SAME MARGINALS, DIFFERENT ERRORS: GAUGE CONTROL FOR PROBABILITY-FLOW ODES

Anonymous authors
Paper under double-blind review

ABSTRACT

Probability-flow ODEs do not uniquely determine how individual samples move. Different velocity fields can generate exactly the same continuous-time marginals. We show that this freedom has a direct numerical consequence: two equivalent velocity fields can incur substantially different discretization errors under the same solver. Based on this observation, we introduce *GaugeFix*, which modifies only the marginal-preserving component of the velocity field while keeping the probability path, numerical solver, and time grid fixed. For Gaussian probability paths, we characterize exactly when the leading covariance error can be cancelled and quantify the unavoidable residual when complete cancellation is impossible. Beyond Gaussian distributions, we derive exact finite-time Wasserstein errors for rotationally symmetric and multi-block paths, and show that GaugeFix improves the distributional convergence order of an order-r Runge–Kutta method to at least $r+1$, and to $r+2$ for symmetric methods. Experiments confirm these predictions and show that the effect persists with learned scores and frozen pretrained diffusion models. On CIFAR-10, GaugeFix improves eight-NFE Euler FID by 14.6% while adding no denoiser evaluations and only a few percent sampling overhead.

1 INTRODUCTION

A probability path specifies where a population of samples should be at each time, but not how each sample must move. Probability-flow ODEs exploit this distinction to express generative sampling as deterministic transport (Song et al., 2021). Related studies have also examined how alternative formulations of a continuous transport or dynamical system affect its numerical discretization (Zheng et al., 2023; Gurfil & Klein, 2006; Libero & Klein, 2026; Lu et al., 2023). Given a density path p_t and a velocity field v_t satisfying

$$\partial_t p_t + \nabla\cdot(p_t v_t) = 0, \tag{1}$$

any perturbation u_t with $\nabla\cdot(p_t u_t) = 0$ preserves every continuous-time marginal. The resulting ODEs can have different trajectories while transporting exactly the same distributions. This *gauge freedom* is established in diffusion models (Horvat & Pfister, 2024). Our starting point is a mismatch between continuous and discrete time: different velocity fields may generate the same continuous-time marginals, yet incur different errors under the same numerical solver. This raises a natural question: can we exploit this freedom to reduce discretization error without changing the solver?

Consider an isotropically scaling Gaussian in two dimensions. The canonical velocity field is the radial field $\dot{x} = cx$, which expands every particle away from the origin at the same exponential rate. Now introduce

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad J^\top = -J, \qquad J^2 = -I.$$

Multiplication by J rotates a vector by 90°. Hence, for any $\omega \in \mathbb{R}$, the modified field $\dot{x} = (cI + \omega J)x$ adds a tangential velocity to the same radial expansion. Its exact flow is $x(t) = e^{ct}e^{\omega t J}x(0)$, where $e^{\omega t J}$ is a rotation matrix. Because an isotropic Gaussian is invariant under rotations, this rotational motion changes particle trajectories but leaves every continuous-time marginal unchanged. Thus the radial and spiral fields are two different representatives of the same probability path.

Despite having identical continuous-time marginals, the two fields behave differently after numerical discretization. One Euler step applied to the spiral field gives

$$x^+ = \big[(1+hc)I + h\omega J\big]x.$$

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.

The two terms $(1 + hc)x$ and $h\omega Jx$ are orthogonal, since $x^\top Jx = 0$, and $\|Jx\| = \|x\|$. Therefore the radius after one Euler step is multiplied by

$$\rho_h(\omega) = \sqrt{(1 + hc)^2 + h^2\omega^2}, \tag{2}$$

whereas the exact continuous flow multiplies the radius by e^{ch}. Expanding factors for small h gives

$$e^{ch} = 1 + ch + \frac{1}{2}c^2h^2 + O(h^3), \qquad \rho_h(\omega) = 1 + ch + \frac{1}{2}\omega^2h^2 + O(h^3).$$

For the radial field, Euler corresponds to $\omega = 0$ and therefore misses the second-order term $\frac{1}{2}c^2h^2$ in the exact radial expansion. The rotational component contributes instead $\frac{1}{2}\omega^2h^2$ to the Euler radius. Choosing $\omega^2 = c^2$ therefore makes this additional contribution exactly match the missing second-order term, so that $\rho_h(\omega) = e^{ch} + O(h^3)$. The solver itself has not changed: we still use Euler with the same step size. The smaller error comes solely from choosing a different continuous-time representative of the same probability path.

This example suggests a constrained form of numerical control. Rather than changing the model, time grid, or integration rule, we replace the original velocity field by a gauge-modified field that preserves the same prescribed marginals. We refer to the original velocity field as *Base* and the gauge-modified field produced by our method as *GaugeFix*. This effect is not specific to Euler. Figure 1 compares Base and GaugeFix for Euler, Heun, RK3, and RK4 in the same two-dimensional Gaussian setting. For each solver, the integration rule and step sizes are identical, while GaugeFix consistently yields lower numerical error. The question is therefore not whether an unrestricted correction can improve sample quality, but whether we can reduce discretization error while leaving the continuous-time probability distributions unchanged.

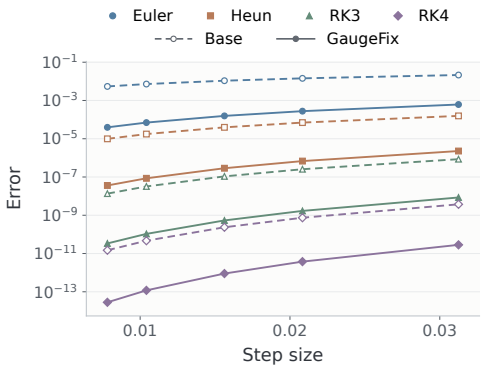

Figure 1: **Same solver, different errors.** Gauge reduces the measured error across step sizes for all four solvers.

Our results give a concrete answer in several settings. For Gaussian scaling paths, complete cancellation is possible precisely when the relevant modes can be paired into two-dimensional rotation blocks. This pairing admits an exact spectral characterization. For time-varying Gaussian paths, explicit sign conditions characterize when cancellation is possible or impossible. Beyond Gaussian distributions, rotational symmetry allows the same mechanism to improve global distributional convergence. Throughout, we measure accuracy at the distribution level: gauge perturbations may change particle trajectories while preserving the same continuous-time distributions. We make three contributions:

- For Gaussian scaling paths, we characterize exactly when a marginal-preserving linear gauge can cancel the leading covariance error, and derive optimal residuals and time-varying Euler/Heun criteria when complete cancellation is impossible.
- Beyond Gaussian distributions, we derive exact finite-time Wasserstein errors and show that gauge control can improve the distributional convergence order of fixed RK solvers, with extensions to nonhomothetic block-torus paths and symmetric RK methods.
- We validate the predicted mechanisms across 90 analytic cases and pretrained diffusion models, using same-field comparisons to distinguish solver-error reduction from changes in the underlying continuous-time distribution.

2 Formulation and Background

This section reviews the gauge structure of probability-flow ODEs, formalizes the distribution-level numerical error studied in this work, and summarizes the most relevant prior work.

2.1 GAUGE FREEDOM IN PROBABILITY-FLOW ODES

We use $\mathcal{N}_{p_t}$ to denote the marginal-preserving perturbations introduced in Section 1. For a prescribed probability path p_t, all velocity fields in the affine class $v_t + \mathcal{N}_{p_t}$ generate the same continuous-time marginals. We refer to the choice of a particular velocity field within this class as a *gauge choice*. Thus, fixing the probability path still leaves freedom in how individual samples are transported.

Our goal is to understand how this freedom interacts with numerical discretization. To do so, we next distinguish numerical errors that change a probability distribution from those that do not. Consider the space $\mathcal{H}_p = L^2(p; \mathbb{R}^d)$. Let $\mathcal{G}_p$ be the closure of gradient fields and let $\mathcal{N}_p$ denote the marginal-preserving fields defined above. Under suitable decay or periodic boundary conditions, weighted integration by parts gives

$$\mathcal{H}_p = \mathcal{G}_p \oplus \mathcal{N}_p. \tag{3}$$

This weighted Hodge decomposition (Horvat & Pfister, 2024; Khelifa et al., 2026) separates a displacement into a component that changes the distribution and one that is invisible to it. For a small displacement e, the first-order change in the density is $q_e = -\nabla \cdot (pe)$. If $e \in \mathcal{N}_p$, then $q_e = 0$, so this component of the displacement is invisible at the distribution level. The weighted Hodge decomposition further shows that the distribution-visible part of a general displacement is precisely its gradient projection $P_{\mathcal{G}_p} e$. We therefore measure the visible numerical error by $\left\| P_{\mathcal{G}_p} e \right\|_{L^2(p)}$. Its equivalent density-space formulation and derivation are given in Appendix A.

We now apply this measure to the local error of a numerical solver. Let $\Phi^w_{t,h}$ denote the exact flow generated by a velocity field w, and let $\Psi^w_{t,h}$ be an order-r numerical step satisfying

$$\Psi^w_{t,h}(x) = \Phi^w_{t,h}(x) + h^{r+1} e^w_{r+1}(x,t) + O(h^{r+2}), \tag{4}$$

where e^w_{r+1} is the leading local discretization error. Its distribution-visible component is therefore $P_{\mathcal{G}_{p_t}} e^w_{r+1}$. For a marginal-preserving family $w_\theta = v + u_\theta$, $u_\theta \in \mathcal{N}_{p_t}$, we define

$$F(\theta) = P_{\mathcal{G}_{p_t}} e^{w_\theta}_{r+1}, \qquad \mathcal{J}(\theta) = \frac{1}{2} \left\| F(\theta) \right\|^2_{L^2(p_t)}. \tag{5}$$

Thus $\mathcal{J}(\theta)$ measures the leading numerical error that is visible at the distribution level. Our theoretical analysis asks how much this quantity can be reduced by changing the gauge while keeping the probability path and the numerical solver fixed.

2.2 RELATED WORK

Fast diffusion sampling has motivated a broad range of specialized numerical methods. DPM-Solver and DPM-Solver++ exploit the semi-linear structure of diffusion ODEs (Lu et al., 2022a;b), while DPM-Solver-v3 constructs an alternative ODE formulation using empirical model statistics (Zheng et al., 2023). Other approaches learn solver parameters (Frankel et al., 2025), use embedded Runge–Kutta errors as guidance (Kong et al., 2026), or incorporate geometric information into the integration rule (Sun et al., 2026). Related numerical ideas include gauge-based truncation-error reduction (Gurfil & Klein, 2006; Libero & Klein, 2026), projection superconvergence (Lu et al., 2023), and phase, amplitude, and symmetry properties of Runge–Kutta methods (Hairer et al., 2006; Chan & Gorgey, 2013; Giri & Sen, 2018; Liu & Brio, 2024).

