NUMERICAL GAUGE FIXING FOR GENERATIVE ODES

Anonymous authors
Paper under double-blind review

ABSTRACT

Sampling from generative ordinary differential equations (ODEs) involves numerically integrating learned velocity fields, often requiring many model evaluations. Existing approaches accelerate sampling by improving solvers and time grids for a fixed velocity field, or by distilling pretrained generators into faster samplers. A complementary degree of freedom arises from gauge equivalence: a prescribed marginal probability path can admit multiple velocity fields with different finite-step numerical errors under the same sampler. In this work, we introduce numerical gauge fixing, a post-training framework that optimizes the choice of velocity field for a frozen pretrained model's marginal path to improve finite-step integration accuracy. Theoretically, we characterize how the choice among gauge-equivalent velocity fields affects finite-step integration error relative to the exact flow of the corrected velocity field. Building on this formulation, we develop a practical compiler that optimizes parameterized corrections by minimizing the discrepancy between coarse and refined rollouts of the same corrected field. Experiments on Gaussian-mixture flows and pretrained image generators show that gauge optimization can reduce finite-step integration error and improve generation quality with the base model, solver, and time grid held fixed. Our work identifies gauge selection as a complementary numerical optimization axis for few-step generative sampling.

1 INTRODUCTION

Diffusion and flow models are widely used for image and video generation (Esser et al., 2024; Ho et al., 2022). Many such models generate samples by solving generative ordinary differential equations (ODEs) (Song et al., 2021b; Lipman et al., 2023). These ODEs describe continuous transport from a simple initial distribution to the model's output distribution through learned velocity fields. In practice, sampling approximates this transport through numerical integration. Because evaluating the velocity field typically requires a neural network call, repeated evaluations incur substantial computational costs, motivating the development of few-step samplers. Reducing integration error under a limited budget of model evaluations is therefore a fundamental problem in sampling from generative ODEs.

Existing acceleration methods primarily improve numerical discretization or learn faster samplers. One line of work treats the pretrained velocity field as fixed and improves sampling efficiency through specialized solvers (Karras et al., 2022; Lu et al., 2022a), optimized time grids (Sabour et al., 2024), or learned integration coefficients (Frankel et al., 2025). Another line uses distillation to learn few-step sampling maps or new velocity fields from pretrained generators, transferring their generative capabilities to models that require fewer sampling steps (Salimans & Ho, 2022; Song et al., 2023). The former optimizes how a given velocity field is integrated, while the latter learns a sampler that performs generation in fewer steps.

Beyond these approaches, the continuous representation of a generative ODE offers a complementary degree of freedom: fixing a marginal probability path does not uniquely determine its velocity field. Multiple *gauge-equivalent* velocity fields can realize the same path, allowing particles to follow different trajectories while inducing identical marginal distributions at every time. This nonuniqueness of continuous representations has been studied in prior work (Horvat & Pfister, 2024; Pao-Huang et al., 2026). We investigate its numerical consequences: with the solver and time grid held fixed, can gauge-equivalent velocity fields exhibit different finite-step integration errors? If so, can selecting a suitable representative improve the numerical accuracy of few-step sampling while preserving the continuous marginal path?

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

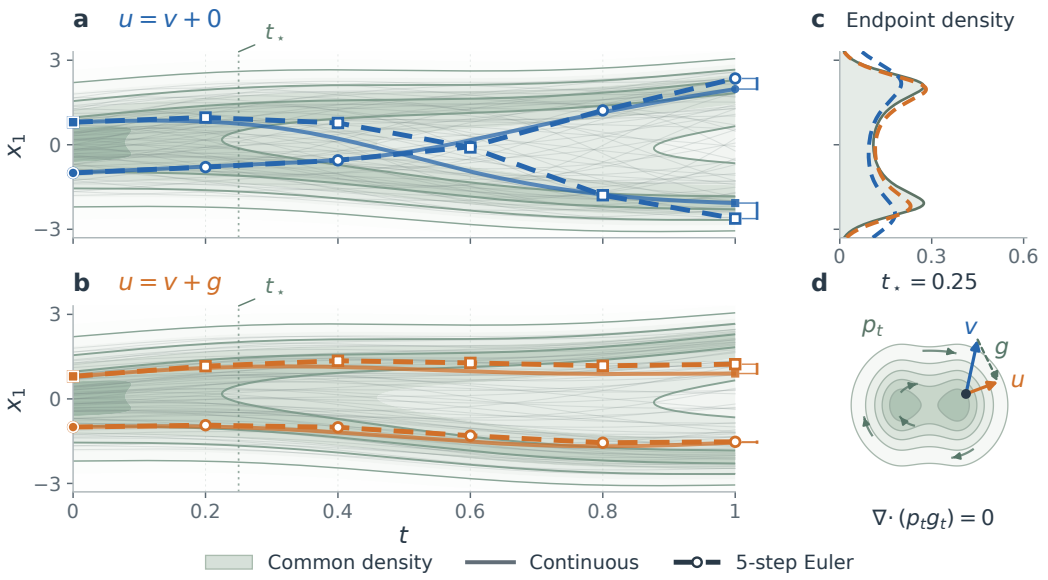

Figure 1: Same marginal path and sampler, different integration errors. Two gauge-equivalent fields use matched initial samples and the same five-step Euler grid. **(a,b)** First-coordinate densities and trajectories; endpoint brackets mark each field's own first-coordinate error. **(c)** First-coordinate endpoint densities. **(d)** Shared two-dimensional density and fixed-time gauge streamlines, with $u = v + g$ and $\mathrm{div}(p_t g_t) = 0$. Construction in Appendix G.

We characterize how integration error varies with the representative and construct gauge-equivalent fields with strictly different accuracy under a fixed Euler grid. We then relate a refined-reference objective to true integration error, identifying the reference-accuracy conditions needed for reliable selection. Building on these results, we introduce *numerical gauge fixing*, a post-training framework for frozen generators. The framework optimizes parameterized corrections by comparing coarse and refined rollouts of the same corrected field, allowing continuous particle trajectories to change while the solver and time grid remain fixed. Comparing each candidate with the original flow would also penalize changes in particle correspondence permitted by gauge equivalence. The same-field objective instead measures how accurately the sampler follows the candidate's own dynamics, providing a consistent criterion for optimization and evaluation. Exact path scores yield marginal-preserving constructions; model-derived carriers make the method practical, with any resulting marginal drift assessed separately. Experiments on Gaussian-mixture flows and pretrained image generators evaluate numerical accuracy, marginal changes, and generation quality under matched sampling configurations.

Our main contributions are as follows:

- **Problem and framework.** We formulate velocity-field representative selection within a fixed marginal probability path as a post-training numerical optimization problem for generative ODEs. We introduce numerical gauge fixing as an optimization axis complementary to solver and time-grid design.
- **Theory and method.** We characterize the effect of gauge selection on finite-step integration error, prove that strict improvement is possible in specific settings, and connect a refined-reference objective to the true integration error. These results motivate a practical method for optimizing parameterized corrections to frozen pretrained models.
- **Experimental validation.** We validate numerical improvement within a fixed marginal path in exact models and evaluate its practical effects on pretrained image generators. Marginal diagnostics and controlled comparisons examine the roles of gauge parameterization and optimization settings.

2 Background and Problem Formulation

2.1 Generative ODEs and Numerical Sampling

Probability-flow diffusion and flow-matching models generate samples by integrating a learned velocity field (Song et al., 2021b; Lipman et al., 2023). Let $v_t(x) = v_\theta(x, t)$ be a frozen model on $\mathbb{R}^d$, and let $\Phi_t[u]$ denote the exact flow of a field u from time 0 to t. Starting from a source law p_0, the model induces

$$\dot{X}_t = v_t(X_t), \qquad X_0 \sim p_0, \qquad p_t = (\Phi_t[v])_{\#}p_0. \tag{1}$$

Here $F_{\#}p$ denotes a pushforward. We take the model path p_t as the reference throughout the paper. It need not coincide with the training interpolation q_t or recover the data distribution at T. We assume positive smooth densities and well-posed flows with finite second moments; Appendix A states the regularity conditions.

Sampling uses a numerical configuration $\xi = (\mathsf{S}, \mathcal{T})$, comprising a solver and a grid $\mathcal{T} = \{0 = t_0 < \cdots < t_K = T\}$. Its endpoint map is $\Psi_\xi[u]$. All settings, including multistep initialization, remain fixed when fields are compared. Neural function evaluations (NFE) count backbone calls and need not equal the number of grid intervals. We use Euclidean state norms and write $\|F\|^2_{L^2(p_0)} = \mathbb{E}_{X_0\sim p_0}\|F(X_0)\|^2$.

2.2 Gauge Equivalence along a Fixed Marginal Path

A marginal path does not uniquely determine the velocity field that realizes it (Horvat & Pfister, 2024; Pao-Huang et al., 2026). The path p satisfies $\partial_t p_t + \nabla\cdot(p_t v_t) = 0$. Adding a correction g_t preserves this continuity equation exactly when $\nabla\cdot(p_t g_t) = 0$. Under the preceding regularity conditions and the same initial law, we therefore define

$$\mathcal{G}(p) = \{g : \nabla\cdot(p_t g_t) = 0 \text{ for every } t\}, \qquad \mathcal{V}(p) = v + \mathcal{G}(p), \tag{2}$$

restricting to admissible fields. Representatives in $\mathcal{V}(p)$ share all one-time marginals, while their particle trajectories and cross-time couplings may differ (Figure 1).

Skew probability-flux constructions (Ma et al., 2015) provide explicit directions in this class.

Proposition 1 (Skew-symmetric gauge construction)**.** *Let $A_t(x)^\top = -A_t(x)$ and suppose $p_t A_t$ has continuous second spatial derivatives. Under Assumption 1, the correction*

$$g_t^A = p_t^{-1}\nabla\cdot(p_t A_t) = \nabla\cdot A_t + A_t s_t, \qquad s_t = \nabla \log p_t, \tag{3}$$

belongs to $\mathcal{G}(p)$. Matrix divergence is row-wise. For $d \geq 2$, nonzero admissible corrections exist with smooth, spatially compactly supported probability flux.

Skew symmetry cancels mixed derivatives in $\operatorname{div}\operatorname{div}(p_t A_t)$, establishing preservation. State-independent A_t reduces the correction to $A_t s_t$; spatial dependence allows a richer search family. Appendix A.2 proves the construction and nonzero existence.

2.3 Numerical Gauge Fixing as an Optimization Problem

With the model path and sampler fixed, we select a representative from an admissible family $\mathcal{K} \subseteq \mathcal{V}(p)$ containing v. For square-integrable endpoint maps, define

$$d_\xi(u) = \|\Psi_\xi[u] - \Phi_T[u]\|_{L^2(p_0)}, \qquad \inf_{u\in\mathcal{K}} \mathcal{J}_\xi(u), \quad \mathcal{J}_\xi(u) = \tfrac{1}{2} d_\xi(u)^2. \tag{4}$$

We call this problem *numerical gauge fixing*. Both maps use the same source input and candidate field. The exact reference changes with u, allowing the particle correspondence to change while measuring each representative's own integration error. Including v retains zero correction; the infimum does not assume an attained optimum.

Since every exact representative has output law p_T, common-source coupling gives

$$W_2\big((\Psi_\xi[u])_{\#}p_0, p_T\big) \leq d_\xi(u). \tag{5}$$

Reducing d_ξ tightens this upper bound, without guaranteeing lower Wasserstein distance or FID. We next study how this error depends on the representative and how a computable objective can guide its reduction.

3 NUMERICAL CONSEQUENCES OF GAUGE SELECTION

The marginal path and numerical configuration remain fixed. We analyze the response to gauge variations, establish the possibility of strict improvement, and derive a reference objective for representative selection. Assumptions and proofs appear in Appendix B.

3.1 FINITE-STEP RESPONSE TO GAUGE VARIATIONS

For an admissible variation $u_\alpha = u + \alpha g \in \mathcal{V}(p)$, both the numerical map and its exact reference depend on α.

Proposition 2 (Gauge variation of integration error)**.** *Let $e_\alpha = \Psi_\xi[u_\alpha] - \Phi_T[u_\alpha]$. Under Assumption 2,*

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathcal{J}_\xi(u_\alpha) = \mathbb{E}[\langle e_\alpha, \partial_\alpha \Psi_\xi[u_\alpha] - \partial_\alpha \Phi_T[u_\alpha]\rangle]\,. \tag{6}$$

The response pairs the integration residual with the difference between discrete and exact sensitivities. Marginal preservation does not force this pairing to vanish. When both signs of a small variation are feasible, a nonzero derivative gives a local descent direction; a zero derivative alone does not establish optimality.

