighted coercivity to retain the local scale and controls the scale mismatch either additively or through relative alignment.

Recent drifting-related works study complementary algorithmic and variational perspectives. Dumont et al. (2026) study learning Monge maps through constrained drifting models: they lift a divergence, such as relative entropy, to the space of transport maps, constrain the dynamics to the set of optimal transport maps, and prove long-time existence and convergence toward the OT map under convexity assumptions. He et al. (2026a) introduce Sinkhorn-Drifting Generative Models, showing that the attraction–repulsion structure of drifting can be viewed as a one-sided surrogate for a Sinkhorn-divergence gradient flow, while two-sided Sinkhorn scaling gives improved identifiability and stability. Zhang et al. (2026) propose Lookahead Drifting Models, which compute several sequential drift terms within each training iteration and regress toward a weighted sum of these terms, with the goal of incorporating higher-order drift information while preserving one-step inference.

More broadly, modern generative modeling is often organized around learned transport dynamics. Diffusion and score-based models learn reverse-time denoising dynamics and generate samples by numerically solving an SDE or probability-flow ODE (Ho et al., 2020; Song et al., 2021), while flow matching and rectified flow learn time-dependent vector fields that transport a base distribution to the data distribution along prescribed probability paths (Albergo & Vanden-Eijnden, 2023; Lipman et al., 2023; Liu et al., 2023). Flow-map and distillation-based approaches instead try to learn longer-time solution operators, or compress many solver steps into one or a few neural evaluations, as in flow-map matching, consistency models, and progressive distillation (Boffi et al., 2025; Song et al., 2023; Salimans & Ho, 2022). From this viewpoint, drifting models can be interpreted as moving part of the inference-time transport computation into training: the long-short flow-map perspective of Li & Zhu (2026) decomposes a global transport into a long-horizon map followed by a short terminal correction, and recovers the drifting field in the limit of a vanishing terminal interval.

The proposed conservative velocity approach is also closely related to deterministic interacting-particle methods based on Stein discrepancies. Stein variational gradient descent (SVGD) was introduced by Liu & Wang (2016) as a deterministic particle method for approximate Bayesian inference, and Liu (2017) interpreted its population limit as a gradient flow of KL divergence under a Stein-induced geometry. Amortized SVGD (Feng et al., 2017) trains a sampler network to reproduce SVGD particle updates, which is conceptually close to the stop-gradient regression view of drifting. At the level of finite-particle analysis, Banerjee et al. (2025) obtain improved SVGD rates by differentiating the joint entropy of the particle law relative to the product target, improving upon earlier results by Shi & Mackey (2023). Extension to a regularized setting was studied by He et al. (2026b). We leverage and extend this high-level entropy-dissipation strategy in our proofs, but the controlled quantities and error terms are specific to drifting: for conservative drifting, the controlled quantity is the mean squared conservative residual with a reciprocal-KDE self-interaction and an $h$-dependent quadrature error; for the non-conservative Laplace method, the rate additionally contains the Laplace scale-mismatch residual.

# C Detailed Set-up and Assumptions for Conservative Drifting

## C.1 Set-up for Conservative Drifting

Let $\rho$ be a positive probability density on $\mathbb{R}^d$. For the conservative drifting method proposed in this paper, $\rho$ should be read as the kernel-smoothed data density $\rho_{\nu,h} = K_h * \nu$. Write

$$s_\rho(z) := \nabla \log \rho(z).$$

Let $K : \mathbb{R}^d \to (0, \infty)$ be a normalized kernel and, for a bandwidth $h > 0$, define

$$K_h(u) := h^{-d} K(u/h).$$

For a particle configuration

$$x = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N,$$

define the kernel density estimate

$$q_x(z) := \frac{1}{N} \sum_{j=1}^{N} K_h(z - x_j), \qquad s_x(z) := \nabla \log q_x(z),$$

and the conservative drift field

$$b_x(z) := s_\rho(z) - s_x(z).$$

The center-evaluation finite-particle ODE for the conservative drifting method is

$$\dot{X}_i(t) = b_i(X(t)), \qquad b_i(x) := b_x(x_i), \qquad i = 1, \dots, N. \tag{25}$$

This is a deterministic interacting particle system: each particle follows an ordinary differential equation, but its velocity depends on the full configuration through the KDE and its score. In this sense, the dynamics are structurally similar to SVGD (Liu & Wang, 2016; Liu, 2017), where particles also evolve deterministically through an empirical-measure-dependent velocity field; the key difference is that the conservative drifting velocity is the KDE-score residual $s_\rho - s_x$, rather than the kernelized Stein velocity used in SVGD.

Let $p_t^N$ denote the joint density of $X(t) = (X_1(t), \dots, X_N(t))$. We measure the joint law against the product target density

$$\rho^{\otimes N}(x) := \prod_{i=1}^{N} \rho(x_i)$$

through the joint relative entropy

$$H_N(t) := \mathrm{KL}(p_t^N \| \rho^{\otimes N}) = \int_{(\mathbb{R}^d)^N} p_t^N(x) \log \frac{p_t^N(x)}{\rho^{\otimes N}(x)} \, \mathrm{d}x. \tag{26}$$

For a vector field $f : \mathbb{R}^d \to \mathbb{R}^d$, define the Stein divergence (Barbour, 1988; Gorham & Mackey, 2015) associated with $\rho$ by

$$\mathcal{A}_\rho f(z) := \nabla \cdot f(z) + s_\rho(z) \cdot f(z). \tag{27}$$

For a configuration $x$, define the empirical Stein drift

$$\mathsf{S}_N(x) := \frac{1}{N} \sum_{i=1}^{N} \mathcal{A}_\rho b_x(x_i), \tag{28}$$

the KDE-smoothed Fisher discrepancy

$$\mathsf{I}_N(x) := \int_{\mathbb{R}^d} \|b_x(z)\|^2 \, q_x(z) \, \mathrm{d}z, \tag{29}$$

and the squared center velocity

$$\mathsf{V}_N(x) := \frac{1}{N} \sum_{i=1}^{N} \|b_x(x_i)\|^2 . \tag{30}$$

Finally set

$$\mathsf{R}_N(x) := \frac{1}{N} \sum_{i=1}^{N} \frac{1}{q_x(x_i)}. \tag{31}$$

### C.2 Assumptions

The following assumptions are used in the entropy calculation and in the conversion from empirical Stein drift to Fisher-type discrepancies.

**Assumption C.1** (Kernel). *The base kernel $K$ is positive, even, normalized, and $C^3$:*

$$K(u) > 0, \qquad K(u) = K(-u), \qquad \int_{\mathbb{R}^d} K(u) \, \mathrm{d}u = 1.$$

*Moreover*

$$m_2(K) := \int_{\mathbb{R}^d} \|u\|^2 \, K(u) \, \mathrm{d}u < \infty,$$

*and $\Delta K(0)$ is finite. Since $K$ is even, $\nabla K(0) = 0$.*

**Assumption C.2** (Target and regular flow)**.** *The density $\rho$ is positive and sufficiently smooth so that $s_\rho$ is $C^2$. The ODE (25) has a unique global solution for the initial law under consideration. The joint density $p_t^N$ is differentiable in $t$ and solves the Liouville equation*

$$\partial_t p_t^N(x) + \sum_{i=1}^N \nabla_{x_i} \cdot \{p_t^N(x) b_i(x)\} = 0.$$

*All integrals appearing below are finite, and the boundary terms in the integrations by parts on $(\mathbb{R}^d)^N$ vanish. A sufficient, but not necessary, condition is rapid decay of $p_t^N(x)b_i(x)$ and of the products obtained by multiplying $p_t^N(x)b_i(x)$ by $\log(p_t^N(x)/\rho^{\otimes N}(x))$. See Banerjee et al. (2025) for related conditions for SVGD.*

**Assumption C.3** (Reciprocal KDE control)**.** *For the time horizon $T > 0$ under consideration, there is a finite constant $\Lambda_T$ such that*

$$\sup_{0\le t\le T} \mathbb{E}_t[\mathsf{R}_N(X(t))] \le \Lambda_T, \tag{32}$$

*where $\mathbb{E}_t$ denotes expectation with respect to $p_t^N$ and $\mathsf{R}_N$ is defined in (31).*

**Remark C.4** (Conditional nature of reciprocal-KDE control)**.** *The reciprocal-KDE assumption is a stability assumption, not a consequence of the entropy identity. The local-occupancy results in Section F give deterministic or high-probability sufficient conditions under which*

$$q_{X(t)}(X_i(t)) \ge \lambda > 0 \qquad \text{for all } i$$

*at a fixed time, and under a propagated occupancy condition. However, propagation arguments that use a flow Lipschitz constant are conditional: the Lipschitz constant of the conservative vector field itself depends on reciprocal KDE quantities such as $1/q_x$, and sometimes on higher inverse powers of $q_x$. Thus there is a potential bootstrap structure.*

*The stability condition is imposed on the full deterministic interval $[0, T]$. A lower-denominator region can be used in a bootstrap argument if its invariance is proved separately. Likewise, a high-probability occupancy event can be combined with a bound on its complement to control the unconditional expectation in (32). The entropy calculation uses the original evolving joint density on this interval. Stopping particles at the boundary of a denominator region may create a singular law, so such a stopped dynamics requires its own transport and boundary analysis.*

**Assumption C.5** (Quadrature smoothness)**.** *For the time horizon $T > 0$, there are finite constants $B_{A,T}(h)$ and $B_{V,T}(h)$ such that, for every configuration $X(t)$ visited by the flow for $0 \le t \le T$,*

$$\sup_{z\in\mathbb{R}^d} \left\| D^2\{\mathcal{A}_\rho b_{X(t)}\}(z) \right\|_{\mathrm{op}} \le B_{A,T}(h), \tag{33}$$

$$\sup_{z\in\mathbb{R}^d} \left\| D^2\{\left\|b_{X(t)}\right\|^2\}(z) \right\|_{\mathrm{op}} \le B_{V,T}(h). \tag{34}$$

*Here $D^2$ denotes the Hessian, and $\|\cdot\|_{\mathrm{op}}$ is the operator norm.*

**Remark C.6** (Bandwidth dependence of the quadrature constants)**.** *The constants $B_{A,T}(h)$ and $B_{V,T}(h)$ may depend on $N, h, T, \rho, K$, and on the region of state space visited by the dynamics. This dependence is not a lower-order issue. Writing*

$$r_x(z) := \log \rho(z) - \log q_x(z), \qquad b_x(z) = \nabla r_x(z),$$

*we have*

$$\mathcal{A}_\rho b_x = \Delta r_x + \nabla \log \rho \cdot \nabla r_x,$$

*and*

$$D^2 \left\|b_x\right\|^2 = 2(Db_x)^\top (Db_x) + 2\sum_{\ell=1}^d b_{x,\ell} D^2 b_{x,\ell}.$$

*Thus the quadrature constants involve log-derivatives of $\rho$ and $q_x$ up to fourth order. If these log-derivatives vary on the kernel scale, for example if*

$$\sup_z \left\| D^j(\log \rho - \log q_x)(z) \right\| = O(h^{-j}), \qquad j = 1, 2, 3, 4,$$

*and the target-score factors have compatible derivative bounds, then* $B_{A,T}(h) + B_{V,T}(h) = O(h^{-4})$. *The target-score bounds are needed because differentiating* $\nabla \log \rho \cdot \nabla r_x$ *also differentiates* $\nabla \log \rho$. *In that regime the quadrature contribution*

$$\{B_{A,T}(h) + B_{V,T}(h)\} m_2(K_h) = O(h^{-2})$$

*scales like* $h^{-2}$, *not* $h^2$, *and shrinking the bandwidth worsens this term. Consequently, the rates below are non-asymptotic inequalities with the quadrature constants displayed. Any asymptotic bandwidth optimization must explicitly account for the* $h$*-dependence of* $B_{A,T}(h)$ *and* $B_{V,T}(h)$.

**Remark C.7** (Sufficient derivative bounds for quadrature growth)**.** *The exponent* $\beta$ *describes the regularity of the drift along the trajectory, including its interaction with the target score. For example, suppose* $0 < \ell_h \leq 1$ *and, uniformly in the configuration, time, and spatial variable,*

$$\left\|D^j r_x\right\|_{\mathrm{op}} \leq C_r \ell_h^{-j}, \quad j = 1, 2, 3, 4, \qquad \left\|D^k s_\rho\right\|_{\mathrm{op}} \leq C_\rho, \quad k = 0, 1, 2.$$

*Here norms of multilinear derivatives use the Euclidean norm in each argument and in the range. For unit vectors* $v, w$, *the product rule gives*

$$\begin{aligned} D^2(s_\rho \cdot \nabla r_x)[v, w] = {} & D^2 s_\rho[v, w] \cdot \nabla r_x + D s_\rho[v] \cdot D^2 r_x[w] \\ & + D s_\rho[w] \cdot D^2 r_x[v] + s_\rho \cdot D^3 r_x[v, w]. \end{aligned}$$

*The four terms have absolute values bounded, respectively, by* $C_\rho C_r \ell_h^{-1}$, $C_\rho C_r \ell_h^{-2}$, $C_\rho C_r \ell_h^{-2}$, *and* $C_\rho C_r \ell_h^{-3}$. *Also* $\left\|D^2 \Delta r_x\right\|_{\mathrm{op}} \leq d C_r \ell_h^{-4}$, *because the Laplacian sums over* $d$ *coordinate directions. Thus*

$$\left\|D^2\{\mathcal{A}_\rho b_x\}\right\|_{\mathrm{op}} \leq (d C_r + 4 C_\rho C_r) \ell_h^{-4}.$$

*Differentiating* $\|\nabla r_x\|^2$ *twice gives*

$$D^2 \|\nabla r_x\|^2 [v, w] = 2 D^2 r_x[v] \cdot D^2 r_x[w] + 2 \nabla r_x \cdot D^3 r_x[v, w].$$

*Each term is bounded by* $2C_r^2 \ell_h^{-4}$, *so* $B_{A,N}(h) + B_{V,N}(h) = O(\ell_h^{-4})$. *Taking* $\ell_h = 1$ *gives the uniform regime; taking* $\ell_h = h^\alpha$ *with* $0 < \alpha < 1/2$ *gives* $\beta = 4\alpha < 2$. *These are sufficient hypotheses to be verified for the target and trajectory under consideration.*

*Global bounded-score conditions do not hold for every smooth kernel. For a finite Gaussian KDE,*

$$s_x(z) = h^{-2} \left\{ \sum_j w_j(z) x_j - z \right\}, \qquad w_j(z) \geq 0, \quad \sum_j w_j(z) = 1.$$

*Writing* $R_x = \max_j \|x_j\|$, *the triangle inequality gives* $\|s_x(z)\| \geq h^{-2}(\|z\| - R_x)$. *Thus its first log derivative is unbounded on* $\mathbb{R}^d$. *In such settings, the scalar integrand bounds in Assumption C.5 must be verified directly, using any cancellations available. Local derivative bounds can also be used when the corresponding quadrature tails are included.*

Finally, we emphasize that our conservative drifting analysis assumes that the smoothing kernel is sufficiently smooth, for example $K \in C^3$ or stronger when the quadrature constants involve higher derivatives. This assumption excludes the exact Laplace kernel $K_h(u) = c_d h^{-d} \exp(-\|u\|/h)$, which is not differentiable at the origin. This is why the exact Laplace kernel is treated separately in Section G through the non-conservative displacement field and the sharp-companion-kernel decomposition. A smooth regularization of the Laplace kernel could be included in the conservative analysis, but the resulting estimates would have constants depending on the regularization scale.

The forthcoming results do not require the target to be log-concave, strongly log-concave, or to satisfy a log-Sobolev, Poincaré, or dissipativity condition. They control time-averaged residual drift and smoothed Fisher discrepancy under the stated regularity assumptions. Converting these quantities into global KL or Wasserstein convergence requires additional functional inequalities or stability estimates.

## D Proof of the Conservative Entropy Identity

The next theorem is the basic finite-particle identity. Differentiating the particle's own KDE contribution produces a self-interaction term in the joint-entropy calculation.