Marginal- or measure-preserving modifications also arise in other dynamical settings. Nonreversible Langevin dynamics uses skew drifts while preserving a stationary distribution (Lelièvre et al., 2013; Zhu et al., 2026), while recent work studies marginal-preserving particle guidance and non-gradient population dynamics (Vinograd et al., 2026; Berman et al., 2026). Convergence and discretization error for probability-flow ODEs have also been studied directly (Gao & Zhu, 2024; Pierret & Galerne, 2025). Our focus is the finite-time distributional error of a fixed numerical solver when the velocity field is varied only within a marginal-preserving family.

3 GAUGE CONTROL FOR GAUSSIAN PROBABILITY FLOWS

We now study when a marginal-preserving change of velocity field can reduce the distributional error of a fixed solver.

3.1 GAUSSIAN SCALING PATHS

Consider

$$p_t = \mathcal{N}(0, \Sigma_t), \qquad \Sigma_t = Se^{2tB}S, \tag{6}$$

where $S \succ 0$ and $B = B^\top$. In whitened coordinates, the eigenvalues of B describe the growth rates of the Gaussian along different directions.

The base linear velocity field has generator SBS^{-1}. We add a rotational component through

$$W = S(B + Q)S^{-1}, \qquad Q^\top = -Q.$$

The matrix Q is the higher-dimensional analogue of the 2×2 rotation matrix used in Section 1. To preserve the same covariance path, the rotation must satisfy $[B, Q] = 0$, where $[B, Q] = BQ - QB$ is the matrix commutator. This condition means that Q only mixes directions with the same growth rate, or equivalently, only directions within the same eigenspace of B.

Now fix an explicit Runge–Kutta method of order r, with stability polynomial

$$P(z) = e^z + \gamma z^{r+1} + O(z^{r+2}), \qquad \gamma \neq 0. \tag{7}$$

Although an admissible Q leaves the exact Gaussian path unchanged, it changes the covariance produced by a finite RK step. We therefore compare the numerical covariance after one step with the exact covariance at the same time. Their difference has a leading term of order h^{r+1}, and our goal is to choose Q so that this leading covariance error vanishes.

Because Q can act only within eigenspaces of B, this cancellation problem separates across groups of directions with the same growth rate. This leads to a pairing condition.

Theorem 1. *Let the distinct nonzero eigenvalues of B have multiplicities m_j. A marginal-preserving linear gauge can completely cancel the leading covariance error if and only if every m_j is even.*

The condition has a simple geometric interpretation. Since a real skew-symmetric matrix acts through two-dimensional rotation blocks, directions inside each eigenspace must be paired. If an eigenspace has odd dimension, one direction is necessarily left unpaired, so complete cancellation is impossible.

For an eigenspace with eigenvalue λ_j, the cancelling rotation frequency can be chosen as

$$\omega_j = |\lambda_j| \tan\left(\frac{\pi}{2(r+1)}\right). \tag{8}$$

When some eigenspaces have odd multiplicity, the leading error cannot be removed completely, but its smallest possible residual is

$$\rho_r^\star = \left(\frac{\sum_{j:\, m_j \text{ odd}} |\lambda_j|^{2(r+1)}}{\sum_j m_j |\lambda_j|^{2(r+1)}}\right)^{1/2}. \tag{9}$$

Thus the obstruction is not the ambient dimension itself, but whether the nonzero growth modes can be paired within equal-rate eigenspaces.

For the isotropic case $B = cI$, this reduces to a particularly simple rule: complete cancellation is possible in even dimension, while an odd dimension leaves one unpaired direction. The full proof is given in Appendix B.

3.2 TIME-VARYING GAUSSIAN PATHS

The previous result assumes that the Gaussian grows with fixed rates. We next allow the scale to vary with time. Consider $p_t = \mathcal{N}(0, s(t)\Sigma)$, and define the instantaneous growth rate

$$c(t) = \frac{1}{2}\frac{d}{dt}\log s(t).$$

A marginal-preserving linear field can again add a rotational component, which we write as

$$\Sigma^{1/2}\big(c(t)I + Q(t)\big)\Sigma^{-1/2}x, \qquad Q(t)^\top = -Q(t).$$

For Euler and Heun, the leading covariance error takes a particularly simple form.

Proposition 2. *For Euler, define*

$$a_{\mathrm{E}}(t) = c(t)^2 + \dot{c}(t).$$

For Heun, when $c(t) \neq 0$, define

$$a_{\mathrm{H}}(t) = \frac{c(t)^2}{3} - \frac{\ddot{c}(t)}{6c(t)}.$$

At a fixed time, if the corresponding quantity $a(t)$ is positive, the leading covariance error can be cancelled on a two-dimensional rotation block by choosing a rotation frequency $\sqrt{a(t)}$. If $a(t) = 0$, adding a rotation does not improve the leading term. If $a(t) < 0$, no real rotational gauge can cancel that term.

These conditions show that the usefulness of a gauge depends not only on the solver, but also on how the Gaussian scale changes over time. An exponentially growing variance satisfies the cancellation condition, whereas for a linearly growing variance Euler and Heun give a negative criterion, so the leading error cannot be cancelled within this gauge family. The derivations are given in Appendix B.

4 Exact global order lifting beyond Gaussian laws

We now show that Gaussianity is not essential: rotational symmetry alone is enough to obtain exact finite-time distributional errors and improved global convergence.

4.1 Circle-invariant laws

We begin with distributions that are unchanged by a continuous rotation. Let

$$J^\top = -J, \qquad J^2 = -I,$$

and assume that a random variable $Z \sim \nu$ satisfies

$$e^{\alpha J} Z \sim \nu \qquad \text{for every } \alpha \in \mathbb{R}.$$

Thus rotating a sample changes its position, but not its distribution. Unlike the Gaussian setting of Section 3, ν need not be Gaussian. We only assume that it has a finite second moment.

To allow a general linear change of coordinates, let $\mu_0 = S_{\#}\nu$, where $S_{\#}\nu$ denotes the distribution of SZ when $Z \sim \nu$. We consider a probability path obtained by uniformly rescaling the samples: $mu_t = (e^{ct} I)_{\#}\mu_0$. That is, every sample is multiplied by the same factor e^{ct} at time t. The velocity field

$$w(x) = S(cI + \omega J)S^{-1}x$$

combines the same radial expansion with an additional rotation. Because the underlying distribution is invariant under this rotation, changing ω changes individual trajectories but leaves the continuous-time probability path μ_t unchanged.

Now apply the same fixed Runge–Kutta method for n steps of size $h = T/n$. A matrix of the form $aI + bJ$ simply rotates the samples and rescales them by a common factor. Because the rotation does not change the distribution, only the scaling error remains at the endpoint. This leads to an exact expression for the final Wasserstein error.

Theorem 3. *Apply n steps of a real explicit-RK stability polynomial P with $h = T/n$. Define $\rho_h = |P(h(c + i\omega))|$ and $m_2 = \int \|x\|^2 \, d\mu_0(x)$. Then*

$$\widehat{\mu}_T = (\rho_h^n I)_{\#}\mu_0, \qquad W_2(\widehat{\mu}_T, \mu_T) = |\rho_h^n - e^{cT}|\sqrt{m_2}. \tag{10}$$

If equation 7 holds, $c \neq 0$, and $m_2 > 0$, the zero gauge has global law order r. Choosing

$$\omega^\star = c \tan\left(\frac{\pi}{2(r+1)}\right) \tag{11}$$

raises this order to at least $r + 1$, while common-particle trajectory order remains r. The law order is exactly $r + 1$ for Euler, Heun, classical RK3, and classical RK4.

A proof is given in Appendix C. We briefly summarize the main idea here. Every numerical step is a rotation followed by a dilation (i.e., a uniform rescaling of all samples). Symmetry removes the rotation in law, leaving only a scalar scale mismatch. Coupling the same latent point under the two rescalings yields the exact Wasserstein expression in Eq. equation 10. The accumulated log-radius error is

$$n \log |P(h(c + i\omega))| - cT = T\gamma \operatorname{Re}(c + i\omega)^{r+1} h^r + O(h^{r+1}). \tag{12}$$

The choice in equation 11 cancels the leading scaling error, while a rotational error remains at the trajectory level. This is why the distribution can converge faster than individual particle trajectories. The exact Wasserstein formula uses standard optimal-transport geometry for uniformly rescaled distributions (Ambrosio et al., 2008; Villani, 2009).

Importantly, the result does not require Gaussianity or full spherical symmetry. It applies to any rotation-invariant distribution with finite second moment, including heavy-tailed, multimodal, compactly supported, and even singular distributions.

4.2 An additional order for symmetric RK methods

The previous result gives at least one additional distributional order for a general RK method. For time-symmetric RK methods, the same gauge can remove one more order of scaling error. Such methods have a stability function satisfying $R(z)R(-z) = 1$, which expresses the time-reversibility of the numerical update at the level of the stability function.

Corollary 4 (Two-order improvement for symmetric RK)**.** *Under the rotational-symmetry assumptions of Theorem 3, let R be a real analytic stability function of even order r satisfying $R(z)R(-z) = 1$. For $c \neq 0$ and $m_2 > 0$, choosing the same gauge frequency equation 11 gives*

$$W_2(\widehat{\mu}_T, \mu_T) = O(h^{r+2}), \qquad \big(\mathbb{E}\left\|\widehat{X}_T - X_T\right\|^2\big)^{1/2} = \Theta(h^r). \tag{13}$$

In particular, implicit midpoint improves from distributional order 2 to 4, while two-stage Gauss–Legendre improves from 4 to 6.

A proof is given in Appendix C. The key additional ingredient is the time symmetry of the solver. The identity $R(z)R(-z) = 1$ means that a forward step and the corresponding backward step cancel at the level of the stability function. Since $R(0) = 1$, $\log R(z)$ is well defined near $z = 0$. Taking logarithms gives

$$\log R(-z) = -\log R(z),$$

so $\log R(z)$ is an odd function and its Taylor expansion contains only odd powers.

For an even-order method of order r, the first error term in this expansion has degree $r + 1$. The gauge choice in equation 11 makes this leading contribution purely rotational, so it does not affect the scale of the distribution. Time symmetry then rules out a term of degree $r + 2$, because even powers cannot appear in $\log R$. The next possible scaling contribution therefore has degree $r + 3$; after accumulation over T/h steps, this yields a global distributional error of order h^{r+2}. This proves the two-order improvement stated in Corollary 4. The full expansion is given in Appendix C.