Representative choice also affects error propagation. For a differentiable one-step method with maps F_i^u, write $x_i = \Phi_{t_i}[u](x_0)$, $z_{i+1} = F_i^u(z_i)$, $z_0 = x_0$, and $e_i = z_i - x_i$. Then

$$e_K = \sum_{i=0}^{K-1} M_{K-1} \cdots M_{i+1} \tau_i, \qquad \tau_i = F_i^u(x_i) - x_{i+1}, \tag{7}$$

where $M_i = \int_0^1 D_x F_i^u(x_i + s e_i)\,\mathrm{d}s$ and an empty product is the identity. A gauge can change both local defects and their propagation, motivating an endpoint objective that accounts for their combined effect. Appendix B.2 gives the resulting stability bound and multistep extension.

3.2 DIFFERENT NUMERICAL ACCURACY ALONG THE SAME MARGINAL PATH

The variation identity describes how error changes. The following construction establishes that gauge-equivalent representatives can attain strictly different accuracy under the same sampler.

Proposition 3 (Different errors for equivalent representatives)**.** *Let $v_t(x) = \lambda x$ on $[0, T]$, with $\lambda > 0$, and let p_0 be a smooth, positive, rotationally symmetric density on $\mathbb{R}^2$ with finite nonzero second moment. For every finite Euler grid with steps $h_i > 0$, there exists a smooth gauge $g_t(x) = \omega(t) J x$, where $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, preserving $p_t = (e^{\lambda t} I)_{\#} p_0$ and satisfying*

$$d_\xi(v + g) = 0 < d_\xi(v) = \left(e^{\lambda T} - \prod_{i=0}^{K-1} (1 + \lambda h_i) \right) \sqrt{\mathbb{E}\|X_0\|^2}. \tag{8}$$

Rotations preserve radial marginals. Choosing their velocities at solver nodes and their integrals between nodes makes each corrected Euler step equal its exact interval map. This proves that finite-step accuracy is not determined by the marginal path. The result covers arbitrary finite Euler grids and non-Gaussian radial laws, without guaranteeing improvement for every model or finite correction family. Appendix B.4 gives the construction and its stability under small perturbations.

3.3 A COMPUTABLE OBJECTIVE FOR REPRESENTATIVE SELECTION

When exact integration is unavailable, a refined numerical map $R[u]$ of the same field provides a computable objective:

$$d_{\mathrm{ref}}(u) = \|\Psi_\xi[u] - R[u]\|_{L^2(p_0)}, \qquad \mathcal{J}_{\mathrm{ref}}(u) = \tfrac{1}{2} d_{\mathrm{ref}}(u)^2. \tag{9}$$

Proposition 4 (Reference accuracy and representative selection)**.** *If $\|R[u] - \Phi_T[u]\|_{L^2(p_0)} \le \varepsilon_{\mathrm{ref}}(u)$, then*

$$|d_{\mathrm{ref}}(u) - d_\xi(u)| \le \varepsilon_{\mathrm{ref}}(u). \tag{10}$$

If this reference bound is at most ε uniformly on $\mathcal{K}$ and $d_{\mathrm{ref}}(\widehat{u}) \le \inf_{u \in \mathcal{K}} d_{\mathrm{ref}}(u) + \delta$ for $\delta \ge 0$, then

$$d_\xi(\widehat{u}) \le \inf_{u \in \mathcal{K}} d_\xi(u) + 2\varepsilon + \delta. \tag{11}$$

A reduction exceeding the sum of two candidates' reference bounds certifies lower true error. The selection guarantee separates reference accuracy from the achieved optimization gap, measured on the root-error scale. A denser rollout alone does not certify a uniform bound, and empirical fitting introduces sampling error. These conditions motivate separate checks of reference precision and optimization in practice. Appendix B.5 provides the proof and squared-loss conversion.

4 A Practical Framework for Numerical Gauge Fixing

We now turn representative selection into a practical optimization procedure. The framework combines a tractable correction family, a model-derived carrier when the path score is unavailable, and differentiation through coarse and refined rollouts of the same field.

4.1 Tractable Gauge Parameterization

Proposition 1 defines feasible corrections through skew potentials. For computation, we choose a finite family, for example $A_\eta(x,t) = \sum_{m=1}^{M} \eta_m B_m(x,t)$ with $B_m^\top = -B_m$ and $A_0 = 0$. When s_t is known, Equation (3) gives exact gauges, including the spatial-divergence term. The basis determines accessible directions and evaluation cost. Our GMM experiments compare Constant, RBF, and Fourier potentials to examine this design choice; the framework supports both state-independent and state-dependent potentials.

4.2 Model-Derived Carriers and Marginal Consistency

For images, we use state-independent $A_\eta(t)$ and a model-derived vector field, the *carrier* $r_{\theta,t}$, to construct

$$\widehat{u}_{\eta,t}(x) = v_{\theta,t}(x) + A_\eta(t) r_{\theta,t}(x). \tag{12}$$

Appendix D derives carriers from standard prediction types. Reusing the base model prediction makes the correction available without additional backbone evaluations. However, even an exact training-path score need not equal the score of the frozen model's induced path. Writing

$$r_{\theta,t} = c_t s_t + \kappa_t, \qquad \widehat{u}_{\eta,t} = v_t + c_t A_\eta(t) s_t + A_\eta(t)\kappa_t \tag{13}$$

with time-only scalar c_t separates an exact gauge component from a residual that can change the marginals. Under the conditions of Appendix C,

$$W_2\big((\Psi_\xi[\widehat{u}_\eta])_{\#} p_0, p_T\big) \le d_{\mathrm{ref}}(\widehat{u}_\eta) + \varepsilon_{\mathrm{ref}}(\widehat{u}_\eta) + B_{\mathrm{car}}(\eta). \tag{14}$$

The terms separate numerical discrepancy, reference error, and carrier-induced drift. Optimizing the first does not control the other two. This distinction motivates the numerical, marginal, and quality evaluations in Section 5; exact gauge guarantees apply only when the carrier is consistent and the regularity conditions hold.

4.3 Same-Field Optimization and Deployment

For source samples $x_j \sim p_0$ and a shared output map H, we minimize

$$\widehat{L}_\xi(\eta) = \frac{1}{nd} \sum_{j=1}^{n} \|H \circ \Psi_\xi[u_\eta](x_j) - H \circ R[u_\eta](x_j)\|^2. \tag{15}$$

Here u_η denotes the chosen corrected field and d is the output dimension. The identity recovers the state-space objective; images use terminal denoising. Both branches start from the same inputs and change with η. Gradients pass through both rollouts, retaining derivatives through frozen backbone inputs, while only correction parameters are updated. Appendices E.2 and C give the normalization and output-map bounds.

Algorithm 1 produces candidate checkpoints, with selection specified in Section 5.1. Deployment uses the selected correction with the original solver and grid. Shared predictions preserve deployment backbone NFE, while operator arithmetic adds overhead; refined integration is an offline fitting cost. Changing the sampling configuration generally requires a new fit.

Algorithm 1 Numerical gauge fixing by same-field optimization

Input: Frozen v_θ; correction family g_η; source p_0; sampler ξ and reference ξ_{ref}; shared output map H; fitting and checkpoint protocol.
Output: Saved correction candidates $\mathcal{C}$.

Set $u_\eta = v_\theta + g_\eta$; initialize η and $\mathcal{C}$ under the protocol.
for each optimization update **do**
  Take a source batch $X = \{x_j\}_{j=1}^n$ under the fitting protocol.
  $Y \leftarrow H \circ \Psi_\xi[u_\eta](X)$; $Z \leftarrow H \circ R_{\xi_{\mathrm{ref}}}[u_\eta](X)$.
  $L \leftarrow (nd)^{-1} \sum_j \|Y_j - Z_j\|^2$.
  Update η using $\nabla_\eta L$ through both branches, keeping θ fixed.
  Save a copy of η in $\mathcal{C}$ at the prescribed checkpoints.
return $\mathcal{C}$.

5 Experiments

We evaluate numerical gauge fixing under exact marginal constraints and on frozen image generators. Gaussian and GMM flows test the numerical mechanism and correction-family design; image experiments assess numerical accuracy, marginal changes, and generation quality. Each Base / corrected comparison fixes the solver, grid, and NFE.

5.1 Experimental Setup

Models and correction families. We use Gaussian and GMM flows with uniform Euler grids, unconditional EDM on CIFAR-10 (Karras et al., 2022), and EDM2-XS on ImageNet-512×512 (Karras et al., 2024). Analytic models use exact skew constructions; images use state-independent operators with 48 and six parameters, respectively. Paired arms share source inputs, and pretrained models remain frozen. Analytic runs use CPU. Each image correction fit takes a few minutes on one RTX 4090, excluding checkpoint quality evaluation.

Optimization and selection. Only correction parameters are fitted with the same-field objective. GMM selects among final fits and zero correction by validation discrepancy. CIFAR-10 selects saved correction checkpoints, including initialization, by FID-5K; zero correction is separately reported as Base. Selected candidates are fixed before evaluation on an independent 50K source stream, with all row metrics using the same candidate. Gaussian results average five starts; GMM reports mean $\pm$ sample SD over three recompilations. Appendix E gives the protocols.

Numerical accuracy and generation quality. MSE is per-coordinate integration error against each field's own exact or refined flow. Gaussian evaluation uses its exact flow, GMM uses RK4-256, and CIFAR-10 uses 512 paired inputs with a 129-NFE same-field Heun reference. Output-to-target SW_2 evaluates analytic generation; FID and KID evaluate 50K CIFAR-10 samples, with KID scaled by 10^3. Reporting numerical and quality metrics together assesses both integration accuracy and agreement with the target distribution.

Marginal differences. Refined corrected and Base samples G_t, B_t give $D_t = 100\widehat{\mathrm{SW}}_2(G_t, B_t)/s_{B,t}$, where $s_{B,t}$ is the Base projected standard deviation. This normalization expresses distributional distance relative to the Base scale, rather than a fraction of displaced probability mass. CIFAR-10 uses 1,024 shared-source inputs and 256 projections. Gaussian values are analytic zeros; GMM uses physical time, while CIFAR-10 uses normalized reference-grid positions, with Output after terminal denoising. These finite-sample comparisons do not certify equivalence between measured nodes. Appendix E specifies the protocols and metric definitions.

5.2 Validation under Exact Marginal Constraints

Known Gaussian and GMM densities let us impose exact gauge constraints and test numerical improvement independently of changes in the target law.

Table 1: Gaussian and GMM with uniform Euler grids. MSE uses each field's own reference; SW_2 compares generated and target samples. D_t measures normalized refined-marginal differences.

Flow	Correction	NFE	MSE ↓		SW_2 ↓		Marginal difference D_t			
			Base	+Corr.	Base	+Corr.	0.25	0.50	0.75	1.00
Gaussian	Rotation	1	0.516	3.03×10^{-19}	0.718	9.59×10^{-11}	0%	0%	0%	0%
GMM	RBF	1	2.089	0.3308 ± 0.0008	1.461	0.385 ± 0.003	$(2.02 \pm 0.02)\%$	$(2.52 \pm 0.36)\%$	$(3.01 \pm 0.57)\%$	$(3.62 \pm 0.58)\%$
GMM	RBF	2	0.418	0.025 ± 0.002	0.628	0.093 ± 0.003	$(2.03 \pm 0.08)\%$	$(2.24 \pm 0.17)\%$	$(2.34 \pm 0.12)\%$	$(2.75 \pm 0.29)\%$
GMM	RBF	4	0.107	0.047 ± 0.023	0.310	0.192 ± 0.059	$(1.30 \pm 0.21)\%$	$(1.67 \pm 0.23)\%$	$(1.67 \pm 0.40)\%$	$(1.88 \pm 0.58)\%$
GMM	RBF	8	0.030	0.021 ± 0.002	0.161	0.135 ± 0.008	$(1.06 \pm 0.06)\%$	$(1.22 \pm 0.09)\%$	$(1.10 \pm 0.07)\%$	$(1.11 \pm 0.06)\%$
GMM	Constant	4	0.107	0.087 ± 0.008	0.310	0.279 ± 0.015	$(1.04 \pm 0.15)\%$	$(1.33 \pm 0.27)\%$	$(1.55 \pm 0.63)\%$	$(1.78 \pm 0.92)\%$
GMM	Fourier	4	0.107	0.056 ± 0.024	0.310	0.206 ± 0.055	$(1.26 \pm 0.11)\%$	$(1.55 \pm 0.28)\%$	$(1.63 \pm 0.43)\%$	$(1.86 \pm 0.58)\%$
GMM	Non-skew	4	0.107	0.01842 ± 0.00005	0.310	0.329 ± 0.028	$(29.79 \pm 1.32)\%$	$(29.68 \pm 2.69)\%$	$(27.10 \pm 2.38)\%$	$(26.17 \pm 1.94)\%$

Gaussian mechanism. The Gaussian row of Table 1 directly realizes Proposition 3. Optimizing a two-parameter rotation reduces one-step Euler MSE from 0.516 to numerical zero, while output-to-target SW_2 falls from 0.718 to below 10^{-10}. Every candidate preserves the Gaussian path, and the selected rotation makes one Euler step match its exact endpoint without additional evaluations. This isolates the numerical mechanism: changing particle dynamics makes the prescribed sampler more accurate along the same continuous marginal path.

