CONVERGENCE OF SINGULAR DYNAMICS DRIVEN BY MIXED LOGARITHMIC HEAT GRADIENTS

KANG LIU[1] AND ENRIQUE ZUAZUA[2,3,4]

Abstract. We study the terminal behavior of singular deterministic and stochastic dynamics driven by convex or extrapolative combinations of logarithmic heat gradients. These equations arise in exact-score guided diffusion sampling, including classifier-free guidance. The rescaled potentials converge to weighted sums of squared support distances, but extrapolative mixing can produce limiting directions that depend on probability masses and lie outside the Clarke subdifferential of the limiting potential. A componentwise chain rule gives dissipation for the limiting differential inclusion. Posterior relative entropy yields finite deterministic kinetic action in logarithmic time, while a conditional free-energy estimate bounds expected energy increments uniformly in the restart time. For compactly supported laws, these estimates give energy convergence and equilibrium localization, almost surely for each fixed stochastic initialization; extrapolation requires a lower small-ball mass bound on the reference law. For finite atomic measures, we prove that the equilibrium set is finite for arbitrary site configurations. Every deterministic trajectory therefore converges to an equilibrium, and stochastic convergence holds almost surely for each fixed initialization, including endpoints at projection interfaces. At empirical endpoints with a unique nearest atom in each active support, deterministic trajectories admit a square-root heat-time expansion with an exponentially small remainder. The exponent 1/2 is sharp when the leading coefficient is nonzero. Square-integrable perturbations in logarithmic time preserve deterministic localization and empirical endpoint convergence for both dynamics. For nested supports, the equilibrium geometry further constrains terminal samples in classifier-free guidance. These results provide a geometric and dynamical framework for understanding terminal convergence in guided diffusion models.

1. Introduction

We study the terminal behavior of singular nonautonomous dynamics driven by mixed logarithmic heat gradients. The problem is motivated by guided diffusion sampling in generative AI, where

[1]Université Bourgogne Europe, CNRS, Institut de Mathématiques de Bourgogne, 21000 Dijon, France

[2]Friedrich–Alexander–Universität Erlangen–Nürnberg, Department of Mathematics, Chair for Dynamics, Control, Machine Learning, and Numerics, 91058 Erlangen, Germany

[3]Universidad Autónoma de Madrid, Departamento de Matemáticas, 28049 Madrid, Spain

[4]Chair of Computational Mathematics, Fundación Deusto, 48007 Bilbao, Basque Country, Spain

Email address: kang.liu@u-bourgogne.fr, enrique.zuazua@fau.de.

2020 *Mathematics Subject Classification.* Primary 34D05, 35B40; Secondary 34A60, 35F21, 35K05, 60H10, 68T07.

Key words and phrases. Singular dynamics, differential inclusions, conditional free energy, posterior relative entropy, geometric asymptotics, score mixing, guided diffusion.

Corresponding author: Kang Liu.

E. Zuazua was partially supported by the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (grant agreement No. 101096251-CoDeFeL); by the Alexander von Humboldt Professorship program; the European Union's Horizon Europe MSCA project ModConFlex (HORIZON-MSCA-2021-DN-01, project 101073558); the Transregio 154 Project "Mathematical Modelling, Simulation and Optimization Using the Example of Gas Networks" of the DFG; the AFOSR 24IOE027 project; the SURE-AI Norwegian Centre for Sustainable, Risk-Averse, and Ethical AI grant 357482, Research Council of Norway; by the Grant PID2023-146872OB-I00-DyCMaMod of MICIU (Spain), by the COST Actions CA24122 – Multiscale Stochastics, Patterns, and Analysis of Combinatorial Environments and CA24136 – Interactions between Control Theory and Machine Learning and by the Basque Government through the consolidated research group "CMRG-Computational Mathematics Research Group" under Grant IT1856-26.

samples are constructed by reversing a noise-adding evolution. This reversal uses the logarithmic gradients of the evolving densities, called scores [24, 40, 41]. Under Gaussian noising, the heat equation regularizes the data laws, and the resulting scores drive sampling in decreasing heat time [31]. Guidance combines these scores to favor a prescribed condition or compose information from different data laws [18, 25, 32].

Endpoint convergence asks whether an individual sampling trajectory has a limiting output as the noise level tends to zero. When the trajectory is stopped at a positive heat time, a convergence rate quantifies the remaining truncation error. This pathwise question complements the study of the generated distribution. It is mathematically delicate because the scores and their spatial derivatives may diverge as heat time vanishes. Consequently, well-posedness at every positive time does not by itself provide a continuous extension to the terminal time.

We consider fixed convex and extrapolative combinations of two exact heat-flow scores. Their rescaled potentials converge to weighted sums of squared support distances, but convergence of the potentials alone does not imply convergence of trajectories. Under extrapolation, limiting directions may depend on probability masses and lie outside the Clarke subdifferential of the limiting potential. We address this difficulty by combining the geometry of the component supports with entropy estimates for the original dynamics. The resulting analysis gives equilibrium localization for compactly supported laws under the stated hypotheses, and point convergence for arbitrary finite atomic configurations, including endpoints at projection interfaces.

1.1. **Heat-flow framework and geometric formulation.** In the heat-flow framework [31], let $u_{0,1}, u_{0,2}$ be compactly supported probability measures and let $u_k = G_t * u_{0,k}$ solve

$$\partial_t u_k = \Delta u_k, \qquad u_k(\cdot, 0) = u_{0,k}, \qquad k = 1, 2, \tag{1.1}$$

where G_t is the Gaussian heat kernel with covariance $2tI_d$. The regularized densities are smooth and positive for $t > 0$, even when the measures are singular. Their logarithmic gradients define the mixed score

$$s^{(\lambda)}(x,t) = \lambda \nabla \log u_1(x,t) + (1-\lambda)\nabla \log u_2(x,t), \qquad \lambda \geq 0. \tag{1.2}$$

Equivalently,

$$h_\lambda = u_1^\lambda u_2^{1-\lambda}, \qquad s^{(\lambda)} = \nabla \log h_\lambda.$$

The product h_λ is a weighted geometric mean for $0 \leq \lambda \leq 1$ and a multiplicative extrapolation for $\lambda > 1$, with a negative power of the reference density. Thus mixing preserves the logarithmic-gradient structure. It does not, however, generally preserve the heat equation: a direct calculation gives

$$(\partial_t - \Delta) h_\lambda = \lambda(1-\lambda)|\nabla \log u_1 - \nabla \log u_2|^2 h_\lambda.$$

We study the deterministic equation

$$\frac{dX_t}{dt} = -s^{(\lambda)}(X_t, t), \qquad X_T = x_T, \tag{1.3}$$

integrated from a fixed $T > 0$ toward $t = 0$. We also consider the stochastic counterpart

$$dX_t^\varepsilon = -(1+\varepsilon)s^{(\lambda)}(X_t^\varepsilon, t)\, dt + \sqrt{2\varepsilon}\, dW_t, \qquad X_T^\varepsilon = x_T, \quad \varepsilon \geq 0, \tag{1.4}$$

with the same direction of time. Here W is Brownian motion in decreasing heat time, as specified in Section 3.2; $\varepsilon = 0$ recovers (1.3). For $\lambda > 1$, this is classifier-free guidance (CFG) [25] when the first law is conditional and the second is the corresponding unconditional law; our analysis also permits unrelated data laws, for which we use the term *extrapolative score mixing*.

In the single-score setting, relative-entropy contraction follows from the common Markov evolution of the reference and generated laws [31, Lemma 5.3]. This argument does not extend directly to score mixtures, since h_λ need not satisfy the heat equation. We instead control the time variation of the mixed potential through posterior relative entropy of the component heat flows, and stochastic energy increments through a conditional free-energy estimate for the transition densities.

Write $A_k = \operatorname{supp} u_{0,k}$ and introduce

$$\begin{aligned} F_\lambda(x,t) &= -4t\left(\lambda \log u_1(x,t) + (1-\lambda)\log u_2(x,t)\right), \\ \Phi_\lambda(x) &= \lambda \operatorname{dist}(x,A_1)^2 + (1-\lambda)\operatorname{dist}(x,A_2)^2. \end{aligned} \tag{1.5}$$

A Laplace–Varadhan-type argument [17, Section 4.3] gives $F_\lambda(\cdot,t) \to \Phi_\lambda$ locally uniformly as $t \to 0^+$; see Lemma 5.3. Derivatives can nevertheless remain singular. Indeed, the covariance formula in Lemma 5.2 exhibits a t^{-2} term in $D_x^2 \log u_k$; after rescaling, $D_x^2 F_\lambda$ may still grow like t^{-1} near projection ties. These ties occur where a support has more than one nearest point.

To analyze the singular endpoint regime, introduce the similarity time

$$\tau = \log(T/t), \qquad t_\tau = Te^{-\tau}, \qquad Y_\tau = X_{t_\tau}, \qquad Y_\tau^\varepsilon = X_{t_\tau}^\varepsilon,$$

with $Y_0 = Y_0^\varepsilon = y_0 = x_T$. The time change preserves geometric paths and transforms $t \to 0^+$ into $\tau \to \infty$. We state the trajectory results in similarity time and then give their original-time interpretations.

1.2. **Main results.** Let $\Pi_k(x) := \operatorname{argmin}_{a \in A_k} |x-a|$ be the nearest-point set of A_k. Define the set of nearest-support equilibria by

$$\mathcal{E}_\lambda = \{x \in \mathbb{R}^d : x \in \lambda \operatorname{conv} \Pi_1(x) + (1-\lambda)\operatorname{conv} \Pi_2(x)\}.$$

For $0 \le \lambda \le 1$, this set coincides with the Clarke critical set of the limiting potential:

$$\mathcal{E}_\lambda = \{x \in \mathbb{R}^d : 0 \in \partial^C \Phi_\lambda(x)\}.$$

Here ∂^C denotes the Clarke subdifferential, defined in Eq. (2.4). A support is called *active* when its score coefficient is nonzero; terms with zero coefficients are omitted throughout.

The main results are as follows.

(i) *Limiting support geometry* (Theorem 2.1). For compact active supports and every $\lambda \ge 0$, $F_\lambda(\cdot,t) \to \Phi_\lambda$ locally uniformly as $t \to 0^+$. For $0 \le \lambda \le 1$, we obtain the complete graphical convergence

$$\operatorname{gph} \nabla F_\lambda(\cdot,t) \longrightarrow \operatorname{gph} \partial^C \Phi_\lambda \qquad (t \to 0^+).$$

For $\lambda > 1$, limiting directions lie in the separated envelope $\lambda \partial^C d_1^2 + (1-\lambda)\partial^C d_2^2$, where $d_k = \operatorname{dist}(\cdot, A_k)$, but may lie outside $\partial^C \Phi_\lambda$ and depend on the probability masses.

(ii) *Deterministic localization for compact data* (Theorem 2.3). For every initial state y_0, the logarithmic-time trajectory Y has an energy limit $\ell \in \mathbb{R}$ and satisfies

$$\Phi_\lambda(Y_\tau) \longrightarrow \ell, \qquad \varnothing \ne \omega(Y) \subseteq \{x \in \mathcal{E}_\lambda : \Phi_\lambda(x) = \ell\}.$$

Here $\omega(Y)$ is the set of accumulation points as $\tau \to \infty$. Thus localization places all accumulation points on one equilibrium energy level, without asserting a single limit. Compactness suffices for convex mixing; for extrapolation, we also assume a lower small-ball mass bound on the negatively weighted reference law.

(iii) *Deterministic endpoints and asymptotics for empirical data* (Theorem 2.5). For finite atomic laws and every $\lambda \ge 0$,

$$\#\mathcal{E}_\lambda < \infty, \qquad Y_\tau \longrightarrow y^* \in \mathcal{E}_\lambda \quad (\tau \to \infty)$$

for every initialization and arbitrary site configurations, including projection-interface endpoints. If each active support has a unique nearest atom at y^*, we call it a *strict-cell endpoint*. In this case, for some $\xi^* \in \mathbb{R}^d$ and $\eta > 0$,

$$Y_\tau = y^* + e^{-\tau/2}\xi^* + O\Big(t_\tau e^{-\eta/t_\tau}\Big) \qquad (\tau \to \infty).$$

Consequently, $|Y_\tau - y^*| \le Ce^{-\tau/2}$ for all $\tau \ge 0$. This is the square-root heat-time rate, sharp when $\xi^* \ne 0$.

(iv) *Stochastic localization and endpoints* (Theorem 3.3). Under the hypotheses of part (ii), for each fixed initial state y_0 and noise parameter $\varepsilon > 0$, almost surely,

$$\Phi_\lambda(Y_\tau^\varepsilon) \longrightarrow \ell^\varepsilon, \qquad \operatorname{dist}(Y_\tau^\varepsilon, \mathcal{E}_\lambda) \longrightarrow 0.$$

All accumulation points lie on the same equilibrium energy level. For finite atomic laws, the whole trajectory converges almost surely to a random equilibrium $y^{\varepsilon,*} \in \mathcal{E}_\lambda$, including the possibility of a projection-interface endpoint.

The refined expansion and rate concern deterministic strict-cell endpoints. Figure 3 also suggests square-root heat-time decay of the stochastic root-mean-square distance to the set of local minimizers in the empirical example.

Square-integrable perturbations in logarithmic time preserve deterministic energy convergence and equilibrium localization, and empirical endpoint convergence for both dynamics (Proposition 3.1 and Remark 3.6). In heat time, the condition becomes weighted square-integrability of the score error along the perturbed sampling trajectory (Remark 3.2).

For classifier-free guidance with a positive-weight conditional component, the supports are nested: $A_1 \subseteq A_2$. When $\lambda > 1$, Corollary 3.7 shows that the local and global minimizers of Φ_λ are exactly A_1, while every equilibrium outside A_1 has a nonunique target projection. Thus empirical terminal samples outside A_1 can occur only at target projection ties.

1.3. **Analytical contributions.** The proofs combine the geometry of nearest-point projections with entropy estimates for the time-dependent dynamics.

Signed geometry and dissipation. The separated projection envelope retains the nearest-point geometry of each support, even when limiting directions depend on probability masses and lie outside $\partial^C \Phi_\lambda$. Applying the chain rule to each squared distance yields dissipation along the signed limiting inclusion, which is the basis for equilibrium localization.

Posterior entropy and deterministic action. Each logarithmic heat gradient is expressed through a posterior mean associated with Gaussian smoothing. The relative entropy of this posterior controls the time derivative of the regularized potential, giving energy convergence and finite deterministic kinetic action in logarithmic time. Compactness of translated paths then yields equilibrium localization.

Conditional free energy and stochastic convergence. Applying Itô's formula to the regularized energy $F_\lambda(Y_\tau^\varepsilon, t_\tau)$ introduces the curvature term $\varepsilon t_\tau \Delta_x F_\lambda(Y_\tau^\varepsilon, t_\tau)$. Under extrapolation, the available upper bound for this term need not be integrable in logarithmic time. We instead combine the expected potential energy of a restarted trajectory with a time-dependent multiple of the entropy of its transition density. After a short smoothing interval for the point-mass initial law, this gives the conditional estimate in Lemma 5.5(ii). The bound is uniform in the restart time; its error decays with heat time and grows at most quadratically in the restart state. Moment bounds make these errors summable. An almost-supermartingale argument then gives almost-sure energy convergence, and dissipation along limiting paths yields equilibrium localization.

Empirical rigidity and endpoints. For each nearest-point pattern, the equidistance and signed barycentric constraints place equilibria in two affine spaces with orthogonal direction spaces. Their intersection therefore contains at most one equilibrium, even for degenerate configurations. There are only finitely many patterns, so the equilibrium set is finite without a genericity assumption. Connectedness of the accumulation set then gives point convergence for both dynamics, including projection-interface endpoints. At a strict-cell endpoint, the posterior means are exponentially close to the selected atoms. The resulting affine equation gives the deterministic asymptotic expansion.

1.4. **Relation to existing work.** *Heat-kernel asymptotics and Hamilton–Jacobi equations.* The small-time heat-kernel formula [42] becomes a support-distance Laplace principle in the convolution setting [17, Section 4.3]. We use its locally uniform form for compactly supported, possibly singular measures. The logarithmic heat transformation and squared-distance geometry also connect with Hamilton–Jacobi and target-set eikonal equations [6, 14, 20]. The covariance identities in Lemma 5.2 give the Euclidean matrix estimate whose trace is the Li–Yau inequality [30]. Convex mixing preserves the upper Hessian bound for the rescaled potential; extrapolation need not preserve it.

Nonsmooth geometry and gradient dynamics. Nearest-point sets and squared distances are treated by convex and nonsmooth analysis [12, 36, 37], including the semiconcavity calculus of [9]. The distinction between convergence of energies and of their gradient evolutions is central to variational flow theory [1, 38, 39]. Here the additional issue is that signed mixing can select directions outside the limiting Clarke subdifferential. We therefore retain the two component subdifferentials when passing to the limiting inclusion.

Long-time dynamics and differential inclusions. Asymptotically autonomous evolution [22, 33] and differential inclusions [3] provide the long-time framework. The trajectory estimates here combine posterior information in the deterministic case with Fokker–Planck free energy [27] and almost-supermartingale convergence [35] in the stochastic case.

Score matching and heat-flow analysis. Score matching and diffusion models provide the generative setting [24, 26, 40, 41, 43]. The Gaussian posterior-mean representation is classical [21, 43]; stochastic sampling also draws on reverse-diffusion theory [2, 23]. The single-score analysis in [31] proves density stability and concentration; its Theorem 3.7(i) establishes deterministic compact-support endpoints for Lebesgue-almost every initialization. Our signed-mixing analysis uses the component posteriors and the actual transition densities.

