BEYOND POPULATION LOSS: STATISTICAL COUPLING DESIGN FOR FLOW MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Changing a flow-matching coupling changes both the regression problem and the statistical cost of learning it. We show that these effects can be separated exactly: for Gaussian endpoints and linear spatial models, cross-covariance fixes the entire population objective, but higher-order dependence still changes finite-sample generation risk. Smooth non-Gaussian couplings strictly outperform the entire jointly Gaussian family at first order. The separation holds for every fixed Euler budget in a constant-field model and, for every scale $s > 1$, in a time-dependent class with a realizable Gaussian reference and exact integration. The latter has an explicit relative improvement exceeding $27/256$. On the constructive side, covariance-constrained quartic transport characterizes the global first-order coupling oracle, with $O(n^{-2})$ regret under stabilized least squares. Finite cell-transport programs retain the continuous Gaussian marginals exactly, converge to the oracle, and admit smooth full-support implementations. Executed fresh-training experiments connect the optimized designs to terminal-risk improvements; at 64 cells, the smooth designs lower first-order coefficients by 44.9% and 65.5% in the two controlled protocols. Matrix reachability certificates and fixed-feature residual scores provide complementary structural results.

1 INTRODUCTION

Flow matching learns a velocity field by regressing on velocities of paths joining source and target samples (Lipman et al., 2023; Tong et al., 2024). The endpoint coupling is a design variable: minibatch transport can straighten paths and reduce gradient variance (Pooladian et al., 2023), while model-aligned coupling uses prediction error rather than geometry alone (Lin et al., 2025). We ask a different question: *can couplings be indistinguishable to the entire population training objective, yet have strictly different statistical costs of generation?*

The answer is yes, even after fixing the endpoint laws, time sampling, model class, training rule, and sampler. Cross-covariance determines every population loss value in a linear spatial model, but not the fourth-order residual moments governing the uncertainty of its fitted parameters. Propagating that uncertainty through the sampler gives a coupling-dependent terminal risk. Consequently, equality of the loss landscape, including all its derivatives, does not imply equal finite-sample generation risk.

Our main results turn this separation into a constructive theory. First, smooth non-Gaussian couplings strictly outperform the *entire* jointly Gaussian family at first order. This holds for every fixed Euler budget in a constant-field model. More importantly, it also holds at *every* Gaussian scale $s > 1$ in a time-dependent class whose Gaussian reference is exactly realizable and whose ODE is integrated exactly: the explicit relative improvement exceeds $27/256$. Second, the global coupling oracle reduces to covariance-constrained quartic transport. A uniform fourth-order expansion gives an $O(n^{-2})$ regret bound, allowing the competing coupling to depend on n. Third, finite cell-transport programs preserve the *continuous Gaussian marginals exactly* and converge to that oracle. Smoothing retains an explicit fraction of the gain, and fresh-training experiments demonstrate the resulting terminal-risk improvements.

Relation to prior work. Model awareness is an established coupling-design principle: MAC aligns couplings with prediction error (Lin et al., 2025), and normalizing-flow distillation supplies another model-informed construction (Berthelot et al., 2026). Source-distribution learning analyzes intrinsic

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.

supervision variance (Kim et al., 2026b); closed-form FM studies stochastic targets (Bertrand et al., 2025). Quantile coupling removes variance in selected projections while quantifying the discrepancy between its training source and sampling prior (Kim et al., 2026a). Here both endpoint laws are held fixed. Covariance preconditioning (Ahamed et al., 2026) and functional FM consistency bounds separating statistical and numerical errors (Shikhman, 2026) address complementary questions.

We use classical normal equations, sandwich covariance and the delta method (White, 1980; van der Vaart, 1998), Gaussian Wasserstein geometry (Bhatia et al., 2019), bounded density perturbations (Sarmanov, 1966), and linearly constrained transport (Zaev, 2014). The contribution is their coupling-specific consequence: *population-sufficient Gaussian couplings can be statistically suboptimal, and optimizing within a population equivalence class yields a constructive terminal-risk oracle*. Exact matrix reachability and population loss-ranking counterexamples are supporting results in Appendices F–G.

2 Population equivalence and statistical risk

Let $\pi \in \Pi(\mu, \nu)$ have fixed centered endpoint marginals. Throughout,

$$X_t = (1-t)X_0 + tX_1, \qquad U = X_1 - X_0, \qquad t \sim \mathrm{Unif}[0,1], \tag{1}$$

with time independent of the endpoints. The population loss is $\mathcal{L}_\pi(v) = \mathbb{E}\|v(X_t,t) - U\|^2$. Its usual decomposition, with $\bar{u}_\pi = \mathbb{E}[U \mid X_t, t]$, is

$$\mathcal{L}_\pi(v) = \underbrace{\mathbb{E}\|v - \bar{u}_\pi\|^2}_{\mathcal{A}_\pi(v)} + \underbrace{\mathbb{E}\,\mathrm{tr}\,\mathrm{Cov}(U \mid X_t, t)}_{\mathcal{N}_\pi}. \tag{2}$$

We keep the model and sampler fixed in each comparison. $K \in \mathbb{N}$ denotes K equal Euler steps; $K = \infty$ denotes exact integration, not a growing-budget limit.

Theorem 1 (Population equivalence for linear spatial models)**.** *Let $\mu = \mathcal{N}(0, \Sigma_0)$, $\nu = \mathcal{N}(0, \Sigma_1)$, with $\Sigma_i \succ 0$. In any fixed class $v_\theta(x,t) = A_\theta(t)x$ with finite losses, $C = \mathbb{E}[X_1 X_0^\top]$ determines the entire population objective. Every feasible C is realized by a jointly Gaussian coupling. Replacing a coupling by that Gaussian representative leaves population training and generation unchanged, under the same deterministic optimizer-selection rule and sampler.*

Proof. At each time, the moments determining the quadratic loss are

$$\mathbb{E}X_t X_t^\top = (1-t)^2\Sigma_0 + t^2\Sigma_1 + t(1-t)(C + C^\top),$$
$$\mathbb{E}UX_t^\top = (1-t)(C - \Sigma_0) + t(\Sigma_1 - C^\top).$$

Also $\mathbb{E}UU^\top = \Sigma_0 + \Sigma_1 - C - C^\top$. Every positive semidefinite block covariance with these marginal blocks defines a Gaussian representative. Identical selected fields and identical samplers give identical output laws. □

For constant matrix fields, Appendix F further gives an exact semidefinite test for reaching a prescribed velocity by coupling design. Population loss, even after subtracting $\mathcal{N}_\pi$, can reverse terminal-error rankings (Appendix G). Our main question is stronger: what can change when the *entire* population objective stays fixed?

2.1 A uniform expansion and an observable residual score

Consider a fixed feature model $v_\theta(x,t) = \Phi(x,t)\theta$, $\Phi \in \mathbb{R}^{d\times p}$, with fixed p. For a chosen coupling, put

$$G = \mathbb{E}\Phi^\top\Phi, \quad \theta = G^{-1}\mathbb{E}\Phi^\top U, \quad r = U - \Phi\theta, \quad q = \Phi^\top r, \quad \Omega = \mathbb{E}qq^\top, \quad \Xi = G^{-1}\Omega G^{-1}. \tag{3}$$

Train by ordinary least squares on n independent triples, stabilized by projection onto a fixed compact convex parameter set. Denote by $F_K(\theta)$ the terminal distribution risk of the specified sampler, and by $\mathfrak{R}_{n,K}(\pi) = \mathbb{E}_{\mathrm{train}}F_K(\widehat{\theta})$ its expected trained risk.

Proposition 2 (Uniform risk expansion). *Suppose $G \succeq g_0 I$, population minimizers remain a positive distance inside the stabilizing set, and $\Phi^\top \Phi$ and q have uniformly bounded eighth moments. Assume F_K is bounded on that set and has uniformly bounded derivatives through order three near the population minimizers. Then, uniformly over the coupling family,*

$$\mathfrak{R}_{n,K}(\pi) = F_K(\theta) + \frac{1}{n}\left[\nabla F_K(\theta)^\top \beta + \frac{1}{2}\operatorname{tr}(\nabla^2 F_K(\theta)\Xi)\right] + O(n^{-3/2}), \tag{4}$$

where

$$\beta = -G^{-1}\mathbb{E}\left[(\Phi^\top\Phi - G)G^{-1}q\right]. \tag{5}$$

If the fourth derivatives of F_K are also uniformly bounded near the minimizers, the remainder improves to $O(n^{-2})$.

Appendix B proves both uniform remainders. The $O(n^{-2})$ refinement retains cancellations of centered cubic sample moments; an absolute third-moment bound alone only gives $O(n^{-3/2})$. Uniformity is essential for optimizing over couplings that can depend on n.

At a population-correct parameter $\theta_\star$, $F_K(\theta_\star) = 0$, $\nabla F_K(\theta_\star) = 0$, and $Q_K = \frac{1}{2}\nabla^2 F_K(\theta_\star) \succeq 0$. Then

$$\mathfrak{R}_{n,K}(\pi) = \frac{\mathcal{S}_K(\pi)}{n} + O(n^{-3/2}), \qquad \mathcal{S}_K(\pi) = \mathbb{E}\|Q_K^{1/2} G^{-1}\Phi^\top (U - v_{\theta_\star})\|^2. \tag{6}$$

This residual score propagates uncertainty through the learner and the sampler. It uses G, the population parameter, and terminal curvature Q_K, rather than the unknown conditional mean or irreducible noise. These quantities must be known or estimated. The remainder in (6) is $O(n^{-2})$ under the fourth-derivative condition, satisfied by both scalar protocols below.

Why supervision variance still matters. Conditioning gives

$$\Omega = \mathbb{E}\Phi^\top \operatorname{Cov}(U \mid X_t, t)\Phi + \mathbb{E}\Phi^\top(\bar{u}_\pi - v_{\theta_\star})(\bar{u}_\pi - v_{\theta_\star})^\top \Phi. \tag{7}$$

Both terms contribute nonnegatively to (6). Lower supervision noise is beneficial when measured in the parameter directions and terminal geometry that matter. Its unweighted trace in (2) is not that quantity. Away from a population-correct endpoint, the bias and curvature terms in (4) cannot in general be dropped.

3 A COUPLING ORACLE REDUCES TO QUARTIC TRANSPORT

We now give a positive selection result over *all* couplings, not just a restricted candidate family. Fix $\mu = \mathcal{N}(0,1)$, $\nu = \mathcal{N}(0, s^2)$, $s > 1$, and the scalar class $v_a(x,t) = ax$. For covariance $c = \mathbb{E}X_0X_1$,

$$G(c) = \frac{1 + s^2 + c}{3}, \quad a(c) = \frac{s^2 - 1}{2G(c)}, \quad L^*(c) = 1 + s^2 - 2c - \frac{(s^2-1)^2}{4G(c)}. \tag{8}$$

Here $R_K(a) = (1 + a/K)^K$ and $R_\infty(a) = e^a$ are the generated scales.

Use the same estimator for every coupling:

$$\widehat{a} = \operatorname{clip}_{[0,3]}\left(\frac{\sum_{i=1}^n X_{t_i}U_i}{\sum_{i=1}^n X_{t_i}^2}\right), \qquad \mathfrak{R}_{n,K}(\pi) = \mathbb{E}\big(R_K(\widehat{a}) - s\big)^2. \tag{9}$$

A zero denominator is assigned slope zero. It occurs with probability zero here. Define

$$\begin{aligned} a_\star = \lambda_K(s), \qquad \lambda_K(s) = K(s^{1/K} - 1), \quad \lambda_\infty(s) = \log s, \\ c_\star = \frac{3(s^2-1)}{2a_\star} - (1 + s^2), \qquad G_\star = G(c_\star). \end{aligned} \tag{10}$$

The condition $|c_\star| < s$ says that an interior covariance gives a population-correct endpoint for this sampler. Put

$$J_a(x,y) = \int_0^1 \left[((1-t)x + ty)\{y - x - a((1-t)x + ty)\}\right]^2 \mathrm{d}t. \tag{11}$$

This is an explicit nonnegative quartic polynomial; its coefficients are in Appendix C.

Theorem 3 (Global first-order coupling oracle)**.** *Fix $s > 1$ and $K \in \mathbb{N} \cup \{\infty\}$ with $|c_\star| < s$. The problem*

$$S_\star = \frac{R'_K(a_\star)^2}{G_\star^2} \min_{\substack{\pi \in \Pi(\mu,\nu) \\ \mathbb{E}_\pi X_0 X_1 = c_\star}} \mathbb{E}_\pi J_{a_\star}(X_0, X_1) \tag{12}$$

has an optimizer $\pi_\star$ and $S_\star > 0$. Moreover,

$$\lim_{n\to\infty} n \inf_{\pi\in\Pi(\mu,\nu)} \mathfrak{R}_{n,K}(\pi) = S_\star, \qquad 0 \le \mathfrak{R}_{n,K}(\pi_\star) - \inf_{\pi\in\Pi(\mu,\nu)} \mathfrak{R}_{n,K}(\pi) = O(n^{-2}). \tag{13}$$

The infimum permits an n-dependent coupling, but the coupling is fixed before drawing fresh independent training triples. Constants depend on fixed s, K, not on π or n.

Proof sketch. Fixed Gaussian marginals give uniform polynomial moments and weak compactness, so the constrained quartic minimum is attained. The fourth-order refinement of Proposition 2 applies uniformly to *all* such couplings: the population slopes lie uniformly inside $[0, 3]$ and G has a positive lower bound. Write the expansion as $F(c) + T(\pi)/n + O(n^{-2})$. The unique zero of F is $c_\star$, and $F(c) \ge \kappa(c - c_\star)^2$ for some $\kappa > 0$. Mixing any coupling with a positively or negatively perfectly correlated coupling restores covariance $c_\star$ at a cost proportional to $|c - c_\star|$. Bounded fourth moments and smooth coefficients then give $T(\pi) \ge S_\star - L|c - c_\star|$. Completing the square yields

$$\mathfrak{R}_{n,K}(\pi) \ge \frac{S_\star}{n} - \frac{L^2}{4\kappa n^2} - \frac{C}{n^2}. \tag{14}$$

At $\pi_\star$, the risk is $S_\star/n + O(n^{-2})$. Positivity follows because $J_{a_\star}(x, y) = 0$ only at $(0, 0)$. Appendix D supplies the compactness, uniformity, and mixture details. □

This is an oracle over couplings for the specified learning protocol, not a minimax lower bound over all estimators. Section 5 turns the continuous optimization into convergent finite programs without replacing the Gaussian endpoint laws by atoms.

4 Gaussian couplings are not statistically sufficient

Let π_ρ^G be the jointly Gaussian coupling with correlation $\rho \in [-1, 1]$ and the stated marginals. Let φ denote the standard Gaussian density. Fix $\tau = 3/4$ and define the bounded, centered, even function

$$h(z) = 2e^{-3z^2/2} - 1, \qquad p_\tau^{(s)}(x, y) = \varphi(x)\frac{\varphi(y/s)}{s}[1 - \tau h(x)h(y/s)]. \tag{15}$$

This Sarmanov-type construction preserves both Gaussian marginals, has covariance zero, and has smooth density at least one quarter of the independent density. Its population objective is therefore identical to that of π_0^G.

4.1 Constant fields and fixed sampler budgets

Theorem 4 (Strict separation for every fixed sampler budget)**.** *For each $K \in \mathbb{N} \cup \{\infty\}$, there is a unique $s_K > 1$ such that*

$$\frac{3}{2}\tanh(\log s_K) = \lambda_K(s_K). \tag{16}$$

Fix these endpoints and the estimator (9)*. Put $a_0 = \lambda_K(s_K)$ and $G_0 = (1 + s_K^2)/3$. The smooth coupling* (15) *and independent Gaussian coupling have the same entire population objective and zero population terminal error, but*

$$\lim_{n\to\infty} n\left[\inf_{\rho\in[-1,1]} \mathfrak{R}_{n,K}(\pi_\rho^G) - \mathfrak{R}_{n,K}(p_\tau^{(s_K)})\right] = D_K := \frac{9\tau R'_K(a_0)^2 a_0^2 s_K^2}{80 G_0^2} > 0. \tag{17}$$

In particular, $s_1 = (3 + \sqrt{17})/4$ and $D_1 = 27/320$.