Nonradial Gaussian mixture. The GMM rows test whether the same freedom remains useful for a nonlinear, nonradial flow. With exact scores and the skew construction of Proposition 1, the RBF family lowers mean MSE and target SW_2 at all four Euler budgets. At one NFE, the RBF correction reduces MSE from 2.089 to 0.3308 and target SW_2 from 1.461 to 0.385. Each budget uses its own fitted correction, demonstrating representative selection for different prescribed discretizations. These results extend the numerical benefit beyond the radial construction while maintaining the exact marginal constraint.

Correction-family design. The four-NFE comparison links attainable accuracy to the search family in Section 4.1. RBF and Fourier potentials yield lower mean MSE and target SW_2 than the Constant family under the same gauge constraint. The families differ in both spatial structure and parameter count; they are evaluated as complete design choices. Their small nonzero empirical marginal distances reflect finite-sample and reference effects. Equal population marginals permit different empirical point clouds because gauge-equivalent flows change particle correspondences, as described in Section 2.2.

Role of the gauge constraint. The Non-skew row tests the role of the constraint itself. Replacing the rotation matrix by the identity in the four-parameter Constant family keeps capacity and the fitting protocol fixed while removing gauge admissibility. The control attains low integration MSE, but its mean target SW_2 is worse than Base and its refined marginals differ by roughly 26–30% of the Base scale. The matched skew family instead improves both MSE and target SW_2, with much smaller measured marginal differences. This contrast supports the constrained formulation in Section 2.3: the gauge structure directs numerical optimization toward representatives of the intended distributional path.

5.3 Image Generation with Frozen Models

We next examine whether representative selection improves sampling from pretrained image models. The model-derived carrier implements the correction, with each comparison fixing the backbone, solver, grid, and NFE.

CIFAR-10 generation. The main block of Table 2 evaluates separately optimized corrections for Euler/$\rho = 7$ at NFE $5, 7, 9$. The corrections reduce FID-50K from 68.69 to 60.72, from 36.38 to 31.57, and from 22.83 to 19.56, respectively; KID-50K improves at all three budgets as well. The selected corrections also lower mean same-field MSE by 0.78–1.28%. Each row evaluates one fixed candidate for all metrics, so the numerical and quality observations concern the same sampler.

Table 2: CIFAR-10 with a frozen EDM model. Base / +Gauge compare the same sampling configuration; Euler on $\rho = 7$ grids forms the main block, followed by five-NFE solver/grid comparisons.

Solver	Grid	NFE	MSE ↓		FID-50K ↓		KID-50K×10^3 ↓		Marginal difference at t ↓				
			Base	+Gauge	Base	+Gauge	Base	+Gauge	0.25	0.50	0.75	1.00	Output
Euler	$\rho = 7$	5	0.071577	0.071022	68.691	60.719	76.164	55.120	0.53%	1.31%	1.73%	1.73%	1.73%
		7	0.046256	0.045665	36.380	31.565	38.792	26.363	0.77%	1.87%	2.33%	2.25%	2.26%
		9	0.032941	0.032585	22.832	19.562	23.580	15.624	0.86%	2.16%	2.87%	2.83%	2.83%
DPM++2M	$\rho = 7$	5	0.022710	0.022473	36.624	31.459	38.157	25.598	0.37%	0.62%	1.59%	1.71%	1.71%
Euler	log	5	0.091647	0.090564	74.160	67.678	79.883	62.581	0.54%	1.34%	1.78%	1.77%	1.77%

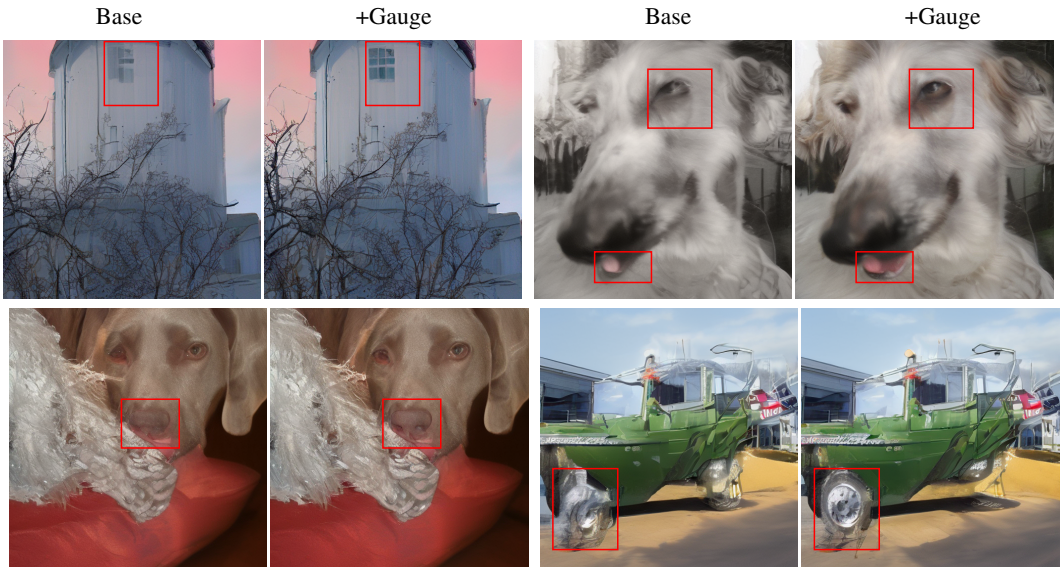

Figure 2: Paired 512×512 generation before and after numerical gauge fixing. Base is left and the optimized +Gauge result is right. Each pair uses frozen EDM2-XS, the same initial noise and class conditioning, and the same 16-NFE sampler. Red boxes highlight corresponding local image regions.

These 48-parameter updates demonstrate practical gains from representative selection at a fixed backbone-evaluation budget.

Marginal behavior. Table 2 also examines the refined distributions associated with these corrections. In the Euler block, normalized marginal differences range from 0.53% to 2.87% of the Base scale across the reported nodes. Thus the quality gains accompany modest measured changes in the refined marginal distributions. Evaluating intermediate states and the terminal output connects the image experiment to the path-based formulation, rather than assessing only endpoint appearance. These finite-node diagnostics complement the numerical measurements in the three-term decomposition of Equation (14); the additional sampling-scale comparisons appear in Appendix F.2.

Solver and grid compatibility. The lower block of Table 2 changes one component of the five-NFE Euler/$\rho = 7$ setting: DPM++2M changes the solver (Lu et al., 2022b), and the logarithmic row changes the grid. Fitting the same correction family to each configuration reduces FID-50K from 36.62 to 31.46 for DPM++2M and from 74.16 to 67.68 for Euler/log; KID and mean numerical MSE also decrease. The gain therefore appears with more than one integration rule and time allocation. This supports the complementary role proposed in the introduction: solver and grid design specify the discretization, while gauge selection adapts the field to that configuration. The correction reuses model predictions without additional backbone calls; measured arithmetic overhead is reported in Appendix F.4.

High-resolution generation. We further apply numerical gauge fixing to EDM2-XS with deterministic DDIM and a fixed rescaled entropic schedule (Stancevic et al., 2025). Figure 2 compares matched initial noise and class conditioning before and after optimizing a six-parameter channel-skew correction. In the displayed examples, the highlighted regions show clearer eye and muzzle contours and a more coherent vehicle-wheel outline, while the overall scene composition remains similar. These paired changes illustrate the effect of representative selection on local structure at 512×512 resolution. The spatially shared operator uses the existing denoiser prediction, retaining the same solver, grid, and 16 backbone evaluations. This qualitative extension demonstrates that the same optimization interface accommodates a higher-resolution generator with a compact correction family; Appendix E.6 specifies its construction.

Additional diagnostics. Appendix F.1 reports paired uncertainty and reference-refinement analyses for the numerical estimates. Appendix F.3 summarizes optimization trajectories, initialization effects, and configurations without quality gains, complementing the fixed-candidate comparisons in Table 2.

6 Related Work

Numerical solvers. DDIM provides deterministic diffusion sampling (Song et al., 2021a); PNDM adapts multistep integration (Liu et al., 2022), while DEIS and DPM-Solver exploit diffusion ODE structure (Zhang & Chen, 2023; Lu et al., 2022a). EDM and DPM-Solver++ improve practical discretizations (Karras et al., 2022; Lu et al., 2022b); UniPC and DPM-Solver-v3 use predictor–corrector updates and model statistics (Zhao et al., 2023; Zheng et al., 2023). Bespoke solvers and S4S learn integration parameters (Shaul et al., 2024; Frankel et al., 2025). These methods change how a field is integrated; our optimization selects the field for a fixed solver.

Time grids. Dynamic programming selects DDPM time grids (Watson et al., 2021). AYS optimizes diffusion sampling schedules (Sabour et al., 2024), GITS uses trajectory regularity (Chen et al., 2024), and entropic schedules allocate steps using an entropy-based criterion (Stancevic et al., 2025). These methods redistribute evaluations; we select a velocity representative for a fixed grid.

Rectified flow and transport-oriented training. Rectified flow learns straighter trajectories and iteratively improves the source–target coupling through reflow (Liu et al., 2023). Improved reflow training (Lee et al., 2024) and minibatch optimal-transport flow matching (Tong et al., 2024) further simplify learned dynamics; InstaFlow combines reflow with distillation (Liu et al., 2024). These approaches learn new flows or couplings. Our ideal search instead preserves the full pretrained marginal path and selects a representative for a fixed sampler.

Distillation. Progressive distillation reduces the number of teacher steps (Salimans & Ho, 2022), while consistency models learn maps that support one- or few-step generation (Song et al., 2023). We instead retain the backbone and optimize integration of each candidate's own dynamics, with no fixed teacher correspondence or replacement sampling map.

Gauge-equivalent representations. Gauge freedom (Horvat & Pfister, 2024) and skew probability-flux constructions (Ma et al., 2015) allow marginal-preserving dynamics. Flux Matching studies shared-path fields and score-model corrections in its Appendices E and C.4 (Pao-Huang et al., 2026). We optimize their finite-step integration accuracy through a same-field objective with the backbone, solver, and grid fixed.

7 Conclusion

Velocity fields realizing the same marginal path can exhibit different finite-step integration errors under a fixed sampler. We characterize this numerical response and connect a computable reference objective to true integration error under the stated regularity and reference-accuracy assumptions. This analysis motivates a post-training framework based on same-field optimization, whose effectiveness and limitations we examine in both exact models and frozen image generators. Together, these results identify continuous representative selection as a numerical optimization axis for few-step generative sampling, complementary to solver and time-grid design.

Reproducibility Statement

To support reproducibility, we provide the source code as a ZIP archive in the supplementary material.

AI Use Statement

We used generative AI tools to assist with language polishing, literature retrieval and discovery, research ideation and execution, and mathematical proofs. The authors take full responsibility for the final content of this paper, including the accuracy of the references, the validity of the mathematical arguments, and the originality and integrity of all AI-assisted contributions.

References

Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In *International Conference on Learning Representations*, 2018. URL https://arxiv.org/abs/1801.01401.

Defang Chen, Zhenyu Zhou, Can Wang, Chunhua Shen, and Siwei Lyu. On the trajectory regularity of ODE-based diffusion sampling. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 7905–7934. PMLR, 2024. URL https://proceedings.mlr.press/v235/chen24bm.html.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 12606–12633. PMLR, 2024. URL https://proceedings.mlr.press/v235/esser24a.html.

Eric Frankel, Sitan Chen, Jerry Li, Pang Wei Koh, Lillian J. Ratliff, and Sewoong Oh. S4S: Solving for a fast diffusion model solver. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 17542–17574. PMLR, 2025. URL https://proceedings.mlr.press/v267/frankel25a.html.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1706.08500.

Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J. Fleet. Video diffusion models. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/39235c56aef13fb05a6adc95eb9d8d66-Abstract-Conference.html.