Empirical denoisers and trajectory endpoints. For a single empirical score, convergence at the forward-noise standard-deviation scale under a strict-Voronoi-cell accumulation hypothesis is proved in [5, Theorem 4.3 and Example 4.4], which also exhibits an exceptional interface endpoint. For flow-matching ODEs, [44, Theorems 5.3–5.4] establishes almost-everywhere terminal convergence under geometric and local mass assumptions, with refined rates for smooth submanifold-supported and finite discrete laws. Our empirical theorem treats signed mixing from every deterministic initialization and arbitrary finite site configurations; the strict-cell assumption concerns the rate rather than endpoint existence.

Guidance and score composition. Guidance and score composition [18, 19, 25, 32] have been analyzed through boundary concentration and score-error sensitivity [11], predictor–corrector structure [7], and stagewise attraction and absorption [8]. For genuine conditional guidance, [4] proves almost-sure terminal convergence on the conditional support under compactness and bounded guidance strength. Support robustness, including exponential-integrator discretizations, is established in [10] for separated compact convex mixture supports with a positive target-density lower bound. Our focus is the terminal dynamics of fixed signed mixtures of possibly unrelated laws. We identify a dissipative nearest-support inclusion even when limiting directions lie outside $\partial^C \Phi_\lambda$, prove endpoint convergence for arbitrary finite configurations, and obtain deterministic strict-cell asymptotics. The nested-support corollary restricts off-target equilibria to target projection ties under support inclusion alone. These results complement conditional-support guarantees and distributional approximation estimates such as [13, 16].

1.5. **Organization.** Section 2 presents the geometric limits and deterministic asymptotics. Section 3 treats perturbations and stochastic asymptotics, and Section 4 gives numerical illustrations. The main results are proved in Section 5, using five lemmas whose proofs appear in Section 6. Section 7 discusses further questions.

2. Geometric limits and deterministic asymptotics

We write $\mathcal{P}(\mathbb{R}^d)$ for Borel probability measures, $\langle\cdot,\cdot\rangle$ for the Euclidean scalar product, and $|\cdot|$ for the associated norm. The parameter $\lambda \geq 0$ is fixed. Terms with zero coefficients are omitted throughout, including in set-valued expressions: $\lambda = 0$ uses only the second law, and $\lambda = 1$ only the first. An *active support* means a support with nonzero coefficient. Compactness hypotheses are specified in the theorem statements and concern only these supports.

The heat regularizations introduced in Eq. (1.1) are

$$u_k(x,t) = \int_{\mathbb{R}^d} G_t(x-a)\, du_{0,k}(a), \qquad G_t(z) = (4\pi t)^{-d/2} \exp\left(-\frac{|z|^2}{4t}\right). \tag{2.1}$$

We recall the regularized and limiting potentials from (1.5):

$$\begin{aligned} F_\lambda(x,t) &= -4t\left(\lambda \log u_1(x,t) + (1-\lambda)\log u_2(x,t)\right), \\ \Phi_\lambda(x) &= \lambda\, \mathrm{dist}(x,A_1)^2 + (1-\lambda)\,\mathrm{dist}(x,A_2)^2, \qquad A_k = \operatorname{supp} u_{0,k}. \end{aligned} \tag{2.2}$$

The relation $\nabla F_\lambda = -4t\, s^{(\lambda)}$ expresses the mixed score through the regularized potential.

Write $d_k(x) = \mathrm{dist}(x, A_k)$ and $\Pi_k(x) = \operatorname{argmin}_{a\in A_k} |x-a|$. Define the nearest-support equilibrium set and the confinement set by

$$\begin{aligned} \mathcal{E}_\lambda &= \left\{x \in \mathbb{R}^d : x \in \lambda \operatorname{conv}\Pi_1(x) + (1-\lambda)\operatorname{conv}\Pi_2(x)\right\}, \\ K_\lambda &= \lambda \operatorname{conv}(A_1) + (1-\lambda)\operatorname{conv}(A_2). \end{aligned} \tag{2.3}$$

The set operations are Minkowski addition and scalar multiplication, including negative coefficients. At projection ties, the convex hulls allow combinations of nearest points. When the active supports are compact, K_λ is compact and convex, and $\mathcal{E}_\lambda \subseteq K_\lambda$ is compact. Its points are *nearest-support equilibria*: the state equals a weighted combination of nearest-point barycentres of the active supports. The equilibrium set can contain nonminimizing points, while the dynamics determines which equilibria occur as trajectory endpoints.

For a locally Lipschitz function h, the Clarke subdifferential is

$$\partial^C h(x) = \operatorname{conv}\left\{\lim_{n\to\infty} \nabla h(x_n) : x_n \to x,\ h \text{ differentiable at } x_n\right\}. \tag{2.4}$$

The squared-distance formula recalled in Lemma 5.1(i) yields

$$\begin{aligned} \widehat{\partial}\Phi_\lambda(x) &:= \lambda \partial^C d_1^2(x) + (1-\lambda)\partial^C d_2^2(x) \\ &= 2\big(x - \lambda \operatorname{conv}\Pi_1(x) - (1-\lambda)\operatorname{conv}\Pi_2(x)\big). \end{aligned} \tag{2.5}$$

We call this the *separated projection envelope*. Thus $x \in \mathcal{E}_\lambda$ if and only if $0 \in \widehat{\partial}\Phi_\lambda(x)$. The envelope is determined entirely by the supports and λ. For a set-valued map $\mathcal{F}$, write $\operatorname{gph}\mathcal{F} = \{(x,p) : p \in \mathcal{F}(x)\}$, with graph distance measured in $\mathbb{R}^d \times \mathbb{R}^d$.

The logarithmic time change

$$\tau = \log(T/t), \qquad t_\tau = Te^{-\tau}, \qquad Y_\tau = X_{t_\tau} \tag{2.6}$$

rewrites Eq. (1.3) as

$$\dot{Y}_\tau = -\frac{1}{4}\nabla F_\lambda(Y_\tau, t_\tau), \qquad Y_0 = y_0, \quad \tau \geq 0. \tag{2.7}$$

Here $y_0 = x_T$, and the inverse relation is $X_t = Y_{\log(T/t)}$. The processes X and Y have the same geometric path and accumulation points. Equal increments of τ correspond to equal multiplicative reductions in heat time. The time change absorbs the explicit inverse-time factor, although spatial derivatives may still grow near projection interfaces.

2.1. **Geometric reduction and autonomous limiting dynamics.** The first theorem identifies the limiting score geometry and the equations satisfied by limits of translated trajectories. Graphical convergence allows the spatial point to vary with heat time and thus captures distinct limiting directions at projection ties. A solution of a differential inclusion below means a locally absolutely continuous curve satisfying the inclusion almost everywhere.

Theorem 2.1 (Geometric reduction)**.** *Let $\lambda \geq 0$ and assume that the active supports are compact.*

(i) *The potentials $F_\lambda(\cdot, t)$ converge locally uniformly to Φ_λ as $t \to 0^+$. If $x_n \to x$, $t_n \to 0^+$, and $\nabla F_\lambda(x_n, t_n) \to p$, then $p \in \widehat{\partial}\Phi_\lambda(x)$. For every nonempty compact $B \subset \mathbb{R}^d$,*

$$\sup_{x \in B} \operatorname{dist}\Big((x, \nabla F_\lambda(x,t)), \operatorname{gph} \widehat{\partial}\Phi_\lambda\Big) \longrightarrow 0. \tag{2.8}$$

If $0 \leq \lambda \leq 1$, then $\widehat{\partial}\Phi_\lambda = \partial^C \Phi_\lambda$ and the graph limit is complete: for every $x \in \mathbb{R}^d$, $p \in \partial^C \Phi_\lambda(x)$, and sequence $t_n \to 0^+$, there exists a sequence $x'_n \to x$ such that

$$\nabla F_\lambda(x'_n, t_n) \longrightarrow p. \tag{2.9}$$

(ii) *For every $T > 0$ and $y_0 \in \mathbb{R}^d$, the solution Y of Eq. (2.7) is unique, global, and precompact. For any sequence $\tau_n \to \infty$, the translates $Y_{\tau_n + \cdot}$ are relatively compact in $C_{\mathrm{loc}}\Big([0,\infty); \mathbb{R}^d\Big)$. Every limit Z takes values in K_λ and satisfies*

$$\dot{Z}_\tau \in -\frac{1}{4}\widehat{\partial}\Phi_\lambda(Z_\tau) \qquad \text{for a.e. } \tau \geq 0. \tag{2.10}$$

Every solution of this inclusion obeys, for $\tau \leq \tau'$ in its domain,

$$\Phi_\lambda(Z_{\tau'}) + 4\int_\tau^{\tau'} |\dot{Z}_s|^2 \, ds = \Phi_\lambda(Z_\tau). \tag{2.11}$$

Proof. See Section 5.2. □

Remark 2.2 (The original-time deterministic process)**.** Under Theorem 2.1, $X_t := Y_{\log(T/t)}$ is the unique solution of (1.3) on $(0, T]$, with $x_T = y_0$. Its range and accumulation set as $t \to 0^+$ coincide with those of Y as $\tau \to \infty$; in particular, X is bounded.

For convex mixing the autonomous inclusion is the Clarke subgradient system of Φ_λ. For extrapolation it is generally an outer geometric description. Attainable field directions can form a strict subset of the envelope, and the inclusion describes a potentially larger class of trajectories. If all active projections are unique on an open set, the field converges to $\nabla\Phi_\lambda$ uniformly on compact subsets of that set.

2.2. **Omega-limit localization for compact supports.** We next prove energy convergence and show that every translated limit is stationary. For a precompact continuous curve, define

$$\omega(Y) = \bigcap_{S>0} \overline{Y([S,\infty))}. \tag{2.12}$$

This is the set of all accumulation points as $\tau \to \infty$. We say that a law u_0 with support A satisfies a *uniform lower small-ball mass condition* if there exist $c > 0$, $\alpha \geq 0$, and $r_0 > 0$ such that

$$\inf_{a \in A} u_0\left(B(a,r)\right) \geq c r^\alpha, \qquad 0 < r \leq r_0. \tag{2.13}$$

This condition is expressed in terms of small-ball probabilities and also applies to singular measures. Finite empirical laws with positive weights satisfy it with $\alpha = 0$. A density bounded below on a compact support with a uniform interior-volume bound satisfies it with $\alpha = d$.

Theorem 2.3 (Omega-limit localization)**.** *Let $\lambda \geq 0$ and assume that the active supports are compact. If $\lambda > 1$, assume in addition that $u_{0,2}$ satisfies Eq.* (2.13)*. For every $T > 0$ and $y_0 \in \mathbb{R}^d$, let Y be the solution of Eq.* (2.7)*.*

(i) *The trajectory has finite kinetic action,*

$$\int_0^\infty |\dot{Y}_\tau|^2 \, d\tau < \infty, \tag{2.14}$$

and there exists $\ell \in \mathbb{R}$ such that

$$F_\lambda(Y_\tau, t_\tau) \longrightarrow \ell, \qquad \Phi_\lambda(Y_\tau) \longrightarrow \ell. \tag{2.15}$$

(ii) *The set $\omega(Y)$ is nonempty, compact, and connected, with*

$$\omega(Y) \subseteq \{x \in \mathcal{E}_\lambda : \Phi_\lambda(x) = \ell\}. \tag{2.16}$$

Proof. See Section 5.3. □

Remark 2.4 (Localization in the original heat time)**.** For $X_t = Y_{\log(T/t)}$, Theorem 2.3 gives

$$\int_0^T t \, |\dot{X}_t|^2 \, dt = \int_0^\infty |\dot{Y}_\tau|^2 \, d\tau < \infty.$$

As $t \to 0^+$, both $F_\lambda(X_t, t)$ and $\Phi_\lambda(X_t)$ converge to ℓ, and all accumulation points of X belong to $\mathcal{E}_\lambda \cap \{\Phi_\lambda = \ell\}$.

For extrapolation, the additional mass hypothesis is imposed only on the negatively weighted reference law; convex mixing, including either single-score case, requires only compactness.

Finite action makes all translated limits stationary, as quantified by Eq. (5.22) in the proof. It controls motion on fixed time windows, while the total trajectory length may be infinite. For empirical data, finiteness of $\mathcal{E}_\lambda$ and connectedness of $\omega(Y)$ yield convergence to a single point, as stated below.

2.3. **Empirical endpoint convergence and asymptotics.** For finite atomic measures, called empirical measures below, the compact-support result strengthens to convergence of each trajectory. Assume

$$u_{0,k} = \sum_{i=1}^{n_k} w_{k,i} \delta_{a_{k,i}}, \qquad w_{k,i} > 0, \qquad \sum_{i=1}^{n_k} w_{k,i} = 1, \tag{2.17}$$

with distinct sites within each support. Set

$$I_k(x) = \operatorname*{argmin}_i |x - a_{k,i}|, \qquad \mathcal{S}_k = \{x : \#I_k(x) > 1\}, \quad k = 1, 2,$$

$$\mathcal{S}_\lambda = \begin{cases} \mathcal{S}_2, & \lambda = 0, \\ \mathcal{S}_1, & \lambda = 1, \\ \mathcal{S}_1 \cup \mathcal{S}_2, & \lambda > 0, \ \lambda \neq 1. \end{cases} \tag{2.18}$$

Outside the active Voronoi skeleton $\mathcal{S}_\lambda$, every active support has a unique nearest atom. We call an endpoint outside this set a *strict-cell endpoint*. The equilibrium set $\mathcal{E}_\lambda$ itself also includes possible ties.

Theorem 2.5 (Empirical convergence and strict-cell asymptotics)**.** *Assume Eq.* (2.17)*, and fix $\lambda \geq 0$ and $T > 0$. For every initial state $y_0 \in \mathbb{R}^d$, the solution Y of Eq.* (2.7) *is unique on $[0, \infty)$ and satisfies:*

(i) *The set $\mathcal{E}_\lambda$ is finite. There exists $y^* \in \mathcal{E}_\lambda$ such that*

$$Y_\tau \longrightarrow y^* \qquad (\tau \to \infty). \tag{2.19}$$

(ii) *If $y^* \notin \mathcal{S}_\lambda$, there exist $\xi^* \in \mathbb{R}^d$ and $\eta > 0$ such that*

$$Y_\tau = y^* + e^{-\tau/2}\xi^* + O\Big(t_\tau e^{-\eta/t_\tau}\Big) \qquad (\tau \to \infty). \tag{2.20}$$

Consequently, there exists $C > 0$ such that

$$|Y_\tau - y^*| \le Ce^{-\tau/2}, \qquad \tau \ge 0. \tag{2.21}$$

Proof. See Section 5.4. □

Remark 2.6 (Original-time form and sharpness)**.** The original-time solution $X_t = Y_{\log(T/t)}$, with $x_T = y_0$, satisfies

$$X_t \longrightarrow x^* := y^* \qquad (t \to 0^+). \tag{2.22}$$

If $y^* \notin \mathcal{S}_\lambda$, (2.21) gives $|X_t - x^*| \le C\sqrt{t/T}$ for every $0 < t \le T$. Let $a_{k,*}$ be the unique nearest atom at this endpoint in each active support. Then $x^* = y^* = \lambda a_{1,*} + (1-\lambda)a_{2,*}$, and the expansion in Theorem 2.5 becomes

$$X_t = x^* + \sqrt{t/T}\,\xi^* + O\Big(te^{-\eta/t}\Big) \qquad (t \to 0^+). \tag{2.23}$$

Thus the heat-time exponent $1/2$ is sharp when $\xi^* \neq 0$; if $\xi^* = 0$, then $|X_t - x^*| = o(t^m)$ for every $m > 0$. The constants may depend on the data, weights, initialization, T, and λ, and may deteriorate as the Voronoi gap approaches zero.

For finite empirical laws, positivity of atomic weights implies Eq. (2.13) directly. The strict-cell hypothesis enters the rate estimate, whereas endpoint convergence holds for arbitrary empirical configurations.

Remark 2.7 (Limits of the conclusions)**.** In one dimension, let $u_{0,1} = (\delta_{-1} + 3\delta_1)/4$, $u_{0,2} = (3\delta_{-1} + \delta_1)/4$, and $\lambda = 2$. Both supports are $\{-1, 1\}$, so $\Phi_2(x) = (1 - |x|)^2$, but

$$\partial^C \Phi_2(0) = [-2, 2], \qquad \widehat{\partial}\Phi_2(0) = [-6, 6], \qquad \nabla F_2(0, t) = -3.$$

Thus the limiting direction at this interface lies outside the Clarke subdifferential of the limiting potential. If instead $u_{0,1} = u_{0,2} = (\delta_{-1} + \delta_1)/2$, then $\nabla F_\lambda(x, t) = 2\{x - \tanh(x/(2t))\}$ for every λ. Its joint limits as $x \to 0$ and $t \to 0^+$ form $[-2, 2]$, strictly smaller than $\widehat{\partial}\Phi_2(0)$. Moreover, the trajectory initialized at zero stays there, although zero is a strict local maximum of $\Phi_\lambda(x) = (1-|x|)^2$. Consequently, convergence to an equilibrium cannot be strengthened to convergence to a local minimum for every initialization.

3. Perturbations and stochastic asymptotics

We extend the convergence results to square-integrable deterministic perturbations and to stochastic sampling. As before, the results are stated in logarithmic time and then translated into heat time.

