Uniform-in-Time Particle Convergence for Projected Gaussian SVGD

Abstract

Liu, Ghosal, Balasubramanian, and Pillai [12] conjectured uniform-in-time mean-field convergence for projected Gaussian Stein variational gradient descent (SVGD) beyond Gaussian targets. We give a quantitative affirmative answer for the continuous-time, exact-Gaussian-moment system with kernel $K_1(x, y) = x^\top y + 1$, assuming $\pi(dx) \propto e^{-V(x)}dx$, $V \in C^2(\mathbb{R}^d)$, and $0 < \alpha I \preceq \nabla^2 V \preceq \beta I$. For Gaussian i.i.d. initialization and every $N \geq d + 1$, the empirical particle law $\zeta_{N,t}$ and Gaussian mean-field law ρ_t satisfy $\mathbb{E}[\sup_{t\geq 0} W_2^2(\zeta_{N,t}, \rho_t)] \leq Cr_d(N)$. The asymptotic orders of $r_d(N)$ are $N^{-1} \log\log N$, $N^{-1}(\log N)^2$, and $N^{-2/d}$ in dimensions one, two, and at least three, respectively; each is sharp in particle number. Thus the time supremum can be placed inside the expectation without losing the Gaussian empirical-measure rate, even at the minimal full-rank sample size. We also obtain high-probability all-time bounds and a last-iterate guarantee to the reverse-KL optimal Gaussian at the same statistical scale. The key deterministic inequality separates initial moment mismatch from the Wasserstein discrepancy of the standardized cloud, using stability in square-root covariance coordinates and exact affine transport. At fixed N, we characterize the discrete limiting cloud: its moments are optimal, but its preserved standardized shape leaves a strictly positive full-law error.

Note. This paper was generated entirely by AI, including MiMo, using an automated research pipeline developed by Chenghua Liu and Hanyu Li.

1 Introduction and prior conjecture

This section states the prior conjecture, explains the obstacle to uniform full-law control, and summarizes our resolution. Detailed definitions and results follow in Section 2.

Stein variational gradient descent (SVGD) transports particles by a kernelized deterministic velocity field [10, 11]. Finite-horizon propagation-of-chaos constants can grow with the horizon, and time-averaged bounds do not control a physical-time (fixed-time or last-iterate) law; uniform-in-time full-law control is therefore delicate. Our result concerns a continuous-time Gaussian-score projection with exact Gaussian expectations, not the ordinary unprojected SVGD particle system.

In the arXiv version of Liu et al. [12], Theorem 3.7 gives a Gaussian-target, pointwise-in-time expectation bound with a constant uniform in time under Gaussian i.i.d. initialization. For general targets, their Theorem 4.3 derives the exact-moment affine representation and is followed by a qualitative conjecture of uniform-in-time mean-field convergence. We formulate a stronger quantitative version, making the sampling model explicit.

Open problem 1.1 (Quantitative exact-moment conjecture)**.** For a possibly non-Gaussian target satisfying $V \in C^2(\mathbb{R}^d)$ and $\alpha I \preceq \nabla^2 V \preceq \beta I$, initialize the projected $K_1(x, y) = x^\top y + 1$ particle system with i.i.d. samples from a full-rank Gaussian law ρ_0. At fixed dimension, can one place the time supremum inside the expectation, retain the sharp Gaussian empirical-measure rate, and obtain a physical-time last-iterate bound to the reverse-KL optimal Gaussian?

Exact moment closure alone does not answer this question. The empirical law retains a standardized shape that affine dynamics transports rather than Gaussianizes, while the moment vector field becomes sensitive near the positive-definite boundary. We resolve both issues under global strong convexity and a bounded continuous Hessian, without assuming a third derivative or a Hessian-Lipschitz modulus:

1. a square-root-coordinate energy yields semiglobal all-time stability on every energy sublevel;
2. a deterministic oracle separates moment mismatch from a time-independent standardized-shape discrepancy;
3. Gaussian matching and log-Wishart estimates give the sharp expected N-rate, a high-probability upper bound, and the endpoint $N = d + 1$; and
4. the exact finite-particle limit retains a positive shape floor.

A scalar counterexample delineates the scope of the first item: the semiglobal Euclidean Lipschitz estimate used here cannot be replaced by one initialization-uniform constant over the full positive-definite cone.

2 Model and main results

This section defines the projected moment and particle dynamics, then states the deterministic oracle, its probabilistic consequences, and the exact finite-particle limit. We close with the assumptions that delimit their scope.

Let $d \geq 1$ be fixed and let

$$\pi(dx) = Z_V^{-1} e^{-V(x)}\, dx, \qquad V \in C^2(\mathbb{R}^d),$$

where $Z_V < \infty$ by strong convexity and, for constants $0 < \alpha \leq \beta < \infty$,

$$\alpha I \preceq \nabla^2 V(x) \preceq \beta I, \qquad x \in \mathbb{R}^d. \tag{2.1}$$

We use W_2 for the quadratic Wasserstein distance on $\mathcal{P}_2(\mathbb{R}^d)$ and $T_{\#}\nu$ for pushforward by T. Vector norms are Euclidean; matrix norms carry subscripts F or op. For $q = (\mu, A) \in \mathbb{R}^d \times \mathrm{Sym}(d)$, the product norm is $\|q\|^2 = |\mu|^2 + \|A\|_{\mathrm{F}}^2$. For every $\Sigma \succ 0$, $\Sigma^{1/2}$ denotes its principal symmetric square root, and

$$\mathcal{Q} = \mathbb{R}^d \times \mathbb{S}^d_{++}, \qquad q = (\mu, A) = (\mu, \Sigma^{1/2}).$$

We also write $\gamma_d = \mathcal{N}(0, I_d)$, set $u_+ = \max\{u, 0\}$ and $\log_+ x = (\log x)_+$, and define $d_{\mathrm{TV}}(P, Q) = \sup_B |P(B) - Q(B)|$. Unless indexed otherwise, $C, c > 0$ may change from line to line and depend only on the fixed problem data in the relevant statement, never on N, t, or a tail parameter.

For $(\mu, \Sigma) \in \mathbb{R}^d \times \mathbb{S}^d_{++}$, define the exact Gaussian moments

$$m(\mu, \Sigma) = \mathbb{E}_{\mathcal{N}(\mu,\Sigma)}[\nabla V(X)], \qquad \Gamma(\mu, \Sigma) = \mathbb{E}_{\mathcal{N}(\mu,\Sigma)}[\nabla^2 V(X)]. \tag{2.2}$$

The Gaussian mean-field moment flow is

$$\dot{\mu} = (I - \Gamma\Sigma)\mu - (1 + |\mu|^2)m, \tag{2.3a}$$

$$\dot{\Sigma} = 2\Sigma - \Sigma(\Sigma\Gamma + \mu m^\top) - (\Gamma\Sigma + m\mu^\top)\Sigma, \tag{2.3b}$$

where $m = m(\mu_t, \Sigma_t)$ and $\Gamma = \Gamma(\mu_t, \Sigma_t)$. We write $\rho_t = \mathcal{N}(\mu_t, \Sigma_t)$ for the solution starting from fixed $\mu_0 \in \mathbb{R}^d$ and $\Sigma_0 \succ 0$.

Given initial points $X_1(0), \ldots, X_N(0)$, set

$$\widehat{\mu}_t = \frac{1}{N}\sum_{i=1}^{N} X_i(t), \qquad \widehat{\Sigma}_t = \frac{1}{N}\sum_{i=1}^{N}(X_i(t) - \widehat{\mu}_t)(X_i(t) - \widehat{\mu}_t)^\top. \tag{2.4}$$

With

$$\widehat{m}_t = m(\widehat{\mu}_t, \widehat{\Sigma}_t), \qquad \widehat{\Gamma}_t = \Gamma(\widehat{\mu}_t, \widehat{\Sigma}_t),$$

the projected particle system is

$$\dot{X}_i(t) = C_t X_i(t) - \widehat{m}_t, \qquad C_t = I - \widehat{\Gamma}_t\widehat{\Sigma}_t - \widehat{m}_t\widehat{\mu}_t^\top. \tag{2.5}$$

This is the Gaussian-score projection associated with K_1; it is not the ordinary, unprojected SVGD particle system. Define

$$\zeta_{N,t} = \frac{1}{N}\sum_{i=1}^{N}\delta_{X_i(t)}.$$

For Theorems 2.2 and 2.4, the initial points are drawn independently from $\mathcal{N}(\mu_0, \Sigma_0)$. Under this Gaussian initialization, all particle paths and standardized quantities are assigned an arbitrary fixed value on the null event $\{\widehat{\Sigma}_0 \not\succ 0\}$. On its complement the paths and Gaussian moments are continuous in t, hence so is $t \mapsto W_2^2(\zeta_{N,t}, \rho_t)$. Consequently

$$\sup_{t\geq 0} W_2^2(\zeta_{N,t}, \rho_t) = \sup_{t\in\mathbb{Q}_+} W_2^2(\zeta_{N,t}, \rho_t), \tag{2.6}$$

so every random supremum below is measurable.

For $Z \sim \gamma_d$, use the constant-free energy representative

$$F(\mu, A) = \mathbb{E}V(\mu + AZ) - \log\det A. \tag{2.7}$$

Then $\mathrm{KL}(\mathcal{N}(\mu, A^2)\|\pi) = F(\mu, A) + c_\pi$, where $c_\pi = \log Z_V - \frac{d}{2}(1 + \log(2\pi))$. Let $q_* = (\mu_*, A_*)$ be the unique minimizer, put $\Sigma_* = A_*^2$ and $\rho_* = \mathcal{N}(\mu_*, \Sigma_*)$, and define

$$\mathcal{K}_R = \{q : F(q) - F(q_*) \leq R\}. \tag{2.8}$$

Section 4 proves that these sublevels are compactly contained in the positive covariance domain.

We first state the deterministic result from which the probabilistic all-time estimates are lifted. Given a deterministic cloud $x_1, \ldots, x_N$ with empirical mean $\widehat{\mu}_0$ and covariance $\widehat{\Sigma}_0 \succ 0$, choose any whitening matrix R_0 satisfying $R_0\widehat{\Sigma}_0R_0^\top = I$ and set

$$z_i = R_0(x_i - \widehat{\mu}_0), \qquad \eta_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{z_i}. \tag{2.9}$$

The value of $W_2(\eta_N, \gamma_d)$ does not depend on the whitening: two whitening matrices differ by a left orthogonal factor, and γ_d is rotation invariant.

Theorem 2.1 (Deterministic localized oracle)**.** *Let $q_0 = (\mu_0, \Sigma_0^{1/2})$, let $q_t = (\mu_t, \Sigma_t^{1/2})$ be the solution of (2.3), set $\rho_t = \mathcal{N}(\mu_t, \Sigma_t)$, and let $\zeta_{N,t}$ be the projected particle solution from an arbitrary* *deterministic cloud with $\widehat{\Sigma}_0 \succ 0$. Put $q_{N,0} = (\widehat{\mu}_0, \widehat{\Sigma}_0^{1/2})$. For every $0 \le R < \infty$ there are finite constants S_R and b_R, depending only on R, d, V, such that, whenever $q_0, q_{N,0} \in \mathcal{K}_R$,*

$$\sup_{t\ge 0} W_2^2(\zeta_{N,t}, \rho_t) \le 2S_R^2 \left\|q_{N,0} - q_0\right\|^2 + 2b_R^2 W_2^2(\eta_N, \gamma_d). \tag{2.10}$$

Both the mean-field and particle solutions are global under these conditions. No independence or distributional assumption on the initial cloud is required.

