Mixture-of-Gaussians-Guided Schedule Design for Brownian Bridge Diffusion Models

Anonymous authors
Paper under double-blind review

Abstract

Brownian Bridge Diffusion Models (BBDM) offer an appealing framework for image restoration and inverse problems by constructing a stochastic bridge from the clean signal directly to the degraded observation, rather than to pure noise. Despite their promise, the choice of bridge schedule is typically inherited from heuristics, and a principled analytical framework for schedule design has been lacking. In this work, we develop such a framework by offering a novel analysis of BBDM reverse dynamics under a Mixture-of-Gaussians (MoG) prior. This setting yields a closed-form ideal posterior and a corresponding MMSE denoiser, while the BBDM-induced reconstruction law is captured analytically through a tractable surrogate. Building on these expressions, we formulate two complementary schedule-design objectives: a Wasserstein criterion targeting perceptual quality and an MSE criterion targeting reconstruction fidelity. Our work exposes an inherent tradeoff between the two and proves the existence of universal schedules for both that are independent of the degradation and prior. Extensive experiments on controlled MoG settings confirm full alignment between theory and practice, and experiments on the FFHQ dataset across inpainting, deblurring, and super-resolution tasks validate the practical value of our schedule-design criteria.

1 Introduction

Diffusion models can be used beyond unconditional generation of images, handling image-to-image translation and solving inverse problems. [1] In the standard DDPM/DDIM formulations (Ho et al., 2020; Song et al., 2021), the forward process gradually corrupts a clean signal towards a terminal state that is close to a Gaussian noise, and the reverse process starts from such a noise and iteratively carves out visual content via a series of denoising steps. When applied to linear inverse problems of the form $y = Hx + n$, where H is a known degradation operator and n is measurement noise, conditional diffusion variants can incorporate this observation by providing y to the denoiser either explicitly or implicitly (Saharia et al., 2021; 2022; Lugmayr et al., 2022; Kawar et al., 2022; Wang et al., 2023; Chung et al., 2023; Song et al., 2023).

Brownian Bridge Diffusion Models (BBDM) offer a different conditional construction (Li et al., 2023). Instead of diffusing a clean signal towards pure noise, BBDM constructs a stochastic bridge from the clean signal to the conditioning observation y. Consequently, the reverse chain starts from the observation itself. For inverse problems, this is an appealing viewpoint: the measurement is not only side information supplied to a network or a guidance term added during inference, but the endpoint of the diffusion path.

A natural question for those two mechanisms is how their inference behavior depends on their schedules. In a conditional DDIM model, the schedule controls the usual data-to-noise path. In BBDM, the bridge schedule controls the path between the clean signal and the observation. These choices directly affect the reconstruction process, yet they are often inherited from heuristics. Our goal is to develop a method-agnostic analytical framework that describes the induced reconstruction laws and uses them for principled schedule design. This paper focuses on BBDM, yet the same framework is also applicable to a conditional DDIM sampler, allowing a direct comparison between the two approaches.

[1] The authors used LLMs for assistance with LaTeX formatting, manuscript presentation and limited code generation and debugging. All scientific ideas, claims, derivations, results, code, and final content were reviewed and approved by the authors.

A useful route towards developing such a framework is to start from a tractable model for the clean-signal distribution. Under a single Gaussian prior, exact analysis becomes possible and can be used to derive closed-form reconstruction laws and schedule-design criteria. Two recent papers by Benita et al. (2025; 2026) follow this route. The first analyzes standard DDPM/DDIM prior sampling under a Gaussian signal model and uses the resulting transfer-function view for schedule design. The second studies training-free posterior sampling methods for inverse problems, including DPS and PiGDM-style samplers (Chung et al., 2023; Song et al., 2023), again under a Gaussian prior and in a setting where the degradation operator and the prior covariance are jointly diagonalizable, sharing a Fourier basis. These works demonstrate the value of exact analysis, but their single-Gaussian and shared-basis assumptions restrict the class of signal models and inverse problems that can be analyzed directly.

In this work we extend the above-described analytical program in two ways: Moving from a single Gaussian prior to a Mixture-of-Gaussians (MoG) one, and enabling an analysis of any linear degradation. The MoG model is richer while still permitting exact posterior formulas. We first derive the exact posterior and the corresponding MMSE denoiser, which gives a complete analytical description at the one-step level. The difficulty appears when this exact MMSE denoiser is inserted into the BBDM reverse sampler: the posterior component probabilities depend on the current latent state. We call the resulting nonlinear sampler the oracle MoG BBDM chain. Although its denoiser is exact, this chain no longer has a fixed affine form that can be unrolled into a simple end-to-end reconstruction law.

To recover tractability, we introduce an approximation that leans on a selected-label surrogate. We first draw a discrete label according to the component probabilities implied by the measurement alone, and then keep this label fixed throughout the reverse chain. Once the label is fixed, the reverse dynamics become an affine Gaussian recursion yielding an explicit reconstruction law. We show that this law can be analyzed in a component-wise basis, applicable for any prior and degradation operator. We further show that this surrogate is mean-exact relative to the corresponding target posterior law $p(x_0 \mid y)$, and its only discrepancy is a deficit in posterior covariance.

The above-described derivations enable two complementary schedule-design objectives: a distributional objective, based on the Wasserstein (W2) distance between the surrogate and the target posterior laws, and an image-to-image objective based on Mean-Squared-Error (MSE) between reconstructions using the two approaches. Whereas the first targets perceptual quality, the second values distortion performance.

By restricting the BBDM bridge's schedule to a bounded four-parameter family, we prove that the MSE objective admits a universal MSE-oriented schedule, independently of the degradation operator, the measurement-noise level, or the prior. Furthermore, we expose an inherent tradeoff between the MSE and W2 losses: Choosing the schedule that increases MSE leads to a decreased W2 and vice-versa.

We conclude this work with a broad experimental study that validates the proposed analysis and its practical value. In controlled MoG settings, where the exact posterior is available, we show that the selected-label approximation closely matches oracle BBDM behavior. We then show that the frozen surrogate captures scheduling trends in trained BBDM models. Finally, we demonstrate on FFHQ that the resulting MSE- and W2 -oriented schedules produce the expected distortion–perception tradeoff across various restoration tasks.

In summary, our contributions are the following: (i) We derive the exact posterior and MMSE denoiser for BBDM under a Mixture-of-Gaussians prior [2]. (ii) Identifying that the MoG reverse process in BBDM loses a global affine structure, we introduce an effective approximation of this computational chain, getting a closed-form expression for the final reconstruction and its dependency on the chosen schedule. (iii) We present two complementary design criteria for the BBDM's schedule, trading perceptual quality versus reconstruction distortion. These enable a reliable and direct optimization of the scheduling parameters of BBDM. (iv) We suggest universal problem-agnostic schedules for the two losses, and validate these in various trained BBDM models on the FFHQ dataset.

Sections 2–6 develop the theory, Section 7 discusses related work, Section 8 presents the experiments, and the appendices contain the full derivations and proofs.

[2] An analogous conditional-DDIM formulation is provided in the Appendix H.

2 Background and Problem Setup

This section fixes the notation and recalls the ingredients of BBDM used in the analysis. We first describe the bridge construction, then the reverse sampling rule, and finally the inverse-problem handled in the present work.

2.1 BBDM as a stochastic bridge

Following Li et al. (2023), we index the subsequence of steps used at inference time by $s \in \{0, \dots, S\}$. The Brownian-bridge forward marginal is

$$q_{BB}(x_s \mid x_0, y) = \mathcal{N}\big(x_s; (1 - m_s)x_0 + m_s y, \delta_s I\big), \tag{1}$$

where $m_0 \cong 0$, $m_S \cong 1$, and $\delta_0 = \delta_S \cong 0$. Equation (1) shows that the bridge starts at x_0 and ends at y: as s increases, the mean moves from the clean signal towards the corresponding measurement, while δ_s controls the randomness around that interpolating path. In our analysis we use the following standard BBDM derived quantities

$$\delta_{s|s-1} = \delta_s - \delta_{s-1}\frac{(1 - m_s)^2}{(1 - m_{s-1})^2}, \tag{2}$$

$$\tilde{\delta}_s = \frac{\delta_{s|s-1} \cdot \delta_{s-1}}{\delta_s}, \qquad \sigma_s = \sqrt{\tilde{\delta}_s}. \tag{3}$$

In the original BBDM formulation, a common reference choice is

$$\delta_s^{\text{ref}} = 2(m_s - m_s^2),$$

motivated by a variance-preservation argument that unrealistically assumes the bridge endpoints are independent (Li et al., 2023). Because clean signals and target measurements are inherently dependent in inverse problems, we keep δ_s general to explicitly optimize its shape rather than fixing it in advance.

Throughout this paper we analyze the implementation setting, also used by the released BBDM code, in which the denoiser at step s takes both x_s and y as inputs.