Proof sketch. Writing $z = \log s$, uniqueness follows because $\tanh(z)/z$ decreases strictly while $(e^{z/K} - 1)/(z/K)$ increases; for $K = \infty$ the latter is one. Gaussian integration gives $\mathbb{E}h(Z) = 0$ and $\mathbb{E}Z^2 h(Z) = -3/4$. Relative to independence, the density perturbation changes only $\mathbb{E}[X_0^2 X_1^2]$ among the relevant fourth moments, reducing it by $9\tau s_K^2/16$. The coefficient of x^2y^2 in J_{a_0} is $a_0^2/5$, giving D_K. Finally, any asymptotically optimal Gaussian correlation must approach zero, the unique population-correct covariance; uniform risk expansion prevents n-dependent correlations from improving its order-$1/n$ constant. See Appendix E. □

For one step the constants are particularly transparent:

$$\lim_n n\mathfrak{R}_{n,1}(\pi_0^G) = \frac{17}{10} + \frac{\sqrt{17}}{5}, \qquad \lim_n n\mathfrak{R}_{n,1}(p_{3/4}^{(s_1)}) = \frac{17}{10} + \frac{\sqrt{17}}{5} - \frac{27}{320}. \tag{18}$$

This is a statistical separation despite equality of all population losses, gradients, and Hessians, not just a reversal of two differently normalized objectives.

Corollary 5 (An open family, not an isolated scale)**.** *For any fixed K, suppose $r = |c_\star|/s < 1/4$. Let*

$$\pi_s = (1 - 2r)p_{3/4}^{(s)} + 2r\,\pi^G_{\mathrm{sign}(c_\star)/2}. \tag{19}$$

At $r = 0$ the second component is absent. This coupling has covariance $c_\star$, density at least one eighth of the independent density, and strictly beats the entire Gaussian family at order $1/n$, with gap

$$D_K(s) = \frac{R_K'(a_\star)^2 a_\star^2 s^2}{5G_\star^2}(1 - 2r)\left(\frac{27}{64} - r\right) > 0. \tag{20}$$

For $K = 1$, the condition holds throughout $(5 + \sqrt{57})/8 < s < 2$.

Each comparison fixes its endpoints and complete training protocol. The scales s_K can differ across budgets. The following result removes the need to tune the endpoint scale and uses a time-dependent field with a realizable Gaussian reference.

4.2 A REALIZABLE TIME-DEPENDENT REFERENCE, WITH EXACT INTEGRATION

Fix *any* $s > 1$ and use the fixed class

$$q_s(t) = (1 - t)^2 + s^2 t^2, \qquad b_s(t) = \frac{(1 + s^2)t - 1}{q_s(t)}, \qquad v_\theta(x, t) = \theta b_s(t)x. \tag{21}$$

Train θ by stabilized least squares with feature $b_s(t)X_t$, using the fixed interval specified in Appendix J. Exact integration generates scale s^θ, so write $\widetilde{\mathfrak{R}}_n(\pi) = \mathbb{E}(s^{\widehat{\theta}} - s)^2$.

Theorem 6 (Separation with a realizable Gaussian reference)**.** *For every $s > 1$, the independent Gaussian coupling has conditional mean $b_s(t)x$ and population optimizer $\theta = 1$. The smooth coupling $p_{3/4}^{(s)}$ has the same entire population objective and the same exact target endpoint. Nevertheless,*

$$\inf_{\rho\in[-1,1]} \widetilde{\mathfrak{R}}_n(\pi_\rho^G) - \widetilde{\mathfrak{R}}_n(p_{3/4}^{(s)}) = \frac{\Delta_s}{n} + O(n^{-2}), \qquad \frac{\Delta_s}{S_s^G} = \frac{27}{256}\left(1 + \frac{1}{I_s}\right) > \frac{27}{256}, \tag{22}$$

where the Gaussian oracle coefficient is

$$G_s = 1 + s^2 - \frac{\pi s}{2}, \qquad I_s = \frac{\pi(1 + s^2)}{4s} - 1, \qquad S_s^G = \frac{s^4(\log s)^2}{G_s^2} I_s. \tag{23}$$

Proof sketch. The conditional-mean formula and $\int_0^1 b_s = \log s$ give realizability and the endpoint. The population parameter is a strictly decreasing fractional-linear function of covariance, with its unique calibrated value at $c = 0$. The coefficient of x^2y^2 in $\int_0^1 [b_s(t)X_t\{U - b_s(t)X_t\}]^2 \mathrm{d}t$ is $\pi(1 + s^2)/(16s) > 0$. The density perturbation reduces only that mixed fourth moment, by $27s^2/64$. Proposition 2 and the same Gaussian-oracle argument as above give (22). Appendix J supplies the integrals, uniformity, and full proof. □

At $s = 2$, the first-order coefficient falls from 2.144563 to 1.683624, a 21.5% reduction. The Gaussian reference has zero pathwise approximation error and no integration error; the non-Gaussian alternative need not have a realizable conditional mean. Thus reference-model misspecification and numerical integration are not prerequisites for the statistical separation. The same quartic-oracle reduction holds for this temporal model, and coordinatewise tensorization gives gap $d\Delta_s$ against arbitrary jointly Gaussian couplings in the diagonal d-parameter class (Appendix J).

5 Finite transport with exact continuous marginals

A point-mass discretization changes the endpoint laws and their fourth moments. Instead, partition each Gaussian marginal into m equal-probability quantile cells and retain its full conditional law inside every cell. Write these laws as μ_i, ν_j, their means as u_i, v_j, and let $d(x, y)$ be the calibrated quartic score from either scalar protocol. Define

$$\pi_P = \sum_{ij} P_{ij}\mu_i \otimes \nu_j, \qquad S_m = \min_{\substack{P\geq 0,\, P\mathbf{1}=P^\top\mathbf{1}=\mathbf{1}/m \\ \sum_{ij} P_{ij}u_i v_j = c_\star}} \sum_{ij} P_{ij}\, \mathbb{E}_{\mu_i\otimes\nu_j} d(X_0, X_1). \tag{24}$$

Every feasible π_P has the *original continuous Gaussian marginals* and covariance $c_\star$. The costs use analytic conditional Gaussian moments, and sampling draws a cell pair followed by independent conditional draws.

Theorem 7 (Constructive approximation of the coupling oracle)**.** *For either scalar protocol with interior calibrated covariance,* (24) *is feasible for all sufficiently large m. Let π_m attain it, and let $S_\star$ be the corresponding continuous quartic optimum. Along dyadic m, with $Z \sim \mathcal{N}(0,1)$ and $I_m(Z)$ its quantile-cell index, put*

$$\delta_m = (1+s)\|Z - \mathbb{E}[Z \mid I_m(Z)]\|_{L^4} \longrightarrow 0.$$

Then S_m is nonincreasing and $0 \leq S_m - S_\star \leq C\delta_m$. For fresh training, writing $\mathfrak{R}_n$ for the relevant protocol,

$$0 \leq \mathfrak{R}_n(\pi_m) - \inf_{\pi\in\Pi(\mu,\nu)} \mathfrak{R}_n(\pi) \leq \frac{C\delta_m}{n} + \frac{C'}{n^2}. \tag{25}$$

The constants are independent of m, n. A feasible η_m-suboptimal LP solution adds η_m/n to the bound.

Proof sketch. Lift the cell masses of a continuous optimizer by independent conditional draws. The L^4 displacement bounds both covariance error and quartic-cost error by $C\delta_m$. A small mixture with the diagonal or antidiagonal cell coupling repairs covariance exactly. Their covariances tend to s and $-s$, so the interior hypothesis bounds the repair weight. Nested partitions contain the preceding feasible family. The uniform $O(n^{-2})$ risk expansion completes the argument; Appendix K gives details. □

At $c_\star = 0$, independent Gaussian smoothing makes the lift smooth and everywhere positive without changing either marginal or covariance:

$$X^\epsilon = \sqrt{1-\epsilon}X + \sqrt{\epsilon}Z_0, \qquad Y^\epsilon = \sqrt{1-\epsilon}Y + s\sqrt{\epsilon}Z_1,$$

where Z_0, Z_1 are independent standard Gaussians, also independent of (X, Y). Its score satisfies the exact identity

$$S(\pi_{m,\epsilon}) = S_G + (1-\epsilon)^2(S_m - S_G). \tag{26}$$

Thus $\epsilon = 0.01$ retains 98.01% of the optimized first-order improvement while giving full support. Here S_G is the independent Gaussian score for the same model and endpoints.

6 Computable design and fresh-training verification

Beyond the two scalar protocols. For fixed features, fixing $\theta_\star, G_\star, Q_K$ makes (6) linear in the coupling under

$$\mathbb{E}_\pi \Phi^\top \Phi = G_\star, \qquad \mathbb{E}_\pi \Phi^\top U = G_\star\theta_\star. \tag{27}$$

Appendix H gives the finite transport dual and a fresh-training candidate-selection guarantee. If G is free, the inverse makes the design nonlinear; estimated moments and terminal curvature contribute additional error.

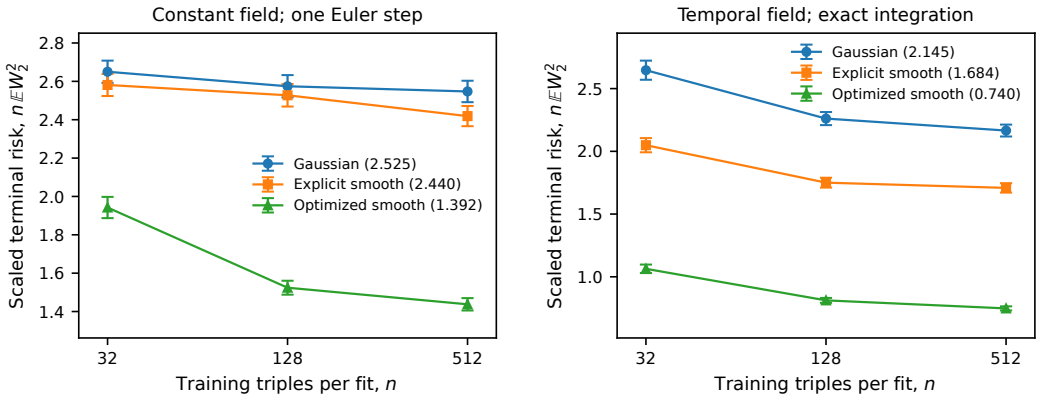

Figure 1: Fresh training with exact Gaussian marginals. Points are n times mean terminal risk, with approximate 95% Monte Carlo intervals; legends report independently computed first-order coefficients. Left: constant velocity and one Euler step. Right: a realizable time-dependent Gaussian reference and exact integration. "Optimized smooth" uses the lifted cell LP with $m = 64$ and Gaussian smoothing $\epsilon = 0.01$. Models and endpoints differ *between* panels; the Gaussian is population-calibrated, not a finite-n oracle.

From an optimized coupling to a trained generator. We solve (24) with $m = 64$ and then smooth with $\epsilon = 0.01$, before drawing any training data. We compare this *optimized smooth* coupling with the calibrated Gaussian and the explicit smooth density (15). Protocol one uses the constant model, $K = 1$, and $s = s_1$; protocol two uses (21), $s = 2$, and exact integration. Each configuration uses 16,384 independent fits for $n \in \{32, 128, 512\}$: 294,912 fits and 66,060,288 training triples in total. The evaluation is the exact terminal distribution error determined by the fitted parameter.

The optimized smooth first-order coefficients are 1.392004 versus 2.524621 for the constant model, and 0.740389 versus 2.144563 for the temporal model: reductions of 44.9% and 65.5% under the *same continuous Gaussian marginals*. At $n = 512$, the observed scaled risks are 1.4377 ± 0.0164 versus 2.5473 ± 0.0285, and 0.7479 ± 0.0084 versus 2.1656 ± 0.0242, respectively (one Monte Carlo standard error). The optimized constant-field design has noticeable finite-n bias at $n = 32$, showing why a first-order coefficient is not an exact small-sample risk.

Certificates and retained checks. Cell programs with $m = 16, 32, 64, 128$ have maximum primal/dual feasibility residual below 4×10^{-15} in the reported runs. Their costs correspond to continuous laws; Theorem 7, rather than visual grid agreement, gives convergence to the continuous optimum. These numerical residuals are floating-point diagnostics. Exact symbolic checks verify the temporal quartic coefficients, smoothing identity, original $27/320$ gap, and noncommuting matrix construction. Appendix L reports the new complete tables and protocol; Appendix I retains the original $K = 1, 8$ experiments and point-grid diagnostics, separately labeled.

7 Limitations and Conclusion

The theory fixes the model, estimator, and sampler. Its statistical oracles optimize couplings, not all possible learning algorithms. The temporal profile is constructed from a known Gaussian reference; the non-Gaussian alternative preserves its population objective but need not preserve pathwise realizability. Coordinatewise extensions do not constitute an unrestricted neural or full-matrix statistical theorem. Couplings are fixed before independent training draws, rather than adaptively matched on the training batch, and accessible terminal curvature is required for general feature scores.

Within these protocols, the distinction is exact: population equivalence fixes the learned field but not its statistical uncertainty. Smooth couplings can beat the full Gaussian family even with a realizable reference and exact integration. A quartic transport objective, a uniform $O(n^{-2})$ oracle bound, and exact-marginal finite approximations turn that distinction into a constructive design principle.

REPRODUCIBILITY STATEMENT

All proofs are included in the appendices. The accompanying archive contains symbolic checks, LP certificates, fixed seeds, original and new replicate-level losses, and figure-generating code. Appendix L specifies the new exact-marginal training experiments, and Appendix I documents the retained original checks. The Gaussian experimental baselines are population-calibrated; the full-family oracle comparisons are proved analytically.

AI USE STATEMENT

A generative AI assistant was used substantially in developing the mathematical formulations, proposing and revising proofs, writing and auditing code, running synthetic numerical checks, reviewing related-work sources, and drafting and editing the manuscript. This assistance was not limited to language polishing. The mathematical claims, generated code, numerical outputs, and citations were checked as described in the appendices; these checks do not constitute formal proof verification or independent human peer review. Responsibility for the correctness, originality, attribution, and submission of this work remains with the submitting human authors.

REFERENCES

Shadab Ahamed, Eshed Gal, Md Shahriar Rahim Siddiqui, Simon Ghyselincks, Moshe Eliasof, and Eldad Haber. Preconditioned flow matching. *arXiv preprint arXiv:2603.02337*, 2026. URL https://arxiv.org/abs/2603.02337.

David Berthelot, Tianrong Chen, Jiatao Gu, Marco Cuturi, Laurent Dinh, Bhavik Chandna, Michal Klein, Josh Susskind, and Shuangfei Zhai. The coupling within: Flow matching via distilled normalizing flows. *arXiv preprint arXiv:2603.09014*, 2026. URL https://arxiv.org/abs/2603.09014.

Quentin Bertrand, Anne Gagneux, Mathurin Massias, and Rémi Emonet. On the closed-form of flow matching: Generalization does not arise from target stochasticity. In *Advances in Neural Information Processing Systems*, 2025. URL https://arxiv.org/abs/2506.03719.

Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019. doi: 10.1016/j.exmath.2018.01.002. URL https://arxiv.org/abs/1712.01504.