**Theorem D.1** (Joint-entropy identity)**.** *Let Assumptions C.1 and C.2 hold. Then, for every $t$ for which the quantities are finite,*

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N(t) = -N\mathbb{E}_t[\mathsf{S}_N(X(t))] - \Delta K_h(0)\,\mathbb{E}_t[\mathsf{R}_N(X(t))]. \tag{35}$$

*Equivalently, with*

$$a_h := \{-\Delta K_h(0)\}_+,$$

*one has the upper bound*

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N(t) \le -N\mathbb{E}_t[\mathsf{S}_N(X(t))] + a_h\mathbb{E}_t[\mathsf{R}_N(X(t))]. \tag{36}$$

*Proof.* Let

$$r_N(x) := \rho^{\otimes N}(x).$$

By definition,

$$H_N(t) = \int p_t^N(x)\log\frac{p_t^N(x)}{r_N(x)}\,\mathrm{d}x.$$

Differentiating under the integral gives

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N(t) = \int \partial_t p_t^N(x)\left\{\log\frac{p_t^N(x)}{r_N(x)} + 1\right\}\,\mathrm{d}x.$$

Since $p_t^N$ is a probability density,

$$\int \partial_t p_t^N(x)\,\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}t}\int p_t^N(x)\,\mathrm{d}x = 0.$$

Thus

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N(t) = \int \partial_t p_t^N(x)\log\frac{p_t^N(x)}{r_N(x)}\,\mathrm{d}x.$$

Using the Liouville equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N(t) = -\sum_{i=1}^N\int \nabla_{x_i}\cdot\{p_t^N(x)b_i(x)\}\log\frac{p_t^N(x)}{r_N(x)}\,\mathrm{d}x.$$

Integrating by parts in $x_i$,

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N(t) = \sum_{i=1}^N\int p_t^N(x)b_i(x)\cdot\nabla_{x_i}\log\frac{p_t^N(x)}{r_N(x)}\,\mathrm{d}x.$$

Now

$$\nabla_{x_i}\log\frac{p_t^N(x)}{r_N(x)} = \nabla_{x_i}\log p_t^N(x) - \nabla_{x_i}\log r_N(x).$$

Since $r_N(x) = \prod_{j=1}^N \rho(x_j)$,

$$\nabla_{x_i}\log r_N(x) = s_\rho(x_i).$$

Therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N(t) = \sum_{i=1}^N\int p_t^N b_i\cdot\nabla_{x_i}\log p_t^N\,\mathrm{d}x - \sum_{i=1}^N\int p_t^N b_i\cdot s_\rho(x_i)\,\mathrm{d}x.$$

For the first integral,

$$p_t^N\nabla_{x_i}\log p_t^N = \nabla_{x_i}p_t^N.$$

Hence

$$\int p_t^N b_i \cdot \nabla_{x_i} \log p_t^N \, \mathrm{d}x = \int b_i \cdot \nabla_{x_i} p_t^N \, \mathrm{d}x = -\int p_t^N \nabla_{x_i} \cdot b_i \, \mathrm{d}x.$$

Consequently,

$$\frac{\mathrm{d}}{\mathrm{d}t} H_N(t) = -\mathbb{E}_t \left[ \sum_{i=1}^N \{ \nabla_{x_i} \cdot b_i(X(t)) + s_\rho(X_i(t)) \cdot b_i(X(t)) \} \right]. \tag{37}$$

It remains to compute $\nabla_{x_i} \cdot b_i(x)$. Since

$$b_i(x) = s_\rho(x_i) - s_x(x_i),$$

we have

$$\nabla_{x_i} \cdot b_i(x) = \nabla \cdot s_\rho(x_i) - \nabla_{x_i} \cdot s_x(x_i).$$

The derivative of $s_x(x_i)$ with respect to $x_i$ contains two pieces: the derivative of the evaluation point and the derivative of the $i$th KDE center. Write

$$s_x^m(z) = \frac{\partial_m q_x(z)}{q_x(z)}, \qquad m = 1, \ldots, d.$$

For fixed $z$,

$$\partial_{x_i^\ell} q_x(z) = -\frac{1}{N} \partial_\ell K_h(z - x_i),$$

and

$$\partial_{x_i^\ell} \partial_m q_x(z) = -\frac{1}{N} \partial_{\ell m}^2 K_h(z - x_i).$$

Therefore

$$\partial_{x_i^\ell} s_x^m(z) = -\frac{1}{N} \frac{\partial_{\ell m}^2 K_h(z - x_i)}{q_x(z)} + \frac{1}{N} \frac{\partial_m q_x(z) \partial_\ell K_h(z - x_i)}{q_x(z)^2}.$$

Set $m = \ell$ and sum over $\ell$. Since $K_h$ is even, $\nabla K_h(0) = 0$, and hence at $z = x_i$,

$$\sum_{\ell=1}^d \partial_{x_i^\ell} s_x^\ell(z) \big|_{z=x_i} = -\frac{1}{N} \frac{\Delta K_h(0)}{q_x(x_i)}.$$

By the chain rule,

$$\nabla_{x_i} \cdot s_x(x_i) = \nabla_z \cdot s_x(z) \big|_{z=x_i} - \frac{1}{N} \frac{\Delta K_h(0)}{q_x(x_i)}.$$

Thus

$$\begin{aligned} \nabla_{x_i} \cdot b_i(x) &= \nabla \cdot s_\rho(x_i) - \nabla_z \cdot s_x(z) \big|_{z=x_i} + \frac{1}{N} \frac{\Delta K_h(0)}{q_x(x_i)} \\ &= \nabla_z \cdot b_x(z) \big|_{z=x_i} + \frac{1}{N} \frac{\Delta K_h(0)}{q_x(x_i)}. \end{aligned}$$

Since $b_i(x) = b_x(x_i)$,

$$\nabla_{x_i} \cdot b_i(x) + s_\rho(x_i) \cdot b_i(x) = \mathcal{A}_\rho b_x(x_i) + \frac{1}{N} \frac{\Delta K_h(0)}{q_x(x_i)}.$$

Substituting this identity into (37) gives

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} H_N(t) &= -\mathbb{E}_t \left[ \sum_{i=1}^N \mathcal{A}_\rho b_{X(t)}(X_i(t)) + \frac{\Delta K_h(0)}{N} \sum_{i=1}^N \frac{1}{q_{X(t)}(X_i(t))} \right] \\ &= -N \mathbb{E}_t[\mathsf{S}_N(X(t))] - \Delta K_h(0) \mathbb{E}_t[\mathsf{R}_N(X(t))]. \end{aligned}$$

This proves (35). Since

$$-\Delta K_h(0) \le \{-\Delta K_h(0)\}_+ = a_h,$$

(36) follows. □

**Remark D.2** (Interpretation of the self-interaction term)**.** *The first term on the right-hand side of (35) is the desired entropy dissipation. The second term is finite-particle specific: it appears because the velocity at $x_i$ depends on the same particle through the KDE denominator. Its sign and size are controlled by $-\Delta K_h(0)$ and by the reciprocal quantity $\mathsf{R}_N$. This is why the subsequent rate theorem requires reciprocal-KDE control. If particles are locally well populated at bandwidth $h$, then $q_x(x_i)$ is bounded below and the self-interaction correction remains of order $1/N$ after the entropy normalization.*

*For Gaussian kernels, the self-interaction correction has a definite positive sign. Indeed, for $K(u) = (2\pi)^{-d/2}\exp(-\|u\|^2/2)$, we have, for each coordinate $j$, $\partial_{jj}K(0) = -K(0) = -(2\pi)^{-d/2}$. Therefore*

$$\Delta K(0) = \sum_{j=1}^{d} \partial_{jj}K(0) = -d(2\pi)^{-d/2} < 0,$$

*and hence $a_1 := \{-\Delta K(0)\}_+ = d(2\pi)^{-d/2} > 0$.*

*For the bandwidth-scaled kernel $K_h(u) = h^{-d}K(u/h)$,*

$$-\Delta K_h(0) = h^{-d-2}a_1.$$

*Thus the reciprocal-KDE term in the conservative finite-particle bound is a genuine positive self-interaction correction for Gaussian kernels.*

# E Proofs and Further Results for Conservative Rates

## E.1 Rate for the Empirical Stein Drift

**Theorem E.1** (Entropy rate for the empirical Stein drift)**.** *Let Assumptions C.1, C.2, and C.3 hold. Then, for every $T > 0$,*

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{S}_N(X(t))]\,\mathrm{d}t \le \frac{H_N(0)}{NT} + \frac{a_h\Lambda_T}{N}, \qquad a_h := \{-\Delta K_h(0)\}_+. \tag{38}$$

*Proof.* By Theorem D.1,

$$H_N'(t) \le -N\mathbb{E}_t[\mathsf{S}_N(X(t))] + a_h\mathbb{E}_t[\mathsf{R}_N(X(t))].$$

Using Assumption C.3,

$$H_N'(t) \le -N\mathbb{E}_t[\mathsf{S}_N(X(t))] + a_h\Lambda_T.$$

Rearrange:

$$N\mathbb{E}_t[\mathsf{S}_N(X(t))] \le -H_N'(t) + a_h\Lambda_T.$$

Integrate over $t \in [0, T]$:

$$N\int_0^T \mathbb{E}_t[\mathsf{S}_N(X(t))]\,\mathrm{d}t \le H_N(0) - H_N(T) + a_h\Lambda_T T.$$

Since relative entropy is nonnegative,

$$H_N(0) - H_N(T) \le H_N(0).$$

Therefore

$$N\int_0^T \mathbb{E}_t[\mathsf{S}_N(X(t))]\,\mathrm{d}t \le H_N(0) + a_h\Lambda_T T.$$

Dividing by $NT$ proves (38). □

**Remark E.2** (What the Stein-drift rate controls)**.** *The quantity $\mathsf{S}_N$ is the empirical Stein drift associated with the vector field $b_x$. It is not written as a square, so (38) is not yet a residual-velocity bound. The role of the quadrature step below is to compare this empirical Stein drift with the KDE-averaged Fisher discrepancy, which is nonnegative and is directly related to the squared velocity of the particles.*

### E.2 Quadrature Conversion to Fisher and Center Velocity

The empirical Stein drift $\mathsf{S}_N$ is not manifestly nonnegative. The next lemma identifies the nonnegative population quantity that it approximates.

**Lemma E.3** (KDE-averaged Stein identity)**.** *For every fixed configuration $x$ for which the integrations by parts are justified,*

$$\int_{\mathbb{R}^d} \mathcal{A}_\rho b_x(z) q_x(z)\,\mathrm{d}z = \mathsf{I}_N(x). \tag{39}$$

*Proof.* By definition of $\mathcal{A}_\rho$,

$$\int q_x \mathcal{A}_\rho b_x = \int q_x \nabla \cdot b_x + \int q_x s_\rho \cdot b_x.$$

Integrating the first term by parts,

$$\int q_x \nabla \cdot b_x = -\int \nabla q_x \cdot b_x.$$

Since $\nabla q_x = q_x s_x$,

$$-\int \nabla q_x \cdot b_x = -\int q_x s_x \cdot b_x.$$

Thus

$$\int q_x \mathcal{A}_\rho b_x = \int q_x (s_\rho - s_x) \cdot b_x.$$

Because $b_x = s_\rho - s_x$,

$$\int q_x \mathcal{A}_\rho b_x = \int q_x \left\|b_x\right\|^2 = \mathsf{I}_N(x).$$

This proves the lemma. □

**Lemma E.4** (Point-evaluation quadrature error)**.** *Let $K_h$ be even and have finite second moment*

$$m_2(K_h) := \int_{\mathbb{R}^d} \|u\|^2 K_h(u)\,\mathrm{d}u.$$

*Under Assumption C.5, for every configuration $X(t)$ visited by the flow,*

$$|\mathsf{S}_N(X(t)) - \mathsf{I}_N(X(t))| \le \frac{1}{2} B_{A,T}(h) m_2(K_h), \tag{40}$$

$$|\mathsf{V}_N(X(t)) - \mathsf{I}_N(X(t))| \le \frac{1}{2} B_{V,T}(h) m_2(K_h). \tag{41}$$

*Proof.* We prove (40); the proof of (41) is identical with $\|b_x\|^2$ replacing $\mathcal{A}_\rho b_x$.

Fix $x$ and define

$$\phi_x(z) := \mathcal{A}_\rho b_x(z).$$

By Lemma E.3,

$$\mathsf{I}_N(x) = \int q_x(z) \phi_x(z)\,\mathrm{d}z.$$

Since

$$q_x(z) = \frac{1}{N} \sum_{i=1}^N K_h(z - x_i),$$

we have

$$\mathsf{I}_N(x) = \frac{1}{N} \sum_{i=1}^N \int K_h(z - x_i) \phi_x(z)\,\mathrm{d}z.$$

Meanwhile,

$$\mathsf{S}_N(x) = \frac{1}{N} \sum_{i=1}^N \phi_x(x_i).$$

Therefore

$$\mathsf{S}_N(x) - \mathsf{I}_N(x) = \frac{1}{N}\sum_{i=1}^{N}\left\{\phi_x(x_i) - \int K_h(z - x_i)\phi_x(z)\,\mathrm{d}z\right\}.$$

Set $u = z - x_i$. Then

$$\int K_h(z - x_i)\phi_x(z)\,\mathrm{d}z = \int K_h(u)\phi_x(x_i + u)\,\mathrm{d}u.$$

Taylor's formula gives

$$\phi_x(x_i + u) = \phi_x(x_i) + \nabla\phi_x(x_i)\cdot u + \int_0^1 (1 - r)u^\top D^2\phi_x(x_i + ru)u\,\mathrm{d}r.$$

Integrating against $K_h(u)\,\mathrm{d}u$, the zeroth-order term gives $\phi_x(x_i)$ because $\int K_h = 1$. The first-order term vanishes because $K_h$ is even, hence $\int uK_h(u)\,\mathrm{d}u = 0$. Thus

$$\left|\int K_h(u)\phi_x(x_i + u)\,\mathrm{d}u - \phi_x(x_i)\right|$$
$$\leq \int K_h(u)\int_0^1 (1 - r)\left\|D^2\phi_x(x_i + ru)\right\|_{\mathrm{op}}\|u\|^2\,\mathrm{d}r\,\mathrm{d}u.$$

Using Assumption C.5,

$$\left|\int K_h(u)\phi_x(x_i + u)\,\mathrm{d}u - \phi_x(x_i)\right| \leq B_{A,T}(h)\int_0^1 (1 - r)\,\mathrm{d}r\int \|u\|^2 K_h(u)\,\mathrm{d}u.$$

Since $\int_0^1 (1 - r)\,\mathrm{d}r = 1/2$,

$$\left|\int K_h(u)\phi_x(x_i + u)\,\mathrm{d}u - \phi_x(x_i)\right| \leq \frac{1}{2}B_{A,T}(h)m_2(K_h).$$

Averaging over $i = 1, \ldots, N$ proves (40). □

**Theorem E.5** (Continuous-time rates for Fisher discrepancy and center velocity)**.** *Let Assumptions C.1, C.2, C.3, and C.5 hold. Then, for every $T > 0$,*

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{I}_N(X(t))]\,\mathrm{d}t \leq \frac{H_N(0)}{NT} + \frac{a_h\Lambda_T}{N} + \frac{1}{2}B_{A,T}(h)m_2(K_h), \tag{42}$$

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t \leq \frac{H_N(0)}{NT} + \frac{a_h\Lambda_T}{N} + \frac{1}{2}\{B_{A,T}(h) + B_{V,T}(h)\}m_2(K_h). \tag{43}$$

*Proof.* From Lemma E.4,

$$\mathsf{I}_N(X(t)) \leq \mathsf{S}_N(X(t)) + \frac{1}{2}B_{A,T}(h)m_2(K_h).$$

Taking expectations, integrating over $[0, T]$, dividing by $T$, and applying Theorem E.1 gives (42).

