CERTIFIED FROZEN-MIDPOINT NEURAL INTEGRATORS: ORIGINAL-ENERGY DISSIPATION AND SECOND-ORDER PHYSICAL TIME

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural gradient-flow solvers can decrease the correct energy while evolving on an incorrect physical clock. With state-dependent mobility, even an exactly solved, energy-dissipative step can be only first order. We develop a certified frozen-midpoint interface that separates three obligations: mobility consistency, full-state nonlinear accuracy, and signed energy work. A predicted midpoint freezes the mobility before optimization, preserving a globally strongly convex average-vector-field subproblem whenever its stated curvature and existence conditions hold. Deterministic oracle envelopes give computable objective-gap and endpoint-radius certificates, without assuming global neural training convergence. We prove variable-step second-order accuracy, finite correction for nonstationary states under explicit solver assumptions, and sharp worst-case tolerance scalings. Applications include coupled positive quantile diffusions and full-covariance Gaussian Wasserstein flows with noncommuting matrices. For smooth quantile reconstructions, a separate posterior theorem certifies error against the continuum evolution, including an explicit spatial defect. We also characterize validation under adaptive sampling and the obstruction at exact equilibrium. This is a purely theoretical study: it establishes conditional guarantees and computability constructions, not empirical speedups, unrestricted-generator convergence, or priority for discrete-gradient energy preservation itself.

1 INTRODUCTION

The physical time of a dissipative evolution is not an optimization-iteration counter. For example, both $\dot{x} = -x$ and $\dot{x} = -2x$ decrease $x^2/2$, but they predict different states at every positive prescribed time. This distinction matters when a learned predictor approximately solves an implicit time step or proposes its mobility. The relevant question is not whether a training loss decreases, but which verifiable conditions on an accepted state imply both the correct dissipation law and the intended physical-time accuracy.

The JKO construction supplies a variational description of Wasserstein evolution (Jordan et al., 1998), and neural implementations approximate its transport or flow maps (Mokrov et al., 2021; Xu et al., 2023). However, first-order minimizing movements, higher-order temporal consistency, and certification of an inexact solve are distinct matters. Our starting point is the classical average-vector-field (AVF) discrete gradient (Celledoni et al., 2012). The mathematical contribution is a coordinated certificate interface for *variable mobility, inexact nonlinear solves, and inexact oracles*, together with models where the required global subproblem statements can actually be proved.

Relation to existing methods. Exact discrete-gradient energy preservation and its order theory are established (Ehrhardt et al., 2025; Eidnes, 2022). High-order original-energy methods also exist (Huang, 2025; Fu et al., 2024; Lautwein et al., 2026). Energetic variational neural discretizations already combine neural spatial approximation and dissipation (Hu et al., 2024; Georgoulis et al., 2023). In particular, UEPI learns discrete-gradient integrators and obtains second order under endpoint symmetry (Celledoni et al., 2025); its Section 4 explicitly makes that distinction. We therefore

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

do not claim that "neural, dissipative, and second order" is new. Our focus is a different theorem: an *imperfect proposal and imperfect oracle* can be accepted through a computable global gap and a separate signed-work test, while a certified midpoint fixes the physical clock. Related inexact JKO analyses (Di Marino et al., 2025), higher-order Wasserstein constructions (Carrillo et al., 2022; Matthes & Plazotta, 2019; Gallouët et al., 2024; Chu & Jacobs, 2025), and reconstruction estimates (Makridakis & Nochetto, 2006) are complementary, not superseded. The appendix gives the precise scope of these comparisons.

2 A FROZEN-MIDPOINT VARIATIONAL STEP

Let H be a finite-dimensional real Hilbert space with a *fixed* inner product, $\mathcal{D} \subset \mathsf{H}$ open and convex, and

$$\dot{u} = F(u) := -K(u)\nabla E(u), \qquad E \in C^3(\mathcal{D}), \quad K \in C^2(\mathcal{D}; \mathcal{L}(\mathsf{H})). \tag{1}$$

Here $K(u)$ is self-adjoint positive definite. All gradients, adjoints, and operator bounds refer to the fixed inner product, not to a silently changing parameter metric. On a specified compact corridor assume $k_-\mathrm{I} \preceq K \preceq k_+\mathrm{I}$ and bounded derivatives. A predictor produces a midpoint z, an endpoint y, and a frozen self-adjoint matrix $B \succ 0$. Its clock certificates are

$$\zeta := \left\| z - x - \tfrac{h}{2}F(x) \right\| \le c_z h^2, \qquad \chi_K := \|B - K(z)\|_{\mathrm{op}} \le c_K h^2. \tag{2}$$

A fully explicit fallback takes $z = x + hF(x)/2$ and $B = K(z)$, provided $z \in \mathcal{D}$. We freeze B *before* optimizing y. It need not equal $K((x+y)/2)$.

For $\delta = y - x$, define

$$g(x,y) = \int_0^1 \nabla E(x + s\delta)\,\mathrm{d}s, \quad r = \delta/h + Bg(x,y), \quad \varepsilon = \|r\|_{B^{-1}}. \tag{3}$$

The exact frozen step has $r = 0$. Its objective is

$$Q^B_{h,x}(y) = \frac{\|y-x\|^2_{B^{-1}}}{2h} + \int_0^1 \frac{E(x + s(y-x)) - E(x)}{s}\,\mathrm{d}s. \tag{4}$$

The integrand extends continuously to $s = 0$. This is an AVF potential, *not* the original JKO objective. If $\nabla^2 E(v) \succeq \lambda_B B^{-1}$ throughout $\mathcal{D}$, then

$$\nabla Q^B_{h,x} = B^{-1}r, \qquad \nabla^2 Q^B_{h,x} \succeq \beta_h B^{-1}, \qquad \beta_h = 1/h + \lambda_B/2. \tag{5}$$

For a base-metric bound $\nabla^2 E \succeq -\ell\mathrm{I}$, the choice $\lambda_B = -\ell\,\|B\|_{\mathrm{op}}$ is valid. Thus small steps can make the subproblem strongly convex even when E is nonconvex. Strong convexity alone does *not* imply an interior minimizer on an open domain; the models below supply the missing existence arguments.

Theorem 2.1 (Global step certificate)**.** *Assume $\beta_h > 0$ and $Q^B_{h,x}$ has an interior minimizer $y^\star$. Then it is the unique global minimizer on $\mathcal{D}$, and every $y \in \mathcal{D}$ satisfies*

$$Q^B_{h,x}(y) - Q^B_{h,x}(y^\star) \le G_h(y) := \frac{\varepsilon^2}{2\beta_h}, \qquad \|y - y^\star\|_{B^{-1}} \le \frac{\varepsilon}{\beta_h}. \tag{6}$$

These are full-state certificates. They remain valid when y is produced by a nonconvex neural parameterization or corrected outside that parameterization.

Proof. Differentiate under the integral in (4); its Hessian contribution is $\int_0^1 s\nabla^2 E(x+s\delta)\,\mathrm{d}s$. Complete the square in the strong-convexity lower bound for $Q(v)$ at y to obtain the gap bound. Apply strong monotonicity to $\nabla Q(y) - \nabla Q(y^\star)$ and use dual Cauchy–Schwarz for the radius. Appendix B proves the differentiation, domain, and norm details. □

2.1 IMPERFECT ORACLES AND SIGNED ENERGY WORK

Suppose the AVF oracle returns $\widetilde{g}$ with a certified bound $\|g - \widetilde{g}\|_B \le \sigma$. Form

$$\widetilde{r} = \delta/h + B\widetilde{g}, \quad \bar{\varepsilon} = \|\widetilde{r}\|_{B^{-1}} + \sigma, \quad \widetilde{D} = \|\widetilde{g}\|^2_B, \quad \widetilde{\omega} = \langle \widetilde{g}, \widetilde{r}\rangle. \tag{7}$$

Algorithm 1 Certified frozen-midpoint neural step (CFM)

Require: Accepted x, proposed physical step h, budgets c_z, c_K, τ, η
1: Propose z, B, y; certify (2), $z, y \in \mathcal{D}$, and $B \succ 0$.
2: On midpoint failure, replace it by the explicit fallback or reject h.
3: **while** either gate in (8) is not certified **do**
4: Refine the oracle and/or correct y in the full state space using $Q^B_{h,x}$.
5: On exhausted work budget, report failure; do not advance the clock.
6: **end while**
7: Optionally certify a continuous reconstruction against a global time-error budget.
8: **return** y, gap and radius bounds, dissipation bound; advance by exactly h.

Our two endpoint gates are

$$\boxed{\bar{\varepsilon} \le \tau h^2}, \qquad \boxed{\widetilde{\omega} + \frac{\sigma}{h} \|\delta\|_{B^{-1}} \le \eta \widetilde{D}}, \qquad \tau > 0, \quad 0 < \eta < 1. \tag{8}$$

Every displayed upper bound used in an actual certificate must be validated, not merely estimated by unguarded floating point.

Theorem 2.2 (Robust dissipation)**.** *Under the oracle enclosure and* (8)*, the energy inequality below holds. If the hypotheses of Theorem 2.1 also hold, so does the gap inequality:*

$$E(y) - E(x) \le -(1-\eta) h \widetilde{D} \le 0, \quad Q(y) - \min Q \le \frac{\bar{\varepsilon}^2}{2\beta_h}. \tag{9}$$

For an exact oracle, the energy law is the identity $E(y) - E(x) = -h \|g\|_B^2 + h \langle g, r \rangle$.

Proof. The chain rule is $E(y) - E(x) = \langle g, \delta \rangle$. Substitute $\delta = -hB\widetilde{g} + h\widetilde{r}$ and bound $\langle g - \widetilde{g}, \delta \rangle \le \sigma \|\delta\|_{B^{-1}}$. Moreover, $\|r\|_{B^{-1}} \le \|\widetilde{r}\|_{B^{-1}} + \sigma$. These prove both claims. A sharper independently enclosed chain defect may replace $\sigma \|\delta\|_{B^{-1}}$ (Appendix C). □

3 Second order on the actual physical clock

The energy gate alone imposes no temporal order. The midpoint gate is also indispensable for a variable mobility. The following theorem does not assume that any Wasserstein resolvent is a contraction.

Assumption 3.1 (Uniform finite-time corridor)**.** For $t_n \le T$, the exact trajectory, accepted endpoints, connecting chords, and certified midpoint proposals lie in a fixed compact subset of $\mathcal{D}$ with a compact neighborhood in $\mathcal{D}$. On that neighborhood, E has bounded derivatives through order three, K through order two, and all frozen B_n satisfy uniform positive spectral bounds. These bounds are independent of the time mesh. They may depend on the spatial model.

Theorem 3.2 (Variable-step physical-time accuracy)**.** *Under Assumption 3.1, there are* $h_0, C_T > 0$ *such that every accepted sequence with* $0 < h_n \le h_0$, $t_{n+1} = t_n + h_n$, *and* $t_n \le T$ *satisfies*

$$\max_{t_n \le T} \|x_n - u(t_n)\| \le C_T \left(\|x_0 - u(0)\| + \sum_{t_j < T} h_j \big[h_j^2 + \bar{\varepsilon}_j + \zeta_j + \chi_{K,j} \big] \right). \tag{10}$$

Consequently (2) *and* (8) *imply second order in* $h_{\max}$ *for matching initial data. No adjacent-step-ratio assumption is required. If* $\beta_h \asymp h^{-1}$*, the sufficient certified gap budget is* $G_h = O(h^5)$.

Proof outline with the leading terms. Write $f = F(x)$ and $H = \nabla^2 E(x)$. Midpoint certification gives $B = K(x) + \frac{h}{2} DK(x)[f] + O(h^2 + \zeta + \chi_K)$. The AVF expansion is $g = \nabla E(x) + H\delta/2 + O(\|\delta\|^2)$. Since $\delta = -hBg + hr$ and the corridor bounds give $\delta = O(h)$,

$$\delta = hf + \frac{h^2}{2} \{ -DK(x)[f] \nabla E(x) - K(x) H f \} + O\big(h^3 + h[\bar{\varepsilon} + \zeta + \chi_K]\big).$$

The braces equal $F'(x)F(x)$, exactly the second Taylor coefficient of the physical flow. Compare with the exact one-step flow and iterate its Lipschitz stability; summing $h_j^3 \le T h_{\max}^2$ proves (10). Appendices D–E give explicit remainder bounds and the bootstrap for $\delta = O(h)$. □

Corollary 3.3 (Uniform physical time under contraction)**.** *If the corridor and local defect constants are uniform for all time and the actual target field satisfies $\langle F(u) - F(v), u - v\rangle \le -\alpha \|u - v\|^2$ with $\alpha > 0$, then*

$$\|x_n - u(t_n)\| \le e^{-\alpha t_n} \|x_0 - u(0)\| + \frac{C e^{\alpha h_{\max}}}{\alpha} h_{\max}^2. \tag{11}$$

This applies, under the stated accepted-step conditions, to strongly convex fixed-resolution quantile flows and the full Gaussian model below. It is a theorem about contraction of the specified continuous field, not arbitrary Wasserstein proximal maps.

Proof. The exact flow is $e^{-\alpha h}$-contractive, so the local comparison gives $e_{n+1} \le e^{-\alpha h_n} e_n + C h_n^3$. Bound the resulting exponentially weighted sum by $e^{\alpha h_{\max}} h_{\max}^2/\alpha$. Appendix R gives the integral comparison and verifies the model constants. □

Why freeze after predicting? Freezing $K(x)$ preserves energy but usually omits the $DK(x)[f]$ term and loses one order. Evaluating $K((x + y)/2)$ can be second order, but its y dependence generally destroys the convex-gradient representation (4). A certified first-order midpoint prediction restores the missing term while keeping B fixed during the inner solve. This design is compatible with classical predictors; neural prediction is optional and must pass the same tests.

Proposition 3.4 (Finite correction, with a nonstationarity qualification)**.** *Fix an admissible x, h, B with $\beta_h > 0$, an interior minimizer $y^\star$, and $\nabla E(x) \ne 0$. Suppose a full-state corrector converges to $y^\star$ and its AVF oracle enclosures shrink to zero. Then it passes* (8) *in finitely many iterations. If the corrector has a linear objective-gap bound on a compact sublevel set, its iteration count is logarithmic in the inverse residual tolerance and the inverse nonstationarity margin. At an exactly certified stationary state, return $y = x$; unresolved stationarity is not a license to advance time.*

The proof uses $g(x, y^\star) \ne 0$, which leaves a strict energy margin $\eta \|g(x, y^\star)\|_B^2$. Appendix G supplies the quantitative threshold and a concrete backtracking corrector. A finite training budget by itself does not satisfy this proposition.

4 Two Wasserstein model classes

4.1 Coupled positive quantiles, not merely a scalar test

On an interval $[-L, L]$, let X_q be piecewise affine on a mass partition with weights $w_i > 0$, endpoints $q_0 = -L, q_N = L$, and gaps $d_i = q_{i+1} - q_i > 0$. The density is w_i/d_i on its corresponding cell. The exact mass matrix A satisfies

$$W_2(\rho_q, \rho_p)^2 = \|q - p\|_A^2, \qquad E_N(q) = \kappa \sum_i \frac{w_i^2}{d_i} + \int_0^1 V(X_q(a))\,\mathrm{d}a. \tag{12}$$

For several components, use the direct-sum mass metric and add a mean interaction $\Phi(m)$, where $m_j = \int_0^1 X_{q_j}$. With $V_j'' \ge a_j$ and $\nabla^2 \Phi \succeq -\ell_\Phi \mathrm{I}$, the energy curvature is at least $\min_j a_j - \ell_\Phi$. Quartic $\Phi(m) = b \|m\|^4/4 - c\|m\|^2/2$ gives a concrete nonlinear, potentially nonconvex coupling.

Theorem 4.1 (Quantile globality and distributional certification)**.** *For the coupled model with quadratic V_j, any frozen positive B in the fixed mass Hilbert space, and $\beta_h > 0$, the AVF objective has a unique interior global minimizer. Its diffusion contribution is*

$$-\sum_{j,i} \frac{\kappa_j w_{j,i}^2}{d_{j,i}(x)} \log \frac{d_{j,i}(y)}{d_{j,i}(x)}. \tag{13}$$

Thus Theorems 2.1–3.2 apply under their respective additional hypotheses, with genuine Wasserstein endpoint bounds. If $B = \mathrm{I}$, the target is the stated semidiscrete Wasserstein flow. A different state-dependent K defines a different gradient metric; measuring its error in W_2 does not make it the standard Wasserstein flow.

The logarithmic divergence in (13) excludes zero gaps; the finite interval bounds all other coordinates. For product measures, squared W_2 distances add across components. The associated marginal-energy continuum model preserves product structure under its smooth coordinatewise flow, even with mean coupling. Appendices I–J give the full derivations, including a distinction between a marginal energy and joint entropy.

4.2 Full-covariance Gaussian flows with variable mobility

Let $u = (m, C) \in \mathbb{R}^d \times \mathbb{S}^d_{++}$, $A \succ 0$, and

$$E(m, C) = \tfrac{1}{2} m^T A m + \tfrac{1}{2} \operatorname{tr}(AC) - \tfrac{1}{2} \log\det C, \quad K_{(m,C)}(v, H) = (v, 2(CH + HC)). \tag{14}$$

In the Euclidean–Frobenius inner product, (1) becomes $\dot{m} = -Am$, $\dot{C} = 2\mathrm{I} - AC - CA$, the Gaussian Wasserstein flow of relative entropy up to an additive constant. No commutation of A and C is imposed. The Wasserstein geometry underlying this formula is classical (Takatsu, 2008; Bhatia et al., 2017); our application is the frozen-mobility certificate and its interior-solvability argument.

Theorem 4.2 (Gaussian convex step and physical order)**.** *For every frozen $Z \succ 0$ and $h > 0$, take $B(v, H) = (v, 2(ZH + HZ))$. Then $Q^B_{h,x}$ has a unique minimizer in $\mathbb{R}^d \times \mathbb{S}^d_{++}$, with $\beta_h = 1/h$. Its covariance AVF gradient is*

$$g_C(C, D) = \tfrac{1}{2} A - \tfrac{1}{2} C^{-1/2} \varphi(C^{-1/2} D C^{-1/2}) C^{-1/2}, \quad \varphi(t) = \frac{\log t}{t - 1}, \quad \varphi(1) = 1. \tag{15}$$

Use a certified midpoint $Z = C + h(2\mathrm{I} - AC - CA)/2 + O(h^2)$. On a fixed entropy sublevel set, accepted steps are second order in Gaussian W_2, and (6) *gives a computable Gaussian endpoint radius after the explicit norm conversion in Appendix M.*

Why the Gaussian existence proof is different. Let $T = C^{-1/2} D C^{-1/2}$ have eigenvalues t_i. The integrated entropy term of Q is $\sum_i p(t_i)$, with

$$p(t) = -\tfrac{1}{2} \int_0^1 \frac{\log(1 - s + st)}{s} \, \mathrm{d}s.$$

The value $p(0)$ is finite, so a logarithmic-barrier divergence argument would be false. The potential instead extends continuously and coercively to positive semidefinite covariances. At a singular candidate $\overline{D}$, move inward to $D_\epsilon = (1 - \epsilon)\overline{D} + \epsilon C$. An eigenvalue $t_i = 0$ contributes $-\varphi(\epsilon)/2 \to -\infty$ to the directional derivative, whereas all nonsingular and quadratic terms remain bounded. Thus the closed-domain minimizer cannot be singular. This proves interior attainment before applying strong convexity. Appendix N gives the domination, coercivity, and derivative details. A compact-interior-sublevel corrector is not automatically available for every Gaussian initializer; a separately certified local contraction supplies a small-step fallback.

A genuine distributional certificate. On an entropy sublevel $C, D \succeq \sigma\mathrm{I}$, a common Gaussian coupling and a Sylvester inequality give

$$W_2(\mathcal{N}(m, C), \mathcal{N}(n, D))^2 \le \|m - n\|^2 + \frac{\|C - D\|_F^2}{4\sigma}.$$

Together with the frozen spectral bound, this converts the computed state radius into a Wasserstein radius. The exact covariance flow has contraction rate at least $2\lambda_{\min}(A)$ in Frobenius norm; the joint mean–covariance field has rate at least $\lambda_{\min}(A)$. Thus Corollary 3.3 gives a uniform physical-time conclusion for this noncommuting model, not only a diagonal Gaussian calculation.

5 A continuum certificate with a visible spatial defect

Finite-dimensional temporal order is not a continuum theorem. To make the difference explicit, consider smooth increasing quantiles $X : [0, 1] \to [-L, L]$ with fixed endpoints and

$$\mathcal{E}(X) = \kappa \int_0^1 \frac{1}{X_a} \, \mathrm{d}a + \int_0^1 V(X) \, \mathrm{d}a, \qquad \mathcal{G}(X) = \kappa \partial_a(X_a^{-2}) + V'(X). \tag{16}$$

Assume a sufficiently regular positive solution $X_t = -\mathcal{G}(X)$ on $[0,T]$ and $V'' \geq \lambda$. These are explicit regularity hypotheses, not conclusions for arbitrary rough densities. Let $U(t)$ be any continuous, piecewise-C^1 positive quantile reconstruction with the same endpoints and $\mathcal{G}(U) \in L^2$. Define its *full* residual $\mathcal{R} = U_t + \mathcal{G}(U)$.

Theorem 5.1 (Continuum physical-time posterior bound)**.** *Under these hypotheses,*

$$W_2(\rho_{U(t)}, \rho_{X(t)}) \leq e^{-\lambda t} \left\| U(0) - X(0) \right\|_{L^2} + \int_0^t e^{-\lambda(t-s)} \left\| \mathcal{R}(s) \right\|_{L^2} \, \mathrm{d}s. \tag{17}$$

If U lies in an affine smooth finite-dimensional space with orthogonal projector P on its variation space, then

$$\left\| \mathcal{R} \right\|^2 = \underbrace{\left\| U_t + P\mathcal{G}(U) \right\|^2}_{\textit{temporal/solve defect}} + \underbrace{\left\| (\mathrm{I} - P)\mathcal{G}(U) \right\|^2}_{\textit{spatial defect}}. \tag{18}$$

Every validated bound on these two terms yields a total continuum error certificate, rather than an unknown additive spatial-error symbol.

Proof. For endpoint-zero variations, integration by parts and monotonicity of $p \mapsto -\kappa p^{-2}$ give $\langle \mathcal{G}(U) - \mathcal{G}(X), U - X \rangle \geq \lambda \left\| U - X \right\|^2$. Apply this to the difference equation, regularize the norm at zero, and integrate the scalar inequality. The one-dimensional quantile isometry converts it to (17). Orthogonality gives (18). Appendices K–L address boundary conditions, regularity, and computable smooth representations. □

A useful quadratic reconstruction for a finite-state vector field is $U_n(s) = x_n + s\delta_n + \frac{h_n}{2} s(s-1)[F(x_{n+1}) - F(x_n)]$. For a smooth quantile Galerkin model, use the projected vector field instead. Its positivity must be certified throughout the slab, not just at endpoints. If the full residual is bounded by b_n on a slab, propagate $e^{\mathrm{up}}_{n+1} = e^{-\lambda h_n} e^{\mathrm{up}}_n + c_\lambda(h_n) b_n$, where $c_\lambda(h) = \int_0^h e^{-\lambda s} \, \mathrm{d}s$. A prescribed global budget can therefore reject an outer step independently of the inner stopping test. Appendix R also proves uniform-in-time second-order error under an additional continuous-flow contraction hypothesis. A continuum second-order conclusion requires the *spatial* defect budget to decrease accordingly; the temporal theorem alone does not supply that fact.

6 Learning, sharpness, and the limits of certification

Learning enters through proposed states, not through an assumed ability to solve nonconvex training globally. A bounded network can propose $z = x + hF(x)/2 + h^2 a_\theta(x,h)$ and an endpoint. This midpoint parameterization automatically has the required scale once its bound and admissibility are certified. A state-space corrector may then repair the endpoint. Appendix O proves how a uniformly accurate predictor reduces a provably convergent corrector's work, without asserting that training attains that accuracy.