Jin-Young Kim, So-Yoon Cho, and Hyun-Gyoon Kim. One-sided quantile coupling for flow matching. *arXiv preprint arXiv:2608.00978*, 2026a. URL https://arxiv.org/abs/2608.00978.

Junwan Kim, Jiho Park, Seonghu Jeon, and Seungryong Kim. Better source, better flow: Learning condition-dependent source distribution for flow matching. *arXiv preprint arXiv:2602.05951*, 2026b. URL https://arxiv.org/abs/2602.05951.

Yexiong Lin, Yu Yao, and Tongliang Liu. Beyond optimal transport: Model-aligned coupling for flow matching. *arXiv preprint arXiv:2505.23346*, 2025. URL https://arxiv.org/abs/2505.23346.

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://arxiv.org/abs/2210.02747.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample flow matching: Straightening flows with minibatch couplings. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 2023. URL https://proceedings.mlr.press/v202/pooladian23a.html.

O. V. Sarmanov. Generalized normal correlation and two-dimensional Fréchet classes. *Doklady Akademii Nauk SSSR*, 168(1):32–35, 1966. In Russian.

Lennon J. Shikhman. Discretization and statistical consistency of functional flow matching. *arXiv preprint arXiv:2608.04531*, 2026. URL https://arxiv.org/abs/2608.04531.

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.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980. doi: 10.2307/1912934.

Danila Zaev. On the Monge–Kantorovich problem with additional linear constraints. *arXiv preprint arXiv:1404.4962*, 2014. URL https://arxiv.org/abs/1404.4962.

A NOTATION, SCOPE, AND PROOF DEPENDENCIES

All expectations in the population formulas include an independent uniform time unless otherwise indicated. A coupling is a joint endpoint law with the two displayed marginals fixed. It is selected before drawing the independent training triples; it is not an adaptive matching computed on those same triples. The quantity $\mathfrak{R}_{n,K}(\pi)$ averages the terminal distribution error over the random training set. It does not include sampling error from an empirical estimate of the generated distribution: in the Gaussian scalar experiments that terminal error is known exactly from the fitted slope.

The population equivalence theorem includes time-dependent linear spatial classes. Statistical results concern two fixed scalar protocols: constant fields with a fixed Euler or exact sampler, and the temporal profile (21) with exact integration. The fixed-feature expansion is conditional on its regularity assumptions. The cell LPs in Section 5 preserve continuous Gaussian marginals and have a proved approximation guarantee. The original point-grid LPs in Appendix I instead optimize discrete laws and are retained as separate diagnostics.

The proof dependencies are as follows. Direct endpoint moments give Theorem 9 and the scalar identities. The uniform least-squares expansion below and the quartic identity give Theorem 3. The same expansion, Gaussian moment identities, and a bounded density perturbation give Theorems 4 and 6, and Corollary 5. Conditional Gaussian lifts and covariance repair give Theorem 7. All constants in the $O(\cdot)$ statistical statements may depend on the fixed sampler, endpoint scale, feature dimension, clipping neighborhood, and moment bounds. They are uniform in the coupling family being optimized. We do not claim uniformity as K or feature dimension grows with n.

B UNIFORM EXPANSION FOR STABILIZED LEAST SQUARES

We give the details needed for optimization over couplings, including couplings that change with sample size. This specializes classical least-squares and delta-method calculations (White, 1980; van der Vaart, 1998) rather than asserting a new general asymptotic method.

B.1 A PRECISE UNIFORM ASSUMPTION SET

Let $\mathcal{P}$ be a family of laws for (Φ, U), with fixed feature dimension p. For each law, write

$$A = \Phi^\top \Phi, \quad G = \mathbb{E}A, \quad \theta = G^{-1}\mathbb{E}\Phi^\top U, \quad q = \Phi^\top(U - \Phi\theta).$$

Assume the following constants can be chosen uniformly over $\mathcal{P}$: $G \succeq g_0 I$ with $g_0 > 0$; a compact convex stabilizing set $\Theta \subset \mathbb{R}^p$ containing the Euclidean ball $B(\theta, \delta)$ for some $\delta > 0$; and

$$\mathbb{E}\|A\|^8 + \mathbb{E}\|q\|^8 \le M_8 < \infty. \tag{28}$$

All matrix norms can be taken as Frobenius norms because p is fixed. The estimator is the projection onto Θ of $\widehat{G}^{-1}\overline{\Phi^\top U}$ when $\widehat{G}$ is invertible. On singular samples choose any fixed element of Θ. Assume F_K is bounded on Θ and has bounded derivatives through order three on a fixed neighborhood of all population minimizers. No smoothness of the projection map at the boundary is needed.

Lemma 8 (Uniform sample-mean moments)**.** *Under* (28)*, for centered copies Z_i of either $A - G$ or q, and $2 \le j \le 8$,*

$$\mathbb{E}\left\| n^{-1}\sum_{i=1}^{n} Z_i \right\|^j \le C_j n^{-j/2}.$$

The constants are uniform over $\mathcal{P}$.

Proof. For each coordinate and even $j \in \{2, 4, 6, 8\}$, expand the jth power of the sum. By centering and independence, a nonzero term cannot have an index occurring only once. Consequently it has at most $j/2$ distinct sample indices, giving at most $C_j n^{j/2}$ terms. Hölder's inequality bounds each expectation by the uniform jth moment. Divide by n^j, sum over the fixed number of coordinates, and interpolate the remaining exponents using Hölder. The normal equation implies $\mathbb{E}q = 0$. □

B.2 PARAMETER BIAS AND COVARIANCE

Set $D = \widehat{G} - G$, $\bar{q} = n^{-1}\sum_i q_i$. On

$$\mathcal{H}_n = \{\|D\| \le g_0/2,\ \|\bar{q}\| \le \delta g_0/2\},$$

the inverse is bounded by $2/g_0$ and the unprojected estimator lies in $B(\theta, \delta)$. Thus projection is inactive and

$$\widehat{\theta} - \theta = G^{-1}\bar{q} - G^{-1}DG^{-1}\bar{q} + \rho_n, \qquad \|\rho_n\| \le C\|D\|^2\|\bar{q}\|. \tag{29}$$

The identity follows by applying the resolvent identity twice; it does not rely on an infinite formal series. Lemma 8 gives $\Pr(\mathcal{H}_n^c) = O(n^{-4})$. Since the projected estimator and F_K are bounded, this event makes a negligible contribution. For example,

$$\mathbb{E}[\|\bar{q}\|\mathbf{1}_{\mathcal{H}_n^c}] \le (\mathbb{E}\|\bar{q}\|^2)^{1/2}\Pr(\mathcal{H}_n^c)^{1/2} = O(n^{-5/2}).$$

Similar Hölder bounds apply to the quadratic terms in (29).

Independence and centering give

$$\mathbb{E}[DG^{-1}\bar{q}] = n^{-1}\mathbb{E}[(A - G)G^{-1}q].$$

Also, $\mathbb{E}\|D\|^2\|\bar{q}\| = O(n^{-3/2})$, by the fourth moment of D and second moment of $\bar{q}$. Hence

$$\mathbb{E}(\widehat{\theta} - \theta) = \frac{\beta}{n} + O(n^{-3/2}), \quad \beta = -G^{-1}\mathbb{E}[(A - G)G^{-1}q]. \tag{30}$$

The leading term $G^{-1}\bar{q}$ has second moment Ξ/n. The difference from that leading term has L^2 norm $O(n^{-1})$: use the fourth moments for $D\bar{q}$, and the eighth and fourth moments for $D^2\bar{q}$. Cauchy–Schwarz therefore gives

$$\mathbb{E}[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)^\top] = \frac{\Xi}{n} + O(n^{-3/2}). \tag{31}$$

Moreover, $\mathbb{E}\|\widehat{\theta} - \theta\|^3 = O(n^{-3/2})$, using the bounded inverse on $\mathcal{H}_n$ and boundedness off it.

Taylor expansion of F_K around θ now has an expected cubic remainder $O(n^{-3/2})$. Substituting (30) and (31) proves Proposition 2. At a smooth nonnegative minimum with value zero, the gradient vanishes and the Hessian is positive semidefinite, proving (6). These arguments also prove all claimed uniformities.

B.3 FOURTH-ORDER SMOOTHNESS IMPROVES THE REMAINDER

Assume additionally that the fourth derivatives of F_K are uniformly bounded near the population minimizers. We prove the $O(n^{-2})$ refinement in Proposition 2; the eighth-moment assumption (28) is unchanged.

For any three coordinates of centered sample averages from the same independent observations,

$$\mathbb{E}[\bar{Z}_a\bar{Z}_b\bar{Z}_c] = \frac{\mathbb{E}[Z_aZ_bZ_c]}{n^2}. \tag{32}$$

Indeed every summand with a sample index occurring only once vanishes. The identity applies to coordinates of D and $\bar{q}$, including products with bounded deterministic matrix coefficients. We use (32), rather than bounding the absolute cubic moment by $O(n^{-3/2})$.

On $\mathcal{H}_n$, one more resolvent step gives, with $\Delta = \widehat{\theta} - \theta$,

$$\Delta = L_n + M_n + V_n + R_n, \quad L_n = G^{-1}\bar{q}, \quad M_n = -G^{-1}DG^{-1}\bar{q},$$
$$V_n = G^{-1}DG^{-1}DG^{-1}\bar{q}, \qquad \|R_n\| \le C\|D\|^3\|\bar{q}\|.$$

The deterministic coefficients are uniformly bounded. By (32), $\mathbb{E}V_n = O(n^{-2})$, while Hölder and Lemma 8 give

$$\mathbb{E}\|D\|^3\|\bar{q}\| \le (\mathbb{E}\|D\|^6)^{1/2}(\mathbb{E}\|\bar{q}\|^2)^{1/2} = O(n^{-2}).$$

Removing the indicator of $\mathcal{H}_n$ from these polynomial terms has smaller order: its complement has probability $O(n^{-4})$, and the needed second moments of the polynomials follow from (28). Boundedness of the projected estimator controls its own contribution there. Consequently,

$$\mathbb{E}\Delta = \frac{\beta}{n} + O(n^{-2}). \tag{33}$$

For the second moment, use $\Delta = L_n + M_n + O(\|D\|^2\|\bar{q}\|)$ on $\mathcal{H}_n$. The expectations of $L_n M_n^\top$ and its transpose are $O(n^{-2})$ by (32). All remaining terms are bounded in expectation by a constant times

$$\mathbb{E}\|D\|^2\|\bar{q}\|^2 \le (\mathbb{E}\|D\|^4)^{1/2}(\mathbb{E}\|\bar{q}\|^4)^{1/2} = O(n^{-2});$$

higher powers of D on $\mathcal{H}_n$ are bounded by the same expression. Thus

$$\mathbb{E}\Delta\Delta^\top = \frac{\Xi}{n} + O(n^{-2}). \tag{34}$$

The third tensor moment of L_n is $O(n^{-2})$ by (32). Since $\|\Delta\| \le C\|\bar{q}\|$ and $\|\Delta - L_n\| \le C\|D\|\|\bar{q}\|$ on $\mathcal{H}_n$, the difference between the third tensor moments of Δ and L_n is bounded by

$$C\mathbb{E}\|D\|\|\bar{q}\|^3 \le C(\mathbb{E}\|D\|^4)^{1/4}(\mathbb{E}\|\bar{q}\|^4)^{3/4} = O(n^{-2}).$$

The complement again contributes smaller order. Finally $\mathbb{E}\|\Delta\|^4 = O(n^{-2})$. Taylor expansion through degree three, with expected fourth-order remainder $O(n^{-2})$, now combines with (33)–(34) to prove

$$\mathbb{E}F_K(\widehat{\theta}) = F_K(\theta) + \frac{1}{n}\left[\nabla F_K(\theta)^\top \beta + \frac{1}{2}\operatorname{tr}(\nabla^2 F_K(\theta)\Xi)\right] + O(n^{-2}).$$

All constants are uniform under the stated assumptions. This refinement is used for both scalar protocols, whose terminal risks are analytic on their fixed stabilizing intervals.

B.4 Why the scalar Gaussian family satisfies the assumptions

Fix $s > 1$. For every coupling of $X_0 \sim \mathcal{N}(0,1)$ and $X_1 \sim \mathcal{N}(0, s^2)$, $c \in [-s, s]$, and

$$\frac{s^2 - s + 1}{3} \le G(c) \le \frac{s^2 + s + 1}{3}.$$

The population slope is positive and satisfies

$$a(c) \le \frac{3(s^2-1)}{2(s^2 - s + 1)} \le \sqrt{3}.$$

The last inequality follows by taking square roots in the polynomial certificate in Appendix G. For fixed s the lower slope bound is also strictly positive, so all minimizers are a uniform positive distance inside $[0, 3]$.

Every moment of X_0, X_1 needed in (28) is uniformly bounded over couplings: mixed absolute moments obey

$$\mathbb{E}|X_0|^u|X_1|^v \le (\mathbb{E}|X_0|^{2u})^{1/2}(\mathbb{E}|X_1|^{2v})^{1/2}.$$

Since X_t, U are linear endpoint functions with bounded time coefficients, $A = X_t^2$ and $q = X_t(U - aX_t)$ meet the uniform moment conditions. The functions $(R_K(a) - s)^2$ have bounded derivatives of every fixed order on $[0, 3]$, for every fixed finite K and for $K = \infty$. This proves the $O(n^{-2})$ expansion over *all* endpoint couplings via the fourth-order refinement, not just over the Gaussian subfamily. The denominator in (9) vanishes only if each $X_{t_i} = 0$. For almost every nonzero endpoint pair, its linear path vanishes at at most one time, an event of zero uniform-time probability.

C Quartic identities and the scalar risk coefficient

Let $x = X_0$, $y = X_1$, and $z_t = (1-t)x + ty$. Expanding and integrating monomials gives

$$\begin{aligned} J_a(x,y) = &\left(\frac{a^2}{5} + \frac{a}{2} + \frac{1}{3}\right)x^4 + \left(\frac{a^2}{5} - \frac{1}{3}\right)(x^3y + xy^3) + \frac{a^2}{5}x^2y^2 \\ &+ \left(\frac{a^2}{5} - \frac{a}{2} + \frac{1}{3}\right)y^4. \end{aligned} \tag{35}$$

A convenient direct derivation uses

$$\int_0^1 z_t^4 \mathrm{d}t = \frac{x^4 + x^3y + x^2y^2 + xy^3 + y^4}{5}, \quad \int_0^1 z_t^3(y-x)\mathrm{d}t = \frac{y^4 - x^4}{4},$$

and

$$\int_0^1 z_t^2(y-x)^2\mathrm{d}t = \frac{x^4 - x^3y - xy^3 + y^4}{3}.$$

These follow either from the antiderivatives of powers of z_t or polynomial integration; the formulas extend by continuity to $x = y$.

For any Gaussian-marginal coupling with covariance c, define

$$V_\pi = \frac{\mathbb{E}_\pi J_{a(c)}(X_0, X_1)}{G(c)^2}, \qquad \beta_\pi = \frac{a(c)\mathbb{E}X_t^4 - \mathbb{E}X_t^3 U}{G(c)^2}. \tag{36}$$

Both depend on observable endpoint moments through degree four. In particular,

$$\mathbb{E}X_t^3 U = \frac{3s^4 - 3}{4}, \qquad \mathbb{E}X_t^4 = \frac{\mathbb{E}(x^4 + x^3y + x^2y^2 + xy^3 + y^4)}{5}.$$

Writing $e = R_K(a(c)) - s$, Proposition 2 becomes

$$\mathfrak{R}_{n,K}(\pi) = F(c) + \frac{T(\pi)}{n} + O(n^{-2}), \quad T(\pi) = 2eR_K'(a)\beta_\pi + [R_K'(a)^2 + eR_K''(a)]V_\pi, \tag{37}$$

where $F(c) = e^2$. For finite K,

$$R_K'(a) = (1 + a/K)^{K-1}, \quad R_K''(a) = \frac{K-1}{K}(1 + a/K)^{K-2},$$

with the second derivative zero for $K = 1$; at $K = \infty$, both derivatives equal e^a. At $c = c_\star$, $T(\pi) = R_K'(a_\star)^2\mathbb{E}J_{a_\star}/G_\star^2$.