4.3 Multiple rotation blocks with different growth rates

So far, every direction has shared the same scaling factor e^{ct}. We now allow different groups of coordinates to grow at different rates. Split the space into orthogonal blocks,

$$E = E_1 \oplus \cdots \oplus E_m,$$

and suppose that each block E_j has its own rotational symmetry, represented by a matrix J_j with

$$J_j^\top = -J_j, \qquad J_j^2 = -I.$$

We assume that the joint distribution is unchanged when the blocks are rotated independently. The blocks themselves need not be statistically independent. Let $S = \operatorname{diag}(S_1, \ldots, S_m)$ and $\mu_0 = S_\# \nu$. We consider the probability path

$$\mu_t = \operatorname{diag}(e^{c_1 t} I, \ldots, e^{c_m t} I)_\# \mu_0, \tag{14}$$

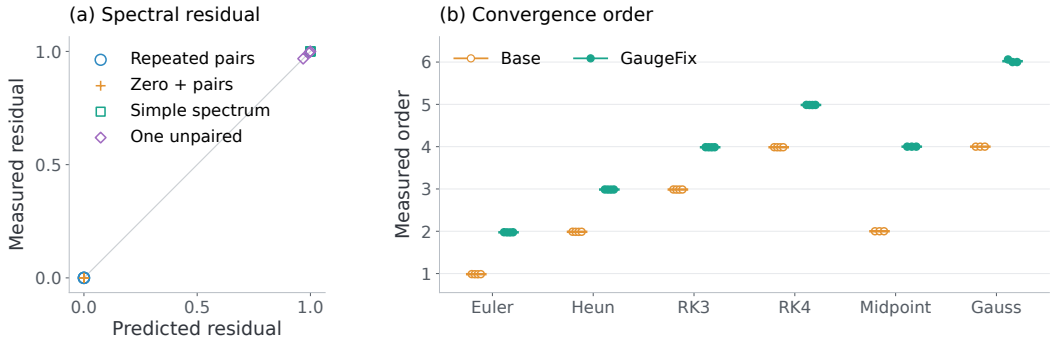

Figure 2: **Cancellation and convergence on analytic paths. (a)** Predicted versus measured normalized leading residuals. **(b)** Measured endpoint distribution orders for Base and GaugeFix. Lower residuals and higher orders are better.

where block j is rescaled at rate c_j. When the c_j are different, the distribution changes shape over time rather than being uniformly rescaled as in Section 4.

Because rotations remain invisible within each block, we can apply the same gauge idea separately to each scaling rate.

Theorem 5. *For block j, use*

$$w_j(x_j) = S_j(c_j I + \omega_j J_j)S_j^{-1}x_j,$$

and define

$$\rho_{j,h} = |P(h(c_j + i\omega_j))|, \qquad m_{2,j} = \int \|x_j\|^2 \, d\mu_0(x).$$

After $n = T/h$ steps,

$$W_2^2(\widehat{\mu}_T, \mu_T) = \sum_{j=1}^{m} m_{2,j}\left(\rho_{j,h}^n - e^{c_j T}\right)^2. \tag{15}$$

Choosing

$$\omega_j^\star = |c_j| \tan\left(\frac{\pi}{2(r+1)}\right)$$

for each block with $c_j \neq 0$ improves the global distributional order from r to at least $r+1$. For Euler, Heun, classical RK3, and classical RK4, the improved order is exactly $r+1$.

The interpretation is the same as in the one-block case. Each block has its own scaling error, and its rotational gauge cancels the leading term independently. Importantly, the blocks need not be probabilistically independent. Arbitrary dependence between their rotation-invariant quantities is allowed. A proof and the exact endpoint representation are given in Appendix C.

5 Experiments

We evaluate our proposed method, GaugeFix, first on constructed probability paths and then on frozen pretrained diffusion models. All learned adapters are frozen before held-out evaluation. Appendix D analyzes how approximate scores can break exact marginal preservation, while Appendix E provides full experimental protocols and additional results.

5.1 Analytic paths: residual cancellation and convergence

Setup. We first use probability paths whose continuous-time solutions and numerical errors can be evaluated directly. For Gaussian paths, $p_t = \mathcal{N}(0, Se^{2tB}S)$, we vary the eigenvalue multiplicities of B to test complete, partial, and impossible cancellation. We also evaluate non-Gaussian rotation-invariant and multi-block paths. *Base* denotes the zero-gauge velocity field; GaugeFix uses the same initial distribution, solver, and step-size grid and changes only its marginal-preserving component. Further settings are given in Appendix E.

Table 1: **Paired pretrained CIFAR-10 comparisons.** Within each solver, Base, GaugeFix, and Free use the same checkpoint, random seeds, and computational budget. Percentages in the FID columns denote improvement over Base. Positive Δgap means a smaller same-field solver gap.

Solver	Base FID	GaugeFix FID	Free FID	GaugeFix Δgap	Free Δgap
Euler	27.995	23.896 (+14.64%)	24.645 (+11.96%)	+15.03%	+10.55%
Heun	3.050	2.996 (+1.77%)	2.878 (+5.61%)	+2.18%	+7.13%

Leading-error cancellation. Figure 2(a) reports the leading covariance error after GaugeFix relative to Base. Paired nonzero growth rates permit complete cancellation and give residuals near zero (*Repeated pairs* and *Zero + pairs*). A simple spectrum leaves no pairable directions and therefore remains near one (*Simple spectrum*). Leaving one direction unpaired gives an intermediate residual (*One unpaired*). The measured values closely follow these predictions.

Endpoint convergence. At the final time T, we measure $W_2(\widehat{\mu}_T, \mu_T)$ between the numerical and exact endpoint distributions and estimate how quickly it decreases as the step size is refined. Figure 2(b) shows that GaugeFix raises the distributional order of Euler, Heun, RK3, and RK4 by approximately one. For symmetric methods, implicit midpoint improves from order 2 to 4 and two-stage Gauss–Legendre from 4 to 6, while their trajectory orders remain unchanged. Additional time-varying and dependent-block checks are reported in Appendix E.

5.2 Toy paths: exact and learned scores

Setup. We next consider two nonlinear two-dimensional probability paths: an eight-component ring Gaussian mixture (GMM8) and a checkerboard mixture. All experiments use an eight-step Euler solver. We first use analytic scores and then replace them with frozen supervised MLP scores. With approximate scores, target quality can change because the continuous-time field itself may drift. We therefore additionally measure *solver error* by comparing coarse and high-accuracy integrations of the same frozen field, as motivated in Appendix D.

Results. Figure 3 summarizes the three measurements. With exact scores, GaugeFix reduces target sliced-W_2 by 65.5% on GMM8 and 64.1% on checkerboard. After replacing the analytic score with a learned score, the target error is still reduced on both paths. More importantly, the same-field comparison shows solver-error reductions of 25.6% and 33.6%, respectively.

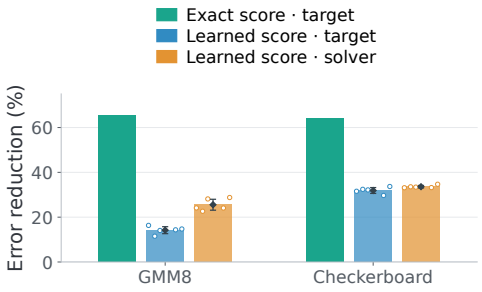

Figure 3: **Toy-path results.** Target-error reduction with exact and learned scores, together with same-field solver-error reduction for learned scores. Circles denote the five held-out seeds for the learned-score experiments; diamonds and error bars show mean $\pm$ standard deviation.

5.3 Pretrained diffusion models

Setup. We evaluate GaugeFix on the public EDM VP checkpoint for CIFAR-10 (Karras et al., 2022). Within each solver, Base and GaugeFix use the same checkpoint, noise schedule, latent seeds, and computational budget. GaugeFix combines 3 RGB channel-rotation operators and 3 spatial finite-difference operators. Each operator is weighted by a combination of 3 basis functions, giving 18 scalar coefficients in total. We optimize these coefficients while keeping the pretrained checkpoint frozen. The correction adds no denoiser evaluations and incurs about 2–3% measured sampling overhead for Euler and Heun.

We report FID50k using 50,000 generated images. With approximate scores, a lower FID does not necessarily imply smaller numerical error, so we also measure a *same-field gap*: the distance between the numerical endpoint and a more accurate integration of the same frozen velocity field. A smaller gap indicates lower solver error. We compare GaugeFix with a matched 18-parameter unconstrained correction (*Free*). Full solver settings and evaluation protocols are given in Appendix E.

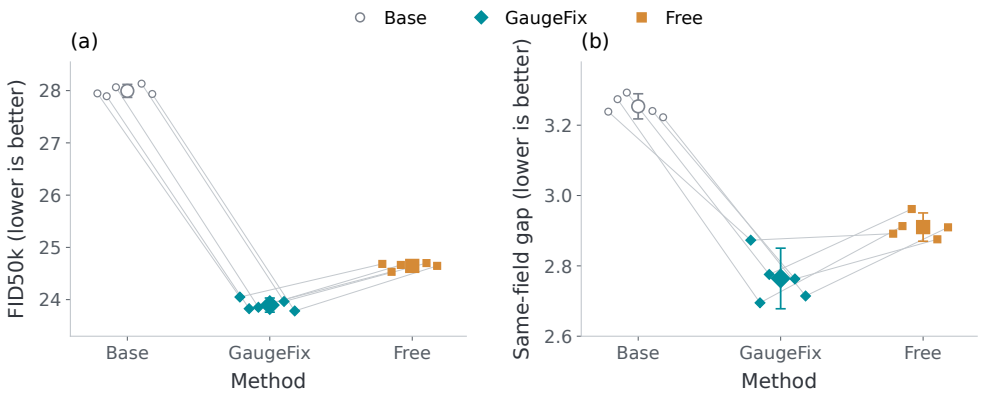

Figure 4: **Pretrained CIFAR-10 robustness.** FID50k and same-field solver gap for Euler across five independent evaluations. Each gray line connects Base, GaugeFix, and the matched 18-parameter Free correction evaluated with the same random seeds. Lower is better.

Table 2: **Adapter component ablation on pretrained CIFAR-10.** Lower FID50k is better.

Variant	Active coefficients	Euler FID50k ↓	Heun FID50k ↓
Base	0	27.995	3.050
Channel only	9	24.639	3.000
Spatial only	9	27.007	3.022
Full GaugeFix	18	**23.896**	2.996

Main results. As shown in Figure 4, GaugeFix consistently reduces both FID50k and the same-field solver gap relative to Base across all five evaluations. It also outperforms the matched Free correction on both metrics. On average, FID50k decreases from 27.995 to 23.896 (14.64%), while the same-field gap is reduced by 15.03%. Table 1 compares GaugeFix with the matched Free correction under both Euler and Heun. GaugeFix gives larger improvements with Euler. With Heun, GaugeFix still reduces both FID and the same-field gap. We further evaluate GaugeFix with DPM-Solver++(2M) (Lu et al., 2022b). The same-field gap decreases by 2.14%. For comparison, ERK-Guid (Kong et al., 2026) reaches FID 2.707 in the Heun setting, compared with 2.996 for GaugeFix. DPM-Solver-v3 (Zheng et al., 2023), evaluated as a standalone solver, reaches FID 3.581. These results show that GaugeFix is not a replacement for specialized diffusion solvers. Its role is to improve discretization by changing the velocity field while keeping the underlying solver fixed.