Christian Horvat and Jean-Pascal Pfister. On gauge freedom, conservativity and intrinsic dimensionality estimation in diffusion models. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/d553f0e0abb80e2a60328d634583bd2e-Abstract-Conference.html.

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, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/a98846e9d9cc01cfb87eb694d946ce6b-Abstract-Conference.html.

Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, and Samuli Laine. Analyzing and improving the training dynamics of diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24174–24184, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Karras_Analyzing_and_Improving_the_Training_Dynamics_of_Diffusion_Models_CVPR_2024_paper.html.

Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://arxiv.org/abs/2405.20320.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2202.09778.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.03003.

Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, and Qiang Liu. InstaFlow: One step is enough for high-quality diffusion-based text-to-image generation. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2309.06380.

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, 2022a. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/260a14acce2a89dad36adc8eefe7c59e-Abstract-Conference.html.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models, 2022b. URL https://arxiv.org/abs/2211.01095.

Yi-An Ma, Tianqi Chen, and Emily B. Fox. A complete recipe for stochastic gradient MCMC. In *Advances in Neural Information Processing Systems*, volume 28, 2015. URL https://proceedings.neurips.cc/paper/2015/hash/9a4400501febb2a95e79248486a5f6d3-Abstract.html.

Peter Pao-Huang, Xiaojie Qiu, and Stefano Ermon. Generative modeling with flux matching, 2026. URL https://arxiv.org/abs/2605.07319.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your steps: Optimizing sampling schedules in diffusion models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 42947–42975. PMLR, 2024. URL https://proceedings.mlr.press/v235/sabour24a.html.

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

Neta Shaul, Uriel Singer, Ricky T. Q. Chen, Matthew Le, Ali Thabet, Albert Pumarola, and Yaron Lipman. Bespoke non-stationary solvers for fast sampling of diffusion and flow models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 44603–44627. PMLR, 2024. URL https://proceedings.mlr.press/v235/shaul24a.html.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a. URL https://arxiv.org/abs/2010.02502.

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*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 32211–32252. PMLR, 2023. URL https://proceedings.mlr.press/v202/song23a.html.

Dejan Stancevic, Florian Handke, and Luca Ambrogioni. Entropic time schedulers for generative diffusion models. *arXiv preprint arXiv:2504.13612*, 2025. URL https://arxiv.org/abs/2504.13612.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. URL https://arxiv.org/abs/2302.00482.

Daniel Watson, Jonathan Ho, Mohammad Norouzi, and William Chan. Learning to efficiently sample from diffusion probabilistic models, 2021. URL https://arxiv.org/abs/2106.03802.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2204.13902.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. UniPC: A unified predictor-corrector framework for fast sampling of diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://arxiv.org/abs/2302.04867.

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. URL https://arxiv.org/abs/2310.13268.

A Regularity and Gauge Construction

A.1 Continuous well-posedness

We state sufficient regularity conditions for the marginal-preservation statements in Section 2.2. Additional assumptions for numerical error variations are specified with the corresponding results.

Assumption 1 (Continuous well-posedness). *On $I = [0, T]$, the density path p is strictly positive and belongs to $C^1(I; C^2(\mathbb{R}^d))$. It solves the continuity equation for the base field v, and p_0 has finite second moment. Every velocity field u and gauge direction considered is continuous in time and continuously differentiable in space. For each such field, there exists $L_u \in L^1(I)$ such that*

$$\sup_{x\in\mathbb{R}^d}\left(\frac{\|u_t(x)\|}{1+\|x\|}+\|D_x u_t(x)\|_{\mathrm{op}}\right)\le L_u(t), \tag{16}$$

where $\|\cdot\|_{\mathrm{op}}$ is the Euclidean operator norm. For a skew-symmetric potential A_t, the matrix field $p_t A_t$ has continuous second spatial derivatives, and the induced correction satisfies the same well-posedness conditions.

These conditions ensure unique global characteristic flows, preservation of finite second moments, and uniqueness of the associated transported probability measures. They also justify the spatial derivative identities used below.

A.2 Proof of Proposition 1

Proof. **Gauge condition.** The base path satisfies $\partial_t p_t + \nabla\cdot(p_t v_t) = 0$. If $v + g$ induces the same path, its continuity equation gives $\partial_t p_t + \nabla\cdot(p_t(v_t + g_t)) = 0$; subtraction yields $\nabla\cdot(p_t g_t) = 0$. Conversely, this identity makes p solve the continuity equation for $v + g$ with the same initial law. Uniqueness under Assumption 1 implies $(\Phi_t[v + g])_{\#}p_0 = p_t$.

Marginal preservation. Let $M_t = p_t A_t$. Since p_t is scalar, $M_t^\top = -M_t$. With the row-divergence convention, the correction satisfies

$$(p_t g_t^A)_i = \sum_{j=1}^{d} \partial_{x_j} M_{t,ij}. \tag{17}$$

Consequently,

$$\begin{aligned}\nabla\cdot(p_t g_t^A) &= \sum_{i,j=1}^{d} \partial_{x_i}\partial_{x_j} M_{t,ij} \\ &= \sum_{i<j} \partial_{x_i}\partial_{x_j}\big(M_{t,ij} + M_{t,ji}\big) = 0.\end{aligned} \tag{18}$$

The diagonal terms vanish by skew symmetry, and the off-diagonal terms cancel because mixed derivatives commute. The product rule gives $p_t^{-1}\nabla\cdot(p_t A_t) = \nabla\cdot A_t + A_t s_t$, and

$$\partial_t p_t + \nabla\cdot\big(p_t(v_t + g_t^A)\big) = 0. \tag{19}$$

Starting from the same initial law p_0, uniqueness under Assumption 1 proves $g^A \in \mathcal{G}(p)$ and gives

$$\big(\Phi_t[v + g^A]\big)_{\#}p_0 = p_t, \qquad t \in [0, T]. \tag{20}$$

State-independent potentials. If A_t is independent of x, then

$$(g_t^A)_i = \frac{1}{p_t}\sum_j A_{t,ij}\partial_{x_j} p_t = \sum_j A_{t,ij}\partial_{x_j}\log p_t. \tag{21}$$

Hence

$$g_t^A = A_t\nabla\log p_t = A_t s_t. \tag{22}$$

Existence of nonzero admissible directions. Suppose $d \geq 2$. Let $J \in \mathbb{R}^{d\times d}$ have entries $J_{12} = -1$ and $J_{21} = 1$, with all other entries zero. Choose $\psi \in C_c^\infty(\mathbb{R}^d)$ such that $J\nabla\psi \not\equiv 0$, and choose a nonzero function $\chi \in C_c^\infty((0,T))$. Define

$$A_t(x) = \frac{\chi(t)\psi(x)}{p_t(x)} J. \tag{23}$$

This potential is skew-symmetric and satisfies

$$p_t A_t = \chi(t)\psi J, \qquad g_t^A = \frac{\chi(t)}{p_t} J\nabla\psi. \tag{24}$$

Its probability flux is therefore

$$p_t g_t^A = \chi(t) J\nabla\psi, \tag{25}$$

which is nonzero, smooth, and supported in a fixed compact spatial set.

To verify admissibility, choose a compact set K containing the support of ψ. Positivity and continuity of p imply

$$\inf_{(t,x)\in[0,T]\times K} p_t(x) > 0. \tag{26}$$

The required derivatives of p_t are bounded on this compact set. Hence g_t^A and its first spatial derivatives are uniformly bounded there. The correction vanishes outside K, so it is globally bounded and globally spatially Lipschitz, uniformly in time. It therefore satisfies Assumption 1, as does $v + g^A$. This establishes the existence of a nonzero admissible correction and completes the proof. □

B Numerical Response and Representative Selection

We analyze the same-field objective in Equation (4) under a fixed marginal path and sampling configuration.

B.1 Assumptions for numerical variations

Write $\Phi_{t\leftarrow s}[u]$ for the exact flow from time s to time t, with $\Phi_t[u] = \Phi_{t\leftarrow 0}[u]$. All comparisons use the same source law and numerical configuration, including any multistep initialization. The following conditions supplement Assumption 1.

Assumption 2 (Differentiable numerical variations)**.** *On each compact parameter interval considered, $u_\alpha = u + \alpha g \in \mathcal{V}(p)$. The exact and numerical endpoint maps are continuously differentiable in α for p_0-almost every input. Their residuals and parameter derivatives have common square-integrable dominating envelopes: there exist $H_0, H_1 \in L^2(p_0)$ such that, throughout the interval,*

$$\|e_\alpha(x)\| \leq H_0(x), \qquad \|\partial_\alpha \Psi_\xi[u_\alpha](x)\| + \|\partial_\alpha \Phi_T[u_\alpha](x)\| \leq H_1(x). \tag{27}$$

For the stepwise identities below, each numerical step map is continuously differentiable in its state and parameter on the sets involved.

These are sufficient conditions for the stated calculations, rather than verified global regularity bounds for the image network. The fixed numerical grid is part of ξ; differentiability is not asserted across discontinuous changes in an adaptive solver's control decisions.

B.2 Gauge variation and error propagation

Proof of the variation identity. By definition, $\mathcal{J}_\xi(u_\alpha) = \frac{1}{2}\mathbb{E}\|e_\alpha(X_0)\|^2$. Its pointwise derivative is

$$\frac{1}{2}\partial_\alpha \|e_\alpha\|^2 = \langle e_\alpha, \partial_\alpha \Psi_\xi[u_\alpha] - \partial_\alpha \Phi_T[u_\alpha]\rangle . \tag{28}$$

The absolute value is bounded by $H_0 H_1 \in L^1(p_0)$ by Cauchy–Schwarz. Differentiating under the expectation therefore gives Equation (6). □

Exact and discrete sensitivities. For $X_t^\alpha = \Phi_t[u_\alpha](X_0)$, the exact sensitivity $Y_t^\alpha = \partial_\alpha X_t^\alpha$ satisfies

$$\dot{Y}_t^\alpha = D_x u_{\alpha,t}(X_t^\alpha) Y_t^\alpha + g_t(X_t^\alpha), \qquad Y_0^\alpha = 0. \tag{29}$$

For a differentiable one-step method, $z_{i+1}^\alpha = F_i^{u_\alpha}(z_i^\alpha)$ and $Z_i^\alpha = \partial_\alpha z_i^\alpha$ give

$$Z_{i+1}^\alpha = D_x F_i^{u_\alpha}(z_i^\alpha) Z_i^\alpha + \partial_\alpha F_i^{u_\alpha}(z_i^\alpha), \qquad Z_0^\alpha = 0, \tag{30}$$

where the last partial derivative holds the state argument fixed. Thus the response in Equation (6) is $\mathbb{E}\langle e_\alpha, Z_K^\alpha - Y_T^\alpha \rangle$. If the segment $\alpha \in [0,1]$ satisfies the assumptions, integration yields

$$\mathcal{J}_\xi(u+g) - \mathcal{J}_\xi(u) = \int_0^1 \mathbb{E}\langle e_\alpha, Z_K^\alpha - Y_T^\alpha \rangle \, \mathrm{d}\alpha. \tag{31}$$

Equality of exact marginal laws does not set the particle sensitivity Y_T^α to zero. Both sensitivities must therefore be retained when measuring the response of each candidate's own integration error.

Propagation of local defects. For a fixed field and input, let $x_i = \Phi_{t_i}[u](x_0)$ and $z_{i+1} = F_i^u(z_i)$, with $z_0 = x_0$. Define

$$e_i = z_i - x_i, \qquad \tau_i = F_i^u(x_i) - x_{i+1}, \qquad M_i = \int_0^1 D_x F_i^u(x_i + s e_i) \, \mathrm{d}s. \tag{32}$$

The fundamental theorem of calculus gives

$$e_{i+1} = F_i^u(z_i) - F_i^u(x_i) + \tau_i = M_i e_i + \tau_i. \tag{33}$$

Expanding from $e_0 = 0$ gives Equation (7). The product multiplying τ_{K-1} is the identity. The formula makes explicit that a change of representative can affect both the defects and the matrices propagating them. If $\|D_x F_i^u\| \le L_i$ along the segments, then $\|M_i\| \le L_i$. If the constants L_i bound the derivatives on all relevant state segments and the defects are square-integrable, taking the $L^2(p_0)$ norm and applying Minkowski's inequality gives

$$d_\xi(u) \le \sum_{i=0}^{K-1} \left(\prod_{j=i+1}^{K-1} L_j \right) \|\tau_i\|_{L^2(p_0)}. \tag{34}$$

For Euler, $D_x F_i^u = I + h_i D_x u_{t_i}$ gives $L_i \le 1 + h_i \ell_i$ under the stated spatial Lipschitz bound.