Proposition 6.1 (Sharp worst-case scales)**.** *For $E(x) = x^2/2$, $B = 1$, choose the inexact AVF residual $r = -ch^p(x+y)/2$, with $c > 0$. Every sufficiently small step dissipates E. For $p = 0$ the iterates approach the wrong rate; for $p = 1$ they are generally only first order; for $p = 2$ they are compatible with second order. Their exact subproblem gaps scale as h^{2p+1} on a nonzero finite-time trajectory. Also, for $E(x) = x$ and $K(x) = x$ on $(0,\infty)$, exactly solving the step with $B = K(x)$ gives explicit Euler, despite energy decrease.*

These examples show why $O(h^2)$ residual and midpoint accuracy, and a sufficient $O(h^5)$ certified gap, are meaningful worst-case requirements. They are not necessity claims for every error sequence; cancellations can improve particular trajectories. Appendix H also gives exact counterexamples for original-energy BDF2 monotonicity, extrapolated endpoints, parameter gradients, and unrestricted finite-sample validation.

For random validation, a reused training sample is not a deterministic oracle enclosure. Appendix Q proves a fresh-sample, adaptively indexed confidence-spending theorem. The resulting claims hold on a specified simultaneous-probability event. A bounded squared-residual Hoeffding certificate can require $O(h^{-8})$ samples to resolve an $O(h^4)$ threshold; this cost is exposed, not hidden. Deterministic interval constructions for the rational quantile forces and smooth polynomial lifts are given in Appendix S.

6.1 THE TARGET ENERGY CANNOT BE REPLACED SILENTLY

If both energy and mobility are learned, write $\widehat{F} = -\widehat{K}\nabla\widehat{E}$ and let δ_F certify $\left\|F - \widehat{F}\right\|$ on the reconstruction. Its true defect is $R = \widehat{R} + \widehat{F} - F$, so a trajectory bound must include the integral of δ_F. Refining the time mesh does not remove a fixed model-error floor. Likewise, a uniform energy-value error gives only $E(y) - E(x) \leq \widehat{E}(y) - \widehat{E}(x) + 2\delta_0$; it gives no gradient bound without further regularity. Appendix V proves these statements and shows how certified surrogate-gradient error can enter the robust AVF envelope. In particular, an oracle floor $\sigma_0 > 0$ makes the stated residual gate unresolved for $h < \sqrt{\sigma_0/\tau}$, regardless of additional inner training.

7 SCOPE AND CONCLUSION

The results do not establish convergence for arbitrary deep generators, uniform-in-dimension conditioning, or an empirical advantage of learned predictors. A certificate must concern the intended energy and mobility, not only a sampled surrogate. The full-covariance Gaussian model covers a genuine variable Wasserstein mobility; the coupled quantile model supplies exact distributional geometry and global positive step solvability; the continuum theorem supplies a conditional, total posterior bound with a visible spatial residual. Together they give a concrete answer to the physical-clock problem: *certify the midpoint mobility, certify full-state solve accuracy, and separately certify signed energy work*. None of these obligations can generally replace the other two.

AI USE STATEMENT

The authors declare that no generative AI tools or LLMs were used to generate text, code, or ideas for this paper. Standard spell-checking and basic grammar-checking tools were used solely for proofreading.

REFERENCES

Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. arXiv:1712.01504, 2017. URL https://arxiv.org/abs/1712.01504.

José A. Carrillo, Katy Craig, Li Wang, and Chaozhen Wei. Primal dual methods for Wasserstein gradient flows. *Foundations of Computational Mathematics*, 22:389–443, 2022. doi: 10.1007/s10208-021-09503-1.

Elena Celledoni, Volker Grimm, Robert I. McLachlan, David I. McLaren, Dion O'Neale, Brynjulf Owren, and G. R. W. Quispel. Preserving energy resp. dissipation in numerical PDEs using the "Average Vector Field" method. *Journal of Computational Physics*, 231(20):6770–6789, 2012. doi: 10.1016/j.jcp.2012.06.022.

Elena Celledoni, Brynjulf Owren, Chong Shen, Baige Xu, and Takaharu Yaguchi. UEPI: Universal energy-behavior-preserving integrators for energy conservative/dissipative differential equations. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/3e467e15cbcbd7b5ecc814cfc27c4ea9-Abstract-Conference.html.

Raymond Chu and Matt Jacobs. Guaranteeing higher order convergence rates for accelerated Wasserstein gradient flow schemes. arXiv:2511.10884, 2025. URL https://arxiv.org/abs/2511.10884.

Simone Di Marino, Emanuele Naldi, and Silvia Villa. Inexact JKO and proximal-gradient algorithms in the Wasserstein space. arXiv:2505.23517, 2025. URL https://arxiv.org/abs/2505.23517.

Simone Di Marino, Sara Farinelli, and Emanuele Naldi. On the stability of proximal operators in Wasserstein spaces under different notions of convexity. arXiv:2607.08209, 2026. URL https://arxiv.org/abs/2607.08209.

Matthias J. Ehrhardt, Erlend S. Riis, Torbjørn Ringholm, and Carola-Bibiane Schönlieb. A geometric integration approach to smooth optimization: foundations of the discrete gradient method. *IMA Journal of Numerical Analysis*, 45(3):1269–1299, 2025. doi: 10.1093/imanum/drae037.

Sølve Eidnes. Order theory for discrete gradient methods. *BIT Numerical Mathematics*, 62:1207–1255, 2022. doi: 10.1007/s10543-022-00909-z.

Zhaohui Fu, Tao Tang, and Jiang Yang. Energy diminishing implicit-explicit Runge–Kutta methods for gradient flows. arXiv:2203.06034, 2024. URL https://arxiv.org/abs/2203.06034. Revised version.

Thomas Gallouët, Andrea Natale, and Gabriele Todeschi. From geodesic extrapolation to a variational BDF2 scheme for Wasserstein gradient flows. *Mathematics of Computation*, 93(350):2769–2810, 2024. doi: 10.1090/mcom/3951.

Emmanuil H. Georgoulis, Michail Loulakis, and Asterios Tsiourvas. Discrete gradient flow approximations of high dimensional evolution partial differential equations via deep neural networks. *Communications in Nonlinear Science and Numerical Simulation*, 117:106893, 2023. doi: 10.1016/j.cnsns.2022.106893.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi: 10.1080/01621459.1963.10500830.

Ziqing Hu, Chun Liu, Yiwei Wang, and Zhiliang Xu. Energetic variational neural network discretizations of gradient flows. *SIAM Journal on Scientific Computing*, 46(4):A2528–A2556, 2024. doi: 10.1137/22M1529427.

Jizu Huang. High-order energy stable discrete variational derivative schemes for gradient flows. *IMA Journal of Numerical Analysis*, 45(4):2329–2360, 2025. doi: 10.1093/imanum/drae062.

Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker–Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998. doi: 10.1137/S0036141096303359.

Lucas Lautwein, Nicole Marheineke, Håkon Noren Myhr, and Kevin Schäfers. Fourth-order energy-consistent average vector field discrete gradient methods for dissipative Hamiltonian systems. arXiv:2609.24692, 2026. URL https://arxiv.org/abs/2609.24692.

Charalambos Makridakis and Ricardo H. Nochetto. A posteriori error analysis for higher order dissipative methods for evolution problems. *Numerische Mathematik*, 104:489–514, 2006. doi: 10.1007/s00211-006-0013-6.

Daniel Matthes and Simon Plazotta. A variational formulation of the BDF2 method for metric gradient flows. *ESAIM: Mathematical Modelling and Numerical Analysis*, 53(1):145–172, 2019. doi: 10.1051/m2an/2018045.

Petr Mokrov, Alexander Korotin, Lingxiao Li, Aude Genevay, Justin M. Solomon, and Evgeny Burnaev. Large-scale Wasserstein gradient flows. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/810dfbbebb17302018ae903e9cb7a483-Abstract.html.

Håkon Noren Myhr and Sølve Eidnes. Derivative-free discrete gradient methods. *Journal of Computational Dynamics*, 11(3):256–273, 2024. doi: 10.3934/jcd.2024004.

Asuka Takatsu. On Wasserstein geometry of the space of Gaussian measures. arXiv:0801.2250, 2008. URL https://arxiv.org/abs/0801.2250.

Chen Xu, Xiuyuan Cheng, and Yao Xie. Normalizing flow neural networks by JKO scheme. In *Advances in Neural Information Processing Systems*, volume 36, pp. 47379–47405, 2023. URL https://papers.neurips.cc/paper_files/paper/2023/hash/93fce71def4e3cf418918805455d436f-Abstract-Conference.html.

PROOF GUIDE

The supplementary theory is organized by dependency rather than by experiment. Appendices A–H establish the finite-state certificate interface and its sharpness. Appendices I–L develop positive distributional models and the continuum comparison. Appendices M–N treat full-covariance Gaussian geometry and solvability. The remaining appendices address learning, adaptation, statistical validation, long-time behavior, validated arithmetic, invariants, and budget allocation. Every theorem's target space and additional assumptions are stated where used.

A SETTING, CERTIFICATE SEMANTICS, AND THEOREM DEPENDENCIES

This appendix fixes conventions that are essential when the mobility is state dependent. A certificate has to refer to a specified state space, metric, energy, and clock. Changing any of these objects changes the assertion being certified. The finite-dimensional results are stated first; the continuum assertions later have additional hypotheses that are not consequences of the finite-dimensional theorems.

A.1 THREE DIFFERENT METRICS

Let H be a finite-dimensional real Hilbert space. Its fixed inner product is denoted $\langle \cdot, \cdot \rangle$, and its norm by $\|\cdot\|$. A coordinate realization may have a fixed mass matrix $A \succ 0$, so that $\langle u, v \rangle = u^T A v$. In that realization the Hilbert gradient is A^{-1} times the coordinate derivative. An operator is self-adjoint if it is self-adjoint in this inner product; it need not be represented by a symmetric array in an arbitrary coordinate basis.

For a fixed positive self-adjoint operator B, define

$$\|v\|_{B^{-1}}^2 = \left\langle B^{-1} v, v \right\rangle, \qquad \|q\|_B^2 = \langle q, Bq \rangle . \tag{19}$$

The letter q in the second expression represents a Hilbert gradient or covector identified by the fixed Riesz map. Thus

$$|\langle q, v \rangle| \le \|q\|_B \, \|v\|_{B^{-1}} . \tag{20}$$

If $b_- \mathrm{I} \preceq B \preceq b_+ \mathrm{I}$, then

$$b_+^{-1/2} \, \|v\| \le \|v\|_{B^{-1}} \le b_-^{-1/2} \, \|v\| , \qquad \|q\|_B \le b_+^{1/2} \, \|q\| .$$

These identities follow by writing $v = B^{1/2} w$ and applying ordinary Cauchy–Schwarz. No differentiation of B is involved: B is frozen during a step.

The third metric is a distributional distance such as W_2. An endpoint radius in $\|\cdot\|_{B^{-1}}$ becomes a distributional radius only after proving a comparison for the representation. Quantile finite elements provide an exact isometry in their mass metric. Gaussian coordinates provide a local comparison on a specified covariance sublevel. A generic generator map provides neither automatically.

A.2 EXACT MATHEMATICS, DETERMINISTIC VALIDATION, AND STATISTICAL VALIDATION

Definition A.1 (Deterministic oracle envelope)**.** For a fixed accepted state x, proposed y, and frozen B, a deterministic AVF envelope consists of $\widetilde{g} \in \mathsf{H}$ and $\sigma \ge 0$ together with a justified bound $\|g(x, y) - \widetilde{g}\|_B \le \sigma$. A validated upper bound for a norm, a spectral radius, or a scalar expression is one whose inequality includes evaluation and rounding errors.

A symbolic identity is one way to justify an envelope. A rational interval calculation is another. A pointwise floating-point evaluation, without error analysis, is not itself such a justification. Theorems that assume envelopes are mathematical implications; an implementation has to establish their premises.

A statistical envelope is different. It satisfies the same inequality on an event of specified probability. Applying a deterministic theorem on that event is legitimate. Calling the resulting assertion deterministic, or choosing a validation sample after observing its favorable outcome without accounting for selection, is not legitimate. Appendix Q gives an explicitly adaptive probability statement.

Definition A.2 (Accepted physical step)**.** An accepted step consists of x, y, h, z, B and the certificate data, with $h > 0$, satisfying the stated admissibility, midpoint, mobility, residual, and work conditions. It changes the physical time by h, regardless of the number of training or correction iterations. A rejected trial changes neither the accepted state nor the accepted physical time.

This definition permits adaptive meshes and adaptive inner work. It does not identify an optimizer iteration with a time step. It also does not identify an extrapolated state with either of the states from which it was formed: an extrapolated state requires its own acceptance test.

A.3 Assumptions that are deliberately separate

There are four logically distinct issues. Smoothness makes the local expansions meaningful. Positive definiteness makes the frozen norms meaningful. A lower Hessian bound makes Q strongly convex for an appropriate h. Finally, an existence argument gives an interior critical point. The first three do not imply the fourth on an arbitrary open domain.

For example, $Q(y) = (y-1)^2$ on $\mathcal{D} = (2, \infty)$ is smooth and strongly convex but has no minimizer in $\mathcal{D}$. A sequence approaching its infimum approaches a boundary point. The global-gap theorem in the main text therefore states interior existence as an independent premise. The quantile barrier and Gaussian inward-derivative arguments prove that premise in two different ways.

A compact corridor is likewise a premise of the general temporal theorem, not a universal consequence of dissipation. In the concrete models, energy sublevels and certified midpoint proximity provide the corridor. In a noncoercive model, energy descent may allow escape to infinity. In an energy with a finite boundary value, it may allow approach to the boundary. The proof must use the actual sublevel geometry rather than the word "stable."

A.4 Dependency map

The global gap theorem depends only on the frozen variational identity, strong convexity, and interior existence. It does not depend on second-order consistency. The energy theorem depends only on the exact chain rule and an oracle envelope. It does not depend on convexity. The temporal theorem depends on a consistent frozen midpoint, small full-state residual, and a finite-time smooth corridor. It does not require a global minimizer at every step if accepted states are otherwise supplied. The global minimizer is useful because it gives a provable way to supply those states and certify their errors.

The continuum theorem depends on an appropriate monotonicity inequality for the full continuum operator and on a smooth admissible reconstruction. It is not obtained by replacing the symbol H by an infinite-dimensional space in the finite-state proof. In particular, the discrete diffusion gradient and its continuum counterpart can have different regularity.

Remark A.3 (What is classical and what is combined here)**.** The AVF discrete chain rule, strong-convexity gradient bound, Taylor consistency argument, and reconstruction comparison principle are established mathematical tools. We prove the versions needed here to expose constants and compatibility conditions. The proposed algorithmic object is their particular certified coupling with a frozen predicted mobility and an inexact oracle. The paper does not make a priority claim for any of the classical inequalities in isolation.

B The frozen AVF potential and global error certificates

B.1 Removing the apparent singularity

Fix $x \in \mathcal{D}$ and write $v = y - x$. For $s > 0$, the fundamental theorem of calculus gives

$$\frac{E(x+sv) - E(x)}{s} = \int_0^1 \langle \nabla E(x + tsv), v \rangle \, \mathrm{d}t. \tag{21}$$

The right side extends continuously to $s = 0$, with value $\langle \nabla E(x), v \rangle$. For y in a compact neighborhood inside $\mathcal{D}$, all segments $x + s(y - x)$ belong to a compact subset of $\mathcal{D}$; this follows from convexity and compactness. Derivatives of E are bounded there. Hence dominated differentiation applies to the extended integral.

Lemma B.1 (First and second variations). *For fixed $h > 0$ and $B \succ 0$, the potential in* (4) *satisfies*

$$\nabla Q(y) = h^{-1}B^{-1}(y-x) + g(x,y),\ \nabla^2 Q(y) \quad = h^{-1}B^{-1} + \int_0^1 s\nabla^2 E(x + s(y-x))\,\mathrm{d}s. \tag{22}$$

Furthermore, $E(y) - E(x) = \langle g(x,y), y - x\rangle$.

Proof. For $s > 0$, differentiation of $[E(x + s(y-x)) - E(x)]/s$ in a direction w gives $\langle \nabla E(x + s(y-x)), w\rangle$. Its derivative in a second direction v is $s\left\langle \nabla^2 E(x + s(y-x))v, w\right\rangle$. The preceding compactness argument permits integration of these derivatives. The kinetic part has gradient $B^{-1}(y-x)/h$ and Hessian B^{-1}/h. Finally, integration of $\frac{\mathrm{d}}{\mathrm{d}s}E(x + s(y-x))$ proves the chain identity. □

There is no term involving DB in this lemma. Introducing a y-dependent mobility inside Q and differentiating it would produce such terms and a different equation. This is the precise reason for freezing after midpoint prediction.

B.2 Curvature, semiconvex energies, and norms

Lemma B.2 (Frozen strong convexity). *Suppose $\left\langle \nabla^2 E(v)w, w\right\rangle \geq \lambda_B \|w\|_{B^{-1}}^2$ for all $v \in \mathcal{D}$. If $\beta_h = 1/h + \lambda_B/2 > 0$, then for $y, v \in \mathcal{D}$,*

$$Q(v) \geq Q(y) + \langle \nabla Q(y), v - y\rangle + \frac{\beta_h}{2}\|v - y\|_{B^{-1}}^2. \tag{23}$$

If instead $\nabla^2 E \succeq -\ell \mathrm{I}$, one may set $\lambda_B = -\ell \|B\|_{\mathrm{op}}$.

Proof. Integrating $s\lambda_B$ over $[0,1]$ in (22) gives the lower Hessian bound. Integrating that bound twice along the segment from y to v yields (23). For the last assertion, $\|w\|^2 \leq \|B\|_{\mathrm{op}} \|w\|_{B^{-1}}^2$, and multiplication by $-\ell$ reverses the appropriate inequality. □

The global lower bound is important. A Hessian estimate measured only at the candidate y certifies neither global strong convexity nor a global objective gap. A local estimate can instead be used within a separately certified localization region, but then one must establish that the global minimizer belongs to that region.

Complete proof of Theorem 2.1. An interior minimizer has $\nabla Q(y^\star) = 0$. Applying (23) at $y^\star$ proves that it is the unique global minimizer. Apply the same inequality at y and take the infimum over $v \in \mathcal{D}$. Enlarging the infimum to all of H only decreases the right side, so

$$Q(y^\star) \geq Q(y) + \inf_{w \in \mathsf{H}} \left\{ \langle \nabla Q(y), w\rangle + \frac{\beta_h}{2}\|w\|_{B^{-1}}^2 \right\}.$$

The quadratic minimum occurs at $w = -B\nabla Q(y)/\beta_h = -r/\beta_h$ and has value $-\|\nabla Q(y)\|_B^2/(2\beta_h)$. Since $\|\nabla Q(y)\|_B = \|r\|_{B^{-1}}$, this is the gap assertion.

For the radius, integrate the Hessian along $[y^\star, y]$ to obtain

$$\beta_h \|y - y^\star\|_{B^{-1}}^2 \leq \langle \nabla Q(y) - \nabla Q(y^\star), y - y^\star\rangle \leq \|r\|_{B^{-1}} \|y - y^\star\|_{B^{-1}}.$$

Division when the distance is nonzero proves the claim; at zero it is immediate. □

Corollary B.3 (Approximate lower bound on the optimum). *The quantity $L_h(y) = Q(y) - \|r\|_{B^{-1}}^2/(2\beta_h)$ is a lower bound on $\min Q$. If $Q(y) \in [Q_-, Q_+]$ and $\|r\|_{B^{-1}} \leq \bar{\varepsilon}$ are validated, then $Q_- - \bar{\varepsilon}^2/(2\beta_h) \leq \min Q \leq Q_+$.*

Proof. The lower statement is the preceding square-completion argument. The upper statement evaluates Q at the admissible candidate. Lowering the evaluated value for the lower bound and raising the residual bound can only weaken, not invalidate, the inequality. □

This form is useful when an application explicitly requests a computable upper-minus-lower objective certificate. It does not require knowing $Q(y^\star)$ or solving a separate dual optimization problem.

B.3 Existence versus local solvability

Proposition B.4 (Local exact branch)**.** *Suppose $x \in \mathcal{D}$ is fixed, $B(h)$ is C^1 near $h = 0$ with $B(0) \succ 0$, and $E \in C^2$. For all sufficiently small h, there is a unique local solution near x of $y - x + hB(h)g(x, y) = 0$, with $y(0) = x$. If, in addition, the global curvature condition of Lemma B.2 holds, this local solution is the unique global minimizer of Q.*

Proof. The left side, denoted $\Psi(h, y)$, is continuously differentiable, vanishes at $(0, x)$, and has $D_y\Psi(0, x) = \mathrm{I}$. The implicit function theorem gives the local branch and its uniqueness. The branch lies in $\mathcal{D}$ for small h because $\mathcal{D}$ is open. Its stationarity for Q and the global strong-convexity inequality prove the final assertion. □

This proposition is sufficient for a local method on a smooth admissible space, but its step threshold depends on the distance to the boundary. The concrete global model results are stronger: they prove existence for every frozen step satisfying their curvature requirement. They should not be confused with a uniform-in-spatial-resolution small-step result.

B.4 A useful transformation

Set $w = B^{-1/2}(y - x)$ and $\widehat{Q}(w) = Q(x + B^{1/2}w)$. Then $\nabla\widehat{Q} = B^{1/2}\nabla Q = B^{-1/2}r$ and $\nabla^2\widehat{Q} \succeq \beta_h\mathrm{I}$. Thus all residual and objective estimates are ordinary Euclidean estimates in these frozen normalized coordinates. This transformation will also be used for correction complexity. It is performed separately on each step and is not a nonlinear coordinate change of the underlying continuous flow.

C Oracle errors, energy work, and robust acceptance

C.1 The exact balance identity

For any admissible pair x, y, not necessarily near an exact solution, the residual definition gives

$$E(y) - E(x) = -h\,\|g\|_B^2 + h\,\langle g, r\rangle\,. \tag{24}$$

The first term is the nominal dissipation and the second is the signed work performed by the solve error. An error orthogonal to g performs no energy work, although it can still change the trajectory. A downhill error can increase the apparent dissipation while changing the physical rate. Both observations explain why an unsigned training loss or an energy test alone is insufficient.

Lemma C.1 (Independent chain and residual envelopes)**.** *Suppose $\|g - \widetilde{g}\|_B \le \sigma$ and $E(y) - E(x) - \langle\widetilde{g}, \delta\rangle \le \chi$, where $\chi \ge 0$ is a separately certified one-sided chain bound. Then*

$$\|r\|_{B^{-1}} \le \|\widetilde{r}\|_{B^{-1}} + \sigma,\ E(y) - E(x) \qquad \le -h\widetilde{D} + h\widetilde{\omega} + \chi. \tag{25}$$

Consequently $\widetilde{\omega} + \chi/h \le \eta\widetilde{D}$ suffices for original-energy dissipation. One valid choice is $\chi = \sigma\,\|\delta\|_{B^{-1}}$.

Proof. Since $r - \widetilde{r} = B(g - \widetilde{g})$, its B^{-1} norm equals $\|g - \widetilde{g}\|_B$. The triangle inequality proves the first claim. Substitute $\delta = -hB\widetilde{g} + h\widetilde{r}$ into the enclosed chain relation to prove the second. Finally, dual Cauchy–Schwarz bounds $\langle g - \widetilde{g}, \delta\rangle$ by $\sigma\,\|\delta\|_{B^{-1}}$. □

The chain envelope need not be derived from the same numerical procedure as the gradient envelope. For example, an interval evaluation of $E(y) - E(x)$ may be much sharper in the direction δ than a uniform gradient enclosure. Conversely, a cancellation-aware difference formula can be preferable to subtracting separately enclosed energies near equilibrium.

Proposition C.2 (Energy-interval alternative)**.** *Suppose $E(x) \in [E_x^-, E_x^+]$ and $E(y) \in [E_y^-, E_y^+]$. The test $E_y^+ - E_x^- \le -(1 - \eta)h\widetilde{D}$ certifies the same energy conclusion as the signed-work gate. It does not, by itself, bound the nonlinear residual or the error relative to an exact time step.*

Proof. The true energy difference is at most $E_y^+ - E_x^-$. The second sentence follows because the inequality contains no information about the vector equation defining the step; the scalar examples in Appendix H satisfy energy decrease with an incorrect rate. □

C.2 An inexact mobility evaluation

Suppose a midpoint z has been certified but $K(z)$ is computed only approximately. It is essential to distinguish two possible implementations. In the method studied here, the actual positive operator B used in the solve is the frozen operator, and $\|B - K(z)\|_{\mathrm{op}} \le \chi_K$ is a consistency certificate. The energy identity uses B exactly as used by the solve. There is no energy error due solely to $B \neq K(z)$, because every positive B dissipates the same E under the exact frozen equation. The mobility discrepancy instead enters the physical-time consistency bound.