Ablation study. The *Channel only* and *Spatial only* variants retain the 9 coefficients associated with channel rotations and spatial differences, respectively. The other 9 coefficients are set to zero without retraining. Table 2 shows that, the channel-only variant recovers most of the improvement over Base. GaugeFix achieves the lowest FID, indicating that retaining the spatial component provides an additional benefit beyond the channel correction alone.

6 Conclusion and Limitations

We showed that choosing among velocity fields with the same continuous-time distributions can reduce numerical error without changing the solver. For Gaussian scaling paths, we characterized when the leading covariance error can be cancelled and quantified the residual otherwise. For rotationally symmetric non-Gaussian paths, we derived exact finite-time Wasserstein errors and established convergence improvements of at least one order, or two for symmetric RK methods. In our CIFAR-10 experiments, GaugeFix improves both FID and the same-field gap with Euler and Heun, while keeping the pretrained checkpoint frozen.

Our exact guarantees rely on the stated linear structure and symmetry assumptions. Extending them to general nonlinear flows with approximate scores is an important direction for future work.

AI USE STATEMENT

We used generative AI tools to assist with refining mathematical formulations, checking proofs, and implementing methods. These tools also assisted with literature searches and manuscript editing. The authors reviewed the mathematical arguments, tested the experimental code, and verified the references against their original sources. We take responsibility for the final manuscript, including its claims, proofs, references, code, figures, and experimental results.

REFERENCES

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser, 2008.

Jules Berman, Tobias Blickhan, and Benjamin Peherstorfer. Leveraging gauge freedom for learning non-gradient population dynamics of stochastic systems. *arXiv preprint arXiv:2605.25107*, 2026.

Mikołaj Bińkowski, Dougal J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In *International Conference on Learning Representations*, 2018.

R. P. K. Chan and Annie Gorgey. Active and passive symmetrization of Runge–Kutta Gauss methods. *Applied Numerical Mathematics*, 67:64–77, 2013. doi: 10.1016/j.apnum.2011.06.013.

Eric Frankel, Sitan Chen, Jerry Li, Pang Wei Koh, Lillian J. Ratliff, and Sewoong Oh. S4S: Solving for a fast diffusion model solver. In *International Conference on Machine Learning*, volume 267, pp. 17542–17574, 2025.

Xuefeng Gao and Lingjiong Zhu. Convergence analysis for general probability flow ODEs of diffusion models in Wasserstein distances. *arXiv preprint arXiv:2401.17958*, 2024.

Subhajit Giri and Shuvam Sen. Phase error analysis of implicit Runge–Kutta methods: Introducing new classes of minimal dissipation low dispersion high order schemes. *arXiv preprint arXiv:1804.07979*, 2018.

Pini Gurfil and Itzik Klein. Mitigating the integration error in numerical simulations of Newtonian systems. *International Journal for Numerical Methods in Engineering*, 68(2):267–297, 2006. doi: 10.1002/nme.1716.

Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer, 2 edition, 2006.

Christian Horvat and Jean-Pascal Pfister. On gauge freedom, conservativity and intrinsic dimensionality estimation in diffusion models. In *International Conference on Learning Representations*, 2024.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577, 2022.

Naïl B. Khelifa, Richard E. Turner, and Ramji Venkataramanan. Diffusion models observe only gradients: A geometric perspective on score matching errors. *arXiv preprint arXiv:2606.06179*, 2026.

Inho Kong, Sojin Lee, Youngjoon Hong, and Hyunwoo J. Kim. Error as signal: Stiffness-aware diffusion sampling via embedded Runge–Kutta guidance. In *International Conference on Learning Representations*, 2026.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Tony Lelièvre, Francis Nier, and Grigorios A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics*, 152:237–274, 2013.

Yaakov Libero and Itzik Klein. Gauge freedom optimization for truncation error reduction in inertial navigation. *arXiv preprint arXiv:2606.17202*, 2026.

Jinjie Liu and Moysey Brio. Energy-superconvergent Runge–Kutta time discretizations. *arXiv preprint arXiv:2405.05448*, 2024.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, volume 35, pp. 5775–5787, 2022a.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022b.

Nan Lu, Wenjun Cai, Yonghui Bo, and Yushun Wang. Superconvergence of projection integrators for conservative system. *Journal of Computational Physics*, 490:112281, 2023. doi: 10.1016/j.jcp.2023.112281.

Emile Pierret and Bruno Galerne. Diffusion models for Gaussian distributions: Exact solutions and Wasserstein errors. In *International Conference on Machine Learning*, 2025.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Chaoqun Sun, Zongjing Fu, Powei Chang, Jinpeng Zhang, Jianxiang Xiang, Yukang Gao, and Chenyu Wang. GeoRK2: Geometry-guided Runge–Kutta integration for diffusion transformer acceleration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9404–9413, 2026.

Cédric Villani. *Optimal Transport: Old and New*. Springer, 2009.

Gal Vinograd, Idan Achituve, and Ethan Fetaya. Diverse sampling in diffusion models with marginal preserving particle guidance. *arXiv preprint arXiv:2605.06553*, 2026.

Kaiwen Zheng, Cheng Lu, Jianfei Chen, and Jun Zhu. DPM-Solver-v3: Improved diffusion ODE solver with empirical model statistics. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Qianyu Julie Zhu, Youssef Marzouk, Konstantinos Spiliopoulos, and Benjamin Zhang. Optimizing irreversible perturbations of the unadjusted Langevin algorithm. *arXiv preprint arXiv:2606.05324*, 2026.

A DISTRIBUTION-LEVEL LOCAL ERROR

This appendix supplies the density-space formulation and derivation used in Section 2. Appendix B gives the Gaussian proofs, Appendix C gives the global Wasserstein proofs, Appendix D treats approximate scores, and Appendix E records the experimental protocols and additional results.

A.1 ASSUMPTIONS AND WEIGHTED DECOMPOSITION

We work on a periodic domain or on $\mathbb{R}^d$ with sufficient decay to remove boundary terms. Let p be a smooth positive probability density. In $\mathcal{H}_p = L^2(p;\mathbb{R}^d)$, let $\mathcal{G}_p$ be the closure of gradients of smooth test functions, and interpret

$$\mathcal{N}_p = \{u \in \mathcal{H}_p : \nabla\cdot(pu) = 0\}$$

in the distributional sense. For every admissible test function φ,

$$\int p\, u^\top \nabla\varphi\, dx = -\langle \nabla\cdot(pu), \varphi\rangle.$$

Thus $\mathcal{N}_p = \mathcal{G}_p^\perp$, which proves the orthogonal decomposition in Eq. equation 3. This is the weighted gradient/divergence-free structure used in the cited gauge and score-error analyses (Horvat & Pfister, 2024; Khelifa et al., 2026).

For the local flow expansion below, the velocity field is spatially C^{r+2} and sufficiently smooth in time. We assume that Eq. equation 4 holds uniformly on the time interval under consideration, with integrable bounds sufficient for the displayed Taylor expansions. The global symmetry results in Appendix C instead allow distributions without densities and require only the stated finite-second-moment assumptions.

A.2 DENSITY CHANGE AND ITS NORM

For a smooth displacement field e, set $T_\varepsilon(x) = x + \varepsilon e(x)$. Testing its pushforward against φ gives

$$\begin{aligned}\int \varphi\, d(T_{\varepsilon\#}p) - \int \varphi p\, dx &= \varepsilon \int p\, \nabla\varphi^\top e\, dx + O(\varepsilon^2)\\ &= -\varepsilon\langle \nabla\cdot(pe), \varphi\rangle + O(\varepsilon^2).\end{aligned}$$

The first-order density change is therefore $q_e = -\nabla\cdot(pe)$. For a zero-mass density perturbation or distribution q, define

$$\|q\|_{-1,p} = \sup_{\varphi:\, \|\nabla\varphi\|_{L^2(p)}>0} \frac{|\langle q,\varphi\rangle|}{\|\nabla\varphi\|_{L^2(p)}}. \tag{16}$$

The supremum is over the same smooth tests as above. This definition is the weighted negative Sobolev norm; for an integrable q, the pairing is $\langle q,\varphi\rangle = \int \varphi q\, dx$.

Proposition 6 (Density-space expression of the visible error)**.** *For $e \in \mathcal{H}_p$ and $q_e = -\nabla\cdot(pe)$,*

$$\|q_e\|_{-1,p} = \left\|P_{\mathcal{G}_p} e\right\|_{L^2(p)}. \tag{17}$$

Proof. Weighted integration by parts and orthogonality give

$$\langle q_e,\varphi\rangle = \int p\, \nabla\varphi^\top e\, dx = \left\langle \nabla\varphi, P_{\mathcal{G}_p} e\right\rangle_{L^2(p)}.$$

Cauchy–Schwarz bounds the quotient in Eq. equation 16 by $\left\|P_{\mathcal{G}_p} e\right\|_{L^2(p)}$. Conversely, the definition of $\mathcal{G}_p$ as a closure permits approximation of its nonzero element $P_{\mathcal{G}_p} e$ by gradients. Normalizing that approximating sequence attains the supremum in the limit. If the projection is zero, both sides vanish. □

A.3 Application to a numerical step

Write $\Phi = \Phi^w_{t,h}$, $\Psi = \Psi^w_{t,h}$, and $e = e^w_{r+1}(\cdot, t)$ as in Eq. equation 4. For a smooth test,

$$\int [\varphi(\Psi(x)) - \varphi(\Phi(x))] p_t(x)\, dx = h^{r+1} \int p_t(x)\, \nabla\varphi(\Phi(x))^\top e(x)\, dx + O(h^{r+2})$$
$$= h^{r+1} \int p_t(x)\, \nabla\varphi(x)^\top e(x)\, dx + O(h^{r+2}).$$

The second equality uses $\Phi(x) = x + O(h)$. Consequently, the leading one-step distribution error is $-h^{r+1}\nabla\cdot(p_t e)$ in the weak sense. The use of the starting density p_t and starting coordinates is valid for this leading coefficient; the replacement contributes only to the next power of h.