Strict positivity and an explicit lower bound. Let $u = y - x$, $w = (x^2, xu, u^2)^\top$, and

$$T_a = \begin{pmatrix} -a & 1 & 0 \\ 0 & -2a & 1 \\ 0 & 0 & -a \end{pmatrix}, \qquad H_3 = \begin{pmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{pmatrix}.$$

The coefficients of $z_t(u - az_t)$ are $T_a w$, so $J_a = w^\top M_a w$ with $M_a = T_a^\top H_3 T_a$. For $a > 0$, $M_a \succ 0$ because H_3 is the Gram matrix of $1, t, t^2$ and T_a is invertible. Direct calculation gives

$$\det M_a = \frac{a^6}{540}, \qquad \operatorname{tr} M_a = \frac{38a^2}{15} - \frac{5a}{2} + \frac{4}{3}.$$

For three positive eigenvalues, $\lambda_{\min}(M_a) \geq 4\det(M_a)/\operatorname{tr}(M_a)^2$. Since $\mathbb{E}X_0^4 = 3$, every admissible coupling satisfies

$$\frac{R_K'(a_\star)^2}{G_\star^2}\mathbb{E}J_{a_\star} \geq \frac{12R_K'(a_\star)^2 \det M_{a_\star}}{G_\star^2 \operatorname{tr}(M_{a_\star})^2} > 0. \tag{38}$$

This is a conservative explicit statistical floor, not a claimed sharp value of the constrained transport optimum.

D Proof of the global coupling oracle

Fix the hypotheses of Theorem 3. We prove existence, uniform oracle regret, and the limiting constant without assuming an optimizer remains in the calibrated covariance class at finite n.

D.1 Compactness and attainment

The set $\Pi(\mu, \nu)$ is tight because both marginals are fixed; its weak limits preserve those marginals, so it is weakly compact. Polynomial moments of degree at most four are continuous under such weak convergence. To justify the unbounded functions, choose any moment order greater than four and use the fixed Gaussian marginal bounds and Hölder to obtain uniform integrability. In particular, covariance is continuous, the constraint $\mathbb{E}X_0X_1 = c_\star$ defines a closed subset, and the quartic objective is continuous there. The subset is nonempty: a Gaussian coupling realizes every $c_\star \in [-s, s]$. Thus an optimizer $\pi_\star$ exists. Strict positivity follows from (38).

D.2 Quadratic separation of the population optimum

The function $a(c)$ in (8) is smooth and strictly decreasing on $[-s, s]$. Its derivative is

$$a'(c) = -\frac{3(s^2-1)}{2(1+s^2+c)^2} < 0.$$

All $a(c)$ are positive and $R_K'(a) > 0$. Consequently the magnitude of the derivative of $R_K(a(c))$ has a positive minimum m_0 on this compact interval. Since $R_K(a(c_\star)) = s$, the mean value theorem gives

$$F(c) \geq \kappa(c - c_\star)^2, \qquad \kappa = m_0^2 > 0. \tag{39}$$

This also proves uniqueness of the calibrated covariance. It does not claim uniqueness of the optimal coupling within that covariance class.

D.3 Repairing covariance by a small mixture

Define the linear, nonnegative functional on all couplings

$$\overline{S}(\pi) = \frac{R_K'(a_\star)^2}{G_\star^2}\mathbb{E}_\pi J_{a_\star}.$$

It is bounded above by a finite constant M uniformly over $\Pi(\mu, \nu)$, by the fixed marginal moment bounds. Let $c = \mathbb{E}_\pi X_0 X_1$. If $c > c_\star$, mix π with the antimonotone Gaussian coupling π_-, having covariance $-s$, using

$$\alpha = \frac{c - c_\star}{c + s}, \qquad \pi^0 = (1-\alpha)\pi + \alpha\pi_-.$$

Then π^0 has covariance $c_\star$ and

$$0 \leq \alpha \leq \frac{|c - c_\star|}{s + c_\star}.$$

If $c < c_\star$, use the perfectly positively correlated coupling π_+ of covariance s and $\alpha = (c_\star - c)/(s - c) \leq |c - c_\star|/(s - c_\star)$. For $c = c_\star$, set $\pi^0 = \pi$. Both denominators in these bounds are strictly positive by the theorem's interior hypothesis. Linearity gives

$$S_\star \leq \overline{S}(\pi^0) \leq \overline{S}(\pi) + M\alpha,$$

so for a uniform $L_0 < \infty$,

$$\overline{S}(\pi) \geq S_\star - L_0|c - c_\star|. \tag{40}$$

The repairing couplings may be singular. They are used in an oracle proof over *all* couplings; the smooth separation construction has no such singularity.

D.4 The uniform statistical comparison

In (37), all coefficients multiplying endpoint fourth moments are smooth functions of c on $[-s, s]$. Their derivatives are bounded, and the fourth moments are uniformly bounded. Replacing $a(c), G(c)$ and the sampler derivatives by their values at $c_\star$, while holding the coupling's endpoint moments fixed, therefore yields

$$|T(\pi) - \overline{S}(\pi)| \leq L_1|c - c_\star|. \tag{41}$$

The term involving β_π is included here: its multiplier e vanishes at $c_\star$, and β_π is uniformly bounded. Combining (37), (39), (40), and (41), for every coupling,

$$\mathfrak{R}_{n,K}(\pi) \geq \kappa(c - c_\star)^2 + \frac{S_\star - L|c - c_\star|}{n} - Cn^{-2}, \qquad L = L_0 + L_1.$$

Minimizing the right side over $|c - c_\star| \geq 0$ gives (14). At the attained constrained optimum,

$$\mathfrak{R}_{n,K}(\pi_\star) = S_\star/n + e_n, \qquad |e_n| \leq Cn^{-2}.$$

Hence the more explicit regret bound is

$$0 \leq \mathfrak{R}_{n,K}(\pi_\star) - \inf_{\pi \in \Pi(\mu,\nu)} \mathfrak{R}_{n,K}(\pi) \leq 2Cn^{-2} + \frac{L^2}{4\kappa n^2}. \tag{42}$$

Multiplying the upper and lower risk bounds by n proves the limit in Theorem 3. This argument treats n-dependent couplings directly and does not interchange an infimum with a pointwise asymptotic expansion.

E PROOFS OF STRICT GAUSSIAN-FAMILY SEPARATION

E.1 A CALIBRATED INDEPENDENT COUPLING FOR EACH FIXED BUDGET

For covariance zero, $a(0) = 3(s^2-1)/(2(1+s^2)) = (3/2)\tanh z$, where $z = \log s$. Population calibration requires $a(0) = \lambda_K(s)$. For finite K, after dividing by $z > 0$ the equation is

$$\frac{3}{2}\frac{\tanh z}{z} = \frac{e^{z/K}-1}{z/K}.$$

The left side decreases continuously from $3/2$ to zero. Indeed $\tanh z - z\,\mathrm{sech}^2 z$ has derivative $2z\,\mathrm{sech}^2 z\tanh z > 0$ and is zero at zero. The right side increases continuously from one to infinity, as follows from its positive-coefficient power series. There is exactly one positive intersection. For exact integration the right side is one, giving the same conclusion. This proves existence and uniqueness of s_K without using numerical root finding.

For $K = 1$, cancel $s - 1 > 0$ in

$$\frac{3(s^2-1)}{2(1+s^2)} = s - 1$$

to obtain $2s^2 - 3s - 1 = 0$, and hence $s_1 = (3+\sqrt{17})/4$. At this point

$$a_0 = s_1 - 1, \quad G_0 = \frac{s_1+1}{2}, \quad \frac{a_0^2 s_1^2}{G_0^2} = 1. \tag{43}$$

E.2 EXACT MOMENTS OF THE SMOOTH CONSTRUCTION

For $Z \sim \mathcal{N}(0,1)$,

$$\mathbb{E}e^{-3Z^2/2} = \frac{1}{2}, \qquad \mathbb{E}Z^2 e^{-3Z^2/2} = \frac{1}{8}.$$

These follow by evaluating the Gaussian integral and its derivative with respect to its quadratic coefficient. Thus $h(z) = 2e^{-3z^2/2} - 1$ is centered, even, bounded in absolute value by one, and satisfies $\mathbb{E}Z^2 h(Z) = -3/4$. For any $0 < \tau < 1$, (15) integrates to one, has both desired marginals, and is bounded below by $(1-\tau)$ times the independent density. Its covariance and the mixed moments of odd degree in either coordinate vanish by parity. Its relevant fourth moments are

$$\mathbb{E}x^4 = 3, \quad \mathbb{E}x^3 y = 0, \quad \mathbb{E}x^2y^2 = s^2\left(1 - \frac{9\tau}{16}\right), \quad \mathbb{E}xy^3 = 0, \quad \mathbb{E}y^4 = 3s^4. \tag{44}$$

For a Gaussian coupling of covariance $c = s\rho$, the corresponding values are

$$3, \qquad 3c, \qquad s^2 + 2c^2, \qquad 3s^2 c, \qquad 3s^4. \tag{45}$$

For example, write $X_1 = cX_0 + \sqrt{s^2 - c^2}Z$ with independent standard Z and expand the monomials; this proves (45), including its continuous endpoint limits.

At $s = s_K$, both the smooth coupling and π_0^G have $c = 0 = c_\star$, and thus the same entire population objective and terminal law. The only changed entry in (35) is $\mathbb{E}x^2y^2$. Since its coefficient is $a_0^2/5$, their statistical constants differ by

$$\frac{R_K'(a_0)^2}{G_0^2}\frac{a_0^2}{5}\frac{9\tau s_K^2}{16} = D_K.$$

At $K = 1$, substituting (43) and $\tau = 3/4$ gives $27/320$. The independent constant from (35) is $17/10 + \sqrt{17}/5$, proving (18).

E.3 WHY OPTIMIZING GAUSSIAN CORRELATION CANNOT REMOVE THE GAP

Here is the argument at any scale where the unique calibrated covariance is interior. Let $\rho_\star = c_\star/s$. Uniform expansion over $\rho \in [-1,1]$ gives

$$\mathfrak{R}_{n,K}(\pi_\rho^G) = F(s\rho) + T_G(\rho)/n + O(n^{-2}),$$

where T_G is continuous, and F has its unique zero at $c_\star$. Choose approximate minimizers ρ_n with risk within n^{-2} of the infimum. Evaluating the objective at $\rho_\star$ shows their risk is $O(1/n)$. Since T_G is bounded, this forces $F(s\rho_n) \to 0$ and therefore $\rho_n \to \rho_\star$. Consequently

$$\liminf_n n\mathfrak{R}_{n,K}(\pi^G_{\rho_n}) \geq \lim_n T_G(\rho_n) = T_G(\rho_\star).$$

The reverse upper bound follows by using $\rho_\star$ itself. We have proved

$$\lim_n n \inf_\rho \mathfrak{R}_{n,K}(\pi^G_\rho) = T_G(\rho_\star). \tag{46}$$

In fact the error is $O(n^{-2})$ before multiplying by n: T_G is smooth on the compact correlation interval, so $T_G(\rho) \geq T_G(\rho_\star) - L_G s|\rho - \rho_\star|$. Combine this with $F(s\rho) \geq \kappa s^2(\rho - \rho_\star)^2$ and complete the square as in (14). Thus the infimum risk equals $T_G(\rho_\star)/n + O(n^{-2})$, and the separation difference equals $D_K/n + O(n^{-2})$.

This proof covers every sequence of correlations, including those approaching singular Gaussian couplings; the population optimum is interior, and the uniform moment argument already includes the endpoints. Applying (46) at $\rho_\star = 0$ completes Theorem 4.

E.4 Proof of the open-family corollary

Put $d = 9\tau/16 = 27/64$ and $r = |c_\star|/s < 1/4$. The mixture (19) has covariance

$$(1-2r)\cdot 0 + 2r \cdot s\,\mathrm{sign}(c_\star)/2 = c_\star.$$

It has a smooth, full-support density bounded below by $(1-2r)(1-\tau) > 1/8$ times the independent density. Its x^4, y^4, x^3y, xy^3 moments agree with those of the calibrated Gaussian. Its remaining fourth moment is

$$\mathbb{E}_{\pi_s} x^2y^2 = s^2[1 - d + (1+2d)r],$$

whereas the calibrated Gaussian has $s^2(1+2r^2)$. Their difference is

$$s^2[d - (1+2d)r + 2r^2] = s^2(1-2r)(d-r) > 0.$$

Substitution into (35), followed by (46), proves (20). At $K = 1$,

$$c_\star = \frac{1}{2} + \frac{3}{2}s - s^2, \qquad \rho_\star = \frac{1}{2s} + \frac{3}{2} - s.$$

The latter function is strictly decreasing. Solving $\rho_\star = 1/4$ gives $(5+\sqrt{57})/8$; solving $\rho_\star = -1/4$ gives 2. This proves the stated open interval. For any other fixed budget, continuity gives a nonempty open neighborhood of s_K with the same property.

F Matrix equivalence, reachability, and terminal curvature

F.1 Constant matrix fields: exact reachability

For $\mu = \mathcal{N}(0, \Sigma_0)$ and $\nu = \mathcal{N}(0, \Sigma_1)$ with $\Sigma_i \succ 0$, consider $v_B(x,t) = Bx$. The joint endpoint law need not be Gaussian. Define

$$\begin{aligned} C &= \mathbb{E}[X_1 X_0^\top], \quad \Gamma(C) = \begin{pmatrix} \Sigma_0 & C^\top \\ C & \Sigma_1 \end{pmatrix}, \\ G(C) &= \frac{\Sigma_0 + \Sigma_1}{3} + \frac{C + C^\top}{6}, \qquad H(C) = \frac{\Sigma_1 - \Sigma_0 + C - C^\top}{2}. \end{aligned} \tag{47}$$

Theorem 9 (Population sufficiency and prescribed-field reachability)**.** *For the preceding model, the following statements hold.*

(i) Every feasible C satisfies $\Gamma(C) \succeq 0$ and

$$\frac{\Sigma_0 + \Sigma_1}{6} \preceq G(C) \preceq \frac{\Sigma_0 + \Sigma_1}{2}. \tag{48}$$

The entire objective and its unique minimizer are

$$\begin{aligned} \mathcal{L}_C(B) &= \mathrm{tr}(BGB^\top) - 2\,\mathrm{tr}(BH^\top) + \mathrm{tr}(\Sigma_0 + \Sigma_1 - C - C^\top), \\ B(C) &= H(C)G(C)^{-1}. \end{aligned} \tag{49}$$

(ii) Every C with $\Gamma(C) \succeq 0$ is realized by a jointly Gaussian coupling. Replacing any coupling by this Gaussian coupling leaves the whole population objective and the generated law unchanged.

(iii) A prescribed $B_\star$ is the unique population minimizer of some coupling if and only if there exists symmetric G such that

$$\begin{aligned} B_\star G + G B_\star^\top &= \Sigma_1 - \Sigma_0, \\ C(G) = 3G - \Sigma_0 - \Sigma_1 + \tfrac{1}{2}(B_\star G - G B_\star^\top), &\qquad \Gamma(C(G)) \succeq 0. \end{aligned} \tag{50}$$

For fixed $B_\star$, this is a semidefinite feasibility problem.