In particular, any sequence of full-rank deterministic or dependent clouds lying in a common sublevel and satisfying $\left\|q_{N,0} - q_0\right\|^2 + W_2^2(\eta_N, \gamma_d) \to 0$ converges uniformly in physical time.

For an unambiguous statement at every admissible N, let

$$r_d(N) = \begin{cases} N^{-1}\{1 + \log\log(e^e + N)\}, & d = 1, \\ N^{-1}(1 + \log N)^2, & d = 2, \\ N^{-2/d}, & d \ge 3. \end{cases} \tag{2.11}$$

Thus, asymptotically, the first two rates are $N^{-1}\log\log N$ and $N^{-1}(\log N)^2$, respectively. The regularization inside the logarithms only makes these rates finite for every admissible sample size.

We now specialize the oracle to Gaussian sampling, first in expectation.

Theorem 2.2 (Uniform-in-time particle convergence)**.** *Assume* (2.1)*, fix $d \ge 1$, $\mu_0 \in \mathbb{R}^d$, and $\Sigma_0 \in \mathbb{S}_{++}^d$, let $N \ge d+1$, and draw the initial particles i.i.d. from $\mathcal{N}(\mu_0, \Sigma_0)$. Then $\widehat{\Sigma}_0 \succ 0$ almost surely, the systems* (2.3) *and* (2.5) *have global solutions, and there is a finite constant C, depending only on d, V, μ_0, Σ_0, such that*

$$\mathbb{E}\left[\sup_{t\ge 0} W_2^2(\zeta_{N,t}, \rho_t)\right] \le C r_d(N). \tag{2.12}$$

In particular, the same bound holds with the supremum outside the expectation. The constant is independent of N and t.

The next proposition shows that the expected rate cannot be improved in its N-dependence, even before the dynamics starts.

Proposition 2.3 (Sharpness in particle number)**.** *Under the Gaussian i.i.d. initialization of Theorem 2.2, for every fixed d, μ_0, and $\Sigma_0 \succ 0$ there is $c > 0$ such that*

$$\mathbb{E}\left[\sup_{t\ge 0} W_2^2(\zeta_{N,t}, \rho_t)\right] \ge c r_d(N), \qquad N \ge d+1. \tag{2.13}$$

Thus the N-dependence in Theorem 2.2 is optimal in every fixed dimension.

Proof. At time zero, $\zeta_{N,0}$ is the empirical measure of N i.i.d. samples from $\rho_0 = \mathcal{N}(\mu_0, \Sigma_0)$. Applying the invertible affine map $x \mapsto \mu_0 + \Sigma_0^{1/2} x$ to the standard Gaussian empirical matching problem, with $\xi_i \overset{\text{iid}}{\sim} \gamma_d$ and $\nu_N = N^{-1}\sum_i \delta_{\xi_i}$, gives

$$\mathbb{E}W_2^2(\zeta_{N,0}, \rho_0) \ge \lambda_{\min}(\Sigma_0)\mathbb{E}W_2^2(\nu_N, \gamma_d).$$

The expectation on the right is bounded below by $c_d r_d(N)$: in dimension one this is the two-sided Gaussian order-statistic estimate of Bobkov and Ledoux [5, Corollary 6.14] for $\mathbb{E}W_2^2$; in dimension two it follows from the nonstandard two-sample Gaussian matching lower bound of Talagrand [16, main result and display (1)] (if ν_N' is an independent copy of ν_N, then $W_2^2(\nu_N, \nu_N') \le 2W_2^2(\nu_N, \gamma_d) + 2W_2^2(\nu_N', \gamma_d)$); and for $d \ge 3$ it is the Gaussian matching lower bound of Ledoux and Zhu [9, Theorem 1.1]. Since time zero is included in the supremum, (2.13) follows. Reducing the constant handles the finitely many nonasymptotic sample sizes. □

The same decomposition also yields exponential concentration around the dimension-dependent statistical scale.

Theorem 2.4 (High-probability all-time convergence)**.** *Under the assumptions of Theorem 2.2, there are constants $C, c, c_0 > 0$, depending only on the fixed problem data, such that for every $N \geq d+1$ and $1 \leq x \leq c_0 N$, there is an event $H_N(x)$ with $\mathbb{P}(H_N(x)^c) \leq Ce^{-cx}$ on which*

$$\sup_{t\geq 0} W_2^2(\zeta_{N,t}, \rho_t) \leq C\left(r_d(N) + \frac{x}{N}\right). \tag{2.14}$$

On $H_N(x)$, simultaneously for every $t \geq 0$,

$$W_2^2(\zeta_{N,t}, \rho_*) \leq C\left\{r_d(N) + \frac{x}{N} + e^{-2ct}\right\}. \tag{2.15}$$

Equivalently, for $Ce^{-cN} \leq \delta \leq 1/2$, the statistical term is $C\{r_d(N) + N^{-1}\log(C/\delta)\}$ with probability at least $1-\delta$, after changing the fixed constants.

Combining all-time particle control with convergence of the mean-field flow gives a last-iterate guarantee rather than a time-averaged one.

Corollary 2.5 (Physical-time last iterate)**.** *Under the assumptions of Theorem 2.2, there are constants $C, c > 0$, depending only on the fixed problem data, such that for all $t \geq 0$,*

$$\mathbb{E}W_2^2(\zeta_{N,t}, \rho_*) \leq C\{r_d(N) + e^{-2ct}\}. \tag{2.16}$$

Consequently, once $t \geq (2c)^{-1}\log_+(1/r_d(N))$, the error is at the statistical scale $O(r_d(N))$.

The deterministic dynamics has an exact limit at every fixed particle number; the residual discrepancy is precisely an affine shape error.

Proposition 2.6 (Exact finite-particle limit and shape floor)**.** *Let a deterministic full-rank cloud and mean-field initialization satisfy $q_{N,0}, q_0 \in \mathcal{K}_R$ for some $R > 0$. There exist $C_R < \infty$, $\lambda_R > 0$, and an invertible matrix B_∞ with $B_\infty B_\infty^\top = \Sigma_*$ such that, for*

$$\zeta_{N,\infty} = (\mu_* + B_\infty \cdot)_\# \eta_N, \tag{2.17}$$

one has

$$W_2^2(\zeta_{N,t}, \zeta_{N,\infty}) \leq C_R e^{-2\lambda_R t}, \tag{2.18}$$

$$\lim_{t\to\infty} W_2(\zeta_{N,t}, \rho_t) = W_2(\zeta_{N,\infty}, \rho_*). \tag{2.19}$$

Moreover,

$$\lambda_{\min}(\Sigma_*) W_2^2(\eta_N, \gamma_d) \leq W_2^2(\zeta_{N,\infty}, \rho_*) \leq \lambda_{\max}(\Sigma_*) W_2^2(\eta_N, \gamma_d). \tag{2.20}$$

The constants C_R and λ_R depend only on R, d, V; the matrix B_∞ may depend on the initial cloud and the whitening coordinates, while the law $\zeta_{N,\infty}$ does not. Thus the moments converge exactly to (μ_*, Σ_*), while every full-rank finite cloud retains a strictly positive full-law error.

This shape floor persists for a fixed non-Gaussian sampling law, even as the number of particles grows.

Corollary 2.7 (Non-Gaussian initialization is not Gaussianized)**.** *On one probability space, let $(X_i)_{i\geq 1}$ be an infinite i.i.d. sequence from a non-Gaussian law P_0 having a density, positive-definite covariance Σ_0, and finite second moment. For each $N \geq d+1$, initialize the particle system with $X_1, \ldots, X_N$ and define $\zeta_{N,\infty}$ by Proposition 2.6 on its almost-sure full-rank event. Write*

$$\eta_0 = \left(x \mapsto \Sigma_0^{-1/2}(x - \mathbb{E}_{P_0} X)\right)_{\#} P_0.$$

Put $D_0 = W_2(\eta_0, \gamma_d) > 0$. Then, almost surely,

$$\lambda_{\min}(\Sigma_*) D_0^2 \leq \liminf_{N\to\infty} W_2^2(\zeta_{N,\infty}, \rho_*) \leq \limsup_{N\to\infty} W_2^2(\zeta_{N,\infty}, \rho_*) \leq \lambda_{\max}(\Sigma_*) D_0^2. \tag{2.21}$$

In particular, the limiting error is bounded away from zero; if $\Sigma_* = \sigma_*^2 I$, it converges to $\sigma_*^2 D_0^2$. Thus Gaussianity is a structural consistency condition for a fixed i.i.d. initialization law in Theorems 2.2 and 2.4, not merely a convenient source of concentration. This does not preclude consistent deterministic or N-dependent designs covered by the oracle.

Remark 2.8 (Scope). The probabilistic theorems concern continuous time, exact evaluation of (2.2), fixed dimension, Gaussian i.i.d. initialization, and the projected system (2.5). The deterministic oracle does allow arbitrary full-rank initialization. These are not dimension-free results or theorems for ordinary SVGD, Euler discretization, or Monte Carlo estimators of m and Γ. Since every centered N-point covariance has rank at most $N-1$, the restriction $N \geq d+1$ is necessary for this unregularized positive-definite dynamics.

3 Proof overview

This section gives the mechanism behind the results and records where each technical step enters. Complete arguments, including the endpoint and rare event calculations, are developed in Sections 4–7.

In square-root coordinates $q = (\mu, A)$, the energy in (2.7) has second variation

$$D^2F(\mu, A)[(u, H), (u, H)] = \mathbb{E}[(u + HZ)^\top \nabla^2 V(\mu + AZ)(u + HZ)] + \operatorname{tr}(A^{-1}HA^{-1}H), \tag{3.1}$$

which is at least $\alpha(|u|^2 + \|H\|_{\mathrm{F}}^2)$. Its sublevels are compactly contained in $\mathcal{Q}$. The moment flow can be written as $\dot{q} = -P(q)\nabla F(q)$ with a positive-definite mobility $P(q)$; on each sublevel, $P(q) \succeq p_R I$. This yields exponential convergence to q_*. For two trajectories, a Gronwall bound controls the common compact sublevel until both enter a fixed neighborhood of q_*. There the linearization is dissipative in the $D^2F(q_*)$-norm, and patching the two regimes proves the all-time stability estimate. Section 4 and Section 5 give the details.

