LIMITS OF FIXED SCORE SCALING IN DIFFUSION MODELS UNDER GAUSSIAN MIXTURES

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion-based image reconstruction involves a perception–distortion (PD) tradeoff. Fixed score scaling adjusts this tradeoff without retraining by applying the same score multiplier at every noise level. For Gaussian sources, doubling the score in the variance-preserving probability-flow ordinary differential equation recovers the posterior mean and minimizes mean-squared error (MSE). However, this result does not guarantee posterior-mean recovery for non-Gaussian sources. For every nontrivial finite Gaussian mixture model (GMM) with positive weights and positive-definite component covariances, we prove a positive lower bound on excess MSE over all real constant multipliers. The bound extends to deterministic denoising diffusion implicit model (DDIM) sampling on sufficiently fine grids over expanding multiplier intervals. To approximate the posterior mean, we propose Simple Endpoint Correction, which combines two constant-scaling endpoints with coefficients selected on a calibration set. Experiments show lower held-out negative log-likelihood for GMMs than for single Gaussians on the evaluated image latents. Endpoint correction reduces posterior-mean error on synthetic GMMs and image-space MSE for DiffC-SD2.1, ResULIC, and ELIC+DDIM/DSG.

1 INTRODUCTION

Diffusion models reconstruct images through their reverse process. Reconstruction involves a perception–distortion (PD) tradeoff, where improving perceptual quality can increase distortion (Blau & Michaeli, 2018). This tradeoff can be adjusted at inference time by modifying the reverse process without retraining (Wang et al., 2025; 2026).

One such modification is fixed score scaling. For a noisy-data density $q_t(x)$, the score $s_t(x) = \nabla_x \log q_t(x)$ is its log-density gradient with respect to the state x. Score scaling replaces $s_t(x)$ with $\lambda s_t(x)$ in the reverse process and fixed scaling uses the same scalar λ at every noise level.

Prior work establishes optimality for inference-time controls under Gaussian assumptions. Wang et al. (2025) scale reverse-sampling variance and prove optimality for Gaussian posteriors under a conditional PD criterion. For score scaling, Wang et al. (2026) show that $\lambda = 2$ in the variance-preserving (VP) probability-flow ordinary differential equation (ODE) recovers the posterior mean for a Gaussian source, minimizing mean-squared error (MSE). This result does not establish whether retuning λ can recover the posterior mean for a non-Gaussian source.

We characterize posterior-mean recovery using finite Gaussian mixture models (GMMs), which have been used as natural-image patch priors (Zoran & Weiss, 2011; 2012). Their scores can be computed from the source parameters, and their posterior means follow from Tweedie's identity (Efron, 2011). We compare these posterior means with fixed score-scaling endpoints without score-estimation error. A Gaussian score is affine, whereas a mixture score depends on component probabilities that vary with state and noise level.

With exact scores and continuous integration, we prove a positive excess-MSE lower bound over all real constant multipliers for every nontrivial finite GMM with positive weights and positive-definite component covariances.

Based on this characterization, we propose Simple Endpoint Correction (SEC), which combines two constant-scaling endpoints to approximate the posterior mean (Figure 1). Their difference defines a

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

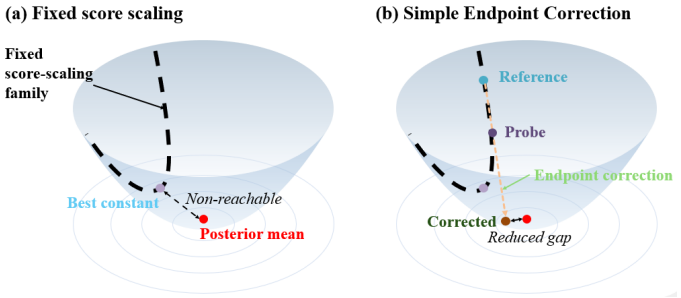


Figure 1: Schematic for nontrivial GMMs. (a) Fixed score scaling cannot recover the posterior mean. (b) Simple Endpoint Correction combines two flow endpoints to approximate the posterior mean.

correction direction. Calibration selects the endpoint pair and correction strength without retraining, and the selected parameters are fixed for evaluation. Our contributions are:

- We prove that constant score scaling recovers the posterior mean within finite GMMs only for a single Gaussian (Theorem 1) and establish stability under parameter and endpoint perturbations (Section 3.2).
- For deterministic denoising diffusion implicit models (DDIMs) with exact predictions, we establish fine-grid bounds on expanding multiplier intervals. A separate bound covers every fixed multistep grid for symmetric scalar two-component GMMs (Section 3.3).
- We introduce Simple Endpoint Correction, which combines two constant-scaling endpoints to approximate the posterior mean without retraining (Section 4).

2 Related work

2.1 VP diffusion and the posterior mean

VP diffusion defines a Gaussian observation model by adding Gaussian noise to data (Song et al., 2021b; Efron, 2011). Let $X_0 \in \mathbb{R}^d$ be a square-integrable clean source and let $(W_t)_{t\geq 0}$ be a d-dimensional Brownian motion independent of X_0. For a locally integrable noise rate $\beta : [0,\infty) \to (0,\infty)$, the VP diffusion is

$$dX_t = -\frac{1}{2}\beta(t)X_t\,dt + \sqrt{\beta(t)}\,dW_t, \qquad \overline{\alpha}(t) = \exp\left(-\int_0^t \beta(u)\,du\right), \tag{1}$$

The marginal at time t is

$$X_t \overset{d}{=} \sqrt{\overline{\alpha}(t)}X_0 + \sqrt{1-\overline{\alpha}(t)}\,\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I_d), \quad \varepsilon \perp X_0. \tag{2}$$

For $t > 0$, let q_t be the density of X_t and $s_t(x) = \nabla_x \log q_t(x)$ its score. For $Y = X_{t_{\text{obs}}}$ with $t_{\text{obs}} > 0$, let $a = \overline{\alpha}(t_{\text{obs}}) \in (0,1)$. The posterior mean minimizes MSE and satisfies Tweedie's identity (Efron, 2011):

$$m(y) := \mathbb{E}[X_0 \mid Y = y] = \frac{y + (1-a)s_{t_{\text{obs}}}(y)}{\sqrt{a}}. \tag{3}$$

2.2 Fixed score scaling

Score scaling adjusts the perception–distortion tradeoff at inference time without retraining (Wang et al., 2026). Fixed score scaling applies the same multiplier $\lambda \in \mathbb{R}$ to the time-dependent score throughout a reverse trajectory. The probability-flow ODE (Song et al., 2021b) becomes

$$\frac{dx_t^{(\lambda)}}{dt} = -\frac{1}{2}\beta(t)x_t^{(\lambda)} - \frac{\lambda}{2}\beta(t)s_t\left(x_t^{(\lambda)}\right), \qquad x_{t_{\text{obs}}}^{(\lambda)} = y, \qquad 0 < t \leq t_{\text{obs}}, \tag{4}$$

s_t is the score of q_t. At $\lambda = 1$, the flow has the same time-marginal densities as the diffusion. For Gaussian and finite-GMM sources, the endpoint $T_\lambda(y) := \lim_{t\downarrow 0} x_t^{(\lambda)}$ exists and $T_\lambda(Y)$ has finite second moment. Its mean-squared distance from $m(Y)$ is

$$G(\lambda) := \mathbb{E}\|T_\lambda(Y) - m(Y)\|^2. \tag{5}$$

The minimum mean-squared error (MMSE) is $\mathrm{MMSE} := \mathbb{E}\|X_0 - m(Y)\|^2$. Conditional orthogonality identifies $G(\lambda)$ as the excess MSE:

$$\mathbb{E}\|X_0 - T_\lambda(Y)\|^2 = \mathrm{MMSE} + G(\lambda). \tag{6}$$

For a Gaussian source, $T_2 = m$ and hence $G(2) = 0$ (Wang et al., 2026).

2.3 Gaussian mixtures

To study this equality beyond a single Gaussian, let X_0 have a finite-GMM density

$$p_0(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Sigma_k), \qquad \pi_k > 0, \quad \sum_{k=1}^{K} \pi_k = 1, \quad \mu_k \in \mathbb{R}^d, \quad \Sigma_k \succ 0, \tag{7}$$

Components with the same pair (μ_k, Σ_k) are merged. The mixture is nontrivial when at least two distinct components remain. The VP marginal q_t is a GMM with an explicit score (Liang et al., 2025). The posterior mean is given by Equation (3).

2.4 Diffusion-based compression

Diffusion-based compression reconstructs images through diffusion decoders (Theis et al., 2022; Yang & Mandt, 2023). Decoder-side controls include approximate optimal transport in latent space (Adrai et al., 2023) and combinations of one-step and multistep denoising outputs (Dornbusch et al., 2025).

3 Posterior-mean error under fixed score scaling

In this section, we established a positive lower bound on posterior-mean error over all constant multipliers for nontrivial finite GMMs (Section 3.1). We then examined its stability under parameter and endpoint perturbations (Section 3.2) and determined when the bound extends to finite-step DDIM (Section 3.3).

3.1 A lower bound for Gaussian mixtures

Theorem 1 (Posterior-mean recovery within finite GMMs)**.** *Fix $t_{\mathrm{obs}} > 0$ with $a \in (0, 1)$. If Equation* (7) *contains at least two distinct Gaussian components, then*

$$\inf_{\lambda \in \mathbb{R}} G(\lambda) > 0. \tag{8}$$

Among finite GMMs with positive-definite component covariances, posterior-mean recovery by constant score scaling holds only for a single Gaussian.

Tail expansions show that equality between the flow endpoint and posterior mean requires doubling the score. At multiplier two, the next-order terms differ, so $G(\lambda) > 0$ for every λ. Continuity of G and a positive lower bound for large $|\lambda|$ establish Equation (8). The proof is in Theorem 18.

3.2 Stability of the lower bound

Small changes to the mixture parameters preserve a positive lower bound on the error after optimizing the multiplier.

Corollary 2 (Stability under mixture-parameter perturbations)**.** *Fix the dimension, component count, and observation time. Let $\vartheta_0 = (\pi_k, \mu_k, \Sigma_k)_{k=1}^K$ be the parameters of a nontrivial finite GMM covered by Theorem 1. Let G_ϑ be the quantity in Equation* (5) *for a source with parameters ϑ. There exist an open neighborhood U of ϑ_0 and a constant $c_{\mathrm{rob}} > 0$ such that*

$$\inf_{\vartheta \in U} \inf_{\lambda \in \mathbb{R}} G_\vartheta(\lambda) \geq c_{\mathrm{rob}} > 0. \tag{9}$$

The neighborhood contains only positive weights and positive-definite covariances. It may become smaller when two components approach each other. The proof is in Theorem 25.

Approximate scores and numerical integration. Score approximation and numerical integration change the flow endpoint. Fix a nonempty bounded multiplier set Λ and set $G_{\Lambda,\min} = \inf_{\lambda\in\Lambda} G(\lambda)$. Let $\widehat{T}_\lambda$ be the endpoint driven by an approximate score and $\widetilde{T}_\lambda$ its numerical approximation. We measure both endpoint errors in root mean square, uniformly over $\lambda \in \Lambda$. Define $\delta_{\mathrm{approx}} = \sup_{\lambda\in\Lambda} \|\widehat{T}_\lambda - T_\lambda\|_{L^2(P_Y)}$ and $\delta_{\mathrm{num}} = \sup_{\lambda\in\Lambda} \|\widetilde{T}_\lambda - \widehat{T}_\lambda\|_{L^2(P_Y)}$, where P_Y is the law of Y. We write $[u]_+ = \max\{u, 0\}$.

Proposition 3 (Effect of score and integration errors)**.** *If the endpoints are measurable and both errors are finite, then*

$$\inf_{\lambda\in\Lambda} \mathbb{E}\|\widetilde{T}_\lambda(Y) - m(Y)\|^2 \geq [\sqrt{G_{\Lambda,\min}} - \delta_{\mathrm{approx}} - \delta_{\mathrm{num}}]_+^2. \tag{10}$$

The lower bound is positive whenever $\delta_{\mathrm{approx}} + \delta_{\mathrm{num}} < \sqrt{G_{\Lambda,\min}}$.

For a nontrivial finite GMM, Theorem 1 gives $G_{\Lambda,\min} > 0$. Proposition 23 bounds the endpoint error caused by score approximation on a compact set. Theorem 22 gives a lower bound under endpoint perturbations.

3.3 Posterior-mean error in deterministic DDIM

Sampling uses finitely many updates. A continuous-time lower bound alone does not establish an error for DDIM. We express its updates in the noise coordinate and use convergence to transfer the bound to fine grids. Divide a VP state by its signal coefficient and parameterize time by the noise-to-signal ratio $r(t) = \sqrt{(1-\overline{\alpha}(t))/\overline{\alpha}(t)}$. For $r = r(t)$, define $Y_r = X_0 + r\varepsilon$ using the clean source and Gaussian noise in Equation (2). Then $Y_r \overset{d}{=} X_t/\sqrt{\overline{\alpha}(t)}$.

Fix a finite observation level $R > 0$, an integer $N \geq 1$, and a strictly increasing grid $0 = r_0 < r_1 < \cdots < r_N = R$. Let $\bar{p}_{r^2}$ be the density of Y_r and $\bar{s}_{r^2} = \nabla \log \bar{p}_{r^2}$ its score, and write $m_r(y) := \mathbb{E}[X_0 \mid Y_r = y]$. The exact noise predictor is $\widehat{\varepsilon}_r(y) := \mathbb{E}[\varepsilon \mid Y_r = y] = -r\,\bar{s}_{r^2}(y)$. Starting from $y_N^{(\lambda)} = y$, the deterministic update with multiplier λ is

$$y_{i-1}^{(\lambda)} = y_i^{(\lambda)} - \lambda(r_i - r_{i-1})\widehat{\varepsilon}_{r_i}\left(y_i^{(\lambda)}\right). \tag{11}$$

At step i, $\widehat{\varepsilon}_{r_i}$ is the exact predictor for the distribution of Y_{r_i} and is evaluated at $y_i^{(\lambda)}$. Equation (11) is an explicit Euler step for the scaled probability-flow ODE in the r coordinate. At $\lambda = 1$, it is deterministic DDIM with $\eta_{\mathrm{DDIM}} = 0$ (Song et al., 2021a). Its endpoint is $T_{\lambda,N}^{\mathrm{DDIM}}(y) := y_0^{(\lambda)}$.

Proposition 4 (Uniform convergence of deterministic DDIM)**.** *Let Y_R have finite second moment. Assume that $(r, y) \mapsto \widehat{\varepsilon}_r(y)$ is continuously differentiable on $[0, R] \times \mathbb{R}^d$ and $\|\widehat{\varepsilon}_r(y)\| \leq C(1 + \|y\|)$ for some $C < \infty$. Let $T_\lambda^{\mathrm{cont}}$ be the endpoint of $dy_r/dr = \lambda\widehat{\varepsilon}_r(y_r)$ solved backward from R to 0. Fix any finite $M > 0$. The following convergence holds uniformly over all grids with maximum spacing at most h:*

$$\sup_{|\lambda|\leq M} \left\|T_{\lambda,N}^{\mathrm{DDIM}}(Y_R) - T_\lambda^{\mathrm{cont}}(Y_R)\right\|_{L^2} \longrightarrow 0 \qquad (h \downarrow 0). \tag{12}$$

Theorem 5 (Continuous lower bound for deterministic DDIM)**.** *Under the assumptions of Proposition 4, let $m_R(Y_R)$ be square-integrable. If*

$$g_* := \inf_{\lambda\in\mathbb{R}} \mathbb{E}\|T_\lambda^{\mathrm{cont}}(Y_R) - m_R(Y_R)\|^2 > 0, \tag{13}$$

then there exist $h_* > 0$, $c > 0$, and a nonincreasing function $L : (0, h_*] \to [1, \infty)$ with $L(h) \to \infty$ as $h \downarrow 0$ such that every grid of $[0, R]$ with $h := \max_i(r_i - r_{i-1}) \leq h_*$ satisfies

$$\inf_{|\lambda|\leq L(h)} \mathbb{E}\|T_{\lambda,N}^{\mathrm{DDIM}}(Y_R) - m_R(Y_R)\|^2 \geq c. \tag{14}$$

Corollary 6 (DDIM lower bound for finite GMMs)**.** *Every nontrivial finite GMM in Equation* (7) *at a finite observation level $R > 0$ satisfies the assumptions of Theorem 5. Hence Equation* (14) *holds for its exact noise predictor and posterior mean.*

Uniform convergence on bounded multiplier intervals gives the lower bound. The proof lets these intervals expand as the grid spacing decreases.

For a symmetric scalar GMM, the next theorem treats every grid with at least two steps and all real multipliers.

Theorem 7 (Fixed-grid result for a symmetric GMM). *Let* $X_0 \in \mathbb{R}$ *follow* $\frac{1}{2}\mathcal{N}(-\mu, \sigma^2_{\text{comp}}) + \frac{1}{2}\mathcal{N}(\mu, \sigma^2_{\text{comp}})$, *where* $\sigma_{\text{comp}} > 0$ *and* $\mu \neq 0$. *For every fixed strictly increasing grid with* $N \geq 2$,

$$\inf_{\lambda \in \mathbb{R}} \mathbb{E}\left|T^{\text{DDIM}}_{\lambda,N}(Y_R) - \mathbb{E}[X_0 \mid Y_R]\right|^2 > 0. \tag{15}$$

For $N = 1$, $T^{\text{DDIM}}_{1,1}(y) = \mathbb{E}[X_0 \mid Y_R = y]$.

For one step, the update with $\lambda = 1$ equals the posterior mean from Tweedie's identity. For $N \geq 2$, analyticity and the behavior as $|\lambda| \to \infty$ give the positive lower bound. Proofs are in Section B.9.

Results for denoising diffusion probabilistic models (DDPMs) are given in Appendix B.9.3.