The generated covariance is $S_K(C) = P_K(B(C))\Sigma_0 P_K(B(C))^\top$, where $P_K(B) = (I + B/K)^K$ and $P_\infty(B) = e^B$. Its exact terminal W_2^2 is the Gaussian Bures distance between $S_K(C)$ and Σ_1. Thus population coupling optimization reduces exactly to finite-dimensional optimization over $\Gamma(C) \succeq 0$. The objective need not be convex. Likewise, failure of (50) for one specified $B_\star$ does not rule out other velocities generating the same target.

F.2 Proof of Theorem 9

The block covariance $\Gamma(C)$ is positive semidefinite for every coupling. Conversely any such block covariance defines a possibly singular joint Gaussian law with the required marginal covariances. Independence is not required.

Integrating the endpoint expansions,

$$\mathbb{E} X_t X_t^\top = \frac{\Sigma_0 + \Sigma_1}{3} + \frac{C + C^\top}{6} = G,$$

$$\mathbb{E} U X_t^\top = \frac{\Sigma_1 - \Sigma_0 + C - C^\top}{2} = H.$$

Expanding the squared Frobenius regression objective gives (49). The PSD block condition, applied to (z, z) and $(z, -z)$, implies

$$-(\Sigma_0 + \Sigma_1) \preceq C + C^\top \preceq \Sigma_0 + \Sigma_1,$$

which proves (48). Thus $G \succ 0$ and $B = HG^{-1}$ is unique. Linear deterministic integration maps a Gaussian initial condition to a Gaussian endpoint, with covariance $P_K(B)\Sigma_0 P_K(B)^\top$. This depends only on C.

If $B_\star G = H$, then taking symmetric parts gives $B_\star G + G B_\star^\top = \Sigma_1 - \Sigma_0$. Since $C + C^\top = 6G - 2(\Sigma_0 + \Sigma_1)$ and $C - C^\top = B_\star G - G B_\star^\top$, we recover (50). Conversely, if (50) holds, the constructed covariance realizes a Gaussian coupling with $G(C) = G$ and

$$H(C) = \tfrac{1}{2}(\Sigma_1 - \Sigma_0) + \tfrac{1}{2}(B_\star G - G B_\star^\top) = B_\star G.$$

Its optimizer is $B_\star$. The bound (48) guarantees $G \succ 0$ automatically; it need not be imposed as a separate strict constraint in the semidefinite feasibility test.

Time-dependent linear spatial classes. The population equivalence argument extends to any fixed class $v_\theta(x, t) = A_\theta(t)x$ for which the expectations exist. At each time, both $\mathbb{E} X_t X_t^\top$ and $\mathbb{E} U X_t^\top$ depend only on C. Thus the entire objective as a function of θ is identical for equal C. Under a deterministic choice of population optimizer and a fixed linear ODE solver, the output law is identical as well. The reachability equations (50) and scalar statistical oracle are not asserted for an arbitrary time-dependent class.

F.3 A full-rank sampler-correct construction

Suppose $\Sigma_0 = I$ and $\Sigma_1 = \mathrm{diag}(s_1^2, \dots, s_d^2)$ with $1 < s_i \le 2$. Fix finite K and set

$$b_i = K(s_i^{1/K} - 1), \qquad c_i = \frac{3(s_i^2 - 1)}{2b_i} - (1 + s_i^2), \quad B_\star = \mathrm{diag}(b_i), \quad C = \mathrm{diag}(c_i).$$

The scalar normal equation gives this $B_\star$, and $P_K(B_\star) = \mathrm{diag}(s_i)$ gives the exact target covariance. To prove nondegeneracy, let $a(c)$ be (8). On $1 < s \le 2$,

$$a(s) < \log s < K(s^{1/K} - 1) \le s - 1 < a(-s). \tag{51}$$

The middle inequalities follow from $e^u > 1 + u$ and the concavity of $s^{1/K}$. For the first,

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[\log s - \frac{3(s^2-1)}{2(s^2+s+1)}\right] = \frac{(s-1)^2(2s^2+5s+2)}{2s(s^2+s+1)^2} > 0,$$

and the expression is zero at $s = 1$. The last follows from $3(s+1) > 2(s^2 - s + 1)$ on the stated range. Since a is strictly decreasing, (51) gives $-s_i < c_i < s_i$. Each 2×2 block covariance is positive definite. The exact-integration construction uses $b_i = \log s_i$ and is also interior.

This property persists on an open set containing noncommuting endpoint covariances. Near the displayed expanding pair, take the symmetric positive definite transport map

$$T = \Sigma_0^{-1/2}(\Sigma_0^{1/2}\Sigma_1\Sigma_0^{1/2})^{1/2}\Sigma_0^{-1/2}.$$

For finite K, let $B = K(T^{1/K} - I)$; for exact integration let $B = \log T$. All eigenvalues of B stay positive in a sufficiently small neighborhood. The Sylvester equation $BG + GB^\top = \Sigma_1 - \Sigma_0$ therefore has a unique symmetric solution varying continuously with the covariances. Its constructed $\Gamma(C(G))$ stays positive definite by continuity. Moreover $P_K(B) = T$, so the terminal covariance remains exact. This proves an open-set statement, not global feasibility for arbitrary covariances.

F.4 An exact noncommuting loss reversal

Take

$$\Sigma_0 = \begin{pmatrix} 1 & 1/4 \\ 1/4 & 1 \end{pmatrix}, \quad T = \mathrm{diag}(3/2, 7/4), \quad \Sigma_1 = T\Sigma_0 T = \begin{pmatrix} 9/4 & 21/32 \\ 21/32 & 49/16 \end{pmatrix}.$$

Their commutator is nonzero. For one Euler step, the prescribed model is $B_\star = T - I$. The reachability construction gives

$$G_\star = \begin{pmatrix} 5/4 & 13/40 \\ 13/40 & 11/8 \end{pmatrix}, \qquad C_\star = \begin{pmatrix} 1/2 & 9/320 \\ 7/64 & 1/16 \end{pmatrix}.$$

The four leading principal minors of $\Gamma(C_\star)$ are

$$1, \quad 15/16, \quad 191039/102400, \quad 448396173/83886080,$$

so the joint Gaussian density is smooth and everywhere positive. The exact loss is $653/128$ and $(I + B_\star)\Sigma_0(I + B_\star)^\top = \Sigma_1$.

The alternative $C_{\mathrm{low}} = 0.9T\Sigma_0$ is also strictly feasible: its Schur complement is $(1 - 0.9^2)\Sigma_1 \succ 0$. Its population optimizer from (49) has loss approximately 0.64181058, while its one-step terminal error is $0.04904102 > 0$. The covariance mismatch is verified directly; the displayed Wasserstein value is numerical evaluation of the standard Gaussian formula

$$d_B^2(S, \Sigma) = \mathrm{tr}(S + \Sigma) - 2\,\mathrm{tr}[(\Sigma^{1/2}S\Sigma^{1/2})^{1/2}]. \tag{52}$$

This formula is standard (Bhatia et al., 2019); the reachability construction and comparison, rather than the formula itself, are the present example.

F.5 Computable terminal sensitivity for the matrix model

At a population-correct $B_\star$, put $P = P_K(B_\star)$, so $P\Sigma_0 P^\top = \Sigma_1$. With $A = I + B_\star/K$, the Fréchet derivative is

$$DP_K(B_\star)[E] = \frac{1}{K}\sum_{\ell=0}^{K-1} A^\ell E A^{K-1-\ell}.$$

For exact integration it is $\int_0^1 e^{(1-t)B_\star} E e^{tB_\star}\,\mathrm{d}t$. The first covariance perturbation is

$$\Delta\Sigma = DP_K[E]\Sigma_0 P^\top + P\Sigma_0 DP_K[E]^\top.$$

Let $\mathscr{L}_{\Sigma_1}(Z) = \Sigma_1 Z + Z\Sigma_1$. The local Bures metric gives the parameter quadratic form

$$Q_K(E) = \frac{1}{2}\,\mathrm{tr}[\Delta\Sigma\,\mathscr{L}_{\Sigma_1}^{-1}(\Delta\Sigma)]. \tag{53}$$

For completeness, diagonalize Σ_1 with eigenvalues $\lambda_i > 0$. The second-order term of (52) is $\frac{1}{2}\sum_{ij}(\Delta\Sigma_{ij})^2/(\lambda_i + \lambda_j)$, which is (53). One can derive it by differentiating the matrix square-root equation twice, or from the Bures metric (Bhatia et al., 2019). Second-order covariance perturbations contribute no separate term because the first derivative of the distance vanishes at equality.

Under column vectorization of B, the feature matrix is $X_t^\top \otimes I$. Its parameter influence reshaped as a matrix is

$$(U - B_\star X_t)X_t^\top G(C)^{-1}.$$

Substituting this matrix into the quadratic form (53) gives the observable score (6). Directions with $\Delta\Sigma = 0$ have zero first-order terminal cost, even when their parameter or velocity error is nonzero. The score therefore differs from unweighted parameter covariance.

G Loss-ranking classification and complementary counterexamples

These results motivate the statistical analysis, but are distinct from it. They show that population terminal bias and numerical integration can each invalidate raw-loss rankings. They do not imply that supervision noise is statistically irrelevant.

G.1 An exact classification for arbitrary endpoint laws

Let μ, ν be centered distributions on $\mathbb{R}^d$ with second moments $1, s^2$, respectively, and $s > 1$. For $v_a(x,t) = ax$ put $c = \mathbb{E}\langle X_0, X_1\rangle$ and $\gamma = \sup_{\pi\in\Pi(\mu,\nu)} \mathbb{E}\langle X_0, X_1\rangle$. Independence is feasible, so $\gamma \geq 0$; Cauchy–Schwarz gives $|c| \leq s$. Direct integration gives

$$\mathbb{E}\|U\|^2 = 1 + s^2 - 2c, \quad \mathbb{E}\|X_t\|^2 = \frac{1 + s^2 + c}{3}, \quad \mathbb{E}\langle X_t, U\rangle = \frac{s^2 - 1}{2}.$$

The resulting quadratic regression gives (8). Its optimized derivative is

$$\frac{\mathrm{d}L^*}{\mathrm{d}c} = -2 + \frac{3(s^2-1)^2}{4(1+s^2+c)^2} \leq -1,$$

because

$$4(s^2 - s + 1)^2 - 3(s^2 - 1)^2 = (s^2 - 4s + 1)^2 \geq 0.$$

Thus smaller raw loss always prefers larger cross-moment in this class.

For any $r > 0$, dividing the first coordinate of a coupling of $r\mu, \nu$ by r gives a bijection with $\Pi(\mu, \nu)$. Expanding squared distance therefore gives

$$W_2^2(r\mu, \nu) = r^2 + s^2 - 2r\gamma = (r - \gamma)^2 + s^2 - \gamma^2. \tag{54}$$

In particular $\gamma = (1 + s^2 - W_2^2(\mu, \nu))/2$. For one-dimensional marginals with quantile functions Q_μ, Q_ν,

$$\gamma = \int_0^1 Q_\mu(u)Q_\nu(u)\mathrm{d}u.$$

Since $a(c) > 0$, all sampler scales are positive, and (54) gives

$$\mathcal{E}_K(c) = (R_K(a(c)) - \gamma)^2 + s^2 - \gamma^2. \tag{55}$$

If $\gamma > 1$, define

$$c_{K,\gamma} = \frac{3(s^2-1)}{2\lambda_K(\gamma)} - (1 + s^2).$$

Where this value lies in the feasible cross-moment interval, $\mathcal{E}_K(c)$ decreases strictly for $c < c_{K,\gamma}$ and increases strictly for $c > c_{K,\gamma}$. Indeed its derivative is $2(R_K(a(c)) - \gamma)R_K'(a(c))a'(c)$, and the

last two factors have negative product. Thus raw-loss ordering is correct on the first side and strictly reversed on the second. If the threshold lies outside the feasible interval, the appropriate boundary is optimal. If $\gamma \le 1$, then $R_K(a(c)) > 1 \ge \gamma$, so generation and raw loss both improve with larger c everywhere.

For a symmetric scale family $\nu = s\mu$, $\gamma = s$ and the feasible cross-moment interval is $[-s, s]$. Inequality (51) gives an interior threshold for every $1 < s \le 2$ and every fixed finite or infinite budget. This includes smooth symmetric mixtures, not just Gaussian endpoints. For $\gamma > 1$, expansion of $\lambda_K(\gamma) = \log\gamma + (\log\gamma)^2/(2K) + O(K^{-2})$ gives

$$c_{K,\gamma} = c_{\infty,\gamma} - \frac{3(s^2-1)}{4K} + O(K^{-2}).$$

The optimal coupling is sampler-dependent even when its endpoint marginals are held fixed.

G.2 A FULL-RANK GAUSSIAN REVERSAL AFTER EXACT NOISE CORRECTION

For scalar Gaussian endpoints with covariance $c \in (-s, s)$, define

$$q_c(t) = (1-t)^2 + 2ct(1-t) + s^2t^2, \quad r_c(t) = c - 1 + (1 + s^2 - 2c)t.$$

The conditional mean is $\bar{u}_c(x,t) = r_c(t)x/q_c(t)$. The linear transform from (X_0, X_1) to (X_t, U) has determinant one. Its covariance determinant remains $s^2 - c^2$, so

$$\mathrm{Var}(U \mid X_t, t) = \frac{s^2 - c^2}{q_c(t)}.$$

Completing the square in q_c and integrating its reciprocal yields

$$\mathcal{N}(c) = \sqrt{s^2 - c^2}\,\arccos(c/s). \tag{56}$$

For instance write $q = At^2 + 2Bt + 1$, $A = 1 + s^2 - 2c$, $B = c - 1$; then $A - B^2 = s^2 - c^2$. The arctangent difference from 0 to 1 is the angle $\arccos(c/s) \in (0, \pi)$, fixing the branch. The noise-corrected optimum is $\mathcal{A}(c) = L^*(c) - \mathcal{N}(c)$.

At $s = 2$, compare $c_A = 6/5$ with $c_B = 19/10$. Both joint covariance matrices are positive definite, with determinants 2.56 and 0.39. The values are

Quantity	$c_A = 1.2$	$c_B = 1.9$
Population slope	0.72580645	0.65217391
Raw optimal loss	1.51129032	0.22173913
Irreducible variance	1.48367235	0.19831642
Exact noise-corrected error	0.02761797	0.02342271
Eight-step terminal W_2^2	0.00001025	0.01641029
Exact-ODE terminal W_2^2	0.00440855	0.00644655

Thus raw loss and exact mean-velocity error both rank the worse generator first. This is not caused solely by numerical discretization. The Euler risks have exact rational forms

$$\left[(541/496)^8 - 2\right]^2, \qquad \left[(199/184)^8 - 2\right]^2.$$

The arccos quantities admit rational alternating-series certificates:

$$\mathcal{N}_A = \frac{8}{5}\sum_{j=0}^{\infty}\frac{(-1)^j}{4^j(2j+1)}, \qquad \mathcal{N}_B = \frac{1}{5}\sum_{j=0}^{\infty}\frac{(-1)^j}{39^j(2j+1)}.$$

Twenty and twenty-one terms bound each value from below and above. The exact-ODE comparisons can likewise be certified by rational Taylor bounds for $e^{45/62}$ and $e^{15/23}$. The verification script includes these interval checks, rather than treating rounded table entries as a proof.

G.3 Standard Gaussian source and a smooth bimodal target

The failure is not restricted to a Gaussian scale target. Take

$$\mu = \mathcal{N}(0,1), \quad \nu = \tfrac{1}{2}\mathcal{N}(-4, 0.05^2) + \tfrac{1}{2}\mathcal{N}(4, 0.05^2), \quad s^2 = 16.0025.$$

The coupling $X_1 = 4\,\mathrm{sign}(X_0) + 0.05Z$, with independent standard Z, gives $\gamma \geq 4\sqrt{2/\pi}$. Let ν_0 put equal mass at -4 and 4. Its maximal cross-moment is $4\sqrt{2/\pi}$, and $W_2(\nu, \nu_0) \leq 0.05$. The quantile formula and Cauchy–Schwarz give

$$4\sqrt{2/\pi} \leq \gamma \leq 4\sqrt{2/\pi} + 0.05.$$

For $c_A = 1, c_B = 2.8$, the raw losses are approximately 5.625677 and 2.878008, and the eight-step scales are 3.194677 and 2.893907. Substitution in (54), over the whole displayed interval for γ, proves $\mathcal{E}_{8,B} - \mathcal{E}_{8,A} > 0.086$. The common irreducible shape error $s^2 - \gamma^2$ cancels; no different error baseline is used.