The particle equation is affine. Its empirical moments solve the same closed ODE, and a fundamental matrix L_t gives

$$X_i(t) = \widehat{\mu}_t + L_t\{X_i(0) - \widehat{\mu}_0\}, \qquad \widehat{\Sigma}_t = L_t \widehat{\Sigma}_0 L_t^\top. \tag{3.2}$$

After whitening the initial cloud, both the empirical law and its moment- matched Gaussian are pushforwards by the same affine map. Consequently,

$$\lambda_{\min}(\widehat{\Sigma}_t) W_2^2(\eta_N, \gamma_d) \leq W_2^2\left(\zeta_{N,t}, \mathcal{N}(\widehat{\mu}_t, \widehat{\Sigma}_t)\right) \leq \lambda_{\max}(\widehat{\Sigma}_t) W_2^2(\eta_N, \gamma_d). \tag{3.3}$$

Combining this with semiglobal moment stability proves the deterministic oracle. Exponential moment convergence also makes the affine coefficient integrable, producing the limit and the two-sided shape floor in Proposition 2.6. These arguments are in Section 6.

For Gaussian initialization, concentration of the sample mean and covariance places the empirical initial state in a deterministic common sublevel, while Gaussian empirical matching controls $W_2(\eta_N, \gamma_d)$. This proves the high-probability theorem directly from the oracle. For the expectation, the rare nearly singular event cannot be discarded; energy decrease reduces its contribution to a log-determinant moment. Bartlett's decomposition gives a sum of $\log \chi_k^2$ variables whose second moments remain bounded even when $N = d + 1$. Cauchy–Schwarz then makes the rare-event contribution exponentially small. Section 7 contains the full localization and matching argument.

4 Reverse-KL geometry in square-root coordinates

This section identifies the gradient-flow structure behind the moment ODE. We first compute its Stein mobility and then use square-root coordinates to obtain strong convexity, compact sublevels, and quantitative convergence.

All matrix tangent and cotangent variables below lie in $\mathrm{Sym}(d)$, equipped with the Frobenius inner product. The product inner product is

$$\langle (a, B), (u, H) \rangle = a^\top u + \mathrm{tr}(BH).$$

For $X \sim \mathcal{N}(\mu, \Sigma)$, the exact reverse-KL energy is

$$\mathcal{E}(\mu, \Sigma) = \mathrm{KL}(\mathcal{N}(\mu, \Sigma) \| \pi) = \mathbb{E}V(X) - \frac{1}{2} \log \det \Sigma + c_\pi. \tag{4.1}$$

Lemma 4.1 (Energy gradient and Stein mobility)**.** *For m and Γ from* (2.2)*,*

$$\nabla_\mu \mathcal{E} = m, \qquad \nabla_\Sigma \mathcal{E} = \frac{1}{2}(\Gamma - \Sigma^{-1}). \tag{4.2}$$

For $(a, B) \in \mathbb{R}^d \times \mathrm{Sym}(d)$, define

$$[\mathcal{M}_{\mu,\Sigma}(a, B)]_\mu = 2B\Sigma\mu + (1 + |\mu|^2)a, \tag{4.3a}$$

$$[\mathcal{M}_{\mu,\Sigma}(a, B)]_\Sigma = \Sigma(2\Sigma B + \mu a^\top) + (2B\Sigma + a\mu^\top)\Sigma. \tag{4.3b}$$

Then

$$\langle (a, B), \mathcal{M}_{\mu,\Sigma}(a, B) \rangle = \left\| 2B\Sigma + a\mu^\top \right\|_\mathrm{F}^2 + |a|^2 . \tag{4.4}$$

In particular, $\mathcal{M}_{\mu,\Sigma}$ is symmetric positive definite, and (2.3) *is precisely*

$$(\dot{\mu}, \dot{\Sigma}) = -\mathcal{M}_{\mu,\Sigma} \nabla \mathcal{E}(\mu, \Sigma). \tag{4.5}$$

Proof. For $H \in \mathrm{Sym}(d)$ and all sufficiently small ε, Gaussian differentiation (Price's identity) gives

$$D_\mu \mathbb{E}V(X)[u] = u^\top m, \qquad \frac{d}{d\varepsilon} \mathbb{E}_{\mathcal{N}(\mu, \Sigma + \varepsilon H)} V(X) \bigg|_{\varepsilon=0} = \frac{1}{2}\, \mathrm{tr}(\Gamma H).$$

The growth needed to differentiate under the integral follows from (2.1); differentiation of the log determinant then proves (4.2).

More generally, direct expansion gives the bilinear identity

$$\left\langle (a, B), \mathcal{M}_{\mu,\Sigma}(a', B') \right\rangle = \left\langle 2B\Sigma + a\mu^\top, 2B'\Sigma + a'\mu^\top \right\rangle_\mathrm{F} + a^\top a'. \tag{4.6}$$

The right-hand side is unchanged when the primed and unprimed variables are interchanged, so $\mathcal{M}_{\mu,\Sigma}$ is self-adjoint. Taking the two arguments equal gives

$$|a|^2 + |\mu|^2 |a|^2 + 4a^\top B\Sigma\mu + 4\operatorname{tr}(B\Sigma^2 B),$$

which is (4.4); it vanishes only when $a = 0$ and $B = 0$ because $\Sigma \succ 0$. Substituting $a = m$ and $B = (\Gamma - \Sigma^{-1})/2$ in $-\mathcal{M}(a, B)$ yields (2.3a) and (2.3b). $\square$

Recall that $q = (\mu, A)$ with $A = \Sigma^{1/2}$ and that F is defined by (2.7). Thus $\mathcal{E}(\mu, A^2) = F(\mu, A) + c_\pi$, so the two energies have identical derivatives in these coordinates.

The next lemma supplies the strong convexity and coercivity used throughout the stability analysis.

Lemma 4.2 (Regularity, strong convexity, and coercivity)**.** *The function F is C^2 on $\mathcal{Q}$ and, for $(u, H) \in \mathbb{R}^d \times \operatorname{Sym}(d)$,*

$$\begin{aligned} D^2F(\mu, A)[(u, H), (u, H)] &= \mathbb{E}\Big[(u + HZ)^\top \nabla^2 V(\mu + AZ)(u + HZ)\Big] \\ &\quad + \operatorname{tr}(A^{-1}HA^{-1}H) \\ &\geq \alpha\big(|u|^2 + \|H\|_{\mathrm{F}}^2\big). \end{aligned} \tag{4.7}$$

Moreover, F has compact sublevel sets in $\mathcal{Q}$ and hence a unique minimizer $q_* = (\mu_*, A_*)$.

Proof. The Hessian bound implies $|\nabla V(x)| \leq C + \beta|x|$ and $|V(x)| \leq C(1 + |x|^2)$. The first two derivatives of $V(\mu + AZ)$ are therefore dominated locally in (μ, A) by integrable polynomials in Z. In particular,

$$DF(\mu, A)[u, H] = \mathbb{E}[\nabla V(\mu + AZ)^\top (u + HZ)] - \operatorname{tr}(A^{-1}H).$$

On a compact parameter neighborhood the integrands in the first and second variations are bounded respectively by $C(1 + |Z|)^2$ and $C(1 + |Z|)^2$. Since $\nabla^2 V$ is bounded and continuous, dominated convergence gives $F \in C^2$ and the displayed second-variation formula. The first term in (4.7) is at least

$$\alpha\mathbb{E}\,|u + HZ|^2 = \alpha(|u|^2 + \|H\|_{\mathrm{F}}^2),$$

and the log-determinant term is nonnegative because $\operatorname{tr}(A^{-1}HA^{-1}H) = \big\|A^{-1/2}HA^{-1/2}\big\|_{\mathrm{F}}^2$.

Let x_V be the unique minimizer of V and set $c_V = V(x_V)$. Strong convexity gives, with $a_1, \ldots, a_d$ the eigenvalues of A,

$$F(\mu, A) \geq c_V + \frac{\alpha}{2}|\mu - x_V|^2 + \sum_{j=1}^d \left(\frac{\alpha}{2}a_j^2 - \log a_j\right). \tag{4.8}$$

The right-hand side diverges if $|\mu| \to \infty$, if an eigenvalue of A tends to zero, or if $\|A\|_{\mathrm{op}} \to \infty$. Thus every sublevel is compactly contained in $\mathcal{Q}$. Existence follows by compactness and uniqueness by (4.7). Under the bijection $\Sigma = A^2$, this minimizer corresponds to the unique reverse-KL optimal Gaussian $\rho_* = \mathcal{N}(\mu_*, \Sigma_*)$, with $\Sigma_* = A_*^2$. $\square$

Let $\Phi(\mu, A) = (\mu, A^2)$. Its differential is

$$J_q(u, H) = (u, AH + HA). \tag{4.9}$$

The Sylvester map $H \mapsto AH + HA$ is invertible on $\mathrm{Sym}(d)$ when $A \succ 0$. The chain rule and Lemma 4.1 therefore transform the moment flow into

$$\dot{q} = f(q) = -P(q)\nabla F(q), \qquad P(q) = J_q^{-1}\mathcal{M}_{\Phi(q)}J_q^{-\mathsf{T}}. \tag{4.10}$$

Here P is smooth, symmetric, and positive definite. Lemma 4.2 gives $\nabla F \in C^1$, and therefore $f = -P\nabla F \in C^1(\mathcal{Q})$. This is the only local ODE regularity used below.

Proposition 4.3 (Sublevel ellipticity and convergence)**.** *For $R \geq 0$, recall*

$$\mathcal{K}_R = \{q \in \mathcal{Q} : F(q) - F(q_*) \leq R\}.$$

There are $M_R, a_R, b_R > 0$ such that on $\mathcal{K}_R$,

$$|\mu| \leq M_R, \qquad a_R I \preceq A \preceq b_R I. \tag{4.11}$$

For $R > 0$, one may take

$$P(q) \succeq p_R I, \qquad p_R = \frac{\min(1, 4a_R^4)}{2(1 + M_R^2)\max(1, 4b_R^2)} > 0. \tag{4.12}$$

If $R > 0$, every solution of (4.10) *starting in $\mathcal{K}_R$ is global, remains in $\mathcal{K}_R$, and satisfies*

$$F(q_t) - F(q_*) \leq e^{-2\alpha p_R t}\{F(q_0) - F(q_*)\}, \tag{4.13}$$

$$\|q_t - q_*\| \leq \sqrt{2R/\alpha}\, e^{-\alpha p_R t}. \tag{4.14}$$

For $R = 0$, the only trajectory is $q_t = q_*$, which is global and stationary.

Proof. For $R = 0$, uniqueness gives $\mathcal{K}_0 = \{q_*\}$, so all claims follow with the stationary trajectory $q_t = q_*$. Assume henceforth that $R > 0$. The bounds (4.11) follow from Lemma 4.2. To prove (4.12), put $s = \lambda_{\min}(\Sigma)$ and $U = 2B\Sigma$. Then

$$\|U\|_{\mathrm{F}}^2 + |a|^2 \leq 2(1 + |\mu|^2)\{\left\|U + a\mu^\top\right\|_{\mathrm{F}}^2 + |a|^2\},$$

whereas

$$\|U\|_{\mathrm{F}}^2 + |a|^2 \geq \min(1, 4s^2)(\|B\|_{\mathrm{F}}^2 + |a|^2).$$

Thus

$$\mathcal{M}_{\mu,\Sigma} \succeq \frac{\min(1, 4s^2)}{2(1 + |\mu|^2)} I.$$

Since $s \geq a_R^2$ and $\|J_q\|_{\mathrm{op}}^2 \leq \max(1, 4b_R^2)$ on $\mathcal{K}_R$, the congruence $P(q) = J_q^{-1}\mathcal{M}_{\Phi(q)}J_q^{-\mathsf{T}}$ gives (4.12): indeed, $\left\|J_q^{-\mathsf{T}}z\right\| \geq \|z\| \,/\, \|J_q\|_{\mathrm{op}}$.