4 Simple Endpoint Correction

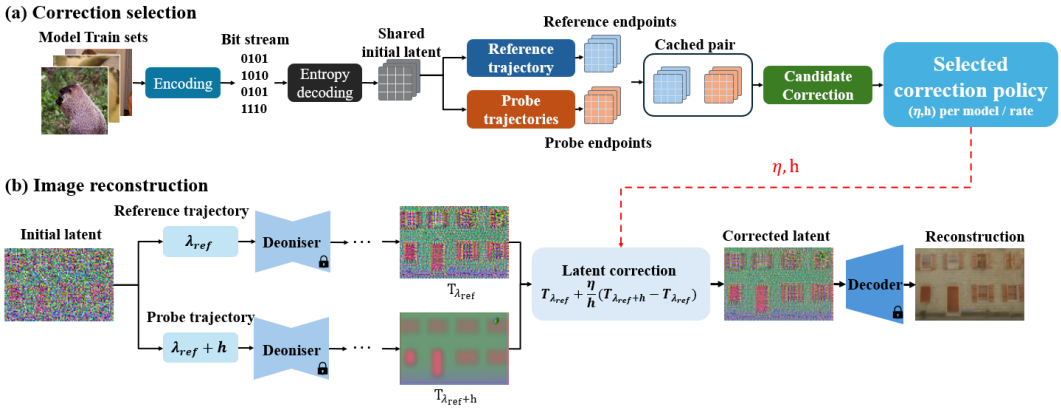


Figure 2: Simple Endpoint Correction. (a) The correction is selected using cached reference and probe endpoints. (b) The selected parameters combine two endpoints before decoding. Model weights remain frozen.

The lower bound in Theorem 1 applies to the constant-scaling endpoint family. Simple Endpoint Correction combines two endpoints, so its output need not lie in this family. We use the same sampler with two multipliers, each held fixed across noise levels. The endpoint difference defines a correction direction in latent space without estimating the posterior mean.

Endpoint correction. Let $T_\lambda(Y)$ denote the computed latent endpoint of a frozen diffusion decoder given initial state Y. The reference trajectory uses the native multiplier λ_{ref}. A probe trajectory uses $\lambda_{\text{ref}} + h$, where $h \neq 0$. Both trajectories use the same bitstream, initial state, solver, sampling steps, and random stream. Define

$$v_h(Y) = \frac{T_{\lambda_{\text{ref}}+h}(Y) - T_{\lambda_{\text{ref}}}(Y)}{h}, \qquad Z_{h,\eta}(Y) = T_{\lambda_{\text{ref}}}(Y) + \eta v_h(Y). \tag{16}$$

The probe offset h determines the endpoint pair, and η controls the correction strength. Setting $\eta = 0$ or $\eta = h$ reproduces the reference or probe endpoint, respectively.

Correction selection. For each model, we calibrate (h, η) on the test split of its training dataset (Figure 2(a)). Candidate sets $\mathcal{H}$ and $\mathcal{E}$ are fixed before calibration, with $0 \notin \mathcal{H}$ and $0 \in \mathcal{E}$. The probe multipliers and constant controls are within the model's supported range. Reference and probe endpoints are cached and reused across candidate correction strengths.

For each model and rate, we select (h, η) on these images to minimize mean image-space MSE without increasing mean learned perceptual image patch similarity (LPIPS) (Zhang et al., 2018) relative to native decoding.

Image reconstruction. The selected pair is fixed for all evaluation images. For each image, we compute its reference and probe endpoints and apply Equation (16) (Figure 2(b)). With the frozen image decoder $\mathcal{D}_\theta$, the reconstruction is $\widehat{X}_{h,\eta}(Y) = \mathrm{clip}_{[0,1]}(\mathcal{D}_\theta(Z_{h,\eta}(Y)))$. Evaluation MSE and LPIPS are reported separately. The calibration constraint does not guarantee the same LPIPS ordering on new images.

Comparisons and cost. For small h and η and a smooth endpoint map, $Z_{h,\eta}(Y) \approx T_{\lambda_{\mathrm{ref}}+\eta}(Y)$. We compare SEC with the effective constant endpoint and a dense constant-multiplier grid to test whether it does more than retune the multiplier. Reconstruction uses two trajectories and one image decoding. The bitstream is unchanged, and the selected parameters are shared decoder settings.

5 Experiments

In this section, we evaluated whether GMMs fit image latents better than a single Gaussian (Section 5.1). We then measured how much posterior-mean error remained after multiplier selection on controlled GMMs (Section 5.2). For scalar GMMs, we examined whether constant score scaling incurred excess distortion at other perception levels (Section 5.3). We also evaluated the PD tradeoff of Simple Endpoint Correction for image compression (Section 5.4).

5.1 Image latent distributions

We compared Gaussian and GMM fits to image latents from the SD2.1 and SDXL variational autoencoders (VAEs) (Rombach et al., 2022; Podell et al., 2024). $K = 32$ gave the lowest development negative log-likelihood (NLL) for all four dataset–encoder pairs. Experimental settings are given in Section D.1.

ImageNet / SD2.1 VAE					ImageNet / SDXL VAE				
Noise σ_{lat}	0	1/3	1	3	Noise σ_{lat}	0	1/3	1	3
Gaussian	0.810	1.070	1.616	2.558	Gaussian	0.868	1.070	1.618	2.558
GMM	0.661	0.930	1.566	2.537	GMM	0.650	0.951	1.573	2.537
COCO / SD2.1 VAE					**COCO / SDXL VAE**				
Noise σ_{lat}	0	1/3	1	3	Noise σ_{lat}	0	1/3	1	3
Gaussian	0.883	1.067	1.614	2.557	Gaussian	0.873	1.072	1.618	2.558
GMM	0.662	0.928	1.566	2.537	GMM	0.659	0.952	1.572	2.538

Table 1: ImageNet and COCO latent NLL for Gaussian and GMM. Values are image-averaged nats per dimension. Lower is better.

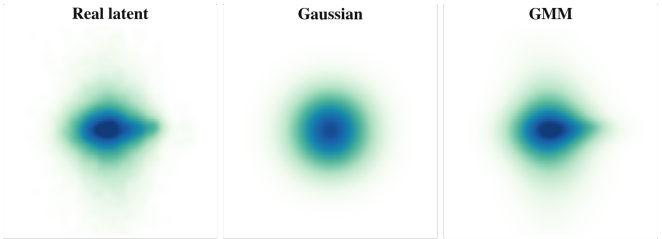


Figure 3: Two-dimensional projection of ImageNet–SD2.1 latents.

GMMs describe the evaluated image latents better than a single Gaussian, both qualitatively in the projected densities (Figure 3) and quantitatively in held-out NLL (Table 1).

5.2 Posterior-mean error for Gaussian and GMM sources

We compared a scalar Gaussian and two-component GMMs with symmetric means, normalized component separation $\Delta = 4$, and component variance ratios $\kappa \in \{1, 16\}$. At each observation-noise level $\sigma_{\mathrm{obs}} \in \{1/3, 1, 3\}$, we selected a constant multiplier $\widehat{\lambda}$ to minimize the evaluated posterior-mean error. Settings and the error metric are given in Section D.2.

Table 2: Minimum evaluated posterior-mean error $100G(\widehat{\lambda})/\,\mathrm{MMSE}$ in percent. Parentheses show the selected constant multiplier $\widehat{\lambda}$.

Source	$\sigma_{\mathrm{obs}} = 1/3$	$\sigma_{\mathrm{obs}} = 1$	$\sigma_{\mathrm{obs}} = 3$
Single Gaussian	≈ 0 (2.000)	≈ 0 (2.000)	≈ 0 (2.000)
Equal-covariance GMM	6.33 (1.718)	25.27 (1.017)	50.76 (1.628)
Unequal-covariance GMM	8.05 (1.691)	27.65 (0.919)	53.66 (1.628)

The Gaussian error is below $4 \times 10^{-28}\%$ at multiplier two. For both GMMs, the MMSE-normalized error increased with observation noise, from 6.33% to 50.76% for equal covariances and from 8.05% to 53.66% for unequal covariances. The unequal-covariance GMM had higher error at all three noise levels. All selected GMM multipliers differed from two. These are selected-multiplier errors; Theorem 1 establishes a positive lower bound over all multipliers.

Endpoint correction on controlled sources. We compared constant scaling and SEC on a Gaussian and three GMM families (symmetric, anisotropic, and heterogeneous) in $d \in \{1, 8\}$ dimensions at observation level $R = 1$. Scores and posterior means were computed from the source parameters. SEC used the selected constant $\widehat{\lambda}$ as its reference multiplier, with separate samples for coefficient selection and evaluation. Source parameters, sample counts, the MMSE-normalized metric $\widehat{G}$, and selected coefficients are given in Section D.3.

Table 3: MMSE-normalized posterior-mean error for selected constant scaling and continuous-time Simple Endpoint Correction (SEC). Lower is better.

Source	d	Constant $\widehat{G}_{\mathrm{const}}$	SEC $\widehat{G}_{\mathrm{SEC}}$
Gaussian	1, 8	≈ 0	≈ 0
Symmetric GMM	1	1.865×10^{-3}	1.361×10^{-3}
Anisotropic GMM	1	1.472×10^{-1}	1.239×10^{-1}
Heterogeneous GMM	1	3.075×10^{-2}	2.604×10^{-2}
Symmetric GMM	8	1.720×10^{-2}	1.129×10^{-2}
Anisotropic GMM	8	2.604×10^{-2}	2.380×10^{-2}
Heterogeneous GMM	8	7.586×10^{-2}	6.804×10^{-2}

Continuous-time SEC reduced posterior-mean error by 8.6–34.37% relative to selected constant scaling across the six GMM conditions (Table 3).

5.3 PD CURVES FOR SCALAR GMMS

The endpoint error leaves the rest of the PD tradeoff unmeasured. Comparing full curves determines where constant scaling incurs additional distortion at the same perception level. For scalar GMMs, Wasserstein distance can be computed from quantiles. For each multiplier, define

$$D_\lambda := \mathbb{E}\|X_0 - T_\lambda(Y)\|^2, \qquad P_\lambda := W_2^2(\mathcal{L}(T_\lambda(Y)), \mathcal{L}(X_0))\,. \tag{17}$$

$\mathcal{L}$ denotes a probability law and W_2 is the Wasserstein-2 distance. Let $m(Y) = \mathbb{E}[X_0 \mid Y]$ and $c^2 = W_2^2(\mathcal{L}(m(Y)), \mathcal{L}(X_0))$. The best distortion at a perception constraint P is $D_{\mathrm{scale}}(P) = \inf_{\lambda: P_\lambda \le P} D_\lambda$. We compare it with $D^\star(P) = \mathrm{MMSE} + (\sqrt{c^2} - \sqrt{P})_+^2$ for $P \in [0, c^2]$ (Freirich et al., 2021). A finite multiplier search gives $\widehat{D}_{\mathrm{scale}}(P)$. We report the maximum of $[\widehat{D}_{\mathrm{scale}}(P) - D^\star(P)]/\,\mathrm{MMSE}$, its integral over P divided by c^2, and its value at $P = c^2$. Source settings and numerical procedures are given in Section D.4.

5.4 ENDPOINT CORRECTION FOR IMAGE COMPRESSION

We evaluated DiffC-SD2.1 (Theis et al., 2022), ResULIC (Ke et al., 2025), and ELIC+DDIM/DSG (Xu et al., 2024) on Kodak, CLIC-64, and Tecnick-100. Ours selects (h, η), while Fixed-h fixes

Table 4: Excess distortion of constant score scaling for scalar sources, in percent of MMSE. P90 denotes the 90th percentile across 300 scalar mixture configurations.

Quantity	Gaussian	GMM median	GMM P90	GMM largest
Maximum over P	≈ 0	2.28	8.77	53.68
Integral over P	≈ 0	2.19	7.26	45.16
Value at $P = c^2$	≈ 0	2.25	8.72	53.68

$h = 0.5$ and selects η. We compare these with native decoding ($\lambda = 1$), the selected constant multiplier, and the effective constant $\lambda_{\text{ref}} + \eta$.

DiffC-SD2.1 selected $h = 1.5$, with $\eta = 0.624$ at low and middle rates and $\eta = 0.512$ at high rate. ResULIC selected $h = 1.0$, with $\eta = 0.43$ at low and middle rates and $\eta = 0.256$ at high rate. ELIC+DDIM/DSG selected $h = 1.5$, with $\eta = 0.735$ at low and middle rates and $\eta = 0.542$ at high rate.

Table 5 reports the arithmetic mean across low, middle, and high rates for each model, method, and dataset. Rate-wise results are in Table 10.

Table 5: Image-compression results (rate mean). Each value is an arithmetic mean over low, middle, and high rates. MSE is multiplied by 10^3. Lower is better. Bold and underline mark the best and second-best unrounded means within each model, dataset, and metric, including ties.

		Kodak		CLIC-64		Tecnick-100	
Model	Method	MSE	LPIPS	MSE	LPIPS	MSE	LPIPS
DiffC-SD2.1	Ours	**20.356**	0.4989	**25.479**	**0.4679**	**33.639**	0.4544
	Fixed-h	20.674	0.5048	26.076	0.4712	34.510	0.4567
	Native ($\lambda = 1$)	29.915	**0.4646**	51.451	0.4779	61.545	**0.4507**
	Constant	29.002	0.5111	48.741	0.4730	58.709	0.4568
	Effective constant	21.591	0.6114	26.899	0.5243	35.454	0.5310
ResULIC	Ours	**10.236**	**0.3560**	**10.498**	**0.3221**	**14.801**	**0.3114**
	Fixed-h	10.628	0.3576	11.081	0.3229	15.460	0.3122
	Native ($\lambda = 1$)	11.271	0.3591	11.782	0.3226	16.598	0.3115
	Constant	11.271	0.3591	11.782	0.3226	16.598	0.3115
	Effective constant	11.271	0.3591	11.782	0.3226	16.598	0.3115
ELIC+DDIM/DSG	Ours	**4.545**	**0.1423**	**3.874**	**0.1128**	**5.199**	**0.1099**
	Fixed-h	4.628	0.1426	3.963	0.1129	5.331	0.1100
	Native ($\lambda = 1$)	4.756	0.1431	4.115	0.1133	5.502	0.1103
	Constant	4.753	0.1430	4.112	0.1131	5.497	0.1101
	Effective constant	4.756	0.1431	4.115	0.1133	5.502	0.1103

Simple Endpoint Correction had the lowest rate-mean MSE for DiffC-SD2.1, ResULIC, and ELIC+DDIM/DSG (Table 5) and the lowest MSE at every evaluated rate (Table 10). For DiffC-SD2.1 and ResULIC, both MSE and LPIPS decreased relative to native decoding at low rates, whereas MSE decreased and LPIPS increased at high rates. For ELIC+DDIM/DSG, both metrics decreased at every evaluated rate. These comparisons use the selected settings. The methods were not compared at equal test-set LPIPS.

6 Conclusion

We proved that constant-scaled VP flows recover the posterior mean within finite GMMs only for a single Gaussian, with a positive minimum excess MSE for nontrivial mixtures. The bound extends to DDIM under the stated grid conditions.

GMMs had lower held-out NLL than single Gaussians on SD2.1 and SDXL image latents. Constant scaling left posterior-mean error on synthetic GMMs and showed excess distortion in scalar PD

comparisons. Simple Endpoint Correction reduced posterior-mean error on synthetic GMMs and image-space MSE relative to native decoding for DiffC-SD2.1, ResULIC, and ELIC+DDIM/DSG.

7 Future Work and Limitations

For DiffC-SD2.1 and ResULIC, SEC reduced MSE, but its LPIPS effect depended on rate. Recovering the posterior mean does not guarantee lower LPIPS, which compares learned image features (Zhang et al., 2018). These results do not establish an optimal MSE–LPIPS curve. Future work could apply Learn then Test (Angelopoulos et al., 2021) to paired LPIPS differences on independent validation images. With i.i.d. data and a known loss bound, this could certify expected LPIPS no higher than native decoding at a chosen confidence level. Native decoding would remain the fallback if no correction qualifies.

Our latent NLL results motivate GMM priors. Gaussian VAE posteriors averaged over a finite image set form a Gaussian mixture (Tomczak & Welling, 2018). Future work could learn such priors and use their posterior means for reconstruction; a GMM prior alone does not remove the constant-scaling bound (Theorem 1). Alternatively, an invertible flow with a Gaussian base could target Gaussian latents (Dinh et al., 2017). A nonlinear decoder need not map a latent posterior mean to the image posterior mean.

Reproducibility Statement

Section 5 describes the density fits and controlled comparisons. Section 4 defines the correction and calibration-set selection rule. The appendix contains the notation, proofs, and experimental settings.

AI-use disclosure

AI tools were used to double-check mathematical proofs, generate a 3D illustration, check English spelling and grammar, and debug experiment code.

References

Theo J. Adrai, Guy Ohayon, Michael Elad, and Tomer Michaeli. Deep optimal transport: A practical algorithm for photo-realistic image restoration. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Anastasios N. Angelopoulos, Stephen Bates, Emmanuel J. Candès, Michael I. Jordan, and Lihua Lei. Learn then test: Calibrating predictive algorithms to achieve risk control. *arXiv preprint arXiv:2110.01052*, 2021. URL https://arxiv.org/abs/2110.01052.

Fan Bao, Chongxuan Li, Jiacheng Sun, Jun Zhu, and Bo Zhang. Estimating the optimal covariance with imperfect mean in diffusion probabilistic models. In *International Conference on Machine Learning*, pp. 1555–1584, 2022a.

Fan Bao, Chongxuan Li, Jun Zhu, and Bo Zhang. Analytic-DPM: An analytic estimate of the optimal reverse variance in diffusion probabilistic models. In *International Conference on Learning Representations*, 2022b.

Yochai Blau and Tomer Michaeli. The perception–distortion tradeoff. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6228–6237, 2018. doi: 10.1109/CVPR.2018.00652.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP. In *International Conference on Learning Representations*, 2017. URL https://arxiv.org/abs/1605.08803.

Jonas Dornbusch, Emanuel Pfarr, Florin-Alexandru Vasluianu, Frank Werner, and Radu Timofte. A simple combination of diffusion models for better quality trade-offs in image denoising. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 895–904, 2025.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Dror Freirich, Tomer Michaeli, and Ron Meir. A theory of the distortion–perception tradeoff in Wasserstein space. In *Advances in Neural Information Processing Systems*, 2021.

Hanzhong Guo, Cheng Lu, Fan Bao, Tianyu Pang, Shuicheng Yan, Chao Du, and Chongxuan Li. Gaussian mixture solvers for diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 25598–25626, 2023.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

Anle Ke, Xu Zhang, Tong Chen, Ming Lu, Chao Zhou, Jiawen Gu, and Zhan Ma. Ultra lowrate image compression with semantic residual coding and compression-aware diffusion. In *International Conference on Machine Learning*, 2025.

Yingyu Liang, Zhizhou Sha, Zhenmei Shi, Zhao Song, Mingda Wan, and Yufa Zhou. Unraveling the smoothness properties of diffusion models: A Gaussian mixture perspective. In *IEEE/CVF International Conference on Computer Vision*, pp. 11436–11446, 2025.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, pp. 740–755, 2014. doi: 10.1007/978-3-319-10602-1_48.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pp. 8162–8171, 2021. URL https://proceedings.mlr.press/v139/nichol21a.html.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations*, 2024.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

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

Lucas Theis, Tim Salimans, Matthew D. Hoffman, and Fabian Mentzer. Lossy compression with gaussian diffusion. *arXiv preprint arXiv:2206.08889*, 2022.