For a multistep method, augment the numerical state with its stored history and compare it with a corresponding exact history. The same argument gives a recursion in this enlarged state; the endpoint error is its projection onto the final state coordinates. A nonzero initial history error contributes a propagated initial-error term, or can be incorporated through startup defects. When differentiating stored model evaluations, their dependence on the gauge parameter must also be retained.

B.3 Local descent and the zero correction

Suppose $u + \alpha g \in \mathcal{K}$ for every $|\alpha| < a$ for some $a > 0$, and let $b = \frac{\mathrm{d}}{\mathrm{d}\alpha} \mathcal{J}_\xi(u + \alpha g)\big|_{\alpha=0}$. Differentiability gives

$$\mathcal{J}_\xi(u + \alpha g) = \mathcal{J}_\xi(u) + \alpha b + o(|\alpha|). \tag{35}$$

If $b \neq 0$, taking $\alpha = -h \operatorname{sign}(b)$ for sufficiently small $h > 0$ lowers the objective. The availability of both signs is essential for this conclusion. At a local minimum the derivative must vanish along every such two-sided feasible direction; along a direction feasible only for positive α, the right derivative must instead be nonnegative. These are necessary conditions, not sufficient conditions for optimality.

Since $v \in \mathcal{K}$, the zero correction is always included and $\inf_{u \in \mathcal{K}} \mathcal{J}_\xi(u) \le \mathcal{J}_\xi(v)$. This does not establish attainment of the infimum or strict improvement. In particular, if $d_\xi(v) = 0$, the base representative is already optimal. More generally, optimality is relative to the selected family. The construction below also shows that a zero first derivative along a direction can coexist with a strictly better representative on that same line.

B.4 Strict improvement under a fixed Euler grid

This section proves Proposition 3 and describes its distributional consequence, stability, and a Gaussian example.

Proof. Let $p_t = (e^{\lambda t} I)_{\#} p_0$ and $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. Rotational symmetry gives $\nabla p_t(x)^\top Jx = 0$, while $\nabla \cdot (Jx) = 0$. Therefore

$$\nabla \cdot \big(p_t \omega(t) Jx\big) = 0 \tag{36}$$

for every smooth ω. The resulting linear fields satisfy the well-posedness conditions on the finite time interval.

For each grid interval, set

$$a_i = 1 + \lambda h_i, \qquad q_i = e^{\lambda h_i}, \qquad b_i = \sqrt{q_i^2 - a_i^2}, \qquad \theta_i = \arctan(b_i/a_i). \tag{37}$$

Since $e^z > 1 + z$ for $z > 0$, $b_i > 0$. Choose a smooth function satisfying

$$\omega(t_i) = b_i/h_i, \qquad \int_{t_i}^{t_{i+1}} \omega(t)\,\mathrm{d}t = \theta_i, \qquad i = 0, \dots, K-1. \tag{38}$$

To construct it explicitly, first choose a polynomial P interpolating all prescribed values at $t_0, \dots, t_{K-1}$. For each interval choose $\beta_i \in C_c^\infty((t_i, t_{i+1}))$ with integral one and define

$$\omega(t) = P(t) + \sum_{i=0}^{K-1} \left(\theta_i - \int_{t_i}^{t_{i+1}} P(s)\,\mathrm{d}s\right) \beta_i(t). \tag{39}$$

The bumps change each interval integral independently and vanish near every grid node, preserving the prescribed values and smoothness.

For $u_t(x) = (\lambda I + \omega(t) J)x$, commutativity of the coefficient matrices implies

$$\Phi_{t_{i+1} \leftarrow t_i}[u](x) = q_i \exp(\theta_i J)x = (a_i I + b_i J)x. \tag{40}$$

The Euler map on the same interval is

$$x + h_i u_{t_i}(x) = \big((1 + \lambda h_i) I + h_i \omega(t_i) J\big)x = (a_i I + b_i J)x. \tag{41}$$

Thus exact and numerical states agree at every grid node, for every input. At intermediate times the exact flow is a rotation followed by the scaling $e^{\lambda t}$, so its law remains p_t. Hence $d_\xi(v + g) = 0$. For the base field,

$$\Phi_T[v](x) = e^{\lambda T} x, \qquad \Psi_\xi[v](x) = \left(\prod_{i=0}^{K-1} a_i\right) x. \tag{42}$$

Their $L^2(p_0)$ distance gives Equation (8), which is positive because $\prod_i a_i < \prod_i q_i = e^{\lambda T}$ and the second moment is nonzero. □

Distributional consequence. The corrected discrete output has law p_T. Write $a = \prod_i a_i$ and $q = e^{\lambda T}$. For any coupling of $Y \sim (aI)_{\#} p_0$ and $Z \sim (qI)_{\#} p_0$, the reverse triangle inequality gives

$$\|Y - Z\|_{L^2} \geq \big|\|Z\|_{L^2} - \|Y\|_{L^2}\big| = (q - a)\sqrt{\mathbb{E}\|X_0\|^2}. \tag{43}$$

The coupling $Y = aX_0$, $Z = qX_0$ attains this bound. Consequently the base numerical distribution has strictly positive Wasserstein error equal to $d_\xi(v)$ in this construction. The difference therefore also appears at the distribution level.

Stability of the strict inequality. For a fixed finite grid, the Euler endpoint matrix depends continuously on the finitely many values $\omega(t_i)$, and the exact endpoint matrix depends continuously on $\int_0^T \omega(t)\,\mathrm{d}t$. Since p_0 has finite second moment, their $L^2(p_0)$ discrepancy is continuous in ω under the uniform norm. It is zero at the constructed function, while the base error is positive. Thus sufficiently small smooth uniform perturbations of ω retain strict improvement and exact marginal preservation. This statement fixes the density path and solver.

A one-step Gaussian illustration. Take $T = \lambda = 1$, $p_0 = \mathcal{N}(0, I_2)$, and one Euler step. Set

$$b = \sqrt{e^2 - 4}, \qquad \theta = \arctan(b/2), \qquad \omega(t) = b + 2(\theta - b)t. \tag{44}$$

Then $\omega(0) = b$ and $\int_0^1 \omega(t)\,\mathrm{d}t = \theta$, so both corrected maps equal $2I + bJ = e\exp(\theta J)$. They induce covariance $e^2 I$, while the zero-gauge Euler error is $\sqrt{2}(e-2)$. This correction belongs to the state-independent skew-score family: $s_t(x) = -e^{-2t}x$ and $A_t = -e^{2t}\omega(t)J$ satisfy $A_t s_t(x) = \omega(t)Jx$.

Along $v + \alpha g$,

$$\mathcal{J}_\xi(v + \alpha g) = (2 - e\cos(\alpha\theta))^2 + (\alpha b - e\sin(\alpha\theta))^2. \tag{45}$$

Its derivative at zero vanishes, although the objective decreases from $(e-2)^2$ at zero to zero at one. Thus the nonzero-derivative criterion is sufficient, but not necessary, for a line of gauge-equivalent fields to contain a better representative. The general construction uses a flexible time function; it does not establish expressibility in any prescribed finite parameterization. For non-Gaussian radial densities, the rotational correction need not belong to the state-independent skew-score family.

B.5 Proof of Proposition 4

Approximate representative selection. If $\varepsilon_{\mathrm{ref}}(u) \le \varepsilon$ uniformly on $\mathcal{K}$ and, for $\delta \ge 0$, $d_{\mathrm{ref}}(\widehat{u}) \le \inf_{u\in\mathcal{K}} d_{\mathrm{ref}}(u) + \delta$, then

$$d_\xi(\widehat{u}) \le \inf_{u\in\mathcal{K}} d_\xi(u) + 2\varepsilon + \delta. \tag{46}$$

Proof. The reverse triangle inequality in $L^2(p_0)$ gives

$$|d_{\mathrm{ref}}(u) - d_\xi(u)| \le \|R[u] - \Phi_T[u]\|_{L^2(p_0)} \le \varepsilon_{\mathrm{ref}}(u). \tag{47}$$

Under the uniform bound and approximate-minimization premise,

$$\begin{aligned} d_\xi(\widehat{u}) &\le d_{\mathrm{ref}}(\widehat{u}) + \varepsilon && (48)\\ &\le \inf_{u\in\mathcal{K}} d_{\mathrm{ref}}(u) + \delta + \varepsilon && (49)\\ &\le \inf_{u\in\mathcal{K}} d_\xi(u) + \delta + 2\varepsilon. && (50) \end{aligned}$$

This argument uses infima and does not require an attained minimum. □

Comparison of two candidates. For u_0, u_1 with respective reference bounds $\varepsilon_0, \varepsilon_1$,

$$d_\xi(u_0) - d_\xi(u_1) \ge d_{\mathrm{ref}}(u_0) - d_{\mathrm{ref}}(u_1) - \varepsilon_0 - \varepsilon_1. \tag{51}$$

Hence a proxy improvement larger than $\varepsilon_0 + \varepsilon_1$ guarantees a true improvement. This pairwise conclusion needs bounds only for the two compared fields.

Squared-loss optimization gaps. The objectives d_{ref} and $\mathcal{J}_{\mathrm{ref}} = \frac{1}{2}d_{\mathrm{ref}}^2$ have the same minimizers, but their approximate-optimization gaps have different scales. If

$$\mathcal{J}_{\mathrm{ref}}(\widehat{u}) \le \inf_{u\in\mathcal{K}} \mathcal{J}_{\mathrm{ref}}(u) + \eta, \qquad \eta \ge 0, \tag{52}$$

and $m = \inf_{u\in\mathcal{K}} d_{\mathrm{ref}}(u)$, then

$$d_{\mathrm{ref}}(\widehat{u}) \le \sqrt{m^2 + 2\eta} \le m + \sqrt{2\eta}. \tag{53}$$

Equation (11) consequently holds with $\delta = \sqrt{2\eta}$. Any additional constant normalization of the training loss must also be accounted for when interpreting such a gap.

Scope of the reference guarantee. The proposition is a population statement conditional on reference precision and the achieved optimization gap. A denser rollout does not by itself certify a uniform error bound over a searched family. The result neither proves convergence of the chosen optimizer nor controls estimation error from finitely many source samples. The norm inequalities also hold for well-posed fields outside $\mathcal{V}(p)$; in that case they quantify each field's own integration error, without establishing preservation of the base marginal path.

C Carrier Consistency and Marginal Drift

We quantify the marginal drift induced by an approximate carrier for the state-independent skew-operator realization of Section 4. This bound does not assert coverage of arbitrary spatially dependent correction families. A small integration discrepancy and a small change of continuous marginals are distinct requirements.

Proposition 5 (Carrier-induced marginal drift)**.** *Let $s_t = \nabla \log p_t$, let A_t be state-independent and skew-symmetric, and let c_t be a scalar depending only on time. Suppose $c_t A_t$ satisfies the potential regularity conditions of Proposition 1, and both $\bar{u}_t = v_t + c_t A_t s_t$ and $\widehat{u}_t = \bar{u}_t + A_t \kappa_t$ satisfy Assumption 1. Assume*

$$\langle x - y, \widehat{u}_t(x) - \widehat{u}_t(y) \rangle \le \ell_t \|x - y\|^2, \qquad \ell \in L^1([0,T]), \tag{54}$$

and that the integral below is finite. Then, with $\widehat{p}_t = (\Phi_t[\widehat{u}])_{\#} p_0$,

$$W_2(\widehat{p}_T, p_T) \le B_{\mathrm{car}} := \int_0^T e^{\int_t^T \ell_\tau \, \mathrm{d}\tau} \|A_t \kappa_t\|_{L^2(p_t)} \, \mathrm{d}t. \tag{55}$$

Proof. Because $c_t A_t$ is state-independent and skew-symmetric, Proposition 1 gives $(\Phi_t[\bar{u}])_{\#} p_0 = p_t$. Couple $\widehat{X}_t = \Phi_t[\widehat{u}](X_0)$ and $\bar{X}_t = \Phi_t[\bar{u}](X_0)$ through the same source. Their difference satisfies

$$\dot{\Delta}_t = \widehat{u}_t(\widehat{X}_t) - \widehat{u}_t(\bar{X}_t) + A_t \kappa_t(\bar{X}_t), \qquad \Delta_0 = 0. \tag{56}$$

The one-sided Lipschitz bound and Cauchy–Schwarz imply

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\Delta_t\|_{L^2} \le \ell_t \|\Delta_t\|_{L^2} + \|A_t \kappa_t\|_{L^2(p_t)} \tag{57}$$

where the norm is positive. Regularizing the norm and passing to the limit gives the same integrated inequality at zero. Gronwall's inequality gives $\|\Delta_T\|_{L^2} \le B_{\mathrm{car}}$. The common-source coupling bounds $W_2(\widehat{p}_T, p_T)$ by this norm. □

For a carrier $r_t = c_t s_t + \kappa_t$, the implemented field is $\widehat{u}_t = v_t + A_t r_t$. Coupling its coarse and exact endpoints, then applying the reference bound and the preceding proposition, gives

$$W_2\big((\Psi_\xi[\widehat{u}])_{\#} p_0, p_T\big) \le d_{\mathrm{ref}}(\widehat{u}) + \varepsilon_{\mathrm{ref}}(\widehat{u}) + B_{\mathrm{car}}. \tag{58}$$

The three terms separate numerical discrepancy, reference error, and carrier-induced marginal drift. If the numerical statistic is a raw per-coordinate mean squared discrepancy E, its root discrepancy is $\sqrt{dE}$ in dimension d. An empirical statistic additionally requires sampling-error control for a population guarantee.