Strong convexity and the fact that the segment from q to the interior minimizer q_* remains in $\mathcal{Q}$ imply the Polyak–Lojasiewicz bound

$$\|\nabla F(q)\|^2 \geq 2\alpha\{F(q) - F(q_*)\}. \tag{4.15}$$

Along the flow,

$$\frac{d}{dt}\{F(q_t) - F(q_*)\} = -\nabla F(q_t)^\top P(q_t)\nabla F(q_t) \leq -2\alpha p_R\{F(q_t) - F(q_*)\}.$$

This proves invariance and (4.13). Strong convexity also gives $F(q) - F(q_*) \geq (\alpha/2)\|q - q_*\|^2$, proving (4.14). Finally, a local solution trapped in the compact set $\mathcal{K}_R \Subset \mathcal{Q}$ cannot explode or reach the positive-definite boundary, so the standard continuation theorem gives global existence. $\square$

5 Semiglobal all-time stability

The next lemma is the deterministic core of the argument. Its constant is uniform in time but is allowed to depend on the initial energy sublevel.

Lemma 5.1 (Semiglobal uniform stability)**.** *For every $0 \leq R < \infty$ there is $S_R < \infty$ such that any two solutions of* (4.10) *with $q_0, \widetilde{q}_0 \in \mathcal{K}_R$ satisfy*

$$\sup_{t \geq 0} \|q_t - \widetilde{q}_t\| \leq S_R \|q_0 - \widetilde{q}_0\| . \tag{5.1}$$

If $R > 0$, there also exist $T_R, D_R, \kappa_R > 0$, depending only on R, d, V, such that

$$\|q_t - \widetilde{q}_t\| \leq D_R e^{-\kappa_R (t - T_R)} \|q_0 - \widetilde{q}_0\| , \qquad t \geq T_R.$$

Proof. The argument has two regimes. Sublevel invariance first gives a common compact set and hence an early-time Gronwall bound; exponential energy decay then brings both trajectories into a neighborhood of q_* where a fixed H_*-norm is contractive. Patching the estimates yields the all-time constant and eventual exponential decay.

The case $R = 0$ is immediate. Suppose $R > 0$ and set

$$H_* = D^2 F(q_*), \qquad P_* = P(q_*), \qquad \kappa_* = \lambda_{\min}(P_*)\lambda_{\min}(H_*) > 0.$$

Since $\nabla F(q_*) = 0$,

$$Df(q_*) = -P_* H_*.$$

In the fixed norm $\|v\|_{H_*}^2 = v^\top H_* v$, write $\mathrm{cond}(H_*) = \lambda_{\max}(H_*)/\lambda_{\min}(H_*)$. Then

$$v^\top H_* Df(q_*) v = -(H_* v)^\top P_* (H_* v) \leq -\kappa_* \|v\|_{H_*}^2 . \tag{5.2}$$

Continuity of Df gives $\varrho > 0$ such that the convex ellipsoid

$$\mathcal{U} = \{q : \|q - q_*\|_{H_*} < \varrho\} \Subset \mathcal{Q}$$

satisfies, for all $q \in \mathcal{U}$ and all v,

$$v^\top H_* Df(q) v \leq -\frac{\kappa_*}{2} \|v\|_{H_*}^2 . \tag{5.3}$$

Indeed, continuity is used here to control the symmetric part of $H_* Df$:

$$\sup_{\|v\|_{H_*} = 1} \left| v^\top H_* \{Df(q) - Df(q_*)\} v \right| \longrightarrow 0 \qquad (q \to q_*).$$

For $e = q - q_*$, the segment formula and $f(q_*) = 0$ give

$$f(q) - f(q_*) = \int_0^1 Df(q_* + \theta e) e \, d\theta.$$

Consequently, whenever the segment lies in $\mathcal{U}$,

$$\frac{d}{dt} \|q_t - q_*\|_{H_*}^2 = 2 \int_0^1 e_t^\top H_* Df(q_* + \theta e_t) e_t \, d\theta \leq -\kappa_* \|e_t\|_{H_*}^2 . \tag{5.4}$$

Since every concentric ellipsoid is convex, the usual first-exit argument applied to (5.4) shows that each closed concentric ellipsoid contained in $\mathcal{U}$ is forward invariant.

If q_t and $\widetilde{q}_t$ lie in the same such ellipsoid, then the segment joining them also lies there and

$$f(q_t) - f(\widetilde{q}_t) = \int_0^1 Df(\widetilde{q}_t + \theta(q_t - \widetilde{q}_t))(q_t - \widetilde{q}_t)\, d\theta.$$

Thus

$$\frac{d}{dt}\, \|q_t - \widetilde{q}_t\|_{H_*}^2 \le -\kappa_* \, \|q_t - \widetilde{q}_t\|_{H_*}^2 . \tag{5.5}$$

By (4.14), all trajectories starting in $\mathcal{K}_R$ enter the ellipsoid of radius $\varrho/2$ no later than

$$T_R = \frac{1}{2\alpha p_R} \log_+ \left(\frac{8\lambda_{\max}(H_*) R}{\alpha \varrho^2} \right). \tag{5.6}$$

Before T_R, all trajectories remain in the convex compact set

$$\mathcal{C}_R = \{(\mu, A) : |\mu| \le M_R,\, a_R I \preceq A \preceq b_R I\} \Subset \mathcal{Q}.$$

Let $\ell_R = \sup_{q \in \mathcal{C}_R} \|Df(q)\|_{\mathrm{op}} < \infty$, where the operator norm is induced by the Euclidean–Frobenius product norm. The segment between the two trajectories remains in $\mathcal{C}_R$, so Gronwall's inequality gives

$$\|q_t - \widetilde{q}_t\| \le e^{\ell_R t} \, \|q_0 - \widetilde{q}_0\| , \qquad 0 \le t \le T_R. \tag{5.7}$$

After T_R, both trajectories remain in the half-radius ellipsoid. Combining (5.5), norm equivalence, and (5.7) gives

$$\|q_t - \widetilde{q}_t\| \le \sqrt{\mathrm{cond}(H_*)}\, e^{-\kappa_*(t - T_R)/2} e^{\ell_R T_R} \, \|q_0 - \widetilde{q}_0\| , \qquad t \ge T_R.$$

Together with (5.7), this proves (5.1) and the claimed eventual exponential decay. In particular, the construction gives the explicit admissible choice

$$S_R = \sqrt{\mathrm{cond}(H_*)}\, e^{\ell_R T_R}. \tag{5.8}$$

□

Remark 5.2 (Dependence of the stability constant). Finiteness of S_R requires no global bound on $\nabla^3 V$: continuity of $D^2 F$ on the relevant compact parameter sets suffices. A closed-form upper bound for S_R, however, requires a quantitative modulus of continuity for $\nabla^2 V$ on those sets. A global bound on $\nabla^3 V$ is one sufficient, but not necessary, condition for such a modulus.

The synchronous coupling $\mu + AZ$ and $\widetilde{\mu} + \widetilde{A}Z$ gives the useful consequence

$$W_2^2(\mathcal{N}(\mu, A^2), \mathcal{N}(\widetilde{\mu}, \widetilde{A}^2)) \le |\mu - \widetilde{\mu}|^2 + \left\| A - \widetilde{A} \right\|_{\mathrm{F}}^2 . \tag{5.9}$$

6 Exact affine decomposition and shape error

We next derive the moment closure and affine representation directly from the particle system. The resulting affine lift separates the evolving moment error from a time-independent discrepancy of the standardized shape.

Lemma 6.1 (Exact closure and affine transport). *If $\widehat{\Sigma}_0 \succ 0$, the particle system has a unique global solution. Its empirical moments solve* (2.3) *from* $(\widehat{\mu}_0, \widehat{\Sigma}_0)$ *and remain positive definite. If L_t solves*

$$\dot{L}_t = C_t L_t, \qquad L_0 = I, \tag{6.1}$$

then

$$X_i(t) = \widehat{\mu}_t + L_t(X_i(0) - \widehat{\mu}_0), \tag{6.2}$$

$$\widehat{\Sigma}_t = L_t \widehat{\Sigma}_0 L_t^\top. \tag{6.3}$$

Proof. We first construct the particle solution, thereby avoiding any continuation argument that presupposes positive definiteness of its empirical covariance. Let

$$q_t^N = (\mu_t^N, A_t^N), \qquad \Sigma_t^N = (A_t^N)^2,$$

be the global solution of (4.10) from $(\widehat{\mu}_0, \widehat{\Sigma}_0^{1/2})$, whose existence follows from Proposition 4.3. Along this trajectory define

$$m_t^N = m(\mu_t^N, \Sigma_t^N), \quad \Gamma_t^N = \Gamma(\mu_t^N, \Sigma_t^N), \quad C_t^N = I - \Gamma_t^N \Sigma_t^N - m_t^N (\mu_t^N)^\top,$$

and let L_t be the global solution of $\dot{L}_t = C_t^N L_t$, $L_0 = I$. Now set

$$X_i(t) = \mu_t^N + L_t\{X_i(0) - \widehat{\mu}_0\}. \tag{6.4}$$

Its empirical mean is μ_t^N. Its empirical covariance

$$\widetilde{\Sigma}_t = L_t \widehat{\Sigma}_0 L_t^\top$$

solves

$$\dot{\widetilde{\Sigma}}_t = C_t^N \widetilde{\Sigma}_t + \widetilde{\Sigma}_t (C_t^N)^\top, \qquad \widetilde{\Sigma}_0 = \widehat{\Sigma}_0.$$

The covariance equation in (2.3) can equivalently be written

$$\dot{\Sigma}_t^N = C_t^N \Sigma_t^N + \Sigma_t^N (C_t^N)^\top.$$

Uniqueness for this linear matrix equation therefore gives $\widetilde{\Sigma}_t = \Sigma_t^N$. Likewise, the mean equation is $\dot{\mu}_t^N = C_t^N \mu_t^N - m_t^N$. Differentiating (6.4) now verifies

$$\dot{X}_i(t) = C_t^N X_i(t) - m_t^N,$$

so the constructed paths solve (2.5) with their own empirical moments.