Jakub M. Tomczak and Max Welling. VAE with a VampPrior. In *International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pp. 1214–1223, 2018. URL https://proceedings.mlr.press/v84/tomczak18a.html.

Yuhan Wang, Suzhi Bi, Ying-Jun Angela Zhang, and Xiaojun Yuan. Traversing distortion–perception tradeoff using a single score-based generative model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2377–2386, 2025.

Yuhan Wang, Suzhi Bi, and Ying-Jun Angela Zhang. Training-free rate–distortion–perception traversal with diffusion. In *International Conference on Machine Learning*, 2026. arXiv:2603.04005.

Tongda Xu, Ziran Zhu, Dailan He, Yanghao Li, Lina Guo, Yuanyuan Wang, Zhe Wang, Hongwei Qin, Yan Wang, Jingjing Liu, and Ya-Qin Zhang. Idempotence and perceptual image compression. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/c801e68207da477bbc44182b9fac1129-Abstract-Conference.html.

Ruihan Yang and Stephan Mandt. Lossy image compression with conditional diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *IEEE International Conference on Computer Vision*, pp. 479–486, 2011. doi: 10.1109/ICCV.2011.6126278.

Daniel Zoran and Yair Weiss. Natural images, gaussian mixtures and dead leaves. In *Advances in Neural Information Processing Systems 25*, 2012.

A NOTATION

Symbols used only within one proof are defined locally.

Table 6: Notation used in the main text.

Symbol	Meaning
Basic notation	
d, $\mathbb{R}^d$, I_d	Signal dimension, Euclidean signal space, and its identity matrix.
$\|x\|$, $\langle x, y\rangle$	Euclidean norm and inner product.
$A \succ 0$	The symmetric matrix A is positive definite.
$\mathbb{E}$, $\mathbb{P}$, D_{KL}	Expectation, probability, and forward Kullback–Leibler divergence.
$\mathcal{N}(\mu, \Sigma)$, $\mathcal{N}(x; \mu, \Sigma)$	Gaussian law and its density at x.
Continuous VP diffusion	
X_0, X_t	Clean random signal and its VP-diffused version at time t.
W_t, ε	Brownian motion and independent standard Gaussian noise.
$\beta(t)$, $\overline{\alpha}(t)$	VP noise rate and cumulative signal-power coefficient.
q_t, $s_t = \nabla \log q_t$	Density of X_t and its score.
t_{obs}, a	Observation time and signal coefficient $a = \overline{\alpha}(t_{\mathrm{obs}})$.
Y, y	Random observation $Y = X_{t_{\mathrm{obs}}}$ and a realization.
$m(y)$	Posterior mean $\mathbb{E}[X_0 \mid Y = y]$.
λ	Constant score multiplier.
Λ	Nonempty set of score multipliers, bounded where stated.
$x_t^{(\lambda)}$, $T_\lambda(y)$	Trajectory and endpoint of the probability-flow ODE with constant score multiplier λ.
$\widehat{T}_\lambda$, $\widetilde{T}_\lambda$	Flow endpoint computed with an approximate score, and its numerical approximation.
$G(\lambda)$, MMSE	Mean-squared distance from the flow endpoint to the posterior mean, and the minimum MSE.
$G_{\Lambda,\min}$	Infimum of $G(\lambda)$ over a multiplier set Λ.
δ_{approx}, δ_{num}	$L^2(P_Y)$ errors from score approximation and numerical integration, maximized over Λ.
τ	Normalized observation-noise variance $(1-a)/a$.
Finite Gaussian mixtures	
p_0	Density of the clean source X_0.
K	Number of distinct Gaussian components.
π_k, μ_k, Σ_k	Weight, mean, and covariance of component k.
ϑ, G_ϑ	A finite-GMM parameter vector and the corresponding mean-squared distance from the flow endpoint to the posterior mean.
Discrete VP diffusion and DDPM	
j, J	Discrete diffusion index and terminal index.
β_j, α_j, $\overline{\alpha}_j$	Step noise variance, $1-\beta_j$, and $\prod_{\ell=1}^{j} \alpha_\ell$.
X_j, x_j	Random diffusion state and a realization.
$q(\cdot \mid \cdot)$	Conditional density in the DDPM chain.
P_J	Marginal law of X_J at the terminal step.
$q_j^{\mathrm{rev}}(x \mid y)$	Reverse conditional density of $X_{j-1} = x$ given $X_j = y$.
$\mathcal{K}_j(y)$, $\overline{\mathcal{K}}_j$	Minimum reverse-step KL divergence over Gaussian model reverse transitions and its expectation over X_j.
$p_j(x_{j-1} \mid x_j)$	A measurable Gaussian model reverse transition at step j.
$\mathcal{Q}$, $\mathcal{P}$	Reverse-process path measure induced by the forward chain and a path measure with Gaussian transitions; both share terminal law P_J.
Deterministic DDIM	
$r(t)$, R, Y_r	Normalized noise-to-signal level $\sqrt{(1-\overline{\alpha}(t))/\overline{\alpha}(t)}$, terminal level, and observation $Y_r = X_0 + r\varepsilon$.

Continued on the next page

Symbol	Meaning
$r_0 < \cdots < r_N$	An N-step grid from 0 to R.
$\bar{p}_{r^2}$, $\bar{s}_{r^2}$	Density of Y_r and its score $\nabla \log \bar{p}_{r^2}$.
$\widehat{\varepsilon}_r(y)$	Exact conditional noise mean $\mathbb{E}[\varepsilon \mid Y_r = y]$.
$y_i^{(\lambda)}$	Deterministic DDIM state at grid point r_i.
$T_{\lambda,N}^{\mathrm{DDIM}}$	Endpoint map of the N-step scaled DDIM update.
$T_\lambda^{\mathrm{cont}}$	Endpoint map of the continuous scaled flow in the normalized DDIM coordinate.
σ_{comp}	Component standard deviation in the symmetric scalar GMM used for the fixed-grid DDIM theorem.
η_{DDIM}	DDIM stochasticity parameter; deterministic DDIM uses $\eta_{\mathrm{DDIM}} = 0$.
h	Maximum grid spacing $\max_i(r_i - r_{i-1})$.
$L(h)$	Multiplier radius in the fine-grid DDIM bound, with $L(h) \to \infty$ as $h \downarrow 0$.
Simple Endpoint Correction (Section 4)	
λ_{ref}	Native decoder multiplier.
h, η	Probe interval and correction strength; here h is not the DDIM grid spacing.
$v_h(Y)$, $Z_{h,\eta}(Y)$	Endpoint difference divided by the probe interval, and corrected latent endpoint.
$\mathcal{D}_\theta$, $\widehat{X}_{h,\eta}$	Frozen image decoder and decoded, clipped reconstruction.
$\mathcal{H}$, $\mathcal{E}$	Candidate sets for probe intervals and correction strengths.
PD evaluation	
D_λ	Reconstruction distortion $\mathbb{E}\|X_0 - T_\lambda(Y)\|^2$.
P_λ	Squared Wasserstein-2 distance between the laws of $T_\lambda(Y)$ and X_0.
$D_{\mathrm{scale}}(P)$	Infimum distortion of constant score scaling under perception constraint P.
$D^\star(P)$	Optimal distortion under perception constraint P.
c^2	Perception distance between the posterior-mean distribution and the source distribution.
Δ, κ, σ_{obs}	Normalized component separation, component variance ratio, and observation-noise standard deviation in the scalar experiment.
σ_{lat}	Isotropic latent-noise standard deviation used in the density-fit experiment.

B PROOFS

B.1 CONTINUOUS-TIME FORMULATION AND REGULARITY

Set $\tau = (1-a)/a$ and $Y_\tau = Y/\sqrt{a}$ in the additive-noise parameterization of Section C. T_λ denotes the additive-coordinate endpoint. By Equation (173), its value at Y_τ equals the clean endpoint of the main-text VP flow at Y. Let $X := X_0 \in \mathbb{R}^d$ have a finite second moment. For $t \geq 0$, define the Gaussian observation marginals

$$Y_t = X + \sqrt{t}\,\varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I_d), \quad \varepsilon \perp X, \qquad \bar{p}_t := p_0 * \mathcal{N}(0, tI_d). \tag{18}$$

t denotes noise variance. The exact score and posterior mean at a fixed $\tau > 0$ are

$$\bar{s}_t(y) := \nabla \log \bar{p}_t(y), \qquad m_\tau(y) = \mathbb{E}[X \mid Y_\tau = y] = y + \tau \bar{s}_\tau(y). \tag{19}$$

Tweedie's formula recovers m_τ from the exact score at the observation level. Section B.8 gives the derivation. The endpoint family is $\{T_\lambda : \lambda \in \mathbb{R}\}$ (Song et al., 2021b; Efron, 2011; Wang et al., 2026).

For each finite $\lambda \in \mathbb{R}$, solve

$$\frac{d}{dt} y_t^{(\lambda)}(y) = -\frac{\lambda}{2} \bar{s}_t(y_t^{(\lambda)}(y)), \qquad y_\tau^{(\lambda)}(y) = y, \qquad t : \tau \downarrow 0, \qquad T_\lambda(y) = y_0^{(\lambda)}(y). \tag{20}$$

The score $\bar{s}_t$ is fixed across multipliers. At $\lambda = 1$, this is the probability-flow ODE for the heat-flow marginals.

Define total squared-error quantities by

$$G_p(\lambda) = \mathbb{E}\|T_\lambda(Y_\tau) - m_\tau(Y_\tau)\|^2, \qquad \mathrm{MMSE}_p(\tau) = \mathbb{E}\|X - m_\tau(Y_\tau)\|^2. \tag{21}$$

Conditional orthogonality gives

$$\mathbb{E}\|X - T_\lambda(Y_\tau)\|^2 = \mathrm{MMSE}_p(\tau) + G_p(\lambda). \tag{22}$$

For the source and observation time in the main text, $m_\tau(Y_\tau) = m(Y)$ and $G_p(\lambda) = G(\lambda)$. Per-coordinate MSE is obtained by dividing by d.

Lemma 8 (Existence, regularity, and local parameter continuity)**.** *For a finite GMM with positive-definite component covariances, equation 20 has a unique solution on $[0, \tau]$ for every finite λ and y. Its endpoint is continuously differentiable in y and continuous in (λ, y). $G_p(\lambda)$ is continuous on $\mathbb{R}$. The same conclusions hold jointly with mixture parameters in a compact family whose component covariance eigenvalues are bounded below by $a > 0$ and whose means have norms bounded by $M < \infty$.*

Proof. The score is a posterior-weighted average of $-(\Sigma_k + tI_d)^{-1}(y - \mu_k)$, so

$$\|\bar{s}_t(y)\| \leq a^{-1}(\|y\| + M), \qquad 0 \leq t \leq \tau. \tag{23}$$

The score is smooth in y and locally uniformly smooth in the parameters. Local uniqueness and the linear-growth bound give existence on the full finite interval. Put $\xi = \tau - t$. The forward-in-ξ equation is $\dot{x}_\xi = (\lambda/2)\bar{s}_{\tau-\xi}(x_\xi)$. For $|\lambda| \leq L$, Grönwall's inequality yields

$$\|x_\xi\| + M \leq (\|y\| + M) e^{L\xi/(2a)}. \tag{24}$$

These bounds give continuous dependence and a common second-moment bound on compact multiplier sets. For a fixed source, dominated convergence proves continuity of G_p. The same argument gives joint continuity in the source. On a compact parameter family, Gaussian densities are bounded by $C\exp(-c\|y\|^2)$, and the squared residual is bounded by $C_L(1+\|y\|^2)$. All constants may depend on the indicated compact sets. No bound uniform over $\lambda \in \mathbb{R}$ is assumed. □

B.2 ISOTROPIC FINITE GMMS

Theorem 9 (Positive excess MSE over all constant multipliers for an isotropic finite GMM)**.** *Fix $d \geq 1$ and $\tau > 0$. Let*

$$\begin{aligned} p_0(x) &= \sum_{k=1}^{K} \pi_k \mathcal{N}(x; 0, \sigma_k^2 I_d), \qquad K \geq 2, \\ \pi_k &> 0, \quad \sum_k \pi_k = 1, \quad 0 < \sigma_1^2 < \cdots < \sigma_K^2 < \infty. \end{aligned} \tag{25}$$

Then no finite constant multiplier makes $T_\lambda = m_\tau$ $\bar{p}_\tau$-almost everywhere, and

$$\inf_{\lambda \in \mathbb{R}} G_p(\lambda) > 0. \tag{26}$$

No unit-variance normalization is required. After merging duplicate variances, a "unique widest component" assumption is unnecessary.

B.2.1 Radial reduction and tail slope

Set $v_k(t) = \sigma_k^2 + t$ and define

$$w_k(R,t) = \frac{\pi_k v_k(t)^{-d/2} \exp[-R^2/(2v_k(t))]}{\sum_j \pi_j v_j(t)^{-d/2} \exp[-R^2/(2v_j(t))]}, \qquad a(R,t) = \sum_k \frac{w_k(R,t)}{v_k(t)}. \tag{27}$$

Then

$$\bar{s}_t(y) = -a(\|y\|,t)y, \qquad \frac{1}{v_K(t)} \le a(R,t) \le \frac{1}{v_1(t)}, \qquad m_\tau(y) = [1 - \tau a(\|y\|,\tau)]y. \tag{28}$$

The ODE preserves each ray. For $y = Re$, $\|e\| = 1$, $R > 0$, write $T_\lambda(Re) = \phi_\lambda(R)e$ and $m_\tau(Re) = \psi(R)e$. In reverse elapsed time $\xi = \tau - t$, the radius satisfies

$$\dot{\mathcal{R}}_\xi = -\frac{\lambda}{2} a(\mathcal{R}_\xi, \tau - \xi)\mathcal{R}_\xi, \qquad \mathcal{R}_0 = R, \qquad \phi_\lambda(R) = \mathcal{R}_\tau > 0. \tag{29}$$

For each fixed finite λ, $Re^{-|\lambda|\tau/(2\sigma_1^2)} \le \mathcal{R}_\xi \le Re^{|\lambda|\tau/(2\sigma_1^2)}$ for all ξ.

Lemma 10 (Tail slopes)**.** *Let $\alpha_K = \sigma_K^2/(\sigma_K^2 + \tau) \in (0,1)$. For every finite λ,*

$$\lim_{R\to\infty} \frac{\phi_\lambda(R)}{R} = \alpha_K^{\lambda/2}, \qquad \lim_{R\to\infty} \frac{\psi(R)}{R} = \alpha_K. \tag{30}$$

Equality of the two endpoint maps implies $\lambda = 2$.

Proof. For each $k < K$, the ratio $w_k(R,t)/w_K(R,t)$ is bounded by $C\exp(-cR^2)$, uniformly for $t \in [0,\tau]$ because $1/v_k(t) - 1/v_K(t)$ has a strictly positive minimum on that interval. Hence $a(R,t) \to 1/v_K(t)$ uniformly in t as $R \to \infty$. Along trajectories started at $R \to \infty$, the radius lower bound extends this limit uniformly in time for each fixed λ. Integrating the logarithmic derivative in equation 29 gives

$$\log \frac{\phi_\lambda(R)}{R} = -\frac{\lambda}{2}\int_0^\tau a(\mathcal{R}_\xi, \tau - \xi)\, d\xi \longrightarrow -\frac{\lambda}{2}\log\frac{\sigma_K^2 + \tau}{\sigma_K^2}.$$

The posterior-mean limit follows from equation 28. Since $\alpha_K \in (0,1)$, equality of the slopes requires $\lambda = 2$. □

B.2.2 Derivatives at the origin

Define

$$\bar{w}_k(t) = \frac{\pi_k v_k(t)^{-d/2}}{h(t)}, \quad h(t) = \sum_k \pi_k v_k(t)^{-d/2}, \quad A(t) = \sum_k \frac{\bar{w}_k(t)}{v_k(t)}, \quad \chi(t) = h(t)^{-2/d}. \tag{31}$$

Lemma 11 (Derivative mismatch at the origin)**.** *For the isotropic finite GMM in Theorem 9,*

$$\begin{aligned} DT_2(0) = JI_d, \qquad Dm_\tau(0) = [1 - \tau A(\tau)]I_d, \\ J = \exp\left(-\int_0^\tau A(t)\, dt\right) = \frac{\chi(0)}{\chi(\tau)} < 1 - \tau A(\tau). \end{aligned} \tag{32}$$

Proof. The origin is a stationary trajectory, and $D\bar{s}_t(0) = -A(t)I_d$. The variational equation, integrated from τ to zero at $\lambda = 2$, gives the formula for J. Since $h'/h = -(d/2)A$, we have $\chi'/\chi = A$. Put $n = d/2$ and $b_k(t) = 1/v_k(t)$. Differentiating the weights gives

$$\bar{w}_k' = -n\bar{w}_k(b_k - A), \qquad A' = -\mathbb{E}_{\bar{w}}[b_k^2] - n\mathrm{Var}_{\bar{w}}(b_k) = -A^2 - (n+1)\mathrm{Var}_{\bar{w}}(b_k).$$

Thus

$$\frac{\chi''(t)}{\chi(t)} = A'(t) + A(t)^2 = -\left(\frac{d}{2}+1\right)\mathrm{Var}_{\bar{w}(t)}\left(\frac{1}{v_k(t)}\right) < 0. \tag{33}$$

Distinct variances and positive weights make this inequality strict for every $t \in [0,\tau]$. Strict concavity implies $\chi(0) < \chi(\tau) - \tau\chi'(\tau)$. Divide by $\chi(\tau) > 0$ to obtain $J < 1 - \tau A(\tau)$. Differentiating the posterior mean at the origin proves the remaining identity. Its coefficient is positive because it equals $\sum_k \bar{w}_k(\tau)\sigma_k^2/(\sigma_k^2+\tau)$. □

B.2.3 Uniform lower bound across multipliers

Lemma 12 (Strict multiplier monotonicity on each ray)**.** *For every $R > 0$ and finite λ, $\partial_\lambda \phi_\lambda(R) < 0$.*

Proof. Let $g(R,\xi) = a(R, \tau-\xi)R > 0$ and differentiate $\dot{\mathcal{R}}_\xi = -(\lambda/2)g(\mathcal{R}_\xi, \xi)$. Variation of constants gives

$$\partial_\lambda \phi_\lambda(R) = -\frac{1}{2}\int_0^\tau g(\mathcal{R}_v, v)\exp\left[-\frac{\lambda}{2}\int_v^\tau \partial_R g(\mathcal{R}_\xi,\xi)\,d\xi\right]dv < 0. \tag{34}$$

All derivatives exist on each finite trajectory by Lemma 8. The sign does not require $\partial_R g$ to have a particular sign. □

Proof of Theorem 9. The tail slope requires $\lambda = 2$, but the derivatives in Lemma 11 differ at that multiplier. Because $\bar{p}_\tau$ is everywhere positive and both maps are continuous, $\bar{p}_\tau$-almost-everywhere equality would imply equality everywhere. Exact recovery therefore fails for every finite multiplier.