Skew symmetry without carrier consistency is insufficient. For the stationary path $p_t = \mathcal{N}(0, I_2)$, take $v = 0$, $s = -x$, $A = J$, and $r = -x + \epsilon J x$. Then

$$Ar = -Jx - \epsilon x, \qquad \nabla \cdot (p_t A r) = -\epsilon \nabla \cdot (p_t x), \tag{59}$$

which is not identically zero when $\epsilon \neq 0$. The resulting covariance is $e^{-2\epsilon t} I$ despite exact skew symmetry. Consistency with a score used during model training does not establish consistency with the actual model-induced law.

Positive-noise truncation and a terminal output map. For the image implementation in Appendices D.4 and E.1, reparameterizing the decreasing interval $[\sigma_{\max}, \sigma_{\min}]$, with $\sigma_{\min} > 0$, by increasing time places its ODE in the preceding setting. For a shared L_H-Lipschitz terminal map H, define

$$\begin{aligned} d^H_{\mathrm{ref}} &= \|H \circ \Psi_\xi[\widehat{u}] - H \circ R[\widehat{u}]\|_{L^2(p_0)}, \\ \varepsilon^H_{\mathrm{ref}} &= \|H \circ R[\widehat{u}] - H \circ \Phi_T[\widehat{u}]\|_{L^2(p_0)}. \end{aligned} \tag{60}$$

The same coupling argument yields

$$W_2\big((H \circ \Psi_\xi[\widehat{u}])_{\#} p_0, H_{\#} p_T\big) \le d^H_{\mathrm{ref}} + \varepsilon^H_{\mathrm{ref}} + L_H B_{\mathrm{car}}. \tag{61}$$

In that implementation, $A_\eta(\sigma_{\min}) = 0$ makes $H(x) = D_\theta(x, \sigma_{\min})$ common to all candidates, and the image loss is measured after this map. No certified L_H, uniform reference error, or carrier bound is supplied for the image network; this inequality identifies the approximation terms without certifying the measured image outputs.

Table 3: Carriers derived from standard model predictions. The factor c_t describes the ideal relation to the *training-path* score s_t^q, not an automatic guarantee for the frozen model's path. The last row assumes $\alpha_t^2 + \sigma_t^2 = 1$.

Prediction	Carrier $r_{\theta,t}(x)$	Ideal factor c_t
Score s_θ	$s_\theta(x,t)$	1
Noise ϵ_θ	$-\epsilon_\theta(x,t)$	σ_t
Data D_θ	$\alpha_t D_\theta(x,t) - x$	σ_t^2
Velocity v_θ	$\alpha_t v_\theta(x,t) - \dot{\alpha}_t x$	$\sigma_t(\sigma_t\dot{\alpha}_t - \alpha_t\dot{\sigma}_t)$
Diffusion v-prediction ν_θ	$-\sigma_t x - \alpha_t \nu_\theta(x,t)$	σ_t

D Carriers across Model Parameterizations

This appendix derives the model interfaces used in Section 4.2. The identities concern ideal predictions for a specified training path. We distinguish that path from the one induced by a frozen model.

D.1 Gaussian interpolation identities

Let

$$X_t = \alpha_t Y + \sigma_t Z, \qquad Z \sim \mathcal{N}(0, I), \qquad Z \perp Y, \tag{62}$$

with differentiable schedules, $\sigma_t > 0$ on the interval considered, and finite conditional moments. Denote its density by q_t and its score by $s_t^q = \nabla \log q_t$. Define the ideal conditional predictions

$$D_t^*(x) = \mathbb{E}[Y \mid X_t = x], \qquad \epsilon_t^*(x) = \mathbb{E}[Z \mid X_t = x], \tag{63}$$

and the interpolation's conditional-mean velocity

$$b_t^*(x) = \dot{\alpha}_t D_t^*(x) + \dot{\sigma}_t \epsilon_t^*(x). \tag{64}$$

These Gaussian conditional paths include standard diffusion and flow-matching constructions (Song et al., 2021b; Lipman et al., 2023). Differentiating the Gaussian conditional density and taking conditional expectations gives

$$s_t^q(x) = -\frac{x - \alpha_t D_t^*(x)}{\sigma_t^2} = -\frac{\epsilon_t^*(x)}{\sigma_t}, \qquad x = \alpha_t D_t^*(x) + \sigma_t \epsilon_t^*(x). \tag{65}$$

The actual reference remains $p_t = (\Phi_t[v_\theta])_{\#} p_0$; we do not assume $q_t = p_t$.

D.2 Carrier conversions

Table 3 lists computable carriers and their ideal relation $r_t^* = c_t s_t^q$. The first three follow directly from Equation (65). For interpolation velocity,

$$\begin{aligned} \alpha_t b_t^*(x) - \dot{\alpha}_t x &= (\alpha_t \dot{\sigma}_t - \sigma_t \dot{\alpha}_t)\epsilon_t^*(x) \\ &= \sigma_t(\sigma_t \dot{\alpha}_t - \alpha_t \dot{\sigma}_t) s_t^q(x). \end{aligned} \tag{66}$$

This identity requires the conditional-mean velocity of the specified independent Gaussian interpolation. It does not follow from an arbitrary generative ODE velocity.

We denote diffusion v-prediction by ν to distinguish it from ODE velocity (Salimans & Ho, 2022). Its ideal value is

$$\nu_t^*(x) = \mathbb{E}[\alpha_t Z - \sigma_t Y \mid X_t = x]. \tag{67}$$

When $\alpha_t^2 + \sigma_t^2 = 1$,

$$\epsilon_t^* = \sigma_t x + \alpha_t \nu_t^*, \qquad D_t^* = \alpha_t x - \sigma_t \nu_t^*. \tag{68}$$

Without that normalization, these conditional means are divided by $\alpha_t^2 + \sigma_t^2$, and the same carrier in Table 3 has factor $c_t = \sigma_t(\alpha_t^2 + \sigma_t^2)$.

D.3 Reusing predictions for the base velocity

For the standard base-velocity parameterizations associated with Equation (62), the following conversions reuse the original prediction for both velocity and carrier. Where the denominators are nonzero,

$$\begin{aligned} v_\theta^{\mathrm{score}} &= \frac{\dot\alpha_t}{\alpha_t}x + \left(\frac{\dot\alpha_t\sigma_t^2}{\alpha_t} - \sigma_t\dot\sigma_t\right)s_\theta, \\ v_\theta^{\mathrm{noise}} &= \frac{\dot\alpha_t}{\alpha_t}x + \left(\dot\sigma_t - \frac{\dot\alpha_t\sigma_t}{\alpha_t}\right)\epsilon_\theta, \\ v_\theta^{\mathrm{data}} &= \frac{\dot\sigma_t}{\sigma_t}x + \left(\dot\alpha_t - \frac{\alpha_t\dot\sigma_t}{\sigma_t}\right)D_\theta, \\ v_\theta^{\nu} &= (\alpha_t\dot\sigma_t - \sigma_t\dot\alpha_t)\nu_\theta, \qquad \alpha_t^2 + \sigma_t^2 = 1. \end{aligned} \tag{69}$$

A direct velocity predictor supplies v_θ itself. These equations describe compatible model interfaces; they do not redefine the base velocity of an arbitrary frozen generator. Existing preconditioning, time coordinates, and fixed conditioning must be respected. For specialized solvers, effective predictions are recovered from the corrected velocity using the solver's original parameterization.

D.4 Linear flow matching and EDM

Linear flow matching. For $\alpha_t = t$ and $\sigma_t = 1 - t$, on $0 \le t < 1$,

$$\begin{aligned} r_{\theta,t}(x) &= t\,v_\theta(x,t) - x, \qquad r_t^* = (1-t)s_t^q, \\ \widehat{u}_{\eta,t}(x) &= v_\theta(x,t) + A_\eta(t)\big(tv_\theta(x,t) - x\big). \end{aligned} \tag{70}$$

This example uses independent Gaussian noise and data. A training procedure that changes their coupling requires a separate check of the carrier identity.

EDM in noise coordinates. With noise scale σ as the integration coordinate, the denoiser defines (Karras et al., 2022)

$$v_{\theta,\sigma}(x) = \frac{x - D_\theta(x,\sigma)}{\sigma}, \qquad r_{\theta,\sigma} = -v_{\theta,\sigma}, \qquad r_\sigma^* = \sigma s_\sigma^q. \tag{71}$$

Thus $\widehat{u}_{\eta,\sigma} = (I - A_\eta)v_{\theta,\sigma}$, implemented by

$$D_\eta(x,\sigma) = D_\theta(x,\sigma) + A_\eta(\sigma)\big(x - D_\theta(x,\sigma)\big). \tag{72}$$

Sampling proceeds toward smaller σ, while the velocity is defined as $\mathrm{d}x/\mathrm{d}\sigma$. After reparameterizing by increasing time t,

$$v_{\theta,t} = \dot\sigma_t\frac{x - D_\theta(x,\sigma_t)}{\sigma_t}. \tag{73}$$

Choosing $r_{\theta,t} = -v_{\theta,t}$ then gives $c_t = \sigma_t\dot\sigma_t$. A time-only factor may be negative without invalidating skew admissibility.

D.5 Consistency, endpoints, and computational scope

Relative to the actual model path, define $\kappa_t = r_{\theta,t} - c_t\nabla\log p_t$. Within the preceding identities,

$$\kappa_t = \big(r_{\theta,t} - c_t\nabla\log q_t\big) + c_t\big(\nabla\log q_t - \nabla\log p_t\big). \tag{74}$$

The two terms reflect prediction error and a difference between the training and model-induced paths. Even ideal predictions need not give $p_t = q_t$ if the initial law or sampling definition differs. Conversion alone therefore does not establish a small κ_t. For state-independent A_η, $\kappa_t = 0$ is sufficient for exact gauge admissibility under the regularity assumptions. Otherwise Appendix C applies conditionally to $A_\eta\kappa_t$. A general continuous normalizing flow outside these model structures requires its own carrier construction.

All derivatives use the current ODE coordinate. Time reparameterization changes the velocity and the associated carrier factor together. Conversions involving $1/\alpha_t$ or $1/\sigma_t$ apply only where these are

defined; endpoints require the model's nonsingular formula, positive-noise truncation, or a justified limit. No inversion of c_t is required, including when $c_t = 0$.

The listed carriers use the current state, known schedules, and existing model predictions. In the corresponding standard parameterizations, the base velocity and carrier share backbone evaluations, while operator arithmetic remains an additional cost. A construction that separately evaluates another score network does not inherit this NFE property.

E Implementation and Experimental Protocols

E.1 The 48-parameter image family

The image state has shape $3 \times 32 \times 32$. On the positive-noise interval, define

$$\tau(\sigma) = \frac{\log(80/\sigma)}{\log(80/0.002)}, \qquad h(\sigma) = 4\tau(1-\tau),$$
$$(\beta_0, \beta_1, \beta_2, \beta_3) = ((1-\tau)^3, 3\tau(1-\tau)^2, 3\tau^2(1-\tau), \tau^3). \tag{75}$$

The implementation clamps τ to $[0, 1]$. Four disjoint radial frequency bands have boundaries 0.125, 0.25, and 0.5 cycles per pixel. A frequency exactly at a boundary is assigned to the higher band. The final band includes all remaining frequencies, and the zero frequency belongs to the first band. The masks are even under frequency inversion.