For completeness, any local particle solution with positive-definite empirical covariance satisfies, by averaging and centering,

$$\dot{\widehat{\mu}} = C_t \widehat{\mu} - \widehat{m}, \qquad \dot{\widehat{\Sigma}} = C_t \widehat{\Sigma} + \widehat{\Sigma} C_t^\top.$$

These are exactly (2.3); uniqueness of the q-flow forces its moments to equal (μ_t^N, Σ_t^N). Its centered particles then solve the same linear equation $\dot{Y}_i = C_t^N Y_i$ with the same initial data, proving uniqueness. Finally,

$$\det L_t = \exp\left(\int_0^t \operatorname{tr} C_s^N \, ds\right) \neq 0$$

for every finite t, so $\Sigma_t^N = L_t \widehat{\Sigma}_0 L_t^\top \succ 0$. Since both the global q-flow and its linear lift exist on every finite interval, this proves all claims, with $(\mu_t^N, \Sigma_t^N) = (\widehat{\mu}_t, \widehat{\Sigma}_t)$. □

Henceforth write

$$q_{N,t} = (\widehat{\mu}_t, \widehat{\Sigma}_t^{1/2}) \tag{6.5}$$

for this empirical-moment trajectory.

Write the initial particles as

$$X_i(0) = \mu_0 + \Sigma_0^{1/2}\xi_i, \qquad \xi_i \overset{\text{iid}}{\sim} \mathcal{N}(0, I_d), \tag{6.6}$$

and define

$$\bar{\xi} = \frac{1}{N}\sum_{i=1}^{N}\xi_i, \qquad S = \frac{1}{N}\sum_{i=1}^{N}(\xi_i - \bar{\xi})(\xi_i - \bar{\xi})^\top. \tag{6.7}$$

For $N \geq d+1$, $S \succ 0$ almost surely. Let

$$z_i = S^{-1/2}(\xi_i - \bar{\xi}), \qquad \eta_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{z_i}, \qquad \gamma_d = \mathcal{N}(0, I_d). \tag{6.8}$$

This is precisely (2.9) with the whitening $R_0 = S^{-1/2}\Sigma_0^{-1/2}$.

The following estimate relates the standardized shape to an ordinary Gaussian empirical measure without using inverse sample-covariance moments.

Lemma 6.2 (Standardized shape error)**.** *Let $N \geq d+1$, let $\xi_1, \dots, \xi_N$ be i.i.d. standard Gaussians, and define $\bar{\xi}, S, z_i, \eta_N$, and γ_d by* (6.7)*–*(6.8)*. Then*

$$\frac{1}{N}\sum_{i=1}^{N}|z_i - \xi_i|^2 = \left\| I - S^{1/2} \right\|_{\mathrm{F}}^2 + \left|\bar{\xi}\right|^2 \leq \|I - S\|_{\mathrm{F}}^2 + \left|\bar{\xi}\right|^2. \tag{6.9}$$

Moreover, there is $C_d < \infty$, depending only on d, such that

$$\mathbb{E}W_2^2(\eta_N, \gamma_d) \leq C_d r_d(N). \tag{6.10}$$

Proof. Since $N^{-1}\sum_i(\xi_i - \bar{\xi}) = 0$, the cross term vanishes in

$$z_i - \xi_i = (S^{-1/2} - I)(\xi_i - \bar{\xi}) - \bar{\xi}.$$

Using the definition of S gives

$$\frac{1}{N}\sum_i \left|(S^{-1/2} - I)(\xi_i - \bar{\xi})\right|^2 = \mathrm{tr}\{(S^{-1/2} - I)S(S^{-1/2} - I)\} = \left\| I - S^{1/2} \right\|_{\mathrm{F}}^2.$$

The scalar inequality $\left|1 - \sqrt{\lambda}\right| \leq |1 - \lambda|$ for $\lambda \geq 0$ proves (6.9).

Let $\nu_N = N^{-1}\sum_i \delta_{\xi_i}$. Coupling by index and using the squared triangle inequality yields

$$\mathbb{E}W_2^2(\eta_N, \gamma_d) \leq 2\mathbb{E}\{\|I - S\|_{\mathrm{F}}^2 + \left|\bar{\xi}\right|^2\} + 2\mathbb{E}W_2^2(\nu_N, \gamma_d).$$

Because NS is Wishart with $N-1$ degrees of freedom, the first expectation is explicit:

$$\mathbb{E}\left|\bar{\xi}\right|^2 = \frac{d}{N}, \qquad \mathbb{E}\|S - I\|_{\mathrm{F}}^2 = \frac{(N-1)d(d+1) + d}{N^2}. \tag{6.11}$$

For the remaining empirical-matching term, the $d = 1$ estimate is the $p = 2$ Gaussian bound of Bobkov and Ledoux [5, Corollary 6.14]; the $d = 2$ estimate is due to Ledoux [8, Theorem 1]; and for $d \geq 3$ the claimed rate follows from Ledoux and Zhu [9, Theorem 1.1, with $p = 2$]:

$$\mathbb{E}W_2^2(\nu_N, \gamma_d) \leq C_d r_d(N).$$

The finitely many sample sizes excluded by an asymptotic formulation of these estimates are absorbed by enlarging C_d. Since $N^{-1} \leq C_d r_d(N)$, this proves (6.10). □

To transfer the standardized discrepancy through the particle dynamics, we record a two-sided bound for every invertible affine map.

Lemma 6.3 (Two-sided affine lifting)**.** *For any full-rank initial cloud, use the whitening and standardized law in* (2.9)*. Then, for every* $t \geq 0$*,*

$$\begin{aligned}\lambda_{\min}(\widehat{\Sigma}_t) W_2^2(\eta_N, \gamma_d) &\leq W_2^2\Big(\zeta_{N,t}, \mathcal{N}(\widehat{\mu}_t, \widehat{\Sigma}_t)\Big) \\ &\leq \lambda_{\max}(\widehat{\Sigma}_t) W_2^2(\eta_N, \gamma_d).\end{aligned} \tag{6.12}$$

Proof. Set $B_t = L_t R_0^{-1}$. Then $B_t B_t^\top = L_t \widehat{\Sigma}_0 L_t^\top = \widehat{\Sigma}_t$, and (6.2) gives

$$\zeta_{N,t} = (x \mapsto \widehat{\mu}_t + B_t x)_\# \eta_N. \tag{6.13}$$

The same affine map pushes γ_d to $\mathcal{N}(\widehat{\mu}_t, \widehat{\Sigma}_t)$. Pushing forward any coupling of η_N and γ_d multiplies its quadratic cost by at most $\|B_t\|_{\mathrm{op}}^2 = \lambda_{\max}(\widehat{\Sigma}_t)$, proving the upper bound. Conversely, pull any coupling of the two image laws back through the invertible affine map and use $|B_t u| \geq \sigma_{\min}(B_t)\,|u|$. Taking the infimum over image couplings proves the lower bound because $\sigma_{\min}(B_t)^2 = \lambda_{\min}(\widehat{\Sigma}_t)$. □

Proof of Theorem 2.1. Lemma 6.1 gives a unique global particle solution and identifies its moments with the q-flow from $q_{N,0}$. Since both initial moment states lie in $\mathcal{K}_R$, Proposition 4.3 and Lemma 5.1 give

$$\sup_{t \geq 0} W_2^2\Big(\mathcal{N}(\widehat{\mu}_t, \widehat{\Sigma}_t), \rho_t\Big) \leq S_R^2 \left\| q_{N,0} - q_0 \right\|^2, \tag{6.14}$$

where we used the synchronous Gaussian coupling (5.9). Invariance of $\mathcal{K}_R$ gives $\lambda_{\max}(\widehat{\Sigma}_t) \leq b_R^2$. The upper bound in Lemma 6.3, followed by the squared triangle inequality, now proves (2.10). □

Proof of Proposition 2.6. The proof first uses exponential moment convergence to show that the affine coefficient is integrable. Its fundamental matrix then converges to an invertible limit, which identifies both the limiting empirical law and its two-sided shape floor.

Write

$$C(q) = I - \Gamma(\mu, A^2) A^2 - m(\mu, A^2) \mu^\top.$$

The coordinate map Φ has invertible differential. Since q_* is an interior minimizer, $\nabla \mathcal{E}(\mu_*, \Sigma_*) = 0$; hence

$$m(\mu_*, \Sigma_*) = 0, \qquad \Gamma(\mu_*, \Sigma_*) = \Sigma_*^{-1}, \qquad C(q_*) = 0. \tag{6.15}$$

The map C is Lipschitz on each $\mathcal{K}_R$. For m, this follows by synchronously coupling the Gaussians and using the β-Lipschitzness of ∇V. For Γ, boundedness of $\nabla^2 V$, Pinsker's inequality, and the Gaussian KL formula give, for $q, \widetilde{q} \in \mathcal{K}_R$,

$$\left\|\Gamma(\mu, A^2) - \Gamma(\widetilde{\mu}, \widetilde{A}^2)\right\|_{\mathrm{F}} \leq 2\sqrt{d}\,\beta\, d_{\mathrm{TV}}\Big(\mathcal{N}(\mu, A^2), \mathcal{N}(\widetilde{\mu}, \widetilde{A}^2)\Big) \leq C_R \left\| q - \widetilde{q} \right\|. \tag{6.16}$$

The last inequality holds because the Gaussian KL divergence is bounded by $C_R \|q-\widetilde{q}\|^2$ on the compact covariance box containing $\mathcal{K}_R$. Set $\lambda_R = \alpha p_R$. Proposition 4.3 and (6.15) therefore imply

$$\|C(q_{N,t})\|_{\mathrm{op}} \le C_R e^{-\lambda_R t}, \qquad \int_0^\infty \|C(q_{N,t})\|_{\mathrm{op}}\, dt < \infty. \tag{6.17}$$

The fundamental matrix from (6.1) consequently satisfies

$$\sup_{t\ge 0}\left(\|L_t\|_{\mathrm{op}} + \left\|L_t^{-1}\right\|_{\mathrm{op}}\right) < \infty.$$

Indeed, this follows by Gronwall for L_t and for $\frac{d}{dt}L_t^{-1} = -L_t^{-1}C(q_{N,t})$. Moreover,

$$L_\infty = I + \int_0^\infty C(q_{N,s})L_s\, ds$$

exists, is invertible, and $\|L_t - L_\infty\|_{\mathrm{F}} \le C_R e^{-\lambda_R t}$. Invertibility is explicit from

$$\sigma_{\min}(L_\infty) = \lim_{t\to\infty}\sigma_{\min}(L_t) \ge \left(\sup_{t\ge 0}\left\|L_t^{-1}\right\|_{\mathrm{op}}\right)^{-1} > 0.$$