For the uniform statement, choose $R_c > 0$ sufficiently small that $\phi_2(R_c) < \psi(R_c)$. By continuity in λ, choose $\eta \in (0,1)$ so that $\lambda_0 = 2-\eta$ satisfies $\phi_{\lambda_0}(R_c) < \psi(R_c)$. At this multiplier, Lemma 10 gives a radius R_t with $\phi_{\lambda_0}(R_t) > \psi(R_t)$. Continuity in R yields compact positive radius intervals $J_c, J_t \subset (0,\infty)$ and numbers $b_c, b_t > 0$ such that

$$\psi(R) - \phi_{\lambda_0}(R) \geq b_c \quad (R \in J_c), \qquad \phi_{\lambda_0}(R) - \psi(R) \geq b_t \quad (R \in J_t).$$

Define the full-dimensional shells $H_i = \{y : \|y\| \in J_i\}$ and their probabilities $P_i = \mathbb{P}(Y_\tau \in H_i) > 0$. Monotonicity implies

$$\lambda \geq \lambda_0 \implies G_p(\lambda) \geq P_c b_c^2, \qquad \lambda \leq \lambda_0 \implies G_p(\lambda) \geq P_t b_t^2.$$

Thus

$$\inf_{\lambda\in\mathbb{R}} G_p(\lambda) \geq \min\{P_c b_c^2, P_t b_t^2\} > 0. \tag{35}$$

Shells are required because a ray has zero probability when $d > 1$. Under the present convention, the endpoint *decreases* with λ, which determines the directions of the multiplier inequalities. □

B.3 Symmetric two-component location GMM

Theorem 13 (Positive excess MSE over all constant multipliers for a symmetric two-component GMM)**.** *Fix $d \geq 2$, $\sigma^2 > 0$, $\tau > 0$, and $\mu \in \mathbb{R}^d \setminus \{0\}$. For*

$$p_0(x) = \tfrac{1}{2}\mathcal{N}(x;\mu,\sigma^2 I_d) + \tfrac{1}{2}\mathcal{N}(x;-\mu,\sigma^2 I_d), \tag{36}$$

no finite constant multiplier exactly recovers the posterior mean, and

$$\inf_{\lambda\in\mathbb{R}} G_p(\lambda) > 0. \tag{37}$$

Proof. Write $v_t = \sigma^2 + t$, $\alpha = \sigma^2/(\sigma^2+\tau) \in (0,1)$, $e = \mu/\|\mu\|$, and $P_\perp = I_d - ee^\top$. Differentiation of the mixture density gives

$$\bar{s}_t(y) = -\frac{y}{v_t} + \frac{\mu}{v_t}\tanh\left(\frac{\mu^\top y}{v_t}\right), \tag{38}$$

$$m_\tau(y) = \alpha y + (1-\alpha)\mu\tanh\left(\frac{\mu^\top y}{v_\tau}\right). \tag{39}$$

Gaussian coordinates require $\lambda = 2$. Since $P_\perp \mu = 0$, the perpendicular ODE is linear, and

$$P_\perp T_\lambda(y) = \alpha^{\lambda/2} P_\perp y, \qquad P_\perp m_\tau(y) = \alpha P_\perp y. \tag{40}$$

The observation covariance in this $(d-1)$-dimensional subspace is $v_\tau P_\perp$. Consequently,

$$G_p(\lambda) \geq (d-1) v_\tau (\alpha^{\lambda/2} - \alpha)^2. \tag{41}$$

This is positive unless $\lambda = 2$.

The origin derivatives differ at $\lambda = 2$. The origin is a stationary trajectory, and

$$D\bar{s}_t(0) = -v_t^{-1} I_d + v_t^{-2} \mu\mu^\top.$$

These matrices commute at different times. Define

$$Q = \|\mu\|^2 \left(\frac{1}{\sigma^2} - \frac{1}{\sigma^2 + \tau} \right) = \frac{\tau \|\mu\|^2}{\sigma^2(\sigma^2 + \tau)} > 0. \tag{42}$$

The endpoint variational equation and the posterior derivative yield

$$e^\top DT_\lambda(0) e = \alpha^{\lambda/2} e^{\lambda Q/2}, \qquad e^\top Dm_\tau(0) e = \alpha(1+Q). \tag{43}$$

At $\lambda = 2$, the first value is αe^Q, which is strictly larger than $\alpha(1+Q)$. A derivative mismatch makes the maps differ at points arbitrarily close to the origin. By continuity, they differ on a nonempty open ball. Since $\bar{p}_\tau > 0$, this proves $G_p(2) > 0$.

Perpendicular coordinates give a uniform lower bound. By Lemma 8, choose $\delta \in (0,1)$ small enough that

$$|\lambda - 2| \leq \delta \quad \Longrightarrow \quad G_p(\lambda) \geq \tfrac{1}{2} G_p(2) > 0.$$

For $|\lambda - 2| \geq \delta$, equation 41 is bounded below by

$$c_\perp(\delta) = (d-1) v_\tau \min\Big\{ (\alpha^{1-\delta/2} - \alpha)^2, (\alpha - \alpha^{1+\delta/2})^2 \Big\} > 0. \tag{44}$$

Thus

$$\inf_{\lambda \in \mathbb{R}} G_p(\lambda) \geq \min\{ \tfrac{1}{2} G_p(2), c_\perp(\delta) \} > 0. \tag{45}$$

The perpendicular bound controls all multipliers outside the chosen neighborhood of $\lambda = 2$. □

Remark (Opposite derivative mismatches in the two examples). For a nontrivial isotropic finite GMM, $DT_2(0)$ is *smaller* than the posterior-mean Jacobian. Along the mixture direction of the symmetric location GMM, it is *larger*. Thus, the derivative error has opposite signs in the two examples.

B.4 Finite common-covariance GMMs

With a shared positive-definite covariance matrix, exact recovery fails without assuming equal weights, two components, or spherical covariance.

Theorem 14 (Failure of exact recovery for finite common-covariance GMMs)**.** *Fix $d \geq 1$ and $\tau > 0$. Let*

$$p_0(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Sigma), \qquad \pi_k > 0, \quad \sum_k \pi_k = 1, \quad \Sigma \succ 0, \tag{46}$$

where $K \geq 2$ and at least two component means are distinct. Merge components with identical means before applying the theorem. For the exact score-scaled ODE in equation 20, no finite constant multiplier $\lambda \in \mathbb{R}$ satisfies

$$T_\lambda = m_\tau \qquad \bar{p}_\tau\text{-almost everywhere}. \tag{47}$$

Consequently $G_p(\lambda) > 0$ for every finite λ, and

$$\inf_{\lambda \in \mathbb{R}} G_p(\lambda) > 0. \tag{48}$$

Let $\varphi(y;\mu,C) := \mathcal{N}(y;\mu,C)$ denote a Gaussian density. Put

$$C_t = \Sigma + tI_d, \qquad w_k(y,t) = \frac{\pi_k\,\varphi(y;\mu_k,C_t)}{\sum_j \pi_j\,\varphi(y;\mu_j,C_t)}, \qquad \bar{\mu}_t(y) = \sum_k w_k(y,t)\mu_k. \tag{49}$$

Then the exact score and posterior mean are

$$\bar{s}_t(y) = -C_t^{-1}\big(y - \bar{\mu}_t(y)\big), \qquad m_\tau(y) = y - \tau C_\tau^{-1}\big(y - \bar{\mu}_\tau(y)\big). \tag{50}$$

Let

$$A = \Sigma C_\tau^{-1} = I_d - \tau C_\tau^{-1}. \tag{51}$$

Lemma 15 (Tail asymptotics require $\lambda = 2$ under exact recovery)**.** *If $T_\lambda = m_\tau$ as maps for some finite λ, then $\lambda = 2$.*

Proof. Diagonalize Σ. Because the means are not all identical, there exists an eigenspace E of Σ on which their orthogonal projections are not all equal. Let ν_E be the corresponding eigenvalue. Choose a unit vector $v \in E$ outside the finitely many hyperplanes $\{v : v^\top(\mu_j - \mu_k) = 0\}$ for pairs whose projections onto E differ. Put

$$a_k = v^\top \mu_k, \qquad a_{\max} = \max_k a_k, \qquad H = \{k : a_k = a_{\max}\}.$$

The maximizing set H may contain several components. On E, A acts by the scalar

$$\alpha_E = \frac{\nu_E}{\nu_E + \tau} \in (0,1). \tag{52}$$

Consider observations $y_R = Rv$ as $R \to \infty$. Since $\bar{\mu}_t(x)$ lies in the convex hull of the finitely many component means, it is uniformly bounded. Variation of constants in the common-covariance ODE therefore gives, for each fixed finite λ,

$$y_t^{(\lambda)}(Rv) = \big(C_t C_\tau^{-1}\big)^{\lambda/2} Rv + O_\lambda(1) \tag{53}$$

uniformly for $t \in [0,\tau]$. With $P_\perp = I_d - vv^\top$ and

$$\beta_\lambda(t) = \left(\frac{\nu_E + t}{\nu_E + \tau}\right)^{\lambda/2},$$

this implies

$$v^\top y_t^{(\lambda)}(Rv) = \beta_\lambda(t)R + O_\lambda(1), \qquad P_\perp y_t^{(\lambda)}(Rv) = O_\lambda(1).$$

The continuous positive function β_λ has a positive minimum on $[0,\tau]$. For $h \in H$ and $k \notin H$, the common covariance gives

$$\log \frac{w_k(y_t^{(\lambda)}(Rv),t)}{w_h(y_t^{(\lambda)}(Rv),t)} = -\frac{(a_{\max} - a_k)\beta_\lambda(t)}{\nu_E + t} R + O_\lambda(1).$$

Because the mixture is finite and $a_{\max} - a_k > 0$, the total weight outside H is $O_\lambda(e^{-c_\lambda R})$ uniformly in t, for some $c_\lambda > 0$. Every component in H has the same v-projection $a_{\max}$, so

$$v^\top \bar{\mu}_t(y_t^{(\lambda)}(Rv)) = a_{\max} + O_\lambda(e^{-c_\lambda R}).$$

Solving the projected flow equation yields

$$v^\top T_\lambda(Rv) = a_{\max} + \alpha_E^{\lambda/2}(R - a_{\max}) + O_\lambda(e^{-c_\lambda R}). \tag{54}$$

The posterior weights at time τ give

$$v^\top m_\tau(Rv) = a_{\max} + \alpha_E(R - a_{\max}) + O(e^{-cR}) \tag{55}$$

for some $c > 0$. Equality of the two maps implies equality of the leading coefficients, $\alpha_E^{\lambda/2} = \alpha_E$, and therefore $\lambda = 2$. □

Lemma 16 (Two-component tail mismatch at $\lambda = 2$)**.** *For the nontrivial mixture equation 46, $T_2 \neq m_\tau$.*

Proof. Choose a vector $v \in \mathbb{R}^d$ outside the finitely many hyperplanes

$$v^\top C_\tau^{-1}(\mu_j - \mu_k) = 0. \tag{56}$$

After relabeling, $\mu_a^\top C_\tau^{-1} v$ is the unique largest projection and $\mu_b^\top C_\tau^{-1} v$ is the unique second largest. Put

$$d_{ab} = \mu_a - \mu_b \neq 0, \qquad y_R = Rv, \tag{57}$$

and let L_R be the unnormalized posterior ratio of component b to component a at (y_R, τ):

$$L_R = \frac{\pi_b \varphi(y_R; \mu_b, C_\tau)}{\pi_a \varphi(y_R; \mu_a, C_\tau)}. \tag{58}$$

Then $L_R \downarrow 0$, and every other component ratio is $o(L_R)$. Consequently

$$\bar{\mu}_\tau(y_R) = \mu_a - d_{ab} L_R + o(L_R), \tag{59}$$

and equation 50 gives

$$m_\tau(y_R) = \mu_a + A(y_R - \mu_a) - \tau C_\tau^{-1} d_{ab}\, L_R + o(L_R). \tag{60}$$

For $\lambda = 2$, the component-a Gaussian reference path is

$$g_t = \mu_a + C_t C_\tau^{-1}(y_R - \mu_a). \tag{61}$$

Cancellation shows that the b/a density ratio along this path is

$$\frac{\pi_b \varphi(g_t; \mu_b, C_t)}{\pi_a \varphi(g_t; \mu_a, C_t)} = L_R h(t), \qquad h(t) = \exp\left[-\frac{1}{2} d_{ab}^\top (C_t^{-1} - C_\tau^{-1}) d_{ab}\right]. \tag{62}$$

The leading tail-exponent order is therefore independent of t. All remaining component ratios are $o(L_R)$ uniformly on $[0, \tau]$. Linearizing the exact trajectory around g_t and using variation of constants yields

$$T_2(y_R) = \mu_a + A(y_R - \mu_a) - \mathcal{M} d_{ab}\, L_R + o(L_R), \tag{63}$$

where

$$\mathcal{M} = \Sigma \int_0^\tau C_t^{-2} h(t)\, dt. \tag{64}$$

The terminal condition at $t = \tau$ determines the sign in equation 63. If η_t is the coefficient of L_R, then $\eta_t' = C_t^{-1}\eta_t + C_t^{-1} d_{ab} h(t)$ and $\eta_\tau = 0$, so $\eta_0 = -\mathcal{M} d_{ab}$.

To compare the two first-order tail corrections, define

$$Q = d_{ab}^\top (\Sigma^{-1} - C_\tau^{-1}) d_{ab} = \tau d_{ab}^\top \Sigma^{-1} C_\tau^{-1} d_{ab} > 0. \tag{65}$$

Since

$$h'(t) = \frac{1}{2} h(t) d_{ab}^\top C_t^{-2} d_{ab}, \qquad h(\tau) = 1, \qquad h(0) = e^{-Q/2}, \tag{66}$$

projection onto $\Sigma^{-1} d_{ab}$ gives

$$(\Sigma^{-1} d_{ab})^\top \mathcal{M} d_{ab} = \int_0^\tau h(t) d_{ab}^\top C_t^{-2} d_{ab}\, dt = 2(1 - e^{-Q/2}), \tag{67}$$

$$(\Sigma^{-1} d_{ab})^\top \tau C_\tau^{-1} d_{ab} = Q. \tag{68}$$

For every $Q > 0$,

$$2(1 - e^{-Q/2}) < Q. \tag{69}$$

Thus $\mathcal{M} d_{ab} \neq \tau C_\tau^{-1} d_{ab}$. The first non-vanishing tail corrections in equation 60 and equation 63 differ, so $T_2(y_R) \neq m_\tau(y_R)$ for all sufficiently large R. □

Lemma 17 (Lower bound for large multipliers)**.** *For the source in Theorem 14, there are $L < \infty$ and $c_\infty > 0$ such that*

$$|\lambda| \geq L \quad \Longrightarrow \quad G_p(\lambda) \geq c_\infty. \tag{70}$$

Proof. Let v be any unit eigenvector of Σ with eigenvalue $\nu > 0$, and put

$$a_- = \min_k v^\top \mu_k, \qquad a_+ = \max_k v^\top \mu_k, \qquad \alpha = \frac{\nu}{\nu + \tau} \in (0, 1). \tag{71}$$

Use reverse elapsed time $\xi = \tau - t$ and write $x_\xi = y^{(\lambda)}_{\tau-\xi}(y)$ and $z_\xi = v^\top x_\xi$. From equation 50,

$$\dot{z}_\xi = -c_\lambda(\xi)\big(z_\xi - \bar{a}_\xi\big), \qquad c_\lambda(\xi) = \frac{\lambda}{2(\nu + \tau - \xi)}, \qquad \bar{a}_\xi = v^\top \bar{\mu}_{\tau-\xi}(x_\xi) \in [a_-, a_+]. \tag{72}$$

The variation-of-constants formula gives, with

$$\beta_\lambda = \exp\left[-\int_0^\tau c_\lambda(\xi)\, d\xi\right] = \alpha^{\lambda/2}, \tag{73}$$

that

$$v^\top T_\lambda(y) = \beta_\lambda v^\top y + \int_0^\tau k_\lambda(\xi)\bar{a}_\xi\, d\xi, \qquad \int_0^\tau k_\lambda(\xi)\, d\xi = 1 - \beta_\lambda, \tag{74}$$

where k_λ has the sign of λ. Hence, for $\lambda > 0$,

$$\beta_\lambda z + (1 - \beta_\lambda)a_- \le v^\top T_\lambda(y) \le \beta_\lambda z + (1 - \beta_\lambda)a_+, \tag{75}$$

and for $\lambda < 0$ the endpoint inequalities reverse in the forcing term,

$$\beta_\lambda z + (1 - \beta_\lambda)a_+ \le v^\top T_\lambda(y) \le \beta_\lambda z + (1 - \beta_\lambda)a_-. \tag{76}$$

$z = v^\top y$.

The posterior mean obeys the fixed-α bounds

$$\alpha z + (1 - \alpha)a_- \le v^\top m_\tau(y) \le \alpha z + (1 - \alpha)a_+. \tag{77}$$

Choose $R > a_+$ so large that

$$g := \alpha R + (1 - \alpha)a_- - a_+ > 0, \tag{78}$$

and let $B = \{y : R \le v^\top y \le R + 1\}$. Since $\bar{p}_\tau$ is a strictly positive Gaussian-mixture density, $P_B := \mathbb{P}(Y_\tau \in B) > 0$.

As $\lambda \to +\infty$, $\beta_\lambda \to 0$. For all sufficiently large positive λ, equation 75 gives on B

$$v^\top T_\lambda(y) \le a_+ + \frac{g}{2}, \qquad v^\top m_\tau(y) \ge a_+ + g, \tag{79}$$

so the squared error is at least $g^2/4$ there.