Similarly,

$$\mathsf{V}_N(X(t)) \leq \mathsf{I}_N(X(t)) + \frac{1}{2}B_{V,T}(h)m_2(K_h).$$

Combining this inequality with (42) gives (43). □

**Remark E.6** (Implication for one-step generation)**.** *Theorem E.5 bounds the average squared speed of the generated particle cloud under the conservative drift. If a time $\tau$ is sampled uniformly from $[0, T]$, then*

$$\mathbb{E}\left[\mathsf{V}_N(X(\tau))\right] = \frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t.$$

*Consequently, a small right-hand side in (43) implies that, at a typical training time, one more explicit drift step changes the empirical generated law by at most*

$$\mathbb{E}\left[W_2^2\left(\mu^N_{X(\tau)}, \frac{1}{N}\sum_{i=1}^{N}\delta_{X_i(\tau)+\eta b_{X(\tau)}(X_i(\tau))}\right)\right] \leq \eta^2\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t.$$

*Thus the theorem is a convergence-to-small-residual-drift statement for one-step generation. With the quadrature comparison, it also controls the smoothed Fisher discrepancy between $q_x$ and $\rho$.*

**Corollary E.7** (Bandwidth form). *Let $K_h(u) = h^{-d}K(u/h)$ and let the assumptions of Theorem E.5 hold. Then*

$$m_2(K_h) = h^2 m_2(K), \qquad \Delta K_h(0) = h^{-d-2}\Delta K(0).$$

*Consequently, with $a_1 := \{-\Delta K(0)\}_+$,*

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t \le \frac{H_N(0)}{NT} + \frac{a_1\Lambda_T}{Nh^{d+2}} + \frac{1}{2}\{B_{A,T}(h) + B_{V,T}(h)\}m_2(K)h^2. \tag{44}$$

*If, in addition, $H_N(0) \le \kappa_0 N$ and $T = N$, then*

$$\frac{1}{N}\int_0^N \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t \le \frac{\kappa_0}{N} + \frac{a_1\Lambda_N}{Nh^{d+2}} + \frac{1}{2}\{B_{A,N}(h) + B_{V,N}(h)\}m_2(K)h^2. \tag{45}$$

*Thus the corresponding root-discrepancy bound is*

$$\left(\frac{1}{N}\int_0^N \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t\right)^{1/2} \le \sqrt{\frac{\kappa_0}{N}} + \sqrt{\frac{a_1\Lambda_N}{Nh^{d+2}}} + h\sqrt{\frac{1}{2}\{B_{A,N}(h) + B_{V,N}(h)\}m_2(K)}. \tag{46}$$

*Proof.* The scaling identities follow by the change of variables $u = hv$:

$$m_2(K_h) = \int \|u\|^2 h^{-d}K(u/h)\,\mathrm{d}u = h^2\int \|v\|^2 K(v)\,\mathrm{d}v = h^2 m_2(K).$$

Also,

$$\Delta K_h(u) = h^{-d-2}(\Delta K)(u/h),$$

so $\Delta K_h(0) = h^{-d-2}\Delta K(0)$. Substituting these identities into (43) gives (44). If $H_N(0) \le \kappa_0 N$ and $T = N$, then

$$\frac{H_N(0)}{NT} \le \frac{\kappa_0 N}{N^2} = \frac{\kappa_0}{N},$$

which gives (45). Finally, (46) follows from $\sqrt{a+b+c} \le \sqrt{a} + \sqrt{b} + \sqrt{c}$, for $a, b, c \ge 0$. □

### E.3 Optimizing the Conservative Bandwidth and The One-step Size

The bandwidth form (45) separates three effects: the entropy initialization term, the finite-particle self-interaction term, and the quadrature term. Since the quadrature constants may depend on $h$, we write them explicitly as functions of the bandwidth. Define

$$A_N(h) := a_1\Lambda_N(h), \qquad C_N(h) := \frac{1}{2}\{B_{A,N}(h) + B_{V,N}(h)\}m_2(K).$$

Then (45) reads

$$\frac{1}{N}\int_0^N \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t \le \frac{\kappa_0}{N} + \frac{A_N(h)}{Nh^{d+2}} + C_N(h)h^2. \tag{47}$$

Thus the commonly stated bandwidth tradeoff is valid only after specifying how $A_N(h)$ and $C_N(h)$ scale with $h$.

Assume that, on the admissible bandwidth range, there are constants $A, C > 0$ and $0 \le \beta < 2$ such that

$$A_N(h) \le A, \qquad C_N(h) \le Ch^{-\beta}. \tag{48}$$

Equivalently, the quadrature constants satisfy

$$B_{A,N}(h) + B_{V,N}(h) = O(h^{-\beta}).$$

Then (47) implies

$$\frac{1}{N}\int_0^N \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t \le \frac{\kappa_0}{N} + \frac{A}{Nh^{d+2}} + Ch^{2-\beta}. \tag{49}$$

The last two terms are minimized at

$$h_{\star,N} := \left( \frac{(d+2)A}{(2-\beta)CN} \right)^{1/(d+4-\beta)}. \tag{50}$$

Indeed, differentiating

$$h \mapsto \frac{A}{Nh^{d+2}} + Ch^{2-\beta}$$

gives

$$-\frac{(d+2)A}{Nh^{d+3}} + (2-\beta)Ch^{1-\beta}.$$

Setting this derivative equal to zero gives (50). At this bandwidth,

$$\frac{A}{Nh_{\star,N}^{d+2}} = \frac{2-\beta}{d+2} Ch_{\star,N}^{2-\beta}.$$

Therefore

$$\frac{1}{N}\int_0^N \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t \le \frac{\kappa_0}{N} + \frac{d+4-\beta}{d+2} C \left( \frac{(d+2)A}{(2-\beta)CN} \right)^{(2-\beta)/(d+4-\beta)}. \tag{51}$$

Thus, under (48), the squared residual velocity is

$$O\left(N^{-(2-\beta)/(d+4-\beta)}\right),$$

and the root residual velocity is

$$O\left(N^{-(2-\beta)/(2(d+4-\beta))}\right).$$

The $h$-uniform quadrature case is the special case $\beta = 0$, which recovers the root rate $N^{-1/(d+4)}$. If $\beta \ge 2$, the quadrature term $h^{2-\beta}$ does not vanish as $h \to 0$, and this quadrature argument does not yield a vanishing residual by bandwidth shrinkage. For example, the natural kernel-scale derivative scaling $B_{A,N}(h) + B_{V,N}(h) = O(h^{-4})$ corresponds to $\beta = 4$ and falls outside the regime of the optimized vanishing-rate statement.

The occupancy condition remains compatible with (50) whenever $Nh_{\star,N}^d \to \infty$ sufficiently quickly. For $0 \le \beta < 2$,

$$Nh_{\star,N}^d \asymp N^{1-d/(d+4-\beta)} = N^{(4-\beta)/(d+4-\beta)},$$

which diverges faster than $\log N$ for fixed $d$ and fixed $\beta < 2$.

The explicit one-step size $\eta$ does not enter the continuous-time entropy inequality. It enters when the learned generator is interpreted through one more explicit correction $z \mapsto z + \eta b_x(z)$. For a fixed $h$, the theorem gives

$$\mathbb{E}\left[W_2^2(\mu_x^N, \widetilde{\mu}_x^N)\right] \le \eta^2 R_{N,h},$$

where $R_{N,h}$ denotes the right-hand side of (47). Taking $h = h_{\star,N}$ gives a residual correction of order

$$\eta\, N^{-(2-\beta)/(2(d+4-\beta))}$$

in root mean square, up to constants. For a fixed configuration, if $b_x$ has Lipschitz constant $L_b(h)$, then the map $z \mapsto z + \eta b_x(z)$ satisfies

$$\|\{z + \eta b_x(z)\} - \{z' + \eta b_x(z')\} - (z - z')\| \le \eta L_b(h)\, \|z - z'\|.$$

Thus $\eta L_b(h) \le c < 1$ controls the deformation of a single frozen-field step. Under an estimate $L_b(h) = O(h^{-2})$, this suggests

$$\eta_{\star,N} \asymp h_{\star,N}^2 \asymp N^{-2/(d+4-\beta)}.$$

For Gaussian kernels, the displacement velocity satisfies $u^{\mathrm{disp}} = h^2 b$, so this conservative-step choice corresponds to a constant-order step size for the displacement formulation. Approximating the interacting-particle ODE over a time horizon also requires control of the full configuration vector field and the accumulated error; see Section H.

**Remark E.8** (Initialization)**.** *Suppose the initial particles are independent with common density $\mu_0$. If $\mathrm{KL}(\mu_0\|\rho) < \infty$, then*

$$p_0^N = \mu_0^{\otimes N} \quad \Longrightarrow \quad H_N(0) = N\mathrm{KL}(\mu_0\|\rho).$$

*Thus the condition $H_N(0) \le \kappa_0 N$ holds with $\kappa_0 = \mathrm{KL}(\mu_0\|\rho)$.*

# F SUFFICIENT CONDITIONS FOR RECIPROCAL KDE CONTROL ON $\mathbb{R}^d$

The reciprocal-KDE condition (32) is needed because the center-evaluation divergence differentiates the $i$th KDE center and produces a factor $1/q_x(x_i)$. This section gives checkable sufficient conditions. The occupancy estimates in this section are elementary and are included for completeness. The deterministic implication is simply that if every particle has enough neighbors in an $O(h)$-ball, then the KDE denominator at that particle is bounded below. The high-probability version follows from the standard binomial Chernoff lower-tail bound, applied conditionally on each particle location, together with a union bound over particles. Such local-neighborhood occupancy arguments are also standard in random geometric graph theory (Penrose, 2003). We use these estimates only as sufficient conditions for reciprocal-KDE control; the finite-particle entropy bounds themselves remain conditional on the stated reciprocal-denominator assumptions.

## F.1 DETERMINISTIC LOCAL-OCCUPANCY CONDITION

**Assumption F.1** (Kernel lower bound near the origin)**.** *There exist constants $r_K > 0$ and $\kappa_K > 0$ such that*

$$K(u) \geq \kappa_K \qquad \text{whenever } \|u\| \leq r_K. \tag{52}$$

*Equivalently,*

$$K_h(u) \geq \kappa_K h^{-d} \qquad \text{whenever } \|u\| \leq r_K h.$$

For a configuration $x$, define the local neighbor count

$$M_i(x;h) := \#\{j \in \{1,\dots,N\} : \|x_j - x_i\| \leq r_K h\}. \tag{53}$$

The count includes $j = i$.

**Proposition F.2** (Local occupancy implies reciprocal control)**.** *Let Assumption F.1 hold. Suppose that a configuration $x$ satisfies*

$$M_i(x;h) \geq \alpha N h^d \qquad \text{for every } i = 1,\dots,N \tag{54}$$

*for some $\alpha > 0$. Then*

$$q_x(x_i) \geq \kappa_K \alpha \qquad \text{for every } i = 1,\dots,N, \tag{55}$$

*and hence*

$$\mathsf{R}_N(x) \leq \frac{1}{\kappa_K \alpha}. \tag{56}$$

*Proof.* Fix $i$. By definition,

$$q_x(x_i) = \frac{1}{N}\sum_{j=1}^{N} K_h(x_i - x_j).$$

For every $j$ with $\|x_j - x_i\| \leq r_K h$, Assumption F.1 gives

$$K_h(x_i - x_j) \geq \kappa_K h^{-d}.$$

Therefore

$$q_x(x_i) \geq \frac{1}{N} M_i(x;h)\kappa_K h^{-d}.$$

Using (54),

$$q_x(x_i) \geq \frac{1}{N}(\alpha N h^d)\kappa_K h^{-d} = \kappa_K \alpha.$$

Taking reciprocals and averaging over $i$ yields (56). □

### F.2 Expectation and High-probability Bounds for I.I.D. Particles

The previous proposition is deterministic. For i.i.d. samples, an exact binomial identity controls the expected reciprocal KDE, while a lower-tail estimate controls all local counts simultaneously.

**Assumption F.3** (Uniform lower local mass)**.** *Let $\mu$ be a probability measure on $\mathbb{R}^d$. For the bandwidth $h$ under consideration, there is a constant $p_0 > 0$ such that*

$$\mu(B(y, r_K h)) \geq p_0 h^d \qquad \text{for } \mu\text{-a.e. } y. \tag{57}$$

**Remark F.4** (One way to verify Assumption F.3)**.** *Suppose $\mu$ has density $f$ on a closed set $S \subset \mathbb{R}^d$, $f(y) \geq m_0 > 0$ for $y \in S$, and $S$ satisfies the thickness condition*

$$\mathrm{Leb}(S \cap B(y, r)) \geq \theta v_d r^d \qquad \text{for all } y \in S \text{ and } 0 < r \leq r_0,$$

*where $v_d$ is the volume of the unit Euclidean ball. If $r_K h \leq r_0$, then*

$$\mu(B(y, r_K h)) \geq m_0 \,\mathrm{Leb}(S \cap B(y, r_K h)) \geq m_0 \theta v_d r_K^d h^d.$$

*Thus Assumption F.3 holds with*

$$p_0 = m_0 \theta v_d r_K^d.$$

**Proposition F.5** (I.i.d. local occupancy)**.** *Let $Y_1, \ldots, Y_N$ be i.i.d. from a probability measure $\mu$ satisfying Assumption F.3. Let*

$$Y = (Y_1, \ldots, Y_N).$$

*Assume $N \geq 2$. Define the event*

$$\mathcal{E}_h := \left\{ M_i(Y; h) \geq \frac{p_0}{4} N h^d \text{ for every } i = 1, \ldots, N \right\}.$$

*Then*

$$\mathbb{P}(\mathcal{E}_h^c) \leq N \exp\left(-\frac{p_0}{16} N h^d\right). \tag{58}$$

*Consequently, under Assumption F.1, on $\mathcal{E}_h$,*

$$\mathsf{R}_N(Y) \leq \frac{4}{\kappa_K p_0}. \tag{59}$$

*Moreover, since $K_h(0) = h^{-d} K(0)$,*

$$\mathsf{R}_N(Y) \leq \frac{N h^d}{K(0)} \qquad \text{always}, \tag{60}$$

*and, under Assumption F.1,*

$$\mathbb{E}[\mathsf{R}_N(Y)] \leq \frac{1}{\kappa_K p_0}. \tag{61}$$

*Proof.* Fix $i$. Conditional on $Y_i = y$, define

$$Z_i := \sum_{j \neq i} \mathbf{1}\{\|Y_j - y\| \leq r_K h\}.$$

Then $Z_i$ is binomial with parameters $N - 1$ and

$$p(y) := \mu(B(y, r_K h)).$$

By Assumption F.3, $p(y) \geq p_0 h^d$ for $\mu$-a.e. $y$. The neighbor count including the particle itself is

$$M_i(Y; h) = 1 + Z_i.$$

Let $m_i := (N - 1)p(y)$. For $N \geq 2$,

$$m_i \geq (N - 1) p_0 h^d \geq \frac{1}{2} N p_0 h^d.$$

If

$$Z_i \geq \frac{1}{2} m_i,$$

then

$$M_i(Y;h) = 1 + Z_i \geq \frac{1}{2}m_i \geq \frac{1}{4}Np_0h^d.$$

Therefore

$$\mathbb{P}\left(M_i(Y;h) < \frac{1}{4}Np_0h^d \mid Y_i = y\right) \leq \mathbb{P}\left(Z_i < \frac{1}{2}m_i \mid Y_i = y\right).$$

For a binomial random variable, the multiplicative Chernoff bound gives

$$\mathbb{P}\left(Z_i < \frac{1}{2}m_i \mid Y_i = y\right) \leq \exp\left(-\frac{m_i}{8}\right).$$

Using $m_i \geq (N-1)p_0h^d \geq Np_0h^d/2$,

$$\exp\left(-\frac{m_i}{8}\right) \leq \exp\left(-\frac{p_0}{16}Nh^d\right).$$

This bound is uniform in $y$. Integrating over $Y_i$ and taking a union bound over $i = 1, \ldots, N$ proves (58).

On $\mathcal{E}_h$, Proposition F.2 applies with $\alpha = p_0/4$, giving (59). For the deterministic self-bound, observe that

$$q_Y(Y_i) = \frac{1}{N}\sum_{j=1}^{N} K_h(Y_i - Y_j) \geq \frac{1}{N}K_h(0) = \frac{K(0)}{Nh^d}.$$

Thus

$$\frac{1}{q_Y(Y_i)} \leq \frac{Nh^d}{K(0)} \qquad \text{for every } i,$$

and averaging over $i$ gives (60).