If multiplication by B itself is inexact, say a computed vector v_B satisfies $\|v_B - B\widetilde{g}\|_{B^{-1}} \le \sigma_B$, use $\widetilde{r} = \delta/h + v_B$. Then the true residual envelope becomes $\|r\|_{B^{-1}} \le \|\widetilde{r}\|_{B^{-1}} + \sigma + \sigma_B$. To preserve the signed-work proof, either recompute the true $B\widetilde{g}$ with an enclosure or include the additional work bound $h\,\|\widetilde{g}\|_B\,\sigma_B$. Omitting this term would confuse an approximate matrix-vector product with the exact frozen metric.

C.3 Norm enclosures and sufficient scalar tests

If a verified lower spectral bound $B \succeq b_-\mathrm{I}$ is available, then $\|\widetilde{r}\|_{B^{-1}} \le \|\widetilde{r}\|/\sqrt{b_-}$. This may be conservative but remains valid. A verified solve for $B^{-1}\widetilde{r}$ can give a sharper bound. For acceptance, a conservative upper bound is harmless for correctness, though it may increase rejection or correction cost.

A simple sufficient exact-oracle energy condition is

$$\|r\|_{B^{-1}} \le \eta_r \|g\|_B\,, \qquad 0 \le \eta_r < 1. \tag{26}$$

It implies $\langle g, r\rangle \le \eta_r \|g\|_B^2$ by dual Cauchy–Schwarz. The signed test is weaker and can accept large errors in favorable directions. The relative norm condition is useful for long-time quantitative decay, where it prevents residuals from dominating near equilibrium. It is not necessary for a single dissipative step.

Lemma C.3 (A sufficient robust relative threshold)**.** *Let $d = \|\widetilde{g}\|_B$. If $d > 0$, $\|\widetilde{r}\|_{B^{-1}} \le qd$, and $\sigma \le qd$ for some $q \ge 0$, then*

$$\widetilde{\omega} + \frac{\sigma}{h}\,\|\delta\|_{B^{-1}} \le (2q + q^2)d^2.$$

Thus the work gate holds whenever $2q + q^2 \le \eta$.

Proof. The residual relation gives $\|\delta\|_{B^{-1}}/h \le d + \|\widetilde{r}\|_{B^{-1}} \le (1+q)d$. Also $\widetilde{\omega} \le d\,\|\widetilde{r}\|_{B^{-1}} \le qd^2$. Add the two inequalities. □

This lemma makes the near-equilibrium requirement explicit: to retain a fixed fractional dissipation margin, the oracle error must eventually be small relative to the gradient, not merely below a fixed absolute tolerance.

C.4 Exact equilibrium and a validation obstruction

At an exact equilibrium x, the exact frozen solution is $y = x$ and $g = 0$. The signed gate has no strict margin: both nominal dissipation and permitted work vanish. If g is represented by an interval containing zero but not reduced to the singleton zero, a naive robust gate may fail forever. This is not evidence that the physical state is changing; it is evidence that the oracle has not resolved stationarity.

There are three mathematically distinct outputs. A certified stationary state can be held fixed. A sufficiently accurate nonstationary state can be advanced after passing the gates. An unresolved case can be reported without advancing the clock. The finite-correction theorem covers the second case. It does not promise finite exact equality testing for arbitrary transcendental data.

More generally, no fixed positive absolute energy-evaluation tolerance can certify strict relative dissipation for all arbitrarily near-equilibrium states. For the scalar quadratic, the true step energy decrease is proportional to hx^2. Choose x so small that this decrease lies below the interval uncertainty. The intervals alone then admit both signs of the energy difference. Additional precision, a structural identity, or an explicitly weakened guarantee is required. This observation is independent of how expressive the neural predictor is.

D MOBILITY PREDICTION AND THE MISSING SECOND-ORDER TERM

D.1 UNIFORM CONSTANTS ON A CORRIDOR

Let $\mathcal{C}$ be a compact set with a compact neighborhood $\mathcal{C}^+ \subset \mathcal{D}$. When comparing different trajectories, enlarge it to contain the convex hull of $\mathcal{C}$ and a small neighborhood. This is possible in finite dimension because $\mathcal{D}$ is open and convex: a uniform ball around $\mathcal{C}$ lies in $\mathcal{D}$, and convex combinations of those balls give a neighborhood of its compact convex hull. We use constants

$$M_0 \geq \sup_{\mathcal{C}^+} \|\nabla E\|, \quad M_1 \geq \sup_{\mathcal{C}^+} \left\|\nabla^2 E\right\|_{\mathrm{op}}, \quad M_2 \geq \sup_{\mathcal{C}^+} \left\|\nabla^3 E\right\|_{\mathrm{op}},$$

and K_0, K_1, K_2 for the corresponding operator bounds on K, DK, D^2K. Multilinear operator norms are induced by the fixed Hilbert norm. In all expansions below, the relevant segments are required to lie in $\mathcal{C}^+$. This can be ensured by taking h small relative to its thickness and using the certified predictor bounds.

Lemma D.1 (Frozen-midpoint expansion)**.** *Let $f = F(x)$, $e_z = z - x - hf/2$, and $e_B = B - K(z)$. If $\|e_z\| = \zeta$ and $\|e_B\|_{\mathrm{op}} \leq \chi_K$, then*

$$B = K(x) + \frac{h}{2} DK(x)[f] + R_B, \quad \|R_B\|_{\mathrm{op}} \leq \chi_K + K_1\zeta + \frac{K_2}{2}(h\|f\|/2 + \zeta)^2. \tag{27}$$

In particular, for bounded $\zeta \leq c_z h^2$ and $\chi_K \leq c_K h^2$, the remainder is bounded by a constant times h^2.

Proof. Taylor's theorem in integral form gives $K(z) = K(x) + DK(x)[z-x] + \int_0^1 (1-s) D^2K(x + s(z-x))[z-x, z-x]\,\mathrm{d}s$. Insert $z - x = hf/2 + e_z$ and add e_B. Taking the operator norm proves the claim. □

The requirement on z is only a first-order approximation of the midpoint: its error is $O(h^2)$. A second-order endpoint predictor is not needed to construct B. This makes a certified midpoint much cheaper to supply than the final implicit state.

Lemma D.2 (Inexact evaluation of the initial vector field)**.** *Suppose $\widehat{f}$ is used in the midpoint predictor and $\left\|\widehat{f} - F(x)\right\| \leq e_F$. If $z = x + h\widehat{f}/2 + h^2 a$ with $\|a\| \leq c_a$, then $\zeta \leq h e_F/2 + c_a h^2$. Thus a sufficient initial-vector-field accuracy is $e_F = O(h)$.*

Proof. Subtract the target $x + hF(x)/2$ and apply the triangle inequality. □

One can demand a tighter initial oracle, but it is important not to demand one silently. Its error is multiplied by $h/2$ before it affects the midpoint. By contrast, the AVF residual enters the state equation multiplied by h and needs the $O(h^2)$ residual scale for second-order accumulated error.

D.2 AVF TAYLOR EXPANSION

Lemma D.3 (Expansion with an explicit remainder)**.** *For $\delta = y - x$ with its segment inside $\mathcal{C}^+$, put $p = \nabla E(x)$ and $H = \nabla^2 E(x)$. Then*

$$g(x, y) = p + \tfrac{1}{2} H\delta + R_g, \qquad \|R_g\| \leq \frac{M_2}{6}\|\delta\|^2. \tag{28}$$

Moreover, $\|g(x,y)\| \leq M_0$ and $\|g(x,y) - p\| \leq M_1\|\delta\|/2$.

Proof. Taylor expansion of $\nabla E(x + s\delta)$ has linear term $sH\delta$ and remainder bounded by $M_2 s^2 \|\delta\|^2 /2$. Integrating over s produces the factors $1/2$ and $1/6$. The other two bounds follow by integrating the zeroth-order and Lipschitz bounds. □

D.3 Why left freezing is insufficient

For comparison, set $B = K(x)$ and solve the frozen AVF equation exactly. The same expansion gives

$$y - x = hF(x) - \frac{h^2}{2} K(x)\nabla^2 E(x)F(x) + O(h^3).$$

The exact flow instead has second coefficient

$$\frac{1}{2}F'(x)F(x) = -\frac{1}{2}DK(x)[F(x)]\nabla E(x) - \frac{1}{2}K(x)\nabla^2 E(x)F(x). \tag{29}$$

The first term is missing unless the model has a special cancellation. Its presence is a statement about physical time, not about dissipation. Positivity of $K(x)$ still gives an exact energy inequality for the left-frozen method.

A midpoint depending implicitly on the endpoint, such as $B = K((x + y)/2)$, supplies the missing term but changes the nonlinear problem. Unless additional structure is available, the Jacobian of $B(y)^{-1}(y - x)/h + g(x, y)$ need not be symmetric, and it is not the Hessian in Lemma B.1. Our frozen prediction retains both the required Taylor coefficient and the tractable potential. We do not assert that implicit midpoint mobility is invalid; it simply needs a different certification argument.

D.4 A learned midpoint with an automatic scale

Let a_θ be a network satisfying a certified output bound $\|a_\theta(x, h)\| \le c_a$ on the deployment domain. Set

$$z_\theta = x + \frac{h}{2}F(x) + h^2 a_\theta(x, h).$$

The midpoint consistency gate holds with $c_z = c_a$, regardless of the training objective. A bounded output activation gives a simple coordinatewise bound, which must then be converted to the fixed Hilbert norm. In a mass metric this conversion includes the mass matrix; in covariance space it includes the Frobenius norm and the chosen symmetric representation.

This construction does not automatically preserve admissibility. A covariance correction can destroy positive definiteness, and a quantile correction can destroy increasing order. Either certify the resulting z_θ or use the explicit fallback for sufficiently small h. The output bound fixes consistency; an independent geometric check fixes admissibility. These two checks must not be conflated.

E Complete variable-step physical-time proof

E.1 Size of an accepted displacement

Assume the uniform corridor hypotheses in the main text, and let $b_+\mathrm{I} \succeq B \succeq b_-\mathrm{I}$. The residual equation is an exact identity even for an inexactly computed state:

$$\delta = -hBg + hr. \tag{30}$$

Because the chord is in the corridor, $\|g\| \le M_0$. Since $\|r\| \le \sqrt{b_+}\bar{\varepsilon}$, it follows directly that

$$\|\delta\| \le h(b_+ M_0 + \sqrt{b_+}\bar{\varepsilon}). \tag{31}$$

Thus $\delta = O(h)$ for accepted gates, with a uniform constant for $h \le h_0$. This estimate does not assume convergence to an exact root and does not invoke a contraction of the nonlinear solver. It uses the accepted endpoint and chord bounds.

E.2 One-step consistency with separately visible errors

Lemma E.1 (Local physical defect)**.** *Under the corridor assumptions and uniform accepted budgets, there is a mesh-independent C such that*

$$\left\|y - x - hF(x) - \tfrac{1}{2}h^2F'(x)F(x)\right\| \le Ch\big(h^2 + \bar{\varepsilon} + \zeta + \chi_K\big). \tag{32}$$

The constant depends on the corridor derivative and spectral bounds and on upper bounds for the accepted budget coefficients.

Proof. Use Lemmas D.1 and D.3. Write $K = K(x)$, $p = \nabla E(x)$, $H = \nabla^2 E(x)$, $f = -Kp$, and $J = DK(x)[f]$. Then

$$\delta = -h(K + hJ/2 + R_B)(p + H\delta/2 + R_g) + hr.$$

The product $-hKp$ equals hf. The bounds $\delta = O(h)$ and $R_g = O(h^2)$ show that

$$\|\delta - hf\| \le Ch\big(h + \bar{\varepsilon} + \zeta + \chi_K\big). \tag{33}$$

Indeed, $B - K = O(h + \zeta + \chi_K)$, $g - p = O(h)$, and the remaining term is hr. Here the quadratic term in ζ in (27) is bounded by a constant times ζ when ζ is uniformly bounded, as it is for accepted gates.

Now retain the two second-order terms $-h^2Jp/2$ and $-hKH\delta/2$. Replacing δ in the latter by hf costs at most $Ch\,\|\delta - hf\| \le C(h^3 + h^2[\bar{\varepsilon} + \zeta + \chi_K])$. The R_Bp term costs $Ch(h^2 + \zeta + \chi_K)$. The KR_g term costs Ch^3. The cross product of $hJ/2 + R_B$ with $H\delta/2 + R_g$ is bounded by the same right side after reducing h_0 if necessary. Finally, $\|hr\| \le h\sqrt{b_+}\bar{\varepsilon}$.

The two retained terms sum to $h^2F'(x)F(x)/2$ by (29). Combining the remainder bounds, and absorbing an extra factor $h \le h_0$ into the constant, gives (32). □

This proof reveals why two different smallness scales appear. A midpoint error affects the mobility and hence the velocity, so its contribution is $h\zeta$. A full-state residual also contributes $h\bar{\varepsilon}$. Both must be $O(h^3)$ per step to obtain a generic accumulated $O(h^2)$ bound.

Lemma E.2 (Exact-flow comparison)**.** *Let $\phi_h(x)$ denote the exact flow starting at a corridor state x. For sufficiently small h, it exists in a common neighborhood and*

$$\left\|\phi_h(x) - x - hF(x) - h^2F'(x)F(x)/2\right\| \le Ch^3.$$

Also $\|\phi_h(x) - \phi_h(v)\| \le e^{Lh}\,\|x - v\|$ for nearby corridor states whose comparison paths remain in that neighborhood, where L bounds $\|F'\|$.

Proof. Since $E \in C^3$ and $K \in C^2$, the vector field is C^2. Along a solution, the first three time derivatives are F, $F'F$, and $F''[F, F] + F'F'F$, all uniformly bounded on a compact neighborhood. The integral Taylor remainder is therefore bounded by a constant times h^3. The Lipschitz estimate follows from the difference equation and Gronwall's inequality. A finite cover of the compact corridor provides a common existence time and bounds. □

For global error comparison, it is enough that the exact flow starting from each numerical endpoint exists for the short step in the common neighborhood. It is not necessary to assume every such flow remains in the original energy sublevel at all future times.

E.3 Accumulation on an arbitrary positive mesh

Complete proof of Theorem 3.2. Set $e_n = \|x_n - u(t_n)\|$. Lemmas E.1 and E.2 imply

$$\|x_{n+1} - \phi_{h_n}(x_n)\| \le Ch_n(h_n^2 + \bar{\varepsilon}_n + \zeta_n + \chi_{K,n}).$$

Comparing the exact flows from x_n and $u(t_n)$ gives

$$e_{n+1} \le e^{Lh_n}e_n + Ch_n(h_n^2 + \bar{\varepsilon}_n + \zeta_n + \chi_{K,n}). \tag{34}$$

Iteration yields

$$e_n \le e^{Lt_n} e_0 + C \sum_{j<n} e^{L(t_n - t_{j+1})} h_j (h_j^2 + \bar{\varepsilon}_j + \zeta_j + \chi_{K,j}).$$

All exponential factors are bounded by e^{LT} after enlarging L to be nonnegative. This proves the main bound. With the accepted gates, the sum is at most $C(1 + \tau + c_z + c_K) \sum_{j<n} h_j^3 \le C(1 + \tau + c_z + c_K) T h_{\max}^2$. Only positivity of h_j and the total-time bound were used. There is no ratio condition between consecutive steps. □

A finite prefix of an adaptive computation is covered by this argument regardless of how the mesh was proposed. If infinitely many accepted steps accumulate before T, the theorem bounds every finite prefix but does not assert that the algorithm reaches T. Nonaccumulation of accepted times is an additional termination question addressed in Appendix P.

E.4 Objective budgets and the correct interpretation of order

If λ_B is uniformly bounded, $\beta_h = h^{-1} + O(1)$ and, for sufficiently small h, $c/h \le \beta_h \le C/h$ with positive constants. Therefore the residual gate implies

$$G_h^{\mathrm{up}} = \frac{\bar{\varepsilon}^2}{2\beta_h} \le C\tau^2 h^5.$$

This is a sufficient budget on the *computable upper certificate*. An assertion that the unknown true gap is $O(h^5)$ is different: strong convexity alone bounds the distance from a gap but does not upper-bound the gradient by that gap without a smoothness bound. On a corridor where $\nabla^2 Q \preceq L_Q B^{-1}$ and $L_Q = O(h^{-1})$, a true gap of order h^5 does imply a residual of order h^2 through the smoothness inequality proved in Appendix G. Without that extra premise, the paper uses the residual-based certificate, not a reversed implication.

Second order here is measured at $t_n = \sum_{j<n} h_j$. Replacing t_n by an effective or learned time after the computation changes the comparison problem. In particular, a trajectory that agrees with the target only after a nontrivial time reparameterization does not satisfy this physical-time conclusion.

E.5 How the corridor is supplied in the applications

For positive quantile finite elements with inverse-gap diffusion energy, accepted energy descent bounds every gap away from zero at fixed spatial resolution. The fixed interval bounds all positions. Chords remain in the same convex positive region. Midpoint certification places z within $O(h)$ of x, so a slightly larger region contains all midpoint proposals for small enough h.

For Gaussian entropy with $A \succ 0$, energy sublevels bound means and bound covariance eigenvalues above and away from zero. The SPD cone is convex, so chords have the same lower eigenvalue bound. The predicted covariance midpoint remains positive for small enough h by its $O(h)$ displacement. On these regions the Gaussian mobility is affine and uniformly positive. These facts verify the corridor without a condition that A and the covariance commute. Their constants still depend on dimension, energy level, and the confinement matrix.

F Continuous reconstruction and a computable trajectory majorant

An endpoint certificate measures distance to the exact frozen step. A trajectory certificate measures distance to the intended evolving solution. The two are different even when both are evaluated in the same norm. This section supplies the second object without a contraction assumption on the numerical step map.

F.1 A COMPARISON LEMMA WITH A ONE-SIDED CONSTANT

Lemma F.1 (Defect-to-trajectory estimate)**.** *Let $u' = F(u)$, and let U be continuous and piecewise continuously differentiable on $[0, T]$. Suppose that for all comparison states visited,*

$$\langle F(v) - F(w), v - w\rangle \le L_o \|v - w\|^2 . \tag{35}$$

If $R(t) = U'(t) - F(U(t))$ is integrable, then

$$\|U(t) - u(t)\| \le e^{L_o t} \|U(0) - u(0)\| + \int_0^t e^{L_o(t-s)} \|R(s)\| \, \mathrm{d}s. \tag{36}$$

The constant L_o may be positive, zero, or negative.

Proof. Let $e = U - u$ and $a_\epsilon = (\|e\|^2 + \epsilon^2)^{1/2}$. At differentiability times,

$$a'_\epsilon \le L_o \frac{\|e\|^2}{a_\epsilon} + \|R\| \frac{\|e\|}{a_\epsilon} \le L_o a_\epsilon + \|R\| + |L_o| \, \epsilon.$$

The last term bounds the difference between $L_o \|e\|^2 / a_\epsilon$ and $L_o a_\epsilon$ for either sign of L_o. Multiply by $e^{-L_o t}$ and integrate across the finitely many or locally finite differentiability intervals; continuity supplies no jump term. Let $\epsilon \downarrow 0$. The same proof extends to absolutely continuous U and almost-everywhere inequalities. □

A Lipschitz bound on F gives a valid L_o. A monotonicity estimate may give a smaller, possibly negative, value. The Gaussian and quantile coordinate flows need not have the same one-sided constant. In particular, dissipativity of the energy does not by itself prove (35) with $L_o \le 0$ in an arbitrarily chosen coordinate norm.

F.2 QUADRATIC RECONSTRUCTION FROM ACCEPTED STATES

For a slab $[t_n, t_n + h]$, write $s = (t - t_n)/h$, $x = x_n$, $y = x_{n+1}$, $\delta = y - x$, and $b = F(y) - F(x)$. Define

$$U(s) = x + s\delta + \tfrac{h}{2} s(s-1) b, \; R(s) \qquad = \delta/h + (s - \tfrac{1}{2}) b - F(U(s)). \tag{37}$$

The polynomial correction vanishes at both endpoints, so the reconstructions join continuously. Its purpose is to reconstruct the second Taylor coefficient along the slab rather than merely interpolate the endpoints linearly. Linear interpolation can have an $O(h)$ pointwise defect even for a second-order endpoint method.

Proposition F.2 (Order of the reconstruction defect)**.** *Under the hypotheses and gates of Theorem 3.2, the reconstruction remains in a common neighborhood for sufficiently small h, and*

$$\sup_{0 \le s \le 1} \|R(s)\| \le C\big(h^2 + \bar{\varepsilon} + \zeta + \chi_K\big). \tag{38}$$

Proof. Put $f = F(x)$ and $J = F'(x)$. Lemma E.1 gives $\delta = hf + h^2 Jf/2 + e_\delta$, where $\|e_\delta\| \le Ch(h^2 + \bar{\varepsilon} + \zeta + \chi_K)$. Taylor expansion of $F(y) - F(x)$ gives $b = J\delta + O(\|\delta\|^2) = hJf + O(h^2 + h[\bar{\varepsilon} + \zeta + \chi_K])$. Consequently

$$U'(t) = \delta/h + (s - \tfrac{1}{2}) b = f + shJf + O(h^2 + \bar{\varepsilon} + \zeta + \chi_K).$$

Also $U(s) = x + shf + O(h^2 + h[\bar{\varepsilon} + \zeta + \chi_K])$, uniformly in s. Taylor expansion of $F(U(s))$ produces the same leading terms $f + shJf$ and a remainder bounded by the displayed right side. Subtraction proves the defect bound. The reconstruction differs from its chord by at most $h\|b\|/8 = O(h^2)$; the corridor thickness therefore guarantees admissibility in a slightly larger neighborhood. □

This proposition covers variable frozen mobility. It does not use the generally false identity $Bg(x, y) = -\int_0^1 F(x + s\delta) \, \mathrm{d}s$ when K varies along the chord.

F.3 A FINITE EVALUABLE UPPER BOUND

A posterior certificate must replace the supremum in (38) by a computable bound. Let L_F, H_F be validated bounds on $\|F'\|$ and $\|F''\|$ over the reconstruction image. Set

$$M_R = H_F(\|\delta\| + \tfrac{h}{2}\|b\|)^2 + L_F h\|b\|. \tag{39}$$

Indeed, $U_s = \delta + h(s-1/2)b$, $U_{ss} = hb$, and $R_{ss} = -F''(U)[U_s, U_s] - F'(U)U_{ss}$, so $\|R_{ss}\| \le M_R$.

Lemma F.3 (Interpolation enclosure)**.** *Let $0 = s_0 < \cdots < s_M = 1$ and let r_i^+ bound $\|R(s_i)\|$. If M_R bounds the second derivative norm, then*

$$\sup_s \|R(s)\| \le \max_{0\le i<M} \left\{ \max(r_i^+, r_{i+1}^+) + \frac{M_R}{8}(s_{i+1} - s_i)^2 \right\}. \tag{40}$$

Proof. On one interval, subtract the linear interpolant of the vector-valued function R. For every unit vector v, the scalar projection of the interpolation error has absolute value at most $M_R(s - s_i)(s_{i+1} - s)/2$, by the scalar integral interpolation remainder. Taking the supremum over unit v gives the vector norm bound, maximized by $M_R(s_{i+1} - s_i)^2/8$. The norm of the interpolant is bounded by the larger endpoint norm because the norm is convex. □

The interpolation theorem requires an actual derivative enclosure, not merely observed smoothness between sample points. Since $M_R = O(h^2)$ for accepted small steps, even a fixed number of subintervals preserves an $O(h^2)$ defect upper bound when the nodal enclosures have comparable accuracy. More subintervals improve sharpness rather than the formal order.

F.4 ADMISSIBILITY THROUGHOUT A SLAB

For piecewise affine quantiles, each gap of $U(s)$ is a quadratic polynomial. Its minimum on $[0, 1]$ is the smaller endpoint value unless its leading coefficient is positive and its critical point lies in $(0, 1)$, in which case that critical value must also be checked. These are finite exact arithmetic operations for rational data.

For Gaussian covariance reconstructions, write $C_U(s) = (1-s)C_x + sC_y + hs(s-1)B_C/2$, where $B_C = F_C(y) - F_C(x)$. If $C_x, C_y \succeq \sigma_0 \mathrm{I}$, the sufficient test $\sigma_0 - h\|B_C\|_{\mathrm{op}}/8 > 0$ proves positivity throughout the slab. An interval subdivision with verified eigenvalue bounds can be sharper. Endpoint positive definiteness alone does not prove positive definiteness of the quadratic correction.

For smooth quantiles, a sufficient derivative test is $\inf_{a,s}[(1-s)X_x'(a) + sX_y'(a)] > h\|b'\|_\infty/8$. Alternatively, a bivariate polynomial or rational enclosure can verify the exact lower derivative bound. This check is part of the continuum certificate, not an optional graphical inspection.