Applying Eq. equation 17 to the marginal-preserving family in Eq. equation 5 gives the equivalent expression

$$\mathcal{J}(\theta) = \frac{1}{2} \left\| -\nabla\cdot \left(p_t e^{w_\theta}_{r+1}\right) \right\|^2_{-1,p_t}.$$

This establishes the density-space interpretation of the objective retained in the main text. The explicit Gaussian cancellation results are proved directly in Appendix B.

B Gaussian probability flows

We prove Theorem 1, the residual formula equation 9, and Proposition 2. Throughout the autonomous analysis, $S \succ 0$, $B = B^\top$, and the gauge is a time-independent linear field, as in Section 3.

B.1 Preserving the covariance path

In coordinates $y = S^{-1}x$, the initial covariance is I. A constant generator M produces covariance $e^{tM}e^{tM^\top}$. Matching its first derivative at zero to the prescribed path e^{2tB} requires $M + M^\top = 2B$, so $M = B + Q$ for a real skew-symmetric Q.

If $BQ = QB$, then

$$e^{t(B+Q)}e^{t(B-Q)} = e^{2tB}.$$

Conversely, equality for all t implies equality of the second derivatives at zero:

$$(B+Q)^2 + 2(B+Q)(B-Q) + (B-Q)^2 = 4B^2 + 2(QB - BQ) = 4B^2.$$

It follows that $BQ = QB$. This proves both necessity and sufficiency of the commutation condition for the autonomous linear family.

B.2 Leading covariance error and spectral pairing

Proof of Theorem 1 and Eq. equation 9. Let $M = B+Q$ with $BQ = QB$. Starting from covariance I in the transformed coordinates, one numerical step maps a sample to $P(hM)y$. Its covariance is therefore $P(hM)P(hM)^\top = P(h(B+Q))P(h(B-Q))$. The exact covariance at the same time is e^{2hB}. Expanding Eq. equation 7 gives

$$P(hM)P(hM^\top) - e^{2hB} = 2\gamma h^{r+1} D_r(Q) + O(h^{r+2}),$$
$$D_r(Q) = \mathrm{Sym}\big((B+Q)^{r+1}\big), \qquad \mathrm{Sym}(H) = \tfrac{1}{2}(H + H^\top). \tag{18}$$

Thus the coefficient to be cancelled is $D_r(Q)$, rather than the entire particle-displacement error.

Since B is symmetric and commutes with Q, each eigenspace of B is invariant under Q. Within an eigenspace of dimension m_j, an orthogonal change of basis writes Q as two-dimensional skew blocks ωJ and, if m_j is odd, at least one zero direction. On a pair with growth rate λ_j,

$$\mathrm{Sym}\big((\lambda_j I + \omega J)^{r+1}\big) = \mathrm{Re}(\lambda_j + i\omega)^{r+1} I_2.$$

Choosing $\omega = |\lambda_j| \tan(\pi/[2(r+1)])$ makes this real part zero for either sign of λ_j. An unpaired direction contributes λ_j^{r+1} and cannot be changed by a skew field on that direction. Every nonzero

eigenspace can therefore be cancelled exactly if and only if its dimension is even. On the zero eigenspace, choose $Q = 0$.

For $B \neq 0$, normalize in the transformed coordinates by

$$\rho_r(Q) = \frac{\|D_r(Q)\|_{\mathrm{F}}}{\|B^{r+1}\|_{\mathrm{F}}}.$$

Every odd-dimensional nonzero eigenspace leaves at least one squared contribution $|\lambda_j|^{2(r+1)}$. Pairing all other directions attains this lower bound, while $\left\|B^{r+1}\right\|_{\mathrm{F}}^2 = \sum_j m_j |\lambda_j|^{2(r+1)}$. This proves Eq. equation 9 and the optimizer in Eq. equation 8. □

The residual in Eq. equation 9 uses the Frobenius norm in the transformed coordinates with initial covariance I. Complete cancellation is unchanged by the invertible transformation S; the numerical value of a normalized residual in another metric need not be the same. If $B = 0$, the zero gauge already has zero error, so this normalized ratio is not defined.

For $B = cI$ with $c \neq 0$, the optimum is zero in even dimension and $1/\sqrt{d}$ in odd dimension. More generally, a zero growth direction can coexist with fully paired nonzero eigenspaces in an odd-dimensional space. It is the multiplicity of each nonzero eigenvalue, not the total dimension alone, that determines cancellation.

B.3 Finite-time covariance accumulation

The matrices M and $M^\top$ commute. Consequently, after $n = T/h$ steps,

$$\widehat{C}_T = \left[P(hM)P(hM^\top)\right]^n.$$

For sufficiently small h, a matrix logarithm yields

$$n \log\big(P(hM)P(hM^\top)\big) = 2TB + 2T\gamma h^r D_r(Q) + O(h^{r+1}).$$

If $D_r(Q) = 0$, then $\widehat{C}_T - e^{2TB} = O(h^{r+1})$. If $D_r(Q) \neq 0$, its nonzero leading term is $2T\gamma h^r e^{2TB} D_r(Q)$, and covariance order r remains. Multiplication by S transfers these estimates to the original covariance. On a fixed finite interval with positive-definite covariances, the Gaussian Wasserstein distance is locally comparable to covariance perturbation norms, yielding the corresponding distributional order estimate.

B.4 Time-varying growth rates

Proof of Proposition 2. For $p_t = \mathcal{N}(0, s(t)\Sigma)$, write $c = \frac{1}{2}\frac{d}{dt}\log s$, $M = cI + Q$, and $Q^\top = -Q$. All coefficients in the following expansions are evaluated at the start of the step. We normalize the current covariance to I.

Euler gives

$$(I + hM)(I + hM^\top) = I + 2chI + h^2(c^2 I + QQ^\top),$$

whereas the exact covariance factor is

$$\frac{s(t+h)}{s(t)} I = \left[1 + 2ch + (\dot{c} + 2c^2)h^2\right] I + O(h^3).$$

The difference is $h^2 D_t^{\mathrm{E}} + O(h^3)$, where

$$D_t^{\mathrm{E}} = QQ^\top - a_{\mathrm{E}} I, \qquad a_{\mathrm{E}} = c^2 + \dot{c}. \tag{19}$$

For Heun, let $H_{t,h}$ be its linear one-step map. Direct expansion gives

$$H_{t,h} = I + hM + \tfrac{h^2}{2}(\dot{M} + M^2) + h^3\left(\tfrac{1}{4}\ddot{M} + \tfrac{1}{2}\dot{M}M\right) + O(h^4),$$
$$\Phi_{t,h} = I + hM + \tfrac{h^2}{2}(\dot{M} + M^2) + \tfrac{h^3}{6}(\ddot{M} + 2\dot{M}M + M\dot{M} + M^3) + O(h^4).$$

Hence

$$H_{t,h} - \Phi_{t,h} = h^3\left(\tfrac{1}{12}\ddot{M} + \tfrac{1}{6}[\dot{M}, M] - \tfrac{1}{6}M^3\right) + O(h^4).$$

Here $[\dot{M}, M] = [\dot{Q}, Q]$ is skew-symmetric, so it does not contribute to twice the symmetric part of the leading defect. The covariance coefficient is therefore

$$D_t^{\mathrm{H}} = cQQ^\top + \left(\frac{\ddot{c}}{6} - \frac{c^3}{3}\right) I = c(QQ^\top - a_{\mathrm{H}} I), \quad a_{\mathrm{H}} = \frac{c^2}{3} - \frac{\ddot{c}}{6c}, \quad c \neq 0. \tag{20}$$

The same expression follows in two dimensions by using $\lambda(t) = c(t) + i\omega(t)$; the matrix calculation also covers a time-dependent skew basis.

For either criterion, $a > 0$ permits cancellation on each paired block by setting its frequency to $\sqrt{a}$. Full cancellation requires even dimension; an odd dimension retains one direction. For $a = 0$, the zero gauge minimizes the leading coefficient. For $a < 0$, $QQ^\top - aI$ has larger eigenvalues than $-aI$, with at least one strict increase if $Q \neq 0$. Thus every Schatten norm of the leading residual increases for a nonzero gauge (with the additional factor $|c|$ for Heun). At $c = 0$, the Heun coefficient is $\ddot{c}\, I/6$, independently of Q. These are pointwise-in-time statements. □

Exponential variance has constant c and recovers the frequencies $|c|$ for Euler and $|c|/\sqrt{3}$ for Heun. For a positive linear variance $s(t) = s_0 + \beta t$, differentiation gives $\dot{c} = -2c^2$ and $\ddot{c} = 8c^3$, so both cancellation parameters equal $-c^2$ whenever the Heun expression is defined.

C Global distributional convergence

This appendix proves Theorem 3, Corollary 4, and Theorem 5. All statements are for a fixed finite $T > 0$ and $h = T/n \to 0$. The coordinate changes S and S_j are invertible, and the distributions have finite second moments. An assertion of an exact nonzero convergence order additionally requires $c \neq 0$ and $m_2 > 0$, or at least one block with $c_j \neq 0$ and $m_{2,j} > 0$.

C.1 One rotational symmetry

Proof of Theorem 3. Since $J^2 = -I$, powers of $cI + \omega J$ remain in the span of I, J. Thus

$$P(h(cI + \omega J)) = a_h I + b_h J, \qquad a_h + ib_h = P(h(c + i\omega)).$$

If $\rho_h = \sqrt{a_h^2 + b_h^2} > 0$, choose ϕ_h so that this matrix is $\rho_h e^{\phi_h J}$. Starting at $X_0 = SZ$ with $Z \sim \nu$, the numerical endpoint is $S\rho_h^n e^{n\phi_h J} Z$. Invariance of ν under $e^{\alpha J}$ removes the rotation in distribution, giving

$$\widehat{\mu}_T = (\rho_h^n I)_{\#}\mu_0.$$

The case $\rho_h = 0$ gives a point mass at zero and satisfies the same identity. The rotation is Euclidean in the Z coordinates; S need not itself preserve Euclidean lengths.

For any distribution μ with finite second moment and $a, b \geq 0$, coupling (aX, bX) with $X \sim \mu$ gives cost $(a-b)^2 \int \|x\|^2 \, d\mu(x)$. Conversely, any coupling (U, V) of $(aI)_{\#}\mu$ and $(bI)_{\#}\mu$ satisfies the reverse triangle bound

$$(\mathbb{E}\,\|U - V\|^2)^{1/2} \geq \left|(\mathbb{E}\,\|U\|^2)^{1/2} - (\mathbb{E}\,\|V\|^2)^{1/2}\right| = |a - b| \left(\int \|x\|^2 \, d\mu(x)\right)^{1/2}.$$

The bounds coincide. Taking $a = \rho_h^n$ and $b = e^{cT}$ proves Eq. equation 10. This is the standard uniform-rescaling identity in optimal transport (Ambrosio et al., 2008; Villani, 2009); it does not require a density.