2.2 Reverse sampling rule

Once a denoiser supplies an estimate $\hat{x}_0(x_s, y)$, the BBDM reverse update along this sampling subsequence can be written as

$$x_{s-1} = a_s \hat{x}_0 + b_s y + c_s x_s + \sigma_s z_s, \qquad z_s \sim \mathcal{N}(0, I), \tag{4}$$

with

$$a_s = (1 - m_{s-1}) - (1 - m_s)\sqrt{\frac{\delta_{s-1} - \sigma_s^2}{\delta_s}}, \tag{5}$$

$$b_s = m_{s-1} - m_s\sqrt{\frac{\delta_{s-1} - \sigma_s^2}{\delta_s}}, \qquad c_s = \sqrt{\frac{\delta_{s-1} - \sigma_s^2}{\delta_s}}. \tag{6}$$

At the first reverse step $s = S$ the coefficient formulas above are interpreted in the limiting endpoint sense:

$$x_{S-1} = (1 - m_{S-1})\hat{x}_0 + m_{S-1} y + \sqrt{\delta_{S-1}} z_S.$$

2.3 Linear measurement model and MOG prior

The present paper studies BBDM in a linear inverse-problem setting. This is an additional modeling layer on top of the original bridge construction. We assume that the clean signal $x_0 \in \mathbb{R}^d$ is observed through the linear model

$$y \mid x_0 \sim \mathcal{N}(Hx_0, \sigma_y^2 I), \tag{7}$$

where $H \in \mathbb{R}^{d\times d}$ is a known degradation operator and $\sigma_y^2 > 0$ is the measurement-noise variance. We further model the data distribution of the clean signal x_0 by a Mixture-of-Gaussians,

$$\begin{aligned} p(x_0) &= \sum_{r=1}^{R} \pi_r \, \mathcal{N}(x_0; \mu_r, \Sigma_r), \\ \Sigma_r &\succ 0, \qquad r = 1, \dots, R, \\ \pi_r &> 0, \qquad \sum_{r=1}^{R} \pi_r = 1. \end{aligned} \tag{8}$$

To separate the continuous and discrete uncertainties, we introduce a latent component variable

$$\mathbb{P}(C = r) = \pi_r, \qquad x_0 \mid (C = r) \sim \mathcal{N}(\mu_r, \Sigma_r). \tag{9}$$

Conditioned on $C = r$, the model remains single Gaussian.

3 Bayes-Optimal Denoiser under MoG Prior

This section derives the posterior distribution $p(x_0 \mid x_s, y)$ under the MoG prior. This posterior is the basic quantity required for the rest of the paper, because BBDM is trained to predict the bridge residual $x_s - x_0$. Under squared loss, the Bayes-optimal residual predictor is

$$\hat{r}_s^{\mathrm{MMSE}}(x_s, y) = \mathbb{E}[x_s - x_0 \mid x_s, y] = x_s - \mathbb{E}[x_0 \mid x_s, y],$$

and therefore the induced Bayes-optimal estimate of the clean signal is

$$\hat{x}_{0,s}^{\mathrm{MMSE}}(x_s, y) \triangleq \mathbb{E}[x_0 \mid x_s, y]. \tag{10}$$

To compute $p(x_0 \mid x_s, y)$, we decompose it over the latent component label:

$$p(x_0 \mid x_s, y) = \sum_{r=1}^{R} p(x_0 \mid x_s, y, C = r) \, \mathbb{P}(C = r \mid x_s, y). \tag{11}$$

Equation (11) splits the derivation into two parts. We first derive the component-conditioned posterior $p(x_0 \mid x_s, y, C = r)$, namely the posterior obtained when the active mixture component is fixed. We then derive the posterior probability $\mathbb{P}(C = r \mid x_s, y)$ of that component and combine the two ingredients. Detailed derivations are collected in Appendix B.

3.1 Component-conditioned posterior

We begin with the posterior obtained when the active component is fixed. Once $C = r$ and the measurement y are known, the posterior $p(x_0 \mid y, C = r)$ is Gaussian. When additionally observing the bridge state x_s, the posterior $p(x_0 \mid x_s, y, C = r)$ remains Gaussian as well.

Lemma 3.1. *Fix a component index r. Conditioned on $C = r$, the posterior after observing the measurement is*

$$p(x_0 \mid y, C = r) = \mathcal{N}\big(x_0; \mu_{r|y}, \Sigma_{r|y}\big),$$

with

$$\Sigma_{r|y}^{-1} = \Sigma_r^{-1} + \frac{1}{\sigma_y^2} H^\top H, \quad \mu_{r|y} = \Sigma_{r|y} \left(\Sigma_r^{-1} \mu_r + \frac{1}{\sigma_y^2} H^\top y \right). \tag{12}$$

Conditioned on the same component, for an interior bridge step $1 \le s \le S - 1$, the posterior after observing both the measurement and the bridge state is

$$p(x_0 \mid x_s, y, C = r) = \mathcal{N}\big(x_0; \mu_{r|s}(x_s, y), \Sigma_{r|s}\big),$$

where

$$\Sigma_{r|s} = \left(\Sigma_r^{-1} + \frac{1}{\sigma_y^2} H^\top H + \frac{(1-m_s)^2}{\delta_s} I\right)^{-1}, \tag{13}$$

$$\mu_{r|s}(x_s, y) = \Sigma_{r|s}\left(\Sigma_r^{-1}\mu_r + \frac{1}{\sigma_y^2} H^\top y + \frac{1-m_s}{\delta_s}(x_s - m_s y)\right). \tag{14}$$

Lemma 3.1 shows that, if the active mixture component were known, then the relevant posterior distributions remain Gaussian, with corresponding posterior means that are affine in (x_s, y).

3.2 Posterior component probabilities and the exact posterior

We now turn to the second ingredient in (11): the posterior probability of each component. We first compute the measurement-only probabilities $\mathbb{P}(C = r \mid y)$, and then refine them after observing the bridge state x_s.

Lemma 3.2. *The posterior probability of component r after observing only y is*

$$\gamma_{r|y}(y) \triangleq \mathbb{P}(C = r \mid y) = \frac{\pi_r \mathcal{N}\big(y; H\mu_r, H\Sigma_r H^\top + \sigma_y^2 I\big)}{\sum_{j=1}^R \pi_j \mathcal{N}\big(y; H\mu_j, H\Sigma_j H^\top + \sigma_y^2 I\big)}. \tag{15}$$

Moreover, conditioned on $(y, C = r)$,

$$p(x_s \mid y, C = r) = \mathcal{N}\big(x_s;\, (1-m_s)\mu_{r|y} + m_s y, (1-m_s)^2\Sigma_{r|y} + \delta_s I\big). \tag{16}$$

therefore, after observing both x_s, y, we get by Bayes' rule and the law of total probability,

$$\gamma_{r|s}(x_s, y) \triangleq \mathbb{P}(C = r \mid x_s, y) = \frac{\gamma_{r|y}(y)\, p(x_s \mid y, C = r)}{\sum_{j=1}^R \gamma_{j|y}(y)\, p(x_s \mid y, C = j)}. \tag{17}$$

Combining (11) with Lemmas 3.1 and 3.2 yields the exact posterior

$$p(x_0 \mid x_s, y) = \sum_{r=1}^R \gamma_{r|s}(x_s, y)\, \mathcal{N}\big(x_0; \mu_{r|s}(x_s, y), \Sigma_{r|s}\big), \tag{18}$$

Since BBDM uses the posterior mean as the Bayes-optimal estimate of x_0 under squared loss, we get,

$$\hat{x}_{0,s}^{\mathrm{MMSE}}(x_s, y) = \mathbb{E}[x_0 \mid x_s, y] = \sum_{r=1}^R \gamma_{r|s}(x_s, y)\, \mu_{r|s}(x_s, y). \tag{19}$$

Equation (19) is the exact MoG analogue of the Gaussian oracle denoiser. Closed-form posterior inference is therefore not lost in the MoG case. What changes is the form of the posterior mean: it is now a weighted sum of affine estimators, with state-dependent component probabilities. This is exactly the mechanism that breaks the global affine reverse law in the next section.

4 The Loss of Global Affinity

The previous section provided the exact oracle denoiser. We now insert that denoiser into the BBDM sampler and inspect the resulting reverse update. We will see that in the MoG case, this recursion cannot be unrolled into a direct and simplified closed-form expression for the final reconstruction in terms of the measurement and the reverse Gaussian innovations. The fully expanded derivation is in Appendix C.

Lemma 4.1. *After substituting the exact MMSE denoiser* (19) *into the BBDM reverse update* (4) *for* $1 \le s \le S-1$*, the oracle MoG BBDM chain becomes*

$$\begin{aligned} x_{s-1} = &\left(c_s I + a_s \frac{1-m_s}{\delta_s} \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s} \right) x_s \\ &+ \left[b_s I + a_s \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s} \left(\frac{1}{\sigma_y^2} H^\top - \frac{m_s(1-m_s)}{\delta_s} I \right) \right] y \\ &+ a_s \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s} \Sigma_r^{-1} \mu_r + \sigma_s z_s. \end{aligned} \tag{20}$$

where $z_s \sim \mathcal{N}(0, I)$.

If $R = 1$, then $\gamma_{1|s}(x_s, y) \equiv 1$, and all coefficients in (20) depend only on the schedule and the model parameters. In this case, the oracle MoG BBDM chain is globally affine. Otherwise, the coefficients multiplying x_s and y depend on the current bridge state through the posterior component probabilities $\gamma_{r|s}(x_s, y)$. Consequently, unlike the single-Gaussian case, consecutive reverse steps do not share fixed linear operators. This prevents us from recursively composing the equations and obtaining a direct closed-form expression for the final reconstruction $\hat{x}_0$ in terms of the measurement and the accumulated reverse Gaussian innovations.

5 The Selected-Label Approximation

Section 4 showed that the oracle MoG BBDM chain loses global affinity only because the posterior component probabilities $\gamma_{r|s}(x_s, y)$ vary with the current state. The selected-label surrogate freezes exactly this source of nonlinearity: after sampling a label from the measurement posterior, it keeps that label fixed throughout the reverse chain.

The approximation proceeds in two steps. First, we choose an auxiliary label $J \in \{1, \dots, R\}$ from the measurement posterior,

$$\mathbb{P}(J = r \mid y) = \gamma_{r|y}(y). \tag{21}$$

Second, once $J = r$ is selected, we keep that label fixed throughout the reverse chain and use the corresponding component-conditioned posterior mean $\mu_{r|s}(x_s, y)$ from Section 3 inside the BBDM update. Detailed derivations are collected in Appendix D.