For each temporal index ℓ and band b, three coefficients specify

$$K_{\ell b} = \begin{pmatrix} 0 & \eta_{\ell b,01} & \eta_{\ell b,02} \\ -\eta_{\ell b,01} & 0 & \eta_{\ell b,12} \\ -\eta_{\ell b,02} & -\eta_{\ell b,12} & 0 \end{pmatrix}, \qquad S_b(\sigma) = h(\sigma) \sum_{\ell=0}^{3} \beta_\ell(\sigma) K_{\ell b}. \tag{76}$$

Each RGB Fourier coefficient is multiplied by its band's S_b. The real-valued implementation uses `rfft2` and `irfft2` with `norm="ortho"`. Equivalently, in the full complex Fourier representation, the transform is unitary and the frequencywise matrices satisfy $S_b^* = -S_b$. The even real masks preserve conjugate symmetry. Hence the operator on real images is skew-adjoint. It is linear and state independent: only time and the learned coefficients determine its action.

The envelope gives $A_\eta(80) = A_\eta(0.002) = 0$. Both the coarse sampler and the reference end their positive-noise integration at $\sigma_{\min} = 0.002$ and then use the shared terminal denoising map. The gauge introduces no additional backbone calls; Fourier transforms and channel operations still have computational cost. Appendix F.4 reports the measured rollout overhead at fixed NFE.

E.2 Same-field gradients and normalization

Source arrays may be fixed or resampled according to the fitting protocol. Checkpoint indices specify whether initialization is retained and which later parameters are saved; the final fit is included. Selection remains separate from candidate generation. GMM compares final fits and zero correction, while CIFAR-10 uses its specified saved correction candidates with Base reported separately.

The image implementation uses Equation (15) with its shared terminal map, normalized by a fixed $B_\xi > 0$. For $y_j(\eta) = H \circ \Psi_\xi[\widehat{u}_\eta](x_j)$ and $z_j(\eta) = H \circ R[\widehat{u}_\eta](x_j)$, where H is the common terminal denoiser specified in Appendix C, define $\widehat{L}_\xi^{\mathrm{impl}} = (B_\xi nd)^{-1} \sum_j \|y_j - z_j\|^2$. Its gradient is

$$\nabla_\eta \widehat{L}_\xi^{\mathrm{impl}} = \frac{2}{B_\xi nd} \sum_j (D_\eta y_j - D_\eta z_j)^\top (y_j - z_j). \tag{77}$$

The implementation differentiates the discretized coarse and reference computations, using activation checkpointing to reduce memory. Backbone parameters are frozen, but derivatives through backbone inputs are retained. DPM++2M history predictions also remain in the graph. Stopping gradients through the reference would remove $D_\eta z_j$ and would not optimize the same objective.

The normalization B_ξ is the average zero-gauge per-coordinate endpoint discrepancy on 512 fixed source samples, evaluated once for each solver/grid configuration. It is positive for these configurations and detached during training. This rescales the loss without changing its minimizers. Numerical tables report the unnormalized discrepancy E, not the normalized training loss or $\mathcal{J}_\xi = \frac{1}{2} d_\xi^2$.

E.3 GAUSSIAN AND GAUSSIAN-MIXTURE EXPERIMENTS

All analytic experiments use uniform Euler grids on $[0, 1]$; NFE equals the number of field evaluations, without equating the arithmetic costs of different families.

Gaussian construction and evaluation. For $X_0 \sim \mathcal{N}(0, I_2)$ and $v_t(x) = x$, the path is $p_t = \mathcal{N}(0, e^{2t} I_2)$. The Rotation family uses $g_\eta(x, t) = [\eta_0 + \eta_1(2t - 1)]Jx$, with $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. Its exact complex endpoint multiplier is $m_\eta = e^{1+i\eta_0}$. If m_{Euler} is the corresponding numerical multiplier, population per-coordinate MSE is $|m_{\mathrm{Euler}} - m_\eta|^2$ and target SW_2 is $\big||m_{\mathrm{Euler}}| - e\big|$. Every parameter value preserves the analytic marginals, giving the zero path entries in Table 1.

The Gaussian row averages five starts $0.01\mathcal{N}(0, I_2)$ with seeds 20260921–20260925, without best-seed selection. L-BFGS-B uses bounds $[-4, 4]$, at most 1,000 iterations, gradient tolerance 10^{-10}, and function tolerance 10^{-15}. The fitting objective compares Euler with same-field RK4-128, normalized by zero-gauge discrepancy; two deterministic states with second moment I_2 evaluate this linear-map loss exactly. All five one-NFE fits meet the stopping tolerances; final evaluation uses the exact map above.

GMM path and analytic fields. Let Y have mixture weights $(0.5, 0.3, 0.2)$, component means $(-2, 0), (2, 0), (0, 2)$, and within-component covariance $0.25 I_2$. With independent $Z \sim \mathcal{N}(0, I_2)$, the interpolation $X_t = tY + (1 - t)Z$ has

$$p_t(x) = \sum_{k=1}^{3} w_k \mathcal{N}(x; t\mu_k, \varsigma_t^2 I_2), \qquad \varsigma_t^2 = (1 - t)^2 + 0.25t^2. \tag{78}$$

Writing $\gamma_k(x, t)$ for component posterior probabilities and $\bar{\mu}(x, t) = \sum_k \gamma_k \mu_k$, the exact score and velocity are

$$\begin{gathered} s_t(x) = -\frac{x - t\bar{\mu}}{\varsigma_t^2}, \qquad v_t(x) = \bar{\mu} + \frac{0.25t - (1 - t)}{\varsigma_t^2}(x - t\bar{\mu}), \\ r_t(x) = tv_t(x) - x = (1 - t)s_t(x). \end{gathered} \tag{79}$$

Gauge families. For a scalar potential f_η, set $A_\eta(x, t) = (1 - t)f_\eta(x, t)J$. The full correction includes the spatial divergence:

$$g_\eta(x, t) = (1 - t)J\big(\nabla_x f_\eta(x, t) + f_\eta(x, t)s_t(x)\big). \tag{80}$$

All families use the four cubic Bernstein time functions $\binom{3}{\ell} t^\ell (1 - t)^{3-\ell}$, $\ell = 0, \ldots, 3$.

- **Constant:** f_η depends only on time, giving four parameters. The potential is spatially constant, while the correction still depends on x through s_t.
- **RBF:** eighteen Gaussian features with centers in $\{-2, 0, 2\}^2$ and widths 1 and 2, each with four time coefficients, giving 72 parameters.
- **Fourier:** a constant plus sine/cosine features at frequencies $\frac{\pi}{4}\{(1, 0), (0, 1), (1, 1), (1, -1)\}$, giving nine spatial features and 36 parameters.

RBF and Fourier potentials are multiplied by the same radial C^3 window: one for $\|x\| \le 6$, zero for $\|x\| \ge 8$, and $1 - 35z^4 + 84z^5 - 70z^6 + 20z^7$ between them, with $z = (\|x\|^2 - 36)/28$. The implementation includes the window gradient in Equation (80). Families and capacities are fixed before the confirmation fits.

Matched Non-skew control. The control replaces J by I in the spatially constant family, giving $u_\eta = v + f_\eta(t)r_t$. Its four coefficients, training and validation arrays, initialization amplitudes, bounds, budgets, and selection rule match the spatially constant gauge. This correction generally has nonzero weighted divergence and is not in $\mathcal{G}(p)$. Its comparison with the spatially constant gauge isolates skew structure; comparison with RBF additionally changes capacity and spatial dependence. The +Corr. columns report this control only in the explicitly labeled Non-skew row.

GMM fitting and selection. Each family and Euler budget is fitted separately, so budget rows are not a convergence study of one fixed field. Each of three recompilations $r = 0, 1, 2$ uses three starts: zero and two nonzero starts calibrated to correction RMS 0.1 and 0.5. Training and validation use 2,048 and 4,096 standard-normal inputs, with seeds $20264110 + 100r$ and $20264111 + 100r$. Initialization draws use $20264112 + 100r +$ start ID. Calibration uses nine equally spaced times with 256 analytic marginal samples each (seed 20264001); coefficients are clipped to $[-2, 2]$, rescaled, and clipped again.

L-BFGS-B uses bounds $[-2, 2]$, `maxiter=150`, `maxfun=240`, function tolerance 10^{-10}, and gradient tolerance 10^{-7}. Both the Euler and same-field RK4-64 branches are differentiated. Validation uses RK4-128 and selects the lowest discrepancy among the three fits and the untouched zero candidate, breaking ties by lower start ID. Numerical discrepancy is normalized by its zero-gauge value. The training loss before normalization is half the per-coordinate MSE reported in Table 1. Quality and path diagnostics do not enter selection. All spatially constant and Non-skew fits meet tolerance; the RBF and Fourier fits stop at the iteration or function-call budget and remain eligible under the same selection rule.

Independent evaluation and reference checks. Endpoint MSE and target SW_2 use 32,768 source samples (seed 20265020), independent target samples (20265021), and 128 unit projections (20265023). MSE uses a same-field RK4-256 reference. Path evaluation uses 8,192 separate inputs (2026092001), 128 projections (2026092004), and shared-source RK4-256 trajectories at $t = 0.25, 0.5, 0.75, 1$. The table reports means and sample SDs over three selected fields; Base is shared. These SDs include fitting and selection variability and are not confidence intervals.

Frozen-field evaluation uses fused float64 analytic kernels without fast-math. The largest RK4-128/256 endpoint RMS difference relative to deployment RMS is 6.93×10^{-7}; DOP853 with relative/absolute tolerances $10^{-10}/10^{-12}$ agrees within RMS 6.83×10^{-9}. Path refinement agrees within RMS 1.09×10^{-7}. The Non-skew checks also pass. Nonzero empirical path distances can remain for exact gauges because different flow maps transport a finite common source set to different empirical measures of the same population law.

E.4 CIFAR-10 OPTIMIZATION AND SAMPLING

Backbone and numerical configurations. We use the unconditional EDM EMA checkpoint `edm-cifar10-32x32-uncond-vp.pkl` with the official implementation at commit `008a4e5`. Source states are $x_0 = 80z$, $z \sim \mathcal{N}(0, I_{3072})$. Backbone arithmetic is float32; states and correction parameters are float64, with TF32 disabled. There is no conditioning, guidance, backbone update, or use of training images in the correction loss. For NFE N, the positive noise nodes are

$$\begin{aligned} \sigma_j^{\rho} &= \left(80^{1/7} + \frac{j}{N-1}(0.002^{1/7} - 80^{1/7})\right)^7, \\ \sigma_j^{\log} &= 80\left(\frac{0.002}{80}\right)^{j/(N-1)}, \qquad j = 0, \ldots, N-1. \end{aligned} \tag{81}$$

The model's noise-rounding routine is applied, followed by a terminal zero. Euler and DPM++2M use one backbone call per positive node; the call at 0.002 supplies the common terminal denoiser. DPM++2M uses first-order startup and the history prediction $(1 + 1/(2r))D_i - D_{i-1}/(2r)$, where $r = \log(\sigma_{i-1}/\sigma_i)/\log(\sigma_i/\sigma_{i+1})$ and both predictions belong to the corrected field.

Training uses 17 positive-noise Heun intervals on a separate $\rho = 7$ grid, followed by terminal denoising, for 35 NFE. Main-table evaluation uses 64 intervals (129 NFE); the refinement check uses 128 (257 NFE). These references follow each candidate's own field. They share their grid across deployment configurations, including the logarithmic coarse grid, and are not formed by subdividing the coarse intervals.

Initialization and optimization. A normal parameter draw is calibrated to a 1% ratio of aggregate mean correction norm to aggregate mean base-velocity norm. Calibration uses 16 inputs (seed 2000000), transported by the base reference to $\sigma = 40, 10, 2, 0.4, 0.08, 0.01$. Euler configurations share the seed-101001 initial tensor. Adam uses learning rate 10^{-3}, default momentum and numerical-stability settings, batch size 16, and gradient clipping at 1.0, with no parameter projection. Training

noise seed 102001 is reset across configurations; normalization uses 512 inputs with seed 1000000. Euler runs for 100 updates and saves every ten. The DPM++2M main candidate comes from the same protocol with five calibrated draws (101001–101005), each trained for 30 updates. FID and marginal drift are not terms in the training loss.

Candidate selection and evaluation inputs. Development FID uses 5,000 source inputs $3000000+i$ and selects among saved correction checkpoints, including the calibrated random initialization. Zero gauge is reported separately as Base. The Euler $\rho = 7$ main candidates use seed 101001 at updates 10, 40, and 70 for NFE 5, 7, and 9; Euler/log uses update 10. DPM++2M/$\rho = 7$ at five NFE uses seed 101005 at update 30, selected by the lowest trained development FID among the comparable initialization-study candidates. All candidates are fixed before confirmation, and each main-table row uses one candidate for every metric.