Let $z = c + i\omega$. Equation equation 7 and a logarithm near $P(0) = 1$ imply

$$n \log P(hz) = Tz + T\gamma z^{r+1} h^r + O(h^{r+1}). \tag{21}$$

Taking the real part yields Eq. equation 12. At zero gauge, the coefficient $T\gamma c^{r+1}$ is nonzero; hence the Wasserstein error has exact order r. With $\omega^\star = c\tan(\pi/[2(r+1)])$, z^{r+1} is nonzero and purely imaginary. Its contribution to the logarithm of the scale vanishes, and Eq. equation 10 gives an error of order at least $r + 1$.

To verify the trajectory assertion, write $z^{r+1} = ib$ with $b \in \mathbb{R} \setminus \{0\}$ at this frequency. The numerical and exact trajectories from the same initial sample satisfy

$$\widehat{X}_T - X_T = T\gamma b\, e^{cT} h^r S e^{\omega^\star TJ} JZ + O_{L^2}(h^{r+1}).$$

Since S, J, and $e^{\omega^\star TJ}$ are invertible and $\mathbb{E}\,\|SZ\|^2 > 0$, the leading L^2 coefficient is nonzero. The common-initial-sample trajectory error therefore remains $\Theta(h^r)$.

For Euler, Heun, classical RK3, and classical RK4, the linear stability polynomial is $P_r(z) = \sum_{k=0}^{r} z^k/k!$, for $r = 1, 2, 3, 4$. Its next logarithmic term is

$$\log P_r(z) = z - \frac{z^{r+1}}{(r+1)!} + \frac{(r+1)z^{r+2}}{(r+2)!} + O(z^{r+3}).$$

At the cancelling frequency, $\mathrm{Re}\, z^{r+2} = \mathrm{Re}(z\, ib) = -b\omega^\star \neq 0$. The next accumulated scale term is thus nonzero and of order h^{r+1}, proving exact distributional order $r+1$ for these four methods. □

C.2 Symmetric Runge–Kutta methods

Proof of Corollary 4. Let R be real analytic near zero, $R(0) = 1$, with exact even linear order r: $R(z) = e^z + \alpha z^{r+1} + O(z^{r+2})$ and $\alpha \neq 0$. The same two-dimensional algebra gives $R(h(cI + \omega J)) = a_h I + b_h J$. The preceding endpoint identity applies with $\rho_h = |R(h(c + i\omega))|$ for sufficiently small h; a rational stability function is evaluated away from its poles.

Choose the analytic logarithm with $\log R(0) = 0$. From $R(z)R(-z) = 1$, continuity at zero gives $\log R(-z) = -\log R(z)$. Therefore

$$\log R(z) = z + \alpha z^{r+1} + \beta z^{r+3} + O(z^{r+5}), \qquad \alpha \in \mathbb{R} \setminus \{0\}.$$

This odd-power property is the usual consequence of time symmetry (Chan & Gorgey, 2013; Hairer et al., 2006). At Eq. equation 11, the z^{r+1} term is purely imaginary. There is no term of degree $r+2$. After T/h steps the first possible scale error is of order h^{r+2}, which proves the Wasserstein bound. The nonzero imaginary term still gives trajectory error $\Theta(h^r)$, by the same argument as in Appendix C.1.

For the two examples in the main text,

$$R_{\mathrm{MP}}(z) = \frac{1 + z/2}{1 - z/2}, \qquad \log R_{\mathrm{MP}}(z) = z + \frac{z^3}{12} + \frac{z^5}{80} + O(z^7),$$
$$R_{\mathrm{G2}}(z) = \frac{1 + z/2 + z^2/12}{1 - z/2 + z^2/12}, \qquad \log R_{\mathrm{G2}}(z) = z - \frac{z^5}{720} - \frac{z^7}{12096} + O(z^9).$$

The displayed next real coefficients are nonzero at the respective first cancelling frequencies. The exact distributional orders are four for implicit midpoint and six for two-stage Gauss–Legendre. □

C.3 Several blocks with different scaling rates

Proof of Theorem 5. Apply the one-block algebra to each E_j. Independent rotation invariance of the joint distribution removes all numerical block angles at once; it does not assert statistical independence of the blocks. With block-diagonal S, the numerical endpoint is

$$\widehat{\mu}_T = \mathrm{diag}(\rho_{1,h}^n I, \ldots, \rho_{m,h}^n I)_{\#}\mu_0. \tag{22}$$

For nonnegative scalings a_j, b_j, any coupling (U, V) of the two block-rescaled distributions obeys, separately on each orthogonal block,

$$\mathbb{E}\,\|U_j - V_j\|^2 \geq \left[(\mathbb{E}\,\|U_j\|^2)^{1/2} - (\mathbb{E}\,\|V_j\|^2)^{1/2}\right]^2 = m_{2,j}(a_j - b_j)^2.$$

Summing gives a lower bound on the total quadratic cost. The coupling $U_j = a_j X_j$, $V_j = b_j X_j$ from the same $X \sim \mu_0$ attains all block bounds simultaneously, including when the block radii are dependent. Taking $a_j = \rho_{j,h}^n$ and $b_j = e^{c_j T}$ proves Eq. equation 15.

On blocks with $c_j \neq 0$, use $\omega_j = |c_j| \tan(\pi/[2(r+1)])$. Its sign does not affect the scale because P has real coefficients. On blocks with $c_j = 0$, use zero gauge. The scalar expansion on each block cancels its leading scale term, so the sum in Eq. equation 15 is $O(h^{2(r+1)})$. For the four explicit methods, a nonstationary block with $m_{2,j} > 0$ has a nonzero next coefficient. The squared block errors cannot cancel each other, giving exact order $r+1$. If every positive-moment block is stationary, the zero-gauge endpoint is already exact instead. □

When two positive-moment blocks have different rates, their second moments are multiplied by different factors $e^{2c_j t}$. No single common scalar rescaling represents the path on a nontrivial time interval. This is the unequal-rate extension described in the main text.

D Approximate scores and reference comparisons

This appendix explains the exact-score construction and the distinction between numerical integration error and a change in the continuous-time distribution. It motivates the two reference comparisons used in Sections 5.2 and 5.3.

D.1 The skew-score construction

For a smooth skew-symmetric matrix field $A_\theta(x,t)$, use the row-divergence convention $(\nabla\cdot A)_i = \sum_j \partial_j A_{ij}$. The correction is

$$u_\theta = A_\theta \widehat{s}_t + \nabla\cdot A_\theta, \qquad A_\theta^\top = -A_\theta. \tag{23}$$

With $\widehat{s}_t = \nabla \log p_t$, the product rule gives

$$p_t u_\theta = \nabla\cdot(p_t A_\theta), \qquad \nabla\cdot(p_t u_\theta) = \sum_{i,j} \partial_i \partial_j (p_t A_{\theta,ij}) = 0.$$

The last identity follows by pairing indices i, j, using antisymmetry and commuting mixed derivatives. This proves marginal preservation when the score is exact.

If $\widehat{s}_t = \nabla \log p_t + \eta_t$, the remaining term is

$$\ell_t := \nabla\cdot(p_t u_\theta) = \nabla\cdot(p_t A_\theta \eta_t). \tag{24}$$

Thus score approximation, rather than the skew identity itself, introduces the additional source term.

D.2 A finite-time bound

Let p_t solve the continuity equation for v_t, and let q_t solve it for $w_t = v_t + u_t$, with $q_0 = p_0$. Setting $\delta_t = q_t - p_t$ gives

$$\partial_t \delta_t + \nabla\cdot(w_t \delta_t) = -\ell_t, \qquad \delta_0 = 0. \tag{25}$$

For a zero-mass signed distribution ℓ, write $\|\ell\|_{\mathrm{Lip}^*} = \sup_{\mathrm{Lip}(\varphi)\le 1} |\langle \ell, \varphi\rangle|$.

Theorem 7 (Endpoint distribution bound)**.** *Suppose w_t has a global flow and is spatially Lipschitz with integrable constant L_t. Assume finite first moments, the stated boundary conditions, and integrability of the right-hand side below. Then*

$$W_1(q_T, p_T) \le \int_0^T \exp\left(\int_s^T L_\tau \, d\tau\right) \|\ell_s\|_{\mathrm{Lip}^*} \, ds. \tag{26}$$

Proof. Let φ be a terminal 1-Lipschitz test and pull it back along the flow of w. The resulting φ_s satisfies the backward transport equation and $\mathrm{Lip}(\varphi_s) \le \exp(\int_s^T L_\tau \, d\tau)$ by Grönwall's inequality. Pairing Eq. equation 25 with φ_s and integrating in time gives

$$\langle \delta_T, \varphi\rangle = -\int_0^T \langle \ell_s, \varphi_s\rangle \, ds.$$

Bound each integrand by the Lipschitz constant times $\|\ell_s\|_{\mathrm{Lip}^*}$, and take the supremum over terminal tests. Kantorovich–Rubinstein duality gives Eq. equation 26. □

D.3 CONNECTION TO THE EXPERIMENTAL METRICS

For a numerical endpoint $\widehat{p}_T$ of the corrected field, the triangle inequality gives

$$W_1(\widehat{p}_T, p_T) \le \underbrace{W_1(\widehat{p}_T, q_T)}_{\text{integration error for the corrected field}} + \underbrace{W_1(q_T, p_T)}_{\text{change in the continuous distribution}} . \tag{27}$$

This is an upper bound, not an additive identity for the errors. It explains why the numerical and reference endpoints of the *same* corrected field are compared separately from the corrected and uncorrected reference endpoints.

In the pretrained experiments, both reference endpoints are computed numerically, and their distances use projected feature moments. The same-field gap is therefore a finite-reference diagnostic, not an exact W_1 or W_2 error. Its reduction and a reduction in target FID are reported separately. Likewise, a reference-distribution difference can contain residual integration error in addition to continuous-time drift. The precise feature map and reference settings are recorded in Appendix E.3.

The identity in Eq. equation 24 assumes that p_t solves the base continuity equation. For a learned base field this comparison is with that field's continuous solution; agreement with the data distribution is a separate question. Finally, finite feature means and covariances do not identify a full probability distribution, so a small feature discrepancy alone is not a certificate of exact marginal preservation.

E EXPERIMENTAL DETAILS AND ADDITIONAL RESULTS

The organization follows the main experiments: analytic paths (Section 5.1), two-dimensional mixtures (Section 5.2), and pretrained diffusion models (Section 5.3). We distinguish the current five-run Euler evaluation from earlier single-run comparisons. Additional tests are reported separately from the main comparisons.