At the terminal update $s = S$, since $x_S = y$, the selected-label denoiser is interpreted as the measurement-only posterior mean $\hat{x}_0 = \mu_{r|y}$. For $1 \le s \le S-1$, substituting $\hat{x}_0 = \mu_{r|s}(x_s, y)$ into the BBDM reverse rule gives, conditioned on $J = r$,

$$x_{s-1} = \left(c_s I + a_s \frac{1-m_s}{\delta_s} \Sigma_{r|s} \right) x_s + \left(b_s I - a_s \frac{m_s(1-m_s)}{\delta_s} \Sigma_{r|s} \right) y + a_s \Sigma_{r|s} \Sigma_{r|y}^{-1} \mu_{r|y} + \sigma_s z_s, \qquad z_s \sim \mathcal{N}(0, I). \tag{22}$$

This is already a major simplification compared with the exact MoG reverse chain of Section 4: once the label is fixed, the one-step update becomes affine in x_s and y, with coefficients determined only by the selected component and the schedule.

The next question is whether these one-step updates can be analyzed in one convenient coordinate system. The key observation is that, for a fixed component r, all matrices $\Sigma_{r|s}$ are obtained from $\Sigma_{r|y}^{-1}$ by adding a scalar multiple of the identity and inverting. Therefore, all of them are diagonalized by the same basis.

Lemma 5.1. *For each fixed component* r*, there exists an orthogonal matrix* U_r *and positive scalars* $\lambda_{r,1}, \dots, \lambda_{r,d}$ *such that*

$$\Sigma_{r|y}^{-1} = U_r \operatorname{diag}(\lambda_{r,1}, \dots, \lambda_{r,d}) U_r^\top. \tag{23}$$

Moreover, for every interior step $1 \le s \le S-1$,

$$\Lambda_{r|s} = \operatorname{diag}\left(\frac{1}{\lambda_{r,1} + \frac{(1-m_s)^2}{\delta_s}}, \dots, \frac{1}{\lambda_{r,d} + \frac{(1-m_s)^2}{\delta_s}} \right) \tag{24}$$

satisfies

$$\Sigma_{r|s} = U_r \Lambda_{r|s} U_r^\top. \tag{25}$$

Lemma 5.1 is the structural replacement for the single global basis available in the one-Gaussian setting. It says that each selected component comes with its own basis U_r, and in that basis every posterior covariance along the reverse chain is diagonal. Unlike Benita et al. (2025; 2026), this construction does not restrict the degradation operator and prior covariance to share a global basis.

We now project the reverse update (22) onto this basis. Define

$$x_s^{U_r} = U_r^\top x_s, \quad y^{U_r} = U_r^\top y, \quad \mu_{r|y}^{U_r} = U_r^\top \mu_{r|y}, \quad z_s^{U_r} = U_r^\top z_s. \tag{26}$$

Because U_r is orthogonal, $z_s^{U_r} \sim \mathcal{N}(0, I)$.

Lemma 5.2. *Conditioned on $J = r$, the projected reverse chain satisfies*

$$x_{s-1}^{U_r} = G_r(s) x_s^{U_r} + N_r(s) y^{U_r} + M_r(s) \mu_{r|y}^{U_r} + \sigma_s z_s^{U_r}, \tag{27}$$

where

$$G_r(s) = c_s I + \frac{a_s(1-m_s)}{\delta_s} \Lambda_{r|s}, \quad N_r(s) = b_s I - \frac{a_s m_s (1-m_s)}{\delta_s} \Lambda_{r|s}, \quad M_r(s) = a_s \Lambda_{r|s} \operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d}). \tag{28}$$

The formulas above apply for $1 \le s \le S-1$. For the terminal reverse update $s = S$, we use the endpoint convention

$$G_r(S) = 0, \quad N_r(S) = m_{S-1} I, \quad M_r(S) = (1 - m_{S-1}) I, \quad \sigma_S^2 = \delta_{S-1}.$$

For all steps, all three matrices are diagonal. Hence, conditioned on $J = r$, the reverse process decouples into d scalar recursions in the basis U_r.

Lemma 5.2 says that once the label is frozen and the process is written in the basis U_r, the reverse chain becomes a diagonal step-by-step update with Gaussian noise. This is exactly what makes it possible to unroll the reconstruction law explicitly.

Corollary 5.3. *Starting from the BBDM initialization $x_S = y$, the fixed-label reverse chain can be unrolled as*

$$\hat{x}_0^{U_r,(r)} = D_1^{(r)} y^{U_r} + D_2^{(r)} \mu_{r|y}^{U_r} + \sum_{i=1}^{S} \left(\prod_{j=1}^{i-1} G_r(j) \right) \sigma_i z_i^{U_r}, \tag{29}$$

where

$$D_1^{(r)} = \prod_{s'=1}^{S} G_r(s') + \sum_{i=1}^{S} \left(\prod_{j=1}^{i-1} G_r(j) \right) N_r(i), \quad D_2^{(r)} = \sum_{i=1}^{S} \left(\prod_{j=1}^{i-1} G_r(j) \right) M_r(i). \tag{30}$$

Consequently, the selection approximation law conditioned on a specific gaussian is,

$$p_{\text{Select}}(\hat{x}_0 \mid y, J = r) = \mathcal{N}\big(\hat{x}_0; \mu_{\text{BBDM}}^{(r)}(y), \Sigma_{\text{BBDM}}^{(r)}\big), \tag{31}$$

with

$$\mu_{\text{BBDM}}^{(r)}(y) = U_r \Big(D_1^{(r)} y^{U_r} + D_2^{(r)} \mu_{r|y}^{U_r} \Big), \tag{32}$$

$$\Sigma_{\text{BBDM}}^{(r)} = U_r \left[\sum_{i=1}^{S} \left(\prod_{j=1}^{i-1} G_r(j) \right) \sigma_i^2 I \left(\prod_{j=1}^{i-1} G_r(j) \right)^\top \right] U_r^\top. \tag{33}$$

After removing the conditioning on J,

$$p_{\text{Select}}(\hat{x}_0 \mid y) = \sum_{r=1}^{R} \gamma_{r|y}(y) \, \mathcal{N}\big(\hat{x}_0; \mu_{\text{BBDM}}^{(r)}(y), \Sigma_{\text{BBDM}}^{(r)}\big). \tag{34}$$

Corollary 5.3 summarizes the full payoff of the selected-label approximation. Each possible selected component produces one explicit Gaussian reconstruction law, analyzed in its own basis. The overall selected-label law is then obtained by mixing these Gaussian outputs with the measurement-posterior probabilities.

The unrolled law contains a schedule-dependent mean term through the matrices $D_1^{(r)}$ and $D_2^{(r)}$. However, the next result shows that these matrices simplify quite significantly.

Proposition 5.4. *Conditioned on the selected label $J = r$, the selected-label BBDM output satisfies*

$$\mu_{\mathrm{BBDM}}^{(r)}(y) \triangleq \mathbb{E}[\hat{x}_0 \mid y, J = r] = \mu_{r|y}. \tag{35}$$

Equivalently, in the notation of Corollary 5.3,

$$D_1^{(r)} = 0, \qquad D_2^{(r)} = I. \tag{36}$$

This proposition shows that the selected-label surrogate does not introduce any component-wise bias. The next proposition shows that the remaining discrepancy is a loss of covariance.

Proposition 5.5. *In the component-wise basis U_r, write*

$$U_r^\top \Sigma_{\mathrm{BBDM}}^{(r)} U_r = \mathrm{diag}\big(\sigma^2_{\mathrm{BBDM},r,1}, \ldots, \sigma^2_{\mathrm{BBDM},r,d}\big), \tag{37}$$

where

$$\sigma_{\mathrm{BBDM},r,k} = \left(\sum_{i=1}^{S} \sigma_i^2 \prod_{j=1}^{i-1} g_{r,k}(j)^2 \right)^{1/2}, \tag{38}$$

and $g_{r,k}(j)$ denotes the k-th diagonal entry of $G_r(j)$. Then

$$\Sigma_{r|y} - \Sigma_{\mathrm{BBDM}}^{(r)} \succeq 0. \tag{39}$$

Equivalently, in the component-wise basis,

$$0 \le \sigma^2_{\mathrm{BBDM},r,k} \le \frac{1}{\lambda_{r,k}}, \qquad k = 1, \ldots, d. \tag{40}$$

Proposition 5.5 shows that the selected-label BBDM sampler is covariance-deficient. For any valid schedule, it may approach the component posterior covariance $\Sigma_{r|y}$ from below, but it cannot exceed it in any direction.

6 Two Schedule Design Objectives

The previous section produced an explicit surrogate reconstruction law. Schedule design now becomes a comparison problem: for a given schedule, how close is the selected-label reconstruction law $p_{\mathrm{Select}}(\hat{x}_0 \mid y)$ to the target measurement posterior $p(x_0 \mid y)$? We begin with the exact posterior after observing only the measurement, then compare it with the selected-label law for the same observation, and finally average the resulting comparison over the measurement distribution. This leads to two complementary objectives. The first is a matched-component Wasserstein upper-bound surrogate for comparing posterior laws. The second is based on expected mean-squared error under an explicit matched-label product coupling and is aligned with low-distortion, pixel-wise reconstruction accuracy. The full derivations appear in Appendix E.

Lemma 6.1. *After observing only the measurement y,*

$$p(x_0 \mid y) = \sum_{r=1}^{R} \mathbb{P}(C = r \mid y)\, p(x_0 \mid y, C = r) = \sum_{r=1}^{R} \gamma_{r|y}(y)\, \mathcal{N}\big(x_0; \mu_{r|y}, \Sigma_{r|y}\big). \tag{41}$$

Lemma 6.1 follows directly from the formulas of Section 3. The important point for what follows is that the true measurement posterior and the selected-label law share the same component weights $\gamma_{r|y}(y)$. They differ only in the Gaussian component attached to each possible label.

Lemma 6.2. *For fixed y,*

$$W_2^2\big(p(x_0 \mid y), p_{\text{Select}}(\hat{x}_0 \mid y)\big) \le \sum_{r=1}^{R} \gamma_{r|y}(y)\, W_2^2\Big(\mathcal{N}(\mu_{r|y}, \Sigma_{r|y}), \mathcal{N}\big(\mu_{\text{BBDM}}^{(r)}(y), \Sigma_{\text{BBDM}}^{(r)}\big)\Big). \tag{42}$$

The bound in Lemma 6.2 follows from the definition of W_2^2 as an infimum over all couplings. Since the true posterior and the selected-label law have the same mixture weights $\gamma_{r|y}(y)$, we build one explicit coupling by using the same random component index in both mixtures and, conditioned on that index, coupling the corresponding Gaussian pair optimally. Evaluating its transport cost therefore gives an explicit upper bound, which reduces the comparison to a weighted sum of Gaussian costs.