For the expectation, retain the actual neighbor count:

$$q_Y(Y_i) \geq \frac{\kappa_K h^{-d}}{N}(1 + Z_i).$$

For $Z \sim \mathrm{Bin}(N-1, p)$ and $p > 0$, the identity $(k+1)^{-1}\binom{N-1}{k} = N^{-1}\binom{N}{k+1}$ gives

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{1+Z}\right] &= \sum_{k=0}^{N-1} \frac{1}{k+1}\binom{N-1}{k}p^k(1-p)^{N-1-k} \\
&= \frac{1}{Np}\sum_{\ell=1}^{N}\binom{N}{\ell}p^\ell(1-p)^{N-\ell} \\
&= \frac{1-(1-p)^N}{Np}.
\end{aligned}$$

The last equality follows by subtracting the $\ell = 0$ term from the binomial expansion of $(p+1-p)^N$. Therefore

$$\mathbb{E}\left[\frac{1}{q_Y(Y_i)}\middle| Y_i = y\right] \leq \frac{h^d}{\kappa_K}\frac{1-(1-p(y))^N}{p(y)} \leq \frac{1}{\kappa_K p_0},$$

where the final inequality uses $p(y) \geq p_0h^d$ and $1-(1-p(y))^N \leq 1$. Integrating over $y$ and averaging over $i$ proves (61). □

**Remark F.6** (Expected control and tail mass)**.** *The same calculation without a uniform lower local mass gives*

$$\mathbb{E}[\mathsf{R}_N(Y)] \leq \frac{h^d}{\kappa_K}\int \frac{1-(1-p_h(y))^N}{p_h(y)}\,\mu(\mathrm{d}y), \qquad p_h(y) := \mu(B(y, r_Kh)). \tag{62}$$

*At $p_h(y) = 0$ the ratio is defined by its limit $N$. For $p > 0$, the finite geometric sum gives*

$$\frac{1-(1-p)^N}{p} = \sum_{k=0}^{N-1}(1-p)^k \leq \min\{N, p^{-1}\}.$$

*Thus (62) retains the contribution of regions with small local mass. Under uniform lower local mass, the expected bound (61) holds without a logarithmic sample-size condition. A condition of order $Nh^d \gtrsim \log N$ is instead used when (58) must control every particle simultaneously.*

### F.3 Propagation of Occupancy Along the Continuous-time Flow

The i.i.d. result controls the reciprocal KDE at a single time. To control it along the flow, one may combine initial occupancy at a slightly smaller scale with a Lipschitz-distortion bound for pairwise distances.

**Assumption F.7** (Lipschitz distortion of the drift)**.** *There exists a deterministic nonnegative integrable function $L_t$ such that, for almost every realized trajectory on $[0,T]$,*

$$\left\|b_{X(t)}(z) - b_{X(t)}(z')\right\| \le L_t \|z - z'\| \qquad \text{for all } z, z' \in \mathbb{R}^d, \tag{63}$$

*for all $0 \le t \le T$. Define*

$$\Gamma_T := \int_0^T L_t \, \mathrm{d}t. \tag{64}$$

**Lemma F.8** (Pairwise distance distortion)**.** *Under Assumption F.7, for every pair $i, j$ and every $0 \le t \le T$,*

$$\|X_i(t) - X_j(t)\| \le e^{\Gamma_T} \|X_i(0) - X_j(0)\|. \tag{65}$$

*Proof.* For $i, j$, set $D_{ij}(t) := X_i(t) - X_j(t)$. From (25),

$$\frac{\mathrm{d}}{\mathrm{d}t} D_{ij}(t) = b_{X(t)}(X_i(t)) - b_{X(t)}(X_j(t)).$$

Using Assumption F.7,

$$\frac{\mathrm{d}}{\mathrm{d}t} \|D_{ij}(t)\| \le L_t \|D_{ij}(t)\|$$

whenever $D_{ij}(t) \neq 0$. The same inequality holds in the upper Dini derivative sense when $D_{ij}(t) = 0$. Gronwall's inequality gives

$$\|D_{ij}(t)\| \le \exp\left(\int_0^t L_s \, \mathrm{d}s\right) \|D_{ij}(0)\| \le e^{\Gamma_T} \|D_{ij}(0)\|.$$

This proves the lemma. □

**Proposition F.9** (Trajectory-level reciprocal control from initial occupancy)**.** *Assume Assumptions F.1 and F.7. Suppose the initial configuration satisfies*

$$\#\{j : \|X_j(0) - X_i(0)\| \le r_K h e^{-\Gamma_T}\} \ge \alpha N h^d e^{-d\Gamma_T} \qquad \text{for every } i. \tag{66}$$

*Then, for every $0 \le t \le T$,*

$$\mathsf{R}_N(X(t)) \le \frac{e^{d\Gamma_T}}{\kappa_K \alpha}. \tag{67}$$

*Proof.* Fix $i$ and $t \in [0,T]$. If

$$\|X_j(0) - X_i(0)\| \le r_K h e^{-\Gamma_T},$$

then by Lemma F.8,

$$\|X_j(t) - X_i(t)\| \le e^{\Gamma_T} r_K h e^{-\Gamma_T} = r_K h.$$

Thus the number of particles within distance $r_K h$ of $X_i(t)$ is at least the number of particles within distance $r_K h e^{-\Gamma_T}$ of $X_i(0)$. By (66),

$$M_i(X(t); h) \ge \alpha N h^d e^{-d\Gamma_T}.$$

The proof of Proposition F.2 gives

$$q_{X(t)}(X_i(t)) \ge \frac{1}{N}\{\alpha N h^d e^{-d\Gamma_T}\}\kappa_K h^{-d} = \kappa_K \alpha e^{-d\Gamma_T}.$$

Taking reciprocals and averaging over $i$ proves (67). □

**Corollary F.10** (I.i.d. initial data plus controlled distortion)**.** *Let $X_1(0), \dots, X_N(0)$ be i.i.d. from a probability measure $\mu$. Suppose that, with*

$$\widetilde{h} := he^{-\Gamma_T},$$

*the lower local mass condition*

$$\mu(B(y, r_K\widetilde{h})) \geq p_0\widetilde{h}^d \qquad \text{for } \mu\text{-a.e. } y \tag{68}$$

*holds. Assume Assumptions F.1 and F.7. Then, with probability at least*

$$1 - N\exp\left(-\frac{p_0}{16}Nh^d e^{-d\Gamma_T}\right), \tag{69}$$

*one has*

$$\sup_{0\leq t\leq T} \mathsf{R}_N(X(t)) \leq \frac{4e^{d\Gamma_T}}{\kappa_K p_0}. \tag{70}$$

*Furthermore,*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} \mathsf{R}_N(X(t))\right] \leq \frac{e^{d\Gamma_T}}{\kappa_K p_0}. \tag{71}$$

*Proof.* Apply Proposition F.5 at bandwidth $\widetilde{h} = he^{-\Gamma_T}$. With probability at least (69), for every $i$,

$$\#\{j : \|X_j(0) - X_i(0)\| \leq r_K he^{-\Gamma_T}\} \geq \frac{p_0}{4}N(he^{-\Gamma_T})^d.$$

This is (66) with $\alpha = p_0/4$. Therefore Proposition F.9 gives (70).

For the expectation, let $M_i^0$ count the initial particles within $r_K\widetilde{h}$ of $X_i(0)$, including particle $i$. The distortion bound puts all these particles within $r_K h$ of $X_i(t)$ for every $t \leq T$. Hence

$$q_{X(t)}(X_i(t)) \geq \frac{\kappa_K h^{-d}}{N}M_i^0, \qquad \sup_{0\leq t\leq T} \mathsf{R}_N(X(t)) \leq \frac{Nh^d}{\kappa_K}\frac{1}{N}\sum_i \frac{1}{M_i^0}.$$

Conditionally on $X_i(0) = y$, the random variable $M_i^0 - 1$ is binomial with parameters $N-1$ and $p_{\widetilde{h}}(y) = \mu(B(y, r_K\widetilde{h}))$. The binomial identity from Proposition F.5 gives

$$\mathbb{E}\left[\frac{1}{M_i^0}\middle| X_i(0) = y\right] = \frac{1 - (1 - p_{\widetilde{h}}(y))^N}{Np_{\widetilde{h}}(y)} \leq \frac{1}{Np_0\widetilde{h}^d}.$$

Taking expectations in the preceding bound for the supremum and using $h^d/\widetilde{h}^d = e^{d\Gamma_T}$ proves (71). □

**Remark F.11** (Use in Theorem E.5)**.** *Under Corollary F.10, Assumption C.3 holds with*

$$\Lambda_T = \frac{e^{d\Gamma_T}}{\kappa_K p_0}.$$

*If $\Gamma_T$ and $p_0^{-1}$ are bounded independently of $N, h, T$, then this expected reciprocal bound is uniform. Establishing such distortion control on $T = N$ is a separate stability problem. For example, a nonnegative Lipschitz bound of order $h^{-2}$ integrated over $[0, N]$ gives a bound of order $Nh^{-2}$ for $\Gamma_N$, which does not give a uniform value of $e^{d\Gamma_N}$.*

### F.4 Reciprocal KDEs for Full-support Initial Densities

The expectation bounds also clarify the role of the spatial tails.

**Proposition F.12** (Full-support reciprocal growth)**.** *Let $Y_1, \dots, Y_N$ be i.i.d. from a bounded continuous density $f$ that is positive on $\mathbb{R}^d$. Suppose $K$ is a nonnegative normalized kernel with $0 < K(0) < \infty$, $h_N \to 0$, and $Nh_N^d \to \infty$. Then*

$$\mathbb{E}[\mathsf{R}_N(Y)] \longrightarrow \infty.$$

*Proof.* For each fixed $y$, conditioning on $Y_1 = y$ gives

$$\mathbb{E}[q_Y(Y_1) \mid Y_1 = y] = \frac{K(0)}{Nh_N^d} + \left(1 - \frac{1}{N}\right)(K_{h_N} * f)(y).$$

The function $v \mapsto 1/v$ is convex on $(0, \infty)$, since its second derivative is $2/v^3 > 0$. Conditional Jensen's inequality and exchangeability therefore imply

$$\mathbb{E}[\mathsf{R}_N(Y)] \geq \int_{\mathbb{R}^d} \frac{f(y)}{K(0)/(Nh_N^d) + (1 - 1/N)(K_{h_N} * f)(y)} \, \mathrm{d}y.$$

After the change of variables in the convolution,

$$(K_{h_N} * f)(y) = \int K(u) f(y - h_N u) \, \mathrm{d}u.$$

Continuity gives $f(y - h_N u) \to f(y)$ for every fixed $y, u$, and the integrand is dominated by $\|f\|_\infty K(u)$. Dominated convergence yields $(K_{h_N} * f)(y) \to f(y)$. The denominator in the displayed lower bound thus tends to $f(y) > 0$, so the integrand tends to one. Fatou's lemma gives

$$\liminf_{N \to \infty} \mathbb{E}[\mathsf{R}_N(Y)] \geq \int_{\mathbb{R}^d} 1 \, \mathrm{d}y = \infty.$$

This proves the proposition. □

**Remark F.13** (Bandwidth optimization and initialization tails)**.** *Proposition F.12 applies, in particular, to Gaussian initialization. For kernels with $a_1 > 0$, the coefficient $a_1 \Lambda_N(h_N)$ in the conservative bound cannot then be uniformly bounded in the regime of the proposition, since the time supremum includes time zero. The non-asymptotic bound continues to hold with this coefficient displayed. Its quantitative growth must be included in any bandwidth optimization. Formula (62) provides an upper bound that retains the local mass in the tails.*

## G Finite-particle Rates for Non-conservative Drifting with Laplace Kernel

We now prove a continuous-time finite-particle rate for the non-conservative displacement-based drifting field with Laplace kernel. The main point is that, unlike the Gaussian kernel, the Laplace mean-shift field is not exactly a score-difference field. Nevertheless, it admits a useful "sharp-score" representation, which yields a finite-particle convergence theorem with an additional scale-mismatch residual. The sharp companion kernel construction used in this section follows the sharp-normalization framework of Franz et al. (2026), specialized to the Laplace kernel. Our use of it is different: rather than replacing the original Laplace normalization by the sharp normalization, we use the sharp kernel to decompose the original non-conservative Laplace field into a preconditioned sharp-score mismatch plus a scale-mismatch residual.

Throughout this section, let

$$K_h(u) := c_d h^{-d} \exp\left(-\frac{\|u\|}{h}\right), \qquad u \in \mathbb{R}^d,$$

where $c_d > 0$ is chosen so that $\int_{\mathbb{R}^d} K_h(u) \, \mathrm{d}u = 1$. For a probability measure $\alpha$ on $\mathbb{R}^d$, define its Laplace KDE

$$Q_{\alpha,h}(z) := \int_{\mathbb{R}^d} K_h(z - y) \, \alpha(\mathrm{d}y),$$

and the non-conservative displacement mean-shift vector

$$M_{\alpha,h}(z) := \frac{\int_{\mathbb{R}^d} (y - z) K_h(z - y) \, \alpha(\mathrm{d}y)}{Q_{\alpha,h}(z)}.$$

The non-conservative Laplace drifting field from a model distribution $\mu$ toward the data distribution $\nu$ is

$$u^{\mathrm{Lap}}_{\nu,\mu,h}(z) := M_{\nu,h}(z) - M_{\mu,h}(z).$$

For a particle configuration $x = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ with $N \geq 2$, let

$$\mu_x^N := \frac{1}{N} \sum_{j=1}^{N} \delta_{x_j}.$$

We analyze the self-masked, or leave-one-out, finite-particle version of the non-conservative Laplace drifting field. For each $i$, define

$$\mu_{x,-i}^N := \frac{1}{N-1} \sum_{j \neq i} \delta_{x_j},$$

and

$$u_{x,-i}(z) := M_{\nu,h}(z) - M_{\mu_{x,-i}^N,h}(z).$$

The leave-one-out form is both analytically convenient and faithful to the self-masking used in the original drifting implementation. Analytically, it ensures that $u_{x,-i}$ depends on $x_i$ only through the evaluation point $z = x_i$, so that

$$\nabla_{x_i} \cdot u_{x,-i}(x_i) = \nabla_z \cdot u_{x,-i}(z)\big|_{z=x_i},$$

and the joint-entropy identity has no reciprocal self-interaction correction. This matches the implementation of Deng et al. (2026, Algorithm 2), where generated samples are reused as negative samples but the diagonal self-interaction is masked.

A full-KDE version without leave-one-out could also be studied, but it would produce an additional self-interaction correction. For the exact Laplace kernel this correction is less transparent because the field is non-conservative and the kernel is not differentiable at the origin. We therefore analyze the self-masked version, which is the version closest to the practical algorithm.

The continuous-time dynamics are

$$\dot{X}_i(t) = u_{X(t),-i}(X_i(t)), \qquad i = 1, \dots, N. \tag{72}$$

The residual drift quantity we want to control is

$$\mathsf{V}_N^{\mathrm{Lap}}(x) := \frac{1}{N} \sum_{i=1}^{N} \|u_{x,-i}(x_i)\|^2. \tag{73}$$

This is the mean squared norm of the non-conservative Laplace drift applied to the generated particles. Hence if one more explicit drift step is applied,

$$\widetilde{x}_i = x_i + \eta u_{x,-i}(x_i),$$

then the identity coupling gives

$$W_2^2 \left( \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}, \frac{1}{N} \sum_{i=1}^{N} \delta_{\widetilde{x}_i} \right) \leq \eta^2 \mathsf{V}_N^{\mathrm{Lap}}(x).$$

Thus a bound on $\mathsf{V}_N^{\mathrm{Lap}}$ controls how much an additional non-conservative Laplace drifting correction changes the generated distribution.