Set $B_t = L_t R_0^{-1}$ and $B_\infty = L_\infty R_0^{-1}$. Because $q_{N,0}\in\mathcal{K}_R$,

$$\left\|R_0^{-1}\right\|_{\mathrm{op}}^2 = \lambda_{\max}(\widehat{\Sigma}_0) \le b_R^2,$$

so $\|B_t - B_\infty\|_{\mathrm{F}} \le C_R e^{-\lambda_R t}$. Passing to the limit in $B_t B_t^\top = \widehat{\Sigma}_t$ gives $B_\infty B_\infty^\top = \Sigma_*$. Since the standardized cloud has zero mean and identity covariance, coupling particles by index yields

$$W_2^2(\zeta_{N,t},\zeta_{N,\infty}) \le |\widehat{\mu}_t - \mu_*|^2 + \frac{1}{N}\sum_{i=1}^N |(B_t - B_\infty)z_i|^2 = |\widehat{\mu}_t-\mu_*|^2 + \|B_t - B_\infty\|_{\mathrm{F}}^2,$$

which proves (2.18). Metric continuity and $\rho_t \to \rho_*$ prove (2.19); applying the two-sided lifting argument to B_∞ proves (2.20). Finally, η_N is finitely atomic and γ_d is non-atomic, so their Wasserstein distance is strictly positive. Equivalently, $O_\infty = \Sigma_*^{-1/2}B_\infty$ is orthogonal: the dynamics optimizes the Gaussian moments but can only rotate, not Gaussianize, the standardized initial shape. If $R_0' = OR_0$ is another whitening, with O orthogonal, then $\eta_N' = O_\#\eta_N$ and $B_\infty' = B_\infty O^\top$. Hence the pushforward in (2.17), and therefore the limiting empirical law, is independent of the chosen whitening coordinates. □

Proof of Corollary 2.7. Let $P_N = N^{-1}\sum_i \delta_{X_i(0)}$ and use the principal empirical whitening in (2.9). The empirical Wasserstein law of large numbers gives $W_2(P_N,P_0)\to 0$ almost surely; the sample mean and covariance converge almost surely as well. Hence $\widehat{\Sigma}_0^{-1/2}\to\Sigma_0^{-1/2}$. Coupling corresponding sample points and then applying the fixed limiting affine map shows

$$W_2(\eta_N,\eta_0) \longrightarrow 0 \qquad \text{almost surely.}$$

On the same probability-one event, every $N \ge d+1$ cloud is full rank and its moment state has finite energy, hence belongs to some sublevel $\mathcal{K}_R$. The two-sided comparison below is independent of that cloud-dependent sublevel. Explicitly, the squared cost of replacing the empirical centering and whitening by their population values tends to zero by the sample second-moment law of large numbers, while the remaining term is at most $\left\|\Sigma_0^{-1/2}\right\|_{\mathrm{op}} W_2(P_N,P_0)$. Taking the lower limit in the lower bound and the upper limit in the upper bound of (2.20) proves (2.21). Here $D_0>0$ because $\eta_0=\gamma_d$ would imply that P_0 itself is Gaussian. □

7 Localization and proof of the main theorem

This section upgrades the deterministic oracle to the expected and high-probability theorems. A fixed well-conditioned event supplies a common energy sublevel, while a log-Wishart estimate controls its rare complement.

Fix $\varepsilon_0 = 1/4$ and define the good event

$$G_N = \{|\bar{\xi}| \leq \varepsilon_0,\ \|S - I\|_{\mathrm{op}} \leq \varepsilon_0\}. \tag{7.1}$$

Lemma 7.1 (Good-event control)**.** *There are constants $C_d, c_d > 0$ such that, for all $N \geq d+1$,*

$$\mathbb{P}(G_N^c) \leq C_d e^{-c_d N}. \tag{7.2}$$

There is a deterministic $R < \infty$, depending only on the fixed problem data, such that on G_N both $q_0 = (\mu_0, \Sigma_0^{1/2})$ and $q_{N,0} = (\widehat{\mu}_0, \widehat{\Sigma}_0^{1/2})$ belong to $\mathcal{K}_R$. Furthermore, there is $C < \infty$, depending only on d, μ_0, Σ_0, such that

$$\mathbb{E}\Big[\mathbf{1}_{G_N} \|q_{N,0} - q_0\|^2\Big] \leq \frac{C}{N}. \tag{7.3}$$

Proof. Since $\sqrt{N}\,\bar{\xi} \sim \mathcal{N}(0, I_d)$, the Gaussian mean has an exponential tail at the fixed threshold ε_0. Furthermore,

$$S - I = \frac{1}{N}\sum_{i=1}^{N}(\xi_i \xi_i^\top - I) - \bar{\xi}\bar{\xi}^\top.$$

Thus

$$\mathbb{P}(\|S - I\|_{\mathrm{op}} > \varepsilon_0) \leq \mathbb{P}\left(\left\|\frac{1}{N}\sum_{i=1}^{N}(\xi_i \xi_i^\top - I)\right\|_{\mathrm{op}} > \frac{\varepsilon_0}{2}\right) + \mathbb{P}(|\bar{\xi}|^2 > \varepsilon_0/2).$$

For each fixed pair of coordinates (j,k), the centered variable $\xi_{ij}\xi_{ik} - \delta_{jk}$ has a subexponential norm bounded by a constant independent of i. Scalar Bernstein's inequality, a union bound over the d^2 entries, and fixed-dimensional norm equivalence bound the first probability by $C_d e^{-c_d N}$; a Gaussian tail gives the same bound for the second. Together with the mean event in (7.1), and after enlarging C_d for the finitely many small sample sizes, this proves (7.2).

On G_N,

$$\widehat{\mu}_0 = \mu_0 + \Sigma_0^{1/2}\bar{\xi}, \qquad \widehat{\Sigma}_0 = \Sigma_0^{1/2} S \Sigma_0^{1/2},$$

and

$$(1 - \varepsilon_0)\lambda_{\min}(\Sigma_0) I \preceq \widehat{\Sigma}_0 \preceq (1 + \varepsilon_0)\lambda_{\max}(\Sigma_0) I.$$

Thus $q_{N,0}$ ranges over a fixed compact subset of $\mathcal{Q}$; continuity of F places it, together with q_0, in a deterministic sublevel $\mathcal{K}_R$.

For positive definite C, D, the identity

$$C - D = C^{1/2}X + XD^{1/2}, \qquad X = C^{1/2} - D^{1/2},$$

and the Frobenius pairing give

$$\langle C - D, X\rangle_{\mathrm{F}} \geq \left(\sqrt{\lambda_{\min}(C)} + \sqrt{\lambda_{\min}(D)}\right)\|X\|_{\mathrm{F}}^2.$$

Cauchy–Schwarz therefore implies

$$\left\|C^{1/2} - D^{1/2}\right\|_{\mathrm{F}} \le \frac{\|C - D\|_{\mathrm{F}}}{\sqrt{\lambda_{\min}(C)} + \sqrt{\lambda_{\min}(D)}}. \tag{7.4}$$

On G_N the denominator has a deterministic positive lower bound. Since

$$|\widehat{\mu}_0 - \mu_0| \le \left\|\Sigma_0^{1/2}\right\|_{\mathrm{op}} |\bar{\xi}|,$$
$$\left\|\widehat{\Sigma}_0 - \Sigma_0\right\|_{\mathrm{F}} \le \|\Sigma_0\|_{\mathrm{op}} \|S - I\|_{\mathrm{F}},$$

the square-root bound (7.4) and the exact moments in (6.11) prove (7.3). □

Proof of Theorem 2.4. We intersect concentration events for the empirical measure, mean, and covariance. These events place all empirical initial states in one deterministic sublevel, so the oracle applies with constants independent of N and x.

Let $\nu_N = N^{-1}\sum_i \delta_{\xi_i}$. Matching corresponding atoms shows that the map

$$(\xi_1, \ldots, \xi_N) \longmapsto W_2(\nu_N, \gamma_d)$$

is $N^{-1/2}$-Lipschitz on $\mathbb{R}^{dN}$. Gaussian concentration and $(\mathbb{E}W_2)^2 \le \mathbb{E}W_2^2 \le C_d r_d(N)$ therefore give

$$\mathbb{P}\left\{W_2^2(\nu_N, \gamma_d) > C\left(r_d(N) + \frac{x}{N}\right)\right\} \le e^{-x}, \qquad x \ge 1. \tag{7.5}$$

Indeed, outside an event of probability e^{-x}, $W_2(\nu_N, \gamma_d) \le \mathbb{E}W_2(\nu_N, \gamma_d) + \sqrt{2x/N}$, and the squared triangle inequality gives (7.5).

Fixed-dimensional Gaussian concentration for $\bar{\xi}$ and scalar Bernstein concentration applied to every entry of $N^{-1}\sum_i(\xi_i\xi_i^\top - I)$ imply that, for $1 \le x \le c_0 N$, outside an event of probability Ce^{-cx},

$$|\bar{\xi}|^2 + \|S - I\|_{\mathrm{F}}^2 \le C\frac{1+x}{N}, \qquad \|S - I\|_{\mathrm{op}} \le \frac{1}{2}. \tag{7.6}$$

Here $c_0 > 0$ is reduced, if needed, so that the second assertion follows from the usual Bernstein scale $C_d\{\sqrt{x/N} + x/N\}$; indeed, $x/N \le c_0$ throughout the stated range. On the intersection of the good events underlying (7.5) and (7.6), the square-root estimate (7.4) gives

$$\|q_{N,0} - q_0\|^2 \le C\frac{1+x}{N}, \qquad W_2^2(\eta_N, \gamma_d) \le C\left(r_d(N) + \frac{x}{N}\right). \tag{7.7}$$

The second bound follows from the residual identity and coupling the atoms z_i with ξ_i. Moreover, all $q_{N,0}$ on these events, together with the fixed q_0, lie in one deterministic energy sublevel $\mathcal{K}_{R_{\mathrm{hp}}}$: the covariance eigenvalues stay in a fixed compact interval, and the sample means stay in a fixed ball because $x \le c_0 N$. Here R_{hp} depends only on the fixed problem data and the chosen c_0, not on N or x. Denote this intersection by $H_N(x)$; the preceding bounds give $\mathbb{P}(H_N(x)^c) \le Ce^{-cx}$.

The deterministic oracle (2.10), the fact that $r_d(N) \ge N^{-1}$, and a union bound prove (2.14). For the last iterate, Proposition 4.3 and the synchronous Gaussian coupling give the deterministic estimate

$$W_2^2(\rho_t, \rho_*) \le \frac{2\{F(q_0) - F(q_*)\}}{\alpha} e^{-2\alpha p_{\overline{R}_{\mathrm{mf}}} t}, \qquad \overline{R}_{\mathrm{mf}} = 1 + F(q_0) - F(q_*). \tag{7.8}$$

Combining this with (2.14) by the squared triangle inequality proves (2.15). Finally take $x = c^{-1}\log(C/\delta)$ and rename the fixed constants to obtain the confidence-level formulation. □

On the good event, Lemmas 5.1 and 7.1, together with (5.9), imply

$$\mathbb{E}\left[\mathbf{1}_{G_N} \sup_{t\geq 0} W_2^2\Big(\mathcal{N}(\widehat{\mu}_t, \widehat{\Sigma}_t), \rho_t\Big)\right] \leq \frac{C}{N}. \tag{7.9}$$

Invariance of $\mathcal{K}_R$ also gives a deterministic upper bound for $\sup_t \lambda_{\max}(\widehat{\Sigma}_t)$ on G_N. Hence Lemmas 6.2 and 6.3 yield

$$\mathbb{E}\left[\mathbf{1}_{G_N} \sup_{t\geq 0} W_2^2\Big(\zeta_{N,t}, \mathcal{N}(\widehat{\mu}_t, \widehat{\Sigma}_t)\Big)\right] \leq C r_d(N). \tag{7.10}$$

The complement of G_N cannot simply be discarded: there the sample covariance can be arbitrarily close to singular. The following lemma is the required weighted tail estimate.