3.1. **Perturbations of the sampling scores.**

Consider the perturbed dynamics

$$\dot{\widetilde{Y}}_\tau = -\frac{1}{4}\nabla F_\lambda(\widetilde{Y}_\tau, t_\tau) + r_\tau, \qquad \widetilde{Y}_0 = \widetilde{y}_0. \tag{3.1}$$

Here r is a prescribed measurable error in logarithmic time.

Proposition 3.1 (Robustness of equilibrium localization)**.** *Assume the hypotheses on λ and the data laws in Theorem 2.3. Fix $T > 0$, $\widetilde{y}_0 \in \mathbb{R}^d$, and $r \in L^2(0, \infty; \mathbb{R}^d)$. Equation* (3.1) *has a unique global bounded Carathéodory solution $\widetilde{Y}$ with the following properties.*

(i) *The trajectory has finite kinetic action,*

$$\int_0^\infty |\dot{\widetilde{Y}}_\tau|^2 \, d\tau < \infty, \tag{3.2}$$

and there exists $\widetilde{\ell} \in \mathbb{R}$ *such that*

$$F_\lambda(\widetilde{Y}_\tau, t_\tau) \longrightarrow \widetilde{\ell}, \qquad \Phi_\lambda(\widetilde{Y}_\tau) \longrightarrow \widetilde{\ell}. \tag{3.3}$$

(ii) *The set* $\omega(\widetilde{Y})$ *is nonempty, compact, and connected, with*

$$\omega(\widetilde{Y}) \subseteq \{x \in \mathcal{E}_\lambda : \Phi_\lambda(x) = \widetilde{\ell}\}. \tag{3.4}$$

Proof. See Section 5.5. □

Remark 3.2 (Errors in the original score)**.** Under the hypotheses on λ and the data in Proposition 3.1, let $\widetilde{X}$ be a locally absolutely continuous solution of $\dot{\widetilde{X}}_t = -s^{(\lambda)}(\widetilde{X}_t, t) - e(\widetilde{X}_t, t)$ on $(0, T]$. If

$$\int_0^T t \, |e(\widetilde{X}_t, t)|^2 \, dt < \infty, \tag{3.5}$$

then $\widetilde{X}$ is bounded and its accumulation points as $t \to 0^+$ belong to $\mathcal{E}_\lambda$. For finite empirical data, finiteness of $\mathcal{E}_\lambda$ and connectedness of $\omega(\widetilde{Y})$ give $\widetilde{Y}_\tau \to \widetilde{y}^* \in \mathcal{E}_\lambda$, hence $\widetilde{X}_t \to \widetilde{x}^* := \widetilde{y}^*$, where $\widetilde{Y}_\tau = \widetilde{X}_{t_\tau}$ and $\widetilde{Y}_0 = \widetilde{X}_T$. Indeed, the change of time (2.6) gives $r_\tau = t_\tau e(\widetilde{X}_{t_\tau}, t_\tau)$ and $\int_0^\infty |r_\tau|^2 \, d\tau = \int_0^T t \, |e(\widetilde{X}_t, t)|^2 \, dt$, so Proposition 3.1 applies. Thus (3.5) is a sufficient condition for localization and empirical endpoint convergence under score approximation along the given trajectory. The selected equilibrium may change, and even an unchanged endpoint may be approached more slowly. For example, take $u_{0,1} = u_{0,2} = \delta_c$, $r_\tau = e^{-\tau/4} v$ with $v \neq 0$, and $\widetilde{Y}_0 = c$. Then

$$\widetilde{Y}_\tau - c = 4(e^{-\tau/4} - e^{-\tau/2})v,$$

so the endpoint remains c, but the leading error has scale $e^{-\tau/4}$, or $(t/T)^{1/4}$ in heat time. Deriving (3.5) from a training loss or a time-discretization error requires additional analysis.

3.2. **Stochastic localization and endpoint convergence.** We use the same decreasing heat time as in the deterministic process. For $\varepsilon \geq 0$, consider the stochastic equation

$$dX_t^\varepsilon = -(1+\varepsilon) s^{(\lambda)}(X_t^\varepsilon, t) \, dt + \sqrt{2\varepsilon} \, dW_t, \qquad X_T^\varepsilon = x_T, \quad 0 < t \leq T. \tag{3.6}$$

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $W_T = 0$ and let W be Brownian motion for a backward filtration $(\mathcal{F}_t)$ with the usual conditions in decreasing time. For $0 < t < t' \leq T$, the increment $W_t - W_{t'}$ is independent of $\mathcal{F}_{t'}$ and is centered Gaussian with covariance $(t' - t) I_d$. Thus (3.6) means

$$X_t^\varepsilon = x_T + (1+\varepsilon) \int_t^T s^{(\lambda)}(X_{t'}^\varepsilon, t') \, dt' + \sqrt{2\varepsilon} \, W_t.$$

A strong solution is an $(\mathcal{F}_t)$-adapted process satisfying this identity almost surely for all $t \in (0, T]$, with sample paths in $C((0, T]; \mathbb{R}^d)$. Uniqueness is pathwise for a prescribed Brownian motion and terminal state. For a single score, $\varepsilon = 1$ is the classical reverse diffusion when initialized with the terminal heat law [2, 23].

The change of time (2.6) gives

$$dY_\tau^\varepsilon = -\frac{1+\varepsilon}{4} \nabla F_\lambda(Y_\tau^\varepsilon, t_\tau) \, d\tau + \sqrt{2\varepsilon t_\tau} \, dW_\tau, \qquad Y_0^\varepsilon = y_0 = x_T. \tag{3.7}$$

We reuse W for the standard Brownian motion in logarithmic time, adapted to the increasing filtration $\mathcal{G}_\tau := \mathcal{F}_{t_\tau}$ [28, 34]. Strong solutions of (3.7) are understood in integral form and have continuous paths on $[0, \infty)$.

For a stochastic process, $\omega(Y^\varepsilon)$ denotes the ordinary omega-limit set (2.12) of each continuous sample path. An $\mathcal{E}_\lambda$-valued random vector belongs to $\mathcal{E}_\lambda$ almost surely.

Theorem 3.3 (Stochastic localization and empirical convergence)**.** *Let $\lambda \geq 0$ and assume that the active supports are compact. If $\lambda > 1$, assume that $u_{0,2}$ satisfies* (2.13)*. Fix $T > 0$, $y_0 \in \mathbb{R}^d$, and $\varepsilon \geq 0$. Equation* (3.7) *has a unique global strong solution Y^ε with continuous, almost surely bounded paths.*

(i) *There is a finite random variable ℓ^ε such that, almost surely,*

$$F_\lambda(Y_\tau^\varepsilon, t_\tau) \longrightarrow \ell^\varepsilon, \qquad \Phi_\lambda(Y_\tau^\varepsilon) \longrightarrow \ell^\varepsilon \qquad (\tau \to \infty), \tag{3.8}$$

$$\varnothing \neq \omega(Y^\varepsilon) \subseteq \{x \in \mathcal{E}_\lambda : \Phi_\lambda(x) = \ell^\varepsilon\}. \tag{3.9}$$

The omega-limit set is compact and connected. In particular, $\operatorname{dist}(Y_\tau^\varepsilon, \mathcal{E}_\lambda) \to 0$ *as* $\tau \to \infty$.

(ii) *If the active laws are finite empirical measures, there is an $\mathcal{E}_\lambda$-valued random vector $y^{\varepsilon,*}$ such that*

$$Y_\tau^\varepsilon \longrightarrow y^{\varepsilon,*} \qquad (\tau \to \infty), \qquad \textit{almost surely.} \tag{3.10}$$

Proof. See Section 5.6. □

Remark 3.4 (The original-time stochastic process)**.** The process $X_t^\varepsilon := Y^\varepsilon_{\log(T/t)}$ solves (3.6), with $x_T = y_0$, and has the same accumulation points as Y^ε. Thus the energy and localization conclusions hold as $t \to 0^+$ as well. For empirical data, setting $x^{\varepsilon,*} := y^{\varepsilon,*}$ gives

$$X_t^\varepsilon \longrightarrow x^{\varepsilon,*} \in \mathcal{E}_\lambda \quad (t \to 0^+), \qquad \text{almost surely.} \tag{3.11}$$

For each fixed initialization, assigning this value at $t = 0$ gives an almost surely continuous extension to $[0, T]$. For general compact laws, the accumulation set is localized to a single equilibrium energy level. In conditional guidance with nested supports, Corollary 3.7 further restricts the possible terminal locations.

Corollary 3.5 (Terminal sampling law)**.** *Assume that the active laws are finite atomic measures, and fix $\lambda \geq 0$, $T > 0$, and $\varepsilon \geq 0$. Let X_T^ε have any law $\mu_T \in \mathcal{P}(\mathbb{R}^d)$, independently of the driving Brownian motion when $\varepsilon > 0$. There is a probability measure μ_0^ε such that, narrowly as $t \to 0^+$,*

$$\operatorname{Law}(X_t^\varepsilon) \rightharpoonup \mu_0^\varepsilon, \qquad \operatorname{supp} \mu_0^\varepsilon \subseteq \mathcal{E}_\lambda.$$

In particular, the terminal law is finitely supported.

Proof. See Section 5.6. □

Remark 3.6 (Square-integrable perturbations for empirical data)**.** For finite empirical active laws, the endpoint conclusion remains valid with an additional prescribed drift $r_\tau\, d\tau$ in (3.7), progressively measurable for $(\mathcal{G}_\tau)$, provided

$$\int_0^\infty |r_\tau|^2\, d\tau < \infty \qquad \text{almost surely.} \tag{3.12}$$

For a well-posed sampler using $s^{(\lambda)} + e$, this is (3.5) almost surely along its path, since $r_\tau = (1+\varepsilon) t_\tau e(\widetilde{X}^\varepsilon_{t_\tau}, t_\tau)$ and

$$\int_0^\infty |r_\tau|^2\, d\tau = (1+\varepsilon)^2 \int_0^T t\, |e(\widetilde{X}_t^\varepsilon, t)|^2\, dt. \tag{3.13}$$

This sufficient condition concerns the error along the perturbed trajectory; it preserves endpoint convergence, but need not preserve the selected equilibrium or its law.

Proof. See Section 5.6. □

In classifier-free guidance [25], a positive-weight conditional component of the reference law gives $A_1 \subseteq A_2$, leading to the following geometric restriction.

Corollary 3.7 (Conditional-support geometry)**.** *Assume that $A_1 \subseteq A_2$ are compact and $\lambda > 1$. The local and global minimizers of Φ_λ are exactly A_1, and every equilibrium outside A_1 has a nonunique projection onto A_1. For finite atomic laws, the endpoints belong to $A_1 \cup \mathcal{S}_1$ for every deterministic initialization and almost surely for each fixed stochastic initialization.*

Proof. Since $d_2 \le d_1$, we have $\Phi_\lambda \ge d_1^2$, with value zero on A_1. For squared distances, the one-sided directional derivative is $(d_k^2)'(x;v) = 2\min_{b\in\Pi_k(x)}\langle x-b, v\rangle$. Thus, for $x \notin A_1$ and $a \in \Pi_1(x)$, Cauchy–Schwarz gives

$$\Phi'_\lambda(x; a-x) \le -2\lambda d_1(x)^2 + 2(\lambda-1)d_1(x)d_2(x) \le -2d_1(x)^2 < 0.$$

Thus the local and global minimizers are exactly A_1. If $x \in \mathcal{E}_\lambda$ has the unique target projection a, write $x = \lambda a + (1-\lambda)b$ with $b \in \operatorname{conv}\Pi_2(x)$. Then $\lambda(x-a) = (\lambda-1)(x-b)$ and $|x-b| \le d_2(x)$, so

$$\frac{\lambda}{\lambda-1} d_1(x) \le d_2(x) \le d_1(x).$$

It follows that $x \in A_1$. The endpoint assertion follows from Theorems 2.5 and 3.3. □

For empirical conditional guidance, a terminal sample outside the target data set A_1 can occur only where several target atoms tie as nearest points. A deterministic trajectory ending at a target atom has a unique nearest point in each support, so Remark 2.6 gives an $O(\sqrt{t})$ stopping error. These conclusions give a geometric restriction under support inclusion alone and complement the conditional-support guarantees in [4, 10].

4. Numerical illustrations of the convergence theory

Two examples illustrate the convergence results: continuous compact supports, and finite atomic supports with deterministic and stochastic endpoints. All computations use exact heat-flow scores and integrate in logarithmic time from $T = 1$ to positive stopping times. For continuous supports, we evaluate the Gaussian integrals and use the Radau solver with relative and absolute tolerances 10^{-10} and 10^{-12}. For the empirical data, posterior weights are evaluated in a numerically stabilized form; deterministic trajectories use Radau, and stochastic trajectories use an exponential step for the linear drift and noise, with implicit treatment of the reference posterior when $\lambda > 1$. The stochastic curves use 2048 steps per decade of heat time and were compared with 512 and 1024 steps per decade using coupled noise increments.

4.1. Continuous supports: geometry and localization.

Consider the segmented supports

$$I = [-1.5, -1.2] \cup [-0.8, 0.8] \cup [1.2, 1.5], \qquad A_1 = I \times \{-1, 1\}, \qquad A_2 = \{-1, 1\} \times I,$$

equipped with normalized arclength measures. These measures satisfy the lower small-ball mass condition (2.13), so Theorems 2.1 and 2.3 apply in both the convex and extrapolative regimes.

Figure 1 superposes trajectories of the time-dependent equation (2.7) on the level sets of its limiting potential Φ_λ. The local-minimum sets are

$$\mathcal{M}_{1/2} = \{-1.1, -0.9, 0.9, 1.1\}^2, \qquad \mathcal{M}_2 = \{-2, 0, 2\} \times \{-1, 1\}.$$

Convex mixing balances proximity to the two supports, whereas extrapolation produces four minima outside their joint convex hull. Initial states are chosen so that the trajectories approach every marked minimum, illustrating the dependence of the endpoint on the initialization. This behavior is consistent with equilibrium localization in Theorem 2.3; the theorem also permits nonminimizing equilibria, as illustrated by Remark 2.7.

Table 1 compares F_λ and its gradient with their limiting geometric counterparts on the fixed grid

$$\mathcal{G} = \{-1.8, -1.6, \dots, 1.8\}^2.$$

The decreasing defects illustrate the potential convergence and field localization of Theorem 2.1. The field is compared with the separated envelope $\widehat{\partial}\Phi_\lambda$, which coincides with $\partial^C\Phi_\lambda$ for convex mixing

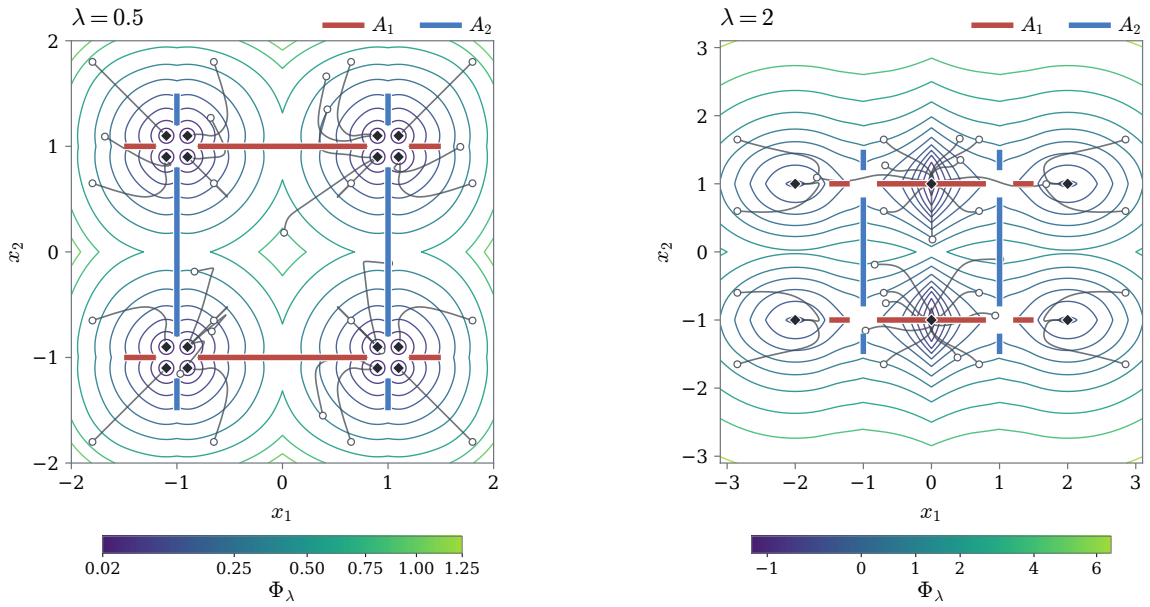


FIGURE 1. Continuous-support geometry and deterministic trajectories. Red and blue segments represent A_1 and A_2; colored contours are level sets of Φ_λ. Gray curves show 28 trajectories from the open-circle initializations to heat time 10^{-6}. Black diamonds mark local minima. Left: convex mixing, $\lambda = 0.5$. Right: extrapolation, $\lambda = 2$.

but may be larger under extrapolation. These fixed-grid diagnostics complement the moving-point graphical formulation of the theorem.

TABLE 1. Maximum potential and field defects on the fixed 19×19 grid $\mathcal{G}$.