### G.1 Sharp-score Representation of the Laplace Mean-shift Field

The Laplace mean-shift field is not itself a KDE score, but it is a scaled score of a companion kernel. Define the sharp companion kernel

$$L_h(u) := h(\|u\| + h) K_h(u). \tag{74}$$

Let

$$R_{\alpha,h}(z) := \int_{\mathbb{R}^d} L_h(z - y)\, \alpha(\mathrm{d}y),$$

and define the sharp-smoothed score

$$\sigma_{\alpha,h}(z) := \nabla \log R_{\alpha,h}(z).$$

Also define the scalar scale factor

$$a_{\alpha,h}(z) := \frac{R_{\alpha,h}(z)}{Q_{\alpha,h}(z)}.$$

**Lemma G.1** (Sharp-score representation)**.** *For every probability measure $\alpha$ for which the following quantities are well-defined,*

$$M_{\alpha,h}(z) = a_{\alpha,h}(z)\sigma_{\alpha,h}(z). \tag{75}$$

*Moreover,*

$$a_{\alpha,h}(z) = h\{\bar{r}_{\alpha,h}(z) + h\}, \tag{76}$$

*where*

$$\bar{r}_{\alpha,h}(z) := \frac{\int_{\mathbb{R}^d} \|y - z\|\, K_h(z-y)\, \alpha(\mathrm{d}y)}{\int_{\mathbb{R}^d} K_h(z-y)\, \alpha(\mathrm{d}y)}$$

*is the Laplace-weighted local mean radius.*

*Proof.* Let $r = \|u\|$. Since

$$L_h(u) = h(r+h)c_d h^{-d} e^{-r/h},$$

the radial derivative is

$$\frac{\mathrm{d}}{\mathrm{d}r}\{h(r+h)c_d h^{-d} e^{-r/h}\} = hc_d h^{-d} e^{-r/h} - h(r+h)c_d h^{-d}\frac{1}{h} e^{-r/h}.$$

Combining the two terms,

$$\frac{\mathrm{d}}{\mathrm{d}r}\{h(r+h)c_d h^{-d} e^{-r/h}\} = c_d h^{1-d} e^{-r/h} - (r+h)c_d h^{-d} e^{-r/h}.$$

Since

$$c_d h^{1-d} e^{-r/h} = hc_d h^{-d} e^{-r/h},$$

we get

$$\frac{\mathrm{d}}{\mathrm{d}r}\{h(r+h)c_d h^{-d} e^{-r/h}\} = \{h - (r+h)\}c_d h^{-d} e^{-r/h} = -rK_h(u).$$

Therefore, for $u \neq 0$,

$$\nabla_u L_h(u) = -rK_h(u)\frac{u}{r} = -uK_h(u).$$

By continuity, the same identity holds at $u = 0$, because both sides vanish there. Taking $u = z - y$, we obtain

$$\nabla_z L_h(z-y) = -(z-y)K_h(z-y) = (y-z)K_h(z-y).$$

Hence

$$\nabla R_{\alpha,h}(z) = \int_{\mathbb{R}^d} \nabla_z L_h(z-y)\, \alpha(\mathrm{d}y) = \int_{\mathbb{R}^d} (y-z)K_h(z-y)\, \alpha(\mathrm{d}y).$$

Dividing by $Q_{\alpha,h}(z)$ gives

$$M_{\alpha,h}(z) = \frac{\nabla R_{\alpha,h}(z)}{Q_{\alpha,h}(z)} = \frac{R_{\alpha,h}(z)}{Q_{\alpha,h}(z)}\frac{\nabla R_{\alpha,h}(z)}{R_{\alpha,h}(z)} = a_{\alpha,h}(z)\sigma_{\alpha,h}(z).$$

This proves (75).

It remains to prove (76). By definition of $L_h$,

$$R_{\alpha,h}(z) = \int_{\mathbb{R}^d} h(\|z-y\| + h)K_h(z-y)\, \alpha(\mathrm{d}y).$$

Thus

$$R_{\alpha,h}(z) = h\int_{\mathbb{R}^d} \|z-y\|\, K_h(z-y)\, \alpha(\mathrm{d}y) + h^2 \int_{\mathbb{R}^d} K_h(z-y)\, \alpha(\mathrm{d}y).$$

The second integral is $Q_{\alpha,h}(z)$. Therefore

$$\frac{R_{\alpha,h}(z)}{Q_{\alpha,h}(z)} = h\frac{\int \|z-y\|\, K_h(z-y)\, \alpha(\mathrm{d}y)}{Q_{\alpha,h}(z)} + h^2.$$

Since $\|z-y\| = \|y-z\|$, this is

$$a_{\alpha,h}(z) = h\bar{r}_{\alpha,h}(z) + h^2 = h\{\bar{r}_{\alpha,h}(z) + h\}.$$

The lemma follows. □

Let

$$Z_{\#,h} := \int_{\mathbb{R}^d} L_h(u)\,\mathrm{d}u.$$

The normalized sharp-smoothed data density is

$$\rho^{\#}_{\nu,h}(z) := \frac{R_{\nu,h}(z)}{Z_{\#,h}}.$$

Since the normalizing constant does not depend on $z$,

$$\nabla \log \rho^{\#}_{\nu,h}(z) = \nabla \log R_{\nu,h}(z) = \sigma_{\nu,h}(z).$$

For the full empirical model, define

$$R_{x,h} := R_{\mu_x^N,h}, \qquad Q_{x,h} := Q_{\mu_x^N,h}, \qquad \sigma_{x,h} := \nabla \log R_{x,h}, \qquad a_{x,h} := \frac{R_{x,h}}{Q_{x,h}}.$$

Also define the full non-conservative Laplace field

$$u_x(z) := M_{\nu,h}(z) - M_{\mu_x^N,h}(z).$$

By Lemma G.1,

$$u_x(z) = a_{\nu,h}(z)\sigma_{\nu,h}(z) - a_{x,h}(z)\sigma_{x,h}(z).$$

Define the sharp score mismatch

$$b^{\#}_x(z) := \sigma_{\nu,h}(z) - \sigma_{x,h}(z),$$

and the Laplace scale-mismatch residual

$$e_x(z) := \{a_{\nu,h}(z) - a_{x,h}(z)\}\sigma_{\nu,h}(z). \tag{77}$$

Then

$$u_x(z) = a_{x,h}(z)b^{\#}_x(z) + e_x(z). \tag{78}$$

This decomposition separates the local model scale from its difference with the data scale. If the scales match, then $e_x = 0$ and the drift is a positive scalar preconditioning of the sharp-score mismatch. More generally, absolute or relative bounds on $e_x$ give the coercivity estimates below.

### G.2 Entropy Identity for the Leave-one-out Non-conservative Field

Let $p_t^N$ denote the joint density of $X(t)$. We measure it against the sharp-smoothed product target

$$(\rho^{\#}_{\nu,h})^{\otimes N}(x) = \prod_{i=1}^{N} \rho^{\#}_{\nu,h}(x_i),$$

and define

$$H^{\#}_N(t) := \mathrm{KL}\left(p_t^N \,\middle\|\, (\rho^{\#}_{\nu,h})^{\otimes N}\right). \tag{79}$$

Let

$$\mathcal{A}_{\#} f(z) := \nabla\cdot f(z) + \sigma_{\nu,h}(z)\cdot f(z)$$

be the Stein divergence associated with $\rho^{\#}_{\nu,h}$. Define

$$\mathsf{S}^{\mathrm{Lap}}_N(x) := \frac{1}{N}\sum_{i=1}^{N} \mathcal{A}_{\#} u_{x,-i}(x_i). \tag{80}$$

**Assumption G.2** (Regularity for the leave-one-out Laplace flow)**.** *The ODE (72) has a unique solution on $[0,T]$, the joint law admits a differentiable density $p_t^N$, and this density solves the Liouville equation associated with $x \mapsto (u_{x,-1}(x_1),\dots,u_{x,-N}(x_N))$. The leave-one-out fields are classically differentiable at their evaluation centers along the trajectory. In particular, the trajectory avoids particle collisions and, for discrete data, data atoms at which the field is not differentiable. The joint entropy and the integrals used below are finite. Differentiation of the entropy under the integral is justified, the entropy is absolutely continuous in time, and its derivative is integrable on $[0,T]$. The joint-space integration by parts has vanishing boundary terms. For spatial integrals of full fields, divergences are understood as weak derivatives; all stated spatial integrations by parts are assumed justified. Local higher differentiability is imposed only where a quadrature lemma uses it.*

**Theorem G.3** (Joint entropy for leave-one-out Laplace drifting)**.** *Under Assumption G.2,*

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N^{\#}(t) = -N\mathbb{E}_t[\mathsf{S}_N^{\mathrm{Lap}}(X(t))]. \tag{81}$$

*Proof.* Let

$$r_N^{\#}(x) := (\rho_{\nu,h}^{\#})^{\otimes N}(x).$$

The joint density $p_t^N$ satisfies

$$\partial_t p_t^N(x) + \sum_{i=1}^N \nabla_{x_i} \cdot \left\{p_t^N(x) u_{x,-i}(x_i)\right\} = 0.$$

By definition,

$$H_N^{\#}(t) = \int p_t^N(x) \log \frac{p_t^N(x)}{r_N^{\#}(x)}\,\mathrm{d}x.$$

Differentiating under the integral gives

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N^{\#}(t) = \int \partial_t p_t^N(x) \left\{\log \frac{p_t^N(x)}{r_N^{\#}(x)} + 1\right\}\,\mathrm{d}x.$$

Since $p_t^N$ is a probability density,

$$\int \partial_t p_t^N(x)\,\mathrm{d}x = 0.$$

Therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N^{\#}(t) = \int \partial_t p_t^N(x) \log \frac{p_t^N(x)}{r_N^{\#}(x)}\,\mathrm{d}x.$$

Using the Liouville equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N^{\#}(t) = -\sum_{i=1}^N \int \nabla_{x_i} \cdot \{p_t^N(x) u_{x,-i}(x_i)\} \log \frac{p_t^N(x)}{r_N^{\#}(x)}\,\mathrm{d}x.$$

Integrating by parts in $x_i$,

$$\frac{\mathrm{d}}{\mathrm{d}t}H_N^{\#}(t) = \sum_{i=1}^N \int p_t^N(x) u_{x,-i}(x_i) \cdot \nabla_{x_i} \log \frac{p_t^N(x)}{r_N^{\#}(x)}\,\mathrm{d}x.$$

Now

$$\nabla_{x_i} \log r_N^{\#}(x) = \nabla \log \rho_{\nu,h}^{\#}(x_i) = \sigma_{\nu,h}(x_i).$$

Hence

$$\nabla_{x_i} \log \frac{p_t^N(x)}{r_N^{\#}(x)} = \nabla_{x_i} \log p_t^N(x) - \sigma_{\nu,h}(x_i).$$

Thus

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}H_N^{\#}(t) = &\sum_{i=1}^N \int p_t^N u_{x,-i}(x_i) \cdot \nabla_{x_i} \log p_t^N\,\mathrm{d}x \\ &- \sum_{i=1}^N \int p_t^N u_{x,-i}(x_i) \cdot \sigma_{\nu,h}(x_i)\,\mathrm{d}x.\end{aligned}$$

Since

$$p_t^N \nabla_{x_i} \log p_t^N = \nabla_{x_i} p_t^N,$$

the first integral in the last display satisfies

$$\int p_t^N u_{x,-i}(x_i) \cdot \nabla_{x_i} \log p_t^N\,\mathrm{d}x = \int u_{x,-i}(x_i) \cdot \nabla_{x_i} p_t^N\,\mathrm{d}x.$$

Integrating by parts again,

$$\int u_{x,-i}(x_i) \cdot \nabla_{x_i} p_t^N \, \mathrm{d}x = -\int p_t^N \nabla_{x_i} \cdot u_{x,-i}(x_i) \, \mathrm{d}x.$$

Because $u_{x,-i}$ depends on $x_i$ only through the evaluation point $z = x_i$, not through the leave-one-out model measure $\mu^N_{x,-i}$, we have

$$\nabla_{x_i} \cdot u_{x,-i}(x_i) = \nabla_z \cdot u_{x,-i}(z)\big|_{z=x_i}.$$

Therefore

$$\frac{\mathrm{d}}{\mathrm{d}t} H_N^{\#}(t) = -\mathbb{E}_t \left[ \sum_{i=1}^N \left\{ \nabla \cdot u_{X(t),-i}(X_i(t)) + \sigma_{\nu,h}(X_i(t)) \cdot u_{X(t),-i}(X_i(t)) \right\} \right].$$

The expression inside braces is exactly

$$\mathcal{A}_{\#} u_{X(t),-i}(X_i(t)).$$

Hence

$$\frac{\mathrm{d}}{\mathrm{d}t} H_N^{\#}(t) = -N \mathbb{E}_t[\mathsf{S}_N^{\mathrm{Lap}}(X(t))],$$

which proves the identity. □

### G.3 Laplace Coercivity with A Scale-mismatch Residual

Let

$$\rho^{\#}_{x,h}(z) := \frac{R_{x,h}(z)}{Z_{\#,h}}$$

be the full sharp-smoothed empirical density. Define the sharp-smoothed drift energy

$$\mathcal{V}_h^{\mathrm{Lap}}(x) := \int_{\mathbb{R}^d} \|u_x(z)\|^2 \, \rho^{\#}_{x,h}(z) \, \mathrm{d}z, \tag{82}$$

and the sharp-smoothed Stein drift

$$\mathcal{J}_h^{\mathrm{Lap}}(x) := \int_{\mathbb{R}^d} \mathcal{A}_{\#} u_x(z) \rho^{\#}_{x,h}(z) \, \mathrm{d}z. \tag{83}$$

**Lemma G.4** (Sharp-smoothed Stein identity)**.** *For every fixed configuration $x$ for which the integrations by parts are justified,*

$$\mathcal{J}_h^{\mathrm{Lap}}(x) = \int_{\mathbb{R}^d} b_x^{\#}(z) \cdot u_x(z) \rho^{\#}_{x,h}(z) \, \mathrm{d}z. \tag{84}$$

*Proof.* By definition,

$$\mathcal{A}_{\#} u_x = \nabla \cdot u_x + \sigma_{\nu,h} \cdot u_x.$$

Therefore

$$\mathcal{J}_h^{\mathrm{Lap}}(x) = \int \rho^{\#}_{x,h} \nabla \cdot u_x + \int \rho^{\#}_{x,h} \sigma_{\nu,h} \cdot u_x.$$

For the first term, integration by parts gives

$$\int \rho^{\#}_{x,h} \nabla \cdot u_x = -\int \nabla \rho^{\#}_{x,h} \cdot u_x.$$

Since

$$\nabla \rho^{\#}_{x,h} = \rho^{\#}_{x,h} \nabla \log \rho^{\#}_{x,h} = \rho^{\#}_{x,h} \sigma_{x,h},$$

we obtain

$$\int \rho^{\#}_{x,h} \nabla \cdot u_x = -\int \rho^{\#}_{x,h} \sigma_{x,h} \cdot u_x.$$

Thus

$$\mathcal{J}_h^{\mathrm{Lap}}(x) = \int \rho^{\#}_{x,h} (\sigma_{\nu,h} - \sigma_{x,h}) \cdot u_x.$$

By definition, $b_x^{\#} = \sigma_{\nu,h} - \sigma_{x,h}$. Hence

$$\mathcal{J}_h^{\mathrm{Lap}}(x) = \int b_x^{\#} \cdot u_x \, \rho^{\#}_{x,h},$$

which proves the claim. □

**Remark G.5** (Spatial regularity of the Laplace field)**.** *The Laplace mean-shift fields are locally Lipschitz, and their spatial divergences in integrals are understood as weak derivatives, represented almost everywhere. At a particle center, the leave-one-out field $u_{x,-i}$ has no self-kernel cusp. Under Assumption G.2, its divergence at $x_i$ is a classical derivative.*

*The full field $u_x$ is used through integrals. Indeed, write its model mean shift near $x_i$ as $P(z)/Q(z)$, omitting the common factor $1/N$. The self-contribution to $Q$ is*

$$K_h(z-x_i) = K_h(0)\left\{1 - \frac{\|z-x_i\|}{h} + O(\|z-x_i\|^2)\right\},$$

*whereas the self-contribution to $P$ is $(x_i - z)K_h(z-x_i)$, which is differentiable at $x_i$. When the remaining contributions are smooth and $P(x_i) \neq 0$, the quotient contains a nonzero multiple of $\|z-x_i\|$ and need not be differentiable there. The sharp-smoothed Stein identity only uses the almost-everywhere divergence and its integration by parts.*