Lemma 7.2 (Weighted control outside the good event)**.** *There are constants $C, c > 0$, depending only on d, V, μ_0, Σ_0, such that*

$$\mathbb{E}\left[\mathbf{1}_{G_N^c} \sup_{t\geq 0} W_2^2(\zeta_{N,t}, \rho_t)\right] \leq Ce^{-cN}, \qquad N \geq d+1. \tag{7.11}$$

Proof. We first dominate the entire path by the initial energy, reduce that energy to polynomial moments and a log determinant, and then use a uniform log-Wishart second moment with Cauchy–Schwarz on G_N^c.

For a probability law ν, write $M_2(\nu) = \int |x|^2 \, d\nu(x)$. The independent coupling through the origin gives

$$W_2^2(\nu, \widetilde{\nu}) \leq 2M_2(\nu) + 2M_2(\widetilde{\nu}). \tag{7.12}$$

For the particle cloud,

$$M_2(\zeta_{N,t}) = |\widehat{\mu}_t|^2 + \operatorname{tr} \widehat{\Sigma}_t = \|q_{N,t}\|^2 .$$

Strong convexity of F and energy decrease along the random moment flow give

$$\sup_{t\geq 0} M_2(\zeta_{N,t}) \leq C\{1 + F(q_{N,0}) - F(q_*)\}. \tag{7.13}$$

The upper Hessian bound on V gives

$$F(q_{N,0}) - F(q_*) \leq C\{1 + |\widehat{\mu}_0|^2 + \operatorname{tr} \widehat{\Sigma}_0 + (-\log\det \widehat{\Sigma}_0)_+\}. \tag{7.14}$$

Indeed, Taylor's theorem and the upper Hessian bound imply $V(x) \leq C(1 + |x|^2)$, so

$$\mathbb{E}V(\widehat{\mu}_0 + \widehat{\Sigma}_0^{1/2} Z) \leq C\{1 + |\widehat{\mu}_0|^2 + \operatorname{tr} \widehat{\Sigma}_0\};$$

the entropy term is $-\frac{1}{2}\log\det \widehat{\Sigma}_0$, whose positive contribution is exactly one half of $(-\log\det \widehat{\Sigma}_0)_+$. Define

$$Z_N = 1 + |\widehat{\mu}_0|^2 + \operatorname{tr} \widehat{\Sigma}_0 + (-\log\det \widehat{\Sigma}_0)_+. \tag{7.15}$$

All log-determinant quantities here and below are evaluated on the almost-sure full-rank event; on its null complement they inherit the fixed extension specified after (2.5).

We next prove

$$\sup_{N\geq d+1} \mathbb{E}Z_N^2 < \infty. \tag{7.16}$$

The polynomial terms have uniform moments of every fixed order: the sample mean is Gaussian with uniformly bounded covariance, while $\operatorname{tr}\widehat{\Sigma}_0$ is bounded above by a fixed multiple of $N^{-1}\sum_i |\xi_i|^2$. For the remaining term, Bartlett's decomposition [1, Lemma 7.2.1] gives

$$NS \sim \mathrm{Wishart}_d(N-1, I), \qquad \det(NS) \overset{d}{=} \prod_{j=1}^{d} Y_{N-j}, \quad Y_k \sim \chi_k^2 \text{ independently.} \tag{7.17}$$

For s in a neighborhood of zero,

$$\mathbb{E}Y_k^s = 2^s \frac{\Gamma(k/2+s)}{\Gamma(k/2)}.$$

Differentiating the cumulant generating function of $\log Y_k$ at zero gives the digamma and trigamma identities

$$\mathbb{E}\log Y_k = \psi(k/2) + \log 2, \qquad \mathrm{Var}(\log Y_k) = \psi_1(k/2)$$

and hence the explicit formula

$$\mathbb{E}\,|\log(Y_k/k)|^2 = \psi_1(k/2) + \{\psi(k/2) + \log 2 - \log k\}^2.$$

Here $\psi_1(k/2) \le \psi_1(1/2)$, and the second term is uniformly bounded by the standard digamma asymptotic $\psi(x) = \log x + O(x^{-1})$, together with the finitely many bounded small values of k. Hence

$$\sup_{k\ge 1} \mathbb{E}\,|\log(Y_k/k)|^2 < \infty. \tag{7.18}$$

Now

$$\log\det S = \sum_{j=1}^{d} \left\{ \log \frac{Y_{N-j}}{N-j} + \log \frac{N-j}{N} \right\}.$$

For fixed d and $N \ge d+1$, the deterministic logarithms are uniformly bounded. Hence (7.18) and $(\sum_{j=1}^d a_j)^2 \le d\sum_{j=1}^d a_j^2$ show that

$$\sup_{N\ge d+1} \mathbb{E}[(-\log\det S)_+^2] < \infty. \tag{7.19}$$

Since $\det\widehat{\Sigma}_0 = \det(\Sigma_0)\det S$, the asserted uniform second-moment bound (7.16) follows.

Proposition 4.3 gives a deterministic uniform bound on $M_2(\rho_t)$. Combining (7.12)–(7.14) therefore gives the pathwise estimate

$$\sup_{t\ge 0} W_2^2(\zeta_{N,t}, \rho_t) \le C Z_N. \tag{7.20}$$

Cauchy–Schwarz, (7.16), and (7.2) now yield

$$\mathbb{E}\left[\mathbf{1}_{G_N^c} \sup_{t\ge 0} W_2^2(\zeta_{N,t}, \rho_t)\right] \le C\mathbb{P}(G_N^c)^{1/2}(\mathbb{E}Z_N^2)^{1/2} \le Ce^{-cN/2}.$$

Renaming $c/2$ proves (7.11). □

The algebraic endpoint is independent of the dynamics and follows from an affine-span characterization.

Proposition 7.3 (Exact full-rank endpoint)**.** *For any N points in $\mathbb{R}^d$, the centered empirical covariance has rank at most $N-1$ and is positive definite if and only if the points affinely span $\mathbb{R}^d$. Consequently $N \ge d+1$ is necessary for the unregularized positive-definite dynamics, and it is sufficient almost surely for i.i.d. initialization from any distribution having a density.*

Proof. If $Y = [x_1 - \bar{x}, \dots, x_N - \bar{x}]$, then $Y\mathbf{1}_N = 0$ and $\widehat{\Sigma} = N^{-1}YY^\top$. Thus $\operatorname{rank}(\widehat{\Sigma}) = \operatorname{rank}(Y) \leq N-1$, and rank d is equivalent to affine spanning. Under a density, affine dependence of any fixed $d+1$ samples is a Lebesgue-null event. $\square$

Remark 7.4 (Logarithmic control at the endpoint). At $N = d+1$, the final Bartlett factor in (7.17) has one degree of freedom. The covariance is positive definite almost surely, but $\mathbb{E}\operatorname{tr}(S^{-1}) = \infty$, whereas $\mathbb{E}[(\log \chi_1^2)^2] < \infty$. This is why logarithmic-energy localization reaches the exact algebraic endpoint.

Proof of Theorem 2.2. On G_N, the squared triangle inequality with the intermediate Gaussian $\mathcal{N}(\widehat{\mu}_t, \widehat{\Sigma}_t)$, followed by (7.9) and (7.10), gives

$$\mathbb{E}\left[\mathbf{1}_{G_N} \sup_{t\geq 0} W_2^2(\zeta_{N,t}, \rho_t)\right] \leq C\{N^{-1} + r_d(N)\} \leq C r_d(N).$$

Lemma 7.2 controls G_N^c. Since $e^{-cN} \leq C r_d(N)$ for every fixed d, the two event contributions prove (2.12). Almost-sure positive definiteness at time zero follows from Proposition 7.3; global existence follows from Lemma 6.1 and Proposition 4.3. $\square$

Proof of Corollary 2.5. Apply Proposition 4.3 to the deterministic initial condition. Since q_* represents ρ_*, (4.14) and (5.9) give

$$W_2^2(\rho_t, \rho_*) \leq C e^{-2ct}$$

for a constant $c > 0$. The squared triangle inequality and Theorem 2.2 prove (2.16). The statistical-floor statement follows by balancing the two terms. $\square$

8 Why a global Euclidean stability constant is impossible

The stability constant in Lemma 5.1 necessarily depends on an energy sublevel. The next proposition rules out a uniform replacement.

Proposition 8.1 (No global Euclidean Lipschitz constant for the scalar flow)**.** *In dimension $d = 1$, there is a smooth, strongly convex potential with globally bounded Hessian for which no finite S can satisfy*

$$\sup_{t\geq 0} |\phi_t(r_0) - \phi_t(\widetilde{r}_0)| \leq S\,|r_0 - \widetilde{r}_0|\,, \qquad r_0, \widetilde{r}_0 > 0,$$

where $\phi_t(r_0)$ is the standard-deviation component of the moment flow started from $(\mu_0, \Sigma_0) = (0, r_0^2)$.

Proof. Fix $\kappa > a > 0$ and $k > 0$, and take

$$V(x) = \frac{\kappa}{2}x^2 + \frac{a}{k^2}\cos(kx). \tag{8.1}$$

Then $\kappa - a \leq V'' \leq \kappa + a$. The subspace $\mu = 0$ is invariant. Writing $r = \sqrt{\Sigma}$ on this subspace gives

$$m = 0, \qquad \Gamma(r) = \kappa - ae^{-k^2r^2/2}, \qquad \dot{r} = g(r) := r\{1 - r^2\Gamma(r)\}. \tag{8.2}$$

In fact,

$$\frac{d}{dr}\{r^2\Gamma(r)\} = 2r\{\kappa - ae^{-k^2r^2/2}\} + ak^2r^3e^{-k^2r^2/2} > 0,$$

because $\kappa > a$. Thus $r^2\Gamma(r)$ is strictly increasing from zero to infinity and has a unique crossing $r_* > 0$ of level one. In particular, $g > 0$ on $(0, r_*)$. Fix $r_c \in (0, r_*)$. For $r_0 \in (0, r_c)$, let $t_{\mathrm{hit}}(r_0)$ be the time at which $\phi_t(r_0)$ first reaches r_c, and then freeze time at $t_0 = t_{\mathrm{hit}}(r_0)$. This time is finite because g is continuous and strictly positive on $[r_0, r_c]$. The variational identity for a scalar autonomous flow is

$$D_r\phi_t(r_0) = \frac{g(\phi_t(r_0))}{g(r_0)}.$$

Moreover, $g(r_0) = r_0(1 + o(1))$ as $r_0 \downarrow 0$. Therefore, with t_0 fixed while differentiating,

$$D_r\phi_{t_0}(r_0) = \frac{g(r_c)}{g(r_0)} \sim \frac{g(r_c)}{r_0} \longrightarrow \infty \qquad (r_0 \downarrow 0). \tag{8.3}$$

Any common Lipschitz constant for all fixed-time maps ϕ_t would bound the left-hand side of (8.3), a contradiction. If the full moment flow had one initialization-uniform Euclidean Lipschitz constant, its restriction to the invariant subspace $\mu = 0$ would have the same property, so the scalar contradiction also rules out that claim. Notice that we have not differentiated the composite $\phi_{t_{\mathrm{hit}}(r_0)}(r_0) = r_c$, whose derivative is zero. □

This obstruction does not contradict Lemma 5.1: a fixed energy sublevel stays a positive distance away from $r = 0$. It explains why near-singular Wishart samples must be controlled probabilistically through energy rather than by a global derivative bound for the flow.