As $\lambda \to -\infty$, $\beta_\lambda \to \infty$. From equation 76, on the same slab

$$v^\top T_\lambda(y) \ge a_+ + \beta_\lambda(R - a_+), \tag{80}$$

whereas equation 77 gives the upper bound $v^\top m_\tau(y) \le \alpha(R+1) + (1-\alpha)a_+$. Hence the projected error is at least one for negative multipliers with sufficiently large $|\lambda|$. Multiplying the smaller of these two pointwise bounds by P_B proves equation 70. □

Proof of Theorem 14. If equality holds $\bar{p}_\tau$-almost everywhere, continuity of both maps and strict positivity of the Gaussian-mixture density imply equality everywhere. By Lemma 15, equality requires $\lambda = 2$, which is excluded by Lemma 16. Hence $G_p(\lambda) > 0$ for every finite λ. By Lemma 17, the mean-squared endpoint error is uniformly positive outside a compact multiplier interval. On the remaining compact interval, continuity from Lemma 8 and pointwise positivity give a strictly positive minimum. This proves equation 48. □

B.5 FINITE GMMS WITH UNEQUAL COVARIANCES

For unequal component covariances, a maximum-variance tail requires $\lambda = 2$ under exact recovery. At this multiplier, a generic tail gives different non-Gaussian corrections. If the two leading components have different covariances, the leading correction comes from an $O(R^{-2})$ terminal boundary layer along the dominant Gaussian trajectory. The score-scaled probability-flow ODE accumulates an asymptotically smaller correction. If the leading components share a covariance, the common-covariance two-component calculation applies.

Theorem 18 (Positive excess MSE over all constant multipliers for every nontrivial finite GMM)**.** *Fix $d \geq 1$ and $\tau > 0$. Let*

$$p_0(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Sigma_k), \qquad \pi_k > 0, \quad \sum_k \pi_k = 1, \quad \Sigma_k \succ 0, \tag{81}$$

with $K \geq 2$ after merging components having the same pair (μ_k, Σ_k). Then no finite constant multiplier $\lambda \in \mathbb{R}$ makes $T_\lambda = m_\tau$ $\bar{p}_\tau$-almost everywhere. Moreover,

$$\inf_{\lambda \in \mathbb{R}} G_p(\lambda) > 0. \tag{82}$$

Thus the single-Gaussian case is the only finite-GMM member of the exact VP family for which a constant score multiplier reaches the posterior mean.

Write

$$C_{k,t} = \Sigma_k + tI_d, \qquad f_k(y,t) = \pi_k \varphi(y; \mu_k, C_{k,t}), \qquad w_k(y,t) = \frac{f_k(y,t)}{\sum_j f_j(y,t)}. \tag{83}$$

Then

$$\bar{s}_t(y) = -\sum_k w_k(y,t) C_{k,t}^{-1}(y - \mu_k), \qquad m_\tau(y) = \sum_k w_k(y,\tau) m_{k,\tau}(y), \tag{84}$$

where the componentwise Gaussian posterior means are

$$m_{k,\tau}(y) = \mu_k + \Sigma_k C_{k,\tau}^{-1}(y - \mu_k) = y - \tau C_{k,\tau}^{-1}(y - \mu_k). \tag{85}$$

Lemma 19 (Maximum-variance tail asymptotics require $\lambda = 2$)**.** *If $T_\lambda = m_\tau$ as maps for a finite multiplier, then $\lambda = 2$.*

Proof. Let

$$S = \max_k \lambda_{\max}(\Sigma_k).$$

Choose an index whose covariance has eigenvalue S and a corresponding unit eigenvector v. Define

$$\mathcal{I}_{\max} = \{k : \Sigma_k v = Sv\}, \qquad a_{\max} = \max_{k \in \mathcal{I}_{\max}} v^\top \mu_k, \qquad H = \{k \in \mathcal{I}_{\max} : v^\top \mu_k = a_{\max}\}.$$

The set H is nonempty. For $k \notin \mathcal{I}_{\max}$ and every $t \in [0, \tau]$,

$$v^\top C_{k,t}^{-1} v > \frac{1}{S+t}, \tag{86}$$

whereas equality holds for $k \in \mathcal{I}_{\max}$. For $k \in \mathcal{I}_{\max} \setminus H$, the quadratic coefficient along the ray is the same, but the linear coefficient is strictly smaller than that of H. Hence, for observations with a large positive v-coordinate and bounded orthogonal coordinate, components outside H are suppressed respectively as e^{-cR^2} or e^{-cR}, uniformly in $t \in [0, \tau]$.

For $k \in H$, symmetry of Σ_k and $\Sigma_k v = Sv$ imply the block factorization

$$\Sigma_k = Svv^\top + P_\perp \Sigma_k P_\perp, \qquad P_\perp = I_d - vv^\top,$$

and

$$v^\top \{-C_{k,t}^{-1}(x - \mu_k)\} = -\frac{v^\top x - a_{\max}}{S+t}. \tag{87}$$

The orthogonal score of every component in H contains no term proportional to $v^\top x$ and has at most linear growth in $P_\perp x$.

Fix a finite λ and start the reverse-time equation $x_\xi = y^{(\lambda)}_{\tau-\xi}(Rv)$ at Rv. Write $z_\xi = v^\top x_\xi$ and $u_\xi = P_\perp x_\xi$. Uniformly for $\xi \in [0,\tau]$,

$$u_\xi = O_\lambda(1), \qquad z_\xi = a_{\max} + \beta_\lambda(\xi)(R - a_{\max}) + o(R), \qquad \beta_\lambda(\xi) = \left(\frac{S+\tau-\xi}{S+\tau}\right)^{\lambda/2}. \tag{88}$$

To prove equation 88, put $b_\lambda = \min_{\xi\in[0,\tau]} \beta_\lambda(\xi) > 0$ and $c_\lambda = b_\lambda/2$. On a tube $z \geq c_\lambda R$, $\|u\| \leq C$, the score satisfies

$$v^\top \bar{s}_t(zv+u) = -\frac{z - a_{\max}}{S+t} + \rho_{R,v}, \qquad \|P_\perp \bar{s}_t(zv+u)\| \leq A\|u\| + B + \|\rho_{R,\perp}\|, \tag{89}$$

where A, B do not depend on R and $\sup \|\rho_R\| =: e_R(C) \to 0$. The mixture restricted to H has the common projected score in equation 87, its orthogonal score has linear growth in u, and every discarded component has exponentially small weight on the tube.

Let $\ell = |\lambda|/2$ and choose $U_\lambda = \ell\tau(B+1)e^{\ell A\tau}$ and $C = U_\lambda + 1$. Until the first exit from this tube, Grönwall's inequality gives $\|u_\xi\| \leq U_\lambda < C$ once $e_R(C) \leq 1$. Variation of constants in the projected equation gives

$$z_\xi = a_{\max} + \beta_\lambda(\xi)(R - a_{\max}) + O_\lambda(e_R(C)) \tag{90}$$

uniformly in ξ. For sufficiently large R, the right-hand side exceeds $c_\lambda R$. Thus neither boundary can be the first exit, which proves equation 88 without assuming commutativity.

At the endpoint,

$$v^\top T_\lambda(Rv) = a_{\max} + \alpha^{\lambda/2}(R - a_{\max}) + o(R), \qquad \alpha = \frac{S}{S+\tau} \in (0,1). \tag{91}$$

The dominance argument at the observation time and Tweedie's formula give

$$v^\top m_\tau(Rv) = a_{\max} + \alpha(R - a_{\max}) + o(R). \tag{92}$$

Equality of the maps therefore requires $\alpha^{\lambda/2} = \alpha$, hence $\lambda = 2$. □

Lemma 20 (Uniform estimates along a dominant Gaussian tail)**.** *Choose a unit vector e such that*

$$e^\top (C_{j,\tau}^{-1} - C_{k,\tau}^{-1}) e \neq 0 \quad \textit{when } \Sigma_j \neq \Sigma_k,$$

and

$$e^\top C_{j,\tau}^{-1}(\mu_j - \mu_k) \neq 0 \quad \textit{when } \Sigma_j = \Sigma_k \textit{ and } \mu_j \neq \mu_k.$$

For $y_R = Re$, order the components by increasing $\kappa_k = e^\top C_{k,\tau}^{-1} e$ and then decreasing $\ell_k = e^\top C_{k,\tau}^{-1}\mu_k$. Let a and b be the first two components, and set

$$L_R = \frac{f_b(y_R,\tau)}{f_a(y_R,\tau)}, \qquad g_t = \mu_a + C_{a,t}C_{a,\tau}^{-1}(y_R - \mu_a). \tag{93}$$

Let $x_t = y_t^{(2)}(y_R)$ and $\delta_t = x_t - g_t$. The estimates below hold as $R \to \infty$.

If $\Sigma_b \neq \Sigma_a$, uniformly for $t \in [0,\tau]$,

$$\sup_{0\leq t\leq\tau} \|\delta_t\| = O(L_R/R), \qquad T_2(y_R) = m_{a,\tau}(y_R) + O(L_R/R). \tag{94}$$

If $\Sigma_b = \Sigma_a =: \Sigma$, put

$$d_{ab} = \mu_a - \mu_b, \qquad C_t = \Sigma + tI_d,$$

and

$$h(t) = \exp\left[-\frac{1}{2} d_{ab}^\top (C_t^{-1} - C_\tau^{-1}) d_{ab}\right]. \tag{95}$$

Then, uniformly for $t \in [0,\tau]$,

$$\sup_t \|\delta_t\| = O(L_R), \qquad w_b(x_t,t) = L_R h(t) + o(L_R), \tag{96}$$

and

$$T_2(y_R) = m_{a,\tau}(y_R) - \mathcal{M} d_{ab} L_R + o(L_R), \qquad \mathcal{M} = \Sigma \int_0^\tau C_t^{-2} h(t)\, dt. \tag{97}$$

Proof. The strict ordering implies that every terminal ratio $L_{k,R} = f_k(y_R, \tau)/f_a(y_R, \tau)$ with $k \neq a, b$ is exponentially $o(L_R)$. Put $P_{k,t} = C_{k,t}^{-1}$ and

$$q_k(x,t) = P_{k,t}(x - \mu_k) - P_{a,t}(x - \mu_a).$$

Subtracting the component-a equation from the mixture equation and applying variation of constants gives

$$\delta_t = -C_{a,t} \int_t^\tau C_{a,s}^{-1} \sum_{k \neq a} w_k(x_s, s) q_k(x_s, s)\, ds. \tag{98}$$

If $\Sigma_b \neq \Sigma_a$, genericity of e gives $\kappa_b > \kappa_a$. Put $D = \Sigma_a - \Sigma_b$ and $w = C_{a,\tau}^{-1} e$. The coefficient of R^2 in the log-density penalty of b relative to a along g_t is

$$H_{ab}(t) = w^\top \{C_{a,t} C_{b,t}^{-1} C_{a,t} - C_{a,t}\} w \tag{99}$$

$$= w^\top \{D + D C_{b,t}^{-1} D\} w. \tag{100}$$

At $t = \tau$, $H_{ab}(\tau) = \kappa_b - \kappa_a > 0$, and

$$H'_{ab}(t) = -\|C_{b,t}^{-1} D w\|^2 < 0. \tag{101}$$

The strict inequality follows because $Dw = 0$ implies $H_{ab}(\tau) = 0$. Compactness of $[0, \tau]$ therefore gives $c > 0$ such that

$$H_{ab}(t) - H_{ab}(\tau) \geq c(\tau - t). \tag{102}$$

The lower-order mean and determinant terms vary by $O(R(\tau - t)) + O(\tau - t)$. Hence, for all sufficiently large R,

$$\frac{f_b(g_t, t)}{f_a(g_t, t)} \leq C L_R e^{-cR^2(\tau - t)/4}. \tag{103}$$

Because b is the second component and $\kappa_b > \kappa_a$, no component other than a has covariance Σ_a. The same calculation for every $k \neq a$ yields

$$\frac{f_k(g_t, t)}{f_a(g_t, t)} \leq C_k L_{k,R} e^{-c_k R^2(\tau - t)}, \qquad L_{k,R} = O(L_R), \tag{104}$$

with $L_{k,R} = o(L_R)$ for $k \neq a, b$.

On the tube $\sup_t \|\delta_t\| \leq R^{-1}$, Gaussian log-ratios satisfy

$$\left| \log \frac{(f_k/f_a)(x_t, t)}{(f_k/f_a)(g_t, t)} \right| \leq C\big(R\|\delta_t\| + \|\delta_t\|^2\big). \tag{105}$$

Thus the exact-path ratios are bounded by fixed multiples of equation 104, while $\|q_k(x_t, t)\| \leq CR$. Equation equation 98 gives

$$\sup_{0 \leq t \leq \tau} \|\delta_t\| \leq \frac{C}{R} \sum_{k \neq a} L_{k,R} = O(L_R/R). \tag{106}$$

For large R, this estimate closes the first-exit argument. Since $g_0 = m_{a,\tau}(y_R)$, the first part follows.

If $\Sigma_b = \Sigma_a = \Sigma$, the exact Gaussian ratio along g_t is $f_b(g_t, t)/f_a(g_t, t) = L_R h(t)$. For every remaining component with covariance Σ, its ratio is a bounded function of t times a terminal ratio that is $o(L_R)$. Components with different covariances satisfy a boundary-layer bound of the form equation 104. Applying equation 98 and equation 105 on the same tube yields

$$\sup_t \|\delta_t\| \leq C \left(\sum_{\substack{k \neq a \\ \Sigma_k = \Sigma_a}} L_{k,R} + \frac{1}{R} \sum_{\Sigma_k \neq \Sigma_a} L_{k,R} \right) = O(L_R). \tag{107}$$

Since $RL_R \to 0$, this estimate closes the bootstrap. It gives $w_b(x_t, t) = L_R h(t) + o(L_R)$ uniformly in t. All other integrated contributions are $o(L_R)$. Their differential remainders are bounded by $CRL_{k,R} = o(L_R)$ for $\Sigma_k \neq \Sigma_a$ and by $CL_{k,R} = o(L_R)$ for $\Sigma_k = \Sigma_a$, $k \neq a, b$. Dividing the perturbation equation by L_R gives

$$\left(\frac{\delta_t}{L_R} \right)' = C_t^{-1} \frac{\delta_t}{L_R} + C_t^{-1} d_{ab} h(t) + o(1), \qquad \delta_\tau = 0, \tag{108}$$

uniformly on $[0, \tau]$. Variation of constants proves equation 97. □

Lemma 21 (Two-component tail mismatch at $\lambda = 2$)**.** *For every mixture in Equation* (81) *with at least two distinct components,* $T_2 \neq m_\tau$.

Proof. Choose a generic unit vector e. For every pair with $\Sigma_j \neq \Sigma_k$, avoid the proper quadratic zero set

$$e^\top (C_{j,\tau}^{-1} - C_{k,\tau}^{-1})e = 0,$$

and for every pair with equal covariance but distinct means, avoid the proper hyperplane

$$e^\top C_{j,\tau}^{-1}(\mu_j - \mu_k) = 0.$$

For $d \geq 2$ a finite union of these proper algebraic subsets cannot cover the sphere. In $d = 1$, the same conclusion follows after duplicate components are merged: unequal variances give unequal quadratic coefficients, and equal variances with unequal means are ordered in one of the two tails.

For $y_R = Re$, write

$$\kappa_k = e^\top C_{k,\tau}^{-1} e, \qquad \ell_k = e^\top C_{k,\tau}^{-1}\mu_k. \tag{109}$$

Since

$$\log f_k(y_R, \tau) = -\tfrac{1}{2}\kappa_k R^2 + \ell_k R + O(1),$$

the components have a strict lexicographic order by increasing κ_k and then decreasing ℓ_k. Let a and b denote the first two components in this order and set

$$L_R = \frac{f_b(y_R, \tau)}{f_a(y_R, \tau)} \longrightarrow 0. \tag{110}$$

All other component ratios are exponentially $o(L_R)$. Apply Lemma 20 to this ordering.

If $\Sigma_b \neq \Sigma_a$, equation 94 gives

$$T_2(y_R) = m_{a,\tau}(y_R) + o(RL_R). \tag{111}$$

The posterior weights at time τ give

$$m_\tau(y_R) = m_{a,\tau}(y_R) + L_R\{m_{b,\tau}(y_R) - m_{a,\tau}(y_R)\} + o(RL_R). \tag{112}$$

By equation 85,

$$m_{b,\tau}(y_R) - m_{a,\tau}(y_R) = \tau R(C_{a,\tau}^{-1} - C_{b,\tau}^{-1})e + O(1). \tag{113}$$

The leading vector is nonzero because its inner product with e equals $\tau(\kappa_a - \kappa_b) \neq 0$. The posterior therefore has a $\Theta(RL_R)$ correction absent from the score-scaled probability-flow endpoint. Hence $T_2(y_R) \neq m_\tau(y_R)$ for all sufficiently large R.

If $\Sigma_b = \Sigma_a =: \Sigma$, then $\kappa_b = \kappa_a$ and $\ell_a > \ell_b$. Put $d_{ab} = \mu_a - \mu_b \neq 0$ and $C_t = \Sigma + tI_d$. The uniform estimate equation 97 gives the flow expansion, whereas the posterior weights give

$$m_\tau(y_R) = m_{a,\tau}(y_R) - \tau C_\tau^{-1} d_{ab} L_R + o(L_R). \tag{114}$$

Set

$$Q = d_{ab}^\top(\Sigma^{-1} - C_\tau^{-1})d_{ab} > 0. \tag{115}$$

Since $h'(t) = \frac{1}{2}h(t)d_{ab}^\top C_t^{-2} d_{ab}$,

$$(\Sigma^{-1}d_{ab})^\top \mathcal{M} d_{ab} = 2(1 - e^{-Q/2}), \tag{116}$$

$$(\Sigma^{-1}d_{ab})^\top \tau C_\tau^{-1} d_{ab} = Q. \tag{117}$$

The strict inequality $2(1 - e^{-Q/2}) < Q$ proves that the two first-order corrections differ. Hence $T_2 \neq m_\tau$. □

Proof of Theorem 18. Both maps are continuous and $\bar{p}_\tau$ is strictly positive. Almost-everywhere equality therefore implies equality everywhere. By Lemma 19, equality at a finite multiplier requires $\lambda = 2$, but Lemma 21 proves inequality at that value. Thus $G_p(\lambda) > 0$ for every finite λ.