	$\lvert F_\lambda - \Phi_\lambda \rvert$		$\operatorname{dist}(\nabla F_\lambda, \widehat{\partial}\Phi_\lambda)$	
t	$\lambda = 0.5$	$\lambda = 2$	$\lambda = 0.5$	$\lambda = 2$
10^{-1}	1.2478	1.7735	0.4104	1.2947
10^{-2}	0.0909	0.1640	0.1559	0.4931
10^{-3}	0.0085	0.0199	0.0505	0.1596
10^{-4}	0.0010	0.0024	0.0160	0.0505

4.2. **Empirical supports: endpoints and convergence scales.** Consider uniform empirical laws on

$$A_1 = \{(-3.0, 2.2), (-1.2, -2.6), (1.0, 1.3)\},$$
$$A_2 = \{(2.8, -3.2), (4.0, 2.6), (0.2, -0.2)\}. \tag{4.1}$$

Figure 2 compares the two dynamics for $\lambda = 0.5, 1, 2.5$, with $\varepsilon = 1$ in the stochastic case. Both use 48 common initial states in $[-9, 9]^2$.

The marked local-minimum set $\mathcal{M}_\lambda \subset \mathcal{E}_\lambda$ contains five weighted centers for $\lambda = 0.5$, equals A_1 for $\lambda = 1$, and contains three points for $\lambda = 2.5$. Two of the latter lie on reference Voronoi edges. The computed paths approach every marked minimum, illustrating the interface endpoints covered by the convergence theorems. Stochastic forcing changes the paths and may change the selected endpoint, while the equilibrium set remains the same.

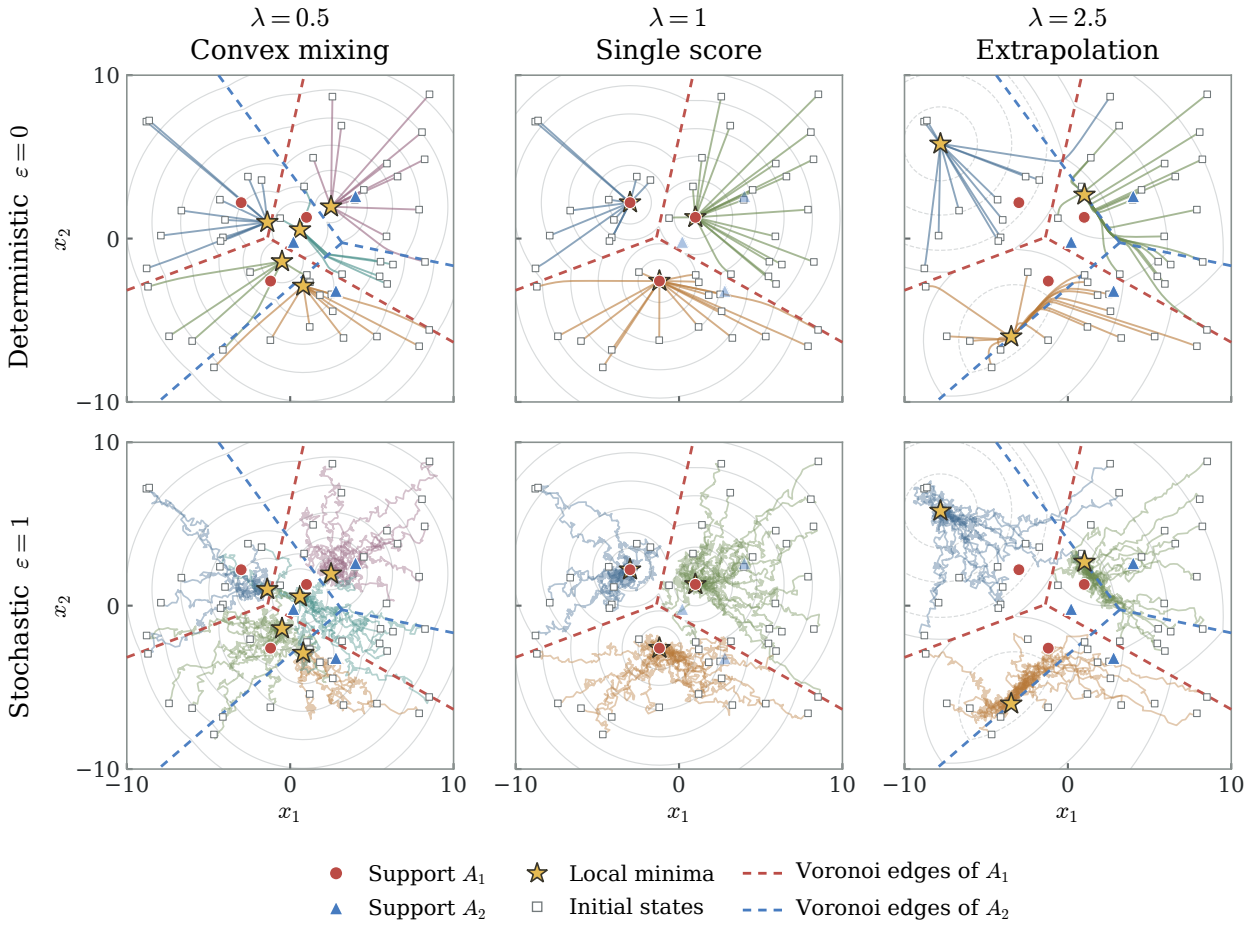


FIGURE 2. Empirical trajectories for (4.1). Columns correspond to $\lambda = 0.5, 1, 2.5$; rows to deterministic and stochastic trajectories. Stars mark local minima; red and blue dashed lines mark the Voronoi edges of A_1 and A_2. The extrapolative case includes two minima on reference interfaces. Path colors identify the approached minimum; faint contours show Φ_λ.

For deterministic strict-cell endpoints, Theorem 2.5 and Remark 2.6 give

$$|Y_\tau - y^*| = O(e^{-\tau/2}), \qquad |X_t - x^*| = O(\sqrt{t}).$$

Figure 3 compares this scale with the root-mean-square (RMS) distance to the marked minima:

$$R_\lambda^\varepsilon(t) = \left(\frac{1}{N_\varepsilon} \sum_{r=1}^{N_\varepsilon} \operatorname{dist}\big(X_t^{\varepsilon,(r)}, \mathcal{M}_\lambda\big)^2 \right)^{1/2}, \qquad N_0 = 48, \quad N_1 = 384. \tag{4.2}$$

The stochastic ensemble uses eight independent noise realizations per initial state.

Over $10^{-7} \le t \le 10^{-4}$, the computed RMS curves exhibit an approximately square-root decay, consistent with the theoretical rate for deterministic strict-cell endpoints. For stochastic trajectories and deterministic projection-interface endpoints, the observed rates remain numerical: the convergence theorems establish endpoint existence in these cases, while the corresponding rate questions remain open.

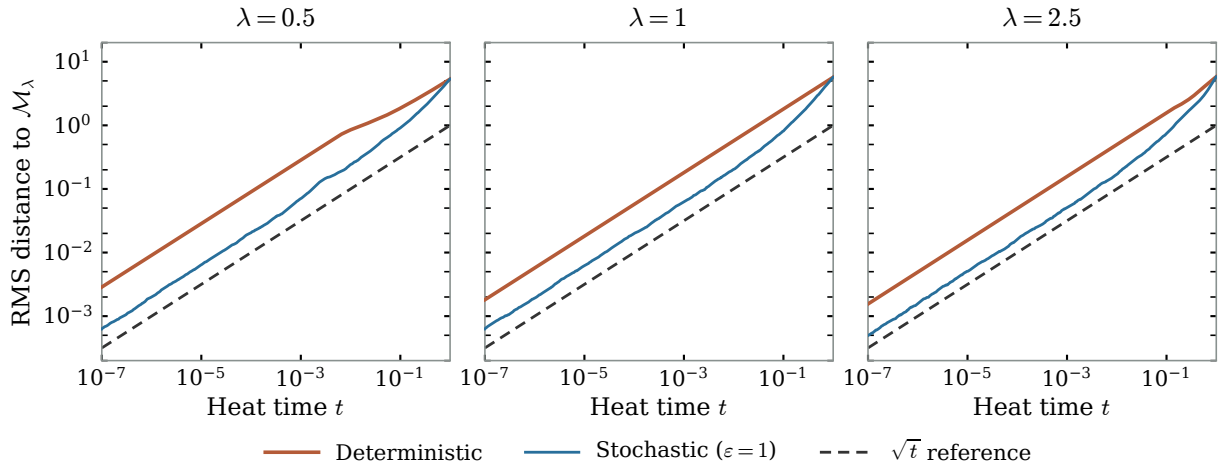


FIGURE 3. Convergence scales for empirical trajectories. Solid curves show the RMS distances $R^0_\lambda(t)$ and $R^1_\lambda(t)$ from (4.2). The dashed curve is the reference $\sqrt{t}$, with unit prefactor. The data are unsmoothed. The $O(\sqrt{t})$ rate is proved for deterministic strict-cell endpoints; stochastic and projection-interface decay scales are numerical observations.

5. Proofs of the main results

We retain the notation of Sections 2 and 3. Throughout, $t_\tau = Te^{-\tau}$ and $\lambda \geq 0$ is fixed; t, t' denote heat time and τ, τ', s, s' logarithmic time.

Posterior quantities and normalized potentials. For a probability measure u_0, write $u = G_t * u_0$ and set

$$d\nu^x_t(a) = \frac{G_t(x-a)}{u(x,t)}\, du_0(a), \qquad m(x,t) = \int a\, d\nu^x_t(a), \qquad f(x,t) = -4t \log u(x,t). \tag{5.1}$$

We also write

$$Q(x,t) = \int |x-a|^2\, d\nu^x_t(a), \qquad \Sigma(x,t) = \int (a - m(x,t)) \otimes (a - m(x,t))\, d\nu^x_t(a).$$

For $u_{0,k}$, the corresponding quantities carry a subscript k. The posterior-mean identity [31, Lemma 5.1 and its proof] gives

$$\begin{aligned} \nabla \log u_k(x,t) &= \frac{m_k(x,t) - x}{2t}, \\ \nabla F_\lambda(x,t) &= 2\left(x - \lambda m_1(x,t) - (1-\lambda) m_2(x,t)\right). \end{aligned} \tag{5.2}$$

Each $m_k(x,t)$ belongs to $\mathrm{conv}(A_k)$, and the deterministic flow has the form

$$\dot{Y}_\tau = \frac{1}{2}(z_\tau - Y_\tau), \qquad z_\tau = \lambda m_1(Y_\tau, t_\tau) + (1-\lambda) m_2(Y_\tau, t_\tau) \in K_\lambda. \tag{5.3}$$

For the energy estimates, remove the spatially constant normalization:

$$\overline{f}(x,t) = f(x,t) - 2dt \log(4\pi t), \qquad \overline{F}_\lambda(x,t) = F_\lambda(x,t) - 2dt \log(4\pi t). \tag{5.4}$$

We write

$$V_\tau(x) := \overline{F}_\lambda(x, t_\tau).$$

The normalization preserves spatial gradients and vanishes as $t \to 0^+$.

Entropy and information bounds. For probability measures, we use the relative entropy

$$\mathrm{KL}(\rho\|\mu) = \int \log\left(\frac{d\rho}{d\mu}\right) d\rho \qquad (\rho \ll \mu),$$

with value $+\infty$ otherwise. For a probability density, set

$$\mathrm{Ent}(\rho) = \int_{\mathbb{R}^d} \rho(x)\log\rho(x)\,dx.$$

For a nonempty compact set B, define

$$H_B(t) := \sup_{x\in B} \mathrm{KL}(\nu_t^x\|u_0), \qquad H_{k,B}(t) := \sup_{x\in B} \mathrm{KL}(\nu_{k,t}^x\|u_{0,k}), \tag{5.5}$$

$$\mathcal{I}_\lambda(x,t) := \lambda\,\mathrm{KL}(\nu_{1,t}^x\|u_{0,1}) + (1-\lambda)\,\mathrm{KL}(\nu_{2,t}^x\|u_{0,2}). \tag{5.6}$$

As before, terms with zero coefficients are omitted.

Restarted dynamics. For fixed $\varepsilon \geq 0$, $Y^{\varepsilon;s,y}$ denotes the solution of (3.7) started at time s from y, and $\mathbb{E}_{s,y}$ its expectation; $\varepsilon = 0$ gives the deterministic flow. We retain $\mathcal{G}_\tau = \mathcal{F}_{t_\tau}$ from Section 3.2 and suppress fixed superscripts within individual proofs.

Stochastic variation of constants. We use the stochastic variation-of-constants (Duhamel) formula [15, Sections 4.4 and 5.1]. For $a > 0$, a progressively measurable process g with locally integrable paths, and a continuous local martingale M, consider

$$dZ_\tau = (-aZ_\tau + g_\tau)\,d\tau + dM_\tau.$$

Then, for $0 \leq s \leq \tau$, almost surely,

$$Z_\tau = e^{-a(\tau-s)}Z_s + \int_s^\tau e^{-a(\tau-s')}g_{s'}\,ds' + \int_s^\tau e^{-a(\tau-s')}\,dM_{s'}. \tag{5.7}$$

The case $M = 0$ is the deterministic formula.

5.1. **Technical ingredients.** The five lemmas below are used in the proofs of the main results. Their proofs are given in Section 6.

Lemma 5.1 (Distance geometry and empirical equilibria)**.** *Assume that the active supports are compact.*

(i) *For every nonempty compact A, the function d_A^2 is locally Lipschitz and 2-semiconcave, with*

$$\partial^C d_A^2(x) = 2\,(x - \mathrm{conv}\,\Pi_A(x))\,. \tag{5.8}$$

The separated envelope is upper semicontinuous with nonempty compact convex values, and

$$\begin{cases} \widehat{\partial}\Phi_\lambda = \partial^C\Phi_\lambda, & 0 \leq \lambda \leq 1, \\ \partial^C\Phi_\lambda \subseteq \widehat{\partial}\Phi_\lambda, & \lambda > 1. \end{cases}$$

For $0 \leq \lambda \leq 1$, the potential Φ_λ is 2-semiconcave.

(ii) *For finite empirical active laws, each collection of nearest-index sets determines at most one equilibrium. In particular,*

$$\#\mathcal{E}_\lambda < \infty.$$

Lemma 5.2 (Heat-flow Hessians and the Li–Yau bound)**.** *Assume that the active supports are compact. For $t > 0$,*

$$\begin{aligned} D_x^2 \log u_k(x,t) &= -\frac{I_d}{2t} + \frac{\Sigma_k(x,t)}{4t^2}, \\ D_x^2 F_\lambda(x,t) &= 2I_d - \frac{1}{t}\left(\lambda\Sigma_1(x,t) + (1-\lambda)\Sigma_2(x,t)\right). \end{aligned} \tag{5.9}$$

In particular,

$$D_x^2 \log u_k \succeq -\frac{I_d}{2t}, \qquad D_x^2 F_\lambda \preceq 2I_d \quad \text{if } 0 \leq \lambda \leq 1.$$

On every strip $\delta \le t \le T$ with $\delta > 0$, the mixed score and ∇F_λ are globally Lipschitz in space and have at most linear growth, uniformly in t.

Lemma 5.3 (Laplace–Varadhan localization)**.**

(i) *Let u_0 have nonempty compact support A. Then*

$$f(\cdot, t) \longrightarrow d_A^2 \qquad \text{locally uniformly as } t \to 0^+.$$

For any sequences $x_n \to x$ and $t_n \to 0^+$,

$$m(x_n, t_n) \longrightarrow a \quad \Longrightarrow \quad a \in \operatorname{conv} \Pi_A(x).$$

(ii) *For finite empirical active laws and $x_0 \notin \mathcal{S}_\lambda$, let $a_{k,*}$ be the active nearest atoms and set*

$$c_* = \lambda a_{1,*} + (1-\lambda) a_{2,*}.$$

There exist a neighborhood U of x_0 and $C, \eta, t_0 > 0$ such that

$$\sup_{x \in U} |\nabla F_\lambda(x,t) - 2(x - c_*)| \le C e^{-\eta/t}, \qquad 0 < t \le t_0. \tag{5.10}$$

Lemma 5.4 (Passage to the geometric inclusion)**.** *Assume compact active supports. Let $\kappa > 0$ and let continuous curves $y, R : [0,\infty) \to \mathbb{R}^d$, with y bounded and $R_0 = 0$, satisfy*

$$y_\tau = y_0 - \frac{\kappa}{4} \int_0^\tau \nabla F_\lambda(y_s, t_s)\, ds + R_\tau. \tag{5.11}$$

If, for every $S > 0$,

$$\sup_{0 \le h \le S} |R_{\tau+h} - R_\tau| \longrightarrow 0 \qquad (\tau \to \infty), \tag{5.12}$$

then, for any $\tau_n \to \infty$, the translates $y_{\tau_n + \cdot}$ are relatively compact in $C_{\mathrm{loc}}([0,\infty); \mathbb{R}^d)$. Every limit z is locally Lipschitz and satisfies

$$\dot z_s \in -\frac{\kappa}{4} \widehat{\partial} \Phi_\lambda(z_s) \qquad \text{for a.e. } s \ge 0.$$

The final lemma controls the time dependence of the potential and the conditional energy increments of stochastic trajectories.

Lemma 5.5 (Deterministic and stochastic energy estimates)**.** *Assume the hypotheses on the data and λ in Theorem 2.3.*

(i) *For every nonempty compact $B \subset \mathbb{R}^d$, there exists a nonnegative function $b_B \in L^1(0,\infty)$ such that*

$$\partial_\tau V_\tau(x) = -4 t_\tau \mathcal{I}_\lambda(x, t_\tau) \le b_B(\tau), \qquad x \in B, \quad \tau \ge 0. \tag{5.13}$$

For $0 \le \lambda \le 1$, one may take $b_B = 0$.