Lemma 6.3. *Using the standard closed form for the quadratic Wasserstein distance between Gaussian measures (Dowson & Landau, 1982; Peyré & Cuturi, 2019), the r-th Gaussian cost in Lemma 6.2 is*

$$W_2^2\Big(\mathcal{N}(\mu_{r|y}, \Sigma_{r|y}), \mathcal{N}\big(\mu_{\text{BBDM}}^{(r)}(y), \Sigma_{\text{BBDM}}^{(r)}\big)\Big) = \sum_{k=1}^{d} \left(\sigma_{\text{BBDM},r,k} - \frac{1}{\sqrt{\lambda_{r,k}}}\right)^2.$$

Lemma 6.3 shows that, after mean exactness is taken into account, the matched-component comparison is purely a covariance comparison. Averaging this expression over the measurement distribution gives the Wasserstein upper-bound schedule objective.

Corollary 6.4. *The matched-component Wasserstein upper-bound selected-label schedule objective is*

$$J_{\text{MOG}}^{W_2} = \sum_{r=1}^{R} \pi_r \sum_{k=1}^{d} \left(\sigma_{\text{BBDM},r,k} - \frac{1}{\sqrt{\lambda_{r,k}}}\right)^2. \tag{43}$$

The Wasserstein objective compares the selected-label law with the true measurement posterior at the level of entire Gaussian components. The next objective will evaluate reconstruction similarity directly through expected mean-squared error, rather than compare full distributions.

Corollary 6.5. *Under the natural product coupling between the matched Gaussian pairs, the expected-MSE objective is*

$$J_{\text{MOG}}^{\text{MSE}} = \sum_{r=1}^{R} \pi_r \sum_{k=1}^{d} \left(\sigma_{\text{BBDM},r,k}^2 + \frac{1}{\lambda_{r,k}}\right). \tag{44}$$

This is the matched-label product-coupling MSE. If the true posterior sample and the selected-label reconstruction are instead assigned independent mixture labels, an additional component-mean term appears. However, the extra term is independent of the schedule. Thus the schedule optimizer is unchanged, although the absolute MSE value is different.

Corollaries 6.4 and 6.5 differ only in how they treat the remaining component covariance. The Wasserstein criterion tries to match the posterior standard deviations $1/\sqrt{\lambda_{r,k}}$, whereas the expected-MSE criterion penalizes the sampler variance itself. Thus the Wasserstein objective favors posterior-distribution matching, while the MSE objective favors concentration around the posterior mean. This observation is closely related to the perception–distortion tradeoff of Blau & Michaeli (2018), which formalizes the tension between distortion-based fidelity and perceptual quality in image restoration.

Bounded schedule family: To design an optimal measurement-aware BBDM schedule for the Mixture-of-Gaussians setting, we seek parameters

$$\Theta = \{(m_s, \delta_s)\}_{s=1}^{S}$$

that minimize either $J_{\text{MOG}}^{W_2}(\Theta)$ or $J_{\text{MOG}}^{\text{MSE}}(\Theta)$. Rather than optimize over all discrete pairs directly, we use the bounded four-parameter family

$$m_s = 1 - (1 - \tau_s^{\alpha})^{\beta}, \quad \delta_s = c\,[4m_s(1 - m_s)]^{\gamma}, \quad \tau_s = \frac{s}{S}, \tag{45}$$

with

$$\alpha \in [1, 2], \quad \beta \in [1, 2], \quad c \in [0.2, 2], \quad \gamma \in [0.2, 2]. \tag{46}$$

This keeps the search low-dimensional while enforcing the Brownian-bridge constraints by construction and avoiding degenerate schedules. Appendix F derives MoG-free edge rules for this family: the MSE variance bound is minimized at $(\alpha, \beta, c, \gamma) = (1, 2, 2, 0.2)$, while the W_2 heuristic uses the opposite edge $(2, 1, 0.2, 2)$ to increase sampler variance toward the posterior spread. These are the problem-agnostic schedules used when reliable high-dimensional MoG covariance estimates are unavailable. Figure 1 visualizes the three BBDM schedules considered in this work for $S = 200$ sampling steps.

Appendix H repeats the same MoG and selected-label pipeline for a conditional DDIM sampler with an oracle posterior denoiser. The comparison isolates which parts of the analysis are specific to the Brownian bridge and which arise from the broader MoG posterior structure.

Finally, Theorem G.2 in Appendix G provides theoretical support for the selected-label approximation itself. In a shared-covariance, linearly separated high-dimensional MoG regime, it proves that if the correct label is frozen, then the selected-label chain remains in states where the exact MoG responsibilities continue to favor that same label with overwhelming probability. At the one step level - the full MOG denoiser versus the selected label denoiser discrepancy is negligible.

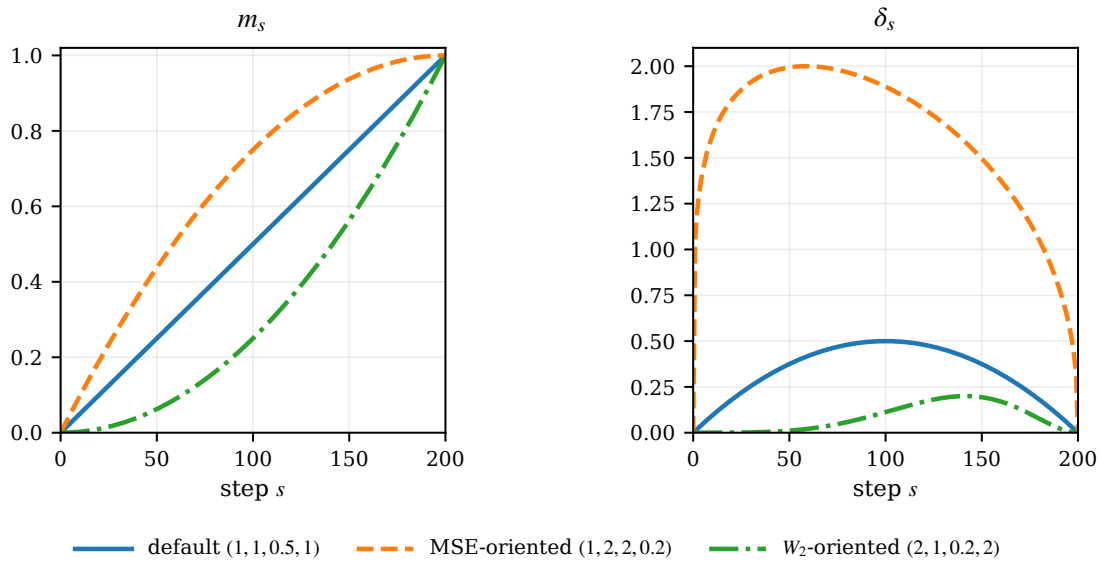

Figure 1: Schedule shapes for the three BBDM schedules used in the FFHQ experiments, shown for $S = 200$ sampling steps. Left: interpolation coefficient m_s. Right: bridge variance coefficient δ_s.

7 Related work

This paper is closest to three lines of work: bridge-based diffusion, diffusion methods for inverse problems, and analytical studies of diffusion inference and schedule design.

Bridge and stochastic-interpolation diffusion. BBDM models image-to-image translation through a stochastic bridge whose endpoint is the target observation (Li et al., 2023). Closely related, I^{2}SB learns diffusion bridges between degraded and clean image distributions for restoration tasks (Liu et al., 2023). More broadly, Schrödinger-bridge and stochastic-interpolant formulations view diffusion models as stochastic paths that transport samples between two endpoint distributions over a fixed time interval (Bortoli et al., 2021; Albergo et al., 2023). These works mainly develop bridge or interpolation constructions and the

corresponding training and sampling procedures. Our focus is not a new bridge architecture, but an analysis of the reconstruction law induced by a BBDM sampler and its schedule.

Diffusion models for inverse problems. A large body of work applies diffusion models to image restoration and inverse problems. Conditional diffusion models such as SR3 and Palette train networks conditioned on degraded observations (Saharia et al., 2021; 2022), whereas zero-shot methods such as RePaint, DDRM, DDNM, DPS, and PiGDM use pretrained diffusion models together with measurement constraints or posterior guidance during sampling (Lugmayr et al., 2022; Kawar et al., 2022; Wang et al., 2023; Chung et al., 2023; Song et al., 2023). The main text focuses on the BBDM setting in which the measurement is built into the bridge endpoint. However, Appendix H further shows that the same MoG and selected-label mechanism is not specific to BBDM by deriving the analogous analysis for a conditional DDIM sampler on which many of those works are built.

Analytical studies of diffusion inference. Noise schedules and sampler parameterizations are known to strongly affect diffusion behavior. Improved DDPMs and EDM-style analyses show that performance depends not only on the denoising network, but also on choices such as the reverse variance, the noise or data parameterization, and the sampler dynamics (Nichol & Dhariwal, 2021; Karras et al., 2022). Closer to our work, Benita et al. (2025; 2026) derive exact reconstruction laws and schedule-design criteria under a single-Gaussian prior. We follow this analytical program but replace the single-Gaussian model by a Mixture-of-Gaussians prior leading to our selected-label surrogate and component-wise basis analysis.