For a uniform lower bound over all $\lambda \in \mathbb{R}$, apply Lemma 24 with $a_{\text{eig}} = \min_k \lambda_{\min}(\Sigma_k)$, $b_{\text{eig}} = \max_k \lambda_{\max}(\Sigma_k)$, and $M = \max_k \|\mu_k\|$. It supplies $L < \infty$ and $c_{\text{tail}} > 0$ such that $G_p(\lambda) \geq c_{\text{tail}}$ for $|\lambda| \geq L$. On the remaining compact interval $[-L, L]$, continuity from Lemma 8 and strict positivity give

$$c_{\text{comp}} := \min_{|\lambda| \leq L} G_p(\lambda) > 0. \tag{118}$$

Consequently, $\inf_{\lambda \in \mathbb{R}} G_p(\lambda) \geq \min\{c_{\text{comp}}, c_{\text{tail}}\} > 0$. This proves equation 82. □

B.6 EXCESS-MSE PERTURBATION BOUND

The bound separates errors in the source, implemented score, and evaluation distribution. Let p_0 be a reference source, let q_0 be a target source with finite second moment, and let $q_\tau = q_0 * \mathcal{N}(0, \tau I_d)$ be its observation density. Use superscripts p, q for their exact scores and posterior means. Let $\widetilde{T}_\lambda$ be the implemented endpoint, not necessarily the exact q-endpoint. For a measurable observation set A, define the *unnormalized* restricted norm

$$\|g\|^2_{\bar{p}_\tau, A} = \int_A \|g(y)\|^2 \bar{p}_\tau(y)\, dy.$$

For a multiplier set $\Lambda \subseteq \mathbb{R}$, set

$$G_{p,A}(\Lambda) = \inf_{\lambda \in \Lambda} \|T^p_\lambda - m^p_\tau\|^2_{\bar{p}_\tau, A}. \tag{119}$$

Theorem 22 (Perturbation stability of a local excess-MSE lower bound)**.** *Suppose $G_{p,A}(\Lambda) > 0$, and suppose*

$$q_\tau(y) \geq \kappa \bar{p}_\tau(y) \quad \textit{for almost every } y \in A, \qquad \kappa > 0. \tag{120}$$

Assume finite error bounds

$$\sup_{\lambda \in \Lambda} \|\widetilde{T}_\lambda - T^p_\lambda\|_{\bar{p}_\tau, A} \leq \varepsilon_F, \qquad \|m^q_\tau - m^p_\tau\|_{\bar{p}_\tau, A} \leq \varepsilon_m. \tag{121}$$

Then

$$\inf_{\lambda \in \Lambda} \int_{\mathbb{R}^d} \|\widetilde{T}_\lambda(y) - m^q_\tau(y)\|^2 q_\tau(y)\, dy \;\geq\; \kappa [\sqrt{G_{p,A}(\Lambda)} - \varepsilon_F - \varepsilon_m]^2_+. \tag{122}$$

The target has positive excess MSE whenever $\varepsilon_F + \varepsilon_m < \sqrt{G_{p,A}(\Lambda)}$. The theorem allows $\Lambda = \mathbb{R}$ only when the endpoint error bound is uniform over all *finite multipliers.*

Proof. Density domination gives

$$\|\widetilde{T}_\lambda - m^q_\tau\|^2_{q_\tau} \geq \kappa \|\widetilde{T}_\lambda - m^q_\tau\|^2_{\bar{p}_\tau, A}.$$

Write the difference inside the latter norm as

$$(T^p_\lambda - m^p_\tau) + (\widetilde{T}_\lambda - T^p_\lambda) + (m^p_\tau - m^q_\tau).$$

The reverse triangle inequality bounds its norm below by $[\sqrt{G_{p,A}(\Lambda)} - \varepsilon_F - \varepsilon_m]_+$, uniformly in λ. Squaring and taking the infimum proves the result. By conditional orthogonality, the left-hand side is the MSE above the *target* source's MMSE, not above the reference source's MMSE. □

Proof of Proposition 3. For each $\lambda \in \Lambda$, the reverse triangle inequality in $L^2(P_Y)$ gives

$$\begin{aligned} \|\widetilde{T}_\lambda - m\|_{L^2(P_Y)} &\geq \|T_\lambda - m\|_{L^2(P_Y)} - \|\widehat{T}_\lambda - T_\lambda\|_{L^2(P_Y)} \\ &\quad - \|\widetilde{T}_\lambda - \widehat{T}_\lambda\|_{L^2(P_Y)} \\ &\geq \sqrt{G_{\Lambda,\min}} - \delta_{\text{approx}} - \delta_{\text{num}}. \end{aligned}$$

Taking the positive part, squaring, and taking the infimum over $\lambda \in \Lambda$ proves Equation (10). This is the same-source case of Theorem 22 with $A = \mathbb{R}^d$, $q = p$, $\kappa = 1$, and $\varepsilon_m = 0$. □

Remark (Half-margin bound). If $\varepsilon_F + \varepsilon_m \leq \frac{1}{2}\sqrt{G_{p,A}(\Lambda)}$, then the excess MSE is at least $\kappa G_{p,A}(\Lambda)/4$. The margin is measured in root-MSE units before it is squared.

B.6.1 ENDPOINT ERROR FROM SCORE ERROR ON BOUNDED MULTIPLIER SETS

Proposition 23 (Sufficient condition based on score error)**.** *Let $\Lambda = [-L, L]$ for $0 < L < \infty$ and let A be compact. Suppose both the reference ODE and the ODE driven by a field $\widetilde{s}_t$ exist and remain in a set $\Omega \supseteq A$, for all starts in A and all $\lambda \in \Lambda$. On this set suppose*

$$\|\bar{s}^p_t(x) - \bar{s}^p_t(x')\| \leq L_s \|x - x'\|, \qquad x, x' \in \Omega, \quad 0 \leq t \leq \tau, \tag{123}$$

$$\sup_{t \in [0,\tau],\, x \in \Omega} \|\widetilde{s}_t(x) - \bar{s}^p_t(x)\| \leq \epsilon_{\text{flow}}, \qquad \sup_{y \in A} \|\bar{s}^q_\tau(y) - \bar{s}^p_\tau(y)\| \leq \epsilon_{\text{post}}. \tag{124}$$

Set

$$C_{\mathrm{flow}} = \begin{cases} (e^{LL_s\tau/2} - 1)/L_s, & L_s > 0, \\ L\tau/2, & L_s = 0, \end{cases} \qquad w_A = \int_A \bar{p}_\tau(y)\, dy. \tag{125}$$

If the computed endpoint has uniform integration error at most ϵ_{num} relative to the $\widetilde{s}$-ODE endpoint, then Theorem 22 applies with

$$\varepsilon_F + \varepsilon_m \le \sqrt{w_A}\left(C_{\mathrm{flow}}\epsilon_{\mathrm{flow}} + \epsilon_{\mathrm{num}} + \tau\epsilon_{\mathrm{post}}\right). \tag{126}$$

Density domination equation 120 and a positive local reference lower bound require separate verification.

Proof. In elapsed time $\xi = \tau - t$, subtract the two trajectory equations. With equal starting points and $|\lambda| \le L$, their difference obeys the integral inequality

$$d(\xi) \le \frac{L}{2}\int_0^\xi \left(L_s d(v) + \epsilon_{\mathrm{flow}}\right) dv.$$

Grönwall gives $d(\tau) \le C_{\mathrm{flow}}\epsilon_{\mathrm{flow}}$. Adding the numerical endpoint error and integrating over A gives the ε_F bound. Tweedie's identity for the target score gives $\|m_\tau^q - m_\tau^p\| \le \tau\epsilon_{\mathrm{post}}$ on A. The two score errors track the implemented field relative to the reference score and the target posterior relative to the reference posterior. □

Sufficient condition for a bounded trajectory set. If both scores obey a common linear-growth bound $\|\bar{s}_t(x)\| \le c_0 + c_1\|x\|$, starts satisfy $\|y\| \le R$, and $|\lambda| \le L$, Grönwall gives a finite trajectory ball. Take this ball as Ω, with the score Jacobian bounded on it. Local Gaussian-mixture parameter neighborhoods satisfy such a growth bound by equation 23.

Compact observation set with a uniform lower bound. For a reference in Theorem 9, the union $H_c \cup H_t$ supplies a compact observation set uniformly over all $\lambda \in \mathbb{R}$. Suppose exact recovery fails for every reference multiplier in a compact Λ and $(\lambda, y) \mapsto T_\lambda^p(y) - m_\tau^p(y)$ is continuous, with $\bar{p}_\tau > 0$. For each λ, choose a point with nonzero residual and an observation ball on which the residual norm is bounded away from zero. Continuity preserves this bound for nearby multipliers. A finite subcover of Λ gives finitely many balls. Their union A satisfies $G_{p,A}(\Lambda) > 0$.

Remark (Extension beyond a bounded multiplier set). The constant in equation 125 grows with L. Extension to all $\lambda \in \mathbb{R}$ requires separate control of large positive and negative multipliers.

B.7 Stability to mixture-parameter perturbations

Consider a fixed finite number K of components and parameter vectors $\vartheta = (\pi_k, \mu_k, \Sigma_k)_{k=1}^K$ with positive weights summing to one and $\Sigma_k \succ 0$. Let

$$p_0^\vartheta(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x; \mu_k, \Sigma_k). \tag{127}$$

Write $G_\vartheta(\lambda)$ for the mean-squared distance under p_0^ϑ between the flow endpoint and posterior mean. This lower bound is locally stable around any finite-GMM reference for which exact recovery fails, including mixtures with non-spherical covariances.

Lemma 24 (Uniform control of large multipliers)**.** *Fix d, K, τ and constants $0 < a \le b < \infty$, $M < \infty$. For every mixture equation 127 satisfying*

$$aI_d \preceq \Sigma_k \preceq bI_d, \qquad \|\mu_k\| \le M, \tag{128}$$

there exist constants $L < \infty$ and $c_{\mathrm{tail}} > 0$, depending only on these bounds, d, and τ, such that

$$|\lambda| \ge L \quad \Longrightarrow \quad G_\vartheta(\lambda) \ge c_{\mathrm{tail}}. \tag{129}$$

Proof. Let $R = \|x\|$, let the posterior component weights be $\gamma_k(x,t)$, and put

$$c_s = \frac{1}{b+\tau}, \qquad b_s = \frac{M}{a}, \qquad R_{\mathrm{eq}} = \frac{b_s}{c_s}, \qquad c_m = \frac{a}{a+\tau}.$$

The exact score satisfies the common inward bound

$$\langle x, \bar{s}_t^{\vartheta}(x)\rangle = -\sum_k \gamma_k x^\top (\Sigma_k + tI_d)^{-1}(x-\mu_k) \le -c_s R^2 + b_s R. \tag{130}$$

In elapsed time ξ, the radial comparison for $\lambda \ge 0$ gives

$$\|T_\lambda^{\vartheta}(y)\| \le R_{\mathrm{eq}} + (\|y\| - R_{\mathrm{eq}})e^{-\lambda c_s \tau/2}. \tag{131}$$

The annulus below has $\|y\| > R_{\mathrm{eq}}$. For $\lambda < 0$, the reverse comparison keeps the radius above R_{eq} and gives

$$\|T_\lambda^{\vartheta}(y)\| \ge R_{\mathrm{eq}} + (\|y\| - R_{\mathrm{eq}})e^{|\lambda| c_s \tau/2}. \tag{132}$$

Interpret the norm differential inequality almost everywhere, or by its upper Dini derivative at a trajectory zero. Such a zero cannot occur in the negative-multiplier comparison.

The component posterior means are $A_k y + (I_d - A_k)\mu_k$, where $A_k = \Sigma_k(\Sigma_k + \tau I_d)^{-1}$ and $c_m I_d \preceq A_k \preceq I_d$. Taking the projection in the direction of y for the lower bound and the norm for the upper bound gives

$$c_m\|y\| - M \le \|m_\tau^{\vartheta}(y)\| \le \|y\| + M. \tag{133}$$

Choose $\ell > 0$ and $R_0 > R_{\mathrm{eq}} + \ell$ such that $c_m R_0 - M \ge R_{\mathrm{eq}} + 3\ell$, and let $R_1 = R_0 + \ell$. Define the fixed annulus $H = \{R_0 \le \|y\| \le R_1\}$. Choose L large enough that

$$(R_1 - R_{\mathrm{eq}})e^{-Lc_s\tau/2} \le \ell, \qquad R_{\mathrm{eq}} + (R_0 - R_{\mathrm{eq}})e^{Lc_s\tau/2} \ge R_1 + M + \ell. \tag{134}$$

This L exists and depends only on these constants. For $\lambda \ge L$ the target norm exceeds the endpoint norm by at least 2ℓ on H. For $\lambda \le -L$ the endpoint norm exceeds the target norm by at least ℓ. Thus the residual norm is at least ℓ in either range.

Every observation density in the class has the lower bound on H

$$\bar{p}_\tau^{\vartheta}(y) \ge [2\pi(b+\tau)]^{-d/2} \exp\left[-\frac{(R_1+M)^2}{2(a+\tau)}\right] =: p_H^{\min} > 0. \tag{135}$$

Therefore $c_{\mathrm{tail}} = \ell^2 p_H^{\min} \operatorname{vol}(H) > 0$ proves the lemma. □

Theorem 25 (Local stability of the uniform excess-MSE lower bound)**.** *Let ϑ_0 parameterize any nontrivial finite GMM covered by Theorem 18. Fix its dimension, component count, and $\tau > 0$. There exist an open parameter neighborhood U of ϑ_0 and $c_{\mathrm{rob}} > 0$ such that*

$$\inf_{\vartheta \in U} \inf_{\lambda \in \mathbb{R}} G_\vartheta(\lambda) \ge c_{\mathrm{rob}} > 0. \tag{136}$$

The neighborhood is relative to positive mixture weights summing to one and positive-definite covariances. Perturbations may change weights, means, and covariance matrices simultaneously. Nearby mixtures need not be centered, symmetric, or isotropic.

Proof. Take a compact parameter neighborhood $\mathcal{P}$ of ϑ_0 satisfying equation 128. The tail lemma supplies common L and $c_{\mathrm{tail}} > 0$ for all $\vartheta \in \mathcal{P}$ and $|\lambda| \ge L$. On the compact multiplier interval $[-L, L]$, the reference excess is continuous and strictly positive. Hence

$$\gamma_0 = \min_{|\lambda| \le L} G_{\vartheta_0}(\lambda) > 0.$$

The joint continuity proved in Lemma 8 is uniform on $\mathcal{P} \times [-L, L]$. Shrink to a neighborhood U of ϑ_0 so that

$$\sup_{|\lambda| \le L} \left|G_\vartheta(\lambda) - G_{\vartheta_0}(\lambda)\right| < \gamma_0/2 \qquad (\vartheta \in U).$$

For these parameters, the excess MSE on the compact interval is at least $\gamma_0/2$, and its complement has excess MSE at least c_{tail}. Take $c_{\mathrm{rob}} = \min\{\gamma_0/2, c_{\mathrm{tail}}\} > 0$. □

B.7.1 One-dimensional location mixtures

Corollary 26 (Positive excess MSE for a scalar symmetric two-component GMM)**.** *For* $p_0(x) = \frac{1}{2}\mathcal{N}(x;\mu_0,\sigma^2) + \frac{1}{2}\mathcal{N}(x;-\mu_0,\sigma^2)$*,* $\mu_0 \neq 0$*,* $\sigma^2,\tau > 0$*,* $\inf_{\lambda\in\mathbb{R}} G_p(\lambda) > 0$*. This uniform lower bound is locally stable under finite-GMM parameter perturbations as in Theorem 25.*

Proof. Set $\alpha = \sigma^2/(\sigma^2+\tau) \in (0,1)$. For each finite λ, the one-dimensional score in equation 38 is a linear Gaussian score plus a bounded term. Variation of constants in the ODE yields $T_\lambda(y) = \alpha^{\lambda/2}y + O_\lambda(1)$ as $y \to +\infty$. $m_\tau(y) = \alpha y + O(1)$. Equality forces $\lambda = 2$, which is excluded by the origin-derivative calculation equation 43. Continuity and a positive observation density imply positive excess at every finite multiplier. Lemma 24 gives a positive lower bound at large multiplier magnitudes, and continuity gives a positive minimum on the remaining compact interval. The same parameter-neighborhood argument completes the proof. □

B.8 Posterior-mean and MSE identities

Let P_0 be the law of X, and let $\varphi_{\tau I_d}$ denote the centered Gaussian density with covariance τI_d. Differentiating the Gaussian convolution gives

$$\nabla \bar{p}_\tau(y) = \int -\frac{y-x}{\tau}\varphi_{\tau I_d}(y-x)\,P_0(dx).$$

Dividing by $\bar{p}_\tau(y) > 0$ proves $\tau\nabla \log \bar{p}_\tau(y) = \mathbb{E}[X \mid Y_\tau = y] - y$. For any square-integrable estimator $H(Y_\tau)$,

$$\mathbb{E}\langle X - m_\tau(Y_\tau), H(Y_\tau) - m_\tau(Y_\tau)\rangle = 0$$

by conditioning. Expanding the square gives equation 22. The same proof applies to the target distribution q_0 in the robustness theorem.

B.9 Finite-step DDIM and DDPM

B.9.1 Fixed-grid DDIM for a symmetric mixture

By symmetry, it suffices to take $\mu > 0$ for the source in Theorem 7. At noise level r, put $v_r = \sigma^2_{\mathrm{comp}} + r^2$. Gaussian conditioning gives

$$m_r(y) := \mathbb{E}[X_0 \mid Y_r = y] = \frac{\sigma^2_{\mathrm{comp}}}{v_r}y + \frac{r^2}{v_r}\mu\tanh\left(\frac{\mu y}{v_r}\right), \tag{137}$$

and

$$\widehat{\varepsilon}_r(y) = \frac{r}{v_r}\left[y - \mu\tanh\left(\frac{\mu y}{v_r}\right)\right], \qquad m_r(y) = y - r\widehat{\varepsilon}_r(y). \tag{138}$$

For a fixed grid, write

$$v_i = \sigma^2_{\mathrm{comp}} + r_i^2, \qquad c_i = \frac{(r_i - r_{i-1})r_i}{v_i} > 0, \qquad b_i(\lambda) = 1 - \lambda c_i. \tag{139}$$

Then one step of Equation (11) is

$$\Phi_i^{(\lambda)}(y) = b_i(\lambda)y + \big(1 - b_i(\lambda)\big)\mu\tanh\left(\frac{\mu y}{v_i}\right), \qquad T^{\mathrm{DDIM}}_{\lambda,N} = \Phi_1^{(\lambda)} \circ \cdots \circ \Phi_N^{(\lambda)}. \tag{140}$$

Proof of Theorem 7. For $N = 1$, Equation (138) gives $T^{\mathrm{DDIM}}_{1,1} = m_R$.