(ii) *For every fixed $\varepsilon > 0$, there exist $s_0, C > 0$ such that, for all $s \ge s_0$ and $y \in \mathbb{R}^d$,*

$$\mathbb{E}_{s,y}[V_{s+1}(Y_{s+1}^{\varepsilon; s, y})] \le V_s(y) + C t_s (1 + |\log t_s|)(1 + |y|^2). \tag{5.14}$$

The constants may depend on ε, λ, T and the data laws, but are uniform in s and y.

5.2. **Proof of Theorem 2.1.**

Step 1 (Localization and graphical recovery). The identities

$$F_\lambda = \lambda f_1 + (1-\lambda) f_2, \qquad \Phi_\lambda = \lambda d_1^2 + (1-\lambda) d_2^2,$$

and Lemma 5.3 give local uniform convergence of the potentials. To identify the field limits, let $x_n \to x$, $t_n \to 0^+$, and suppose that $\nabla F_\lambda(x_n, t_n) \to p$. Each active posterior mean lies in a fixed compact support hull. Along a common subsequence,

$$m_k(x_n, t_n) \longrightarrow a_k \in \operatorname{conv} \Pi_k(x).$$

The mean-shift identity therefore yields

$$p = 2\,(x - \lambda a_1 - (1-\lambda)a_2) \in \widehat{\partial}\Phi_\lambda(x).$$

For the uniform graph-distance assertion on a compact set B, a contrary sequence would satisfy

$$x_n \in B, \qquad t_n \to 0^+, \qquad \operatorname{dist}\Big((x_n, \nabla F_\lambda(x_n,t_n)), \operatorname{gph}\widehat{\partial}\Phi_\lambda\Big) \geq \delta$$

for some $\delta > 0$. Compactness and the bounded posterior means give a subsequence converging to (x,p) in the envelope graph, contradicting this lower bound. This proves (2.8).

For graphical recovery, assume $0 \leq \lambda \leq 1$, fix $x \in \mathbb{R}^d$ and $p \in \partial^C \Phi_\lambda(x)$, and let $t_n \to 0^+$ be any sequence. Choose $B = \overline{B}(x,r)$ with $r > 0$ and define on this ball

$$J_n(y) = F_\lambda(y,t_n) - \langle p, y-x\rangle - 2|y-x|^2,$$
$$J(y) = \Phi_\lambda(y) - \langle p, y-x\rangle - 2|y-x|^2.$$

The curvature and semiconcavity bounds in Lemmas 5.1 and 5.2 give

$$D^2 J_n \preceq -2I_d, \qquad J(y) \leq J(x) - |y-x|^2.$$

Let x_n' be the unique maximizer of J_n on B and set

$$\delta_n := \sup_{y\in B} |F_\lambda(y,t_n) - \Phi_\lambda(y)|.$$

Then

$$J(x) - \delta_n \leq J_n(x) \leq J_n(x_n') \leq J(x) - |x_n' - x|^2 + \delta_n.$$

Hence $|x_n' - x|^2 \leq 2\delta_n \to 0$. For large n, the maximizer lies in the interior of B, and its first-order condition gives

$$\nabla F_\lambda(x_n', t_n) - p - 4(x_n' - x) = 0, \qquad \nabla F_\lambda(x_n', t_n) \longrightarrow p.$$

This proves (2.9). On a region with unique active projections, the envelope is a continuous singleton. The same subsequence argument on each compact subset gives uniform convergence of the fields there.

Step 2 (Confinement and limiting trajectories). On each bounded logarithmic-time interval, Lemma 5.2 gives a unique solution. Set

$$z_s = \lambda m_1(Y_s,t_s) + (1-\lambda) m_2(Y_s,t_s) \in K_\lambda.$$

Applying (5.7) with $a = 1/2$ and $M = 0$ gives

$$Y_\tau = e^{-\tau/2} y_0 + \frac{1}{2}\int_0^\tau e^{-(\tau-s)/2} z_s\, ds \in e^{-\tau/2} y_0 + (1 - e^{-\tau/2}) K_\lambda. \tag{5.15}$$

The normalized kernel average k_τ belongs to K_λ. Hence Y stays in $\operatorname{conv}(\{y_0\} \cup K_\lambda)$, extends globally, and has bounded derivative.

Let p_0 be the projection of y_0 onto K_λ. Since $e^{-\tau/2} p_0 + (1 - e^{-\tau/2}) k_\tau \in K_\lambda$, we obtain

$$\operatorname{dist}(Y_\tau, K_\lambda) \leq e^{-\tau/2} |y_0 - p_0| = e^{-\tau/2} \operatorname{dist}(y_0, K_\lambda). \tag{5.16}$$

Apply Lemma 5.4 with $\kappa = 1$ and $R = 0$. Every translated limit solves (2.10); the last estimate places its whole range in K_λ.

Step 3 (Signed dissipation). Let Z solve (2.10). On each compact time interval, local boundedness of the envelope makes Z and every active composition $d_k^2 \circ Z$ Lipschitz. Fix a time s at which these curves are differentiable and the inclusion holds; such times form a set of full measure.

For any $p_k \in \partial^C d_k^2(Z_s)$, (6.1) gives

$$d_k^2(Z_{s+h}) - d_k^2(Z_s) \leq \langle p_k, Z_{s+h} - Z_s\rangle + |Z_{s+h} - Z_s|^2.$$

At the chosen time,

$$Z_{s+h} - Z_s = h\dot{Z}_s + o(|h|), \qquad \frac{|Z_{s+h} - Z_s|^2}{|h|} \longrightarrow 0.$$

Division by $h > 0$ gives $(d_k^2 \circ Z)'(s) \le \langle p_k, \dot{Z}_s \rangle$; division by $h < 0$ gives the reverse inequality. Consequently,

$$(d_k^2 \circ Z)'(s) = \langle p_k, \dot{Z}_s \rangle \qquad \text{for every } p_k \in \partial^C d_k^2(Z_s).$$

At each such time, the inclusion supplies component vectors satisfying

$$\lambda p_1 + (1-\lambda)p_2 = -4\dot{Z}_s.$$

Multiplying the component identities by their signed coefficients gives

$$\frac{d}{ds}\Phi_\lambda(Z_s) = \langle \lambda p_1 + (1-\lambda)p_2, \dot{Z}_s \rangle = -4|\dot{Z}_s|^2. \tag{5.17}$$

The composition $\Phi_\lambda \circ Z$ is absolutely continuous, so integration proves (2.11). The componentwise equality is what preserves dissipation when $1 - \lambda$ is negative.

5.3. **Proof of Theorem 2.3.**

Step 1 (The characteristic energy and reference information). The trajectory stays in $B = \mathrm{conv}(\{y_0\} \cup K_\lambda)$ by (5.15). The chain rule and Lemma 5.5(i) give

$$\frac{d}{d\tau}V_\tau(Y_\tau) = -4|\dot{Y}_\tau|^2 + R_\lambda(\tau), \qquad R_\lambda(\tau) := -4t_\tau \mathcal{I}_\lambda(Y_\tau, t_\tau). \tag{5.18}$$

For convex mixing, nonnegativity of relative entropy gives $R_\lambda \le 0$. For $\lambda > 1$, the target contribution remains nonpositive, and the reference contribution satisfies

$$R_\lambda(\tau) \le 4(\lambda - 1)t_\tau \,\mathrm{KL}(\nu_{2,t_\tau}^{Y_\tau} \| u_{0,2}) \le 4(\lambda - 1)t_\tau H_{2,B}(t_\tau).$$

Accordingly, set

$$b(\tau) = \begin{cases} 0, & 0 \le \lambda \le 1, \\ 4(\lambda - 1)t_\tau H_{2,B}(t_\tau), & \lambda > 1. \end{cases}$$

Thus $R_\lambda \le b$ and $b \ge 0$. For $\lambda > 1$, the bound (6.9) and $d\tau = -dt/t$ give

$$\int_0^\infty b(\tau)\, d\tau = 4(\lambda - 1)\int_0^T H_{2,B}(t)\, dt < \infty.$$

Indeed, $H_{2,B}(t) \le C_B(1 + |\log t|)$ near zero and is bounded on compact positive-time intervals.

Step 2 (Corrected energy and finite action). Define

$$\mathcal{L}(\tau) = \overline{F}_\lambda(Y_\tau, t_\tau) + \int_\tau^\infty b(s)\, ds. \tag{5.19}$$

This function is locally absolutely continuous and satisfies

$$\mathcal{L}'(\tau) = -4|\dot{Y}_\tau|^2 + R_\lambda(\tau) - b(\tau) \le -4|\dot{Y}_\tau|^2 \tag{5.20}$$

for almost every τ. Local uniform convergence to Φ_λ and continuity on positive-time strips bound the potential below on B. Hence $\mathcal{L}$ is bounded below and decreases to a finite limit ℓ.

Integrating the differential inequality over $[\tau, S]$ yields

$$4\int_\tau^S |\dot{Y}_s|^2\, ds \le \mathcal{L}(\tau) - \mathcal{L}(S).$$

Letting $S \to \infty$ proves

$$4\int_\tau^\infty |\dot{Y}_s|^2\, ds \le \mathcal{L}(\tau) - \ell. \tag{5.21}$$

In particular, the action is finite and its tail tends to zero. Moreover,

$$F_\lambda(Y_\tau, t_\tau) = \mathcal{L}(\tau) - \int_\tau^\infty b(s)\, ds + 2dt_\tau \log(4\pi t_\tau) \longrightarrow \ell.$$

Uniform convergence of $F_\lambda(\cdot, t)$ on B then gives $\Phi_\lambda(Y_\tau) \to \ell$, proving (2.15).

Step 3 (Stationary limits and equilibrium localization). For each $S > 0$, absolute continuity and Cauchy–Schwarz give

$$\begin{aligned} \sup_{0\le h\le S} |Y_{\tau+h} - Y_\tau| &\le \int_\tau^{\tau+S} |\dot Y_s|\, ds \\ &\le \sqrt{S}\left(\int_\tau^\infty |\dot Y_s|^2\, ds\right)^{1/2} \longrightarrow 0. \end{aligned} \tag{5.22}$$

Take $z \in \omega(Y)$ and choose $\tau_n \to \infty$ with $Y_{\tau_n} \to z$. The estimate implies

$$Y_{\tau_n+\cdot} \longrightarrow (s \mapsto z) \quad \text{locally uniformly on } [0,\infty).$$

By Theorem 2.1, this constant curve solves the limiting inclusion. Therefore

$$0 \in \widehat{\partial}\Phi_\lambda(z), \qquad z \in \mathcal{E}_\lambda, \qquad \Phi_\lambda(z) = \ell,$$

where the last identity follows from continuity and the energy limit.

Finally, the closed tails $B_S = \overline{Y([S,\infty))}$ are nested, nonempty compact connected sets. Their intersection $\omega(Y)$ has these same properties, proving (2.16).

5.4. **Proof of Theorem 2.5.**

Step 1 (Finite equilibria give point convergence). Positive atomic weights imply the small-ball condition with exponent zero, so the compact-support theorem applies for every $\lambda \ge 0$. Its localization conclusion and Lemma 5.1(ii) give

$$\varnothing \ne \omega(Y) \subseteq \mathcal{E}_\lambda, \qquad \omega(Y) \text{ connected}, \qquad \#\mathcal{E}_\lambda < \infty.$$

Hence $\omega(Y) = \{y^*\}$, and precompactness implies $Y_\tau \to y^*$ from every initial state. The inverse time change gives (2.22) with $x^* := y^*$.

Step 2 (The affine tail and its asymptotic coefficient). If $y^* \notin \mathcal{S}_\lambda$, convergence places the tail in the neighborhood of Lemma 5.3(ii). The equilibrium relation gives $c_* = y^*$, so, for $e_\tau = Y_\tau - y^*$ and $\tau \ge \tau_0$,

$$\dot e_\tau = -\frac{1}{2} e_\tau + r_\tau, \qquad |r_\tau| \le Ce^{-\eta/t_\tau} = Ce^{-\eta e^\tau/T}.$$

Formula (5.7), applied from time τ_0, gives

$$e_\tau = e^{-\tau/2}\left(e^{\tau_0/2} e_{\tau_0} + \int_{\tau_0}^\tau e^{s/2} r_s\, ds\right). \tag{5.23}$$

With $t' = Te^{-s}$, the weighted remainder satisfies

$$\int_{\tau_0}^\infty e^{s/2}|r_s|\, ds \le C\sqrt{T}\int_0^{t_{\tau_0}} (t')^{-3/2} e^{-\eta/t'}\, dt' < \infty.$$

The limiting coefficient is

$$\xi^* = e^{\tau_0/2} e_{\tau_0} + \int_{\tau_0}^\infty e^{s/2} r_s\, ds, \qquad e^{\tau/2} e_\tau \longrightarrow \xi^*.$$

Subtracting (5.23) from this definition gives the exact remainder formula

$$e_\tau - e^{-\tau/2}\xi^* = -e^{-\tau/2}\int_\tau^\infty e^{s/2} r_s\, ds.$$

For $t = t_\tau$, use the same substitution and $(t')^{-3/2} \le \sqrt{t}\,(t')^{-2}$ on $(0,t]$:

$$\begin{aligned}\left|e_\tau - e^{-\tau/2}\xi^*\right| &\le C\sqrt{t}\int_0^t (t')^{-3/2} e^{-\eta/t'}\,dt' \\ &\le Ct\int_0^t (t')^{-2} e^{-\eta/t'}\,dt' = \frac{C}{\eta}\,te^{-\eta/t}.\end{aligned}$$

This proves (2.20). The expansion bounds $e^{\tau/2}|Y_\tau - y^*|$ for large τ; continuity on the remaining compact time interval gives (2.21) for all $\tau \ge 0$ after enlarging C. Since $e^{-\tau/2} = \sqrt{t_\tau/T}$, the same expansion gives (2.23). The local balance gives $y^* = \lambda a_{1,*} + (1-\lambda)a_{2,*}$. If $\xi^* \neq 0$, then

$$e^{\tau/2}|Y_\tau - y^*| \longrightarrow |\xi^*| > 0, \qquad \frac{|X_t - x^*|}{\sqrt{t}} \longrightarrow \frac{|\xi^*|}{\sqrt{T}}.$$

Thus the leading scale is attained, with constants depending on the trajectory's entrance into the strict cell.

5.5. **Proof of Proposition 3.1.**

Step 1 (Confinement under square-integrable forcing). Positive-time Lipschitz bounds and $r \in L^1_{\mathrm{loc}}$ give a unique maximal Carathéodory solution. With $z_s = \lambda m_1(\widetilde{Y}_s, t_s) + (1-\lambda)m_2(\widetilde{Y}_s, t_s) \in K_\lambda$, (5.7) gives

$$\begin{aligned}\widetilde{Y}_\tau &= e^{-\tau/2}\widetilde{y}_0 + \frac{1}{2}\int_0^\tau e^{-(\tau-s)/2} z_s\,ds + \mathcal{R}(\tau), \\ \mathcal{R}(\tau) &= \int_0^\tau e^{-(\tau-s)/2} r_s\,ds.\end{aligned} \tag{5.24}$$

The first two terms lie in the convex hull of $\{\widetilde{y}_0\} \cup K_\lambda$. The forcing contribution satisfies

$$|\mathcal{R}(\tau)| \le \left(\int_0^\tau e^{-(\tau-s)}\,ds\right)^{1/2} \|r\|_{L^2(0,\tau)} \le \|r\|_{L^2(0,\infty)}.$$

Thus the solution is global and confined to a compact ball B. For $\tau \ge S$, split the convolution at S:

$$|\mathcal{R}(\tau)| \le e^{-(\tau-S)/2}|\mathcal{R}(S)| + \|r\|_{L^2(S,\infty)}.$$

Letting $\tau \to \infty$ and then $S \to \infty$ gives $\mathcal{R}(\tau) \to 0$. The convex-hull estimate yields

$$\operatorname{dist}(\widetilde{Y}_\tau, K_\lambda) \le e^{-\tau/2}\operatorname{dist}(\widetilde{y}_0, K_\lambda) + |\mathcal{R}(\tau)| \longrightarrow 0. \tag{5.25}$$

Step 2 (Absorption of the energy error). The bounded drift and $r \in L^2$ give $\dot{\widetilde{Y}} \in L^2_{\mathrm{loc}}$, so the chain rule applies. The identity (5.13) applies along the perturbed curve. Let $b \ge 0$ be the integrable bound from Section 5.3 on B, so that

$$\widetilde{R}_\lambda(\tau) = -4t_\tau \mathcal{I}_\lambda(\widetilde{Y}_\tau, t_\tau) \le b(\tau).$$

The identity $\nabla F_\lambda = -4(\dot{\widetilde{Y}} - r)$ and the estimate

$$4\langle r_\tau, \dot{\widetilde{Y}}_\tau\rangle \le 2|r_\tau|^2 + 2|\dot{\widetilde{Y}}_\tau|^2$$

give

$$\begin{aligned}\frac{d}{d\tau}\overline{F}_\lambda(\widetilde{Y}_\tau, t_\tau) &= -4|\dot{\widetilde{Y}}_\tau|^2 + 4\langle r_\tau, \dot{\widetilde{Y}}_\tau\rangle + \widetilde{R}_\lambda(\tau) \\ &\le -2|\dot{\widetilde{Y}}_\tau|^2 + 2|r_\tau|^2 + b(\tau).\end{aligned} \tag{5.26}$$

Thus the corrected energy

$$\widetilde{\mathcal{L}}(\tau) = \overline{F}_\lambda(\widetilde{Y}_\tau, t_\tau) + \int_\tau^\infty (2|r_s|^2 + b(s))\,ds \tag{5.27}$$

is locally absolutely continuous, bounded below, and satisfies

$$\widetilde{\mathcal{L}}'(\tau) \le -2|\dot{\widetilde{Y}}_\tau|^2 \qquad \text{for a.e. } \tau.$$

It decreases to a finite limit $\widetilde{\ell}$. Integration first over a finite interval and then to infinity yields

$$2\int_\tau^\infty |\dot{\widetilde{Y}}_s|^2\,ds \le \widetilde{\mathcal{L}}(\tau) - \widetilde{\ell}. \tag{5.28}$$

The error tail and Gaussian normalization vanish as $\tau \to \infty$. Local uniform convergence of the potentials on B then gives both energy limits in (3.3).