E.1 ANALYTIC PATHS

Models and measurements. The Gaussian experiments use zero-mean paths $p_t = \mathcal{N}(0, Se^{2tB}S)$ with $S \succ 0$ and $B = B^\top$. The initial covariance is S^2. The scalar-growth tests use dimensions one through six and non-diagonal positive-definite covariance matrices. The spectral tests vary the multiplicities of the nonzero eigenvalues of B. The normalized residual is

$$\frac{\left\|\mathrm{Sym}((B+Q)^{r+1})\right\|_{\mathrm{F}}}{\|B^{r+1}\|_{\mathrm{F}}},$$

with Q chosen by the block construction in Appendix B. The measured points in Figure 2(a) evaluate this leading matrix coefficient numerically. They are not Monte Carlo estimates of finite-step endpoint error. Coincident cases overlap in the plot.

Convergence experiments use double precision and fixed step-size grids. The order is fitted by least squares to log endpoint Wasserstein error against log step size. The exact distributional formulas in Appendix C provide the endpoint error for the circle and multi-block tests. The archived suite contains 90 parameter/solver configurations: 24 scalar-growth Gaussian, 16 spectral, 12 time-varying, 16 non-Gaussian one-circle, 16 multi-block, and six symmetric-RK cases. The archived pass flags refer to the original residual and slope tolerances.

Spectral configurations. The four patterns displayed in Figure 2(a) are listed in Table 3. Each is tested with Euler, Heun, RK3, and RK4. For the fully cancellable cases the matrix evaluation leaves residuals of at most 2.1×10^{-15}; the uncancellable case gives one. For the partially cancellable case, the predicted and measured residuals agree to the shown precision.

Non-Gaussian distributions. The one-circle experiments use a linearly transformed Student's t distribution with five degrees of freedom, a uniform unit ball, a mixture on two shells, and a distribution on a simultaneous circle orbit. The four multi-block configurations are: a Student's t block and a shell block; three blocks using a disk, a Student's t distribution with seven degrees of

Table 3: Spectral configurations in Figure 2(a). The four numbers in the last column are ordered as Euler, Heun, RK3, RK4. Residuals are normalized by the corresponding zero-gauge coefficient.

Configuration	Dimension	Predicted normalized residuals
Repeated pairs	4	$0,\ 0,\ 0,\ 0$
Zero + pairs	5	$0,\ 0,\ 0,\ 0$
Simple spectrum	4	$1,\ 1,\ 1,\ 1$
One unpaired	3	$0.967881,\ 0.993861,\ 0.998864,\ 0.999791$

freedom, and a shell mixture; four circle-orbit blocks; and a two-block torus with perfectly correlated radii. Their block growth rates are unequal. A separate discrete optimal-transport check enumerates assignments for anticorrelated radii and compares the minimum cost with Eq. equation 15.

Table 4: Distributional convergence orders. Explicit-RK values average four multi-block configurations. Symmetric-RK values average $c = 0.3, 0.7, 1.1$ in the one-circle setting. These are fitted orders, not uncertainty intervals.

Setting	Solver	Base	GaugeFix
Multiple blocks	Euler	0.985	1.975
	Heun	1.988	2.987
	RK3	2.987	3.986
	RK4	3.986	4.986
Symmetric RK	Implicit midpoint	2.000	4.000
	Two-stage Gauss–Legendre	4.000	6.020

For the two-stage Gauss method, the three GaugeFix estimates are 6.061084, 5.999562, and 6.000020; their mean gives 6.020 in Table 4. The expected asymptotic order is six. The corresponding trajectory orders for midpoint and Gauss are two and four. These fitted results are distinct from the four explicit-method curves in the introductory Gaussian example, whose Base/GaugeFix order pairs are $0.990/1.983$, $1.991/2.991$, $2.991/3.990$, and $3.990/4.990$.

Time-dependent scales. Table 5 records the twelve local checks of Proposition 2. The sign quantity is a_{E} or a_{H}, not an endpoint error. When the Base coefficient is zero, a normalized residual would be $0/0$ and is left undefined. When $c = 0$ for Heun, the sign formula is undefined, but its covariance coefficient is still given by $\ddot{c}\, I/6$ in Appendix B.4.

Table 5: Time-varying Gaussian checks. The residual is the optimized leading covariance coefficient relative to Base. "Undefined" is not counted as complete cancellation of a nonzero Base error.

Solver	Case	d	Sign quantity	Predicted	Measured
Euler	Exponential, even	2	0.490000	0	0
	Exponential, odd	3	0.490000	0.577350	0.577350
	Accelerating	4	0.460000	0	0
	Linear variance, even	2	-0.160000	1	1
	Linear variance, odd	3	-0.160000	1	1
	Zero Base coefficient	3	0	Undefined	
Heun	Exponential, even	2	0.163333	0	0
	Exponential, odd	3	0.163333	0.577350	0.577350
	Curved scale	4	0.115714	0	1.7×10^{-16}
	Linear variance	2	-0.160000	1	1
	$c = 0$, nonzero defect	3	Undefined	1	1
	Zero Base coefficient	3	$\simeq 0$	Undefined	

E.2 TWO-DIMENSIONAL MIXTURE PATHS

Paths, scores, and references. The experiments use Gaussian-mixture paths with fixed centers and exponentially changing component variance. GMM8 arranges eight component centers on a ring; the checkerboard arrangement is the second main-text path. An additional smoothed two-moons arrangement is retained below. The global-correction experiments use 640 training and 1,280 evaluation samples. Analytic scores supervise an MLP, which is frozen before the correction is evaluated. The relative validation score RMSEs are 7.50% for GMM8, 7.65% for checkerboard, and 9.70% for moons.

Exact-score experiments use eight-step Euler and a 256-step RK4 numerical reference evaluated with the analytic score. "Analytic" here refers to the score, not an exact RK4 integration. The learned-score protocol fits the correction against a 128-step reference of the same frozen field and uses a separate 32-step preservation loss. Thus the two score settings use their respective fitting protocols, rather than being a score-only ablation of an otherwise identical training run.

Reported errors. Target error compares the numerical endpoint with the target reference. The same-field error compares the numerical endpoint with a finer integration of that same corrected field. The displayed toy comparisons use sliced-W_2, obtained by averaging one-dimensional projected Wasserstein errors; it is distinct from the exact W_2 formulas of the analytic-path experiments. Other archived diagnostics include empirical optimal transport, MMD, covariance error, and density-defect RMS.

Table 6 gives the global-correction measurements behind Figure 3, together with the additional moons result. Exact-score percentages come from one evaluation. Learned-score values average five runs, with seeds 91000–91004 for GMM8 and 91100–91104 for checkerboard. The standard deviations follow the archived $\mathrm{ddof} = 0$ convention; they are not confidence intervals. The reference ratio is the corrected high-step target error divided by the Base high-step target error, not the same-field solver error.

Table 6: Global corrections on two-dimensional mixtures. Positive reductions are improvements over Base. The last three columns are five-run means $\pm$ standard deviations. Moons is an additional test, not part of the two-dataset main-text figure.

	Exact-score	Learned score		
Path	Target red. (%)	Target red. (%)	Solver red. (%)	Reference ratio
GMM8	65.5	14.2 ± 1.6	25.6 ± 2.4	1.023 ± 0.017
Checkerboard	64.1	31.9 ± 1.3	33.6 ± 0.6	1.178 ± 0.025
Moons	13.7	0.2 ± 0.4	1.0 ± 0.1	1.007 ± 0.002

Additional spatial correction. For the moons arrangement, the archived spatial experiment uses $a(x,t)$ with 18 polynomial coefficients and the two-dimensional identity

$$u = a(x,t)J\nabla \log p + J\nabla a, \qquad pu = J\nabla(ap). \tag{28}$$

This is Eq. equation 23 with $A = aJ$ and is exactly marginal-preserving for the analytic score. The experiment also evaluates the same form with the frozen learned score. The recorded training and initial evaluation sizes are 768 and 1,920; the separate five-run spatial comparison uses 1,280 samples per run. Its mean target/solver-error reductions are $2.7\%/15.3\%$ for the spatial correction and $22.9\%/70.2\%$ for the parameter-matched spatial Free control. The global correction in those same spatial-comparison runs gives approximately $-0.1\%/1.0\%$; this is a different evaluation group from Table 6.

Original selection criteria. The original exact-score protocol required more than 10% target sliced-W_2 reduction on at least two paths, improvement over five norm-matched random gauges, and reduced visible density-defect RMS. The learned-score protocol required more than 10% same-field error reduction, no target degradation, improvement over the random controls, and a high-step target-error ratio below 1.25 on at least two paths. The exact-score moons run did not meet the density-defect criterion. These original criteria are recorded for reproducibility; they are not redefined by the selection of two paths for the main-text figure.

E.3 PRETRAINED MODELS, IMPLEMENTATION, AND METRICS

Checkpoints and samplers. The main experiments use the public NVLabs EDM unconditional VP CIFAR-10 checkpoint at 32×32 resolution (Karras et al., 2022). Additional evaluations use the unconditional VP AFHQv2 checkpoint at 64×64 and conditional CIFAR-10. For the paired corrections, each solver uses the same checkpoint, coarse schedule, input noise samples, and denoiser-evaluation count across Base, GaugeFix, and Free. Table 7 gives the numerical settings omitted from the main-text table. Schedule selection examines $\rho \in \{3, 5, 7, 9, 12, 16, 24, 40\}$ where applicable and is completed before the frozen evaluation.

Table 7: Sampling and reference settings for the main CIFAR-10 comparisons. NFE counts denoiser evaluations. Budgets and schedules are matched within each solver, not across different solver rows.

Role	Solver	NFE	Schedule ρ
Coarse integration	Euler (8 steps)	8	12
Coarse integration	Heun (9 steps)	17	9
Coarse integration	DPM-Solver++(2M)	17	9
Reference integration	Heun (18 steps)	35	7

An earlier five-step Heun test (nine denoiser evaluations) produced approximately 25.1% saturated pixels before image conversion, compared with about 0.002% in the tested Euler setting. The reported Heun comparison uses nine steps. This observation concerns those particular settings and is not a minimal-stability threshold over all step counts.

The 18 coefficients. Let K_1, K_2, K_3 be the RGB channel-rotation operators and K_4, K_5, K_6 the periodic spatial finite-difference operators. The operator bank is skew-adjoint. Each operator has three coefficients, corresponding to three basis functions of log noise. In bank notation the combination is

$$A_\theta(t) = \alpha \sum_{\ell=1}^{3} \sum_{k=1}^{6} \theta_{\ell k} b_\ell(t) K_k, \qquad \alpha = 0.15$$

for the archived basic-bank configuration. Thus channel and spatial groups each contain nine coefficients. The completion evaluation uses the original constant state features and frozen coefficients. It adds no denoiser calls during sampling. Free uses 18 coefficients without the skew-score marginal constraint. Parameter count and evaluation settings are matched; this does not assert identical historical training sample counts or optimization cost.