Smooth full-support couplings with these cross-moments can be realized using a Gaussian copula. Let Z, E be independent standard Gaussians, $V = \rho Z + \sqrt{1-\rho^2}E$, and

$$X_0 = Z, \qquad X_1 = Q_\nu(\Phi(V)).$$

Here Φ is the standard Gaussian distribution function. Since $\mathbb{E}[Z \mid V] = \rho V$, this coupling has cross-moment $\rho\gamma$. Choose $\rho = c/\gamma$, which lies strictly inside $(-1, 1)$. Both marginal densities are positive and smooth; the Gaussian copula density is positive and smooth as well. This proves that the example can be realized without atoms or a singular joint covariance.

G.4 A reversal with no model approximation error

Fix the Gaussian endpoints $\mathcal{N}(0,1), \mathcal{N}(0,4)$ and the single, fixed time-dependent model class

$$\mathcal{V} = \{v_\eta(x,t) = r_\eta(t)x/q_\eta(t) : -2 < \eta < 2\}.$$

For every Gaussian coupling of covariance c, the true conditional mean belongs to this class at $\eta = c$. Hence the optimal approximation error is zero and optimal raw loss equals (56). Exact integration is correct for every c.

Three equal Euler steps at times $0, 1/3, 2/3$ have total scale

$$P_3(c) = \frac{5(c+2)(c+8)}{4(4c+17)}. \tag{57}$$

This follows by multiplying $1 + r_c(t)/(3q_c(t))$ at the three times. At $c_A = -8/5$ and $c_B = 0$,

$$\mathcal{N}_A = \frac{6}{5}\arccos(-4/5) < \pi = \mathcal{N}_B,$$

since $\arccos(-4/5) < 5\pi/6$, but

$$P_3(c_A) = \frac{16}{53}, \quad P_3(c_B) = \frac{20}{17}, \quad (P_3(c_A) - 2)^2 = \frac{8100}{2809} > \frac{196}{289} = (P_3(c_B) - 2)^2.$$

At c_A, $\min_t q_c(t) = 36/205 > 0.175$ and $\sup_t |r_c(t)/q_c(t)| \leq 41/12$. To see the latter, write $q = A(t + B/A)^2 + (s^2 - c^2)/A$ and $r = A(t + B/A)$; the unrestricted maximum of $|r|/q$ is $A/(2\sqrt{s^2 - c^2}) = 41/12$. Thus the fixed example does not approach a collapsed path.

There is also an exact ranking classification in this realizable class. Differentiation of (57) gives

$$P_3'(c) = \frac{5(2c^2 + 17c + 53)}{2(4c+17)^2} > 0,$$

and P_3 increases from 0 to 2 on $(-2, 2)$. Hence the three-step error is strictly decreasing. On the other hand,

$$\mathcal{N}'(c) = -1 - \frac{c}{\sqrt{4 - c^2}}\arccos(c/2).$$

Its unique zero is $c_{\mathrm{noise}} = 2\cos\theta_0$, where $\tan\theta_0 = -\theta_0$ and $\theta_0 \in (\pi/2, \pi)$. Uniqueness follows from strict increase of $\tan\theta + \theta$ on that interval. Below c_{noise} noise and terminal error have opposite rankings; above it they agree. As $c = -2 + \epsilon$ with $\epsilon \downarrow 0$, $\mathcal{N}(c) = 2\pi\sqrt{\epsilon} + O(\epsilon)$ while $P_3(c) = 5\epsilon/6 + O(\epsilon^2)$, so the loss tends to zero and the three-step error tends to four. This limiting statement does not have a uniform path-Lipschitz bound; it is not used in the fixed full-rank comparison or statistical separation.

G.5 Projection bias and integration error can cancel

In the scalar model the total error decomposes exactly as

$$\mathcal{E}_K(c) = s^2 - \gamma^2 + \left[(e^{a(c)} - \gamma) + (R_K(a(c)) - e^{a(c)})\right]^2.$$

The cross term can be negative, and can cancel the two signed errors at a sampler-calibrated covariance. Therefore adding two nonnegative squared error terms is an upper-bound device, not an exact ranking criterion. Even for a time-dependent scalar linear field, a velocity perturbation $e(t)x$ changes terminal scale through $\exp(\int e(t)\mathrm{d}t)$, whereas its pathwise squared error depends on $\int q(t)e(t)^2\mathrm{d}t$. A nonzero perturbation with integral zero can preserve the terminal scale. This explains structurally why mean-velocity L^2 error and terminal error need not have the same ordering.

H Computable transport design and candidate selection

H.1 A finite linear program and its dual

For discrete endpoint laws $\mu = \sum_i a_i \delta_{x_i}$ and $\nu = \sum_j b_j \delta_{y_j}$, let $P_{ij} \geq 0$ denote coupling masses. The scalar calibrated design problem is

$$\min_{P \geq 0} \sum_{ij} P_{ij} d_{ij}, \qquad P\mathbf{1} = a, \quad P^\top \mathbf{1} = b, \quad \sum_{ij} P_{ij} x_i y_j = c_\star, \tag{58}$$

where $d_{ij} = R_K'(a_\star)^2 J_{a_\star}(x_i, y_j)/G_\star^2$. Its finite-dimensional dual is

$$\max_{\alpha,\beta,\lambda} \sum_i a_i \alpha_i + \sum_j b_j \beta_j + \lambda c_\star, \qquad \alpha_i + \beta_j + \lambda x_i y_j \leq d_{ij}. \tag{59}$$

Weak duality follows by multiplying the inequality by P_{ij} and summing. For a feasible finite problem, boundedness and finite linear-program duality give equal primal and dual optima. This is an application of standard transport with additional linear constraints (Zaev, 2014), not a new duality theorem.

For the general fixed-feature design, define pairwise moments

$$G_{ij} = \int_0^1 \Phi(x_{ij,t}, t)^\top \Phi(x_{ij,t}, t)\mathrm{d}t, \quad b_{ij} = \int_0^1 \Phi(x_{ij,t}, t)^\top (y_j - x_i)\mathrm{d}t,$$

where $x_{ij,t} = (1-t)x_i + t y_j$. Impose $\sum P_{ij} G_{ij} = G_\star$ and $\sum P_{ij} b_{ij} = G_\star \theta_\star$. Use as cost

$$d_{ij} = \int_0^1 \|Q_K^{1/2} G_\star^{-1} \Phi(x_{ij,t}, t)^\top (y_j - x_i - v_{\theta_\star}(x_{ij,t}, t))\|^2 \mathrm{d}t.$$

The finite optimization is again linear, with corresponding matrix/vector multipliers in its dual. Infeasibility has the usual linear separation certificate. Such a certificate concerns the specified moments and model, not every possible generator of the endpoint law. If the endpoint moments, Gram matrix, or terminal geometry are estimated, their errors must be propagated into feasibility and objective uncertainty; the exact formulas do not remove these estimation requirements.

H.2 An explicit optimizer in a sign-symmetric subclass

Consider Gaussian endpoints and covariance zero at the calibrated scale $s = s_K$. Restrict to joint laws invariant under either coordinate's sign change separately. Then $\mathbb{E}x^3 y = \mathbb{E}xy^3 = 0$ and the marginal fourth moments are fixed. By (35), the only variable term in the first-order risk is a positive multiple of $\mathbb{E}x^2y^2$. Set $R_0 = x^2$ and $R_1 = y^2/s^2$, each distributed as χ_1^2. Thus the optimal design within this subclass minimizes $\mathbb{E}R_0 R_1$.

The minimizer is the antimonotone quantile coupling

$$R_0 = Q_{\chi_1^2}(W), \qquad R_1 = Q_{\chi_1^2}(1-W), \qquad W \sim \mathrm{Unif}[0,1]. \tag{60}$$

Assign independent fair signs to their square roots and multiply the second by s. The resulting endpoints have exactly the required Gaussian laws and separate sign symmetry. To prove optimality,

in a finite equal-mass discretization, exchanging two similarly ordered pairs lowers product cost because

$$ab + a'b' - (ab' + a'b) = (a' - a)(b' - b) \geq 0 \quad (a \leq a',\ b \leq b').$$

Repeated exchanges give reverse order. Approximate the continuous nonnegative marginals by truncated quantiles and pass to the limit using their finite second moments and uniform integrability. This gives (60). At $K = 1$,

$$\mathcal{S}_1(\pi) = \frac{3}{2} + \frac{\sqrt{17}}{5} + \frac{1}{5}\mathbb{E}R_0R_1.$$

The one-dimensional integral in (60) is approximately 0.1213432189, giving coefficient 2.348889769. The quoted quadrature error estimate is numerical, not an interval-arithmetic certificate. This coupling is singular as a joint law and is *not* asserted to solve the unrestricted quartic problem, in which the two cubic mixed moments can also vary. The smooth full-support construction proves that statistical gains do not require this singular comparator.

H.3 An observable finite-candidate guarantee

Suppose M candidate couplings share the same population-correct parameter and meet the uniform assumptions of Proposition 2. They may have different Gram matrices if those matrices are included correctly in each score. Let S_j be (6), and suppose an independent pilot procedure returns $\widehat{S}_j$ satisfying

$$\max_{1 \leq j \leq M} |\widehat{S}_j - S_j| \leq \epsilon.$$

Choose $\widehat{j} \in \arg\min_j \widehat{S}_j$, then train on n fresh triples from that coupling. On the pilot event,

$$\mathfrak{R}_{n,K}(\pi_{\widehat{j}}) \leq \min_j \mathfrak{R}_{n,K}(\pi_j) + \frac{2\epsilon}{n} + 2Cn^{-3/2}. \tag{61}$$

Indeed $S_{\widehat{j}} \leq \min_j S_j + 2\epsilon$, and the uniform expansion converts scores to fresh-training risks. No training/selection independence is silently assumed: it is enforced by the fresh sample requirement.

One concrete pilot estimator is median-of-means on the observable squared residual cost. Suppose its variance is at most v^2 for each candidate and m pilot triples are available per candidate. Take an odd number of blocks $b \geq 8\log(2M/\delta)$, with $m \geq 2b$, and use $\lfloor m/b \rfloor$ observations in each block. A block mean deviates by more than $2v\sqrt{2b/m}$ with probability at most $1/4$, by Chebyshev. The probability that at least half the independent blocks fail is at most $e^{-b/8}$, by a Bernoulli Hoeffding bound. A union bound therefore gives, with probability at least $1 - \delta$,

$$\epsilon = 2v\sqrt{2b/m}.$$

This uses Mm pilot triples in addition to the final training set. Errors from estimating $G, \theta_\star, Q_K$ are not included in this simple display and must be added to ϵ when those quantities are not analytically available.

I Executed numerical checks and reproducibility

I.1 Design and accounting

All numerical experiments are synthetic and low-dimensional. Endpoints and times are sampled independently across triples and across training replications. There is no optimization loop: each scalar model is fitted by the exact empirical least-squares ratio and clipped to $[0, 3]$. Each generated Gaussian distribution is evaluated by the exact squared difference of its standard deviation from the target scale; no empirical Wasserstein estimator or extra generation sample is used.

For each $K \in \{1, 8\}$ and $n \in \{32, 128, 512\}$, we compare the independent Gaussian at its sampler-calibrated s_K, the smooth density with $\tau = 3/4$, and the antimonotone-radius comparator. There are $65{,}536$ replications per configuration for $K = 1$ and $8{,}192$ for $K = 8$, totaling $663{,}552$ independent fits and $148{,}635{,}648$ training triples. These counts describe only completed runs included in the archive. Preliminary smoke tests and interrupted larger runs are excluded from the reported results.

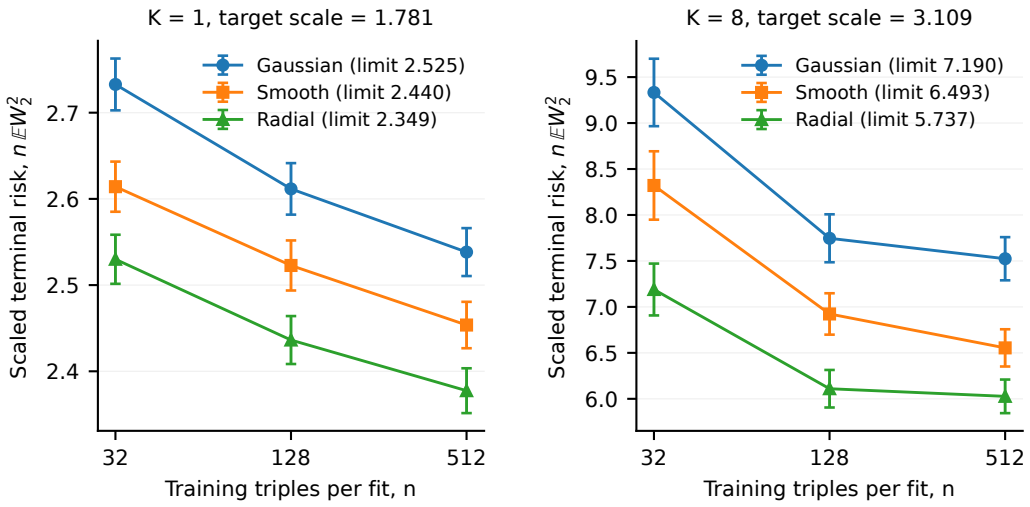

Figure 2: Retained original experiments. Gaussian, explicit smooth, and antimonotone-radius couplings at $K = 1, 8$. Bars are approximate 95% Monte Carlo intervals; legend entries give analytic first-order coefficients. Endpoint scales differ between panels. These data are separate from the new cell-design experiments in Figure 1.

The smooth density is sampled by rejection from independent Gaussian endpoints: accept with probability $(1 - \tau h(x)h(y/s))/(1 + \tau)$. Its acceptance probability is exactly $1/(1 + \tau) = 4/7$, since the centered perturbation integrates to zero. Antimonotone radii are sampled by Gaussian quantiles at $(1 + W)/2$ and $1 - W/2$, then independent fair signs. The implementation records seeds, the clipping count, each replicate's terminal loss, software versions, and elapsed execution times. Endpoint-level draws are regenerated from seeds rather than retained.

Error bars are mean $\pm$ 1.96 sample standard deviation$/\sqrt{\text{replications}}$ for the scaled terminal risks. These are approximate Monte Carlo confidence intervals for the training-risk expectation, not intervals describing variability of a single trained model. The Gaussian simulation baseline is population-calibrated, not the unknown finite-n best Gaussian correlation. Theorems 4 and 3, rather than the simulations, justify statements about such oracles.

I.2 Full Monte Carlo table

The scaled risks approach their analytic coefficients with finite-sample bias and Monte Carlo fluctuations. We do not require every observed point to fall within a confidence interval centered on the asymptotic coefficient: the coefficient is a limit, not the exact expectation at that n. Nor does the experiment establish a uniform small-n ranking theorem.

I.3 Sampler-dependent analytic coefficients

The scales s_K are numerically computed roots of an equation with analytically proved unique positive solution. The table compares different endpoint scales across budgets; it must not be read as a same-distribution efficiency comparison between samplers. The $K = \infty$ row is an exact-ODE mathematical risk coefficient, not a numerical solver run.