8 Experiments

The experiments follow a three-stage progression. First, in controlled MoG models, where the posterior and oracle samplers are clearly available, we validate the selected-label theory and compare the BBDM sampler with a conditional DDIM sampler. Second, on MNIST (LeCun et al., 1998), where fitting a MoG is still feasible, we test whether the frozen surrogate can faithfully capture scheduling parameter trends that align with BBDM with learned denoiser. We also show in this case that modeling the prior with more Gaussians directly improves the frozen model reconstruction capabilities. Finally, on FFHQ (Karras et al., 2019), where reliable high-dimensional MoG fitting is impractical, we use the MoG-free MSE-oriented and W_2-oriented schedule heuristics of Appendix F. Additional plots and qualitative examples are found in Appendix I.

8.1 Theory validation in the shared-covariance regime

We begin with synthetic shared-covariance MoG data. The self-consistency result in Appendix G predicts that, when the measurement separation grows with the dimension, the responsibilities should become nearly one-hot and the frozen selected label should remain stable along its own reverse chain. The synthetic study confirms empirically that as the ambient dimension increases, the realized separation statistic approaches its theoretical threshold, while the selected-label error, the denoiser discrepancy, and the final reconstruction discrepancy all decrease rapidly. Empirically, once the measurement margin is large enough, the exact MoG reverse process behaves almost as if a single component had been selected in advance. The full four-panel validation plot appears in Appendix I.

8.2 Blended tradeoff under oracle MoG and frozen reverse chains

We next use the random means, shared-covariance toy model to visualize the tradeoff induced by blending the two schedule objectives of Section 6. For different λ values we optimize

$$J_\lambda = (1-\lambda)J_{\mathrm{MOG}}^{W_2} + \lambda J_{\mathrm{MOG}}^{\mathrm{MSE}}$$

over the bounded schedule family and then evaluate the chosen schedule on two empirical axes over $n_y = 100$ measurements: mean-squared error to the true signal and the sliced W_2 distance (Bonneel et al., 2015) between reconstruction samples and posterior samples. The selected-label surrogate reconstructions are shown as dots and the oracle MoG BBDM reconstructions as crosses. For each measurement y, the frozen label is sampled from the measurement posterior and is kept fixed along the frozen chain.

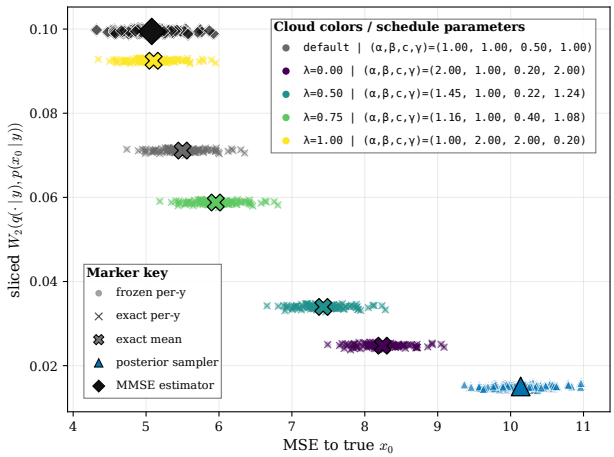

Figure 2: Shared-covariance toy experiment for schedules optimized with J_λ. The prior is a uniform-weight MoG with $R = 32$ components in dimension $d = 512$. The component means are sampled independently as $\mu_{r,k} \sim \mathrm{Unif}[-1, 1]$, the inverse problem is denoising with $H = I$ and $\sigma_y = 0.1$, and all components share the diagonal covariance $\Sigma = \mathrm{diag}(\mathrm{geomspace}(0.5, 2, d))$. The BBDM chains use $S = 20$ reverse steps. The plot evaluates the schedules optimized for $\lambda \in \{0, 0.5, 0.75, 1\}$, using 256 samples per measurement and 256 random projections for the sliced W_2 estimate. The direct posterior sampler and posterior MMSE estimator are included as schedule-independent Bayesian references. The direct posterior sampler has nonzero empirical sliced-W_2 because it is compared to independent finite posterior reference samples.

8.3 BBDM–DDIM comparison

We next compare BBDM and conditional DDIM as a function of the number of reverse steps S. We use the same shared-covariance MoG setting as above, with $R = 32$, $d = 512$, uniform mixture weights, $\mu_{r,k} \sim \mathrm{Unif}[-1, 1]$, and $\Sigma = \mathrm{diag}(\mathrm{geomspace}(0.5, 2, d))$. The inverse problem is denoising, $H = I$, with $\sigma_y = 0.5$.For each $S \in \{2, \ldots, 1000\}$, we evaluate the two normalized selected-label objectives

$$\frac{1}{d} J_{\mathrm{MOG}}^{W_2} = \frac{1}{d} \sum_{k=1}^{d} \left(\sigma_{\mathrm{model},k}(S) - \frac{1}{\sqrt{\lambda_k}} \right)^2,$$

$$\frac{1}{d} J_{\mathrm{MOG}}^{\mathrm{MSE}} = \frac{1}{d} \sum_{k=1}^{d} \left(\sigma_{\mathrm{model},k}^2(S) + \frac{1}{\lambda_k} \right), \qquad \lambda_k = \Sigma_{kk}^{-1} + \sigma_y^{-2}.$$

We compare the default conditional DDIM schedule, where

$$\beta_t = 10^{-4} + \frac{t-1}{T-1}(2 \cdot 10^{-2} - 10^{-4}), \qquad t = 1, \ldots, T,$$

and

$$\bar{\alpha}_0 = 1, \qquad \bar{\alpha}_t = \prod_{u=1}^{t} (1 - \beta_u).$$

with three BBDM schedules: the default bridge schedule $(1, 1, 0.5, 1)$, the MSE-oriented schedule $(1, 2, 2, 0.2)$, and the W_2-oriented schedule $(2, 1, 0.2, 2)$.

Figure 3 shows the complementary behavior of the two objectives. Under W_2, the default BBDM schedule has a small-S advantage over DDIM, while the W_2-oriented schedule gives the lowest loss over most of the sampling range. Under MSE, the ranking reverses: the MSE-oriented BBDM schedule gives the lowest loss throughout, whereas the W_2-oriented schedule incurs a larger MSE because it preserves more posterior variance. These results agree with the roles of the two objectives: W_2 favors posterior-spread matching, while MSE favors concentration around the posterior mean.

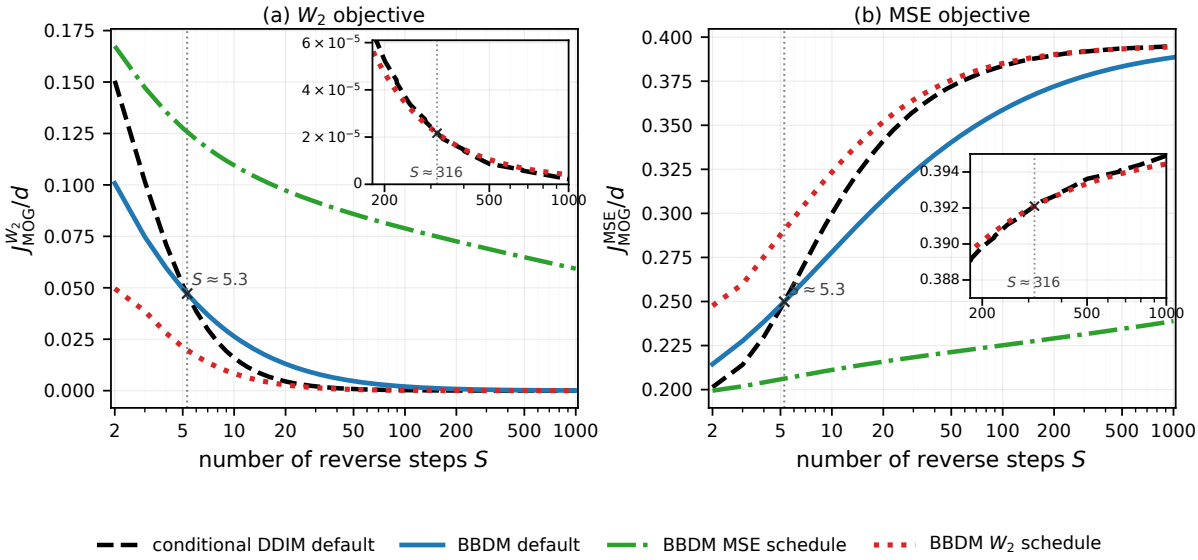

Figure 3: Step-wise analytic comparison between conditional DDIM and BBDM schedules in the shared-covariance MoG denoising setting with $\sigma_y = 0.5$. Panel (a) reports the normalized matched-component Wasserstein objective $J^{W_2}_{\mathrm{MOG}}/d$, while panel (b) reports the normalized matched-label product-coupling MSE $J^{\mathrm{MSE}}_{\mathrm{MOG}}/d$. The default BBDM and DDIM curves cross near $S \simeq 5.3$. The W_2-oriented BBDM schedule gives the lowest Wasserstein loss over most of the range, whereas the MSE-oriented schedule gives the lowest MSE throughout.

8.4 MNIST validation and transfer to trained BBDM models

We use MNIST to test whether the selected-label surrogate can guide schedule design for trained BBDM models. The models address Fourier low-pass measurements of the form $y = Hx_0 + n$, with retained frequency fraction V and $n \sim \mathcal{N}(0, \sigma_y^2 I)$.

The detailed surrogate validation is reported in Appendix I. Figure 6 evaluates whether the surrogate reproduces the schedule trends of a trained BBDM, while Figure 7 examines how the surrogate depends on the richness of the fitted MoG prior. Together, these experiments show that the surrogate provides a sufficiently informative model for guiding the schedule search.

We then optimize the surrogate MSE objective for five degradation settings and train corresponding BBDM models on the MNIST training set for 40 epochs using either the default or selected schedule. The search consistently returns the MSE-oriented schedule $(\alpha, \beta, c, \gamma) \approx (1, 2, 2, 0.2)$.