Quality evaluation uses 50,000 inputs $5000000 + i$; numerical evaluation and path diagnostics use its first 512 and 1,024 inputs. Base and corrected arms are paired. These inputs are separate from training, calibration, normalization, and development, but the three confirmation metrics share inputs and are not independent replications. The Euler $\rho = 7$ quality measurements are reused for the same fixed candidates, without reselection after numerical or path evaluation.

E.5 Evaluation metrics

Numerical error and image quality. CIFAR-10 MSE averages squared pixel-coordinate differences between coarse and same-field reference outputs after the common terminal denoiser. Improvement is $100(E_{\text{base}} - E_\eta)/E_{\text{base}}$. Paired 95% intervals are $\bar{\Delta} \pm 1.96 s_\Delta/\sqrt{n}$, where Δ is the per-input Base-minus-corrected MSE. They measure input-sampling uncertainty conditional on a fixed checkpoint, not optimization-seed variability.

The EDM Inception detector yields 2,048-dimensional features after conversion to $\text{uint8}(\text{clip}(127.5x + 128, 0, 255))$. FID (Heusel et al., 2017) uses the official EDM CIFAR-10 reference statistics. KID (Bińkowski et al., 2018) uses 50,000 cached real-image features, kernel $(a^\top b/2048 + 1)^3$, and the mean of 50 unbiased estimates on 1,000 real and 1,000 generated features per subset (seed 43001, without replacement within a subset). KID is scaled by 10^3; subset-sampling errors are not independent training-run uncertainty.

Sliced Wasserstein distance and marginal normalization. For two equal-sized sample sets U, V, let $u_{\ell,(i)}, v_{\ell,(i)}$ be independently sorted projections onto fixed unit directions ϑ_ℓ. We use

$$\widehat{\text{SW}}_2(U, V) = \left[\frac{1}{Ln} \sum_{\ell=1}^{L} \sum_{i=1}^{n} (u_{\ell,(i)} - v_{\ell,(i)})^2 \right]^{1/2}. \tag{82}$$

For refined corrected and Base samples from the same source batch A, the main tables report

$$D_t = 100 \frac{\widehat{\text{SW}}_2(G_{A,t}, B_{A,t})}{s_{B,t}}, \qquad s^2_{B,t} = \frac{1}{L} \sum_{\ell=1}^{L} \widehat{\text{Var}}(\vartheta_\ell^\top B_{A,t}), \tag{83}$$

with variance denominator $n - 1$. This is a marginal distance relative to the Base projected scale, not a paired-particle displacement or a percentage of probability mass.

CIFAR-10 uses 1,024 inputs, 256 projections (seed 7000001), and 129-NFE reference states at indices $16, 32, 48, 64$, followed by the terminal output. The displayed time is

$$t(\sigma) = \frac{80^{1/7} - \sigma^{1/7}}{80^{1/7} - 0.002^{1/7}}. \tag{84}$$

The four positive nodes are approximately $17.5278, 2.51522, 0.169753, 0.002$; Output follows terminal denoising. This coordinate differs from the logarithmic basis coordinate τ. GMM uses physical interpolation time. All percentages are finite-sample diagnostics, not whole-path equivalence certificates.

Table 4: Numerical confirmation for the fixed CIFAR-10 main-table candidates. Intervals concern Base-minus-corrected per-coordinate MSE, conditional on the selected checkpoint.

Solver	Grid	NFE	Gain (%)	Paired 95% interval
Euler	$\rho = 7$	5	+0.78	$[0.0002239, 0.0008860]$
Euler	$\rho = 7$	7	+1.28	$[0.000167, 0.001016]$
Euler	$\rho = 7$	9	+1.08	$[-0.000188, 0.000900]$
DPM++2M	$\rho = 7$	5	+1.04	$[-0.0000585, 0.0005320]$
Euler	log	5	+1.18	$[0.000765, 0.001403]$

E.6 Skew construction for high-resolution generation

For the EDM2-XS latent state $x \in \mathbb{R}^{4\times64\times64}$, we use the model-derived carrier

$$r_\theta(x, \sigma) = \frac{D_\theta(x, \sigma) - x}{\sigma}. \tag{85}$$

Let e_i denote the ith standard basis vector of $\mathbb{R}^4$. Six trainable coefficients $\eta = (\eta_{ij})_{1\le i<j\le4}$ parameterize the channel matrix

$$K_\eta = \sum_{1\le i<j\le4} \eta_{ij}(e_i e_j^\top - e_j e_i^\top), \qquad K_\eta^\top = -K_\eta. \tag{86}$$

The same matrix acts at every spatial location. On the positive-noise interval $0.002 \le \sigma \le 80$, define

$$\tau(\sigma) = \frac{\log(80/\sigma)}{\log(80/0.002)}, \qquad A_\eta(\sigma) = 4\tau(\sigma)(1 - \tau(\sigma))K_\eta. \tag{87}$$

The time envelope makes the correction vanish at both ends of this interval while retaining skew symmetry. At latent location (h, w), the correction is

$$[g_\eta(x, \sigma)]_{:,h,w} = A_\eta(\sigma)[r_\theta(x, \sigma)]_{:,h,w}. \tag{88}$$

This spatially shared channel operator uses the existing denoiser prediction and adds no backbone evaluations.

F Supplementary Experimental Evidence

F.1 Numerical confirmation and reference sensitivity

Table 4 supplements the main-table means with paired uncertainty on the same 512 confirmation inputs and 129-NFE references. The intervals exclude zero for Euler at five and seven NFE and for Euler/log; nine-NFE Euler and five-NFE DPM++2M remain uncertain.

For the original two five-NFE $\rho = 7$ candidates, a 257-NFE check was prespecified when changing ref35 to ref129 changed the sign of the gain, the ref129 interval included zero, or matched-subset endpoint sampling-floor ratios changed by at least 10%. The DPM++2M candidate triggered the interval criterion; Euler triggered none. On the same 256-input DPM++2M subset, gains at reference NFE 35, 129, and 257 are 1.60843%, 0.84354%, and 0.75308%, with all three paired intervals including zero. Ref129-to-ref257 endpoint MSE is 1.890×10^{-6} for Base and 2.123×10^{-5} for the corrected field. This checks sensitivity without supplying a bound on the remaining reference error or replacing the primary 512-input results.

F.2 Sampling scale of the marginal diagnostics

For the two five-NFE $\rho = 7$ candidates, an additional diagnostic compares corrected samples with an independent Base anchor, using 1,024 samples and 256 directions. It divides this distance by the median of ten Base/Base distances sharing the same anchor. The source starts are $A = 5000000$, $B = 6000000$, and $C_j = 6000000 + 10000j$, $j = 1, \ldots, 10$; source noise is generated per input in float32 on CPU and cast to float64. Pixel/feature projection seeds are 7000001/7000002. Pixels use unclipped states; endpoint features use the quality-evaluation conversion.

Table 5: Sampler time in milliseconds per batch of 32, mean ± sample SD over 40 pairs. Overhead is the relative difference of mean corrected and Base times.

Solver	NFE	Base	+Gauge	Overhead (%)
Euler	5	157.355 ± 1.041	161.671 ± 4.407	+2.74
DPM++2M	5	161.770 ± 2.866	164.324 ± 3.725	+1.58

Across the five nodes, pixel ratios are 0.9866–1.0073 for Euler and 0.9853–1.0061 for DPM++2M. Endpoint feature ratios are 0.9935 and 0.9914. These ratios place the distances near the observed sampling scale, but are distinct from the shared-source percentages in Equation (83). The shared-anchor comparisons are not independent replications, and their median is neither an equivalence threshold nor evidence about unmeasured times.

F.3 Optimization and checkpoint selection

The quality-based selection rule is motivated by the separation between numerical error and generation quality. In the Euler five-NFE development run, update 100 lowers ref35 discrepancy by 5.58% relative to Base, but raises FID-5K from 73.666 to 109.578. In the original DPM++2M/log five-NFE development run, the best saved correction checkpoint is the random initialization: its discrepancy gain is 1.201%, yet FID-5K is 49.968 versus 47.273 for Base. Thus neither continued optimization nor applying a correction to a different sampler guarantees better quality.

The calibrated DPM++2M initialization study uses five draws, 30 updates, and fixed training and evaluation streams. At five NFE, mean FID-5K changes from 65.910 ± 12.483 at initialization to 38.159 ± 2.374 at update 30, versus 41.493 for Base; all five starts lower numerical discrepancy relative to Base. At seven NFE, mean FID changes from 17.102 ± 0.654 to 17.872 ± 0.254, although all starts lower discrepancy relative to initialization. Nine- and ten-NFE quality gains are mixed. Unscaled draws produce much larger discrepancy and severe image clipping, motivating the small initialization amplitude. These SDs describe initialization variability conditional on shared streams; the main table separately evaluates a fixed selected candidate on the 50K stream.

F.4 Inference timing at fixed NFE

Table 5 reports the two five-NFE $\rho = 7$ candidates used in the main comparison. Measurements use an idle RTX 4090, PyTorch 2.3.1/CUDA 12.1, the uncompiled sampler, and batches of 32, with two global warmup pairs, six configuration warmup pairs, and 40 measured pairs. Arms share pregenerated GPU noise and alternate execution order. Wall time includes rollout and trailing CUDA synchronization; model loading, noise generation, image conversion, and feature extraction are excluded. The result measures sampler overhead at fixed NFE, not end-to-end speedup at matched quality.

G Analytic Construction of Figure 1

Figure 1 is an analytic illustration, separate from the learned-model experiments and the radial construction in Proposition 3. For $Z \sim \mathcal{N}(0, I_2)$, let

$$X_t^\omega = H_t(R_{\omega t} Z), \qquad H_t(z) = \begin{pmatrix} \mu(t) + a(t) z_1 + c(t) \tanh(2 z_1) \\ b(t) z_2 \end{pmatrix}, \tag{89}$$

where R_θ is a planar rotation, $a(t) = 1 - 0.35t$, $b(t) = 1 - 0.4t$, $c(t) = 1.5t$, and $\mu(t) = 0.18 \sin(2\pi t)$. The maps H_t are invertible on $[0, 1]$, and rotations preserve the standard Gaussian. Thus the corresponding velocity fields for $\omega = 2.0$ and $\omega = 0.6$ induce exactly the same full two-dimensional marginal path. The background displays its first-coordinate marginal, rather than a density slice or a nontrivial one-dimensional gauge.

The corresponding velocity fields are

$$v_t^\omega(H_t(z)) = \partial_t H_t(z) + \omega D H_t(z) J z, \qquad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{90}$$

Taking $v_t = v_t^{2.0}$ and $u_t = v_t^{0.6}$ gives $g_t = u_t - v_t$, with

$$g_t(H_t(z)) = -1.4\, D H_t(z) J z, \qquad \operatorname{div}(p_t g_t) = 0. \tag{91}$$

The weighted-divergence identity follows from the shared marginal path and continuity equation.

Trajectories and endpoint errors. Both representatives use Euler on the grid $0, 0.2, \ldots, 1$. The highlighted initial states are $(-1, -0.5)$ and $(0.8, 0.7)$; 64 additional continuous trajectories use four radii and sixteen angular positions. Solid curves are continuous trajectories, and dashed curves with hollow nodes are Euler paths. At $t = 1$, small filled markers denote continuous endpoints. Brackets connect each Euler endpoint to its own field's continuous endpoint in the first coordinate, at the actual coordinate scale. Their lengths are $|x^\omega_{1,\mathrm{Euler}}(1) - x^\omega_{1,\mathrm{exact}}(1)|$, not full-state norms or uncertainty intervals. Endpoint Euler densities use 256×256 Gaussian quantile midpoint quadrature. The exact and numerical density curves have the same Gaussian smoothing bandwidth, 0.1, with no renormalization to the displayed window.

The fixed-time gauge field. Panel (d) uses the full two-dimensional density at $t_\star = 0.25$, whereas (a,b) display its first-coordinate marginal over time. The vertical time marker links these views and does not add an Euler evaluation. The velocity triangle is evaluated at $H_{t_\star}(0.9, 0.4)$, with a common vector scale of 1.2; the translated g arrow closes the identity $u = v + g$. The green arcs are fixed-time gauge streamlines, of the form $s \mapsto H_{t_\star}(R_{-1.4s} z)$, rather than trajectories through the time-dependent field. Three outer arcs use latent radius 1.95, and two inner arcs use radius 0.8, all oriented clockwise. The inner arcs cross density contours: weighted-flux conservation does not require g_t to be tangent to density level sets.

This analytic illustration demonstrates a numerical difference without asserting improvement for every particle or time, or a corresponding improvement in generation quality. Source data retain the full-state errors at all Euler nodes; these are not plotted as a separate error panel in the current figure.