For $N \geq 2$, $T^{\mathrm{DDIM}}_{\lambda,N} \neq m_R$ for every real λ. For $\lambda = 0$, the endpoint is the identity, whereas m_R is not. Fix $\lambda \neq 0$ and define

$$p_i = \frac{\pi v_i}{2\mu}, \qquad f_i(u) = b_i(\lambda)u + \lambda c_i\mu\tan\left(\frac{\mu u}{v_i}\right). \tag{141}$$

Since $\tanh(iu) = i\tan(u)$, $\Phi_i^{(\lambda)}(iu) = if_i(u)$. The values $0 < p_1 < \cdots < p_N$ are the first positive poles of the tangent terms.

Set $H_N = f_N$ and $s_N = p_N$. H_N is real analytic on $[0, s_N)$, satisfies $H_N(0) = 0$, and diverges in magnitude as $s \uparrow s_N$. Inductively, suppose H_{i+1} is analytic on $[0, s_{i+1})$, vanishes at zero, and diverges at s_{i+1}. Let s_i be the first positive point at which

$$|H_{i+1}(s_i)| = p_i. \tag{142}$$

Continuity and divergence imply $0 < s_i < s_{i+1}$. The composition $H_i = f_i \circ H_{i+1}$ has a pole at s_i: its tangent term has a pole with nonzero coefficient $\lambda c_i \mu$, while its linear term is finite there. Hence

$$T^{\mathrm{DDIM}}_{\lambda,N}(is) = iH_1(s) \tag{143}$$

has a pole at is_1 with $0 < s_1 < p_N$. By Equation (137), the nearest imaginary-axis poles of m_R are at $\pm ip_N$, so m_R is analytic at is_1. Equality on the real axis would force equality of their analytic continuations, a contradiction.

To make the error lower bound uniform over λ, define

$$h_N(y) = y - \mu \tanh\left(\frac{\mu y}{v_N}\right). \tag{144}$$

Because the real hyperbolic tangent is bounded, induction through Equation (140) gives, in $L^2(Y_R)$,

$$T^{\mathrm{DDIM}}_{\lambda,N}(y) = (-1)^N \left(\prod_{i=1}^{N} c_i\right) \lambda^N h_N(y) + O\big(|\lambda|^{N-1}(1 + |y|)\big). \tag{145}$$

Therefore

$$\frac{\mathbb{E}|T^{\mathrm{DDIM}}_{\lambda,N}(Y_R) - m_R(Y_R)|^2}{|\lambda|^{2N}} \longrightarrow \left(\prod_{i=1}^{N} c_i\right)^2 \mathbb{E}|h_N(Y_R)|^2 > 0. \tag{146}$$

This expectation is positive because h_N is nonzero in the tails. Thus the posterior-mean approximation error diverges as $|\lambda| \to \infty$. It is continuous in λ by dominated convergence on compact multiplier intervals. Since the observation density is positive and the two real-analytic maps never agree identically, the posterior-mean approximation error is positive for every λ. Continuity and the large-$|\lambda|$ bound prove Equation (15).

□

B.9.2 DDIM CONVERGENCE AND THE FINE-GRID LOWER BOUND

Proof of Proposition 4. Fix $M < \infty$ and write $h_i = r_i - r_{i-1}$. Linear growth and the continuous and discrete Grönwall inequalities bound every exact and Euler state by $C_M(1 + \|y\|)$, uniformly over $|\lambda| \le M$ and all grids. On a compact set $A \subset \mathbb{R}^d$, the trajectories remain in a common compact set, where the vector field and its first derivatives are bounded. The Euler error estimate then gives

$$\sup_{|\lambda| \le M} \sup_{y \in A} \|T^{\mathrm{DDIM}}_{\lambda,N}(y) - T^{\mathrm{cont}}_{\lambda}(y)\| \le C_{M,A} R h. \tag{147}$$

For a closed ball A, the growth bound and Equation (147) yield

$$\begin{aligned} &\sup_{|\lambda| \le M} \mathbb{E}\|T^{\mathrm{DDIM}}_{\lambda,N}(Y_R) - T^{\mathrm{cont}}_{\lambda}(Y_R)\|^2 \\ &\qquad \le C^2_{M,A} R^2 h^2 + 4C^2_M \mathbb{E}\big[(1 + \|Y_R\|)^2 \mathbf{1}_{\{Y_R \notin A\}}\big]. \end{aligned} \tag{148}$$

The tail term vanishes as A expands since Y_R has finite second moment. Taking $h \downarrow 0$ and then expanding A proves Equation (12) uniformly over grids. □

Proof of Theorem 5. For $M < \infty$, let $\epsilon_M(h)$ be the left-hand side of Equation (12), with the supremum also taken over grids of maximum spacing at most h. Then $\epsilon_M(h) \to 0$. The reverse triangle inequality gives

$$\inf_{|\lambda| \le M} \|T^{\mathrm{DDIM}}_{\lambda,N}(Y_R) - m_R(Y_R)\|_{L^2} \ge \sqrt{g_*} - \epsilon_M(h). \tag{149}$$

For each integer $m \ge 1$, choose $\bar{h}_m > 0$ so that $\epsilon_m(h) \le \sqrt{g_*}/2$ for $h \le \bar{h}_m$. Set $h_m = \min\{1/m, \bar{h}_1, \dots, \bar{h}_m\}$ and $L(h) = \max\{m \in \mathbb{N} : h \le h_m\}$. Then $L(h)$ is finite, is nonincreasing in h, and tends to infinity as $h \downarrow 0$. Taking $M = L(h)$ in Equation (149) proves Equation (14) with $h_* = h_1$ and $c = g_*/4$. □

Proof of Corollary 6. For a finite GMM,

$$\widehat{\varepsilon}_r(y) = r \sum_{k=1}^{K} w_k(y, r^2)(\Sigma_k + r^2 I_d)^{-1}(y - \mu_k), \tag{150}$$

where the weights are defined in Equation (83) with $t = r^2$. This expression extends continuously with $\widehat{\varepsilon}_0(y) = 0$, is continuously differentiable, and satisfies the growth condition in Proposition 4. The observation and posterior mean have finite second moments. Theorem 1 and Equation (173) give Equation (13). Applying Theorem 5 proves the corollary. □

B.9.3 Gaussian approximation of DDPM reverse conditionals

The DDPM forward process. DDPM uses a discrete VP process (Ho et al., 2020). For a terminal index $J \geq 1$ and a schedule $\beta_j \in (0, 1)$, set $\alpha_j = 1 - \beta_j$ and $\overline{\alpha}_j = \prod_{\ell=1}^{j} \alpha_\ell$ for $j = 1, \ldots, J$. Write x_j for a realization of X_j and $q(\cdot \mid \cdot)$ for a conditional density. The forward chain satisfies

$$q(x_j \mid x_{j-1}) = \mathcal{N}\big(x_j; \sqrt{\alpha_j} x_{j-1}, \beta_j I_d\big)\,, \qquad q(x_j \mid x_0) = \mathcal{N}\big(x_j; \sqrt{\overline{\alpha}_j} x_0, (1 - \overline{\alpha}_j) I_d\big)\,. \tag{151}$$

Gaussian approximation. DDPMs approximate reverse conditional distributions with Gaussian transitions (Ho et al., 2020), whose variances can be learned (Nichol & Dhariwal, 2021). Analytic-DPM and Extended Analytic-DPM derive optimal reverse variances and covariances (Bao et al., 2022b;a), while Gaussian Mixture Solvers use mixture transitions (Guo et al., 2023).

For GMM sources, we minimize the forward Kullback–Leibler (KL) divergence over the mean and covariance of Gaussian reverse transitions.

Consider the discrete VP chain in Equation (151). Let $q_j^{\mathrm{rev}}(x \mid y)$ denote the conditional density of X_{j-1} given $X_j = y$. Write $D_{\mathrm{KL}}(P \| Q)$ for the KL divergence from P to Q. At each state y, define the minimum over Gaussian means μ and covariances C by

$$\mathcal{K}_j(y) := \inf_{\mu \in \mathbb{R}^d,\, C \succ 0} D_{\mathrm{KL}}\big(q_j^{\mathrm{rev}}(\cdot \mid y) \,\|\, \mathcal{N}(\mu, C)\big)\,. \tag{152}$$

Assume that each reverse conditional has a density, finite second moments, a positive-definite covariance, and finite differential entropy.

Let P_J denote the law of X_J and write $\overline{\mathcal{K}}_j := \mathbb{E}[\mathcal{K}_j(X_j)]$. Let $\mathcal{Q}$ be the reverse Markov path measure with terminal marginal P_J and transitions q_j^{rev}. Let $\mathcal{P}$ range over reverse Markov path measures with the same terminal marginal and measurable Gaussian transitions $p_j(x_{j-1} \mid x_j)$.

Assume also that $\overline{\mathcal{K}}_j < \infty$ for all j.

Theorem 27 (Gaussian approximation of GMM reverse conditionals)**.** *For the fixed schedule in Equation* (151) *and every nontrivial finite GMM in Equation* (7)*, $q_j^{\mathrm{rev}}(\cdot \mid y)$ is a nontrivial finite GMM at every step j and state $y \in \mathbb{R}^d$. Therefore,*

$$\mathcal{K}_j(y) > 0, \qquad \overline{\mathcal{K}}_j > 0, \qquad \inf_{\mathcal{P}} D_{\mathrm{KL}}(\mathcal{Q} \,\|\, \mathcal{P}) = \sum_{j=1}^{J} \overline{\mathcal{K}}_j > 0. \tag{153}$$

For a single Gaussian source, all reverse conditionals are Gaussian, and both the reverse-step and path-space minima are zero.

Gaussian projection and path decomposition. Fix j and y, write $Q_y = q_j^{\mathrm{rev}}(\cdot \mid y)$, and let m_y, C_y be its mean and covariance. Taking the expectation of the Gaussian log-density under this distribution gives

$$D_{\mathrm{KL}}(Q_y \,\|\, \mathcal{N}(\mu, C)) = D_{\mathrm{KL}}(Q_y \,\|\, \mathcal{N}(m_y, C_y)) + D_{\mathrm{KL}}(\mathcal{N}(m_y, C_y) \,\|\, \mathcal{N}(\mu, C)). \tag{154}$$

The second term proves the moment projection and shows that $\mathcal{K}_j(y) = 0$ exactly for a Gaussian conditional. For path measures with terminal law P_J, the KL chain rule gives

$$D_{\mathrm{KL}}(\mathcal{Q} \,\|\, \mathcal{P}) = \sum_{j=1}^{J} \mathbb{E}_{X_j} D_{\mathrm{KL}}\big(q_j^{\mathrm{rev}}(\cdot \mid X_j) \,\|\, p_j(\cdot \mid X_j)\big)\,. \tag{155}$$

Conditional moments are measurable in y. Their moment-matched transitions are therefore admissible and minimize every summand, proving the path identity in Equation (153).

Lemma 28 (Reverse conditionals of finite GMMs)**.** *For every nontrivial finite GMM in Equation* (7)*, every step j, and every $y \in \mathbb{R}^d$, the reverse conditional has the finite-GMM representation in Equation* (160) *with positive weights and at least two distinct Gaussian components.*

Proof. Introduce a component label $L \in \{1, \dots, K\}$ for Equation (7). Under the forward chain, set $\overline{\alpha}_0 := 1$. For $j = 0, \dots, J$,

$$X_j \mid L = k \sim \mathcal{N}(m_{k,j}, C_{k,j}), \qquad m_{k,j} = \sqrt{\overline{\alpha}_j}\mu_k, \qquad C_{k,j} = \overline{\alpha}_j \Sigma_k + (1 - \overline{\alpha}_j) I_d. \tag{156}$$

The posterior component weights are

$$w_{k,j}(y) = \frac{\pi_k \mathcal{N}(y; m_{k,j}, C_{k,j})}{\sum_\ell \pi_\ell \mathcal{N}(y; m_{\ell,j}, C_{\ell,j})}. \tag{157}$$

Gaussian conditioning within a component gives

$$\nu_{k,j}(y) = m_{k,j-1} + \sqrt{\alpha_j} C_{k,j-1} C_{k,j}^{-1}(y - m_{k,j}), \tag{158}$$

$$V_{k,j} = C_{k,j-1} - \alpha_j C_{k,j-1} C_{k,j}^{-1} C_{k,j-1}, \qquad V_{k,j}^{-1} = C_{k,j-1}^{-1} + \frac{\alpha_j}{\beta_j} I_d. \tag{159}$$

Hence the reverse conditional density is

$$q_j^{\mathrm{rev}}(x \mid y) = \sum_{k=1}^{K} w_{k,j}(y) \mathcal{N}(x; \nu_{k,j}(y), V_{k,j}). \tag{160}$$

The map $C \mapsto (C^{-1} + \alpha_j \beta_j^{-1} I_d)^{-1}$ is injective, so unequal source covariances give unequal reverse covariances. If two source components have the same covariance, their distinctness implies unequal means, and

$$\nu_{k,j}(y) - \nu_{\ell,j}(y) = \beta_j C_j^{-1} \sqrt{\overline{\alpha}_{j-1}} (\mu_k - \mu_\ell) \neq 0, \tag{161}$$

where C_j is their common time-j covariance. All weights in Equation (157) are positive. Thus Equation (160) has at least two distinct Gaussian components for every y. □

Proof of Theorem 27. By Lemma 28, every reverse conditional is a positive finite mixture with at least two distinct Gaussian components. A positive finite mixture of distinct Gaussian distributions is not a single Gaussian. Project onto a generic direction that separates the component mean–variance pairs. Equality with a scalar Gaussian would equate their moment-generating functions. Sending the argument to $+\infty$ first selects the largest variance and then the largest mean; equality would force the total weight of that single parameter pair to be one, contradicting the remaining positive mixture weights.

Define $m_{j-1|j}(y) := \mathbb{E}[X_{j-1} \mid X_j = y]$ and $C_{j-1|j}(y) := \mathrm{Cov}(X_{j-1} \mid X_j = y)$. Consequently,

$$\mathcal{K}_j(y) = H_{\mathrm{diff}}\big(\mathcal{N}(m_{j-1|j}(y), C_{j-1|j}(y))\big) - H_{\mathrm{diff}}\big(q_j^{\mathrm{rev}}(\cdot \mid y)\big) > 0. \tag{162}$$

$H_{\mathrm{diff}}(\cdot)$ denotes differential entropy. The finitely many within-component covariances bound $C_{j-1|j}(y)$ below, while the affine component means bound it above by $C(1 + \|y\|^2) I_d$. Entropy concavity bounds the mixture entropy below by a constant. Hence

$$0 < \mathcal{K}_j(y) \leq C + C \log(1 + \|y\|^2). \tag{163}$$

The marginal law of X_j is a finite GMM, so the right-hand side is integrable and $0 < \overline{\mathcal{K}}_j < \infty$. The moment projection and KL chain rule prove Equation (153). For a single Gaussian source, joint Gaussianity makes each reverse conditional Gaussian, so the minimum reverse-step and path-space KL divergences are zero. □

C EQUIVALENT NOISE PARAMETERIZATIONS

The mappings below give equivalent endpoints for the VP flow, additive Gaussian noise, and linear-interpolation parameterizations. The interpolation parameter u is distinct from the normalized noise level r.

C.1 LINEAR INTERPOLATION AND ADDITIVE GAUSSIAN NOISE

Let $X \in \mathbb{R}^d$ be a clean signal and $\varepsilon \sim \mathcal{N}(0, I_d)$ be independent Gaussian noise. The linear interpolant is

$$Z_u = (1-u)\varepsilon + uX, \qquad v_u(z) = \mathbb{E}[X - \varepsilon \mid Z_u = z], \qquad 0 < u \le 1. \tag{164}$$

Let p_u^{lin} be the density of Z_u, and write $s_u^{\mathrm{lin}}(z) = \nabla_z \log p_u^{\mathrm{lin}}(z)$. Gaussian conditioning gives

$$v_u(z) = \frac{z}{u} + \frac{1-u}{u} s_u^{\mathrm{lin}}(z). \tag{165}$$

For $u_0 \in (0,1)$ and $\rho \in \mathbb{R}$, the controlled ODE with state $h_u^{(\rho)}$ is

$$\frac{dh_u^{(\rho)}}{du} = (2-\rho)v_u(h_u^{(\rho)}) - (1-\rho)\frac{h_u^{(\rho)}}{u} = \frac{h_u^{(\rho)}}{u} + (2-\rho)\frac{1-u}{u} s_u^{\mathrm{lin}}(h_u^{(\rho)}), \qquad h_{u_0}^{(\rho)} = z. \tag{166}$$

Set

$$\begin{aligned} r &= \frac{1-u}{u}, \qquad t = r^2, \qquad u = \frac{1}{1+r}, \\ y_t &= \frac{h_u^{(\rho)}}{u}, \qquad \lambda = 2-\rho, \qquad R = \frac{1-u_0}{u_0}, \quad \tau = R^2. \end{aligned} \tag{167}$$

Write $\bar{p}_t$ and $\bar{s}_t = \nabla \log \bar{p}_t$ for the density and score of $X + \sqrt{t}\,\varepsilon$ indexed by noise variance t. Since $Z_u/u = X + r\varepsilon$, the densities and scores satisfy

$$p_u^{\mathrm{lin}}(z) = u^{-d}\bar{p}_t(z/u), \qquad s_u^{\mathrm{lin}}(z) = u^{-1}\bar{s}_t(z/u), \qquad \frac{dt}{du} = -\frac{2(1-u)}{u^3}. \tag{168}$$

Differentiating $y_t = h_u^{(\rho)}/u$ in equation 166 gives

$$\frac{dy_t}{du} = \lambda \frac{1-u}{u^3}\bar{s}_t(y_t), \qquad \frac{dy_t}{dt} = -\frac{\lambda}{2}\bar{s}_t(y_t). \tag{169}$$

Reindex the trajectory by $r = \sqrt{t}$ as $y_r^{(\lambda)} := y_{r^2}$. Then $dy_r^{(\lambda)}/dr = -\lambda r\, \bar{s}_{r^2}(y_r^{(\lambda)})$. Denote its endpoint by T_λ^{add}, its posterior mean by m^{add}, and its mean-squared endpoint error by G^{add}. At $u = 1$, the spatial scaling is one, so the two reconstructions are identical. Denote the interpolation endpoint by $T_\rho^{\mathrm{lin}}(z)$ and the posterior mean of X conditioned on $Z_{u_0} = z$ by $m^{\mathrm{lin}}(z)$. Then

$$T_\rho^{\mathrm{lin}}(z) = T_{2-\rho}^{\mathrm{add}}(z/u_0), \qquad m^{\mathrm{lin}}(z) = m^{\mathrm{add}}(z/u_0), \tag{170}$$

and hence

$$\mathbb{E}\|T_\rho^{\mathrm{lin}}(Z_{u_0}) - m^{\mathrm{lin}}(Z_{u_0})\|^2 = G^{\mathrm{add}}(2-\rho). \tag{171}$$

The maps $u_0 \mapsto R$ and $\rho \mapsto \lambda$ are bijections on their domains. They therefore preserve the endpoint and its mean-squared error.