Table 1 reports paired reconstruction metrics: PSNR and structural similarity index measure (SSIM) (Wang et al., 2004), calculated from piq library, together with an external pretrained MNIST classifier negative log-likelihood (`super-j/vit-base-mnist`) that measures whether the reconstructed digit remains semantically recognizable. We do not use Fréchet Inception Distance (FID) (Heusel et al., 2017) or Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al., 2018) in this experiment: FID is primarily a distribution-level natural-image metric, while LPIPS relies on deep perceptual features that are not well matched to low-resolution grayscale digit images. The optimized schedule improves PSNR and NLL in all five settings. SSIM improves in four settings, and decreases only slightly on the baseline pair.

Table 1: MNIST blur: default schedule versus the MSE-optimized schedule.

V	σ_y	Sch.	PSNR↑	SSIM↑	NLL↓
0.10	0.10	def.	25.360	**0.9637**	0.04250
		opt.	**26.335**	0.9485	**0.04128**
0.10	0.20	def.	23.233	0.9270	0.05477
		opt.	**24.558**	**0.9385**	**0.04840**
0.30	0.10	def.	30.565	0.9680	0.03150
		opt.	**30.818**	**0.9870**	**0.03140**
0.05	0.10	def.	22.928	0.9360	0.06640
		opt.	**24.238**	**0.9470**	**0.05940**
1.00	0.50	def.	23.736	0.9460	0.04300
		opt.	**25.568**	**0.9640**	**0.03440**

8.5 MoG-free heuristics for real-world datasets

We now evaluate the MoG-free schedule heuristics of Appendix F on FFHQ images at resolution $256 \times 256 \times 3$. Unless stated otherwise, all models are trained for 60 epochs on 54,000 training images, with 7,000 validation images and 7,000 test images, and evaluated with 200 sampling steps. We compare the default BBDM schedule with the MSE-oriented schedule $(\alpha, \beta, c, \gamma) = (1, 2, 2, 0.2)$ and the W_2-oriented schedule $(\alpha, \beta, c, \gamma) = (2, 1, 0.2, 2.0)$, using FID, LPIPS, PSNR and SSIM. FID is computed using `torch-fidelity` via the `fidelity` command, while LPIPS is computed using the `lpips` package with the learned AlexNet model.

For reference, we also train architecture-matched conditional DDIM baseline. This baseline uses the same UNet architecture, training data, optimizer, training budget, EMA, and evaluation protocol as the BBDM models. The only algorithmic change is the diffusion path and sampler: conditional DDIM uses the standard data-to-noise forward process, receives y only as denoiser conditioning, starts sampling from Gaussian noise, and uses deterministic DDIM sampling.

We consider three families of linear inverse problems. The blur setting uses a Fourier low-pass operator with retained frequency fraction V. The super-resolution setting uses

$$y = U_s D_s x_0 + n, \qquad n \sim \mathcal{N}(0, \sigma_y^2 I),$$

where D_s downsamples by a factor s in each spatial dimension and U_s upsamples the resulting low-resolution image back to the original grid using nearest-neighbor interpolation. The distributed inpainting setting uses

$$y = M_p x_0 + n, \qquad n \sim \mathcal{N}(0, \sigma_y^2 I),$$

where M_p is a fixed binary mask that keeps a fraction p of the RGB pixels. Qualitative results are shown in Appendix I.

The results show that in most settings the conditional DDIM baseline improves substantially over the default BBDM schedule in FID and LPIPS. However, the W_2-oriented BBDM schedule still achieves the best FID in all FFHQ inverse problems, and usually gives the best LPIPS as well.

The distortion metrics follow a complementary pattern. The MSE-oriented BBDM schedule gives the best PSNR and SSIM in every FFHQ setting, while both the W_2-oriented BBDM schedule and conditional DDIM tend to sacrifice distortion for more perceptual reconstructions.

The additional sampling-step ablation in Appendix Table 5 shows that the same trend between the different BBDM schedules is mostly present also for other amounts of sampling steps.

Table 2: FFHQ blur: BBDM schedules and conditional DDIM comparison.

V	σ_y	Method	FID↓	LPIPS↓	PSNR↑	SSIM↑
0.10	0.10	def.	11.527	0.0709	30.656	0.866
		MSE	21.807	0.1039	**31.792**	**0.888**
		W_2	**5.007**	**0.0562**	29.996	0.839
		cDDIM	7.034	0.0603	29.891	0.843
0.03	0.10	def.	15.942	0.1266	27.217	0.773
		MSE	25.073	0.1731	**28.547**	**0.806**
		W_2	**9.374**	0.1138	27.324	0.756
		cDDIM	12.400	**0.1108**	26.800	0.748
0.30	0.10	def.	7.880	0.0396	33.372	0.914
		MSE	13.166	0.0511	**34.871**	**0.934**
		W_2	**2.967**	**0.0278**	32.468	0.893
		cDDIM	4.099	0.0295	32.731	0.901
0.10	0.20	def.	17.294	0.0985	29.471	0.841
		MSE	26.820	0.1222	**30.815**	**0.868**
		W_2	**5.996**	**0.0754**	28.958	0.811
		cDDIM	8.006	0.0775	28.714	0.811

Table 3: FFHQ super-resolution: BBDM schedules and conditional DDIM comparison.

Scale	σ_y	Method	FID↓	LPIPS↓	PSNR↑	SSIM↑
4×	0.10	def.	11.720	0.0851	28.888	0.830
		MSE	23.997	0.1276	**30.168**	**0.860**
		W_2	**5.066**	**0.0721**	28.526	0.806
		cDDIM	8.538	0.0766	28.325	0.806
8×	0.10	def.	19.950	0.1623	25.018	0.716
		MSE	34.414	0.2049	**26.060**	**0.754**
		W_2	**12.258**	0.1533	24.837	0.696
		cDDIM	18.165	**0.1510**	24.624	0.688

Table 4: FFHQ distributed inpainting: BBDM schedules and conditional DDIM comparison.

p	σ_y	Method	FID↓	LPIPS↓	PSNR↑	SSIM↑
0.25	0.10	def.	14.289	0.0794	29.802	0.863
		MSE	19.209	0.0855	**31.311**	**0.889**
		W_2	**5.026**	**0.0670**	27.890	0.808
		cDDIM	9.591	0.0711	28.969	0.836
0.125	0.10	def.	15.081	**0.1061**	27.245	0.807
		MSE	22.808	0.1210	**28.790**	**0.844**
		W_2	**11.099**	0.1265	24.816	0.726
		cDDIM	15.731	0.1089	26.592	0.777

9 Conclusion

This work develops an analytical framework for schedule design in Brownian Bridge Diffusion Models. Under a Mixture-of-Gaussians prior, the selected-label surrogate reduces the effect of the bridge schedule to an explicit covariance-shaping problem. The resulting objectives clarify the distortion–perception tradeoff: MSE-oriented schedules favor concentration near the posterior mean, whereas Wasserstein-oriented schedules favor matching the posterior spread.

A central direction for future work is to find a way to solve the derived MoG schedule objectives directly over the full valid schedule space for high dimensional datasets. This would replace the bounded parametric search and MoG-free directional heuristics used here with optimization procedures that more faithfully target the actual covariance criteria.

References

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023.

Roi Benita, Michael Elad, and Joseph Keshet. Designing scheduling for diffusion models via spectral analysis. *arXiv preprint arXiv:2502.00180*, 2025.

Roi Benita, Michael Elad, and Joseph Keshet. Analyzing and guiding zero-shot posterior sampling in diffusion models. *arXiv preprint arXiv:2602.07715*, 2026.

Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6228–6237, 2018.

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023.

D. C. Dowson and B. V. Landau. The fréchet distance between multivariate normal distributions. *Journal of Multivariate Analysis*, 12(3):450–455, 1982.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

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

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4401–4410, 2019.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. BBDM: Image-to-image translation with brownian bridge diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1952–1961, 2023.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A. Theodorou, Weili Nie, and Anima Anandkumar. I^2SB: Image-to-image schrödinger bridge. In *International Conference on Machine Learning (ICML)*, 2023.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11461–11471, 2022.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning (ICML)*, 2021.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019.

Mark Rudelson and Roman Vershynin. Hanson–wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18(82):1–9, 2013.

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *arXiv preprint arXiv:2104.07636*, 2021.

Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. *ACM Transactions on Graphics*, 2022.

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

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations (ICLR)*, 2023.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

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

A Generic Gaussian conditioning identities

This appendix records a standard Gaussian conditioning identity that is used repeatedly throughout the paper.

Posterior under an affine Gaussian observation

Let

$$x \sim \mathcal{N}(\mu, \Sigma),$$

and suppose that the observation $z \in \mathbb{R}^m$ is generated by

$$z = Ax + b + \eta, \qquad \eta \sim \mathcal{N}(0, R), \qquad \eta \perp x, \tag{47}$$

where $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$, and $R \succ 0$.

Then the posterior distribution of x given z is Gaussian:

$$x \mid z \sim \mathcal{N}(\mu_{x|z}, \Sigma_{x|z}), \tag{48}$$

with

$$\Sigma_{x|z}^{-1} = \Sigma^{-1} + A^\top R^{-1} A, \tag{49}$$

$$\mu_{x|z} = \Sigma_{x|z}\big(\Sigma^{-1}\mu + A^\top R^{-1}(z - b)\big). \tag{50}$$

Proof. By Bayes' rule,

$$p(x \mid z) \propto p(z \mid x)\, p(x).$$

Using (47),

$$p(z \mid x) \propto \exp\left(-\frac{1}{2}(z - Ax - b)^\top R^{-1}(z - Ax - b)\right),$$

and since $x \sim \mathcal{N}(\mu, \Sigma)$,

$$p(x) \propto \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right).$$

Therefore,

$$p(x \mid z) \propto \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu) - \frac{1}{2}(z - Ax - b)^\top R^{-1}(z - Ax - b)\right).$$

Expanding only the terms that depend on x, and ignoring the factor of $\frac{1}{2}$ we obtain

$$\begin{aligned}
&(x - \mu)^\top \Sigma^{-1}(x - \mu) + (z - Ax - b)^\top R^{-1}(z - Ax - b) \\
&= x^\top \Sigma^{-1} x - 2x^\top \Sigma^{-1}\mu + x^\top A^\top R^{-1} A x - 2x^\top A^\top R^{-1}(z - b) + \text{const}.
\end{aligned}$$

Grouping the quadratic and linear terms in x gives

$$x^\top\big(\Sigma^{-1} + A^\top R^{-1} A\big)x - 2x^\top\big(\Sigma^{-1}\mu + A^\top R^{-1}(z - b)\big) + \text{const}.$$

This is the canonical form of a Gaussian density in x. Hence the posterior precision matrix is

$$\Sigma_{x|z}^{-1} = \Sigma^{-1} + A^\top R^{-1} A,$$

and the posterior mean is

$$\mu_{x|z} = \Sigma_{x|z}\big(\Sigma^{-1}\mu + A^\top R^{-1}(z - b)\big).$$

This proves (49)–(50).