**Remark G.6** (Growth of the empirical scale)**.** *The scale factor has the universal lower bound $a_{x,h} \geq h^2$. Its upper bound depends on the evaluation region. To see this, let $R_x := \max_j \|x_j\|$ and define*

$$w_j(z) := \frac{K_h(z-x_j)}{\sum_k K_h(z-x_k)}.$$

*Since the weights are positive and sum to one, the triangle inequality gives*

$$\|z\| - R_x \leq \sum_j w_j(z)\,\|z-x_j\| = \bar{r}_{x,h}(z) \leq \|z\| + R_x.$$

*Consequently,*

$$h(\|z\| - R_x + h) \leq a_{x,h}(z) \leq h(\|z\| + R_x + h).$$

*In particular, $a_{x,h}$ grows without bound as $\|z\| \to \infty$. The weighted energy below retains the local scale inside the integral.*

Define the weighted drift energy and weighted scale mismatch by

$$\mathcal{W}_h^{\mathrm{Lap}}(x) := \int_{\mathbb{R}^d} \frac{\|u_x(z)\|^2}{a_{x,h}(z)} \rho_{x,h}^{\#}(z)\,\mathrm{d}z, \tag{85}$$

$$\mathcal{D}_h^{\mathrm{Lap}}(x) := \int_{\mathbb{R}^d} \frac{\|e_x(z)\|^2}{a_{x,h}(z)} \rho_{x,h}^{\#}(z)\,\mathrm{d}z. \tag{86}$$

**Assumption G.7** (Weighted Laplace scale alignment)**.** *For all configurations visited by the flow on $[0,T]$, the quantities $\mathcal{V}_h^{\mathrm{Lap}}$, $\mathcal{J}_h^{\mathrm{Lap}}$, $\mathcal{W}_h^{\mathrm{Lap}}$, and $\mathcal{D}_h^{\mathrm{Lap}}$ are finite. There is a constant $\Delta_h \geq 0$ such that*

$$\mathcal{D}_h^{\mathrm{Lap}}(x) \leq \Delta_h^2. \tag{87}$$

*The constants may depend on $N, h, T$ and the initial law; any uniformity used in a rate statement is specified separately.*

**Lemma G.8** (Weighted Laplace coercivity)**.** *Under Assumption G.7,*

$$\mathcal{J}_h^{\mathrm{Lap}}(x) \geq \frac{1}{2}\mathcal{W}_h^{\mathrm{Lap}}(x) - \frac{1}{2}\mathcal{D}_h^{\mathrm{Lap}}(x) \geq \frac{1}{2}\mathcal{W}_h^{\mathrm{Lap}}(x) - \frac{1}{2}\Delta_h^2. \tag{88}$$

*If, for some $0 \leq \theta_h < 1$,*

$$\mathcal{D}_h^{\mathrm{Lap}}(x) \leq \theta_h \mathcal{W}_h^{\mathrm{Lap}}(x), \tag{89}$$

*then*

$$\mathcal{J}_h^{\mathrm{Lap}}(x) \geq (1 - \sqrt{\theta_h})\mathcal{W}_h^{\mathrm{Lap}}(x). \tag{90}$$

*Proof.* Since $a_{x,h} > 0$, the decomposition $u_x = a_{x,h} b_x^{\#} + e_x$ gives

$$b_x^{\#} \cdot u_x = \frac{\|u_x\|^2}{a_{x,h}} - \frac{e_x \cdot u_x}{a_{x,h}}.$$

For any vectors $v, w$, the inequality $0 \le \|v-w\|^2 = \|v\|^2 + \|w\|^2 - 2v\cdot w$ implies $v\cdot w \le (\|v\|^2 + \|w\|^2)/2$. Applying it to $e_x, u_x$, dividing by $a_{x,h}$, and integrating gives

$$\int \frac{e_x \cdot u_x}{a_{x,h}} \rho^{\#}_{x,h} \le \frac{1}{2}\mathcal{D}^{\mathrm{Lap}}_h(x) + \frac{1}{2}\mathcal{W}^{\mathrm{Lap}}_h(x).$$

Lemma G.4 now yields the first inequality in (88). The bound $\mathcal{D}^{\mathrm{Lap}}_h \le \Delta^2_h$ gives the second.

For the relative estimate, Cauchy–Schwarz with respect to the measure $a^{-1}_{x,h}\rho^{\#}_{x,h}\,\mathrm{d}z$ gives

$$\left|\int \frac{e_x \cdot u_x}{a_{x,h}} \rho^{\#}_{x,h}\,\mathrm{d}z\right| \le \sqrt{\mathcal{D}^{\mathrm{Lap}}_h(x)\mathcal{W}^{\mathrm{Lap}}_h(x)}.$$

Using (89), the right side is at most $\sqrt{\theta_h}\mathcal{W}^{\mathrm{Lap}}_h(x)$. Substituting this bound in the identity for $\mathcal{J}^{\mathrm{Lap}}_h$ proves (90). □

**Assumption G.9** (Local scale and tail energy)**.** *For every configuration visited on $[0,T]$, there is a measurable set $\Omega_x \subset \mathbb{R}^d$ such that*

$$a_{x,h}(z) \le L_h \quad (z \in \Omega_x), \qquad \int_{\Omega_x^c} \|u_x(z)\|^2 \rho^{\#}_{x,h}(z)\,\mathrm{d}z \le \tau_{h,N}, \tag{91}$$

*where $0 < L_h < \infty$ and $\tau_{h,N} \ge 0$ are independent of the configuration and time on the stated horizon.*

**Corollary G.10** (Localized unweighted coercivity)**.** *Under Assumptions G.7 and G.9,*

$$\mathcal{V}^{\mathrm{Lap}}_h(x) \le 2L_h\mathcal{J}^{\mathrm{Lap}}_h(x) + L_h\mathcal{D}^{\mathrm{Lap}}_h(x) + \tau_{h,N}. \tag{92}$$

*More generally, if $a_{x,h} \ge \lambda_h > 0$ everywhere and $\int \|e_x\|^2 \rho^{\#}_{x,h} \le \delta^2_h$, then*

$$\mathcal{J}^{\mathrm{Lap}}_h(x) \ge \frac{\mathcal{V}^{\mathrm{Lap}}_h(x) - \tau_{h,N}}{2L_h} - \frac{\delta^2_h}{2\lambda_h}.$$

*For the Laplace kernel one may always take $\lambda_h = h^2$.*

*Proof.* Positivity of the weighted integrand and the local upper bound give

$$\mathcal{W}^{\mathrm{Lap}}_h(x) \ge \int_{\Omega_x} \frac{\|u_x\|^2}{a_{x,h}} \rho^{\#}_{x,h} \ge \frac{1}{L_h}\int_{\Omega_x} \|u_x\|^2 \rho^{\#}_{x,h} \ge \frac{\mathcal{V}^{\mathrm{Lap}}_h(x) - \tau_{h,N}}{L_h}.$$

Substitute this inequality in the first bound of (88) and multiply by $2L_h$ to obtain (92). Under the additional lower bound,

$$\mathcal{D}^{\mathrm{Lap}}_h(x) = \int \frac{\|e_x\|^2}{a_{x,h}} \rho^{\#}_{x,h} \le \frac{1}{\lambda_h}\int \|e_x\|^2 \rho^{\#}_{x,h} \le \frac{\delta^2_h}{\lambda_h}.$$

Using this in the same inequality proves the final statement. □

### G.4 From the Sharp-smoothed Coercivity to Particle Rates

The entropy identity controls the leave-one-out, center-evaluated Stein quantity. The weighted coercivity estimate controls a full-field integral. The following assumptions record their comparison and the comparison of the two unweighted energies.

**Assumption G.11** (Quadrature and leave-one-out approximation)**.** *For all configurations visited on $[0,T]$, there are constants $\varepsilon_{S,h,N}, \varepsilon_{V,h,N} \ge 0$ such that*

$$\left|\mathsf{S}^{\mathrm{Lap}}_N(x) - \mathcal{J}^{\mathrm{Lap}}_h(x)\right| \le \varepsilon_{S,h,N}, \tag{93}$$

$$\left|\mathsf{V}^{\mathrm{Lap}}_N(x) - \mathcal{V}^{\mathrm{Lap}}_h(x)\right| \le \varepsilon_{V,h,N}. \tag{94}$$

**Remark G.12.** *The point evaluations in Assumption G.11 involve only the leave-one-out fields. The full field appears in integrals, where its weak divergence is sufficient. Lemma G.21 bounds the approximation errors using local smoothness of each leave-one-out field near its evaluation center, explicit quadrature tails, and integrated differences of field values. These quantities may depend on $N, h, T$. The rate statements retain their dependence; denominator control at the centers alone does not give a uniform approximation rate.*

**Theorem G.13** (Continuous-time finite-particle rates for Laplace drifting)**.** *Let Assumptions G.2, G.7, and G.11 hold on $[0, T]$. Then*

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathcal{W}_h^{\mathrm{Lap}}(X(t))]\,\mathrm{d}t \le \frac{2H_N^{\#}(0)}{NT} + \Delta_h^2 + 2\varepsilon_{S,h,N}. \tag{95}$$

*If Assumption G.9 also holds, then*

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{V}_N^{\mathrm{Lap}}(X(t))]\,\mathrm{d}t \le \frac{2L_h H_N^{\#}(0)}{NT} + L_h\Delta_h^2 + 2L_h\varepsilon_{S,h,N} + \tau_{h,N} + \varepsilon_{V,h,N}. \tag{96}$$

*In particular, if $H_N^{\#}(0) \le \kappa_0 N$ and $T = N$, then*

$$\frac{1}{N}\int_0^N \mathbb{E}_t[\mathsf{V}_N^{\mathrm{Lap}}(X(t))]\,\mathrm{d}t \le \frac{2\kappa_0 L_h}{N} + L_h\Delta_h^2 + 2L_h\varepsilon_{S,h,N} + \tau_{h,N} + \varepsilon_{V,h,N}. \tag{97}$$

*Proof.* Integrating Theorem G.3 gives

$$H_N^{\#}(T) - H_N^{\#}(0) = -N\int_0^T \mathbb{E}_t[\mathsf{S}_N^{\mathrm{Lap}}(X(t))]\,\mathrm{d}t.$$

Since $H_N^{\#}(T) \ge 0$, rearrangement and division by $NT$ yield

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{S}_N^{\mathrm{Lap}}(X(t))]\,\mathrm{d}t \le \frac{H_N^{\#}(0)}{NT}. \tag{98}$$

By Lemma G.8 and Assumption G.11,

$$\mathcal{W}_h^{\mathrm{Lap}}(x) \le 2\mathcal{J}_h^{\mathrm{Lap}}(x) + \mathcal{D}_h^{\mathrm{Lap}}(x) \le 2\mathsf{S}_N^{\mathrm{Lap}}(x) + 2\varepsilon_{S,h,N} + \Delta_h^2.$$

Take expectations, integrate over $[0, T]$, divide by $T$, and apply (98). This proves (95).

For the center-velocity estimate, Corollary G.10 and both comparisons in Assumption G.11 give

$$\begin{aligned}\mathsf{V}_N^{\mathrm{Lap}}(x) &\le \mathcal{V}_h^{\mathrm{Lap}}(x) + \varepsilon_{V,h,N}\\ &\le 2L_h\mathcal{J}_h^{\mathrm{Lap}}(x) + L_h\mathcal{D}_h^{\mathrm{Lap}}(x) + \tau_{h,N} + \varepsilon_{V,h,N}\\ &\le 2L_h\mathsf{S}_N^{\mathrm{Lap}}(x) + 2L_h\varepsilon_{S,h,N} + L_h\Delta_h^2 + \tau_{h,N} + \varepsilon_{V,h,N}.\end{aligned}$$

After expectation and time averaging, (98) proves (96). Finally, $H_N^{\#}(0)/(NT) \le \kappa_0/N$ when $T = N$, which gives (97). □

**Corollary G.14** (Rates under relative scale alignment)**.** *Suppose the regularity and approximation assumptions of Theorem G.13 hold, and replace the absolute mismatch bound by (89), uniformly along the trajectory, with $0 \le \theta_h < 1$. Then*

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathcal{W}_h^{\mathrm{Lap}}(X(t))]\,\mathrm{d}t \le \frac{1}{1-\sqrt{\theta_h}}\left\{\frac{H_N^{\#}(0)}{NT} + \varepsilon_{S,h,N}\right\}.$$

*Under Assumption G.9, also*

$$\frac{1}{T}\int_0^T \mathbb{E}_t[\mathsf{V}_N^{\mathrm{Lap}}(X(t))]\,\mathrm{d}t \le \frac{L_h}{1-\sqrt{\theta_h}}\left\{\frac{H_N^{\#}(0)}{NT} + \varepsilon_{S,h,N}\right\} + \tau_{h,N} + \varepsilon_{V,h,N}.$$

*Proof.* The relative coercivity bound and the Stein comparison imply

$$(1-\sqrt{\theta_h})\mathcal{W}_h^{\mathrm{Lap}}(x) \le \mathcal{J}_h^{\mathrm{Lap}}(x) \le \mathsf{S}_N^{\mathrm{Lap}}(x) + \varepsilon_{S,h,N}.$$

Expectation, time integration, and (98) prove the first bound after division by $1-\sqrt{\theta_h} > 0$. The localization argument gives $\mathcal{V}_h^{\mathrm{Lap}} \le L_h \mathcal{W}_h^{\mathrm{Lap}} + \tau_{h,N}$. Combining it with $\mathsf{V}_N^{\mathrm{Lap}} \le \mathcal{V}_h^{\mathrm{Lap}} + \varepsilon_{V,h,N}$ and the first bound proves the second. □

**Remark G.15** (Meaning of the Laplace rates)**.** *The weighted estimate retains the model's local Laplace scale inside the energy. The center-velocity estimate uses a local upper scale and an explicit tail energy to recover the quantity applied in an additional drift step. Absolute scale alignment contributes the term $L_h\Delta_h^2$. Relative alignment allows the mismatch to be absorbed into the dissipative term, as in Corollary G.14. Thus the absolute mismatch bound describes one sufficient route to a small residual. It is not a lower bound on the limiting drift, and does not imply a positive asymptotic error floor.*

**Remark G.16** (Local scale variation and circulation)**.** *The sharp score $\sigma_{\alpha,h}$ is a spatial gradient. On a smooth two-dimensional open set, the decomposition gives*

$$\operatorname{curl} u_x = \nabla a_{x,h} \times b_x^{\#} + \nabla(a_{\nu,h} - a_{x,h}) \times \sigma_{\nu,h},$$

*where $\nabla a \times f := \partial_1 a\, f_2 - \partial_2 a\, f_1$. Indeed, the product rule gives $\operatorname{curl}(af) = \nabla a \times f + a \operatorname{curl} f$, and both $b_x^{\#}$ and $\sigma_{\nu,h}$ have zero curl. The mismatch $e_x = (a_{\nu,h} - a_{x,h})\sigma_{\nu,h}$ depends on differences between the data and model scales. Circulation depends on spatial derivatives of those scales. These are related but distinct quantities. Even when $e_x = 0$, the term $\nabla a_{x,h} \times b_x^{\#}$ may be nonzero. For the Gaussian companion kernel, $a_{\alpha,h} \equiv h^2$; the mismatch and both scale-derivative terms vanish.*

### G.5 Checkable Forms of The Laplace Assumptions

We first record the normalization and tails of the sharp kernel. Let

$$\overline{L}_h(u) := \frac{L_h(u)}{Z_{\#,h}}.$$

**Lemma G.17** (Sharp-kernel moments, scores, and tails)**.** *The Laplace companion kernel satisfies*

$$Z_{\#,h} = (d+1)h^2, \qquad \int \|u\|^2 \,\overline{L}_h(u)\,\mathrm{d}u = d(d+3)h^2. \tag{99}$$