I.4 Discrete transport certificates

For $m \in \{16, 32, 64\}$, take m standard Gaussian midpoint quantiles and rescale them to have empirical variance exactly one; use s_1 times these nodes for the target. Both marginal weight vectors are uniform. The LP minimizes the quartic coefficient with covariance exactly zero in its mathematical specification. All pair costs are integrated analytically in time. The independent comparison uses exactly the same grid, marginal weights, and normalization. The archive stores

Table 1: Full simulation results. Values are n times terminal risk; $\pm$ denotes one Monte Carlo standard error. The final column is the first-order limit, not a fitted value.

K	n	Coupling	Scaled risk $\pm$ SE	Limit
1	32	Gaussian	2.7328 ± 0.0153	2.5246
1	32	Smooth	2.6142 ± 0.0148	2.4402
1	32	Radial	2.5299 ± 0.0145	2.3489
1	128	Gaussian	2.6116 ± 0.0152	2.5246
1	128	Smooth	2.5228 ± 0.0148	2.4402
1	128	Radial	2.4363 ± 0.0142	2.3489
1	512	Gaussian	2.5383 ± 0.0142	2.5246
1	512	Smooth	2.4537 ± 0.0137	2.4402
1	512	Radial	2.3775 ± 0.0133	2.3489
8	32	Gaussian	9.3330 ± 0.1872	7.1903
8	32	Smooth	8.3209 ± 0.1898	6.4926
8	32	Radial	7.1892 ± 0.1439	5.7372
8	128	Gaussian	7.7464 ± 0.1331	7.1903
8	128	Smooth	6.9232 ± 0.1148	6.4926
8	128	Radial	6.1097 ± 0.1042	5.7372
8	512	Gaussian	7.5240 ± 0.1198	7.1903
8	512	Smooth	6.5541 ± 0.1033	6.4926
8	512	Radial	6.0268 ± 0.0932	5.7372

Table 2: Sampler-calibrated scales and analytic first-order risk coefficients. The gap is against the entire Gaussian family in the asymptotic theorem.

K	s_K	Gaussian	Smooth	Gap
1	1.780776	2.524621	2.440246	0.084375
2	2.267106	3.645892	3.405789	0.240103
4	2.744304	5.282527	4.811445	0.471082
8	3.109358	7.190316	6.492641	0.697675
16	3.343294	8.836067	7.971381	0.864686
64	3.549310	10.624585	9.599146	1.025439
∞	3.624946	11.371310	10.283701	1.087609

the primal coupling matrix, equality-constraint dual vector, objective matrix, and endpoint nodes for each grid. Numerical checks recompute marginal errors, covariance residual, nonnegativity, dual slacks, and the primal–dual objective gap. The reported residuals are floating-point diagnostics. The Gaussian midpoint grid does not reproduce the Gaussian fourth moment exactly, so its independent coefficient differs from 2.524621125 even after variance normalization. No error bound transferring these discrete optima to the continuous Gaussian optimum is claimed.

I.5 Exact and numerical verification inventory

The symbolic script independently integrates the quartic residual and bias polynomials; checks the Gaussian-scale identities and the exact $27/320$ constant; evaluates the smooth-density moments; factors the open-family gap; verifies all four rational principal minors, the normal equations, and the exact endpoint of the noncommuting matrix construction; and verifies the raw-loss monotonicity certificate. It also uses rational interval bounds for the scalar noise-corrected and exact-ODE reversals, and rational arithmetic for the realizable three-step example. Numerical checks separately test root calibration for seven budgets, the open-family formulas on a 41-point grid, and the local Bures quadratic form by a finite perturbation. The distinction between exact checks and numerical corroboration is recorded in the machine-readable output.

The proof of the uniform expansion, constrained-optimum attainment, and coupling-oracle regret is mathematical, not inferred from simulations or symbolic spot checks. All tests passing is evidence of implementation consistency, not an independent formal proof certificate.

Table 3: Quartic transport on fixed empirical endpoint grids. Residual is the largest of marginal/covariance equality errors and dual feasibility violations; certificates use floating-point arithmetic.

m	Variables	Independent	Optimized	Reduction	Residual
16	256	2.084862	1.261523	39.49%	3.4×10^{-15}
32	1024	2.241414	1.280991	42.85%	8.9×10^{-16}
64	4096	2.347878	1.291028	45.01%	1.8×10^{-15}

I.6 Reproduction commands

From the archive root, install the pinned numerical dependencies, then run:

```
python code/verify_exact.py
python code/run_experiments.py --deterministic-only
python code/run_experiments.py --case 0
# Repeat --case 1 through --case 5.
python code/assemble_results.py
python code/make_figures.py
bash build.sh
```

The complete rerun is optional for compiling the manuscript: retained CSV/NPZ outputs and pre-rendered figures are included. A shorter smoke run writes separate outputs through `-quick`. No internet access is needed to execute the numerical code or compile from the included sources once its dependencies are installed.

J A realizable temporal reference and its coupling oracle

We prove Theorem 6, extend the global coupling oracle to this protocol, and give a coordinatewise extension. Throughout $s > 1$ is fixed. Define q_s, b_s by (21), $r_s(t) = (1+s^2)t-1$, and $w(t) = t(1-t)$.

J.1 Population training, stabilization, and unique calibration

For an arbitrary coupling with covariance $c \in [-s, s]$,

$$\mathbb{E}X_t^2 = q_s(t) + 2cw(t), \qquad \mathbb{E}UX_t = r_s(t) + cw'(t).$$

Under the independent Gaussian coupling, Gaussian conditioning gives $\mathbb{E}[U \mid X_t, t] = b_s(t)X_t$. The identity $b_s = q_s'/(2q_s)$ gives

$$\int_0^1 b_s(t)\mathrm{d}t = \tfrac{1}{2}\log\frac{q_s(1)}{q_s(0)} = \log s.$$

Thus exact integration of $\dot{x} = \theta b_s(t)x$ produces $s^\theta x$, and the independent Gaussian reference is both pathwise realizable and population-correct at $\theta = 1$.

Put

$$A_s = \int_0^1 b_s(t)w'(t)\mathrm{d}t, \qquad B_s = 2\int_0^1 b_s(t)^2 w(t)\mathrm{d}t, \qquad G_s = \int_0^1 b_s(t)^2 q_s(t)\mathrm{d}t.$$

The population Gram matrix and normal-equation right side are $G_s + cB_s$ and $G_s + cA_s$. Therefore

$$\theta(c) = \frac{G_s + cA_s}{G_s + cB_s}. \tag{62}$$

A uniform positive lower bound for the Gram matrix is

$$g_s = \int_0^1 b_s(t)^2\big((1-t) - st\big)^2\mathrm{d}t > 0, \qquad M_s = 1 + \frac{s+1}{\sqrt{g_s}}, \qquad \Theta_s = [-M_s, M_s]. \tag{63}$$

Indeed $\mathbb{E}X_t^2 \geq ((1-t)-st)^2$, and the integrand defining g_s is positive except at finitely many times. Cauchy–Schwarz and $\mathbb{E}U^2 \leq (s+1)^2$ give $|\theta(c)| \leq (s+1)/\sqrt{g_s}$, so every population parameter is at least distance one from the boundary of Θ_s. The estimator used in the theorem and experiments is

$$\widehat{\theta} = \mathrm{clip}_{\Theta_s}\left(\frac{\sum_{i=1}^n b_s(t_i)X_{t_i}U_i}{\sum_{i=1}^n b_s(t_i)^2 X_{t_i}^2}\right). \tag{64}$$

A zero denominator is assigned parameter zero; independent continuous time makes this event have probability zero. The profile is bounded for each fixed s, and the Gaussian marginals give all uniform polynomial moments. Thus Proposition 2, including its $O(n^{-2})$ refinement, applies over all endpoint couplings.

Integration by parts, using $w(0) = w(1) = 0$ and $b_s' = (1+s^2)/q_s - 2b_s^2$, yields

$$A_s - B_s = -\int_0^1 \left(b_s' + 2b_s^2\right) w \mathrm{d}t = -(1+s^2)\int_0^1 \frac{w(t)}{q_s(t)}\mathrm{d}t < 0.$$

Consequently

$$\theta'(c) = \frac{G_s(A_s - B_s)}{(G_s + cB_s)^2} < 0,$$

and $c = 0$ is the unique covariance giving $\theta = 1$ and the correct endpoint. On the compact interval $[-s, s]$, the derivative of $s^{\theta(c)}$ is bounded away from zero in magnitude. The mean value theorem gives

$$F_s(c) := (s^{\theta(c)} - s)^2 \geq \kappa_s c^2 \tag{65}$$

for a fixed $\kappa_s > 0$.

J.2 Exact quartic cost and statistical constants

Define the nonnegative homogeneous quartic

$$\widetilde{J}_s(x,y) = \int_0^1 \left[b_s(t)((1-t)x + ty)\{y - x - b_s(t)((1-t)x + ty)\}\right]^2 \mathrm{d}t. \tag{66}$$

Writing $\alpha = 1 - t$ and $\beta = t$, the residual identity is

$$y - x - b_s(t)(\alpha x + \beta y) = \frac{-s^2\beta x + \alpha y}{q_s(t)}.$$

Direct integration gives

$$\begin{aligned}\widetilde{J}_s(x,y) = &\left[\frac{\pi s(1+s^2)}{32} - \frac{s^2}{6}\right]x^4 + \frac{s^2-1}{6}x^3 y + \frac{\pi(1+s^2)}{16s}x^2y^2 \\ &+ \frac{1-s^2}{6s^2}xy^3 + \left[\frac{\pi(1+s^2)}{32s^3} - \frac{1}{6s^2}\right]y^4.\end{aligned} \tag{67}$$

For completeness, the substitution $z = st/(1-t)$ transforms the coefficient of x^2y^2 into

$$\frac{1}{s}\int_0^\infty \frac{(sz-1)^2(z^4 - 4z^2 + 1)}{(1+z^2)^4}\mathrm{d}z = \frac{\pi(1+s^2)}{16s}.$$

All five coefficients follow from the same substitution and the elementary beta integrals

$$\int_0^\infty \frac{z^j}{(1+z^2)^4}\mathrm{d}z = \frac{\Gamma((j+1)/2)\Gamma((7-j)/2)}{12}, \qquad 0 \leq j \leq 6.$$

These identities are also checked symbolically in the accompanying code.

The relations $r_s^2 = (1+s^2)q_s - s^2$ and $\int_0^1 q_s^{-1}\mathrm{d}t = \pi/(2s)$ give $G_s = 1 + s^2 - \pi s/2$. Integrating $b_s' = (1+s^2)/q_s - 2b_s^2$ and using $b_s(0) = -1$, $b_s(1) = 1$ gives

$$\int_0^1 b_s(t)^2 \mathrm{d}t = \frac{\pi(1+s^2)}{4s} - 1 = I_s > 0.$$

At a calibrated coupling, the terminal quadratic form is $Q_s = s^2(\log s)^2$, hence the score is

$$\widetilde{S}_s(\pi) = \frac{s^2(\log s)^2}{G_s^2}\mathbb{E}_\pi \widetilde{J}_s(X_0, X_1). \tag{68}$$

For the independent Gaussian, $U - b_s(t)X_t$ is independent of X_t conditional on time and has variance $s^2/q_s(t)$. Thus its residual-score moment is $s^2 I_s$, proving the expression S_s^G in (23).

The density $p_{3/4}^{(s)}$ has covariance zero and therefore the identical population objective. By (44), it changes only the x^2y^2 entry among the relevant mixed fourth moments, decreasing it by $27s^2/64$. Since the coefficient of this monomial in (67) equals $(I_s + 1)/4$,

$$\Delta_s = \frac{s^2(\log s)^2}{G_s^2}\frac{I_s + 1}{4}\frac{27s^2}{64} = S_s^G \frac{27}{256}\left(1 + \frac{1}{I_s}\right) > 0.$$

The fourth-order uniform expansion gives $\widetilde{\mathfrak{R}}_n(\pi) = F_s(c) + T_s(\pi)/n + O(n^{-2})$. Along the Gaussian family $T_s(\pi_\rho^G)$ is smooth on $[-1, 1]$ because Gaussian fourth moments are polynomials in $s\rho$ and the Gram matrix is bounded below. Combining its Lipschitz bound with (65), exactly as in Appendix E, gives

$$\inf_\rho \widetilde{\mathfrak{R}}_n(\pi_\rho^G) = \frac{S_s^G}{n} + O(n^{-2}), \qquad \widetilde{\mathfrak{R}}_n(p_{3/4}^{(s)}) = \frac{S_s^G - \Delta_s}{n} + O(n^{-2}).$$

This proves Theorem 6, including comparisons to n-dependent Gaussian correlations.

J.3 The temporal global coupling oracle

Theorem 10 (Global oracle for the temporal protocol)**.** *For the fixed temporal class, exact sampler, and estimator* (64)*, let*

$$\widetilde{S}_\star = \min_{\substack{\pi \in \Pi(\mu,\nu) \\ \mathbb{E}_\pi X_0 X_1 = 0}} \widetilde{S}_s(\pi).$$

The minimum is attained by some $\widetilde{\pi}_\star$, is strictly positive, and

$$\inf_{\pi \in \Pi(\mu,\nu)} \widetilde{\mathfrak{R}}_n(\pi) = \frac{\widetilde{S}_\star}{n} + O(n^{-2}), \qquad 0 \le \widetilde{\mathfrak{R}}_n(\widetilde{\pi}_\star) - \inf_{\pi \in \Pi(\mu,\nu)} \widetilde{\mathfrak{R}}_n(\pi) = O(n^{-2}).$$

Proof. The weak compactness and moment-continuity argument of Appendix D applies verbatim to (66). This cost vanishes only at $(0, 0)$: its integrand at $t = 0$ forces $xy = 0$; when exactly one endpoint is nonzero, the integrand is positive on an interval inside $(0, 1)$. Positivity and compact attainment therefore give $\widetilde{S}_\star > 0$.

Extend (68) linearly to all couplings, holding its calibrated coefficients fixed. Mixing with the perfectly correlated or antimonotone coupling repairs covariance to zero with weight at most $|c|/s$. Uniform fourth moments imply

$$\widetilde{S}_s(\pi) \ge \widetilde{S}_\star - L_0|c|.$$

The coefficients of $T_s(\pi)$ are smooth functions of c multiplying endpoint moments of degree four; b_s is fixed and bounded, and $G_s + cB_s$ stays positive. At $c = 0$, $T_s = \widetilde{S}_s$, since the terminal gradient vanishes. Hence

$$|T_s(\pi) - \widetilde{S}_s(\pi)| \le L_1|c|.$$

Combining these inequalities with (65) and the uniform $O(n^{-2})$ expansion yields

$$\widetilde{\mathfrak{R}}_n(\pi) \ge \kappa_s c^2 + \frac{\widetilde{S}_\star - (L_0 + L_1)|c|}{n} - Cn^{-2}.$$

Complete the square, and evaluate the upper bound at $\widetilde{\pi}_\star$. This proves both assertions, including n-dependent competitors. □

J.4 Coordinatewise extension against all Gaussian cross-covariances

For fixed d, use endpoint laws $\mathcal{N}(0, I_d)$ and $\mathcal{N}(0, s^2 I_d)$ and the fixed class

$$v_\theta(x,t) = b_s(t)\,\mathrm{diag}(\theta_1,\dots,\theta_d)x.$$

Least squares separates into d estimators of the form (64), with coordinatewise clipping. Exact integration gives covariance $\mathrm{diag}(s^{2\widehat{\theta}_1},\dots,s^{2\widehat{\theta}_d})$, so terminal W_2^2 is the sum of the d scalar risks.