F.5 PROPAGATION AND INTERPRETATION

If K_n denotes the right side of (40), Lemma F.1 gives

$$B_{n+1} = e^{L_o h_n} B_n + c_{-L_o}(h_n) K_n, \qquad c_{-L_o}(h) = \int_0^h e^{L_o s}\,\mathrm{d}s.$$

An initial upper bound B_0 then propagates to all accepted endpoints. Bounds at interior times use the corresponding partial integral. Every quantity in the recurrence, including an exponential evaluated in finite precision, needs an upper enclosure to obtain a deterministic computation.

The recurrence compares with the exact solution of the finite-state vector field. For a continuum statement, replace its defect by the full continuum residual, as done in Appendix K. Reusing a finite-state majorant while omitting the spatial defect would certify a different target.

G A FULL-STATE CORRECTOR, FINITE ACCEPTANCE, AND WORK BOUNDS

G.1 AN EXACT-ORACLE BACKTRACKING CONSTRUCTION

This section proves a concrete existence-of-correction result. It is not an assertion that every finite-budget Newton implementation converges. Work in the normalized coordinates $w = B^{-1/2}(y - x)$

from Appendix B, and write $q(w) = Q(x + B^{1/2}w)$. Suppose q is β-strongly convex on its open convex domain Ω and has an interior minimizer $w^\star$. Let the initial sublevel $S = \{w \in \Omega : q(w) \le q(w_0)\}$ be compactly contained in Ω.

Choose an Armijo parameter $a \in (0,1)$, a reduction factor $\gamma \in (0,1)$, and a trial step $\bar{s} > 0$. At iteration k, test $s = \bar{s}\gamma^j$, $j = 0, 1, \ldots$, until the state is admissible and

$$q(w_k - s\nabla q(w_k)) \le q(w_k) - as \left\|\nabla q(w_k)\right\|^2. \tag{41}$$

Then accept that correction. An exactly zero gradient ends the correction.

Theorem G.1 (Global convergence of the backtracking corrector)**.** *Under the preceding hypotheses, the line search terminates at every iteration. There is $s_{\min} > 0$, independent of k, such that all accepted correction steps have $s_k \ge s_{\min}$. Moreover,*

$$q(w_k) - q(w^\star) \le (1-\nu)^k [q(w_0) - q(w^\star)] \tag{42}$$

for some $\nu \in (0,1)$, and $w_k \to w^\star$.

Proof. Choose $\rho > 0$ so that the closed ρ-neighborhood of S is a compact subset of Ω. Bound $\|\nabla q\|$ on S by G and $\left\|\nabla^2 q\right\|$ on the neighborhood by L. If $G = 0$, the unique point reached is already the minimizer. Otherwise, $s \le \rho/G$ keeps the entire trial segment inside that neighborhood. Taylor's formula gives $q(w - s\nabla q(w)) \le q(w) - s(1 - Ls/2)\left\|\nabla q(w)\right\|^2$. Thus $s \le 2(1-a)/L$ suffices for Armijo, with the usual interpretation if $L = 0$. Backtracking therefore accepts a step at least

$$s_{\min} = \gamma \min\{\bar{s}, \rho/G, 2(1-a)/L\} > 0.$$

Every accepted step stays in S, so the same bounds apply at every iteration.

The strong-convexity gap bound gives $\left\|\nabla q(w_k)\right\|^2 \ge 2\beta[q(w_k) - q(w^\star)]$. Combining with Armijo yields a geometric decrease. To state a factor strictly between zero and one without a degenerate endpoint case, take $\nu = \min\{a\beta s_{\min}, 1/2\}$. The decrease is at least ν times the current gap. Finally, strong convexity bounds $\left\|w_k - w^\star\right\|^2 \le 2[q(w_k) - q(w^\star)]/\beta$, proving convergence. □

All evaluations in this theorem are exact. A verified line search with interval enclosures can implement the same strict tests away from equality by refining precision. At equality or unresolved stationarity it may return an unresolved status. A practical Newton method can replace this fallback, but it must have its own convergence assumptions or be safeguarded by a proven descent mechanism.

G.2 From objective contraction to residual contraction

Lemma G.2 (Gradient upper bound on a convex sublevel)**.** *Suppose S is a convex sublevel as above and $\left\|\nabla^2 q\right\| \le L_Q$ on S. Then*

$$\|\nabla q(w)\| \le L_Q \sqrt{2[q(w) - q(w^\star)]/\beta}, \qquad w \in S. \tag{43}$$

If, additionally, the standard gradient-descent trial $w - \nabla q(w)/L_Q$ is in a region where the L_Q smoothness inequality holds, the sharper bound $\left\|\nabla q(w)\right\|^2 \le 2L_Q[q(w) - q(w^\star)]$ holds.

Proof. The segment from $w^\star$ to w lies in S by convexity. Integrating the Hessian gives $\|\nabla q(w)\| \le L_Q \|w - w^\star\|$. Combine with the strong-convexity distance bound. For the sharper statement, apply the smoothness inequality to the stated trial and compare its objective value with the global minimum. □

The first inequality has the hypotheses needed on an open barrier domain and is sufficient for logarithmic work estimates. We do not automatically use the sharper whole-space inequality when its trial point might leave the admissible domain. Since both L_Q and β scale like h^{-1} on a fixed compact corridor, either valid form yields the same h exponent in the corresponding sufficient objective budget.

G.3 WHY NONSTATIONARY CORRECTION EVENTUALLY PASSES BOTH GATES

Lemma G.3 (Strict margin at the exact step)**.** *For an interior exact frozen step $y^\star$, if $\nabla E(x) \neq 0$, then $d_\star = \|g(x, y^\star)\|_B > 0$. If $\left\|B^{1/2}\nabla^2 E(v)B^{1/2}\right\| \leq L_B$ on the chord from x to $y^\star$, then*

$$d_\star \geq \frac{\|\nabla E(x)\|_B}{1 + hL_B/2}. \tag{44}$$

Proof. If $g(x, y^\star) = 0$, the exact equation gives $y^\star = x$, hence $g(x, x) = \nabla E(x) = 0$, a contradiction. For the quantitative statement, $\|g(x, y^\star) - \nabla E(x)\|_B \leq L_B \|y^\star - x\|_{B^{-1}}/2$. The exact equation gives $\|y^\star - x\|_{B^{-1}} = hd_\star$. Rearrange the triangle inequality. □

Proof of Proposition 3.4. Let $y_k \to y^\star$ and let the oracle errors $\sigma_k \to 0$. Continuity gives $g(x, y_k) \to g(x, y^\star)$, while $\|\widetilde{g}_k - g(x, y_k)\|_B \leq \sigma_k$ gives the same limit for $\widetilde{g}_k$. Thus $\widetilde{r}_k \to 0$, $\bar{\varepsilon}_k \to 0$, and $\widetilde{D}_k \to d_\star^2 > 0$. The left side of the work test tends to zero, including its oracle term because h is fixed and δ_k is bounded. Its right side tends to $\eta d_\star^2 > 0$. Both tests therefore hold strictly for all sufficiently large k. □

No uniform nonstationarity margin over all states was asserted. A sequence approaching equilibrium can require increasing precision or correction work even though every nonstationary state is individually covered.

G.4 A QUANTITATIVE SUFFICIENT STOPPING THRESHOLD

Let L_B also bound the frozen-normalized energy Hessian on the correction sublevel chords. Since the AVF gradient is $L_B/2$-Lipschitz in its second endpoint,

$$\|g(x, y) - g(x, y^\star)\|_B \leq (L_B/2)\, \|y - y^\star\|_{B^{-1}}\,.$$

Choose $q > 0$ such that $2q + q^2 < \eta$. Suppose the correction is close enough that $\|y - y^\star\|_{B^{-1}} \leq d_\star/(2L_B)$, interpreting this restriction as void if $L_B = 0$, and that $\sigma \leq d_\star/8$. Then $\|\widetilde{g}\|_B \geq d_\star - d_\star/4 - d_\star/8 \geq d_\star/2$. It is consequently sufficient to enforce

$$\|\widetilde{r}\|_{B^{-1}} \leq \min\{\tau h^2/2, qd_\star/2\}, \qquad \sigma \leq \min\{\tau h^2/2, qd_\star/2, d_\star/8\}. \tag{45}$$

Lemma C.3 proves the energy gate and the first terms prove the residual gate. If desired, use the lower bound in (44) instead of the unknown $d_\star$ to obtain a sufficient prescribed target. A validated lower bound for $\|\nabla E(x)\|_B$ is then required.

For a fully explicit sufficient gap target, set

$$s_* = \min\{\tau h^2/4, qd_\star/4\}, \qquad \Delta_{\mathrm{tar}} = \min\left\{\frac{\beta d_\star^2}{8L_B^2}, \frac{\beta s_*^2}{2L_Q^2}\right\},$$

where the first entry is interpreted as $+\infty$ when $L_B = 0$. A true objective gap at most Δ_{tar} gives the required state distance and true residual at most s_* by strong convexity and Lemma G.2. Refining the oracle to $\sigma \leq \min\{s_*, d_\star/8\}$ gives $\|\widetilde{r}\|_{B^{-1}} \leq 2s_*$ and hence (45). Consequently a corrector with factor $1 - \nu$ needs at most

$$\max\left\{0, \left\lceil \frac{\log(\Delta_0/\Delta_{\mathrm{tar}})}{-\log(1-\nu)} \right\rceil\right\}$$

iterations, where Δ_0 is any upper bound on its initial gap. This is logarithmic in the inverse residual target and inverse nonstationarity margin. The prefactor still contains conditioning through ν, and may deteriorate with spatial resolution. Since $L_Q, \beta \asymp h^{-1}$ on a fixed corridor, the residual part of this target has the familiar h^5 scale. A numerical implementation can use the residual certificate directly instead of evaluating this a priori sufficient bound.

G.5 Neural initialization and safeguarded work

A proposed neural state can lower the initial objective gap and hence the correction count, but the convergence theorem does not depend on that happening. If the proposal has a higher objective than a known admissible fallback, use the fallback. Comparing two validated objective intervals may certify which is smaller; overlapping intervals justify refining them or conservatively choosing the fallback. This safeguard guarantees that a poor learned initializer does not worsen the proven initial-sublevel bound, although inference itself still has a cost. The paper makes no empirical claim about the cost balance.

H Sharp scalar examples and failures of surrogate certificates

These examples are exact analytic constructions. They are not numerical experiments or refutations of qualified stability theorems in the cited literature. Each identifies a specific implication that cannot be used without additional hypotheses.

H.1 Dissipation with an incorrect clock

Take $E(x) = x^2/2$, $B = 1$, $g(x, y) = (x + y)/2$, and prescribe $r = -ch^p g$ with $c > 0$ and $p \geq 0$. The residual equation becomes

$$\frac{y - x}{h} = -(1 + ch^p)\frac{x + y}{2}, \qquad y = R_p(h)x, \quad R_p(h) = \frac{1 - a_h}{1 + a_h}, \quad a_h = \frac{h}{2}(1 + ch^p).$$

For every $h > 0$, $a_h > 0$ and $|R_p(h)| < 1$. Hence the original energy strictly decreases whenever $x \neq 0$. The signed residual work is $gr = -ch^p g^2 \leq 0$, so an energy-only gate accepts it even though the rate may be wrong.

For sufficiently small h, $a_h < 1$ and the real logarithm is available. The identity $\log[(1 - a)/(1 + a)] = -2(a + a^3/3 + a^5/5 + \cdots)$ gives

$$\frac{\log R_p(h)}{h} = -(1 + ch^p) - \frac{h^2}{12}(1 + ch^p)^3 + O(h^4(1 + ch^p)^5). \tag{46}$$

Along meshes $h = T/N$, the endpoint is $R_p(h)^N x_0$. For $p = 0$ it tends to $e^{-(1+c)T}x_0$, not $e^{-T}x_0$. For $p = 1$, its leading exponent error is $-cTh$, so for $x_0 \neq 0$ and $T > 0$ the trajectory error is $\Theta(h)$. For $p = 2$, the leading exponent error is $-(c+1/12)Th^2$, giving $\Theta(h^2)$. Thus energy decrease does not even imply consistency, and a residual of generic order h does not preserve order two.

The scalar frozen objective is a quadratic with Hessian $\beta_h = 1/h + 1/2$. Its exact residual equals its derivative, so

$$Q(y) - Q(y^\star) = \frac{r^2}{2\beta_h}. \tag{47}$$

On a nonzero finite-time trajectory, $|g|$ is bounded above and below for sufficiently small h. Therefore the exact gap is $\Theta(h^{2p+1})$. This makes the sufficient fifth-order gap scaling sharp in a worst-case class, while leaving room for specially canceling residual sequences.

H.2 An absolute residual gate can allow an energy increase

At the equilibrium $x = 0$, let $y = \alpha h^3$ for $\alpha > 0$. Then $g = y/2$ and $r = \alpha h^2 + \alpha h^3/2$. For any fixed $\tau > \alpha$, all sufficiently small steps satisfy $|r| \leq \tau h^2$. Nevertheless $E(y) - E(0) = \alpha^2 h^6/2 > 0$. The increase is small but has the wrong sign. Consequently the time-accuracy tolerance does not imply exact original-energy monotonicity near equilibrium.

H.3 Left-frozen mobility is generically first order

On $(0, \infty)$ take $E(x) = x$ and $K(x) = x$. The physical flow is $\dot{x} = -x$. Because $g = 1$, the exactly solved left-frozen method $B = x$ is $y = x - hx$. It stays positive and decreases E for $0 < h < 1$, but its amplification factor $1 - h$ is explicit Euler and has first-order global error.

The predicted midpoint $z = x - hx/2$ instead gives $B = z$ and $y = x - hx + h^2x/2$, the second Taylor polynomial of the exact flow. The comparison isolates the missing derivative-of-mobility term: there is no nonlinear-solve error or quadrature error in either method.

Likewise, using a fixed incorrect factor $B = cK$ in a constant-mobility quadratic flow gives an energy-dissipative approximation to $\dot{x} = -cx$, not to $\dot{x} = -x$. A successful solve with the wrong metric is not a certificate for the intended physical evolution.

H.4 BDF2 AND THE ORIGINAL ENERGY

For $\dot{x} = -x$ and step size $h = 3$, initialize BDF2 with backward Euler: $x_0 = 1$, $x_1 = 1/4$. The recurrence is $(3 + 2h)x_{n+1} = 4x_n - x_{n-1}$. It gives $x_2 = 0$ and $x_3 = -1/36$. Hence the original scalar energy $x^2/2$ increases from step two to step three. This example does not contradict stability in a modified two-step energy, convergence for small h, or the hypotheses of variational metric BDF2 results (Matthes & Plazotta, 2019; Gallouët et al., 2024). It only rules out an unconditional assertion of original-energy monotonicity for this unconstrained scalar recurrence.

H.5 EXTRAPOLATION CAN DESTROY AN ENERGY CERTIFICATE

For exact scalar AVF, $R(h) = (1 - h/2)/(1 + h/2)$. A Richardson endpoint formed from two half steps and one full step is $y_{\text{ext}} = [4R(h/2)^2 - R(h)]x/3$. Both constituent computations dissipate $x^2/2$ at every positive step size. At $h = 40$,

$$R(40) = -19/21, \quad R(20) = -9/11, \quad \frac{4R(20)^2 - R(40)}{3} = \frac{9103}{7623} > 1.$$

Thus the extrapolated endpoint increases the original energy. A time-error estimator may use the constituent states, but an extrapolated state cannot inherit their certificate. Our algorithm never accepts such a state without rechecking it.

H.6 A SMALL PARAMETER GRADIENT IS NOT A FULL-STATE CERTIFICATE

Consider the convex state objective $Q(y) = (y - 1)^2/2$ and parameterization $y_\theta = \theta^2$. At $\theta = 0$, $\frac{\mathrm{d}}{\mathrm{d}\theta} Q(y_\theta) = 2\theta(\theta^2 - 1) = 0$. However, $\nabla Q(0) = -1$ and the global state-space gap is $1/2$. The parameter derivative has lost the direction in which the state should move. Redundant or rank-deficient neural parameterizations can exhibit precisely this failure. A lower bound on the relevant Jacobian singular values would be additional information; it cannot be inferred from the small parameter gradient itself.

Even a globally optimal restricted model may fail to approximate the full-state step. If a model outputs only $y = 0$ in the same example, its restricted optimization error is zero while its full-state gap is positive. Training convergence and representation error are separate obligations.

H.7 FINITE EVALUATIONS CANNOT CERTIFY AN UNRESTRICTED FIELD

Proposition H.1 (Deterministic finite-query obstruction)**.** *No deterministic algorithm using only finitely many point values and finitely many derivative values of an otherwise unrestricted smooth scalar residual on* $(0, 1)$ *can certify a finite universal upper bound on its* L^2 *norm from those observations alone.*

Proof. Run the algorithm on the identically zero residual and record the finite set of query points. There is an open interval avoiding every point. Choose a nonzero smooth compactly supported bump ψ in that interval. For every amplitude $M > 0$, the residual $M\psi$ and all its derivatives agree with zero at every recorded point. The deterministic adaptive query path is therefore unchanged, by induction on the sequence of observations. Yet its L^2 norm is $M \|\psi\|_{L^2}$, which is unbounded as $M \to \infty$. A bound depending only on those observations cannot be universally valid. □

The obstruction disappears only after adding information, such as a derivative envelope, a finite-dimensional function class with validated coefficients, or an explicit probability model with fresh validation. It does not say that useful empirical estimates are impossible; it says that they are not unrestricted deterministic certificates.

H.8 A pseudoinverse can hide an invariant violation

Let $B = \mathrm{diag}(1, 0)$ and $E(x_1, x_2) = x_1^2/2$. The intended dynamics leaves x_2 unchanged. A proposed step with residual $r = (0, 1)$ has zero seminorm $r^T B^\dagger r$ but changes x_2 by h. Thus a pseudoinverse residual alone misses drift in the kernel. A valid degenerate-mobility extension must enforce the invariant affine leaf separately, as in Appendix T.

I Positive quantile models and exact transport geometry

I.1 A density representation on a nonuniform mass grid

Fix $0 = a_0 < a_1 < \cdots < a_N = 1$, with $w_i = a_{i+1} - a_i > 0$. Let $q_0 = -L$, $q_N = L$, and let the interior vector q belong to

$$\mathcal{D}_N = \{(q_1, \ldots, q_{N-1}) : -L < q_1 < \cdots < q_{N-1} < L\}.$$

Define X_q by affine interpolation of (a_i, q_i). On (a_i, a_{i+1}), its slope is d_i/w_i, where $d_i = q_{i+1} - q_i > 0$. Pushing uniform mass on $(0, 1)$ through X_q gives the density

$$\rho_q(x) = \frac{w_i}{d_i}, \qquad q_i < x < q_{i+1}. \tag{48}$$

There is no point mass at the nodes, because the reference mass of every individual node is zero. Total mass is $\sum_i w_i = 1$, and positivity is a representation property. In particular, evaluating $\int \rho_q^2$ is not evaluating a diffusion energy on an empirical atomic measure, where that integral would not have the same meaning.

Lemma I.1 (Quantile transport identity)**.** *For any two increasing quantiles X, Y with finite second moments, $W_2(X_\# \mathrm{Unif}, Y_\# \mathrm{Unif})^2 = \int_0^1 |X(a) - Y(a)|^2 \, \mathrm{d}a$. For the present common finite-element mass grid, this equals $(q - p)^T A (q - p)$, where*

$$A_{jj} = (w_{j-1} + w_j)/3, \qquad A_{j,j+1} = A_{j+1,j} = w_j/6. \tag{49}$$

All other entries are zero, and $A \succ 0$ on the interior coordinates.

Proof. For strictly increasing continuous quantiles, the coupling $(X(a), Y(a))$ is the graph of the increasing map $T = Y \circ X^{-1}$. Let ψ be a convex primitive of T. Fenchel's inequality $xy \le \psi(x) + \psi^*(y)$ bounds the expected product for every coupling with these marginals. Equality holds on $y = T(x)$, so this coupling maximizes the product and minimizes $\mathbb{E}|x - y|^2$, whose other two terms depend only on the marginals. Approximation by strictly increasing quantiles and convergence in L^2 gives the general finite-second-moment assertion; alternatively the same argument uses a monotone subgradient when quantiles have flat pieces.

For the finite-element statement, write $X_q - X_p = \sum_{j=1}^{N-1} (q_j - p_j)\phi_j$, with endpoint-zero hat functions ϕ_j. The integral of ϕ_j^2 over its two cells is $(w_{j-1} + w_j)/3$, and the integral of $\phi_j \phi_{j+1}$ is $w_j/6$. No nonadjacent supports overlap. The resulting Gram matrix is positive definite because the hat functions are linearly independent. □

The endpoint-zero qualification is important. The fixed endpoint values contribute affine terms to a quadratic confinement energy, but they do not become optimization variables. In particular, the mass metric in the step equation is the interior Gram matrix.

I.2 Energy and its fixed-metric gradient

For $\kappa > 0$,

$$\kappa \int_{-L}^{L} \rho_q(x)^2 \, \mathrm{d}x = \kappa \sum_i \frac{w_i^2}{d_i}, \qquad \int V(x) \rho_q(x) \, \mathrm{d}x = \int_0^1 V(X_q(a)) \, \mathrm{d}a.$$

Let D be the incidence matrix mapping interior variations to gap variations; fixed endpoints contribute only an affine offset to the gaps. The coordinate diffusion gradient and Hessian are

$$D^T \left(-\frac{\kappa w_i^2}{d_i^2} \right)_i, \qquad D^T \, \mathrm{diag} \left(\frac{2\kappa w_i^2}{d_i^3} \right)_i D. \tag{50}$$

The Hilbert gradient in the mass inner product is A^{-1} times the first expression. Therefore the standard semidiscrete Wasserstein gradient system is $\dot{q} = -A^{-1}\nabla_{\mathrm{coord}} E_N(q)$. It is $\dot{q} = -\nabla E_N(q)$ when the gradient notation already refers to the mass Hilbert space. These are the same equation, not two different metrics.

If $V'' \geq a$, then for an interior variation v,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \int_0^1 V(X_q + tX_v)\,\mathrm{d}a\bigg|_{t=0} = \int_0^1 V''(X_q) X_v^2\,\mathrm{d}a \geq a\,\|v\|_A^2 .$$

The diffusion Hessian is positive semidefinite, so this gives the claimed mass-metric lower curvature. The proof includes nonuniform cells and does not replace the exact mass matrix by a diagonal approximation.

I.3 Mean coupling across several marginals

Let there be d components, possibly with different mass grids and interval lengths. Use the direct sum of their mass Hilbert spaces. For each component let $m_j = \int_0^1 X_{q_j}(a)\,\mathrm{d}a$, and define

$$E(q) = \sum_{j=1}^{d} \left[\kappa_j \sum_i \frac{w_{j,i}^2}{d_{j,i}} + \int_0^1 V_j(X_{q_j})\,\mathrm{d}a \right] + \Phi(m_1, \ldots, m_d). \tag{51}$$

The mean map has linear variation $Mv = (\int X_{v_j})_j$. Cauchy–Schwarz on a unit-mass interval gives $\|Mv\|_{\mathbb{R}^d} \leq \|v\|_{\mathsf{H}}$. If $\nabla^2\Phi \succeq -\ell_\Phi \mathrm{I}$, then

$$\left\langle \nabla^2(\Phi \circ m)v, v \right\rangle \geq -\ell_\Phi \|Mv\|^2 \geq -\ell_\Phi \|v\|_{\mathsf{H}}^2 .$$

Thus $\nabla^2 E \succeq (\min_j a_j - \ell_\Phi)\mathrm{I}$ in the fixed direct-sum mass metric. For $\Phi(m) = b\,\|m\|^4/4 - c\,\|m\|^2/2$, $b, c \geq 0$,

$$\nabla^2\Phi(m) = b(\|m\|^2 \mathrm{I} + 2mm^T) - c\mathrm{I} \succeq -c\mathrm{I}.$$

The coupling can therefore be nonconvex while the frozen AVF subproblem remains globally strongly convex for sufficiently small h.