*For every probability measure $\alpha$,*

$$\|\sigma_{\alpha,h}(z)\| \le h^{-1}.$$

*For $s \ge 0$, let*

$$p_d(s) := e^{-s}\left\{\frac{d}{d+1}\sum_{k=0}^{d}\frac{s^k}{k!} + \frac{1}{d+1}\sum_{k=0}^{d-1}\frac{s^k}{k!}\right\}.$$

*Then*

$$\int_{\|u\|>hs} \overline{L}_h(u)\,\mathrm{d}u = p_d(s). \tag{100}$$

*Proof.* If $U$ has density $K_h$, radial integration gives

$$\mathbb{E}[g(\|U\|/h)] = c_d\left|\mathbb{S}^{d-1}\right|\int_0^\infty g(s)s^{d-1}e^{-s}\,\mathrm{d}s$$

for every bounded measurable $g$. For $m \ge 1$, integration by parts gives

$$\int_0^\infty s^m e^{-s}\,\mathrm{d}s = m\int_0^\infty s^{m-1}e^{-s}\,\mathrm{d}s,$$

with initial value $\int_0^\infty e^{-s}\,\mathrm{d}s = 1$. Thus the integral of $s^m e^{-s}$ is $m!$. Taking $g = 1$ in the radial formula yields $c_d\left|\mathbb{S}^{d-1}\right| = 1/(d-1)!$. Consequently $S := \|U\|/h$ has density $s^{d-1}e^{-s}/(d-1)!$ and $\mathbb{E}S^k = (d+k-1)!/(d-1)!$ for nonnegative integers $k$. It follows that

$$\mathbb{E}S = d, \quad \mathbb{E}S^2 = d(d+1), \quad \mathbb{E}S^3 = d(d+1)(d+2).$$

Since $L_h(u) = h^2(1 + \|u\|/h)K_h(u)$,

$$Z_{\#,h} = h^2(1 + \mathbb{E}S) = (d+1)h^2,$$

and

$$\int \|u\|^2 \overline{L}_h(u)\,\mathrm{d}u = h^2 \frac{\mathbb{E}S^2 + \mathbb{E}S^3}{d+1} = d(d+3)h^2.$$

The sharp-kernel gradient satisfies

$$\frac{\nabla L_h(u)}{L_h(u)} = -\frac{u}{h(\|u\| + h)}.$$

Its norm is at most $1/h$, including at $u = 0$ by continuity. Furthermore,

$$\sigma_{\alpha,h}(z) = \int \frac{L_h(z-y)}{R_{\alpha,h}(z)} \frac{\nabla L_h(z-y)}{L_h(z-y)}\,\alpha(\mathrm{d}y).$$

The coefficients define a probability measure, so the triangle inequality gives $\|\sigma_{\alpha,h}\| \le 1/h$.

For the tail, radial integration gives

$$\int_{\|u\|>hs} \overline{L}_h(u)\,\mathrm{d}u = \frac{1}{(d+1)(d-1)!} \int_s^\infty (t^d + t^{d-1})e^{-t}\,\mathrm{d}t.$$

Writing $I_m(s) = \int_s^\infty t^m e^{-t}\,\mathrm{d}t$, integration by parts gives $I_m(s) = s^m e^{-s} + mI_{m-1}(s)$ and $I_0(s) = e^{-s}$. Successive substitution in this recurrence yields $I_m(s) = m!e^{-s}\sum_{k=0}^m s^k/k!$. Substituting the formulas for $I_d$ and $I_{d-1}$ proves (100). □

**Corollary G.18** (An integrated shell-alignment condition)**.** *Suppose, uniformly along the trajectory,*

$$\|\sigma_{\nu,h}\|_\infty \le G_h, \qquad \int |\bar{r}_{\nu,h}(z) - \bar{r}_{x,h}(z)|^2\,\rho^{\#}_{x,h}(z)\,\mathrm{d}z \le \delta^2_{r,h}.$$

*Then the weighted mismatch bound holds with*

$$\Delta_h^2 = G_h^2 \delta^2_{r,h}.$$

*One may always use $G_h = h^{-1}$, while a smaller bound can be used when the data score has additional regularity.*

*Proof.* The sharp-score decomposition gives

$$e_x(z) = h\{\bar{r}_{\nu,h}(z) - \bar{r}_{x,h}(z)\}\sigma_{\nu,h}(z).$$

Since $a_{x,h} \ge h^2$,

$$\frac{\|e_x(z)\|^2}{a_{x,h}(z)} \le G_h^2 \,|\bar{r}_{\nu,h}(z) - \bar{r}_{x,h}(z)|^2\,.$$

Integration against $\rho^{\#}_{x,h}$ proves the stated mismatch bound. The universal choice of $G_h$ follows from Lemma G.17. □

**Proposition G.19** (Confinement and sharp-kernel tails)**.** *Suppose $\nu$ is supported in $B(0, R_\nu)$ and, along the trajectory, $\max_i \|X_i(t)\| \le R_T$ with deterministic $R_T < \infty$. For every $r > 0$, Assumption G.9 holds with*

$$\Omega_x = B(0, R_T + r), \qquad L_h = h(2R_T + r + h), \qquad \tau_{h,N} = (R_\nu + R_T)^2 p_d(r/h). \tag{101}$$

*Proof.* For $z \in \Omega_x$, each distance $\|z - x_j\|$ is at most $2R_T + r$. Its kernel-weighted average satisfies the same upper bound. Thus $a_{x,h}(z) \le h(2R_T + r + h)$.

Write $M_{\alpha,h}(z) = m_{\alpha,h}(z) - z$, where $m_{\alpha,h}$ is the kernel-weighted mean of $y$. The support bounds imply $\|m_{\nu,h}(z)\| \le R_\nu$ and $\|m_{\mu_x^N,h}(z)\| \le R_T$. The $-z$ terms cancel in their difference, giving $\|u_x(z)\| \le R_\nu + R_T$. Also,

$$\rho^{\#}_{x,h}(z) = \frac{1}{N}\sum_i \overline{L}_h(z - x_i).$$

If $\|z\| > R_T + r$, then $\|z - x_i\| > r$ for every $i$. Therefore

$$\int_{\Omega_x^c} \rho_{x,h}^{\#}(z)\,\mathrm{d}z \le \frac{1}{N}\sum_i \int_{\|z-x_i\|>r} \overline{L}_h(z-x_i)\,\mathrm{d}z$$
$$= p_d(r/h).$$

Multiplying by $(R_\nu + R_T)^2$ proves the tail-energy bound. □

**Remark G.20** (Finite-horizon confinement)**.** *For bounded data support, a deterministic initial bound $\max_i \|X_i(0)\| \le R_0$ gives a finite-horizon bound without assuming a uniform density lower bound. Let $R(t) = \max_i \|X_i(t)\|$. The leave-one-out weighted model mean lies in the convex hull of the other centers, so $\left\|u_{X(t),-i}(X_i(t))\right\| \le R_\nu + R(t)$. Integration for each particle, followed by maximization, yields*

$$R(t) \le R_0 + \int_0^t \{R_\nu + R(s)\}\,\mathrm{d}s.$$

*Gronwall's inequality therefore gives $R(t) \le (R_0 + R_\nu)e^t - R_\nu$. This ensures finite constants on a fixed horizon where the regular flow exists. Uniform constants on a growing horizon require a stronger confinement estimate.*

We next compare the leave-one-out particle sums with the full-field integrals. For a fixed configuration, set

$$k_i(z) := \overline{L}_h(z - x_i), \qquad f_i(z) := \mathcal{A}_{\#} u_{x,-i}(z),$$
$$g_i(z) := \|u_{x,-i}(z)\|^2, \qquad d_i(z) := u_{x,-i}(z) - u_x(z).$$

The functions $f_i$ are defined almost everywhere and take their classical values at $x_i$. Define

$$q_S(x) := \frac{1}{N}\sum_i \left| f_i(x_i) - \int k_i(z) f_i(z)\,\mathrm{d}z \right|,$$
$$q_V(x) := \frac{1}{N}\sum_i \left| g_i(x_i) - \int k_i(z) g_i(z)\,\mathrm{d}z \right|,$$
$$\ell_1(x) := \frac{1}{N}\sum_i \int k_i(z)\,\|d_i(z)\|\,\mathrm{d}z,$$
$$\ell_V(x) := \frac{1}{N}\sum_i \int k_i(z)\{\|u_{x,-i}(z)\| + \|u_x(z)\|\}\,\|d_i(z)\|\,\mathrm{d}z.$$

**Lemma G.21** (Integrated leave-one-out comparison)**.** *Assume the displayed integrals are finite and integration by parts against each $k_i$ is justified for $d_i$. Then*

$$\left| \mathsf{S}_N^{\mathrm{Lap}}(x) - \mathcal{J}_h^{\mathrm{Lap}}(x) \right| \le q_S(x) + \frac{2}{h}\ell_1(x), \tag{102}$$
$$\left| \mathsf{V}_N^{\mathrm{Lap}}(x) - \mathcal{V}_h^{\mathrm{Lap}}(x) \right| \le q_V(x) + \ell_V(x). \tag{103}$$

*Proof.* Since $\rho_{x,h}^{\#} = N^{-1}\sum_i k_i$ and the Stein operator is linear,

$$\frac{1}{N}\sum_i \int k_i f_i - \mathcal{J}_h^{\mathrm{Lap}}(x) = \frac{1}{N}\sum_i \int k_i\,\mathcal{A}_{\#} d_i.$$

Integration by parts gives, for each $i$,

$$\int k_i\,\mathcal{A}_{\#} d_i = \int k_i\,\nabla\cdot d_i + \int k_i \sigma_{\nu,h}\cdot d_i$$
$$= -\int \nabla k_i \cdot d_i + \int k_i \sigma_{\nu,h}\cdot d_i$$
$$= \int k_i(\sigma_{\nu,h} - \nabla \log k_i)\cdot d_i.$$

Lemma G.17 bounds both score norms by $1/h$. Thus the absolute value of the last expression is at most $(2/h)\int k_i \|d_i\|$. Averaging over $i$ gives

$$\left|\frac{1}{N}\sum_i \int k_i f_i - \mathcal{J}_h^{\mathrm{Lap}}(x)\right| \le \frac{2}{h}\ell_1(x).$$

The difference between $\mathsf{S}_N^{\mathrm{Lap}}$ and the averaged integral on the left is at most $q_S(x)$. The triangle inequality proves (102).

For the energy comparison, the identity $\|v\|^2 - \|w\|^2 = (v-w)\cdot(v+w)$ gives

$$\left|\|u_{x,-i}\|^2 - \|u_x\|^2\right| \le \|d_i\|\,(\|u_{x,-i}\| + \|u_x\|).$$

Multiplying by $k_i$, integrating, and averaging yields

$$\left|\frac{1}{N}\sum_i \int k_i g_i - \mathcal{V}_h^{\mathrm{Lap}}(x)\right| \le \ell_V(x).$$

The difference between $\mathsf{V}_N^{\mathrm{Lap}}$ and the averaged integral is at most $q_V(x)$. The triangle inequality proves (103). □

**Lemma G.22** (Local quadrature with an explicit tail)**.** *Let $f$ be twice continuously differentiable on a neighborhood of $\overline{B}(x_i, r_i)$, where $r_i > 0$, and assume $\sup_{z\in B(x_i,r_i)} \left\|D^2 f(z)\right\|_{\mathrm{op}} \le B_i$. Suppose the tail integral below is finite. Then*

$$\left|f(x_i) - \int k_i(z) f(z)\,\mathrm{d}z\right| \le \frac{B_i}{2}\int_{\|u\|\le r_i} \|u\|^2\,\overline{L}_h(u)\,\mathrm{d}u + \int_{\|u\|>r_i} \overline{L}_h(u)\,|f(x_i+u) - f(x_i)|\,\mathrm{d}u. \tag{104}$$

*Proof.* Change variables $z = x_i + u$ and split the difference integral into $\|u\| \le r_i$ and its complement. In the ball, Taylor's formula is

$$f(x_i+u) - f(x_i) = \nabla f(x_i)\cdot u + \int_0^1 (1-s) u^\top D^2 f(x_i + su) u\,\mathrm{d}s.$$

The ball and $\overline{L}_h$ are symmetric, so the integral of the linear term over the ball vanishes. Taking absolute values of the remaining integral and using $\int_0^1 (1-s)\,\mathrm{d}s = 1/2$ gives the first term of (104). Taking the absolute value inside the integral over the complement gives the second term. Adding the two estimates proves the result. □

For a collision-free configuration whose centers lie outside the finite support of the data, choose

$$0 < r_i < \min\left\{\min_{j\ne i}\|x_i - x_j\|, \operatorname{dist}(x_i, \operatorname{supp}\nu)\right\}.$$

Each closed ball then has a neighborhood avoiding all other particle centers and all data atoms. Both $f_i$ and $g_i$ are smooth on this neighborhood. Applying Lemma G.22 to each $f_i$ and averaging bounds $q_S$; applying it to each $g_i$ bounds $q_V$. By (99), their local second-moment contributions are at most

$$\frac{d(d+3)h^2}{2N}\sum_i B_i$$

with the appropriate Hessian bounds for each family. The tail integrals remain explicit. For a continuous data measure, the same conclusion holds whenever the data contribution has the stated local differentiability. These comparisons require no derivative of the full field at a particle center.

**Corollary G.23** (Quadrature tails under bounded support)**.** *Suppose $\nu$ is supported in $B(0, R_\nu)$ and all centers lie in $B(0, R_T)$. Write $B = R_\nu + R_T$. Wherever the weak derivative is represented pointwise,*

$$|f_i(z)| \le \frac{(2d+1)B}{h}, \qquad 0 \le g_i(z) \le B^2.$$

*The same bounds hold at the classical evaluation point $x_i$. Thus the tail term in (104) is at most $2(2d+1)Bh^{-1}p_d(r_i/h)$ for $f_i$, and at most $B^2 p_d(r_i/h)$ for $g_i$.*

*Proof.* Let $\alpha$ be supported in $B(0, R_\alpha)$ and let $m_{\alpha,h}(z)$ be its kernel-weighted mean of $y$. Differentiating the quotient defining this mean gives, almost everywhere,

$$Dm_{\alpha,h}(z) = \int \frac{K_h(z-y)}{Q_{\alpha,h}(z)}(y - m_{\alpha,h}(z))(\nabla_z \log K_h(z-y))^\top \, \alpha(\mathrm{d}y).$$

The Laplace score has norm $1/h$ away from its origin, and $\|y - m_{\alpha,h}(z)\| \le 2R_\alpha$. The triangle inequality for the integral therefore gives $\|Dm_{\alpha,h}(z)\|_{\mathrm{op}} \le 2R_\alpha/h$. The formula and bound are understood almost everywhere; differentiation under the integral follows from the locally bounded denominator and the bounded weak derivatives of the kernel.