For any jointly Gaussian endpoint coupling, coordinate pair j is Gaussian with some correlation $\rho_j \in [-1,1]$. Off-diagonal cross-covariances may correlate parameter estimates but do not affect the expectation of this sum. Therefore the infimum over *all* jointly Gaussian couplings equals $d\inf_\rho \widetilde{\mathfrak{R}}_n(\pi^G_\rho)$: the lower bound holds coordinatewise, and block-diagonal cross-covariance realizes the upper bound. The product of d copies of $p^{(s)}_{3/4}$ consequently has gap

$$\frac{d\Delta_s}{n} + O_d(n^{-2}).$$

This product coupling has smooth, positive density and the same entire population objective as independent Gaussian endpoints. Similarly, the all-coupling oracle infimum tensorizes, because any coordinate-pair marginal is an admissible scalar coupling and products realize any selected scalar design. The model is coordinatewise; this is not a claim for unrestricted learned matrix fields.

K Exact-marginal cell transport, smoothing, and oracle regret

We prove Theorem 7. The cost d is either $R'_K(a_\star)^2 J_{a_\star}/G_\star^2$ or the temporal cost in (68); both are continuous, nonnegative homogeneous quartics. The endpoint laws remain $\mathcal{N}(0,1)$ and $\mathcal{N}(0,s^2)$ throughout this section.

K.1 Finite feasibility and nested continuous families

Let A_i be standard Gaussian quantile cells with probability $1/m$, and $B_i = sA_i$. Put $\mu_i = \mu(\cdot \mid A_i)$ and $\nu_i = \nu(\cdot \mid B_i)$, with means u_i and $v_i = su_i$. The row and column constraints in (24) give exactly the original marginal laws. Conditional independence within a cell pair gives

$$\mathbb{E}_{\pi_P} X_0 X_1 = \sum_{ij} P_{ij} u_i v_j.$$

The diagonal and antidiagonal cell matrices, each with m entries equal to $1/m$, have covariances sa_m and $-sa_m$, where

$$a_m = \frac{1}{m}\sum_{i=1}^m u_i^2 = \mathbb{E}\big(\mathbb{E}[Z \mid I_m(Z)]\big)^2 \longrightarrow 1.$$

This convergence holds as equal-probability cells refine, and in particular along dyadic partitions. Since $|c_\star| < s$, convex mixtures of these two matrices make the LP feasible for every sufficiently large m. The finite feasible polytope is compact, and the conditional quartic costs are finite, so a minimizer exists.

For a dyadic refinement, each conditional law is the equal mixture of its two children. A coarse matrix P therefore represents exactly the same continuous joint law at the finer resolution by assigning $P_{ij}/4$ to each of its four child cell pairs. The continuous feasible families are nested, implying $S_{2m} \le S_m$. Since they are contained in the full calibrated coupling set, $S_m \ge S_\star$.

K.2 Approximation and exact covariance repair

Write $\bar{X}_m = \mathbb{E}[X \mid I_m(X)]$, and define $\bar{Y}_m$ analogously. Conditional quantile means converge in L^4 to the Gaussian variables. One direct justification is to truncate the quantile function away from probability zero and one, approximate the resulting bounded continuous function on the compact interval, and then control the tails with Jensen's inequality and the Gaussian fourth moment. This proves

$$\|X - \bar{X}_m\|_{L^4} + \|Y - \bar{Y}_m\|_{L^4} = \delta_m \longrightarrow 0.$$

For any coupling π, keep its cell-pair masses, but redraw $X^{[m]}, Y^{[m]}$ independently from the two conditional marginal laws given that cell pair. This realizes the lift $\pi^{[m]}$. Both the original and redrawn endpoints have the same marginal cell-error moments, so Minkowski's inequality gives

$$\|X - X^{[m]}\|_{L^4} + \|Y - Y^{[m]}\|_{L^4} \le 2\delta_m.$$

Cauchy–Schwarz therefore bounds covariance displacement by $C_s\delta_m$. A quartic polynomial satisfies

$$|d(z) - d(z')| \le C_d\|z - z'\|(\|z\| + \|z'\|)^3.$$

Hölder's inequality and the fixed marginal fourth moments imply

$$|\mathbb{E}_\pi XY - \mathbb{E}_{\pi^{[m]}} XY| \le C\delta_m, \qquad |\mathbb{E}_\pi d - \mathbb{E}_{\pi^{[m]}} d| \le C\delta_m, \tag{69}$$

uniformly over all endpoint couplings.

Apply this lift to a continuous optimizer $\pi_\star$, and denote the lifted covariance by c_m. For large m, $sa_m - |c_\star| \ge (s - |c_\star|)/2$. If $c_m > c_\star$, mix the lifted matrix with the antidiagonal matrix using weight

$$\alpha = \frac{c_m - c_\star}{c_m + sa_m} \le \frac{2|c_m - c_\star|}{s - |c_\star|}.$$

If $c_m < c_\star$, use the diagonal matrix with weight $(c_\star - c_m)/(sa_m - c_m)$, obeying the same bound. The result remains a conditional-Gaussian cell lift and now has covariance *exactly* $c_\star$. The costs of all such lifts are uniformly bounded by their Gaussian marginal fourth moments. Hence the repair adds at most $C\delta_m$ to the objective. Together with (69), this gives

$$0 \le S_m - S_\star \le C\delta_m.$$

All these statements concern the exact mathematical LP. Floating-point implementations record feasibility and primal–dual residuals separately.

K.3 Fresh-training regret and approximate LP solutions

For every calibrated lift, the appropriate uniform scalar expansion gives

$$\mathfrak{R}_n(\pi_m) = \frac{S_m}{n} + O(n^{-2}).$$

Theorem 3 or Theorem 10 gives the matching all-coupling infimum $S_\star/n + O(n^{-2})$. Their constants are uniform over the full Gaussian-marginal coupling family, hence independent of the cell count. Subtraction proves (25). If a feasible LP solution is within η_m of S_m, its statistical regret increases by at most η_m/n.

The quantization error can itself be computed from the conditional moments. If $m_{r,i} = \mathbb{E}[Z^r \mid A_i]$, then

$$\delta_m^4 = (1+s)^4 \frac{1}{m} \sum_i \left(m_{4,i} - 4m_{1,i}m_{3,i} + 6m_{1,i}^2 m_{2,i} - 3m_{1,i}^4\right).$$

Choosing a deterministic resolution with $\delta_{m(n)} = O(n^{-1})$ and an LP objective tolerance $\eta_{m(n)} = O(n^{-1})$ recovers the $O(n^{-2})$ oracle regret. This is a statistical approximation guarantee, not a polynomial-time bound for the required resolution.

K.4 Smooth, full-support oracle approximations at zero covariance

Let $\mathbb{E}XY = 0$ and introduce the independent Gaussian smoothing from Section 5. Its marginals are exactly $\mathcal{N}(0,1)$ and $\mathcal{N}(0,s^2)$; its covariance remains zero. For $0 < \epsilon < 1$, convolution with a nondegenerate Gaussian gives a smooth, everywhere positive joint density. Put $r = \sqrt{1-\epsilon}$. Expanding independent Gaussian moments gives

$$\mathbb{E}(X^\epsilon)^3 Y^\epsilon = r^4 \mathbb{E}X^3Y, \qquad \mathbb{E}X^\epsilon (Y^\epsilon)^3 = r^4 \mathbb{E}XY^3,$$
$$\mathbb{E}(X^\epsilon)^2 (Y^\epsilon)^2 = r^4 \mathbb{E}X^2Y^2 + (1 - r^4)s^2.$$

The marginal fourth moments stay fixed. The extra terms in the two cubic-mixed moments are proportional to $\mathbb{E}XY$, hence vanish. Applying these identities to any homogeneous quartic score proves (26).

In particular, $0 \le S(\pi_{m,\epsilon}) - S_\star \le C\delta_m + 2\epsilon(S_G - S_\star)$. Its trained regret is at most

$$\frac{C\delta_m + 2\epsilon(S_G - S_\star) + \eta_m}{n} + \frac{C'}{n^2},$$

with the η_m term included for a feasible approximate cell optimizer. Thus letting $\delta_m, \epsilon, \eta_m = O(n^{-1})$ yields smooth full-support $O(n^{-2})$ oracle approximations. With the fixed experimental value $\epsilon = 0.01$, the smoothing instead retains exactly 98.01% of the cell optimizer's first-order advantage over independence.

K.5 Analytic cell costs and sampling

Let z_i be standard Gaussian quantile boundaries and φ its density. Conditional moments obey

$$m_{0,i} = 1, \quad m_{1,i} = m[\varphi(z_{i-1}) - \varphi(z_i)],$$

$$m_{r,i} = m[z_{i-1}^{r-1}\varphi(z_{i-1}) - z_i^{r-1}\varphi(z_i)] + (r-1)m_{r-2,i}, \qquad r = 2, 3, 4,$$

where the Gaussian boundary terms at infinity are zero. For $d(x, y) = \sum_{r=0}^{4} d_r x^{4-r} y^r$, the cell cost is $\sum_r d_r m_{4-r,i} s^r m_{r,j}$. No point-mass replacement or time quadrature is used for either protocol.

To sample a lift, draw (i, j) with probabilities P_{ij}, then independent $U, V \sim \mathrm{Unif}[0, 1]$ and set

$$X = \Phi^{-1}\left(\frac{i - 1 + U}{m}\right), \qquad Y = s\Phi^{-1}\left(\frac{j - 1 + V}{m}\right).$$

Here Φ is the standard Gaussian distribution function. The implementation then adds independent Gaussian smoothing. The LP and its smoothing parameter are chosen before all subsequent training data are drawn.

L Exact-marginal design experiments and additional verification

L.1 Protocols, computation, and data accounting

The experiments in Figure 1 are new, and are retained separately from the original experiments in Appendix I. They use $16{,}384$ independent replications for each of three sample sizes, three couplings, and two protocols, giving $294{,}912$ fits and $66{,}060{,}288$ training triples. Each training time is uniform on $[0, 1]$ and independent of its endpoint pair. All endpoint pairs are independent across training triples. LP construction precedes sampling; there is no pilot selection on the evaluation training sets.

The constant-field protocol uses $s = (3 + \sqrt{17})/4$, one Euler step, and clipping to $[0, 3]$. The temporal protocol uses $s = 2$, the profile (21), exact integration, and the fixed symmetric interval (63), numerically $[-3.7755111467, 3.7755111467]$. These are distinct fixed protocols, so cross-panel risk values are not a comparison of model or sampler efficiency.

The Gaussian baseline is independent and population-calibrated in both protocols. The explicit smooth coupling is (15) with $\tau = 3/4$. The optimized smooth coupling is a 64-cell conditional-Gaussian lift followed by independent Gaussian smoothing with $\epsilon = 0.01$. Linear programs also run at $m = 16, 32, 128$ to evaluate the construction at different resolutions. The time-dependent quartic coefficients and all conditional Gaussian moments are analytic; no numerical time quadrature estimates the LP costs. The one-dimensional integral defining the clipping radius is evaluated numerically with absolute and relative tolerance 10^{-12}.

The programs use SciPy's HiGHS linear-program solver with primal and dual feasibility tolerances 10^{-9}. Retained solutions have much smaller observed residuals, reported in Table 5. Joint sign symmetrization of the solution preserves the marginal constraints, covariance, and homogeneous quartic objective. The sampler uses inverse Gaussian distribution functions within selected cells, followed by independent Gaussian perturbations. Exact marginal statements refer to the mathematical construction; the implementation uses ordinary floating-point arithmetic and records feasibility residuals.

For each fitted parameter, terminal W_2^2 is evaluated as the exact squared standard-deviation difference: $(1+\widehat{a}-s)^2$ or $(s^{\widehat{\theta}}-s)^2$. There is no empirical distribution-distance estimator or additional generation

Table 4: New fresh-training results. Values are n times terminal risk; $\pm$ is one Monte Carlo standard error. All configurations use $16{,}384$ fits. The last column is computed from the coupling moments, not fitted to these data.

Protocol	Coupling	n	Scaled risk $\pm$ SE	Limit
Constant	Gaussian	32	2.6500 ± 0.0296	2.5246
Constant	Explicit smooth	32	2.5812 ± 0.0294	2.4402
Constant	Optimized smooth	32	1.9423 ± 0.0283	1.3920
Constant	Gaussian	128	2.5744 ± 0.0297	2.5246
Constant	Explicit smooth	128	2.5274 ± 0.0300	2.4402
Constant	Optimized smooth	128	1.5242 ± 0.0185	1.3920
Constant	Gaussian	512	2.5473 ± 0.0285	2.5246
Constant	Explicit smooth	512	2.4189 ± 0.0268	2.4402
Constant	Optimized smooth	512	1.4377 ± 0.0164	1.3920
Temporal	Gaussian	32	2.6467 ± 0.0389	2.1446
Temporal	Explicit smooth	32	2.0489 ± 0.0289	1.6836
Temporal	Optimized smooth	32	1.0635 ± 0.0173	0.7404
Temporal	Gaussian	128	2.2611 ± 0.0266	2.1446
Temporal	Explicit smooth	128	1.7503 ± 0.0201	1.6836
Temporal	Optimized smooth	128	0.8120 ± 0.0099	0.7404
Temporal	Gaussian	512	2.1656 ± 0.0242	2.1446
Temporal	Explicit smooth	512	1.7085 ± 0.0191	1.6836
Temporal	Optimized smooth	512	0.7479 ± 0.0084	0.7404

Table 5: Cell transport under the original continuous Gaussian marginals. The lifted LP score is before smoothing; the smooth score uses $\epsilon = 0.01$. Reduction compares the smooth score to the independent Gaussian under the same protocol. Feasibility diagnostics use floating-point arithmetic.

Protocol	m	Lifted LP	Smooth	Reduction	Residual
Constant	16	1.543050	1.562583	38.11%	8.9×10^{-16}
Constant	32	1.416553	1.438604	43.02%	1.8×10^{-15}
Constant	64	1.369007	1.392004	44.86%	1.7×10^{-15}
Constant	128	1.349515	1.372899	45.62%	2.7×10^{-15}
Temporal	16	0.726763	0.754977	64.80%	2.8×10^{-16}
Temporal	32	0.717096	0.745503	65.24%	1.1×10^{-15}
Temporal	64	0.711879	0.740389	65.48%	3.2×10^{-15}
Temporal	128	0.709475	0.738033	65.59%	4.4×10^{-16}

noise. Error bars are 1.96 times the sample standard deviation of the scaled replicate risks divided by the square root of the replication count. They quantify Monte Carlo uncertainty in the expectation, not a guarantee for one trained generator. Neither the plots nor the asymptotic theorems assert every-small-n optimality of the lifted design.

L.2 Reproducibility, exact algebra, and limitations of the checks

The archive records all seeds, sample counts, clipping counts, per-fit terminal losses, LP primal and dual vectors, conditional moments, coefficients, and software versions. Random seeds are $27091601 + 10000i + 100j + k$, where protocol $i \in \{0, 1\}$, sample-size index $j \in \{0, 1, 2\}$, and coupling index $k \in \{0, 1, 2\}$. Arrays are retained in `revision_replicate_risks.npz`; the CSV tables and metadata files are generated directly from these arrays. Original data are not overwritten or pooled with the new runs.

The additional symbolic script verifies the temporal residual identity, its transformed integral coefficients, the Gaussian score and relative gap, and the smoothing moment identities. Numerical tests independently compare the temporal coefficients to direct quadrature, recheck the continuous-cell

primal–dual certificates and Gaussian marginal moments, and replay retained training losses from their seeds. The uniform statistical expansion, covariance repair, and oracle approximation are proved mathematically; they are not inferred from simulations or from successful symbolic spot checks.

From the archive root, the additional results are reproduced by:

```
python code/verify_revision.py
python code/run_revision_experiments.py
python code/assemble_revision.py
python -m pytest -q code
bash build.sh
```

The original commands in Appendix I remain available. Compiling the paper only needs the retained tables and figures, not a rerun of any experiment.