I.4 Product measures and the associated continuum model

Lemma I.2 (Tensorization of squared transport distance)**.** *For product measures $\mu = \bigotimes_j \mu_j$ and $\nu = \bigotimes_j \nu_j$ on Euclidean product space with squared product distance, $W_2(\mu,\nu)^2 = \sum_j W_2(\mu_j,\nu_j)^2$.*

Proof. Every joint coupling induces a coupling of each coordinate pair. The total expected squared distance is the sum of the coordinate costs, each bounded below by its optimal transport cost. Conversely, the product of optimal coordinate couplings is a joint coupling with cost equal to the sum. This proves equality. □

The corresponding continuum marginal energy on a joint measure is

$$\mathcal{E}(\mu) = \sum_j \left[\kappa_j \int \rho_j^2 + \int V_j \rho_j \right] + \Phi(m),$$

where ρ_j are the marginal densities. Its first variation is a sum of coordinate functions, $\sum_j [2\kappa_j \rho_j(x_j) + V_j(x_j) + \partial_j\Phi(m)x_j]$. Thus its smooth Wasserstein velocity has coordinate j equal to $-2\kappa_j \partial_{x_j}\rho_j - V_j'(x_j) - \partial_j\Phi(m)$, depending only on x_j and time-dependent marginal quantities. A smooth coordinatewise flow map sends a product initial law to a product law, even though the means interact through time. Equivalently, the product of smooth marginal solutions satisfies the joint continuity equation with that velocity. Uniqueness of such a smooth flow, when used to identify the evolution, is an additional standard well-posedness premise.

This is a coupled high-dimensional distributional model, but not a claim about arbitrary correlations or joint-density diffusion $\int \rho(x)^2\,\mathrm{d}x$. The latter has a different first variation and does not reduce to (51). The distinction prevents a separable representation theorem from being overstated as a general high-dimensional nonlinear diffusion theorem.

J Quantile step globality, closed oracles, and sublevel control

J.1 The inverse-gap integral and its potential

For a fixed positive initial gap a and final gap b, direct integration gives

$$\int_0^1 \frac{1}{[(1-s)a+sb]^2}\,\mathrm{d}s = \frac{1}{ab}. \tag{52}$$

For $a \neq b$, this follows by integrating the derivative of $[(1-s)a+sb]^{-1}$; for $a = b$, both sides equal a^{-2}. Thus the AVF diffusion gradient in coordinate form is $D^T(-\kappa w_i^2/[d_i(x)d_i(y)])_i$.

The AVF potential contribution of one gap is

$$\int_0^1 \frac{\kappa w^2/[(1-s)a+sb] - \kappa w^2/a}{s}\,\mathrm{d}s = -\frac{\kappa w^2}{a}\log\frac{b}{a}. \tag{53}$$

Indeed, the integrand simplifies to $-\kappa w^2(b-a)/(a[a+s(b-a)])$, whose antiderivative gives the formula. Its first derivative in b is $-\kappa w^2/(ab)$, exactly the AVF force. Its second derivative is $\kappa w^2/(ab^2) > 0$. Unlike a numerical quadrature rule, these formulas hold without a step-size or gap-ratio approximation.

J.2 Quadratic confinement and quartic coupling

For $V_j(x) = a_j x^2/2 + \ell_j x + c_j$, integration of the quadratic energy along a chord is exact. In fixed Hilbert notation, a quadratic-affine energy $E_c(q) = \frac{1}{2}\langle q, H_c q\rangle + \langle b_c, q\rangle + c_c$ contributes

$$\langle H_c x + b_c, \delta\rangle + \frac{1}{4}\langle \delta, H_c\delta\rangle$$

to the AVF potential and $H_c(x+y)/2 + b_c$ to its AVF gradient. Boundary terms are contained in b_c after expressing the full quantile in interior coordinates.

For the quartic mean coupling, let $m = m(x)$ and $d = m(y) - m(x)$. Since $\nabla\Phi(m) = b\,\|m\|^2 m - cm$, its averaged gradient is

$$\begin{aligned}\int_0^1 \nabla\Phi(m+sd)\,\mathrm{d}s = b\big[\,\|m\|^2 m + \tfrac{1}{2}(\|m\|^2 d + 2\langle m, d\rangle\, m) \\ + \tfrac{1}{3}(2\langle m, d\rangle\, d + \|d\|^2 m) + \tfrac{1}{4}\|d\|^2 d\big] - c(m + d/2).\end{aligned} \tag{54}$$

Applying the adjoint of the mean map gives the state gradient. The identity is obtained by multiplying a quadratic polynomial by a linear polynomial and integrating each power of s. It supplies a rational oracle whenever coefficients and states are rational. The potential itself is a quartic polynomial in d and can be obtained by integrating $[\Phi(m+sd) - \Phi(m)]/s$ termwise. There is no uncomputed integral in either object.

J.3 Existence of a global positive step

Theorem J.1 (Global quantile minimizer)**.** *For the coupled energy* (51) *with positive diffusion coefficients, smooth coupling bounded on the finite box of means, and quadratic confinement, fix $x \in \mathcal{D}$ and $B \succ 0$. Suppose the global curvature bound yields $\beta_h > 0$. Then $Q_{h,x}^B$ has a unique global minimizer in $\mathcal{D}$. The theorem applies in particular to the quartic coupling above.*

Proof. The closure of the ordered coordinate domain is compact. The kinetic contribution is continuous and nonnegative on this closure. Quadratic confinement and the smooth mean-coupling potential are bounded there, because all chord states and means lie in compact boxes. By (53), if any final gap tends to zero while its initial gap remains fixed positive, its potential contribution tends to $+\infty$. The other logarithmic contributions are bounded below because all final gaps have a finite upper bound. Thus every finite sublevel of Q is compactly contained in $\mathcal{D}$.

Choose a minimizing sequence with values below $Q(x) + 1$. Compactness of that sublevel gives a convergent subsequence with an interior limit. Continuity gives attainment at the limit. The global strong-convexity bound from Lemma B.2 proves uniqueness and globality. Interior attainment also gives the exact frozen-step equation through Lemma B.1. □

For an energy lower curvature $m = \min_j a_j - \ell_\Phi$ and verified $b_-\mathrm{I} \preceq B \preceq b_+\mathrm{I}$, an explicit valid frozen constant is $\lambda_B = mb_-$ when $m \geq 0$, and $\lambda_B = mb_+$ when $m < 0$. This gives a directly checkable positive-curvature step condition.

The fixed initial gaps matter in the barrier coefficient. This existence proof is valid for each positive x and does not provide a uniform lower gap over all initial states. Nor does it say that a fixed-size step has $\beta_h > 0$ for an arbitrarily strong nonconvex coupling. The curvature condition remains explicit.

J.4 Accepted energy sublevels and a corridor

Let $E(q) \leq E_0$ and suppose the sum of all nondiffusion terms is bounded below by C_{low} on the finite box. The diffusion terms are nonnegative, so

$$0 < \frac{\kappa_j w_{j,i}^2}{d_{j,i}(q)} \leq E_0 - C_{\mathrm{low}}.$$

If the sublevel is nonempty, the right side is positive. Therefore

$$d_{j,i}(q) \geq \frac{\kappa_j w_{j,i}^2}{E_0 - C_{\mathrm{low}}}. \tag{55}$$

Because a convex combination of two gaps with this bound has the same bound, accepted endpoints and chords remain in a compact positive-gap region. All energy derivatives are bounded there. Their bounds can be computed from the incidence matrix, cell weights, and powers of the lower gap bound in (50).

At fixed resolution, (55) verifies the energy part of the temporal corridor. As the mass grid is refined, $w_{j,i}^2$ shrinks, and this elementary bound can deteriorate. A theorem uniform in resolution would require stronger density regularity or a different estimate. We do not suppress this dependence.

The same barrier argument makes Q sublevels compact, so the safeguarded corrector in Appendix G applies. In the mass-normalized coordinates its Hessian is positive definite, with explicitly bounded eigenvalues on each sublevel. This proves an implementable fallback, not merely existence of a favorable endpoint.

J.5 Exact distributional radius and model identity

If $B = \mathrm{I}$ as an operator in the mass Hilbert space, the radius in Theorem 2.1 is exactly W_2 for one marginal and the product W_2 for the coupled representation. For general frozen B with $B \preceq b_+\mathrm{I}$, it yields

$$W_2(\rho_y, \rho_{y^\star}) \leq \sqrt{b_+}\,\frac{\bar{\varepsilon}}{\beta_h}.$$

This is a legitimate distributional error bound for the states even when the target mobility is a non-identity generalized mobility. It does not identify that generalized flow with the standard Wasserstein gradient flow. The latter identification is justified here by the mass-metric formulation only when the target operator is the identity.

J.6 A tridiagonal-plus-low-rank structure

For the single-component quadratic-confinement model with $B = \mathrm{I}$, the coordinate Hessian of Q is a mass matrix term plus $D^T \operatorname{diag}(\kappa w_i^2/[d_i(x)d_i(y)^2])D$. Both are tridiagonal. Across components, a mean coupling adds a matrix whose rank is at most the number of components, since it factors through the mean map. Thus a state-space correction is a structured linear algebra problem despite the nonlinear density representation. This observation establishes the availability of a tractable certificate kernel; it is not a measured runtime advantage over another algorithm.

K CONTINUUM DIFFUSION: GEOMETRY, BOUNDARY TERMS, AND POSTERIOR CONTROL

K.1 THE SMOOTH QUANTILE EQUATION

Consider the density energy $\mathcal{F}(\rho) = \kappa \int_{-L}^{L} \rho^2 \,\mathrm{d}x + \int_{-L}^{L} V\rho \,\mathrm{d}x$. Let $X(t,a)$ be a strictly increasing quantile, so $\rho(t, X(t,a)) = 1/X_a(t,a)$. Changing variables $x = X(t,a)$ gives (16). For a smooth variation H with $H(0) = H(1) = 0$,

$$D\mathcal{E}(X)[H] = -\kappa \int_0^1 X_a^{-2} H_a \,\mathrm{d}a + \int_0^1 V'(X) H \,\mathrm{d}a \tag{56}$$

$$= \int_0^1 \left[\kappa \partial_a (X_a^{-2}) + V'(X)\right] H \,\mathrm{d}a. \tag{57}$$

The boundary term vanishes because the quantile endpoints are fixed. Therefore its L^2 gradient is $\mathcal{G}(X) = \kappa \partial_a (X_a^{-2}) + V'(X)$ and the smooth equation is

$$X_t = 2\kappa X_{aa}/X_a^3 - V'(X). \tag{58}$$

This is equivalent, wherever the derivatives and change of variables are valid, to $\rho_t = \kappa \partial_{xx}(\rho^2) + \partial_x(\rho V')$ with no flux at the endpoints. One way to see the equivalence is to differentiate the cumulative-distribution identity along a mass label. The material velocity is $-\partial_x(2\kappa\rho + V)$; using $\partial_x \rho = -X_{aa}/X_a^3$ gives (58).

A fixed-endpoint smooth solution must satisfy the corresponding boundary compatibility of its velocity. We assume such a solution exists on the time interval of comparison. The theorem below does not establish existence, uniqueness in every weak class, instantaneous regularization, or a positive lower density bound for arbitrary initial data. These would be separate analytic results.

K.2 PRECISE REGULARITY SUFFICIENT FOR THE COMPARISON

A convenient sufficient class consists of quantiles X, U continuous in time with values in $H^2(0,1)$, absolutely continuous in L^2 on each time slab, with common fixed endpoints and with X_a, U_a bounded above and below by positive constants on each finite comparison interval. Suppose $V \in C^2([-L,L])$, $V'' \geq \lambda$, and the time derivatives and gradients needed below lie in $L_t^1 L_a^2$. These hypotheses guarantee that the integration by parts and the difference equation hold in L^2 for almost every time. Stronger classical smoothness is also sufficient.

The lower derivative bound is an admissibility and regularity premise, not a bound inferred from finitely many unguarded sample values. For a computed reconstruction it can be checked by the interval constructions in Appendix L. The exact solution's regularity remains an explicit premise of the theorem.

Lemma K.1 (Monotonicity of the full quantile operator)**.** *For admissible U, X with the same endpoints,*

$$\langle \mathcal{G}(U) - \mathcal{G}(X), U - X \rangle_{L^2} \geq \lambda \left\| U - X \right\|_{L^2}^2 . \tag{59}$$

Proof. Let $H = U - X$, so H has zero trace. The diffusion contribution, after integration by parts, is

$$-\kappa \int_0^1 (U_a^{-2} - X_a^{-2})(U_a - X_a) \,\mathrm{d}a \geq 0,$$

because the scalar function $p \mapsto p^{-2}$ is decreasing for $p > 0$. The potential contribution satisfies $(V'(U) - V'(X))(U - X) \geq \lambda |U - X|^2$ pointwise by the integral mean-value formula. Integrating proves the result. Approximation by smooth functions, or the Sobolev integration-by-parts formula under the stated regularity, justifies the same calculation in the sufficient class above. □

The sign of λ need not be positive. A semiconvex potential gives an exponentially growing comparison factor, but still yields a valid finite-time posterior estimate.

Theorem K.2 (Total continuum posterior certificate)**.** *Let $X_t = -\mathcal{G}(X)$ and let U be an admissible continuous reconstruction in the preceding class. Define $\mathcal{R} = U_t + \mathcal{G}(U)$. If $r(t) \geq \|\mathcal{R}(t)\|_{L^2}$ almost everywhere, then*

$$W_2(\rho_{U(t)}, \rho_{X(t)}) \leq e^{-\lambda t} e_0 + \int_0^t e^{-\lambda(t-s)} r(s)\,\mathrm{d}s, \qquad e_0 \geq \|U(0) - X(0)\|_{L^2}. \tag{60}$$

For slabwise bounds $r(t) \leq b_n$ on $[t_n, t_{n+1}]$, the recurrence

$$e_{n+1}^{\mathrm{up}} = e^{-\lambda h_n} e_n^{\mathrm{up}} + c_\lambda(h_n) b_n, \qquad c_\lambda(h) = \begin{cases} (1 - e^{-\lambda h})/\lambda, & \lambda \neq 0, \\ h, & \lambda = 0, \end{cases} \tag{61}$$

provides endpoint bounds, with analogous partial-slab bounds.

Proof. The difference equation is $(U - X)_t = -[\mathcal{G}(U) - \mathcal{G}(X)] + \mathcal{R}$. Apply Lemma K.1 and the regularized-norm argument from Lemma F.1, now in L^2. The proof uses only inner products and absolutely continuous norms, so it remains valid in this Hilbert space under the stated domain conditions. It gives the integral bound in L^2. Lemma I.1 identifies that norm with W_2. Integrating a constant upper bound on each slab gives (61). □

This is a certificate against the continuum evolution, rather than against a reference numerical solution or the next implicit endpoint. Its validity is independent of the algorithm that constructed U, provided the admissibility and full-residual premises are verified.

K.3 Coupled continuum marginals

For the coupled product model, define $\mathcal{G}_j(X) = \kappa_j \partial_a(X_{j,a}^{-2}) + V_j'(X_j) + \partial_j \Phi(m)$. For two states, the mean-coupling contribution to the monotonicity pairing is

$$\langle \nabla\Phi(m(U)) - \nabla\Phi(m(X)), m(U) - m(X) \rangle \geq -\ell_\Phi \|m(U) - m(X)\|^2.$$

The mean map has norm at most one, so the total operator is $(\min_j a_j - \ell_\Phi)$-monotone in the direct-sum L^2 norm. The previous proof then gives a product-Wasserstein posterior bound with that constant, by Lemma I.2. This extension includes the nonlinear quartic mean coupling without pretending that a general joint density has a product quantile representation.

K.4 What the theorem does not hide

If only $U_t + P\mathcal{G}(U)$ is evaluated, one has not evaluated $\mathcal{R}$. The missing term may be large even when the neural or Galerkin time step is solved exactly. Likewise, if U is only piecewise affine in the mass variable, $\partial_a(U_a^{-2})$ can contain interface distributions, and $\mathcal{G}(U)$ need not be an L^2 function. The theorem cannot be applied to that reconstruction without a smoothing or a different weak residual norm and comparison argument. The next appendix supplies a smooth finite-dimensional route and explicitly retains the spatial term.

L Smooth representations and an explicit spatial-defect budget

L.1 A smooth affine approximation space

Let $X_{\mathrm{ref}}(a) = -L + 2La$ and let V_N be a finite-dimensional subspace of $C^2([0,1])$ with zero endpoint traces. Work in the affine space $X_{\mathrm{ref}} + V_N$ and the open convex domain

$$\mathcal{D}_N^{\mathrm{sm}} = \{X_{\mathrm{ref}} + v : v \in V_N,\ \min_{a \in [0,1]} (X_{\mathrm{ref}} + v)'(a) > 0\}.$$

The fixed Hilbert metric is the L^2 Gram matrix of the chosen basis of V_N. Its orthogonal projector is P_N. Restricting $\mathcal{E}$ to this space gives Hilbert gradient $P_N\mathcal{G}(X)$, because the first variation paired with any basis variation is (57). The semidiscrete gradient flow is therefore $X_t = -P_N\mathcal{G}(X)$.

The domain is convex because positivity of the derivative is preserved by convex combinations. Its openness follows from norm equivalence in a finite-dimensional smooth space. A lower potential

curvature $V'' \geq \lambda$ gives the same lower Hessian bound for the restricted energy. For small h, the local branch proposition gives an interior AVF step from any fixed admissible state. If $\beta_h > 0$, that root is the global minimizer within this smooth domain. We do not claim a barrier divergence for every possible boundary derivative degeneracy in every smooth basis.

L.2 Orthogonal decomposition of the full residual

Proposition L.1 (Exact temporal–spatial residual split)**.** *Let $U(t) \in X_{\mathrm{ref}} + V_N$ be absolutely continuous, with $U_t \in V_N$ and $\mathcal{G}(U) \in L^2$. Then*

$$\|U_t + \mathcal{G}(U)\|_{L^2}^2 = \|U_t + P_N\mathcal{G}(U)\|_{L^2}^2 + \|(\mathrm{I} - P_N)\mathcal{G}(U)\|_{L^2}^2 . \tag{62}$$

Proof. Decompose $\mathcal{G}(U)$ into its orthogonal projection and orthogonal complement. The first summand, together with U_t, lies in V_N; the second lies in $V_N^{\perp}$. The Pythagorean identity proves the formula. □

The orthogonality is lost if a collocation projector or an unverified empirical Gram matrix is used in place of the exact L^2 projector. In that case, one may use a triangle inequality plus projector-error bounds, but not the exact squared sum.

Corollary L.2 (A conditional total second-order conclusion)**.** *Suppose a family of admissible reconstructions has $e_0 \leq C_0 h_{\max}^2$ and, on each slab,*

$$\|U_t + P_N\mathcal{G}(U)\|_{L^2} \leq C_t h_{\max}^2, \qquad \|(\mathrm{I} - P_N)\mathcal{G}(U)\|_{L^2} \leq C_s h_{\max}^2,$$

with constants uniform over that family. Then its total continuum Wasserstein error on $[0, T]$ is at most $[e^{|\lambda|T}C_0 + Te^{|\lambda|T}(C_t^2 + C_s^2)^{1/2}]h_{\max}^2$.

Proof. Use Proposition L.1 to bound the full residual by $(C_t^2 + C_s^2)^{1/2}h_{\max}^2$ and substitute into Theorem K.2. The simple bound $e^{-\lambda(t-s)} \leq e^{|\lambda|T}$ gives the stated constant. □

The corollary is intentionally conditional on a spatial residual budget. It does not assert that an arbitrary fixed space, or an arbitrary refinement schedule, meets that budget. A posterior adaptivity can use the evaluated spatial term to decide whether refinement is necessary. Establishing that a particular refinement procedure terminates with dimension-independent constants is another theorem, not a consequence of temporal order alone.

L.3 Positive polynomial quantiles

A useful constructive class uses a polynomial derivative. Let $p(a) = \sum_{i=0}^{m} c_i \binom{m}{i} a^i (1-a)^{m-i}$ with $c_i > 0$, and define

$$X(a) = -L + 2L\frac{\int_0^a p(b)\,\mathrm{d}b}{\int_0^1 p(b)\,\mathrm{d}b}. \tag{63}$$

Then $X(0) = -L$, $X(1) = L$, and $X'(a) \geq 2L \min_i c_i / \int_0^1 p > 0$ because the Bernstein basis is nonnegative and sums to one. This gives an admissible smooth proposal with a directly verifiable derivative lower bound. Rational positive coefficients give a polynomial with rational coefficients after normalization.

There are two distinct optimization domains. The set of all polynomial quantiles of a prescribed degree with positive derivative is an open convex domain in its affine coefficients. The subset represented by positive Bernstein coefficients is a sufficient-positivity parameterization and can be smaller. A minimum over that smaller parameter set need not be the full-state minimum. The state-space certificate is therefore evaluated in the full affine polynomial variation space, or else the result must explicitly be a constrained-domain certificate with the appropriate normal-cone condition. We use the former convention.

Positive Bernstein coefficients are also sufficient but not necessary for a polynomial to be positive at a given degree. Subdivision can tighten their range enclosure. For a strictly positive polynomial on a compact interval, sufficiently fine subdivision yields positive lower Bernstein bounds: uniform

continuity and the convergence of subdivided control bounds to the function range justify this fact. An implementation can therefore refine until positivity is certified, while retaining an unresolved output if a prescribed work budget is exhausted or the function touches zero.

L.4 Computing the projected and full residuals

For polynomial V, polynomial quantile U with $U_a \geq \sigma > 0$, and polynomial time reconstruction, the continuum gradient

$$\mathcal{G}(U) = -2\kappa U_{aa}/U_a^3 + V'(U)$$

is a rational function of the mass and slab variables with denominator bounded away from zero. Its squared residual and its inner products against basis functions can be enclosed by deterministic interval quadrature on $[0,1]^2$. The projection coefficients solve the positive Gram system $Ac = (\langle \mathcal{G}(U), \phi_j \rangle)_j$. A rational Gram matrix can be inverted exactly, or its linear solve can be enclosed. The temporal residual then includes the same verified coefficient errors.

One may compute the full norm directly instead of subtracting nearly equal projected norms. The identity $\|(\mathrm{I} - P_N)\mathcal{G}(U)\|^2 = \|\mathcal{G}(U)\|^2 - \|P_N\mathcal{G}(U)\|^2$ is exact, but interval subtraction can be inefficient when the two terms are close. Directly enclosing $\mathcal{G}(U) - \sum_j c_j \phi_j$ often avoids this cancellation. Either method is valid if all enclosures are rigorous.

L.5 Time reconstruction and positivity

Apply the quadratic reconstruction of Appendix F to the finite vector field $F_N = -P_N\mathcal{G}$. Since V_N is linear, the entire reconstruction belongs to the affine space. Its mass derivative is a polynomial in the slab variable with coefficients depending smoothly on mass. A sufficient condition is

$$\min_a \min\{X_x'(a), X_y'(a)\} > \frac{h}{8} \left\| (F_N(y) - F_N(x))' \right\|_\infty .$$

When this is too conservative, subdividing in both variables gives a sharper certificate. Passing the endpoint positivity test alone is insufficient for the corrected polynomial curve.

The continuum residual uses $U_t + \mathcal{G}(U)$, whereas the finite-state reconstruction calculation uses $U_t + P_N\mathcal{G}(U)$. Combining them through (62) is precisely what prevents the analysis from hiding spatial approximation error inside an unspecified constant.

L.6 Initial data and changing spaces

An initial projection or learned fit can have nonzero Wasserstein error, even if it preserves mass and positivity. The comparison theorem includes that initial error explicitly. A certificate for it can use the quantile L^2 identity when the target quantile is known with an enclosure.

If the approximation space changes between slabs, a reprojected or remeshed state need not agree with the previous endpoint. One can either enforce continuity in a common representation or treat the transfer as a jump J_n. The triangle inequality then adds $\|J_n\|_{L^2}$ to the current error bound before the next slab propagation. Ignoring transfer jumps would invalidate the otherwise correct continuous reconstruction argument. A second-order total budget must include their accumulated contribution as well.

M Full-covariance Gaussian Wasserstein geometry

M.1 Coordinates and the metric tensor

Let $\mu = \mathcal{N}(m, C)$ with $m \in \mathbb{R}^d$ and $C \in \mathbb{S}^d_{++}$. The coordinate Hilbert space is $\mathbb{R}^d \times \mathbb{S}^d$ with $\langle (v, H), (w, J) \rangle = v^T w + \mathrm{tr}(HJ)$. In particular, the off-diagonal covariance coordinates have the Frobenius multiplicity, not an unweighted vectorization norm.