Since $u_{x,-i} = m_{\nu,h} - m_{\mu^N_{x,-i},h}$, it follows that $\|u_{x,-i}\| \le B$ and $\|Du_{x,-i}\|_{\mathrm{op}} \le 2B/h$. The trace bound $|\mathrm{tr}\, A| \le d\,\|A\|_{\mathrm{op}}$ and $\|\sigma_{\nu,h}\| \le 1/h$ give

$$|f_i| \le |\nabla\cdot u_{x,-i}| + \|\sigma_{\nu,h}\|\,\|u_{x,-i}\| \le \frac{2dB}{h} + \frac{B}{h}.$$

Also $g_i = \|u_{x,-i}\|^2$ lies in $[0, B^2]$. At a classical evaluation point the derivative bound follows from the same local Lipschitz bound. Hence $|f_i(z) - f_i(x_i)| \le 2(2d+1)B/h$ almost everywhere, and $|g_i(z) - g_i(x_i)| \le B^2$. Integration outside the ball and (100) prove both tail estimates. □

**Lemma G.24** (Leave-one-out perturbations of field values)**.** *For fixed $i$ and $z$, set*

$$c_i(z) = K_h(z - x_i), \qquad D_i(z) = \sum_{j\ne i} K_h(z - x_j).$$

*Then*

$$d_i(z) = \frac{c_i(z)}{D_i(z) + c_i(z)}\{x_i - z - M_{\mu^N_{x,-i},h}(z)\}. \tag{105}$$

*At $z = x_i$ this also reads*

$$d_i(x_i) = -\frac{K_h(0)}{D_i(x_i)} M_{\mu^N_x,h}(x_i). \tag{106}$$

*If $Q_{\mu^N_{x,-i},h}(x_i) \ge \lambda > 0$, then*

$$\|d_i(x_i)\| \le \frac{c_d}{(N-1)\lambda h^d}\left\|M_{\mu^N_x,h}(x_i)\right\|.$$

*Proof.* Let $P_i(z) = \sum_{j\ne i}(x_j - z)K_h(z - x_j)$. The full model mean shift is $(P_i + (x_i - z)c_i)/(D_i + c_i)$, and the leave-one-out model mean shift is $P_i/D_i$. Their difference is

$$\frac{P_i + (x_i - z)c_i}{D_i + c_i} - \frac{P_i}{D_i} = \frac{c_i}{D_i + c_i}\left(x_i - z - \frac{P_i}{D_i}\right).$$

The data terms cancel in $u_{x,-i} - u_x$, leaving exactly this full-minus-leave-one-out model difference. This proves (105). At $z = x_i$, the full model mean shift is $P_i/(D_i + K_h(0))$, so the same difference equals $-K_h(0)P_i/\{D_i(D_i + K_h(0))\}$, which proves (106). Finally, $D_i(x_i) = (N-1)Q_{\mu^N_{x,-i},h}(x_i) \ge (N-1)\lambda$ and $K_h(0) = c_d h^{-d}$. Taking norms proves the last bound. □

The value identity also bounds $\ell_1$ and $\ell_V$ when the denominators and the local weighted moments are controlled throughout the regions of integration. The denominator at a center is controlled by the same local-occupancy mechanism as in the conservative analysis.

**Lemma G.25** (Leave-one-out Laplace denominator control)**.** *Fix $r_0 > 0$. For the Laplace kernel,*

$$K_h(u) \ge c_d h^{-d} e^{-r_0} \qquad \textit{whenever } \|u\| \le r_0 h.$$

*For a configuration $x$, define*

$$n_i^{(-)}(x) := \#\{j \ne i : \|x_j - x_i\| \le r_0 h\}.$$

*If*

$$n_i^{(-)}(x) \ge \alpha(N-1)h^d \qquad \textit{for every } i,$$

*then*

$$Q_{\mu^N_{x,-i},h}(x_i) \ge c_d e^{-r_0}\alpha \qquad \textit{for every } i.$$

*Proof.* By definition,

$$Q_{\mu^N_{x,-i},h}(x_i) = \frac{1}{N-1}\sum_{j\neq i} K_h(x_i - x_j).$$

For every $j \neq i$ satisfying $\|x_j - x_i\| \le r_0 h$,

$$K_h(x_i - x_j) = c_d h^{-d} \exp\left(-\frac{\|x_i - x_j\|}{h}\right) \ge c_d h^{-d} e^{-r_0}.$$

Therefore

$$Q_{\mu^N_{x,-i},h}(x_i) \ge \frac{1}{N-1} n_i^{(-)}(x) c_d h^{-d} e^{-r_0}.$$

Using $n_i^{(-)}(x) \ge \alpha(N-1)h^d$,

$$Q_{\mu^N_{x,-i},h}(x_i) \ge \frac{1}{N-1}\{\alpha(N-1)h^d\} c_d h^{-d} e^{-r_0} = c_d e^{-r_0}\alpha.$$

This proves the claim. □

### G.6 Bandwidth and Step-size Choices for Laplace Drifting

For $H_N^{\#}(0) \le \kappa_0 N$ and $T = N$, write the right-hand side of (97) as

$$\mathcal{R}_{N,h} := \frac{2\kappa_0 L_h}{N} + L_h \Delta_h^2 + 2L_h \varepsilon_{S,h,N} + \tau_{h,N} + \varepsilon_{V,h,N}. \tag{107}$$

The bandwidth enters through all five terms. The local scale and tail energy can be made explicit under confinement. For example, with the hypotheses of Proposition G.19, choose $r = ch\log N$ for a fixed $c > 0$ and $N \ge 2$. Then

$$L_h = h\{2R_T + ch\log N + h\}, \qquad \tau_{h,N} = (R_\nu + R_T)^2 p_d(c\log N).$$

For fixed $d, c$, the finite polynomial in (100) gives

$$\tau_{h,N} = O\big((R_\nu + R_T)^2\{1 + (\log N)^d\}N^{-c}\big).$$

The remaining terms require weighted shell alignment and the integrated approximation estimates. A sequence $h = h_N$ gives a vanishing root residual whenever

$$\frac{L_{h_N}}{N} \to 0, \qquad L_{h_N}\Delta_{h_N}^2 \to 0, \qquad L_{h_N}\varepsilon_{S,h_N,N} \to 0, \qquad \tau_{h_N,N} + \varepsilon_{V,h_N,N} \to 0,$$

with bounded $\kappa_0$ and all hypotheses valid on $[0, N]$. Under relative alignment, one instead uses Corollary G.14, retaining the factor $(1 - \sqrt{\theta_{h_N}})^{-1}$.

The condition $Nh^d \gtrsim \log N$ can ensure simultaneous local occupancy under an appropriate lower local mass condition. It does not alone control the quadrature tails, the integrated leave-one-out errors, or confinement over $[0, N]$. Thus the bandwidth is chosen by balancing the terms in (107) over a range where their dependence on $N$ and $h$ is controlled.

The explicit step size enters through $\widetilde{x}_i = x_i + \eta u_{x,-i}(x_i)$. If $\tau$ is independent and uniform on $[0, N]$, the identity coupling gives

$$\mathbb{E}W_2^2(\mu^N_{X(\tau)}, \widetilde{\mu}^N_{X(\tau)}) \le \eta^2 \mathbb{E}\mathsf{V}_N^{\mathrm{Lap}}(X(\tau)) \le \eta^2 \mathcal{R}_{N,h}.$$

Taking square roots yields a root mean squared correction at most $\eta\sqrt{\mathcal{R}_{N,h}}$. For approximation of the interacting-particle ODE, the relevant Lipschitz constant is that of the full configuration vector field. The admissible step size must account for this constant and the time horizon as described in Section H.

### G.7 Comparison Between the Conservative and Non-conservative Results

The conservative field $b_x = s_\rho - s_x$ is a score mismatch, and its KDE-averaged Stein drift equals the smoothed Fisher discrepancy. The joint-entropy argument and quadrature comparison give

$$\frac{1}{N}\int_0^N \mathbb{E}_t[\mathsf{V}_N(X(t))]\,\mathrm{d}t \le \frac{\kappa_0}{N} + \frac{a_1\Lambda_N(h)}{Nh^{d+2}} + C_N(h)h^2.$$

The reciprocal-KDE term comes from differentiation of a particle's own contribution to the full model KDE. The exact expectation bounds in Section F control this term under lower local mass and distortion conditions. The full-support initialization result identifies a regime in which its coefficient must grow with $N$.

The optimized conservative root rate is $N^{-1/(d+4)}$ when $\kappa_0$, $a_1\Lambda_N(h)$, and $C_N(h)$ are uniformly bounded. If instead $C_N(h) = O(h^{-\beta})$, with $0 \le \beta < 2$ and the other constants uniformly bounded, the root rate is $N^{-(2-\beta)/(2(d+4-\beta))}$.

For Laplace drifting, the decomposition $u_x = a_{x,h}b_x^{\#} + e_x$ gives a weighted coercivity relation. The scale $a_{x,h}$ grows in the tails, so the unweighted center-velocity estimate uses local scale control and tail energy:

$$\frac{1}{N}\int_0^N \mathbb{E}_t[\mathsf{V}_N^{\mathrm{Lap}}(X(t))]\,\mathrm{d}t \le \frac{2\kappa_0 L_h}{N} + L_h\Delta_h^2 + 2L_h\varepsilon_{S,h,N} + \tau_{h,N} + \varepsilon_{V,h,N}.$$

Absolute alignment bounds the weighted mismatch additively; relative alignment absorbs it into coercivity. The leave-one-out dynamics remove the reciprocal self-interaction correction from the joint entropy identity. Their comparison with full-field integrals is expressed through local leave-one-out quadrature and integrated field-value perturbations.

The reference densities also differ. The conservative analysis uses $\rho = K_h * \nu$, while the Laplace analysis uses

$$\rho_{\nu,h}^{\#}(z) = \frac{1}{(d+1)h^2}\int L_h(z-y)\,\nu(\mathrm{d}y), \qquad L_h(u) = h(\|u\|+h)K_h(u).$$

The relation $\nabla L_h(z-y) = (y-z)K_h(z-y)$ connects the displacement field to this sharp-smoothed score. Both analyses control time-averaged residual drift under explicit regularity and stability conditions. The additional arguments needed to obtain distributional convergence or uniform finite-step guarantees are separate from these entropy estimates.

## H Stop-gradient Training and The Particle ODE

We briefly explain how the continuous-time particle dynamics arise from the practical stop-gradient training rule. Let $g_\theta : \mathcal{Z} \to \mathbb{R}^d$ be the generator and let

$$x_i^k := g_{\theta_k}(\xi_i), \qquad i = 1, \dots, N,$$

be the generated batch at training step $k$, for fixed latent variables $\xi_1, \dots, \xi_N$. Let $v_{x^k}$ denote the drift field computed from the current generated batch

$$x^k = (x_1^k, \dots, x_N^k)$$

and the data batch. In the original displacement-based drifting method (Deng et al., 2026), $v_{x^k}$ is the non-conservative field $u^{\mathrm{disp}}_{\nu,\mu_{x^k},h}$; in the conservative method studied in this paper, it is the KDE-score field $b_{x^k}$. The practical target for the $i$-th generated sample is

$$y_i^k := \mathrm{sg}\{x_i^k + \eta v_{x^k}(x_i^k)\}.$$

Here sg denotes the stop-gradient operator. For any quantity $a(\theta)$ appearing in the computational graph,

$$\mathrm{sg}\{a(\theta)\} = a(\theta) \quad \text{in the forward pass,} \qquad \nabla_\theta\,\mathrm{sg}\{a(\theta)\} = 0 \quad \text{in the backward pass.}$$

Thus sg preserves the numerical value of its argument but removes all derivatives through that argument. In particular, the target $y_i = \mathrm{sg}\{x_i + \eta v_x(x_i)\}$ is evaluated as $x_i + \eta v_x(x_i)$, but is treated as a constant when differentiating the regression loss with respect to the generator parameters.

The corresponding regression loss is

$$\mathcal{L}_k(\theta) := \frac{1}{2N} \sum_{i=1}^{N} \left\| g_\theta(\xi_i) - y_i^k \right\|^2 . \tag{108}$$

The stop-gradient operation means that $y_i^k$ is treated as a constant when differentiating $\mathcal{L}_k$ with respect to $\theta$. Thus the derivative of the loss is

$$\nabla_\theta \mathcal{L}_k(\theta) = \frac{1}{N} \sum_{i=1}^{N} D_\theta g_\theta(\xi_i)^\top \{ g_\theta(\xi_i) - y_i^k \},$$

with no derivative through the map $x^k \mapsto v_{x^k}$. Therefore the training step is a frozen-target regression problem: the generator is trained to move its current outputs toward the fixed points

$$x_i^k + \eta v_{x^k}(x_i^k).$$

The idealized exact-regression limit is the limit in which this regression problem is solved exactly at every step and the generator class is rich enough to interpolate the frozen targets on the sampled latent variables. In that limit, the next generator satisfies

$$g_{\theta_{k+1}}(\xi_i) = y_i^k = x_i^k + \eta v_{x^k}(x_i^k), \qquad i = 1, \dots, N.$$

Consequently the generated particles obey the explicit Euler update

$$x_i^{k+1} = x_i^k + \eta v_{x^k}(x_i^k), \qquad i = 1, \dots, N. \tag{109}$$

If the regression is not exact, then (109) holds with an additional approximation error

$$e_i^k := g_{\theta_{k+1}}(\xi_i) - \{ x_i^k + \eta v_{x^k}(x_i^k) \}.$$

Our analysis neglects this optimization and approximation error and studies the idealized particle system (109).

Finally, define $t_k = k\eta$ and let $X_i^\eta(t)$ be the piecewise-linear interpolation of the Euler iterates. Under the usual consistency and stability conditions for explicit Euler schemes, for example local Lipschitz continuity of the interacting-particle vector field

$$F_i(x) := v_x(x_i), \qquad x = (x_1, \dots, x_N),$$

the interpolation $X^\eta(t)$ converges as $\eta \to 0$ on finite time intervals to the solution of

$$\dot{X}_i(t) = F_i(X(t)) = v_{X(t)}(X_i(t)), \qquad i = 1, \dots, N.$$

Here, $v_x$ can either be the conservative KDE-score field or the non-conservative leave-one-out Laplace field. These are the continuous-time particle dynamics analyzed in the main text. Thus stop-gradient is used to justify the frozen-field Euler update; once this particle ODE is assumed, the entropy identities and convergence bounds are properties of the resulting interacting-particle flow.

We emphasize that the passage from the stop-gradient Euler update to the continuous-time ODE is an idealized modeling step. Let $F(x) := (F_1(x), \dots, F_N(x))$. The exact-regression stop-gradient update is

$$x^{k+1} = x^k + \eta F(x^k).$$

Suppose $F$ is $L_{h,\mathcal{D}}$-Lipschitz on a region $\mathcal{D}$ containing the Euler iterates and the exact trajectory, and $\sup_{\mathcal{D}} \|F\| \le B_{h,\mathcal{D}}$. For brevity write $L = L_{h,\mathcal{D}}$ and $B = B_{h,\mathcal{D}}$. The exact solution satisfies, for $s \in [0, \eta]$, $\|X(t+s) - X(t)\| \le Bs$, and therefore

$$\begin{aligned} \|X(t+\eta) - X(t) - \eta F(X(t))\| &\le \int_0^\eta \|F(X(t+s)) - F(X(t))\| \, \mathrm{d}s \\ &\le \int_0^\eta LBs \, \mathrm{d}s = \frac{1}{2} LB\eta^2. \end{aligned}$$

Writing $e_k = \left\|x^k - X(k\eta)\right\|$ with $e_0 = 0$, the Lipschitz bound for one Euler step gives

$$e_{k+1} \le (1 + L\eta)e_k + \frac{1}{2}LB\eta^2.$$

For $L > 0$, repeated substitution and the finite geometric sum yield

$$e_k \le \frac{1}{2}LB\eta^2 \sum_{j=0}^{k-1}(1 + L\eta)^j = \frac{B\eta}{2}\{(1 + L\eta)^k - 1\} \le \frac{B\eta}{2}(e^{Lk\eta} - 1).$$

If $L = 0$, the field is constant on $\mathcal{D}$ and Euler is exact. Consequently, for fixed $T, h, \mathcal{D}$,

$$\max_{k\eta \le T} \left\|x^k - X(k\eta)\right\| \le \frac{B_{h,\mathcal{D}}\eta}{2}(e^{L_{h,\mathcal{D}}T} - 1).$$

The constants involve derivatives of the full configuration map, including its dependence on all KDE centers. When $h$ and $T$ vary with $N$, a small value of $\eta L_{h,\mathcal{D}}$ alone does not ensure that this global bound vanishes. A sufficient condition from the displayed estimate is

$$\eta B_{h,\mathcal{D}}(e^{L_{h,\mathcal{D}}T} - 1) \longrightarrow 0.$$

Stronger dynamical stability can give a more favorable dependence on the horizon. The continuous-time entropy rates are therefore interpreted with their independent regularity and stability conditions when applied to a finite-step training procedure.

If the regression step is not exact, the particle update has the form $x^{k+1} = x^k + \eta F(x^k) + e^k.$, as mentioned above. Under the same Lipschitz assumptions, the discrete-to-continuous error contains an additional accumulated term of size

$$\sum_{k\eta \le T} e^{L_{h,\mathcal{D}}(T - k\eta)} \left\|e^k\right\|.$$

Thus tracking the practical neural-network training dynamics would require bounds on optimization error, approximation error, minibatch noise, and the $h$-dependent stability constant. These effects are outside the scope of the continuous-time entropy analysis.