B Full derivation of the exact MoG posterior and the MMSE denoiser

This appendix expands Section 3 step by step. The objective is to compute the posterior distribution

$$p(x_0 \mid x_s, y)$$

and its posterior mean, which is the MMSE-optimal estimator of x_0.

Model

We assume that the clean signal $x_0 \in \mathbb{R}^d$ is drawn from the mixture

$$p(x_0) = \sum_{r=1}^{R} \pi_r \, \mathcal{N}(x_0; \mu_r, \Sigma_r), \qquad \pi_r > 0, \qquad \sum_{r=1}^{R} \pi_r = 1. \tag{51}$$

The measurement is generated through the linear model

$$y \mid x_0 \sim \mathcal{N}(Hx_0, \sigma_y^2 I), \tag{52}$$

and the BBDM forward process at step s is

$$x_s = (1 - m_s)x_0 + m_s y + \sqrt{\delta_s}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{53}$$

Roadmap of the derivation

The MoG case contains two sources of uncertainty:

- the *continuous* uncertainty of x_0 within each Gaussian component.
- the *discrete* uncertainty of which component generated the sample.

To separate them, we introduce a latent component variable

$$C \in \{1, \dots, R\},$$

and write the prior hierarchically as

$$\mathbb{P}(C = r) = \pi_r, \qquad x_0 \mid (C = r) \sim \mathcal{N}(\mu_r, \Sigma_r). \tag{54}$$

Our target is the posterior

$$p(x_0 \mid x_s, y).$$

Using the law of total probability over C, we decompose it as

$$p(x_0 \mid x_s, y) = \sum_{r=1}^{R} p(x_0 \mid x_s, y, C = r)\, \mathbb{P}(C = r \mid x_s, y). \tag{55}$$

Thus, it is enough to compute two ingredients:

1. the component-conditioned posterior

$$p(x_0 \mid x_s, y, C = r),$$

2. the posterior component probabilities

$$\mathbb{P}(C = r \mid x_s, y),$$

The key Gaussian conditioning identity used throughout is recorded in Appendix A.

We proceed in the following order:

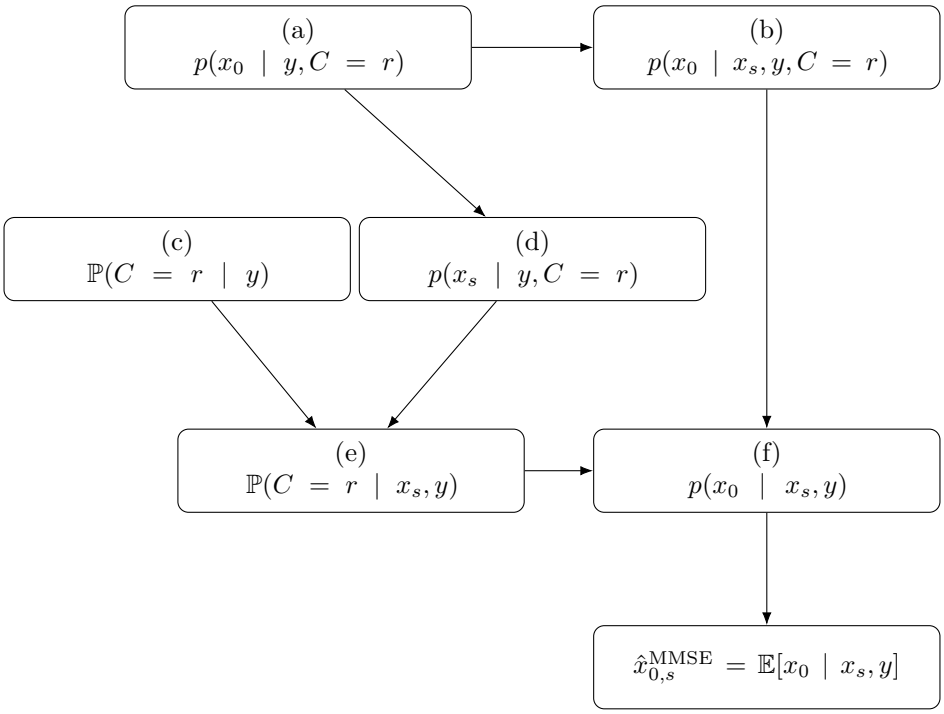

Figure 4: Dependency graph for the exact MoG posterior derivation.

(a) Component-conditioned posterior

In this subsection, we derive

$$p(x_0 \mid y, C = r).$$

Fix a component index r. Conditioned on $C = r$, we have

$$x_0 \mid (C = r) \sim \mathcal{N}(\mu_r, \Sigma_r),$$

and the measurement model is

$$y = Hx_0 + n, \qquad n \sim \mathcal{N}(0, \sigma_y^2 I).$$

This is an affine Gaussian observation model of the form treated in Appendix A. Applying the generic posterior formula with

$$x = x_0, \qquad z = y, \qquad A = H, \qquad b = 0, \qquad R = \sigma_y^2 I,$$

we obtain

$$x_0 \mid (y, C = r) \sim \mathcal{N}(\mu_{r|y}, \Sigma_{r|y}),$$

where

$$\Sigma_{r|y}^{-1} = \Sigma_r^{-1} + \frac{1}{\sigma_y^2} H^\top H, \tag{56}$$

$$\mu_{r|y} = \Sigma_{r|y} \left(\Sigma_r^{-1} \mu_r + \frac{1}{\sigma_y^2} H^\top y \right). \tag{57}$$

Thus, once both y and the component index r are fixed, the posterior over x_0 is Gaussian.

(b) Component-conditioned posterior with x_s

In this subsection, we derive

$$p(x_0 \mid x_s, y, C = r),$$

namely the posterior of x_0 after observing both y and the bridge variable x_s, while keeping the component index fixed.

From part (a), conditioned on $(y, C = r)$ we have

$$x_0 \mid (y, C = r) \sim \mathcal{N}(\mu_{r|y}, \Sigma_{r|y}).$$

The bridge variable satisfies

$$x_s = (1 - m_s)x_0 + m_s y + \sqrt{\delta_s}\,\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I).$$

Hence, conditioned on $(y, C = r)$, x_s is again an affine Gaussian observation of x_0. Applying the same generic identity from Appendix A with

$$x = x_0, \qquad z = x_s, \qquad A = (1 - m_s)I, \qquad b = m_s y, \qquad R = \delta_s I,$$

and prior $x_0 \mid (y, C = r) \sim \mathcal{N}(\mu_{r|y}, \Sigma_{r|y})$, we obtain

$$x_0 \mid (x_s, y, C = r) \sim \mathcal{N}\big(\mu_{r|s}(x_s, y), \Sigma_{r|s}\big),$$

where

$$\Sigma_{r|s} = \left(\Sigma_{r|y}^{-1} + \frac{(1 - m_s)^2}{\delta_s} I\right)^{-1}, \tag{58}$$

$$\mu_{r|s}(x_s, y) = \Sigma_{r|s}\left(\Sigma_{r|y}^{-1}\mu_{r|y} + \frac{1 - m_s}{\delta_s}(x_s - m_s y)\right). \tag{59}$$

Using

$$\Sigma_{r|y}^{-1} = \Sigma_r^{-1} + \frac{1}{\sigma_y^2} H^\top H, \qquad \Sigma_{r|y}^{-1}\mu_{r|y} = \Sigma_r^{-1}\mu_r + \frac{1}{\sigma_y^2} H^\top y,$$

we equivalently obtain

$$\Sigma_{r|s} = \left(\Sigma_r^{-1} + \frac{1}{\sigma_y^2} H^\top H + \frac{(1 - m_s)^2}{\delta_s} I\right)^{-1}, \tag{60}$$

$$\mu_{r|s}(x_s, y) = \Sigma_{r|s}\left(\Sigma_r^{-1}\mu_r + \frac{1}{\sigma_y^2} H^\top y + \frac{1 - m_s}{\delta_s}(x_s - m_s y)\right). \tag{61}$$

Thus, for every fixed component r, the posterior of x_0 given $(x_s, y, C = r)$ is Gaussian with explicitly known mean and covariance.

(c) Posterior component probabilities

In this subsection, we derive

$$\gamma_{r|y}(y) \triangleq \mathbb{P}(C = r \mid y),$$

the posterior probability that component r generated the sample after observing the measurement y.

By Bayes' rule,

$$\mathbb{P}(C = r \mid y) = \frac{\mathbb{P}(C = r)\, p(y \mid C = r)}{p(y)}. \tag{62}$$

We already know that $\mathbb{P}(C = r) = \pi_r$, so it remains to compute $p(y \mid C = r)$.