A negative ρ corresponds to a positive score multiplier; for example, $\rho = -0.15$ gives $\lambda = 2.15$.

C.2 VARIANCE-PRESERVING DIFFUSION

Let $\overline{\alpha}(s) > 0$ satisfy $\overline{\alpha}'(s) = -\beta(s)\overline{\alpha}(s)$ and consider the Gaussian-corruption marginal $\sqrt{\overline{\alpha}(s)}X + \sqrt{1-\overline{\alpha}(s)}\,\varepsilon$. Let q_s^{VP} be its density. On an interval where this reparameterization is invertible, set $t = (1-\overline{\alpha}(s))/\overline{\alpha}(s)$ and $y_t = h_s/\sqrt{\overline{\alpha}(s)}$. The score-scaled variance-preserving ODE is

$$\frac{dh_s}{ds} = -\frac{\beta(s)}{2}h_s - \frac{\lambda\beta(s)}{2}\nabla_h \log q_s^{\mathrm{VP}}(h_s). \tag{172}$$

Table 7: Equivalent linear-interpolation and additive-noise parameterizations.

Object	Linear interpolant	Additive noise
Observation level	signal fraction $u_0 \in (0,1)$	noise-to-signal level $R = (1-u_0)/u_0$
Observation	z	$y = z/u_0$
Control	ρ	$\lambda = 2-\rho$
Probability flow	$\rho = 1$	$\lambda = 1$
Gaussian posterior mean	$\rho = 0$	$\lambda = 2$

Using $\nabla_h \log q_s^{\mathrm{VP}}(h) = \bar{s}_t(h/\sqrt{\overline{\alpha}(s)})/\sqrt{\overline{\alpha}(s)}$ and $dt/ds = \beta(s)/\overline{\alpha}(s)$ gives $dy_t/dt = -(\lambda/2)\bar{s}_t(y_t)$. Taking $r = \sqrt{t}$ recovers the normalized additive flow above.

Fix $s_{\mathrm{obs}} > 0$ and write $a_{\mathrm{obs}} = \overline{\alpha}(s_{\mathrm{obs}}) \in (0,1)$. If $T^{\mathrm{VP}}_{\lambda,s_{\mathrm{obs}}}$ denotes the endpoint of Equation (172) initialized at h and $m^{\mathrm{VP}}_{s_{\mathrm{obs}}}(h) = \mathbb{E}[X \mid \sqrt{a_{\mathrm{obs}}}X + \sqrt{1-a_{\mathrm{obs}}}\,\varepsilon = h]$, then

$$\begin{aligned} T^{\mathrm{VP}}_{\lambda,s_{\mathrm{obs}}}(h) &= T^{\mathrm{add}}_{\lambda}(h/\sqrt{a_{\mathrm{obs}}}), \\ m^{\mathrm{VP}}_{s_{\mathrm{obs}}}(h) &= m^{\mathrm{add}}(h/\sqrt{a_{\mathrm{obs}}}). \end{aligned} \tag{173}$$

The clean endpoint is unchanged because $\overline{\alpha}(0) = 1$. Under the corresponding random observations, the two endpoint errors are identical, so their mean-squared endpoint errors agree. In the main text, $s_{\mathrm{obs}} = t_{\mathrm{obs}}$ and $T_\lambda = T^{\mathrm{VP}}_{\lambda,t_{\mathrm{obs}}}$.

D EXPERIMENTAL SETTINGS

D.1 IMAGE LATENT DISTRIBUTIONS

Datasets and encoders. We used all $50{,}000$ images from the ImageNet-1K ILSVRC 2012 validation split (Deng et al., 2009) and all $5{,}000$ images from the MS COCO 2017 validation split (`val2017`) (Lin et al., 2014). We partitioned ImageNet-1K into $40{,}000$ fitting, $5{,}000$ selection, and $5{,}000$ evaluation images, and COCO into $4{,}000$, 500, and 500 images, respectively. Images were encoded with frozen Stable Diffusion 2.1 (SD2.1) and Stable Diffusion XL (SDXL) VAEs (Rombach et al., 2022; Podell et al., 2024).

Latent patches and noise. Each image contributed 64 nonoverlapping $4 \times 8 \times 8$ latent patches, corresponding to 64×64 image regions. Each patch had dimension $d = 256$. Isotropic noise with standard deviation $\sigma_{\mathrm{lat}} \in \{0, 1/3, 1, 3\}$ was added before the whitening transform estimated from the fitting partition. The transform was fixed for the selection and evaluation partitions.

Density fitting and component selection. We fitted full-covariance Gaussians and GMMs on the fitting partition. We selected the component count from $K \in \{2, 4, 8, 12, 16, 24, 32, 48, 64\}$ using development NLL on the selection partition. All four dataset–encoder pairs selected $K = 32$. Table 1 reports held-out NLL on the evaluation partition, averaged by image in nats per latent dimension.

D.2 POSTERIOR-MEAN ERROR UNDER CONSTANT SCALING

Sources and observation noise. Table 2 compares a scalar Gaussian with two-component GMMs whose means are symmetric. The GMMs use normalized component separation $\Delta = 4$ and component variance ratios $\kappa \in \{1, 16\}$. For scalar component variances σ_k^2, $\kappa = \max_k \sigma_k^2 / \min_k \sigma_k^2$: $\kappa = 1$ gives the equal-covariance condition and $\kappa = 16$ the unequal-covariance condition. We evaluated observation-noise standard deviations $\sigma_{\mathrm{obs}} \in \{1/3, 1, 3\}$.

Multiplier selection and error. For each source and noise level, $\widehat{\lambda}$ is the constant multiplier with the smallest evaluated posterior-mean error $G(\lambda)$ in Equation (5). The table reports $100G(\widehat{\lambda})/\,\mathrm{MMSE}$, with $\widehat{\lambda}$ in parentheses. By Equation (6), a value of $q\%$ means reconstruction MSE of $(1 + q/100)\,\mathrm{MMSE}$. These noise-sweep conditions are separate from the SEC conditions in Section D.3.

D.3 ENDPOINT CORRECTION ON CONTROLLED SOURCES

Source distributions. We constructed each source in Table 8 in $d \in \{1, 8\}$ dimensions, with equal component weights and diagonal component covariances. Each source was standardized to $\mathbb{E}[X_0] = 0$ and $\mathrm{tr}(\mathrm{Cov}(X_0))/d = 1$. This fixes the mean variance per dimension, not the covariance matrix.

Table 8: Source generation settings for the controlled SEC experiment. The variance ratio is the maximum divided by the minimum component variance across all components and coordinates.

Source	Components	Mean-separation setting	Variance ratio
Gaussian	1	0	1
Symmetric GMM	2	1	1
Anisotropic GMM	2	2	8
Heterogeneous GMM	4	2	16

The one-dimensional symmetric GMM is

$$X_0 \sim \tfrac{1}{2}\mathcal{N}(-1/\sqrt{2}, 1/2) + \tfrac{1}{2}\mathcal{N}(1/\sqrt{2}, 1/2). \tag{174}$$

In one dimension, Anisotropic GMM denotes two Gaussian components with different variances.

Observations and flow endpoints. We sampled X_0 from the specified source and added independently generated Gaussian noise to obtain the normalized observation from Section 3.3 at $R = 1$,

$$Y_R = X_0 + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I_d), \quad \varepsilon \perp X_0. \tag{175}$$

The posterior mean $m_R(y) = \mathbb{E}[X_0 \mid Y_R = y]$ and posterior covariance were computed from the known mixture parameters by Gaussian conditioning. For each multiplier, we numerically integrated

$$\frac{dz_r}{dr} = -\lambda r \nabla \log \bar{p}_{r^2}(z_r), \qquad z_R = y, \qquad R = 1, \tag{176}$$

from $r = 1$ to 0. Here $\bar{p}_{r^2}$ is the density of $X_0 + r\varepsilon$, and its score was computed from the GMM. The flow endpoint $T_\lambda^{\mathrm{cont}}(y)$ was used for constant scaling and for the two endpoints in Equation (16).

Coefficient selection and evaluation. For each source and dimension, we generated 256 observations for coefficient selection and a separate set of 1,024 observations for evaluation. One constant $\widehat{\lambda}$ or one pair (h, η) was applied to all evaluation observations and coordinates. For SEC, the reference multiplier was $\lambda_{\mathrm{ref}} = \widehat{\lambda}$, so the probe multiplier was $\widehat{\lambda} + h$. This differs from the native reference used in the image-compression experiment. Table 9 lists the selected GMM coefficients.

For $N = 1024$ evaluation observations $Y_{R,i}$, we measured

$$\begin{aligned} \widehat{\mathrm{MMSE}} &= \frac{1}{N} \sum_{i=1}^{N} \mathrm{tr}(\mathrm{Cov}(X_0 \mid Y_R = Y_{R,i}))\,, \\ \widehat{G}(T) &= \frac{\frac{1}{N} \sum_{i=1}^{N} \|T(Y_{R,i}) - m_R(Y_{R,i})\|^2}{\widehat{\mathrm{MMSE}}}. \end{aligned} \tag{177}$$

Here T is the computed constant or SEC endpoint map. We write $\widehat{G}_{\mathrm{const}} = \widehat{G}(T_{\widehat{\lambda}}^{\mathrm{cont}})$ and $\widehat{G}_{\mathrm{SEC}} = \widehat{G}(Z_{h,\eta})$. The values in Table 3 are ratios to estimated MMSE, not percentages or errors divided by dimension. Given a reported relative reduction ρ in percent, we computed the displayed SEC error as

$$\widehat{G}_{\mathrm{SEC}} = \widehat{G}_{\mathrm{const}} \left(1 - \frac{\rho}{100}\right). \tag{178}$$

The SEC entries inherit the rounding of the reported constant errors and reduction percentages.

Table 9: Coefficients selected on samples separate from the evaluation set for each GMM and dimension. SEC uses the selected constant as its reference multiplier, $\lambda_{\mathrm{ref}} = \widehat{\lambda}$.

Source	d	Constant $\widehat{\lambda}$	SEC h	SEC η
Symmetric GMM	1	2.125	1.5	1.250760
Anisotropic GMM	1	1.5	-1	0.288631
Heterogeneous GMM	1	2.25	0.5	1.131576
Symmetric GMM	8	1.875	-1	0.555736
Anisotropic GMM	8	2.0	1	0.960294
Heterogeneous GMM	8	1.75	1.5	0.948126

D.4 PD CURVES FOR SCALAR MIXTURES

Source configurations. We evaluated all 300 combinations of

$$\begin{aligned} K &\in \{2, 4, 8, 16\}, \qquad \Delta \in \{0, 0.5, 1, 2, 4\}, \\ \kappa &\in \{1, 2, 4, 8, 16\}, \quad \sigma_{\mathrm{obs}} \in \{1/3, 1, 3\}. \end{aligned}$$

Each source had K components with weights $1/K$, equally spaced means μ_k, and logarithmically spaced variances σ_k^2 with $\max_k \sigma_k^2 / \min_k \sigma_k^2 = \kappa$. We centered and rescaled each source to mean zero and variance one. The separation was the component-mean RMS divided by the pooled within-component standard deviation,

$$\Delta = \frac{\sqrt{K^{-1} \sum_{k=1}^{K} \mu_k^2}}{\sqrt{K^{-1} \sum_{k=1}^{K} \sigma_k^2}}.$$

The observation was $Y = X_0 + \sigma_{\text{obs}}\varepsilon$, with $\varepsilon \sim \mathcal{N}(0,1)$ independent of X_0. The 12 configurations with $\Delta = 0$ and $\kappa = 1$ reduce to a single Gaussian and remain in the 300-configuration aggregate. We also evaluated a separate $\mathcal{N}(0,1)$ control at each of the three noise levels; these controls were excluded from the mixture aggregates.

Numerical evaluation. We searched 801 multipliers in $[-2, 6]$ at spacing 0.01. Flow endpoints used the analytic GMM score, float64 arithmetic, and 512 fourth-order Runge–Kutta (RK4) steps. Quantile integrals used 65,536-point Gauss–Legendre quadrature on $u \in (0,1)$, with 64 bisection iterations for CDF inversion. Expectations were computed by quantile integration without Monte Carlo sampling.

PD statistics. Here P is squared Wasserstein-2 distance, not LPIPS. For each configuration, we used the P coordinates of the Pareto points from the multiplier search, without a uniform P grid. At each point, we subtracted $D^\star(P)$ from the score-scaling MSE and divided by MMSE, as in Section 5.3. We integrated this normalized regret between Pareto points using the trapezoidal rule. After the last Pareto point, we held the reconstruction fixed and integrated the resulting regret analytically up to $P = c^2$. The integrated statistic was divided by c^2. We computed the maximum evaluated regret, integrated regret, and regret at $P = c^2$ for each configuration, then took the median, 90th percentile, and largest value of each statistic across the 300 configurations. Table 4 reports these values multiplied by 100.

D.5 Rate-wise image-compression results

Table 10 reports the individual rate results used in Table 5. The main-text means assign equal weight to low, middle, and high rates within each model, method, and dataset.

Coefficient-selection images did not overlap with the Kodak, CLIC-64, or Tecnick-100 evaluation images.

Image preprocessing. Reconstructions were cropped to their original dimensions and clipped to $[0,1]$ before evaluation. ResULIC used edge padding to multiples of 64. The 512×512 crops and Kodak images required no additional padding. Bitrate in bpp used the original image dimensions. All candidates for the same image and rate shared the bitstream and initial random stream.

Image metrics. MSE was the mean squared RGB error over all channels and pixels in $[0,1]$, averaged over images. PSNR was computed as $-10\log_{10}(\text{MSE})$ for each image and then averaged. LPIPS used the AlexNet backbone with inputs mapped from $[0,1]$ to $[-1,1]$ by $2x-1$. SSIM and MS-SSIM used `pytorch_msssim` with `data_range=1`.

Table 10: Image-compression results at three rates. MSE is multiplied by 10^3. Lower is better. Bold and underline mark the best and second-best unrounded values within each model, rate, and metric, including ties.

			Kodak		CLIC-64		Tecnick-100	
Model	Method	Rate	MSE	LPIPS	MSE	LPIPS	MSE	LPIPS
DiffC-SD2.1	Ours	Low	**42.722**	**0.7303**	**57.298**	**0.7851**	**74.476**	**0.7582**
		Middle	**13.250**	0.5995	**14.749**	0.4909	**20.443**	0.4761
		High	**5.095**	0.1668	**4.389**	0.1276	**5.999**	0.1291
	Fixed-h	Low	43.515	0.7461	58.763	0.7932	76.589	0.7630
		Middle	13.400	0.6013	15.039	0.4926	20.894	0.4779
		High	5.106	0.1670	4.426	0.1278	6.046	0.1292
	Native ($\lambda = 1$)	Low	68.340	0.7696	124.534	0.8030	145.794	0.7683
		Middle	16.279	**0.4574**	25.356	0.5033	32.740	**0.4549**
		High	5.125	**0.1667**	4.464	**0.1275**	6.101	**0.1289**
	Constant	Low	68.340	0.7696	124.534	0.8030	145.794	0.7683
		Middle	13.542	0.5971	17.226	**0.4885**	24.232	0.4733
		High	5.125	**0.1667**	4.464	**0.1275**	6.101	**0.1289**
	Effective constant	Low	44.891	0.9299	60.781	0.8730	78.738	0.9193
		Middle	14.757	0.7375	15.453	0.5722	21.523	0.5448
		High	5.125	**0.1667**	4.464	**0.1275**	6.101	**0.1289**
ResULIC	Ours	Low	**14.436**	**0.5248**	**16.152**	**0.4925**	**22.787**	**0.4639**
		Middle	**8.986**	0.3069	**8.446**	0.2744	**11.856**	0.2656
		High	**7.287**	0.2363	**6.896**	0.1993	**9.759**	0.2048
	Fixed-h	Low	15.527	0.5285	17.500	0.4938	24.261	0.4649
		Middle	9.056	0.3078	8.810	0.2753	12.310	0.2666
		High	7.301	0.2364	6.933	0.1995	9.809	0.2050
	Native ($\lambda = 1$)	Low	17.364	0.5357	19.026	0.4958	26.769	0.4662
		Middle	9.123	**0.3055**	9.336	**0.2729**	13.132	**0.2638**
		High	7.325	**0.2361**	6.985	**0.1991**	9.892	**0.2045**
	Constant	Low	17.364	0.5357	19.026	0.4958	26.769	0.4662
		Middle	9.123	**0.3055**	9.336	**0.2729**	13.132	**0.2638**
		High	7.325	**0.2361**	6.985	**0.1991**	9.892	**0.2045**
	Effective constant	Low	17.364	0.5357	19.026	0.4958	26.769	0.4662
		Middle	9.123	**0.3055**	9.336	**0.2729**	13.132	**0.2638**
		High	7.325	**0.2361**	6.985	**0.1991**	9.892	**0.2045**
ELIC+DDIM/DSG	Ours	Low	**6.086**	**0.2194**	**5.664**	**0.1811**	**7.721**	**0.1714**
		Middle	**4.252**	**0.1218**	**3.387**	**0.0939**	**4.497**	**0.0937**
		High	**3.296**	**0.0858**	**2.570**	**0.0633**	**3.378**	**0.0647**
	Fixed-h	Low	6.312	0.2195	5.850	0.1813	7.986	0.1715
		Middle	4.270	0.1224	3.459	0.0941	4.617	0.0938
		High	3.301	0.0859	2.580	0.0633	3.389	0.0647
	Native ($\lambda = 1$)	Low	6.664	0.2196	6.167	0.1817	8.349	0.1717
		Middle	4.297	0.1236	3.585	0.0947	4.744	0.0942
		High	3.308	0.0860	2.594	0.0634	3.414	0.0648
	Constant	Low	6.664	0.2196	6.167	0.1817	8.349	0.1717
		Middle	4.287	0.1232	3.577	0.0943	4.733	0.0938
		High	3.307	0.0861	2.591	0.0634	3.408	0.0647
	Effective constant	Low	6.664	0.2196	6.167	0.1817	8.349	0.1717
		Middle	4.297	0.1236	3.585	0.0947	4.744	0.0942
		High	3.308	0.0860	2.594	0.0634	3.414	0.0648