The Gaussian Wasserstein metric and its Bures covariance distance are classical (Takatsu, 2008; Bhatia et al., 2017). The following derivation records the metric factors needed for the algorithm. An affine gradient velocity has the form $v(x) = \dot{m} + S(x - m)$ with $S = S^T$. Its covariance

derivative is $\dot{C} = SC + CS$. Define the Lyapunov operator $\mathcal{L}_C(S) = CS + SC$. It is positive self-adjoint on symmetric matrices, since

$$\operatorname{tr}(S\mathcal{L}_C(S)) = 2\operatorname{tr}(CS^2) > 0 \quad (S \neq 0).$$

Thus the unique symmetric velocity matrix is $S = \mathcal{L}_C^{-1}(\dot{C})$. Its kinetic energy is

$$\mathbb{E}\,\|v(X)\|^2 = \|\dot{m}\|^2 + \operatorname{tr}(SCS) = \|\dot{m}\|^2 + \tfrac{1}{2}\operatorname{tr}(\dot{C}\,\mathcal{L}_C^{-1}\dot{C}). \tag{64}$$

The affine symmetric field is the Wasserstein tangent field for Gaussian paths: it satisfies the continuity equation and is a gradient field; equivalently, it is the minimal kinetic field for the prescribed Gaussian tangent. Formula (64) therefore identifies the inverse metric, relative to Euclidean–Frobenius coordinates, as

$$K_{(m,C)}(v, H) = (v, 2\mathcal{L}_C(H)) = (v, 2(CH + HC)). \tag{65}$$

This factor of two is essential. Replacing it by $\mathcal{L}_C$ would change the covariance clock.

M.2 Entropy flow without a commutation assumption

For $A \succ 0$, the relative entropy of $\mathcal{N}(m, C)$ with respect to $\mathcal{N}(0, A^{-1})$ differs by a constant from

$$E(m, C) = \tfrac{1}{2}m^T Am + \tfrac{1}{2}\operatorname{tr}(AC) - \tfrac{1}{2}\log\det C.$$

Its fixed-coordinate gradient is $(Am, (A - C^{-1})/2)$. Applying (65) gives

$$\dot{m} = -Am, \qquad \dot{C} = 2\mathrm{I} - AC - CA. \tag{66}$$

The same covariance equation follows directly from the Ornstein–Uhlenbeck Fokker–Planck equation. Thus this is a genuine state-dependent Wasserstein mobility, rather than an artificial diagonal preconditioner imposed on a Euclidean problem.

Lemma M.1 (Exact Gaussian trajectory)**.** *For $C_0 \succ 0$,*

$$m(t) = e^{-At}m_0, \tag{67}$$

$$C(t) = e^{-At}C_0e^{-At} + A^{-1}(\mathrm{I} - e^{-2At}). \tag{68}$$

This covariance is positive definite for all $t \geq 0$.

Proof. Differentiate the two formulas. The derivative of the first covariance term is $-AC_1 - C_1A$. The second term equals $2\int_0^t e^{-2As}\,\mathrm{d}s$ and satisfies the required inhomogeneous equation. Initial values agree, and uniqueness for the linear matrix ODE gives the solution. The congruence $e^{-At}C_0e^{-At}$ is positive definite for finite t, and the integral term is positive semidefinite. No step interchanges C_0 with A. □

The exact formula is an analytic model identity used in proofs. The purely theoretical manuscript contains no Gaussian benchmark or claimed learned acceleration.

M.3 Frozen mobility and its spectral constants

For a midpoint covariance $Z \succ 0$, define $B(v, H) = (v, 2(ZH + HZ))$. If $\sigma_Z\mathrm{I} \preceq Z \preceq M_Z\mathrm{I}$, then

$$\min\{1, 4\sigma_Z\}\mathrm{I} \preceq B \preceq \max\{1, 4M_Z\}\mathrm{I}. \tag{69}$$

Indeed, the covariance quadratic form is $4\operatorname{tr}(ZH^2)$, between $4\sigma_Z\,\|H\|_F^2$ and $4M_Z\,\|H\|_F^2$. The mean block is the identity. Diagonalizing Z alone shows the covariance eigenvalues are $2(z_i + z_j)$ in a symmetric matrix basis. This diagonalization is used only to describe B, not to diagonalize the entire evolving problem simultaneously.

For a midpoint error $\|Z - Z_{\mathrm{target}}\|_F \leq e_Z$, the corresponding operator error is at most $4e_Z$, because $\|2((Z - Z_{\mathrm{target}})H + H(Z - Z_{\mathrm{target}}))\|_F \leq 4\,\|Z - Z_{\mathrm{target}}\|_{\mathrm{op}}\,\|H\|_F$. Thus a validated covariance midpoint error of order h^2 meets the general mobility consistency requirement. Positivity follows for small h from the initial lower eigenvalue bound and the $O(h)$ predicted displacement; it can also be checked directly.

M.4 Explicit entropy sublevel bounds

Let $a = \lambda_{\min}(A) > 0$ and $f_a(t) = at/2 - \log t/2$ for $t > 0$. This function tends to $+\infty$ at zero and infinity and has minimum $m_a = (1 + \log a)/2$ at $t = 1/a$. If $\lambda_i(C)$ are the covariance eigenvalues, then

$$E(m, C) \geq \frac{a}{2} \|m\|^2 + \sum_{i=1}^{d} f_a(\lambda_i(C)). \tag{70}$$

Consequently a nonempty sublevel $E \leq E_0$ obeys

$$\|m\|^2 \leq \frac{2(E_0 - dm_a)}{a}, \qquad f_a(\lambda_i(C)) \leq E_0 - (d-1)m_a.$$

The scalar sublevel of f_a on the right is a compact interval $[\sigma, M] \subset (0, \infty)$; its endpoints can be bounded by validated scalar root bracketing. Hence $\sigma \mathrm{I} \preceq C \preceq M\mathrm{I}$. The same interval bounds any chord between two states in the sublevel, because the SPD order interval is convex. This proves a compact corridor for accepted energy-decreasing endpoints at fixed E_0, A, d.

The midpoint can lie just outside the entropy sublevel without causing a problem: its certified $O(h)$ proximity puts it inside a slightly larger compact SPD corridor. The temporal theorem only needs that larger corridor, not dissipation of the midpoint itself.

M.5 Converting coordinate radii into Wasserstein radii

For two Gaussian laws, the exact distance is

$$W_2(\mathcal{N}(m, C), \mathcal{N}(n, D))^2 = \|m - n\|^2 + \mathrm{tr}(C + D - 2(C^{1/2} D C^{1/2})^{1/2}). \tag{71}$$

We need only an upper comparison, which has a short independent proof.

Lemma M.2 (A covariance lower-bound conversion)**.** *If $C, D \succeq \sigma \mathrm{I}$ with $\sigma > 0$, then*

$$W_2(\mathcal{N}(m, C), \mathcal{N}(n, D))^2 \leq \|m - n\|^2 + \frac{1}{4\sigma} \|C - D\|_F^2 . \tag{72}$$

Proof. Couple $m + C^{1/2}Z$ and $n + D^{1/2}Z$ with the same standard Gaussian Z. Its cost is $\|m - n\|^2 + \left\|C^{1/2} - D^{1/2}\right\|_F^2$, an upper bound on optimal cost. Set $S = C^{1/2}$, $T = D^{1/2}$, and $X = S - T$. The exact identity $SX + XT = C - D$ holds without commutation. Pair it with X in Frobenius inner product. Since $S, T \succeq \sqrt{\sigma}\mathrm{I}$, $2\sqrt{\sigma} \|X\|_F^2 \leq \langle X, SX + XT \rangle_F \leq \|X\|_F \|C - D\|_F$. Divide and square to obtain the result. □

For an accepted state y and exact frozen endpoint $y^\star$, both belong to the initial entropy sublevel: the exact endpoint dissipates by the chain identity and the accepted state dissipates by its gate. Let b_+ be the bound in (69). The global certificate therefore yields

$$W_2(\mu_y, \mu_{y^\star}) \leq \max\{1, (2\sqrt{\sigma})^{-1}\} \sqrt{b_+} \frac{\bar{\varepsilon}}{\beta_h}. \tag{73}$$

The constants are computable from the initial entropy level, A, and the midpoint spectral enclosure. They are not universal dimension-free constants. The same conversion applied to the temporal coordinate error proves Gaussian Wasserstein order two on the actual clock.

N The Gaussian AVF oracle and interior solvability

The principal subtlety in this application is not the formal AVF gradient. It is existence of an interior global minimizer. The integrated entropy potential has a finite value at a singular covariance, so a barrier-divergence argument would be incorrect.

N.1 A NONCOMMUTATIVE CHORD INTEGRAL

For initial covariance $C \succ 0$ and final covariance $D \succ 0$, set $T = C^{-1/2} D C^{-1/2}$. Along the chord,

$$C_s = (1-s)C + sD = C^{1/2}[(1-s)\mathrm{I} + sT]C^{1/2}.$$

Hence

$$\int_0^1 C_s^{-1}\,\mathrm{d}s = C^{-1/2}\left[\int_0^1 ((1-s)\mathrm{I} + sT)^{-1}\,\mathrm{d}s\right] C^{-1/2}. \tag{74}$$

Only the one symmetric positive matrix T is diagonalized. Scalar integration of each eigenvalue gives

$$\int_0^1 \frac{1}{1-s+st}\,\mathrm{d}s = \varphi(t), \qquad \varphi(t) = \begin{cases} \log t/(t-1), & t \neq 1, \\ 1, & t = 1. \end{cases}$$

Substitution proves (15). This congruence formula is not the invalid replacement of C_s^{-1} by $[(1-s)C^{-1} + sD^{-1}]$, and it does not require C and D to commute.

The scalar function is smooth and positive for $t > 0$. Near $t = 1$, its power series is $\varphi(1+z) = 1 - z/2 + z^2/3 - z^3/4 + \cdots$ for $|z| < 1$. This removes cancellation in a numerical evaluation. A deterministic enclosure may instead integrate the inverse chord directly, avoiding matrix logarithms altogether.

N.2 CONVEXITY IN THE FIXED COORDINATE SPACE

For symmetric H,

$$D^2[-\tfrac{1}{2}\log\det C][H,H] = \tfrac{1}{2}\operatorname{tr}(C^{-1}HC^{-1}H) = \tfrac{1}{2}\left\|C^{-1/2}HC^{-1/2}\right\|_F^2 \geq 0. \tag{75}$$

The confinement is affine in C, and the mean energy has positive Hessian A. Thus E is convex in the Euclidean–Frobenius coordinates. For every frozen $B \succ 0$ of the specified form and every $h > 0$, the AVF potential has curvature at least B^{-1}/h. This proves strong convexity with $\beta_h = 1/h$ but still leaves interior existence to prove.

N.3 A CONTINUOUS EXTENSION TO THE SINGULAR BOUNDARY

For $t \geq 0$, define

$$p(t) = -\frac{1}{2}\int_0^1 \frac{\log(1-s+st)}{s}\,\mathrm{d}s. \tag{76}$$

At $t = 0$, the improper integral is finite: on $s \leq 1/2$, $|\log(1-s)|/s$ is bounded, and on $s \geq 1/2$, it is bounded by $2|\log(1-s)|$, which is integrable. On every compact interval $0 \leq t \leq M$, the same split gives an integrable domination after accounting for the uniformly bounded positive logarithm when $M > 1$. Dominated convergence proves continuity of p on $[0, M]$.

The entropy part of the AVF potential is precisely $\sum_i p(t_i)$, where t_i are eigenvalues of $T = C^{-1/2} D C^{-1/2}$. Eigenvalues depend continuously on symmetric matrices, so this expression extends continuously to every $D \succeq 0$. All kinetic, mean, and confinement terms are already continuous there. Thus Q has a continuous extension to the closed domain $\mathbb{R}^d \times \mathbb{S}_+^d$.

For interior y, convexity of E gives $E(x + s(y-x)) - E(x) \geq s\,\langle \nabla E(x), y - x\rangle$. Consequently

$$Q(y) \geq \frac{\|y-x\|_{B^{-1}}^2}{2h} + \langle \nabla E(x), y - x\rangle\,. \tag{77}$$

The inequality extends to the closed domain by continuity. Its right side is a positive quadratic plus an affine function, hence coercive. The continuous extended potential therefore attains a minimum on the closed domain: any minimizing sequence is bounded and has a convergent subsequence in that domain.

N.4 EXCLUDING EVERY SINGULAR MINIMIZER

Lemma N.1 (Inward derivative at a singular covariance)**.** *Let $\overline{D} \succeq 0$ be singular and let the mean be fixed. Along $D_\epsilon = (1-\epsilon)\overline{D} + \epsilon C$, $0 < \epsilon \leq 1$, the derivative of the entropy part of Q tends to $-\infty$ as $\epsilon \downarrow 0$. The derivatives of the other terms remain bounded.*

Proof. Let $t_i \geq 0$ be eigenvalues of $\overline{T} = C^{-1/2}\overline{D}C^{-1/2}$. Then $C^{-1/2}D_\epsilon C^{-1/2} = (1-\epsilon)\overline{T} + \epsilon \mathrm{I}$ has eigenvalues $t_i(\epsilon) = (1-\epsilon)t_i + \epsilon$ in the same eigenbasis. For $t > 0$, differentiating (76) gives $p'(t) = -\varphi(t)/2$. Therefore the derivative is

$$-\frac{1}{2}\sum_i (1-t_i)\varphi(t_i(\epsilon)). \tag{78}$$

At least one t_i equals zero. Its contribution is $-\varphi(\epsilon)/2$, tending to $-\infty$ since $\varphi(\epsilon) = \log\epsilon/(\epsilon-1) \to +\infty$. Contributions for $t_i > 0$ have finite limits. The kinetic and confinement terms are polynomial or affine on the bounded line segment and have bounded derivatives. The mean remains fixed. □

Theorem N.2 (Interior Gaussian minimizer for every frozen step)**.** *For every initial $C \succ 0$, $h > 0$, and frozen $Z \succ 0$, the Gaussian AVF potential has a unique global minimizer with positive definite covariance.*

Proof. The continuous coercive extension has a minimizer on the closed domain. Suppose its covariance were singular. By Lemma N.1, the potential along the inward segment has strictly negative derivative for all sufficiently small positive ϵ. Its derivative is integrable at zero because the potential is continuous there and the only singular terms are logarithmic. Equivalently, integrate the explicit scalar formulas for $p(t_i(\epsilon))$. It follows that an interior point on this segment has strictly smaller objective than the boundary candidate, a contradiction. Hence the minimizer is interior. The strong-convexity bound proves uniqueness and globality on the open domain, and stationarity gives the exact frozen AVF equation. □

The theorem proves existence for arbitrary h for this *frozen* convex Gaussian subproblem. Second-order consistency of the time integration still requires a valid midpoint and sufficiently small steps in the temporal theorem. Unconditional solvability and small-step accuracy are distinct assertions.

N.5 A CORRECTION CAVEAT AND A LOCAL REMEDY

Although the global minimum is interior, not every Gaussian Q sublevel is compactly contained in the open SPD cone: its continuous boundary values are finite. Therefore the compact-interior-sublevel premise of Theorem G.1 is not automatic for an arbitrary Gaussian initializer. One can enforce a spectral box containing the exact minimizer, use a line search with a proved interior-preserving convergence theorem, or initialize within a verified local strongly convex neighborhood and use a local contraction.

For sufficiently small h on a compact initial entropy corridor, a local fallback is straightforward. The implicit-function argument gives the exact step within $O(h)$ of x, and the Hessian of Q in frozen coordinates is $h^{-1}\mathrm{I} + O(1)$ on a fixed small ball. The map $w \mapsto w - h\nabla\widehat{Q}(w)$ then has derivative $O(h)$ on that ball. Starting from the midpoint-consistent explicit first-order endpoint and taking the ball large enough to contain its $O(h^2)$ error gives a contraction into itself for small h. To be explicit, if the map derivative is bounded by $q < 1$ on a closed ball of radius R about w_0, and its displacement at the center is at most $(1-q)R$, then it maps the ball into itself and converges to its unique fixed point. These two inequalities can be validated from corridor derivatives and the residual at w_0.

This supplies a small-step provable corrector without falsely transferring the quantile logarithmic-barrier argument to Gaussian sublevels. A practical implementation can use a more efficient safeguarded solver, but its global convergence is not claimed solely from convexity on an open cone.

N.6 AN ENCLOSURE WITHOUT MATRIX LOGARITHMS

Assume $C, D \succeq \sigma\mathrm{I}$ and let $H = D - C$. For $A(s) = C + sH$,

$$\frac{\mathrm{d}}{\mathrm{d}s}A(s)^{-1} = -A(s)^{-1}HA(s)^{-1}, \quad \frac{\mathrm{d}^2}{\mathrm{d}s^2}A(s)^{-1} = 2A(s)^{-1}HA(s)^{-1}HA(s)^{-1}.$$

Thus the second derivative operator norm is at most $2\|H\|_{\mathrm{op}}^2/\sigma^3$, and a Frobenius bound follows by multiplying by $\sqrt{d}$. Composite midpoint quadrature with mesh width Δs has matrix-norm integration error at most this bound times $(\Delta s)^2/24$. Exact rational inversions at rational midpoints, together with rational upper bounds on σ^{-1} and $\|H\|$, give a deterministic AVF envelope. The covariance gradient has an additional factor $1/2$, and conversion to the B norm uses (69). This construction proves computability even when an eigendecomposition routine itself has no validated error guarantee.

O NEURAL PROPOSALS, REPRESENTATION EFFECTS, AND CONDITIONAL WORK REDUCTION

O.1 WHAT LEARNING IS AND IS NOT RESPONSIBLE FOR

The method allows a neural map to propose a midpoint, a frozen mobility approximation, and an endpoint. Its correctness does not require that training minimize a nonconvex objective globally. Instead, the deployed state is checked against a full-state equation and corrected when necessary. The output after correction may lie outside the image of the original network. This is a deliberate algorithmic choice: imposing an arbitrary restrictive generator image would reintroduce a representation error that the full-state theorem does not eliminate.

A useful formal training objective is the frozen potential $Q_{h,x}^B(y_\theta)$, with admissibility imposed by a representation or checked separately. Minimizing that objective over parameters is not the same as finding its global minimum over states. The certificate is evaluated after the state is proposed, not inferred from training loss decrease. A small parameter gradient needs a verified Jacobian lower bound before it can imply a small state gradient.

O.2 CONDITIONAL BENEFIT OF AN ACCURATE PREDICTOR

Proposition O.1 (Prediction error to initial certificate)**.** *Suppose a predictor satisfies $\|y_\theta - y^\star\|_{B^{-1}} \le a_\theta$ and the segment between these states lies in a region where $\left\|B^{1/2}\nabla^2 Q B^{1/2}\right\|_{\mathrm{op}} \le L_Q$. Then*

$$\|r(y_\theta)\|_{B^{-1}} \le L_Q a_\theta, \qquad Q(y_\theta) - Q(y^\star) \le \frac{L_Q}{2} a_\theta^2. \tag{79}$$

If $a_\theta = O(h^{q+1})$ and $L_Q = O(h^{-1})$, the initial residual is $O(h^q)$ and the initial gap is $O(h^{2q+1})$.

Proof. In frozen normalized coordinates, integrate the Hessian from the stationary minimizer to the predictor. This gives the gradient bound. Taylor's formula with zero linear term at the minimizer gives the objective bound. Substitute the assumed scalings. □

The proposition is conditional on an actual predictor approximation property. A finite training set does not establish it uniformly over a deployment domain. It nevertheless states precisely what a learning theorem would need to provide to obtain a correction-complexity advantage.

For a corrector with gap factor $\rho \in (0,1)$, a sufficient initial gap Δ_0 and sufficient target gap Δ_{tar} imply at most $\max\{0, \lceil \log(\Delta_0/\Delta_{\mathrm{tar}})/[-\log\rho] \rceil\}$ corrections, under the gradient and distance conversions of Appendix G. Reducing Δ_0 improves this bound. It says nothing about the cost of training, inference, validation, or linear solves. Those costs would need to be measured in an empirical study, which is not part of this paper.

O.3 A SIMPLE CLASSICAL CORRECTION BENCHMARK IN THEORY

Consider $T(y) = x - hBg(x,y)$ on a fixed convex ball contained in the admissible corridor. In frozen coordinates its derivative in the endpoint has norm at most $hL_B/2$, where L_B bounds the normalized energy Hessian. If $q = hL_B/2 < 1$ and T maps the ball into itself, fixed-point iteration converges geometrically with factor q.

If an initial explicit predictor has state error $O(h^2)$, one such correction has state error $O(h^3)$ and hence residual $O(h^2)$ on a corridor with $L_Q = O(h^{-1})$. More corrections can meet any prescribed constant in the residual budget. The energy gate still needs to be checked, particularly near equilibrium. This elementary fact is a useful theoretical baseline: a learned predictor does not acquire a speed advantage simply by being paired with a corrector. Its value would be in supplying a better initial state, a larger admissible step, or cheaper certified information than a strong classical alternative.

O.4 Positive gap logits and their exact reach

For one quantile component, let $d_i = 2L\exp(\theta_i)/\sum_j \exp(\theta_j)$ and set $q_j = -L + \sum_{i<j} d_i$. Every finite logit vector yields a positive quantile with fixed endpoints. Conversely, every positive gap vector of total length $2L$ is represented by $\theta_i = \log d_i$ plus an arbitrary common constant. Thus unrestricted logits cover the full positive-gap state domain.

This surjectivity does not transfer to an arbitrary neural network producing the logits: its image may be smaller. Nor does it remove ill conditioning near a zero gap. The next result gives a conditional quantitative statement for the free-logit map itself.

Proposition O.2 (A logit Jacobian lower bound)**.** *Restrict logits to the zero-sum subspace $\mathbf{1}^\perp$ and suppose every gap satisfies $d_i \geq d_{\min} > 0$. The Jacobian J from zero-sum logits to interior positions, in Euclidean norms, has smallest singular value at least $d_{\min}/2$. Hence for a differentiable state objective,*

$$\|\nabla_{\mathrm{coord}} Q(q)\| \leq \frac{2}{d_{\min}} \|\nabla_\theta Q(q(\theta))\| . \tag{80}$$

Proof. The gap Jacobian is $S = \mathrm{diag}(d) - dd^T/(2L)$, mapping $\mathbf{1}^\perp$ into itself. For $v \perp \mathbf{1}$, put $\bar{v}_d = (\sum_i d_i v_i)/(2L)$. Then

$$v^T S v = \sum_i d_i (v_i - \bar{v}_d)^2 \geq d_{\min} \sum_i (v_i - \bar{v}_d)^2 \geq d_{\min} \|v\|^2 .$$

The last inequality uses $\sum_i v_i = 0$. Thus the smallest singular value of S on that subspace is at least $d_{\min}$. Let T be cumulative summation from zero-sum gap variations to interior position variations. Its inverse on that space is the incidence map D, which has Euclidean operator norm at most two: the sum of squared consecutive differences with zero endpoint extensions is at most four times the sum of squared interior entries. Hence $\|Tw\| \geq \|w\|/2$. Since $J = TS$, its smallest singular value is at least $d_{\min}/2$. The adjoint has the same nonzero singular values, and the chain rule gives the gradient assertion. □

For the mass metric and frozen mobility, one obtains the conservative full-state residual conversion

$$\|r\|_{B^{-1}} \leq \frac{2\sqrt{b_+}}{d_{\min}\sqrt{\lambda_{\min}(A)}} \|\nabla_\theta Q\| .$$

Here ∇_θ is the free-logit derivative, not the derivative with respect to all weights of a deep network. The latter introduces another Jacobian and needs another lower singular-value bound. This result identifies a circumstance in which a parameter derivative can become a valid certificate; it also makes clear why an unchecked parameter derivative generally cannot.

O.5 Learning a mobility without changing the target

A network may output a positive B by a Cholesky or spectral parameterization. Positivity guarantees a well-defined frozen metric and nominal energy dissipation, but not the right physical flow. To target K, its output must additionally satisfy the operator discrepancy bound in (2). A regularizer encouraging closeness to $K(z)$ is not a substitute for that bound. A deterministic certificate can compare the two operators with interval entries or a verified spectral norm bound.

For Gaussian flows, a simpler option is to learn only Z and construct $B = 2\mathcal{L}_Z$ analytically. For quantile standard Wasserstein flow, $B = \mathrm{I}$ needs no learning at all. These model-specific choices remove an unnecessary source of clock error and leave learning focused on endpoint proposal or admissible midpoint correction.

O.6 Representation constraints and exact invariants

If a model enforces linear invariants, the full-state correction should occur within their invariant affine leaf, not in the unrestricted ambient coordinates. If it imposes additional nonlinear restrictions without a proved geometric metric, the state equation changes. Positivity parameterizations, invariant leaves, and arbitrary low-dimensional generator manifolds therefore require different analyses. Appendix T handles the fixed-linear case; a general learned manifold with changing rank is outside the present theorem.