Step 3 (Equilibrium localization). The tail-action estimate implies, for each $S > 0$,

$$\sup_{0\le h\le S} |\widetilde{Y}_{\tau+h} - \widetilde{Y}_\tau| \le \sqrt{S}\left(\int_\tau^\infty |\dot{\widetilde{Y}}_s|^2\,ds\right)^{1/2} \longrightarrow 0.$$

To identify the constant limiting curves, apply Lemma 5.4 with $\kappa = 1$ and $R_\tau = \int_0^\tau r_s\,ds$. Its forcing condition follows from

$$\sup_{0\le h\le S} |R_{\tau+h} - R_\tau| \le \sqrt{S}\,\|r\|_{L^2(\tau,\infty)} \longrightarrow 0.$$

If $\widetilde{Y}_{\tau_n} \to z$, the constant translated limit solves the inclusion, and energy convergence gives

$$0 \in \widehat{\partial}\Phi_\lambda(z), \qquad z \in \mathcal{E}_\lambda, \qquad \Phi_\lambda(z) = \widetilde{\ell}.$$

The closed-tail argument proves (3.4). For empirical data, Lemma 5.1(ii) makes this connected set a singleton, proving point convergence in Remark 3.2.

5.6. **Proofs of Theorem 3.3 and Corollary 3.5.** We first prove Theorem 3.3. For $\varepsilon = 0$, the assertions are Theorems 2.3 and 2.5. Assume $\varepsilon > 0$ and write

$$Y = Y^\varepsilon, \qquad \kappa = 1 + \varepsilon, \qquad a = \kappa/2.$$

We use Lemma 5.5(ii) to prove energy convergence, then apply the pathwise compactness and dissipation arguments.

Step 1 (Bounded paths and second moments). The accumulated noise is

$$M_\tau = \int_0^\tau \sqrt{2\varepsilon t_s}\,dW_s, \qquad \mathbb{E}|M_\tau|^2 = 2\varepsilon dT(1 - e^{-\tau}). \tag{5.29}$$

By the Itô isometry, M is a continuous L^2-bounded martingale and converges almost surely [34]. For each continuous noise path, $H = Y - M$ solves

$$\dot{H}_\tau = -\frac{\kappa}{4}\nabla F_\lambda(H_\tau + M_\tau, t_\tau), \qquad H_0 = y_0, \qquad Y = H + M.$$

On finite intervals, Lemma 5.2 gives a continuous, globally Lipschitz drift with linear growth. The solutions agree on overlaps and depend only on the past noise, yielding the unique global strong solution [28].

With $z_s = \lambda m_1(Y_s, t_s) + (1-\lambda)m_2(Y_s, t_s) \in K_\lambda$, (5.7) and integration by parts give

$$\begin{aligned} Y_\tau &= e^{-a\tau}y_0 + a\int_0^\tau e^{-a(\tau-s)}z_s\,ds + C_\tau, \\ C_\tau &= \int_0^\tau e^{-a(\tau-s)}\,dM_s = M_\tau - a\int_0^\tau e^{-a(\tau-s)}M_s\,ds. \end{aligned} \tag{5.30}$$

Whenever $M_\tau \to M^*$, its exponential average has the same limit. Indeed, for $\tau \ge S$,

$$\left|a\int_0^\tau e^{-a(\tau-s)}(M_s - M^*)\,ds\right| \le e^{-a(\tau-S)}\sup_{0\le s\le S}|M_s - M^*| + \sup_{s\ge S}|M_s - M^*|.$$

First let $\tau \to \infty$, then $S \to \infty$. It follows that $C_\tau \to 0$. Since the first two terms in (5.30) belong to $e^{-a\tau}y_0 + (1 - e^{-a\tau})K_\lambda$, we obtain

$$\operatorname{dist}(Y_\tau, K_\lambda) \le e^{-a\tau} \operatorname{dist}(y_0, K_\lambda) + |C_\tau| \longrightarrow 0. \tag{5.31}$$

The same formula bounds the path almost surely. Moreover,

$$\mathbb{E}|C_\tau|^2 = 2\varepsilon d \int_0^\tau e^{-2a(\tau-s)} t_s \, ds \le \frac{\varepsilon d T}{a},$$

so compactness of K_λ also gives

$$\sup_{\tau \ge 0} \mathbb{E}|Y_\tau|^2 < \infty. \tag{5.32}$$

Step 2 (Convergence of the energy at integer times). The variable Y_n is $\mathcal{G}_n$-measurable. Compact support and Jensen's inequality give

$$\tfrac{1}{2}|x|^2 - C_0 \le V_\tau(x), \qquad |V_\tau(x)| \le C_0(1 + |x|^2), \tag{5.33}$$

uniformly in τ; see Section 6.5. Thus $U_n := V_n(Y_n) + C_0$ is nonnegative and integrable by (5.32).

Conditionally on $\mathcal{G}_n$, the process restarts from Y_n at time n. The Markov property and Lemma 5.5(ii) give, for all $n \ge n_0$ with n_0 sufficiently large,

$$\mathbb{E}[U_{n+1} \mid \mathcal{G}_n] \le U_n + B_n, \qquad B_n := C t_n (1 + |\log t_n|)(1 + |Y_n|^2). \tag{5.34}$$

The errors B_n are nonnegative and $\mathcal{G}_n$-measurable. The uniform moment bound and $t_n = Te^{-n}$ imply

$$\sum_{n \ge n_0} \mathbb{E}B_n \le C \left(1 + \sup_{\tau \ge 0} \mathbb{E}|Y_\tau|^2\right) \sum_{n \ge n_0} t_n(1 + |\log t_n|) < \infty.$$

By Tonelli's theorem, the nonnegative sum $\sum_{n \ge n_0} B_n$ belongs to L^1. We can therefore add the remaining error to the energy through its conditional expectation:

$$A_n := \mathbb{E}\left[\sum_{k=n}^{\infty} B_k \,\middle|\, \mathcal{G}_n\right], \qquad L_n := U_n + A_n. \tag{5.35}$$

These variables are nonnegative and integrable. The tower property, $\mathcal{G}_n \subseteq \mathcal{G}_{n+1}$, and measurability of B_n give

$$\mathbb{E}[A_{n+1} \mid \mathcal{G}_n] = \mathbb{E}\left[\sum_{k=n+1}^{\infty} B_k \,\middle|\, \mathcal{G}_n\right] = A_n - B_n,$$
$$\mathbb{E}[L_{n+1} \mid \mathcal{G}_n] \le U_n + B_n + A_n - B_n = L_n.$$

Thus A_n and L_n are nonnegative supermartingales and have finite almost-sure limits [34, 35]. Moreover,

$$\mathbb{E}A_n = \sum_{k=n}^{\infty} \mathbb{E}B_k \longrightarrow 0.$$

Writing $A_\infty = \lim_{n\to\infty} A_n \ge 0$, Fatou's lemma gives

$$0 \le \mathbb{E}A_\infty \le \liminf_{n\to\infty} \mathbb{E}A_n = 0.$$

Hence $A_n \to 0$ almost surely. Consequently,

$$V_n(Y_n) = L_n - A_n - C_0 \longrightarrow \ell^\varepsilon \qquad \text{almost surely}$$

for a finite random variable ℓ^ε. Finally, local uniform convergence $V_n \to \Phi_\lambda$ and pathwise boundedness imply

$$|V_n(Y_n) - \Phi_\lambda(Y_n)| \longrightarrow 0 \qquad \text{almost surely.}$$

Therefore

$$\Phi_\lambda(Y_n) \longrightarrow \ell^\varepsilon \qquad \text{almost surely.} \tag{5.36}$$

Step 3 (All times and equilibrium localization). Fix a continuous path for which the bounds and limits above hold; this is the case almost surely. The remaining argument is deterministic. The integrated equation is

$$Y_\tau = y_0 - \frac{\kappa}{4}\int_0^\tau \nabla F_\lambda(Y_s, t_s)\, ds + M_\tau. \tag{5.37}$$

Convergence of M implies, for every $S > 0$,

$$\sup_{0\le h\le S} |M_{\tau+h} - M_\tau| \longrightarrow 0. \tag{5.38}$$

Lemma 5.4 therefore gives compactness of the translated paths and identifies each limit z as a solution of

$$\dot z_s \in -\frac{\kappa}{4}\widehat\partial \Phi_\lambda(z_s).$$

The signed chain rule in Section 5.2 yields

$$\Phi_\lambda(z_{\tau'}) + \frac{4}{\kappa}\int_\tau^{\tau'} |\dot z_s|^2\, ds = \Phi_\lambda(z_\tau), \qquad 0 \le \tau \le \tau'. \tag{5.39}$$

For a limit of integer translates $Y_{n_j+\cdot}$ on $[0,1]$, (5.36) gives

$$\Phi_\lambda(z_0) = \Phi_\lambda(z_1) = \ell^\varepsilon.$$

The dissipation identity forces z to be constant. It follows that

$$\sup_{0\le h\le 1} |Y_{n+h} - Y_n| \longrightarrow 0. \tag{5.40}$$

A violating sequence would have a nonconstant limiting segment. Uniform continuity of Φ_λ on the compact path closure extends the energy limit to all $\tau \to \infty$. Local uniform convergence of V_τ and the vanishing normalization give (3.8).

Every translated limit now has constant energy and is stationary by (5.39). The closed-tail argument gives

$$\varnothing \ne \omega(Y) \subseteq \mathcal{E}_\lambda \cap \{\Phi_\lambda = \ell^\varepsilon\}, \qquad \omega(Y) \text{ compact and connected.}$$

This is (3.9). If the distance to $\mathcal{E}_\lambda$ failed to tend to zero, boundedness would give an accumulation point outside this closed set, a contradiction.

Step 4 (Point convergence for empirical data). For finite empirical active laws, Lemma 5.1(ii) makes $\mathcal{E}_\lambda$ finite. Its nonempty connected subset $\omega(Y)$ is therefore a singleton. Boundedness then gives convergence of the entire path. The limit along integer times is measurable; assigning a fixed equilibrium on the exceptional null set defines the random vector $y^{\varepsilon,*}$ in (3.10). The inverse time change gives Remark 3.4.

Proof of Corollary 3.5. Let $Z = X_T^\varepsilon$ have law μ_T and set $Y_\tau = X_{t_\tau}^\varepsilon$. Use M from (5.29) when $\varepsilon > 0$, and put $M = 0$ when $\varepsilon = 0$. On each finite interval, the equation for $H = Y - M$ in Step 1 has a globally Lipschitz drift. Continuous dependence on the initial state and the continuous noise path gives a jointly measurable solution map.

For each fixed initial state, endpoint convergence follows from the theorem just proved, or from Theorem 2.5 when $\varepsilon = 0$. Since Z and M are independent, Fubini's theorem gives $Y_\tau \to Y_\infty \in \mathcal{E}_\lambda$ almost surely for any initial law μ_T. The limit along integer times, assigned a fixed equilibrium on the exceptional null set, is measurable. Set $X_0^\varepsilon = Y_\infty$ and $\mu_0^\varepsilon = \mathrm{Law}(X_0^\varepsilon)$. Bounded convergence gives

$$\mathbb{E} f(X_t^\varepsilon) \longrightarrow \mathbb{E} f(X_0^\varepsilon), \qquad f \in C_b(\mathbb{R}^d),$$

which proves narrow convergence of the laws. Since $\mathcal{E}_\lambda$ is finite and $X_0^\varepsilon \in \mathcal{E}_\lambda$ almost surely, $\operatorname{supp}\mu_0^\varepsilon \subseteq \mathcal{E}_\lambda$. □

Proof of Remark 3.6. Fix a path on which M converges and $r \in L^2(0,\infty)$, and put $R_\tau = M_\tau + \int_0^\tau r_s\,ds$. The random ODE for $\widetilde{Y} - R$ gives the strong solution as in Step 1. Formula (5.7) and the convolution estimate in Section 5.5, with kernel $e^{-a(\tau-s)}$, give confinement. Moreover,

$$\sup_{0\le h\le S}|R_{\tau+h}-R_\tau| \le \sup_{0\le h\le S}|M_{\tau+h}-M_\tau| + \sqrt{S}\,\|r\|_{L^2(\tau,\infty)} \longrightarrow 0.$$

The convergent function M is uniformly continuous, while $|\int_\tau^{\tau+h} r_s\,ds| \le \sqrt{h}\,\|r\|_{L^2(0,\infty)}$. Thus R is uniformly continuous; boundedness of the drift along the confined path gives the same property for $\widetilde{Y}$. Lemma 5.4 gives limiting paths satisfying (5.39). Put $\psi(\tau) = \Phi_\lambda(\widetilde{Y}_\tau)$. If its lower and upper limits differed, the finite set $\Phi_\lambda(\mathcal{E}_\lambda)$ would allow a choice

$$\eta \in (\liminf\psi, \limsup\psi)\setminus\Phi_\lambda(\mathcal{E}_\lambda), \qquad b \in (\liminf\psi,\eta).$$

Choose $s_j\to\infty$ with $\psi(s_j)=b$ and let $\tau_j = \inf\{\tau > s_j : \psi(\tau)=\eta\}$. These times are finite. Uniform continuity gives $h>0$ with $\tau_j - s_j > h$ for all j; hence $\psi\le\eta$ on $[\tau_j-h,\tau_j]$ and $\psi(\tau_j)=\eta$. For a limiting path z on $[0,h]$,

$$\Phi_\lambda(z_s)\le\eta=\Phi_\lambda(z_h), \qquad 0\le s\le h.$$

Dissipation forces constant energy and hence a constant path in $\mathcal{E}_\lambda$, contradicting the choice of η. Thus ψ converges. Dissipation makes all limiting paths stationary; finiteness and connectedness give a single random endpoint. Substitution of $r_\tau = (1+\varepsilon)t_\tau e(\widetilde{X}^\varepsilon_{t_\tau}, t_\tau)$ and $d\tau = -dt/t$ gives (3.13). □

6. Proofs of the technical lemmas

We retain the notation of Section 5 and prove the five lemmas in the order in which they were stated.

6.1. **Distance geometry and empirical rigidity.**

Proof of Lemma 5.1.

Step 1 (Distance calculus and the separated envelope). Apply the compact maximum rule [37, Theorem 10.31] to

$$d_A(x)^2 = |x|^2 - 2\max_{a\in A}\left(\langle x,a\rangle - \frac{|a|^2}{2}\right).$$

The distance formula and semiconcavity follow; in particular,

$$d_A(y)^2 \le d_A(x)^2 + \langle p, y-x\rangle + |y-x|^2, \qquad p\in\partial^C d_A^2(x); \tag{6.1}$$

see [9, Chapters 2–3]. For $0\le\lambda\le 1$, the convex sum rule applied to $|\cdot|^2 - d_k^2$ yields $\widehat{\partial}\Phi_\lambda = \partial^C\Phi_\lambda$ and 2-semiconcavity of Φ_λ. The Clarke sum rule gives $\partial^C\Phi_\lambda \subseteq \widehat{\partial}\Phi_\lambda$ for signed coefficients [12, 36]. Compactness of the nearest-point sets gives a closed graph and local boundedness of the envelope, hence upper semicontinuity.

Step 2 (Finiteness of empirical equilibria). Suppose first that $\lambda\notin\{0,1\}$. Fix nonempty index sets $I_k\subset\{1,\dots,n_k\}$ and choose $i_k\in I_k$ for $k=1,2$. Define

$$S = \operatorname{span}\{a_{k,i}-a_{k,i_k} : k=1,2,\ i\in I_k\}.$$

A point realizing these nearest-index sets is equidistant from the corresponding atoms. Expanding the squared distances gives

$$2\langle x, a_{k,i}-a_{k,i_k}\rangle = |a_{k,i}|^2 - |a_{k,i_k}|^2, \qquad k=1,2,\quad i\in I_k. \tag{6.2}$$

The solution set M, when nonempty, is an affine space with direction $S^\perp$. The equilibrium condition also places x in

$$C = \lambda \operatorname{conv}\{a_{1,i} : i \in I_1\} + (1-\lambda)\operatorname{conv}\{a_{2,i} : i \in I_2\}.$$

Each component convex hull is contained in its chosen atom plus the span of the corresponding differences. Hence

$$C \subseteq \lambda a_{1,i_1} + (1-\lambda)a_{2,i_2} + S.$$

For two points $x, y \in M \cap C$, the affine constraints and this last inclusion give, respectively,

$$x - y \in S^\perp, \qquad x - y \in S, \qquad \text{and therefore } x = y.$$

Each nearest-index pattern contains at most one equilibrium, and there are finitely many patterns. At $\lambda = 0$ or 1, the same argument uses only the active support. □

6.2. **Heat-flow Hessians and the Li–Yau bound.**

Proof of Lemma 5.2. The single-law Hessian identity is established in [31, Lemma 5.1 and its proof]; its signed combination gives (5.9). Positivity of the covariance matrices yields both curvature bounds. Since $\partial_t \log u_k = \Delta \log u_k + |\nabla \log u_k|^2$, taking the trace gives the Euclidean Li–Yau estimate [30]:

$$|\nabla \log u_k|^2 - \partial_t \log u_k = -\Delta \log u_k \le \frac{d}{2t}. \tag{6.3}$$

Compact supports bound the covariances uniformly. The Hessian formulas therefore give the positive-time Lipschitz bounds, while (5.2) gives linear growth. □

6.3. **Gaussian localization and empirical rates.**

Proof of Lemma 5.3. The scalar limit is the fixed-measure Laplace–Varadhan principle [17, Section 4.3]. We give the locally uniform estimate and the posterior localization needed here.