9 An explicit nonquadratic example

This section verifies the assumptions and identifies the variational optimum for a concrete non-quadratic target. It illustrates that the theorem is not confined to Gaussian potentials.

Example 9.1 (Cosine perturbation of a Gaussian target). Let

$$V_\varepsilon(x) = \frac{1}{2}x^2 + \varepsilon(1 + \cos x), \qquad 0 < \varepsilon \le 0.1. \tag{9.1}$$

Then $1 - \varepsilon \le V_\varepsilon''(x) \le 1 + \varepsilon$, so Theorem 2.2 applies. If $s > 0$ denotes the variance of $\mathcal{N}(\mu, s)$, then

$$m(\mu, s) = \mu - \varepsilon e^{-s/2}\sin\mu, \qquad \Gamma(\mu, s) = 1 - \varepsilon e^{-s/2}\cos\mu. \tag{9.2}$$

The moment flow is

$$\dot\mu = (1 - \Gamma s)\mu - (1 + \mu^2)m, \qquad \dot s = 2s(1 - \Gamma s - \mu m). \tag{9.3}$$

Up to an additive constant, its reverse-KL energy is

$$E_\varepsilon(\mu, s) = \frac{1}{2}(\mu^2 + s - \log s) + \varepsilon e^{-s/2}\cos\mu, \tag{9.4}$$

and

$$\begin{pmatrix}\dot\mu \\ \dot s\end{pmatrix} = -\begin{pmatrix}1 + \mu^2 & 2s\mu \\ 2s\mu & 4s^2\end{pmatrix}\nabla E_\varepsilon(\mu, s). \tag{9.5}$$

The mobility determinant is $4s^2 > 0$. With $a_0 = e^{-s/2}$,

$$\nabla^2 E_\varepsilon = \begin{pmatrix}1 - \varepsilon a_0\cos\mu & (\varepsilon a_0/2)\sin\mu \\ (\varepsilon a_0/2)\sin\mu & 1/(2s^2) + (\varepsilon a_0/4)\cos\mu\end{pmatrix}. \tag{9.6}$$

The upper-left entry is at least $1-\varepsilon$, and its Schur complement is bounded below as follows:

$$\begin{aligned}
&\frac{1}{2s^2} + \frac{\varepsilon a_0}{4}\cos\mu - \frac{(\varepsilon a_0 \sin\mu/2)^2}{1-\varepsilon a_0 \cos\mu} \\
&\quad \geq \frac{1}{2s^2} - \frac{\varepsilon e^{-s/2}}{4} - \frac{\varepsilon^2 e^{-s}}{4(1-\varepsilon)} \\
&\quad \geq \frac{1}{2s^2}\left(1 - 8e^{-2}\varepsilon - \frac{2e^{-2}\varepsilon^2}{1-\varepsilon}\right) > 0.
\end{aligned} \tag{9.7}$$

Indeed, $s^2 e^{-s/2} \leq 16e^{-2}$ and $s^2 e^{-s} \leq 4e^{-2}$. Hence E_ε is strictly convex and coercive, since the perturbation in (9.4) is bounded while $\mu^2 + s - \log s$ diverges at every boundary of $\mathbb{R} \times (0,\infty)$. Furthermore,

$$\partial_\mu E_\varepsilon(0,s) = 0, \qquad \partial_s E_\varepsilon(0,s) = \frac{1}{2}\left(1 - \frac{1}{s} - \varepsilon e^{-s/2}\right).$$

Let $h(s) = s(1-\varepsilon e^{-s/2})$. Then

$$h'(s) = 1 - \varepsilon e^{-s/2} + \frac{\varepsilon s}{2} e^{-s/2} > 1-\varepsilon > 0, \qquad h(0+) = 0, \qquad h(\infty) = \infty.$$

Thus there is a unique $s_* > 0$ with $h(s_*) = 1$. The point $(0, s_*)$ is stationary and, by strict convexity, is the unique minimizer; equivalently,

$$s_*\{1 - \varepsilon e^{-s_*/2}\} = 1. \tag{9.8}$$

This gives a fully worked, genuinely nonquadratic target with an explicit scalar characterization of the optimum.

10 Related work and positioning

This section distinguishes the projected exact-moment problem from ordinary SVGD and from time-averaged propagation of chaos. It also identifies the empirical-matching results used only as statistical inputs.

SVGD was introduced by Liu and Wang [11], and its gradient-flow and mean-field viewpoints were developed by Liu [10], Lu et al. [13], and Duncan et al. [6]. Nonasymptotic and finite-particle analyses include Korba et al. [7] and Shi and Mackey [14]. These works concern ordinary SVGD and do not themselves give uniform physical-time W_2 stability for the projected system (2.5).

The closest starting point is Liu et al. [12], whose exact-moment closure and affine representation lead to the prior conjecture discussed above. Banerjee et al. [3] prove finite-particle KSD and W_2 bounds for ordinary SVGD, including bilinear-plus-Matérn kernels, with long-time statements for time-averaged particle laws. The broad KSD and W_1 conclusions of Balasubramanian et al. [2], and their in-probability W_2 conclusion, concern time-averaged laws uniformly over deterministic averaging horizons, rather than physical-time last iterates; their finite-feature results have a different scope. Banerjee and Kim [4] obtain physical-time results for unprojected, Langevin-regularized SVGD with Brownian noise, whereas our system is deterministic and uses an exact Gaussian-score projection. Stein and Li [15] analyze generalized bilinear kernels for Gaussian targets. None of these results directly yields the fixed-K_1, non-Gaussian, projected full-law theorem here.

The dimension-dependent statistical input comes from Gaussian empirical matching [5, 8, 9, 16]. Our contribution is not a new matching estimate: it shows that the nonlinear all-time dynamics adds no particle-number penalty to the sharp expected time-zero scale.

11 Implications and limitations

The deterministic oracle applies to each fixed full-rank cloud, with constants determined by a sublevel containing both initial moment states. A sequence of deterministic, dependent, or variance-reduced designs inherits a uniform guarantee when it lies in one common sublevel and its moment and standardized-shape errors vanish. Gaussian independence is used only to obtain the sharp expected N-rate and the stated high-probability upper bound.

At fixed N, the affine dynamics converges to a discrete law with moments (μ_*, Σ_*) but preserves standardized shape up to rotation. Thus moment convergence must not be confused with Gaussianization. The nonquadratic target in Section 9 shows that the theory is not limited to Gaussian potentials.

The restrictions are substantive. Constants depend on d, V, μ_0, Σ_0 and deteriorate near the positive-definite boundary. Section 8 proves, for a smooth scalar potential with bounded Hessian, that the Euclidean standard-deviation flow has no single initialization-uniform Lipschitz constant over $(0, \infty)$. This does not exclude other metrics or weighted stability mechanisms. No dimension-free, Euler-discretization, or estimated-moment claim is made. Finally, ρ_* is the reverse-KL Gaussian approximation, not the non-Gaussian target itself, so approximation bias is separate from particle error.

12 Conclusion

Square-root geometry and affine whitening turn the non-Gaussian projected Gaussian-SVGD problem into a stable moment flow plus a preserved shape term. This yields a deterministic all-time oracle, a sharp expected N-rate, a high-probability upper bound at the corresponding scale, validity at the minimal count $N = d + 1$, and an exact description of the finite-particle limit. The result resolves the strengthened continuous exact-moment conjecture under global strong convexity and a bounded continuous Hessian, without a third-derivative or Hessian-Lipschitz assumption.

References

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis.* Third edition, Wiley-Interscience, Hoboken, NJ, 2003.

[2] K. Balasubramanian, S. Banerjee, and A. Korba. Uniform-in-time propagation-of-chaos for Stein variational gradient descent. arXiv:2607.00149, 2026.

[3] S. Banerjee, K. Balasubramanian, and P. Ghosal. Improved finite-particle convergence rates for Stein variational gradient descent. In *International Conference on Learning Representations*, 2025.

[4] S. Banerjee and D. Kim. Quantitative target convergence and uniform-in-time propagation of chaos for Langevin-regularized SVGD. arXiv:2608.28827, 2026.

[5] S. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Memoirs of the American Mathematical Society*, 261(1259), 2019. doi:10.1090/memo/1259.

[6] A. Duncan, N. Nüsken, and L. Szpruch. On the geometry of Stein variational gradient descent. *Journal of Machine Learning Research*, 24(56):1–39, 2023.

[7] A. Korba, A. Salim, M. Arbel, G. Luise, and A. Gretton. A non-asymptotic analysis for Stein variational gradient descent. In *Advances in Neural Information Processing Systems*, volume 33, pages 4672–4682, 2020.

[8] M. Ledoux. On optimal matching of Gaussian samples. *Journal of Mathematical Sciences*, 238(4):495–522, 2019. doi:10.1007/s10958-019-04253-6.

[9] M. Ledoux and J.-X. Zhu. On optimal matching of Gaussian samples III. *Probability and Mathematical Statistics*, 41(2):237–265, 2021. doi:10.37190/0208-4147.41.2.3.

[10] Q. Liu. Stein variational gradient descent as gradient flow. In *Advances in Neural Information Processing Systems*, volume 30, pages 3115–3123, 2017.

[11] Q. Liu and D. Wang. Stein variational gradient descent: A general purpose Bayesian inference algorithm. In *Advances in Neural Information Processing Systems*, volume 29, pages 2378–2386, 2016.

[12] T. Liu, P. Ghosal, K. Balasubramanian, and N. S. Pillai. Towards understanding the dynamics of Gaussian–Stein variational gradient descent. In *Advances in Neural Information Processing Systems*, volume 36, pages 61234–61291, 2023. doi:10.52202/075280-2676; arXiv:2305.14076.

[13] J. Lu, Y. Lu, and J. Nolen. Scaling limit of the Stein variational gradient descent: The mean field regime. *SIAM Journal on Mathematical Analysis*, 51(2):648–671, 2019. doi:10.1137/18M1187611.

[14] J. Shi and L. W. Mackey. A finite-particle convergence rate for Stein variational gradient descent. In *Advances in Neural Information Processing Systems*, volume 36, pages 26831–26844, 2023. doi:10.52202/075280-1166.

[15] V. Stein and W. Li. Towards understanding accelerated Stein variational gradient flow: Analysis of generalized bilinear kernels for Gaussian target distributions. arXiv:2509.04008, version 2, 2025.

[16] M. Talagrand. Scaling and non-standard matching theorems. *Comptes Rendus Mathématique*, 356(6):692–695, 2018. doi:10.1016/j.crma.2018.04.018.