P Outer time budgets, rejection semantics, and finite progress

P.1 A global posterior budget

Suppose the comparison constant is $\lambda \geq 0$ in the dissipative form of (61). For matching initial data, accepting only reconstructions with $b_n \leq \epsilon_{\mathrm{time}}/T$ gives

$$e_n^{\mathrm{up}} \leq \sum_{j<n} h_j b_j \leq \epsilon_{\mathrm{time}}, \qquad t_n \leq T.$$

This is a nonasymptotic physical-time guarantee. It is separate from the inner objective gap and remains meaningful when the local truncation constant is unknown. For $\lambda < 0$, the sufficient uniform bound $b_n \leq \epsilon_{\mathrm{time}}/(Te^{|\lambda|T})$ gives the same conclusion. A nonzero initial error consumes its propagated share of the budget before any step is taken.

Sharper allocations use the actual recurrence rather than this uniform bound. For example, reserve a nonnegative remaining budget and evaluate a candidate's contribution $c_\lambda(h_n)b_n$ before accepting it. Any such allocation is valid if the final propagated sum is certified. The budget can depend on already accepted states and past certificates; it cannot ignore a contribution merely because that contribution is inconveniently large.

P.2 Nested stopping conditions

The inner residual test answers whether the accepted endpoint approximates the defining frozen step to the accuracy needed by the time integrator. The work test answers whether it dissipates the intended original energy. The reconstruction test answers whether the full evolution error meets a prescribed budget. A candidate may pass the first two and fail the third because the physical step is too long. In a continuum problem it may also fail because the spatial representation is inadequate, even at arbitrarily small time step.

A sensible controller therefore distinguishes correction, oracle refinement, time-step reduction, and spatial refinement. Additional inner optimization should not be used indefinitely to repair a spatial residual or a truncation error that persists at the exact inner root. The decomposition in Appendix L makes that distinction observable.

P.3 A conditional no-accumulation theorem

Theorem P.1 (Finite progress of a halving controller)**.** *Fix $T > 0$. Suppose the trial subroutine, as actually executed, returns an admissible endpoint and reconstruction passing all requested tests in finite work for every $0 < h \leq h_\star$. Suppose every larger trial also terminates in finite work, either with acceptance or rejection. At each accepted state, propose $h_{\mathrm{prop}} = \min\{h_{\max}, 2h_{\mathrm{prev}}, T - t_n\}$, with a fixed positive initial proposal for the first step. On a failed trial, halve h and retry. Suppose a passing trial is accepted. Then the controller reaches T in finitely many accepted steps and finitely many trials.*

Proof. Let $\delta > 0$ be smaller than the initial proposal, $h_{\max}/2$, and $h_\star/2$. Apart from a final remainder shorter than δ, induction shows that accepted steps are at least δ. Indeed, a proposed step is at least δ whenever the preceding accepted step is at least δ and the remaining time is at least δ. If rejected trials are halved, the first trial below $h_\star$ passes; its size is greater than $h_\star/2 \geq \delta$, unless the initial proposal was already smaller, in which case it was at least δ and passes. The first accepted

step obeys the same argument. Thus at most $\lceil T/\delta \rceil + 1$ accepted steps are needed. Each step has finitely many halvings, and the hypothesis supplies finite work for its successful trial. Every failed trial has finite work by hypothesis, so total trial work is finite. □

The theorem is conditional on actual small-step success of the trial routine. Existence of a certifiable state does not imply that a fixed, insufficient inner work cap will find it. The theorem does not apply to a cap that causes rejection of arbitrarily small steps. In the finite-state setting, the order-two reconstruction bound and a compact corridor supply the outer-defect part of this condition, while exact or refined oracles and a suitable corrector supply the inner part. In a continuum computation, a spatial residual larger than the requested budget violates the condition; reducing h alone need not help. The result therefore does not hide a universal termination theorem for arbitrary neural architectures or spatial spaces.

P.4 Dyadic physical time and reproducibility

One practical way to avoid floating-point ambiguity in the accepted clock is to restrict trial lengths to rational dyadic multiples of a fixed unit and to represent the remaining time exactly. Then the statement $t_{n+1} = t_n + h_n$ is an exact arithmetic identity. Other representations are possible, but any error in the physical time itself should be accounted for. If the reported time differs from the exact sum by Δt and the reference velocity is bounded by M_F, a further state-comparison term $M_F|\Delta t|$ is sufficient.

Rational clock bookkeeping does not validate the state solve or the energy calculation by itself. It removes one independent uncertainty. Conversely, a rigorously validated state step with an incorrectly reported physical time does not answer the intended evolution question.

P.5 Step doubling is a proposal mechanism, not a certificate

A full step and two half steps can suggest a new time-step size by comparing endpoints. Such a discrepancy is often useful, but without an estimator reliability theorem it is not a rigorous upper bound on the local error. In the proposed method, it may propose h; acceptance still uses the verified reconstruction and inner tests when a rigorous global budget is requested.

The state from two accepted half steps has its own physical clock and energy certificates. A Richardson extrapolation of the full and half-step states is a different state and can increase the original energy, as shown exactly in Appendix H. No extrapolated endpoint inherits a certificate merely because its ingredients were certified.

P.6 Restarting after a transfer or a changed model

If a remeshing, projection, or model replacement changes the represented distribution, record that change as a transfer error. In a common Hilbert representation with jump J, the posterior bound updates by $e^{\mathrm{up}} \leftarrow e^{\mathrm{up}} + \|J\|$. In a distributional representation without common coordinates, a verified Wasserstein bound on the transfer is sufficient by the triangle inequality. Changing the target energy or mobility is more fundamental: it changes the evolution equation and requires a modeling-error term or a new target statement. A controller must not silently treat such a change as an ordinary inner correction.

Q Fresh-sample validation, adaptive selection, and sampling costs

This section states probabilistic alternatives to deterministic envelopes. Its guarantees are conditional-probability statements with explicit failure budgets, not substitutes for exact interval proofs.

Q.1 A BOUNDED SCALAR VALIDATION LEMMA

Let a candidate define a measurable nonnegative quantity $Z(\xi) \in [0, M^2]$ under a specified probability law for ξ. For example, Z can be the squared norm of a continuum residual at a uniformly drawn mass location, or at a uniformly drawn point of a space–time slab. Let $\mu = \mathbb{E}Z$. For independent samples $Z_1, \dots, Z_m$ and $\delta \in (0, 1)$, define

$$U_m = \frac{1}{m} \sum_{i=1}^{m} Z_i + M^2 \sqrt{\frac{\log(1/\delta)}{2m}}. \tag{81}$$

Then $\mathbb{P}\{\mu \le U_m\} \ge 1 - \delta$ by Hoeffding's inequality (Hoeffding, 1963).

For completeness, the underlying exponential bound can be proved by convexly bounding e^{tZ} by its chord on $[0, M^2]$ and optimizing after centering. The resulting moment bound is $\mathbb{E}e^{t(Z-\mathbb{E}Z)} \le e^{t^2 M^4/8}$. Independence multiplies these bounds for the sum. Markov's inequality applied to the negative centered sum gives $\mathbb{P}\{\mu - \overline{Z} \ge a\} \le \inf_{t>0} \exp(-tma + mt^2 M^4/8) = \exp(-2ma^2/M^4)$. Taking $a = M^2\sqrt{\log(1/\delta)/(2m)}$ proves the assertion.

The bound M must itself be justified over the entire validation domain. The largest observed residual is not a valid replacement for an a priori range bound without a different probabilistic theorem.

Q.2 ADAPTIVE CANDIDATES AND CONFIDENCE SPENDING

Theorem Q.1 (Simultaneous fresh-sample certificates)**.** *Index every validation call, including rejected candidates and retried calls, by $j = 1, 2, \dots$. Let the candidate, range bound M_j, sample count m_j, and validation law be measurable with respect to the history $\mathcal{F}_{j-1}$. Using fresh randomness independent of $\mathcal{F}_{j-1}$, draw samples that are conditionally independent with the specified conditional law. Choose deterministic $\delta_j > 0$ with $\sum_j \delta_j \le \delta$. Then with probability at least $1 - \delta$, every bound* (81) *at every call is valid simultaneously. The same conclusion holds for any adaptively selected accepted subsequence of those calls.*

Proof. Conditional on $\mathcal{F}_{j-1}$, the candidate function and all validation parameters are fixed, so the scalar Hoeffding result applies with conditional failure probability at most δ_j. Taking expectations gives the same unconditional failure bound. A union bound over all calls gives total failure probability at most $\sum_j \delta_j \le \delta$. On the complement event, every call is valid; choosing an accepted subsequence based on the observed results cannot invalidate an assertion already true for all calls. □

A convenient choice is $\delta_j = 6\delta/(\pi^2 j^2)$. Spending a failure budget only on accepted calls is not the theorem above: a candidate can undergo many rejected validation attempts before a favorable one is selected. All attempts must be included, or a separately proved sequential-confidence method must be used.

Reusing a fixed validation set after fitting the candidate to it destroys the conditional-independence premise. Uniform convergence over a specified function class or a valid reusable-holdout analysis could replace it, but neither follows from small empirical loss alone.

Q.3 APPLYING VALIDATION TO A CONTINUUM SLAB

Suppose U is fixed before the validation draw on a slab of physical length h. Draw s uniformly on $[0, 1]$ and a uniformly on $[0, 1]$, and let $Z(s, a) = |\mathcal{R}(t_n + hs, a)|^2$. Then

$$\mathbb{E}Z = \frac{1}{h} \int_{t_n}^{t_{n+1}} \|\mathcal{R}(t)\|_{L^2}^2 \, \mathrm{d}t.$$

If U_m bounds this expectation, Cauchy–Schwarz in time gives $\int_{t_n}^{t_{n+1}} \|\mathcal{R}(t)\|_{L^2} \, \mathrm{d}t \le h\sqrt{U_m}$. Therefore a probabilistic counterpart of the endpoint recurrence is

$$e_{n+1}^{\mathrm{up}} = e^{-\lambda h} e_n^{\mathrm{up}} + h e^{\max\{0, -\lambda\} h} \sqrt{U_m}. \tag{82}$$

On the simultaneous event in Theorem Q.1, this recurrence bounds the continuum endpoint error for every accepted slab. It uses an integrated residual, not an unjustified claim that sampled values bound the supremum in time.

Admissibility, range bounds, and any energy gate still need their own validation. Sampling the squared residual does not prove positive quantile slope between sample points, and it does not certify the sign of the exact energy difference. A combined algorithm may use deterministic geometric and energy checks together with a probabilistic residual bound, but its final error statement must retain the overall probability qualifier.

Q.4 The cost of resolving a second-order residual scale

Suppose one wants to certify $\mu \le \tau^2 h^4$. If the observed empirical mean is at most $\tau^2 h^4/2$, the Hoeffding allowance in (81) is at most the other half whenever

$$m \ge \frac{2M^4}{\tau^4 h^8} \log(1/\delta). \tag{83}$$

Thus this particular conservative certificate can require $O(h^{-8})$ samples. This is a sufficient sample count conditional on the observed empirical margin, not a guarantee that an arbitrary candidate has a small mean. It is also not an information-theoretic lower bound: variance-sensitive methods may improve it under appropriate conditions.

Proposition Q.2 (A rare-spike validation lower bound)**.** *Consider independent uniform validation of a bounded nonnegative function $Z \le M^2$. Suppose a procedure certifies $\mathbb{E}Z \le \epsilon$ whenever all m observations are zero. For any $q \in (0,1)$ with $qM^2 > \epsilon$, uniform validity with failure probability at most δ requires*

$$m \ge \frac{\log(1/\delta)}{-\log(1-q)}. \tag{84}$$

In particular, choosing q proportional to h^4 gives an $\Omega(h^{-4}\log(1/\delta))$ obstruction for resolving a squared-residual threshold proportional to h^4.

Proof. Let $Z = M^2 \mathbf{1}_A$ on a measurable set of mass q. Its mean exceeds ϵ, but all observations are zero with probability $(1-q)^m$. On that event the stated procedure issues a false certificate. Therefore $(1-q)^m \le \delta$, which rearranges to (84). Smooth plateau functions with slightly adjusted support and amplitude give the same asymptotic obstruction in a smooth class without a uniform derivative bound. □

This lower bound and the Hoeffding sufficient count are deliberately different. Their gap is not hidden by calling the Hoeffding exponent optimal. The deterministic finite-query obstruction in Appendix H is stronger in another direction: without regularity information, no finite observations give an unrestricted deterministic certificate at all.

Q.5 Validation laws and the intended norm

Sampling locations from a nonuniform law certifies a norm under that law. To recover a target Lebesgue or reference-measure norm, use the appropriate importance weights and prove bounds on the weighted random variable. An unknown or unbounded density ratio can destroy the boundedness premise. Likewise, sampling parameter gradients certifies a parameter-space quantity, not the full-state residual, unless the geometric conversion is separately justified.

R Near-equilibrium accuracy, energy decay, and uniform-time bounds

R.1 A relative residual condition removes an artificial energy floor

This section first assumes exact AVF oracles. Let E_* be the minimum energy in the relevant invariant region, and assume a uniform frozen-metric Polyak–Łojasiewicz inequality

$$\|\nabla E(x)\|_B^2 \ge 2\mu(E(x) - E_*), \qquad \mu > 0. \tag{85}$$

Suppose also $\left\|B^{1/2}\nabla^2 E B^{1/2}\right\| \le L_B$ on every accepted chord and $\mu \le L_B$. Instead of only the signed gate, impose the sufficient relative norm condition $\|r\|_{B^{-1}} \le \eta_r \|g\|_B$, with $0 \le \eta_r < 1$.

Theorem R.1 (Energy decay at physical times)**.** *Every such step satisfies*

$$E(y) - E_* \le \left(1 - \frac{2\mu(1-\eta_r)h}{[1 + hL_B(1+\eta_r)/2]^2}\right)(E(x) - E_*). \tag{86}$$

The factor lies in $[0, 1]$*. For a sequence with* $h_n \le h_{\max}$ *and uniform constants,* $E(x_n) - E_* \le e^{-\gamma t_n}(E(x_0) - E_*)$*, where* $\gamma = 2\mu(1-\eta_r)/[1 + h_{\max}L_B(1+\eta_r)/2]^2$*.*

Proof. The residual equation gives $\|\delta\|_{B^{-1}} \le h(1+\eta_r)\|g\|_B$. The AVF gradient differs from the initial gradient by at most $L_B \|\delta\|_{B^{-1}}/2$, so

$$\|\nabla E(x)\|_B \le [1 + hL_B(1+\eta_r)/2]\|g\|_B .$$

Combine this with (85) and the exact energy decrease $E(y) - E(x) \le -(1-\eta_r)h\|g\|_B^2$ to obtain (86). To check the factor, the maximum of $h/(1+ah)^2$ for $a > 0$ is $1/(4a)$. With $a = L_B(1+\eta_r)/2$, the subtracted quantity is at most $\mu(1-\eta_r)/(L_B(1+\eta_r)) \le 1$. Finally, this quantity is at least γh_n for every step. Iterating $1 - s \le e^{-s}$ gives the physical-time exponential bound. □

A sufficient combined tolerance is

$$\|r\|_{B^{-1}} \le \min\{\tau h^2, \eta_r \|g\|_B\}. \tag{87}$$

Its residual-based gap certificate is bounded by $\min\{\tau^2 h^4, \eta_r^2 \|g\|_B^2\}/(2\beta_h)$. For $\beta_h \asymp h^{-1}$, this is the minimum of an $O(h^5)$ time budget and a gradient-dependent $O(h\|g\|_B^2)$ equilibrium budget. The latter tightens as equilibrium is approached. A fixed absolute residual threshold alone need not preserve a strict energy decay rate all the way to equilibrium.

With imperfect oracles, Lemma C.3 gives a sufficient robust relative gate. A quantitative decay theorem can then be obtained by replacing the exact gradient norm with its lower enclosure and tracking the oracle constants. We do not silently apply the exact-oracle decay factor to an approximate gradient.

R.2 When the PL premise follows from the model

If E is m-strongly convex in the fixed Hilbert norm along the region joining each state to its minimizer, the same square-completion argument as before gives $E(x) - E_* \le \|\nabla E(x)\|^2/(2m)$. If $B \succeq b_-\mathrm{I}$, then (85) holds with $\mu = mb_-$. A compatible L_B follows from an upper Hessian bound and $B \preceq b_+\mathrm{I}$.

For quantiles with sufficiently strongly convex confinement relative to mean coupling, $m = \min_j a_j - \ell_\Phi > 0$ is valid. For Gaussian entropy on a covariance interval $\sigma\mathrm{I} \preceq C \preceq M\mathrm{I}$, the fixed-coordinate Hessian is bounded below by $\min\{\lambda_{\min}(A), 1/(2M^2)\}\mathrm{I}$ and above by $\max\{\lambda_{\max}(A), 1/(2\sigma^2)\}\mathrm{I}$. These follow from (75). Thus the PL and smoothness assumptions are explicit on the accepted entropy corridor.

R.3 Uniform-in-time trajectory accuracy under actual contraction

Energy decay alone does not imply closeness of trajectories at a fixed time. For that stronger statement assume the target vector field satisfies

$$\langle F(u) - F(v), u - v\rangle \le -\alpha\|u - v\|^2, \qquad \alpha > 0, \tag{88}$$

throughout a common invariant comparison corridor, with time-independent derivative and certificate constants. Assume all accepted local physical defects are bounded by Ch_n^3 and that the exact short flows used in comparison remain in this corridor.

Theorem R.2 (A uniform physical-time error bound)**.** *Under these assumptions, for arbitrary positive steps with* $h_n \le h_{\max}$*,*

$$\|x_n - u(t_n)\| \le e^{-\alpha t_n}\|x_0 - u(0)\| + \frac{Ce^{\alpha h_{\max}}}{\alpha}h_{\max}^2. \tag{89}$$

No finite terminal horizon appears in the second term.

Proof. Exact flow contraction from (88) gives the recurrence $e_{n+1} \le e^{-\alpha h_n} e_n + C h_n^3$. Iteration and $h_j^3 \le h_{\max}^2 h_j$ give

$$e_n \le e^{-\alpha t_n} e_0 + C h_{\max}^2 \sum_{j<n} h_j e^{-\alpha(t_n - t_{j+1})}.$$

For $s \in [t_j, t_{j+1}]$, the endpoint exponential is at most $e^{\alpha h_{\max}} e^{-\alpha(t_n - s)}$. Integrating and summing bounds the sum by $e^{\alpha h_{\max}} \int_0^{t_n} e^{-\alpha(t_n - s)} \, \mathrm{d}s \le e^{\alpha h_{\max}}/\alpha$. □

This theorem invokes contraction of the *continuous vector field in a specified fixed norm*. It does not assume contraction of arbitrary Wasserstein proximal maps. Distinctions between convexity notions and proximal stability are important in Wasserstein geometry; see, for example, Di Marino et al. (2026).

For a strongly convex quantile energy with identity target mobility, $F = -\nabla E$ satisfies (88) with its strong-convexity constant. Fixed-resolution energy sublevels supply time-independent derivative bounds. The corresponding ODE solution is global: its energy derivative is nonpositive, the inverse-gap energy keeps it in a compact subset of the open ordered domain, and a smooth finite-dimensional ODE can be continued while its solution remains in such a compact set. For Gaussian coordinates, the linear difference field is $(-A\Delta m, -A\Delta C - \Delta C A)$, whose pairing with $(\Delta m, \Delta C)$ is at most $-a \|\Delta m\|^2 - 2a \|\Delta C\|_F^2$, with $a = \lambda_{\min}(A)$. Hence it is a-contractive in Euclidean–Frobenius norm. The entropy corridor and the norm conversion in Lemma M.2 turn (89) into a uniform Gaussian Wasserstein error bound, under the same uniform accepted-step hypotheses.

These are genuinely stronger conclusions than a fixed-T estimate, but they require additional contraction and uniform-corridor information. A nonconvex mean coupling without a positive net curvature need not satisfy them.

S Deterministic computability and validated arithmetic

S.1 What needs to be enclosed

A complete deterministic accepted-step certificate consists of admissibility bounds, a positive spectral bound for the frozen metric, a midpoint consistency bound, an AVF error envelope, a residual upper bound, and a signed-work upper bound. A global trajectory certificate additionally encloses the reconstruction throughout each slab and its defect norm or integrated defect. The theoretical inequalities remain valid with conservative enclosures; they do not require every expression to be evaluated in closed form.

Finite binary floating-point numbers can be interpreted as exact rational inputs. This does not validate how those inputs were generated, but it makes a post-processing claim about those particular stored states mathematically well defined. One can then verify whether they satisfy the inequalities. The verification code accompanying this theoretical paper uses exact rational arithmetic for selected identities, state-space acceptance tests, and bounded-work quantile and full-covariance reference steps. It is not a formal proof assistant and does not establish every analytic theorem by machine checking.

S.2 Rational interval operations

Represent an interval by rational endpoints $[a, b]$ with $a \le b$. Addition, subtraction, multiplication, and division by an interval avoiding zero can be performed with exact endpoint arithmetic and the usual interval extremum rules. The resulting interval contains every exact value of the expression for inputs in the supplied intervals. Induction on an expression tree proves this enclosure property.

For an exact nonnegative rational r and integer precision $p \ge 1$, a dyadic square-root upper bound is

$$u_p = 2^{-p} \min\{k \in \mathbb{N} : k^2 \ge 2^{2p} r\}. \tag{90}$$

Integer square-root and rational comparisons compute the minimum. Then $\sqrt{r} \le u_p < \sqrt{r} + 2^{-p}$, except that equality at the left is possible. No floating-point rounding assumption is needed. This operation provides certified norm upper bounds for rational quadratic forms.

For positive rational x, set $t = (x-1)/(x+1)$, so $|t| < 1$. The logarithm series yields

$$\log x = 2\sum_{k=0}^{m-1} \frac{t^{2k+1}}{2k+1} + R_m, \qquad |R_m| \le \frac{2|t|^{2m+1}}{(2m+1)(1-t^2)}. \tag{91}$$

The remainder estimate follows by replacing each denominator in the omitted series by its smallest value and summing a geometric series. All endpoints are rational, and the remainder tends to zero. Thus every positive rational logarithm has arbitrarily narrow rational enclosures. Range reduction can accelerate convergence but is not necessary for computability.

S.3 Positive matrices and linear solves

For a rational symmetric matrix, exact LDL^T elimination with positive pivots certifies positive definiteness. Rational Gaussian elimination gives an exact inverse when nonsingularity has been certified. A simple lower eigenvalue bound for $P \succ 0$ is $\lambda_{\min}(P) \ge 1/\left\|P^{-1}\right\|_F$, where the Frobenius norm receives a dyadic upper bound. An upper bound is $\lambda_{\max}(P) \le \operatorname{tr} P$. These estimates can be conservative but are fully explicit.

For a generic coordinate form of the frozen step, let $P \succ 0$ be the kinetic matrix and let g_c be the coordinate AVF covector. Then

$$r = \delta/h + P^{-1}g_c, \quad \varepsilon^2 = r^T P r, \quad D = g_c^T P^{-1} g_c, \quad \omega = g_c^T r. \tag{92}$$

This formulation covers a nonidentity fixed Hilbert Gram matrix without hiding it in a vectorization convention. For mass-coordinate quantiles, P is the mass matrix. For Gaussian symmetric coordinates, the Frobenius Gram matrix must be included when converting the frozen Lyapunov mobility to P.

The residual and energy gates can be verified using rational norm upper bounds and rational comparisons. A strict comparison eventually becomes decidable as interval widths shrink. An equality may remain unresolved unless a symbolic identity establishes it. A correct implementation reports that distinction instead of treating a tolerance in the validator as an exact theorem.

S.4 Quantile oracle and energy evaluation

For rational positive gaps, the diffusion AVF forces in (52) and the quartic mean force in (54) are rational. The quadratic-confinement terms are rational as well. Thus the exact AVF gradient, residual quadratic forms, and signed energy-work identity can be verified without numerical quadrature or logarithm evaluation. The original inverse-gap energy is also rational for this model.

The potential Q contains logarithms of gap ratios. Formula (91) gives its objective-value interval whenever one is needed. However, the residual-based gap certificate itself only needs the residual and a lower curvature constant, not a numerical evaluation of $Q(y)$ or $Q(y^\star)$. This separation reduces validation cost and avoids unnecessary cancellation in a nearly minimized objective.