Official target FID. FID50k uses 50,000 generated images per run, the official 2,048-dimensional Inception features and CIFAR-10 reference statistics. The completion runs reuse the original evaluator, image conversion, float64 feature-moment accumulation, and SciPy matrix square root. Batch size is 256; the final batch has the exact remaining size. Within each run, all compared methods receive the same latent batches.

Same-field gap and reference-distribution change. For each method θ, write $\widehat{\mu}_{c,\theta}$ and $\widehat{\mu}_{f,\theta}$ for the coarse and reference endpoint distributions. Let D be Gaussian FID on the fixed projected Inception features. The diagnostics are

$$G_\theta = D(\widehat{\mu}_{c,\theta}, \widehat{\mu}_{f,\theta}), \qquad L_\theta = D(\widehat{\mu}_{f,\theta}, \widehat{\mu}_{f,0}). \tag{29}$$

G_θ is the main-text same-field gap. L_θ compares corrected and Base reference distributions and is recorded separately. In particular, the GaugeFix coarse endpoint is not compared to the Base reference when computing G_θ.

Each gap run uses 4,096 images per endpoint. Official features are projected by the fixed 256×2048 row-normalized Gaussian matrix, generated in CUDA float64 with seed 20260821. The 35-evaluation Heun reference in Table 7 is shared with the archived Heun and DPM++ audits. It is a finite numerical reference, not an exact flow. These projected-feature measurements are not official target FID50k and do not certify a full-distribution Wasserstein error; see Appendix D.3.

Frozen evaluations and statistics. The current Euler comparison uses five evaluation seeds: 110000, 120000, 130000, 140000, and 150000. Pilot seed 100000 is excluded from their averages. GaugeFix and Free coefficients are frozen before these evaluations; the completion run performs no fitting, schedule search, coefficient selection, or seed filtering. Before reusing archived Base/GaugeFix FID values, a complete Base FID50k run at seed 110000 was replayed. Its absolute difference from the archived FID was 1.45×10^{-12}, below the predeclared tolerance 10^{-3}.

For a lower-is-better quantity M, the percentage reduction for run i is $\delta_i = 100(1 - M_{\theta,i}/M_{0,i})$. Reported reductions average these paired percentages, rather than taking a ratio of rounded means. For five pretrained runs, the sample standard deviation uses $\mathrm{ddof} = 1$; the 95% interval is $\bar{\delta} \pm t_{0.975,4} s_\delta / \sqrt{5}$. It describes evaluation variation for a fixed fitted correction, not variation from retraining. Heun and DPM++ use the archived single FID run at seed 190000 and single gap run at seed 160000. No five-run interval is assigned to those entries.

E.4 Pretrained results and component ablation

Paired comparisons. Table 8 reports the source values behind the main comparison, with FID rounded here to four decimals. The Euler Free values are the new five-run evaluation, not the earlier pilot result. The main Euler gap reduction is now based on an independent same-field evaluation. DPM++ is retained as an additional solver setting, and the standalone DPM-Solver-v3 and Heun-based ERK-Guid results are distinguished from GaugeFix modifications.

Table 8: CIFAR-10 target FID and same-field gap. Euler entries average five runs; other paired rows use one FID run and one gap run. FID50k uses 50,000 images, whereas gap uses 4,096 images and projected features. "Not evaluated" denotes missing measurements, not zero improvement.

Solver	Correction	FID50k ↓	Gap ↓	Gap reduction
Euler	Base	27.9948	3.253591	Reference
	GaugeFix	23.8955	2.764185	15.03%
	Free	24.6453	2.910265	10.55%
Heun	Base	3.0496	0.253244	Reference
	GaugeFix	2.9957	0.247726	2.18%
	Free	2.8785	0.235176	7.13%
	ERK-Guid	2.7066	Not evaluated	
DPM++(2M)	Base	2.6858	0.164193	Reference
	GaugeFix	2.6885	0.160684	2.14%
DPM-Solver-v3	Standalone	3.5808	Not evaluated	

The five-run Euler FID reductions are 14.64% for GaugeFix (95% interval $[14.12\%, 15.16\%]$) and 11.96% for Free ($[11.68\%, 12.25\%]$). Their same-field gap reductions are 15.03% ($[12.11\%, 17.96\%]$) and 10.55% ($[9.74\%, 11.36\%]$), respectively. These intervals are for each correction relative to Base, not for the paired difference between GaugeFix and Free. For DPM++, the reported 2.14% gap reduction accompanies a 0.10% increase in target FID. No frozen DPM++-specific Free adapter was available for the completion study, so an Euler or Heun adapter was not relabeled as that control.

Component ablation. In Table 2, *Channel only* retains the 3×3 channel-coefficient block and sets the spatial block to zero. *Spatial only* does the reverse. The remaining coefficients are kept at their fitted values, without retraining or rescaling. Base zeros both blocks, and Full GaugeFix retains all 18 coefficients. The reported Euler FID values are 27.995, 24.639, 27.007, and 23.896 in that order; the Heun values are 3.050, 3.000, 3.022, and 2.996. This tests the contribution of components of the frozen full adapter, rather than the best separately fitted nine-parameter adapter.

Sampling cost. The recorded single-50,000-image method/Base time ratios are 1.02 for Euler + GaugeFix and 1.03 for Heun + GaugeFix, which supply the approximately 2–3% statement in Section 5.3. The other recorded ratios are 1.00 for Euler + Free, 1.16 for Heun + Free, 1.52 for ERK-Guid, and 0.99 for DPM++ + GaugeFix. These are sampling-run measurements, not repeated timing estimates or adapter-fitting costs; the 0.99 observation is not used to claim a speedup.

The later completion evaluation used an NVIDIA L20X with PyTorch 2.7.0a0+ecf3bae40a.nv25.02 and CUDA 12.8. Its peak allocated memory was 4.730 GiB. Completed gap runs took 34.86 minutes in total, and completed FID generation/feature accumulation took 33.52 minutes, excluding loading and final matrix square roots. The GPU was shared. These later totals are recorded separately and are not the source of the earlier overhead ratios. No fitting was performed in the completion evaluation.

Other quality metrics. Table 9 retains the additional feature metrics from the earlier evaluation. KID is the degree-three polynomial-MMD estimate (Bińkowski et al., 2018). Precision and recall use the archived approximate third-nearest-neighbor metric in a fixed 128-dimensional feature space, following Kynkäänniemi et al. (2019). These are separate measurements, not replacements for the five-run official FID.

Table 9: Additional CIFAR-10 feature measurements from the archived runs. Precision and recall use the approximate feature-space implementation described in the text. No repeated-evaluation intervals are supplied for these values.

Solver	Correction	KID ↓	Precision ↑	Recall ↑
Euler	Base	0.02876	0.5879	0.4728
	GaugeFix	0.02332	0.6112	0.5066
	Free	0.02480	0.5983	0.4881
Heun	Base	0.0010756	0.6795	0.6348
	GaugeFix	0.0010404	0.6741	0.6329
	Free	0.0009092	0.6734	0.6412

E.5 Additional solvers and transfer settings

External references. The DPM-Solver-v3 experiment (Zheng et al., 2023) uses official code commit `b615d70`, released EDM–CIFAR-10 empirical statistics over $[0.002, 80]$, third order, and uniform log-SNR spacing. Its standalone FID50k of 3.5808 uses eight denoiser evaluations and seed 100000. The separate frozen-adapter transfer test uses 17 evaluations and seeds 270000–310000 in increments of 10000; pilot 260000 is excluded.

ERK-Guid (Kong et al., 2026) uses the implementation ported from commit `ef4f1a7`, following the cited Eq. 19 and Algorithm 1. Checks cover the previous-step embedded-Euler buffer, stiffness and direction estimates, threshold, quadratic correction, initial-step skip, final Euler update, and denoiser-call count. A 28-point search uses 4,096 samples and 17 evaluations. Parameters $(2.0, 0.3)$ are then frozen; the subsequent three projected-FID reductions are 9.56%, 6.60%, and 8.23%. Its official FID50k is reported in Table 8.

Five-run projected-feature evaluations. Table 10 records the previously plotted robustness checks without mixing them with official FID50k. Each row uses five runs. AFHQv2 uses 2,048 samples per run in this table; the others use 4,096. The 4,096-sample AFHQv2 same-field check is a separate evaluation. The intervals are computed from paired reductions with the same Student-t convention as above.

Table 10: Additional projected-FID evaluations. Positive values indicate a reduction relative to the matched uncorrected sampler. These metrics use projected features and smaller sample sets than official FID50k.

Configuration	Samples/run	Mean reduction (%)	95% interval (%)
CIFAR-10, Euler	4096	13.545	$[10.740, 16.350]$
CIFAR-10, Heun	4096	0.442	$[-0.588, 1.472]$
CIFAR-10, DPM++	4096	-0.114	$[-0.654, 0.425]$
CIFAR-10, DPM-v3 transfer	4096	0.213	$[-1.246, 1.672]$
AFHQv2, DPM++	2048	0.047	$[-0.738, 0.832]$

Composition and conditional checkpoints. Frozen ERK-Guid + GaugeFix changes projected FID from 0.3690 to 0.3801; the corresponding Free composition gives 0.4097. A separately declared ERK-conditioned search uses the same 18 gauge parameters, 1,024 samples per two-sided finite-difference probe, and 4,096 independent samples for radius selection. Its selected candidate reduces projected FID by 0.69%, below the original 2% threshold for advancing to FID50k; its recorded high-reference ratio is 0.993.

For conditional CIFAR-10, labels are balanced and deterministic. Frozen transfer from the unconditional Heun adapter gives a 0.762% projected-FID reduction and a 1.610% increase in the same-field gap. A checkpoint-specific search examines 18 coordinates and 20 direction/radius candidates. Its best quality candidate reduces the proxy by 3.917% with a 13.248% increase in the gap. The best gap candidate reduces that gap by 0.073% with slightly increased target error. These candidates did not meet the original joint advancement criteria, so no conditional FID50k was run.

The AFHQv2 same-field evaluation gives a 0.104% gap increase. Its projected-quality and same-field criteria were not met, so official FID50k was not run. These outcomes retain the original stopping rules; missing FID50k entries are not interpreted as measured zero gains.

Candidate selection. The archived Heun and DPM++ searches use finite differences of the actual same-field feature gap together with projected target FID in a small trust region. Separate evaluation seeds check the gap, and an independent-sample reference null is used for the reference-shift criterion. The earlier FID-only Heun candidate increases the gap by 15.91%. It, the curvature-selected DPM++ candidate, and the AFHQv2 records remain in the artifact collection. The main comparisons use the frozen selected adapters.