Step 1 (The locally uniform Laplace estimate). For $h > 0$, define

$$b_A(h) = \inf_{a \in A} u_0(B(a,h)).$$

Cover A by finitely many balls $B(a_j, h/2)$ centered on the support. For each $a \in A$, some j satisfies $|a - a_j| < h/2$, whence $B(a_j, h/2) \subset B(a,h)$. Each covering ball has positive mass, so

$$b_A(h) \ge \min_j u_0(B(a_j, h/2)) > 0.$$

Fix x, set $D = d_A(x)$, and choose $a_x \in \Pi_A(x)$. Restricting the Gaussian integral to $B(a_x, h)$ and using $|x - a| \ge D$ on A gives

$$b_A(h)(4\pi t)^{-d/2} e^{-(D+h)^2/(4t)} \le u(x,t) \le (4\pi t)^{-d/2} e^{-D^2/(4t)}.$$

Thus, for $\overline{f} = f - 2dt\log(4\pi t)$,

$$0 \le \overline{f}(x,t) - d_A(x)^2 \le 2d_A(x)h + h^2 - 4t\log b_A(h). \tag{6.4}$$

On a compact set B, put $D_B = \sup_{x\in B} d_A(x) < \infty$. For each fixed $h > 0$, the preceding bound gives

$$\limsup_{t\to 0^+} \sup_{x \in B} |\overline{f}(x,t) - d_A(x)^2| \le 2D_B h + h^2.$$

Letting $h \to 0^+$ proves local uniform convergence of $\overline{f}$. Since $2dt\log(4\pi t) \to 0$, the same conclusion holds for f.

Step 2 (Posterior localization with a moving center). Fix x_*, put $P = \Pi_A(x_*)$, and let $U_\rho = \{a \in A : \operatorname{dist}(a, P) < \rho\}$ for $\rho > 0$. When $A \setminus U_\rho$ is nonempty, it is compact and disjoint from P, so

$$\min_{a \in A\setminus U_\rho} \left(|x_* - a|^2 - d_A(x_*)^2\right) > 0.$$

Continuity then gives a bounded neighborhood $\mathcal{U}$ of x_* and $\gamma > 0$ such that

$$|x-a|^2 \geq d_A(x)^2 + \gamma, \qquad x \in \mathcal{U}, \quad a \in A \setminus U_\rho.$$

Choose $h > 0$ with $2Mh + h^2 \leq \gamma/2$, where $M = \sup_{x\in\mathcal{U}} d_A(x)$. The excluded Gaussian mass is at most $(4\pi t)^{-d/2} e^{-(d_A(x)^2+\gamma)/(4t)}$. Dividing it by the lower bound for $u(x,t)$ from Step 1 yields

$$\nu_t^x(A \setminus U_\rho) \leq b_A(h)^{-1} e^{-\gamma/(8t)}, \qquad x \in \mathcal{U}. \tag{6.5}$$

Let $C = \operatorname{conv} P$. The distance to C is convex, is at most ρ on U_ρ, and is at most $\operatorname{diam}(A)$ on A. Jensen's inequality gives

$$\operatorname{dist}(m(x,t), C) \leq \int_A \operatorname{dist}(a, C)\, d\nu_t^x(a) \leq \rho + \operatorname{diam}(A)\nu_t^x(A \setminus U_\rho). \tag{6.6}$$

When $U_\rho = A$, the same bound holds with zero excluded mass. For $x_n \to x_*$ and $t_n \to 0^+$, these estimates apply for all sufficiently large n and give

$$\limsup_{n\to\infty} \operatorname{dist}(m(x_n, t_n), C) \leq \rho.$$

Letting $\rho \to 0^+$ and using that C is closed proves the assertion about every cluster point of the bounded posterior means.

Step 3 (The empirical strict-cell estimate). The empirical posterior probabilities are

$$p_{k,i}(x,t) = \frac{w_{k,i} e^{-|x-a_{k,i}|^2/(4t)}}{\sum_j w_{k,j} e^{-|x-a_{k,j}|^2/(4t)}}, \qquad m_k(x,t) = \sum_i p_{k,i}(x,t) a_{k,i}. \tag{6.7}$$

There are finitely many competing atoms, and each has a positive squared-distance gap at x_0. Continuity gives a neighborhood U and a common $\Delta > 0$ such that

$$|x-a_{k,i}|^2 - |x-a_{k,*}|^2 \geq \Delta, \qquad x \in U, \quad i \neq *.$$

Keeping the nearest Gaussian term in the denominator yields

$$p_{k,i}(x,t) \leq \frac{w_{k,i}}{w_{k,*}} \exp\left(-\frac{|x-a_{k,i}|^2 - |x-a_{k,*}|^2}{4t}\right) \leq \frac{w_{k,i}}{w_{k,*}} e^{-\Delta/(4t)}.$$

Since the posterior weights sum to one,

$$m_k(x,t) - a_{k,*} = \sum_{i\neq *} p_{k,i}(x,t)(a_{k,i} - a_{k,*}),$$

$$|m_k(x,t) - a_{k,*}| \leq \left(\sum_{i\neq *} \frac{w_{k,i}}{w_{k,*}} |a_{k,i} - a_{k,*}|\right) e^{-\Delta/(4t)}.$$

Combining the component bounds in the mean-shift identity proves (5.10) with $\eta = \Delta/4$. Singleton supports contribute zero error; when every active support is a singleton, the identity is exactly affine. □

6.4. **Passage to limiting paths.**

Proof of Lemma 5.4.

Step 1 (Compactness of the translated curves). Condition (5.12) with $S = 1$, together with continuity on compact intervals, makes R uniformly continuous. Boundedness of y and the posterior-mean formula give $Q_0 < \infty$ with

$$|\nabla F_\lambda(y_s, t_s)| \leq Q_0 \qquad (s \geq 0).$$

For $0 \leq \tau \leq \tau'$, the integral equation gives

$$|y_{\tau'} - y_\tau| \leq \frac{\kappa}{4} Q_0 (\tau' - \tau) + |R_{\tau'} - R_\tau|.$$

The translates are bounded and equicontinuous. Arzelà–Ascoli and diagonal extraction give, along a subsequence of any $\tau_n \to \infty$,

$$y_{\tau_n+\cdot} \longrightarrow z \quad \text{locally uniformly on } [0,\infty).$$

Set $q_n(s) = \nabla F_\lambda(y_{\tau_n+s}, t_{\tau_n+s})$. After a further diagonal extraction, $q_n \rightharpoonup^* q$ in L^∞ on every bounded window. For $0 \le \tau < \tau'$, the translated equation reads

$$y_{\tau_n+\tau'} - y_{\tau_n+\tau} = -\frac{\kappa}{4}\int_\tau^{\tau'} q_n(s)\,ds + R_{\tau_n+\tau'} - R_{\tau_n+\tau}.$$

The final difference tends to zero by (5.12). Passing to the limit gives

$$z_{\tau'} - z_\tau = -\frac{\kappa}{4}\int_\tau^{\tau'} q(s)\,ds, \qquad \dot z_s = -\frac{\kappa}{4} q(s) \quad \text{for a.e. } s.$$

In particular, z is locally Lipschitz.

Step 2 (The weak-star constraint). For each fixed s, Lemma 5.3 and the posterior-mean formula place every cluster point of $q_n(s)$ in $C(s) := \widehat{\partial}\Phi_\lambda(z_s)$. Since $q_n(s)$ is bounded,

$$\operatorname{dist}(q_n(s), C(s)) \longrightarrow 0.$$

The sets $C(s)$ are nonempty, compact, and convex, with a common bound on each finite interval I. For $\xi \in \mathbb{R}^d$, their support function

$$h(s,\xi) = \max_{p\in C(s)} \langle \xi, p\rangle$$

is measurable by upper semicontinuity of the envelope and continuity of z. For every nonnegative $\varphi \in L^1(I)$,

$$\int_I \varphi(s)\langle \xi, q_n(s)\rangle\,ds \le \int_I \varphi(s) h(s,\xi)\,ds + |\xi| \int_I \varphi(s)\operatorname{dist}(q_n(s), C(s))\,ds.$$

The last integrand is dominated by a constant times φ. Dominated convergence and weak-star convergence consequently imply

$$\int_I \varphi(s)\big(\langle \xi, q(s)\rangle - h(s,\xi)\big)\,ds \le 0.$$

First take $\xi \in \mathbb{Q}^d$ and then use continuity in ξ. On one full-measure set,

$$\langle \xi, q(s)\rangle \le h(s,\xi) \quad (\xi \in \mathbb{R}^d), \qquad q(s) \in C(s).$$

The last inclusion is the support-function characterization of closed convex sets. Together with Step 1, it proves the lemma. □

6.5. **Deterministic and stochastic energy estimates.** We derive the posterior-information bound for part (i), then prove part (ii) using transition-density free energy.

Proof of part (i) *of Lemma 5.5.*

Step 1 (The information identity). For fixed x and $t > 0$, set

$$Z(x,t) = \int e^{-|x-a|^2/(4t)}\,du_0(a) > 0, \qquad \overline{f}(x,t) = -4t\log Z(x,t).$$

The posterior second moment $Q(x,t)$ is finite by Gaussian domination. Taking the posterior expectation of $\log(d\nu_t^x/du_0) = -|x-a|^2/(4t) - \log Z$ gives

$$\mathrm{KL}(\nu_t^x \| u_0) = -\frac{Q(x,t)}{4t} - \log Z(x,t).$$

Differentiation under the Gaussian integral gives $\partial_t Z/Z = Q/(4t^2)$. Consequently,

$$\overline{f}(x,t) = Q(x,t) + 4t\,\mathrm{KL}(\nu_t^x\|u_0), \qquad \partial_t \overline{f}(x,t) = 4\,\mathrm{KL}(\nu_t^x\|u_0). \tag{6.8}$$

Step 2 (The lower small-ball estimate). Assume (2.13). Since $Q(x,t) \geq d_A(x)^2$, (6.4) with $h = t$ gives, for $0 < t \leq t_B := \min\{r_0, 1\}$,

$$\begin{aligned} 0 \leq 4t\,\mathrm{KL}(\nu_t^x \| u_0) &= \overline{f}(x,t) - Q(x,t) \\ &\leq \overline{f}(x,t) - d_A(x)^2 \leq 2t d_A(x) + t^2 - 4t\log(ct^\alpha). \end{aligned} \tag{6.9}$$

On a compact set B, this gives $H_B(t) \leq C_B(1 + |\log t|)$, hence integrability near zero. Continuity of the Gaussian integrals and their ratios makes H_B bounded on compact positive-time intervals. For arbitrary x, division by $4t$ also yields the global bound

$$\mathrm{KL}(\nu_t^x \| u_0) \leq \tfrac{1}{2} d_A(x) + \tfrac{t}{4} - \log c - \alpha \log t \leq C(1 + |x| + |\log t|). \tag{6.10}$$

Step 3 (Potential variation). For each compactly supported law, Jensen's inequality gives

$$d_k(x)^2 \leq \overline{f}_k(x,t) \leq \int |x-b|^2 \, du_{0,k}(b).$$

Both bounds equal $|x|^2 + O(|x|+1)$, uniformly in t. The identity $\lambda + (1-\lambda) = 1$ and the mean-shift formula give the first three bounds below. Moreover, $0 \leq 4t\,\mathrm{KL}(\nu_{k,t}^x \| u_{0,k}) = \overline{f}_k(x,t) - Q_k(x,t) \leq \overline{f}_k(x,t)$, so (6.8) also controls the time derivative:

$$\begin{aligned} V_\tau(x) \geq \tfrac{1}{2}|x|^2 - C, \qquad |V_\tau(x)| \leq C(1+|x|^2), \\ |\nabla V_\tau(x)| \leq C(1+|x|), \qquad |\partial_\tau V_\tau(x)| \leq C(1+|x|^2). \end{aligned} \tag{6.11}$$

For $\lambda > 1$, apply (6.10) to the reference law and use nonnegativity of the target information. For $0 \leq \lambda \leq 1$, both information terms are nonnegative. The posterior identity therefore gives, for all x and sufficiently large τ,

$$\partial_\tau V_\tau(x) = -4t_\tau \mathcal{I}_\lambda(x, t_\tau) \leq C t_\tau (1 + |x| + |\log t_\tau|). \tag{6.12}$$

The left-hand side is nonpositive in the convex regime.

For a nonempty compact set B, define

$$b_B(\tau) = \begin{cases} 0, & 0 \leq \lambda \leq 1, \\ 4(\lambda-1) t_\tau H_{2,B}(t_\tau), & \lambda > 1. \end{cases}$$

Nonnegativity of relative entropy and (6.8) give (5.13). In the extrapolative case, Step 2 and the substitution $t = Te^{-\tau}$ yield

$$\int_0^\infty b_B(\tau)\, d\tau = 4(\lambda-1) \int_0^T H_{2,B}(t)\, dt < \infty.$$

Thus b_B is integrable, proving part (i). □

Part (ii) *of Lemma 5.5.* Fix $\varepsilon > 0$ and a restart (s, y), and put

$$\kappa = 1 + \varepsilon, \qquad a = \kappa/2, \qquad q = t_s = Te^{-s}.$$

Write $Y = Y^{\varepsilon;s,y}$ and let ρ_τ be its density for $\tau > s$. It satisfies

$$\partial_\tau \rho_\tau = \frac{\kappa}{4} \operatorname{div}(\rho_\tau \nabla V_\tau) + \varepsilon t_\tau \Delta \rho_\tau, \qquad \rho_s = \delta_y. \tag{6.13}$$

We first control the entropy produced by smoothing the point mass over a short interval, and then use a free-energy inequality on the remaining unit-time interval. All spatial integrals are over $\mathbb{R}^d$. Constants below are independent of y and sufficiently large s, with $\varepsilon > 0$ fixed.

Step 1 (Potential and restarted moment bounds). Take s large enough that $q \leq 1/2$ and (6.11) and (6.12) apply. Along Y, put

$$z_\tau = \lambda m_1(Y_\tau, t_\tau) + (1-\lambda) m_2(Y_\tau, t_\tau) \in K_\lambda, \qquad R_\lambda = \max_{z \in K_\lambda} |z|.$$

Since the drift equals $a(z_\tau - Y_\tau)$, the Duhamel formula (5.7) yields, for $s \le \tau \le s+1$,

$$Y_\tau = e^{-a(\tau-s)}y + a\int_s^\tau e^{-a(\tau-s')}z_{s'}\,ds' + \int_s^\tau e^{-a(\tau-s')}\sqrt{2\varepsilon t_{s'}}\,dW_{s'}.$$

The drift convolution is bounded by $R_\lambda(1-e^{-a(\tau-s)})$. As $t_{s'} \le q$ on this interval, Itô's isometry gives

$$\begin{aligned} \mathbb{E}_{s,y}|Y_\tau|^2 &= \int |x|^2\rho_\tau(x)\,dx \le C(1+|y|^2), \\ \mathbb{E}_{s,y}|Y_\tau - y|^2 &\le C\big[(\tau-s)^2(1+|y|^2) + q(\tau-s)\big]. \end{aligned} \tag{6.14}$$

Step 2 (Short-time smoothing and finite entropy). Choose $h = q$ and compare Y, on $[s, s+h]$, with

$$dZ_\tau = -aZ_\tau\,d\tau + \sqrt{2\varepsilon t_\tau}\,dW_\tau, \qquad Z_s = y.$$

By (5.7), the law γ_h of Z_{s+h} is Gaussian with mean $e^{-ah}y$ and covariance $\sigma_h^2 I_d$, where

$$\sigma_h^2 = 2\varepsilon\int_s^{s+h} e^{-2a(s+h-s')}t_{s'}\,ds', \qquad c_\varepsilon q^2 \le \sigma_h^2 \le C_\varepsilon q^2.$$

Thus the elapsed time and the diffusivity are both of order q, giving a variance of order q^2.