A positivity certificate is simply a lower bound on every gap. For a quadratic reconstruction, exact critical-point checks in the slab variable suffice, as described in Appendix F. The full continuum theorem cannot be applied to its piecewise affine mass representation; a smooth lift requires the additional procedure below.

S.5 Smooth polynomial residuals

Suppose $U(a, s)$ is a polynomial quantile reconstruction with rational coefficients, $U_a \ge \sigma > 0$ on the slab, and V is polynomial. The full residual is rational, with denominators that are powers of U_a. Subdivide the compact rectangle into rational boxes and use interval arithmetic to enclose the residual and its derivatives on each box. Positivity makes all divisions valid.

Interval range sums then enclose integrals of continuous quantities, including the squared residual. As the maximal box diameter tends to zero, the interval widths tend to zero by uniform continuity of the rational expression and its bounded denominators. More efficient derivative-based quadrature estimates can be used, but this elementary subdivision argument already proves deterministic computability to any positive requested absolute tolerance.

For an orthogonal projection, enclose the basis inner products and solve the fixed Gram system with enclosures. Include these coefficient uncertainties in the temporal and spatial residuals. If the target initial quantile is also given by a validated smooth expression, the initial L^2 discrepancy is handled by the same procedure.

S.6 Gaussian oracle and matrix entropy

For rational SPD C, D, the direct inverse-chord quadrature in Appendix N uses rational matrix inverses at rational nodes and an explicit derivative remainder. It therefore gives a deterministic AVF envelope without a validated eigendecomposition. The original Gaussian energy uses $\log\det C$; its determinant is a positive rational number and (91) encloses its logarithm.

When symmetric matrices are stored by upper-triangular entries, their Frobenius inner product has weight two on off-diagonal entries. A validator must preserve those weights in kinetic matrices, norms, and adjoints. Otherwise it may certify the wrong mobility while all its scalar arithmetic is exact. The exact covariance algebra checks in the supplementary verification script include a noncommuting rational example to expose this issue.

S.7 Meaning of the provided verification code

The included tests check finite rational identities, sufficient gates, explicit counterexamples, and symbolic polynomial derivatives. They do not certify the convergence of a large neural training run, validate an unprovided empirical data set, or replace the analytic arguments. They are useful because they make several delicate signs, exponents, and matrix factors independently executable. The source package reports the commands and observed test output separately from the paper's mathematical statements.

Executable coverage. The reference package separates rational interval algebra, the generic kinetic-matrix certificate, structured model oracles, and two bounded-work correction routines. The Gaussian routine uses full symmetric covariance matrices and the Frobenius Gram weights; it never assumes commutation with the confinement. Both routines return an explicit failure status if a midpoint, positivity test, or work limit is unresolved, and no physical-time advancement is implied by failure. A supplied verification run passes 51 finite-arithmetic unit tests and 25 symbolic identity checks. These counts describe executable checks, not 76 independent proofs of the analytic results. No convergence-rate measurement, training run, or empirical performance comparison is included.

T Fixed linear invariants and constant-kernel mobilities

T.1 Restriction to an invariant affine leaf

Let $C_0 : \mathsf{H} \to \mathbb{R}^r$ be linear and let $\mathcal{A} = \{x : C_0 x = b_0\}$ be a nonempty affine space. Its variation space is $H_0 = \ker C_0$, with orthogonal projector P_0. Suppose the target mobility is self-adjoint positive semidefinite with a *fixed* kernel $H_0^{\perp}$, and its restriction to H_0 is uniformly positive on the corridor. Then the flow $F(x) = -K(x)\nabla E(x)$ is tangent to $\mathcal{A}$.

Choose $x_{\mathrm{ref}} \in \mathcal{A}$ and write $x = x_{\mathrm{ref}} + v$, $v \in H_0$. The restricted energy is $E_0(v) = E(x_{\mathrm{ref}} + v)$, with gradient $P_0\nabla E(x)$. Restricting K to H_0 gives a positive definite mobility there. The admissible domain is the relatively open convex set of v for which $x_{\mathrm{ref}} + v \in \mathcal{D}$.

Theorem T.1 (Leafwise transfer of the certificate theory)**.** *Suppose midpoint and endpoint proposals lie in $\mathcal{A}$, and the frozen mobility has the same fixed kernel and is positive on H_0. All finite-dimensional certificate, order, and correction results apply on H_0 with the restricted energy, mobility, and curvature constants. In ambient notation the kinetic norm is the pseudoinverse norm, but it is used only on residuals and displacements in H_0.*

Proof. The restricted problem satisfies exactly the positive-definite assumptions of the preceding theory. Its AVF gradient is $P_0 g$, because projection is fixed and commutes with the chord integral. Since $\delta \in H_0$, $E(y) - E(x) = \langle g, \delta\rangle = \langle P_0 g, \delta\rangle$. The restricted inverse of B agrees with its ambient Moore–Penrose inverse on H_0. Thus the restricted residual, potential, and energy identity coincide

with their ambient formulas after the required projections. The proofs apply without alteration on this fixed Hilbert subspace. □

This covers fixed mass or moment constraints when the mobility and representation enforce them exactly. It does not cover an arbitrary changing-rank neural pullback metric or a state-dependent kernel distribution. In those settings the invariant space may rotate, pseudoinverse derivatives can be singular, and the frozen norm may fail to control relevant errors.

T.2 Why the kernel condition is an additional gate

A pseudoinverse norm vanishes on the kernel. Therefore checking only that norm cannot detect a residual component normal to the invariant leaf. The explicit two-coordinate counterexample in Appendix H demonstrates drift in a nominally conserved variable with zero pseudoinverse residual. To use Theorem T.1, one must check $C_0 y = C_0 x$ and the corresponding midpoint condition, or build them into the parameterization and exact arithmetic.

If an approximate invariant defect is allowed instead, it must appear separately in the trajectory comparison. With a jump-like defect j_n in the conserved coordinates, a norm or distributional bound on its accumulated effect is needed. Simply declaring the pseudoinverse norm to be the error metric would discard a physical state variable rather than control it.

T.3 A coordinate construction without irrational bases

For rational linear constraints, choose a rational full-column-rank matrix Z spanning H_0 and write $x = x_{\mathrm{ref}} + Zq$. The induced fixed Gram matrix is $G = Z^T Z$ (or $Z^T A Z$ in a mass Hilbert metric), which is rational and positive definite. The restricted kinetic matrix, energy covector, and norm calculations can then be assembled directly in these coordinates. An orthonormal basis is unnecessary; one only has to retain the Gram matrix.

This construction is useful for exact validation because orthonormalizing a rational basis would usually introduce irrational entries. It also illustrates a recurring principle of the paper: a nonorthonormal coordinate representation is harmless when the metric is carried explicitly, but harmful when it is silently replaced by an identity matrix.

T.4 Compatibility with positivity

Linear invariants do not imply positivity. The intersection of an affine leaf with a positive-gap domain is relatively open and convex, so the restricted strong-convexity theory remains meaningful. If the energy retains a divergent inverse-gap AVF barrier on every feasible boundary approach, the quantile existence proof extends to this intersection, provided it is nonempty. If the leaf forces a zero gap, the domain is empty and no such step exists. Feasibility is therefore an explicit preliminary check, not a byproduct of minimizing a loss.

U General order-to-gap scaling and optimal inner-budget allocation

U.1 A general conditional exponent rule

The fifth-order sufficient gap scale is a special case of a general perturbation principle. The following theorem does not construct a higher-order energy method; such methods have their own order and structure conditions (Eidnes, 2022; Huang, 2025; Lautwein et al., 2026). It states how a verified inexact solve can preserve an order that has already been proved for an exact method.

Theorem U.1 (Conditional order-to-gap transfer)**.** *Suppose an exact step $\Psi_h(x)$ has local error at most Ch^{p+1} relative to a smooth physical flow, and is the interior minimizer of a potential satisfying*

$$Q_{h,x}(y) - Q_{h,x}(\Psi_h(x)) \geq \frac{\beta_h}{2} \|y - \Psi_h(x)\|_h^2, \qquad \beta_h \geq c/h.$$

Assume $\|v\| \le C_m \|v\|_h$ uniformly, and the exact physical flow has a uniform finite-time Lipschitz bound. If accepted states have certified true-gap upper bounds G_n, then

$$\max_{t_n \le T} \|x_n - u(t_n)\| \le C_T \left(e_0 + \sum_j h_j^{p+1} + \sum_j \sqrt{h_j G_j} \right). \tag{93}$$

In particular, $G_j \le C_G h_j^{2p+1}$ suffices for order p on arbitrary positive meshes with bounded total time.

Proof. The growth inequality and certified gap give $\|x_{n+1} - \Psi_{h_n}(x_n)\| \le C_m \sqrt{2G_n/\beta_{h_n}} \le C\sqrt{h_n G_n}$. Add the exact method's local error and compare exact physical flows starting from x_n and $u(t_n)$. Iterating the resulting Gronwall recurrence proves (93). If $G_n = O(h_n^{2p+1})$, each square-root term is $O(h_n^{p+1})$, whose sum is at most $T h_{\max}^p$ times a constant. □

The proof uses stability of the exact physical evolution, not a blanket contractivity property of numerical Wasserstein resolvents. For the present AVF construction $p = 2$, the residual-based gap upper bound supplies $G_n = O(h_n^5)$. The theorem does not imply that a high-order method automatically has a globally convex potential or an original-energy certificate.

U.2 Worst-case sharpness of the abstract scaling

To see that the exponent cannot generically be weakened without extra cancellation, take the exact scalar flow step $\Psi_h(x) = e^{-h}x$ and the quadratic potential $Q_{h,x}(y) = (y - e^{-h}x)^2/(2h)$, whose curvature is $1/h$. Perturb the endpoint to $y = (e^{-h} - ch^{q+1})x$, with $c > 0$ and $q > 0$. For sufficiently small h, its amplification lies between zero and one, so the original energy $x^2/2$ decreases. Its exact gap is $c^2 h^{2q+1} x^2/2$. At time $T = Nh$, the logarithmic amplification rate is $-1 - ch^q + O(h^{q+1} + h^{2q+1})$, yielding a generic $\Theta(h^q)$ global error. Thus permitting a lower gap exponent can lower the order even when the unperturbed step is exact and energy remains decreasing.

This constructed potential is used only to establish the abstract perturbation obstruction; it is not claimed to be the AVF potential. The AVF-specific scalar construction in Appendix H already establishes the relevant $p = 2$ sharpness within the method studied here.

U.3 A finite-horizon allocation problem

Suppose a prescribed positive mesh and a priori comparison estimate yield an accumulated inner-error allowance $\sum_{n=0}^{N-1} h_n \epsilon_n \le B_{\text{in}}$, with admissible per-step caps $0 < \epsilon_n \le u_n$. These caps may encode the order scale, a known nonstationarity margin, and an oracle budget. Suppose a proven correction-work upper model is

$$W(\epsilon) = \sum_{n=0}^{N-1} a_n \log \frac{A_n}{\epsilon_n}, \qquad a_n > 0, \quad A_n \ge u_n. \tag{94}$$

This logarithmic dependence is motivated by geometric correction convergence. It is a mathematical work model, not a measured cost law for a neural optimizer.

Proposition U.2 (Optimal allocation for the logarithmic model)**.** *If $B_{\text{in}} < \sum_n h_n u_n$, the unique minimizer of* (94) *subject to the stated constraints has*

$$\epsilon_n^\star = \min\left\{ u_n, \frac{a_n}{\lambda h_n} \right\}, \qquad \sum_n h_n \epsilon_n^\star = B_{\text{in}}, \tag{95}$$

where a positive λ solving the last equation exists. The resulting vector is unique. If $B_{\text{in}} \ge \sum_n h_n u_n$, all caps are optimal.

Proof. The objective is strictly convex on positive coordinates, diverges when any coordinate tends to zero, and decreases in each coordinate. Therefore an optimum exists with positive coordinates and is unique. In the first case the budget is saturated; otherwise an unsaturated coordinate could be

increased to lower the objective. The Lagrange stationarity condition is $-a_n/\epsilon_n + \lambda h_n + \nu_n = 0$, with $\nu_n \geq 0$ active only at the upper cap. For an uncapped coordinate it gives $a_n/(\lambda h_n)$; the capped cases yield the minimum formula. The weighted sum is continuous and nonincreasing in λ, tends to the sum of caps as $\lambda \downarrow 0$, and tends to zero as $\lambda \to \infty$. Thus a positive solution exists. Strict convexity gives a unique allocation even if a multiplier is not unique at a degenerate breakpoint. The second case follows from coordinatewise monotonicity. □

The proposition offers a theoretical way to avoid uniformly oversolving every inner problem. Larger a_n allocates a looser tolerance to a more expensive step, unless its accuracy or energy cap is active. The physical-time weighting h_n remains explicit. Replacing it by an unweighted sum would optimize a different error bound.

In an online method, future a_n and caps may be unknown, so this finite-horizon optimum cannot be claimed to be attained. One may use a conservative remaining-budget rule or recompute an allocation for known future steps. The deterministic acceptance gates remain authoritative regardless of the allocation heuristic. Near equilibrium, a small energy cap can dominate the nominal h_n^2 time cap; the work model correctly exposes that additional precision requirement rather than treating it as a training anomaly.

V Learned energies, modeling defects, and nonvanishing error floors

The preceding algorithm allows learned proposals but certifies the intended energy and mobility. A different use of learning is to replace those objects themselves by surrogates. This section makes the resulting additional error explicit. It prevents a guarantee for a surrogate flow from being misreported as a guarantee for the physical one.

V.1 Separating modeling error from integration error

Let $F = -K\nabla E$ be the target field and let $\widehat{F} = -\widehat{K}\nabla\widehat{E}$ be a surrogate on a common corridor. Suppose verified bounds there give

$$\left\|K - \widehat{K}\right\|_{\mathrm{op}} \leq \delta_K, \quad \left\|\nabla E - \nabla\widehat{E}\right\| \leq \delta_g, \quad \|\nabla E\| \leq M_g, \quad \left\|\widehat{K}\right\|_{\mathrm{op}} \leq \widehat{K}_0. \tag{96}$$

Then $\left\|F - \widehat{F}\right\| \leq \delta_F := \delta_K M_g + \widehat{K}_0\delta_g$. This follows by inserting and subtracting $-\widehat{K}\nabla E$.

Theorem V.1 (Posterior error with a surrogate field)**.** *Suppose the target field satisfies the one-sided bound* (35)*. For an admissible reconstruction U, let $\widehat{R} = U' - \widehat{F}(U)$ and suppose $\left\|F(U) - \widehat{F}(U)\right\| \leq \delta_F(t)$. Then*

$$\|U(t) - u(t)\| \leq e^{L_o t}e_0 + \int_0^t e^{L_o(t-s)}[\left\|\widehat{R}(s)\right\| + \delta_F(s)]\,\mathrm{d}s. \tag{97}$$

Proof. The true reconstruction residual is $R = U' - F(U) = \widehat{R} + \widehat{F}(U) - F(U)$. Bound its norm by the sum and apply Lemma F.1. The theorem also holds for an appropriate continuum Hilbert operator when the same domain and regularity premises are supplied. □

For constant $\delta_F > 0$, making the numerical residual arbitrarily small does not remove the integral of δ_F. On a finite horizon, a generic $O(h_{\max}^2)$ total error therefore requires the modeling defect to be of that order as well, or a sharper structure-specific cancellation argument. Under a genuinely contractive target with rate $\alpha > 0$, the model contribution is bounded by δ_F/α uniformly in time, but it is still a nonvanishing floor.

V.2 Certifying a true AVF step from a surrogate gradient

For a frozen step, suppose $\widehat{g} = \int_0^1 \nabla \widehat{E}(x + s\delta)\,\mathrm{d}s$ is computed with numerical envelope σ_{num} in the B norm. If $\left\|\nabla E(v) - \nabla \widehat{E}(v)\right\|_B \le \sigma_{\mathrm{model}}$ along the chord, then

$$\left\|g - \widetilde{\widehat{g}}\right\|_B \le \sigma_{\mathrm{model}} + \sigma_{\mathrm{num}}. \tag{98}$$

The proof is the triangle inequality and integration of the uniform model bound. This combined quantity can be used as σ in the robust acceptance theorem. The global gap bound still needs a curvature certificate for the *true* potential. Convexity of a surrogate potential alone does not prove the true global gap assertion.

If the combined envelope has a fixed positive lower floor σ_0, the residual gate cannot pass for $h < \sqrt{\sigma_0/\tau}$, because $\bar{\varepsilon} \ge \sigma_0$. This is a limitation of the supplied information, not necessarily of the candidate state: the envelope may be conservative. Either improve the surrogate or its enclosure, weaken the requested guarantee, or report that the target tolerance is unresolved. Repeated inner optimization of the same surrogate cannot reduce a fixed modeling envelope.

A midpoint constructed from $\widehat{F}(x)$ has a different scaling: its contribution to the midpoint error is only $h\left\|\widehat{F}(x) - F(x)\right\|/2$. Thus a first-stage field error of $O(h)$ is sufficient for an $O(h^2)$ midpoint. This does not relax the endpoint AVF oracle requirement, where the gradient error appears directly in the residual norm. Different algorithmic locations can require different oracle accuracies.

V.3 Energy-value accuracy alone is insufficient for physical time

Suppose merely $|\widehat{E}(v) - E(v)| \le \delta_0$ on a region. Then

$$E(y) - E(x) \le \widehat{E}(y) - \widehat{E}(x) + 2\delta_0. \tag{99}$$

Thus a sufficiently large verified surrogate energy decrease can certify a true decrease, but an arbitrarily small surrogate decrease need not do so. Near equilibrium the true dissipative margin shrinks, and fixed value uncertainty can dominate it.

There is also no gradient bound from uniform value accuracy alone. On a fixed interval, take $\widehat{E}_\omega(x) = E(x) + \delta_0 \sin(\omega x)$. The value error is at most δ_0, while the derivative error has supremum $\delta_0\omega$, unbounded as $\omega \to \infty$. Hence a small uniform energy-fit error does not certify the physical velocity. Additional derivative regularity or an explicitly validated gradient approximation is needed.

For a convex E and an arbitrarily oscillatory surrogate, the surrogate need not remain convex. This example makes no contrary claim. Its purpose is to identify exactly what cannot be inferred from energy values without another premise. Requiring a convex surrogate is one possible extra structural restriction, but its quantitative consequences would need their own estimates on the relevant domain.

V.4 Model refinement and time refinement

Suppose a sequence of learned models has a certified gradient error $\delta_g(N)$ and mobility error $\delta_K(N)$, with N denoting model resolution or a verified approximation level rather than an optimizer epoch count. Theorem V.1 supplies the total budget $O(h_{\max}^2) + O(\delta_K(N)M_g + \widehat{K}_0\delta_g(N))$ on a fixed horizon, subject to uniform corridor constants. Balancing these two terms is a justified target for joint model and time refinement.

It is not enough to increase N or run more training epochs and assume those certified errors decrease. The needed quantities are uniform or posterior bounds for the deployed states. Polynomial coefficient enclosures, structured analytic models, or a separately justified statistical event can provide such bounds. Unrestricted finite sampled losses cannot, by Proposition H.1.

V.5 Implication for a neural physical-time solver

The strongest guarantee in this paper is obtained when the intended energy and mobility are evaluated or enclosed independently of the predictor. Learning can then accelerate proposals without

changing what is being certified. When the energy or mobility is also learned, the preceding results add explicit modeling terms and expose a possible certification floor. This distinction is particularly important when a network appears to produce smooth, decreasing-energy trajectories: those visual properties may belong to a different dynamical system.

W Claim audit, relation to prior work, and scope of the theory

W.1 A compact claim-to-proof map

Claim	Premises and proof location
Global frozen-step gap	Fixed positive mobility, global curvature, interior minimizer; Appendix B. No neural training assumption.
Original-energy dissipation	An exact chain rule and a validated oracle or chain envelope; Appendix C. Convexity is not needed for the energy identity.
Physical order two	Certified midpoint and mobility, full-state residual scale, uniform finite-time corridor; Appendices D–E.
Finite inner correction	Nonstationarity, convergence of the actual full-state corrector, shrinking oracle envelopes; Appendix G. Exact stationarity is handled separately.
Positive quantile globality	Fixed mass grid, positive inverse-gap diffusion, bounded box terms, positive frozen curvature; Appendices I–J.
Full Gaussian globality	Positive initial and frozen covariance, convex Gaussian entropy, inward boundary derivative; Appendices M–N.
Continuum posterior error	A sufficiently regular positive exact solution, an admissible smooth reconstruction, full residual; Appendices K–L.
Adaptive statistical validity	Predictable candidates and sample counts, fresh draws, range bounds, confidence spent on all calls; Appendix Q.
Uniform-time order	Actual continuous-flow contraction and uniform-in-time corridor constants, in addition to local order; Appendix R.
Degenerate fixed invariants	A fixed kernel and exactly enforced affine leaf; Appendix T. No changing-rank extension.

W.2 Prior ingredients and the intended distinction

The exact AVF method and discrete-gradient energy identities predate this work (Celledoni et al., 2012; Ehrhardt et al., 2025). Order constructions for discrete gradients, high-order discrete variational derivative methods, and high-order dissipative AVF developments demonstrate that high order and energy structure are not new in isolation (Eidnes, 2022; Huang, 2025; Lautwein et al., 2026). Inexact nonlinear solves also arise in this literature, including derivative-free discrete-gradient work (Myhr & Eidnes, 2024). We do not claim the first analysis of an inexact energy integrator.

Neural gradient-flow methods include variational transport and normalizing-flow approaches (Mokrov et al., 2021; Xu et al., 2023), energetic variational neural discretizations (Hu et al., 2024), and neural PDE approximations (Georgoulis et al., 2023). UEPI explicitly combines neural structure with discrete gradients and addresses the role of symmetry in second-order accuracy (Celledoni et al., 2025). Consequently the broad description "an energy-dissipative second-order neural integrator" is not a priority claim of this manuscript.

The intended distinction is the specified *acceptance contract*: a frozen predicted mobility with a physical-clock consistency bound, a global full-state AVF subproblem certificate where existence is actually proved, a signed original-energy test robust to oracle error, and a trajectory certificate that identifies whether its target is a finite model or a continuum equation. The noncommuting Gaussian interior-solvability argument and the explicit smooth-quantile spatial residual make this contract concrete beyond a scalar constant-metric example.

Higher-order Wasserstein schemes have different variational constructions and hypotheses (Carrillo et al., 2022; Matthes & Plazotta, 2019; Gallouët et al., 2024; Chu & Jacobs, 2025). Inexact JKO analysis addresses related approximation questions for minimizing movements (Di Marino et al.,

2025). Our Q is not the JKO functional, so its gap bound should not be presented as a new theorem about every JKO subproblem. Reconstruction comparison is itself an established a posteriori technique (Makridakis & Nochetto, 2006); here it is instantiated with a certified frozen-midpoint residual and, separately, a full smooth quantile operator.

W.3 Limits that matter for using the results

The general finite-state temporal theorem is local in its regularity assumptions and finite-dimensional in its constants. The positive quantile corridor can deteriorate as cell sizes shrink. The Gaussian corridor depends on dimension, confinement, and entropy level. The continuum posterior theorem includes its regularity premise and full spatial defect; it does not establish an a priori approximation rate for arbitrary neural spaces.

The learned component is optional. No theorem here shows that neural training attains the uniform approximation property used in a conditional work bound. No experiment establishes faster correction, better generalization, or lower wall-clock cost. Those omissions are deliberate in this pure-theory version, not claims that such empirical advantages follow from the proofs.

The deterministic arithmetic constructions prove that selected model certificates can be computed to arbitrary positive accuracy under the stated positivity and regularity conditions. They do not guarantee that a finite preset floating-point precision resolves every stationary or equality case. The statistical alternative has explicit failure probability and sampling assumptions. A combined implementation must state which type of guarantee it actually uses.

W.4 Independent review and reproducibility boundary

The mathematical content is presented as proofs with explicit premises, not as a formally verified theorem library. The supplementary exact-arithmetic and symbolic checks cover selected finite identities and counterexamples. They are deliberately separated from analytic assertions involving compactness, existence, regularity, or adaptive infinite sequences.

The accompanying source package contains the full theorem statements and proofs, bibliography, build instructions, and these verification kernels. It contains no empirical plots presented as evidence and no inherited training measurements from an earlier draft. Before external submission, the named authors should independently review the proofs, confirm the intended novelty relative to the full relevant literature, verify authorship and institutional details, and approve the required disclosure. The manuscript does not imply that a conference has accepted it or that a correctness guarantee has been issued by an external reviewer.