Let $\mathbb{P}$ and $\mathbb{Q}$ denote the path laws of Y and Z on $[s, s+h]$. Their drifts differ by $az_\tau(x)$, where

$$z_\tau(x) = \lambda m_1(x,t_\tau) + (1-\lambda)m_2(x,t_\tau) \in K_\lambda.$$

Thus the normalized drift difference satisfies, uniformly in x,

$$|\eta_\tau(x)|^2 := \frac{a^2|z_\tau(x)|^2}{2\varepsilon t_\tau} \le \frac{a^2R_\lambda^2}{2\varepsilon t_\tau}.$$

The uniform bound ensures that the associated stochastic exponential is a true martingale; by uniqueness, the changed law is that of Y. Girsanov's relative-entropy identity [29], followed by contraction to the terminal marginals, gives

$$\begin{aligned} \mathrm{KL}(\rho_{s+h}\|\gamma_h) \le \mathrm{KL}(\mathbb{P}\|\mathbb{Q}) &= \frac{1}{2}\mathbb{E}_{s,y}\int_s^{s+h}|\eta_\tau(Y_\tau)|^2\,d\tau \\ &\le \frac{a^2R_\lambda^2}{4\varepsilon}\frac{e^h-1}{q} \le \frac{a^2R_\lambda^2e^{1/2}}{4\varepsilon}, \qquad h = q \le 1/2. \end{aligned} \tag{6.15}$$

Expanding the Gaussian log-density, and dropping its nonpositive quadratic term, we obtain

$$\begin{aligned} \mathrm{Ent}(\rho_{s+h}) &= \mathrm{KL}(\rho_{s+h}\|\gamma_h) + \int \rho_{s+h}\log\gamma_h \\ &\le \frac{a^2R_\lambda^2e^{1/2}}{4\varepsilon} - \frac{d}{2}\log(2\pi\sigma_h^2) \le C(1+|\log q|). \end{aligned} \tag{6.16}$$

We also use the elementary lower bound, obtained by comparison with the standard Gaussian,

$$\mathrm{Ent}(\rho) \ge -\frac{d}{2}\log(2\pi) - \frac{1}{2}\int |x|^2\rho(x)\,dx. \tag{6.17}$$

The moment estimate and these two bounds make the entropy finite at $s+h$. For any fixed $\tau_0 > s$, the same Gaussian comparison over $[s, \tau_0]$ gives finite entropy at τ_0; choosing $h = q$ above is what makes the bound uniform in s, y.

Step 3 (Free-energy dissipation). On a fixed interval $[\tau_0, s+1]$, with $\tau_0 > s$, the transition density is smooth and positive, and the diffusion is uniformly elliptic. Moreover, the drift has at most linear

growth and ΔV_τ is bounded. The Fokker–Planck equation (6.13) gives the local identity

$$\partial_\tau(\rho_\tau \log \rho_\tau) = \operatorname{div}\Big(\frac{\kappa}{4}\rho_\tau \log \rho_\tau \nabla V_\tau + \varepsilon t_\tau(1+\log\rho_\tau)\nabla\rho_\tau\Big) + \frac{\kappa}{4}\rho_\tau \Delta V_\tau - \varepsilon t_\tau \frac{|\nabla\rho_\tau|^2}{\rho_\tau}.$$

Using the finite entropy and second-moment bounds from Steps 1–2, the convex entropy truncation and spatial localization argument of [31, Appendix B] justifies integration over $\mathbb{R}^d$ and yields

$$\operatorname{Ent}(\rho_\tau) + \int_{\tau_0}^{\tau} \varepsilon t_{s'} \underbrace{\int |\nabla \log \rho_{s'}|^2 \rho_{s'}\, dx}_{\text{Fisher information}}\, ds' = \operatorname{Ent}(\rho_{\tau_0}) + \frac{\kappa}{4}\int_{\tau_0}^{\tau}\int \Delta V_{s'}\, \rho_{s'}\, dx\, ds'. \tag{6.18}$$

The entropy lower bound (6.17) then implies finite integrated Fisher information. Together with the linear growth of ∇V_τ, this justifies the subsequent integrations by parts.

Define the weighted free energy

$$\theta_\tau = \frac{4\varepsilon t_\tau}{\kappa}, \qquad \mathscr{H}_\tau = \int V_\tau \rho_\tau + \theta_\tau \operatorname{Ent}(\rho_\tau). \tag{6.19}$$

The Fokker–Planck equation (6.13) becomes

$$\partial_\tau \rho_\tau = \frac{\kappa}{4}\operatorname{div}(\rho_\tau \nabla(V_\tau + \theta_\tau \log \rho_\tau)).$$

The time-derivative bound in (6.11) and the second-moment estimate justify differentiation of the potential term. Since $\theta_\tau' = -\theta_\tau$ and $\int \partial_\tau \rho_\tau = 0$, differentiation and integration by parts yield

$$\frac{d}{d\tau}\mathscr{H}_\tau = -\frac{\kappa}{4}\int |\nabla V_\tau + \theta_\tau \nabla \log \rho_\tau|^2 \rho_\tau + \int \partial_\tau V_\tau\, \rho_\tau - \theta_\tau \operatorname{Ent}(\rho_\tau). \tag{6.20}$$

The dissipation term is nonpositive even under signed mixing. Moreover, $q/e \le t_\tau \le q$ on $[s, s+1]$, so (6.12), (6.14), and (6.17) give

$$\int \partial_\tau V_\tau\, \rho_\tau \le Cq(1+|\log q|)(1+|y|^2),$$
$$-\theta_\tau \operatorname{Ent}(\rho_\tau) \le Cq(1+|y|^2).$$

Consequently,

$$\frac{d}{d\tau}\mathscr{H}_\tau \le Cq(1+|\log q|)(1+|y|^2) \qquad \text{for a.e. } \tau \in (s, s+1]. \tag{6.21}$$

Here C is independent of the restart time s and state y.

Step 4 (The conditional one-step bound). By the global gradient bound in (6.11),

$$|V_s(x) - V_s(y)| \le C(1+|y|)|x-y| + C|x-y|^2.$$

The second estimate in (6.14), with $h = q$, gives $\mathbb{E}_{s,y}|Y_{s+h} - y|^2 \le Cq^2(1+|y|^2)$. This controls the spatial change in potential. Integrating (6.12) controls its time change and yields

$$\int V_{s+h}\rho_{s+h} \le V_s(y) + Cq(1+|\log q|)(1+|y|^2).$$

Since $\theta_{s+h} \le Cq$, the upper entropy bound (6.16) gives

$$\mathscr{H}_{s+h} \le V_s(y) + Cq(1+|\log q|)(1+|y|^2). \tag{6.22}$$

Integrating (6.21) on $[s+h, s+1]$ provides the same upper bound for $\mathscr{H}_{s+1}$. Finally, (6.17) and (6.14) imply

$$\begin{aligned}\mathbb{E}_{s,y} V_{s+1}(Y_{s+1}) &= \mathscr{H}_{s+1} - \theta_{s+1} \operatorname{Ent}(\rho_{s+1}) \\ &\le \mathscr{H}_{s+1} + Cq(1+|y|^2) \\ &\le V_s(y) + Cq(1+|\log q|)(1+|y|^2),\end{aligned}$$

which is (5.14), because $q = t_s$. □

7. Conclusions and perspectives

We have identified the terminal asymptotics of deterministic and stochastic dynamics driven by convex and extrapolative combinations of logarithmic heat gradients. In logarithmic time, limits of translated trajectories satisfy a generalized gradient inclusion associated with the geometric potential Φ_λ, a weighted sum of squared support distances. For convex mixing, this is the Clarke subgradient system of Φ_λ. Under extrapolation, we retain the component subdifferentials separately, since limiting directions need not belong to $\partial^C \Phi_\lambda$. A componentwise chain rule still gives dissipation. Together with posterior-entropy and conditional free-energy estimates, it yields energy convergence and equilibrium localization under the stated compact-support hypotheses.

For finite atomic laws, finiteness of the equilibrium set turns this localization into point convergence for every deterministic initialization and almost surely for each fixed stochastic initialization, including projection-interface endpoints. Thus the singular sampling dynamics has a well-defined terminal output in this setting. At deterministic strict-cell endpoints, the expansion gives an $O(\sqrt{t})$ stopping error with an exponentially small remainder after subtraction of the leading term; the exponent $1/2$ is sharp when its coefficient is nonzero. The perturbation results show which convergence properties survive square-integrable drift errors in logarithmic time, while the nested-support corollary relates the equilibrium geometry to classifier-free guidance. Several questions remain.

(i) *Stochastic rates and fluctuations.* Energy control does not quantify the distance to a random endpoint. Pathwise rates, fluctuation laws, and mean-square errors require different estimates. Mean-square bounds also require control of paths that enter their limiting regime late. Figure 3 suggests square-root heat-time scaling of the stochastic RMS distance to the marked local-minimum set $\mathcal{M}_\lambda$, as measured in (4.2). A rate for the distance to the individual random endpoint would quantify the stopping error of stochastic samplers.

(ii) *Asymptotics at projection interfaces.* At projection ties, several atoms can contribute at leading order, and the positive gap underlying the strict-cell expansion vanishes. Analyzing this competition and possible differences between tangential and normal scales would extend the rate theory and clarify how support geometry affects approach to the terminal sample.

(iii) *Point convergence for general compact data.* An equilibrium energy level can contain a continuum, so localization alone need not yield a single limit. Additional nearest-point geometry or gradient inequalities compatible with the componentwise inclusion could establish terminal outputs beyond finite atomic laws.

(iv) *Endpoint selection and terminal distributions.* Equilibrium geometry does not determine reachability or selection probabilities, which depend on the full dynamics, initialization, and noise. For empirical data, determining these probabilities would identify the weights of the terminal law and, for genuine conditional guidance, clarify its relation to the intended conditional distribution.

(v) *Learning errors and time discretization.* The weighted square-integrability conditions in Remarks 3.2 and 3.6 preserve deterministic localization and empirical endpoint convergence for both dynamics, but need not preserve rates, selected equilibria, or terminal laws. These conditions concern errors along generated trajectories. Relating them to averaged training

losses and numerical step sizes requires estimates that remain valid through the singular regime.

Statements and declarations

Data availability. All numerical examples are synthetic and specified in Section 4. Numerical outputs and scripts are available from the corresponding author upon reasonable request.

Competing interests. The authors declare that they have no conflict of interest.

References

1. Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2 ed., Lectures in Mathematics ETH Zürich, Birkhäuser, Basel, 2008.
2. Brian D. O. Anderson, *Reverse-time diffusion equation models*, Stochastic Processes and their Applications **12** (1982), no. 3, 313–326.
3. Jean-Pierre Aubin and Hélène Frankowska, *Set-valued analysis*, Birkhäuser, Boston, 1990.
4. Iskander Azangulov, Peter Potaptchik, Qinyu Li, Eddie Aamari, George Deligiannidis, and Judith Rousseau, *Adaptive diffusion guidance via stochastic optimal control*, Proceedings of the 29th International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol. 300, PMLR, 2026, pp. 4087–4095.
5. Ricardo Baptista, Agnimitra Dasgupta, Nikola B. Kovachki, Assad Oberai, and Andrew M. Stuart, *Memorization and regularization in generative diffusion models*, 2025, arXiv:2501.15785v2.
6. Martino Bardi and Italo Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton–Jacobi–Bellman equations*, Systems & Control: Foundations & Applications, Birkhäuser, Boston, 1997.
7. Arwen Bradley and Preetum Nakkiran, *Classifier-free guidance is a predictor-corrector*, Transactions on Machine Learning Research (2025).
8. Jian-Feng Cai, Zhengyi Su, and Chao Wang, *Particle dynamics of flow matching and classifier-free guidance from a stagewise geometry perspective*, 2026, arXiv:2609.06947v1.
9. Piermarco Cannarsa and Carlo Sinestrari, *Semiconcave functions, Hamilton–Jacobi equations, and optimal control*, Progress in Nonlinear Differential Equations and Their Applications, vol. 58, Birkhäuser, Boston, 2004.
10. Ruijia Cao, Yuchen Wu, and Nisha Chandramoorthy, *On the robustness of distribution support under diffusion guidance*, 2026, arXiv:2605.07220v2.
11. Muthu Chidambaram, Khashayar Gatmiry, Sitan Chen, Holden Lee, and Jianfeng Lu, *What does guidance do? a fine-grained analysis in a simple setting*, Advances in Neural Information Processing Systems, vol. 37, 2024, pp. 84968–85005.
12. Frank H. Clarke, *Optimization and nonsmooth analysis*, Classics in Applied Mathematics, vol. 5, Society for Industrial and Applied Mathematics, Philadelphia, 1990.
13. Giovanni Conforti, Alain Durmus, and Marta Gentiloni Silveri, *KL convergence guarantees for score diffusion models under minimal data assumptions*, SIAM Journal on Mathematics of Data Science **7** (2025), no. 1, 86–109.
14. Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bulletin of the American Mathematical Society **27** (1992), no. 1, 1–67.
15. Giuseppe Da Prato and Jerzy Zabczyk, *Stochastic equations in infinite dimensions*, 2 ed., Encyclopedia of Mathematics and its Applications, vol. 152, Cambridge University Press, Cambridge, 2014.
16. Valentin De Bortoli, *Convergence of denoising diffusion models under the manifold hypothesis*, Transactions on Machine Learning Research (2022).
17. Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, 2 ed., Applications of Mathematics, vol. 38, Springer, New York, 1998.
18. Prafulla Dhariwal and Alexander Nichol, *Diffusion models beat GANs on image synthesis*, Advances in Neural Information Processing Systems, vol. 34, 2021, pp. 8780–8794.
19. Yilun Du, Conor Durkan, Robin Strudel, Joshua B. Tenenbaum, Sander Dieleman, Rob Fergus, Jascha Sohl-Dickstein, Arnaud Doucet, and Will Sussman Grathwohl, *Reduce, reuse, recycle: Compositional generation with energy-based diffusion models and MCMC*, Proceedings of the 40th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 202, 2023, pp. 8489–8510.
20. Wendell H. Fleming and H. Mete Soner, *Controlled Markov processes and viscosity solutions*, 2 ed., Stochastic Modelling and Applied Probability, vol. 25, Springer, New York, 2006.
21. Keinosuke Fukunaga and Larry D. Hostetler, *The estimation of the gradient of a density function, with applications in pattern recognition*, IEEE Transactions on Information Theory **21** (1975), no. 1, 32–40.

22. Victor A. Galaktionov and Juan Luis Vázquez, *A stability technique for evolution partial differential equations: A dynamical systems approach*, Progress in Nonlinear Differential Equations and Their Applications, vol. 56, Birkhäuser, Boston, 2004.
23. Ulrich G. Haussmann and Étienne Pardoux, *Time reversal of diffusions*, The Annals of Probability **14** (1986), no. 4, 1188–1205.
24. Jonathan Ho, Ajay Jain, and Pieter Abbeel, *Denoising diffusion probabilistic models*, Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 6840–6851.
25. Jonathan Ho and Tim Salimans, *Classifier-free diffusion guidance*, 2022, arXiv:2207.12598.
26. Aapo Hyvärinen, *Estimation of non-normalized statistical models by score matching*, Journal of Machine Learning Research **6** (2005), no. 24, 695–709.
27. Richard Jordan, David Kinderlehrer, and Felix Otto, *The variational formulation of the Fokker–Planck equation*, SIAM Journal on Mathematical Analysis **29** (1998), no. 1, 1–17.
28. Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, 2 ed., Graduate Texts in Mathematics, vol. 113, Springer, New York, 1991.
29. Christian Léonard, *Girsanov theory under a finite entropy condition*, Séminaire de Probabilités XLIV, Lecture Notes in Mathematics, vol. 2046, Springer, Berlin, 2012, pp. 429–465.
30. Peter Li and Shing-Tung Yau, *On the parabolic kernel of the Schrödinger operator*, Acta Mathematica **156** (1986), 153–201.
31. Kang Liu and Enrique Zuazua, *A PDE perspective on generative diffusion models*, 2026, arXiv:2511.05940v3.
32. Nan Liu, Shuang Li, Yilun Du, Antonio Torralba, and Joshua B. Tenenbaum, *Compositional visual generation with composable diffusion models*, Computer Vision – ECCV 2022, Springer, 2022, pp. 423–439.
33. Lawrence Markus, *Asymptotically autonomous differential systems*, Contributions to the Theory of Nonlinear Oscillations, Volume III, Annals of Mathematics Studies, vol. 36, Princeton University Press, 1956, pp. 17–29.
34. Daniel Revuz and Marc Yor, *Continuous martingales and Brownian motion*, 3 ed., Grundlehren der mathematischen Wissenschaften, vol. 293, Springer, Berlin, 1999.
35. Herbert Robbins and David Siegmund, *A convergence theorem for non negative almost supermartingales and some applications*, Optimizing Methods in Statistics (Jagdish S. Rustagi, ed.), Academic Press, New York, 1971, pp. 233–257.
36. R. Tyrrell Rockafellar, *Convex analysis*, Princeton Mathematical Series, vol. 28, Princeton University Press, Princeton, NJ, 1970.
37. R. Tyrrell Rockafellar and Roger J.-B. Wets, *Variational analysis*, Grundlehren der mathematischen Wissenschaften, vol. 317, Springer, Berlin, 1998.
38. Etienne Sandier and Sylvia Serfaty, *Gamma-convergence of gradient flows with applications to Ginzburg–Landau*, Communications on Pure and Applied Mathematics **57** (2004), no. 12, 1627–1672.
39. Sylvia Serfaty, *Gamma-convergence of gradient flows on Hilbert and metric spaces and applications*, Discrete and Continuous Dynamical Systems **31** (2011), no. 4, 1427–1451.
40. Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli, *Deep unsupervised learning using nonequilibrium thermodynamics*, Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 37, 2015, pp. 2256–2265.
41. Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, *Score-based generative modeling through stochastic differential equations*, International Conference on Learning Representations, 2021.
42. S. R. S. Varadhan, *On the behavior of the fundamental solution of the heat equation with variable coefficients*, Communications on Pure and Applied Mathematics **20** (1967), no. 2, 431–455.
43. Pascal Vincent, *A connection between score matching and denoising autoencoders*, Neural Computation **23** (2011), no. 7, 1661–1674.
44. Zhengchao Wan, Qingsong Wang, Gal Mishne, and Yusu Wang, *Elucidating flow matching ODE dynamics via data geometry and denoisers*, Proceedings of the 42nd International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 267, PMLR, 2025, pp. 62020–62083.