Conditioned on $C = r$, we have

$$x_0 \sim \mathcal{N}(\mu_r, \Sigma_r), \qquad y = Hx_0 + n, \qquad n \sim \mathcal{N}(0, \sigma_y^2 I).$$

By an affine-Gaussian marginal identity, it follows that

$$p(y \mid C = r) = \mathcal{N}\big(y;\, H\mu_r,\ H\Sigma_r H^\top + \sigma_y^2 I\big). \tag{63}$$

Substituting (63) into (62), we obtain

$$\gamma_{r|y}(y) \triangleq \mathbb{P}(C = r \mid y) = \frac{\pi_r \mathcal{N}\big(y;\, H\mu_r,\ H\Sigma_r H^\top + \sigma_y^2 I\big)}{\sum_{j=1}^{R} \pi_j \mathcal{N}\big(y;\, H\mu_j,\ H\Sigma_j H^\top + \sigma_y^2 I\big)}, \qquad \sum_{r=1}^{R} \gamma_{r|y}(y) = 1. \tag{64}$$

(d) Marginal likelihood

In this subsection, we derive

$$p(x_s \mid y, C = r),$$

the conditional marginal likelihood of the bridge variable x_s after observing y and fixing the component index r.

From part (a), conditioned on $(y, C = r)$ we have

$$x_0 \sim \mathcal{N}(\mu_{r|y}, \Sigma_{r|y}),$$

and the forward process is

$$x_s = (1 - m_s)x_0 + m_s y + \sqrt{\delta_s}\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I).$$

Again using an affine-Gaussian marginal identity, we obtain

$$p(x_s \mid y, C = r) = \mathcal{N}\big(x_s;\, (1 - m_s)\mu_{r|y} + m_s y,\ (1 - m_s)^2 \Sigma_{r|y} + \delta_s I\big). \tag{65}$$

(e) Posterior component probabilities

In this subsection, we derive

$$\gamma_{r|s}(x_s, y) \triangleq \mathbb{P}(C = r \mid x_s, y),$$

the posterior probability that component r generated the sample after observing both y and x_s.

By Bayes' rule,

$$\mathbb{P}(C = r \mid x_s, y) = \frac{\mathbb{P}(C = r, x_s \mid y)}{p(x_s \mid y)}. \tag{66}$$

Using the product rule on the numerator,

$$\mathbb{P}(C = r, x_s \mid y) = \mathbb{P}(C = r \mid y)\, p(x_s \mid y, C = r).$$

Therefore,

$$\mathbb{P}(C = r \mid x_s, y) = \frac{\mathbb{P}(C = r \mid y)\, p(x_s \mid y, C = r)}{\sum_{j=1}^{R} \mathbb{P}(C = j \mid y)\, p(x_s \mid y, C = j)}. \tag{67}$$

Substituting the expressions from parts (c) and (d), we obtain

$$\gamma_{r|s}(x_s, y) = \frac{\gamma_{r|y}(y)\, \mathcal{N}\big(x_s;\, (1 - m_s)\mu_{r|y} + m_s y,\ (1 - m_s)^2 \Sigma_{r|y} + \delta_s I\big)}{\sum_{j=1}^{R} \gamma_{j|y}(y)\, \mathcal{N}\big(x_s;\, (1 - m_s)\mu_{j|y} + m_s y,\ (1 - m_s)^2 \Sigma_{j|y} + \delta_s I\big)}, \qquad \sum_{r=1}^{R} \gamma_{r|s}(x_s, y) = 1. \tag{68}$$

Thus, the role of x_s is to refine the measurement-only probabilities $\gamma_{r|y}(y)$ into the posterior responsibilities $\gamma_{r|s}(x_s, y)$.

(f) Exact posterior and MMSE-optimal denoiser

We now combine the component-conditioned posteriors from part (b) with the posterior responsibilities from part (e) to obtain the full posterior $p(x_0 \mid x_s, y)$ and its posterior mean. The formula for the full posterior will be:

$$p(x_0 \mid x_s, y) = \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \mathcal{N}\big(x_0; \mu_{r|s}(x_s, y), \Sigma_{r|s}\big). \tag{69}$$

Thus, the exact posterior remains a Mixture-of-Gaussians, with component means $\mu_{r|s}(x_s, y)$, component covariances $\Sigma_{r|s}$, and weights $\gamma_{r|s}(x_s, y)$.

Finally, since the Bayes-optimal denoiser under squared loss is the posterior mean, we obtain

$$\hat{x}_{0,s}^{\mathrm{MMSE}}(x_s, y) = \mathbb{E}[x_0 \mid x_s, y] = \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \mu_{r|s}(x_s, y). \tag{70}$$

C Full derivation of the oracle-induced reverse update

This appendix expands the algebra behind Section 4. The goal is to start from the exact posterior mean derived in Appendix B, substitute it into the BBDM sampler, and collect the resulting coefficients of x_s, y, and the constant term.

Using the notation of Section 2, the BBDM reverse update for $1 \le s \le S-1$ is

$$x_{s-1} = a_s \hat{x}_0 + b_s y + c_s x_s + \sigma_s z_s, \qquad z_s \sim \mathcal{N}(0, I), \tag{71}$$

where

$$a_s = (1 - m_{s-1}) - (1 - m_s)\sqrt{\frac{\delta_{s-1} - \sigma_s^2}{\delta_s}}, \qquad b_s = m_{s-1} - m_s\sqrt{\frac{\delta_{s-1} - \sigma_s^2}{\delta_s}}, \qquad c_s = \sqrt{\frac{\delta_{s-1} - \sigma_s^2}{\delta_s}}. \tag{72}$$

From Appendix B, the exact MMSE denoiser is

$$\hat{x}_{0,s}^{\mathrm{MMSE}}(x_s, y) = \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \mu_{r|s}(x_s, y), \tag{73}$$

with component-conditioned posterior mean

$$\mu_{r|s}(x_s, y) = \Sigma_{r|s}\left(\Sigma_r^{-1}\mu_r + \frac{1}{\sigma_y^2}H^\top y + \frac{1 - m_s}{\delta_s}(x_s - m_s y)\right). \tag{74}$$

Substituting (73) into (71) gives

$$x_{s-1} = a_s \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \mu_{r|s}(x_s, y) + b_s y + c_s x_s + \sigma_s z_s. \tag{75}$$

Substituting (74) into (75) yields

$$\begin{aligned} x_{s-1} &= a_s \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\Sigma_{r|s}\left(\Sigma_r^{-1}\mu_r + \frac{1}{\sigma_y^2}H^\top y + \frac{1 - m_s}{\delta_s}(x_s - m_s y)\right) + b_s y + c_s x_s + \sigma_s z_s \\ &= a_s \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s}\Sigma_r^{-1}\mu_r + \frac{a_s}{\sigma_y^2}\sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s}H^\top y \\ &\quad + a_s\frac{1 - m_s}{\delta_s}\sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s}(x_s - m_s y) + b_s y + c_s x_s + \sigma_s z_s. \end{aligned} \tag{76}$$

Next expand the term containing $(x_s - m_s y)$:

$$\begin{aligned} x_{s-1} = a_s \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s}\Sigma_r^{-1}\mu_r + \frac{a_s}{\sigma_y^2}\sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s}H^\top y \\ + a_s\frac{1 - m_s}{\delta_s}\sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s}x_s - a_s\frac{m_s(1 - m_s)}{\delta_s}\sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s}y \\ + b_s y + c_s x_s + \sigma_s z_s. \end{aligned} \tag{77}$$

Finally collect the coefficients of x_s and y:

$$
\begin{aligned}
x_{s-1} = &\left[c_s I + a_s \frac{1-m_s}{\delta_s} \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s}\right] x_s \\
&+ \left[b_s I + \frac{a_s}{\sigma_y^2} \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s} H^\top - a_s \frac{m_s(1-m_s)}{\delta_s} \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s}\right] y \\
&+ a_s \sum_{r=1}^{R} \gamma_{r|s}(x_s, y)\, \Sigma_{r|s} \Sigma_r^{-1} \mu_r + \sigma_s z_s, \qquad z_s \sim \mathcal{N}(0, I).
\end{aligned}
\tag{78}
$$

D Selected-label approximation, problem-adapted basis, and the induced Gaussian law

This appendix derives the fixed-label reverse law used in Section 5. It begins from the frozen-label BBDM update, moves to the component-wise basis in which the dynamics become diagonal, and writes the explicit Gaussian law induced by the selected-label approximation. It ends with proving the mean exactness and covariance deficit claims in 5.4 and 5.5.

D.1 Fixed-label reverse update

The selected-label approximation draws an auxiliary label $J \in \{1, \dots, R\}$ from the measurement posterior,

$$\mathbb{P}(J = r \mid y) = \gamma_{r|y}(y), \tag{79}$$

and then keeps that label fixed along the reverse chain.

Condition on $J = r$. The approximation uses the component-conditioned posterior mean

$$\mu_{r|s}(x_s, y) = \Sigma_{r|s} \left(\Sigma_{r|y}^{-1} \mu_{r|y} + \frac{1 - m_s}{\delta_s} (x_s - m_s y) \right), \tag{80}$$

with

$$\Sigma_{r|s} = \left(\Sigma_{r|y}^{-1} + \frac{(1 - m_s)^2}{\delta_s} I \right)^{-1}. \tag{81}$$

At $s = S$, we use $\hat{x}_0 = \mu_{r|y}$. The formulas below are for $1 \le s \le S - 1$:

Insert $\hat{x}_0 = \mu_{r|s}(x_s, y)$ into the BBDM update

$$x_{s-1} = a_s \hat{x}_0 + b_s y + c_s x_s + \sigma_s z_s, \qquad z_s \sim \mathcal{N}(0, I). \tag{82}$$

This gives

$$\begin{aligned} x_{s-1} &= a_s \Sigma_{r|s} \left(\Sigma_{r|y}^{-1} \mu_{r|y} + \frac{1 - m_s}{\delta_s} (x_s - m_s y) \right) + b_s y + c_s x_s + \sigma_s z_s \\ &= a_s \Sigma_{r|s} \Sigma_{r|y}^{-1} \mu_{r|y} + a_s \frac{1 - m_s}{\delta_s} \Sigma_{r|s} x_s - a_s \frac{m_s (1 - m_s)}{\delta_s} \Sigma_{r|s} y + b_s y + c_s x_s + \sigma_s z_s. \end{aligned} \tag{83}$$

Collecting the x_s and y terms yields

$$\begin{aligned} x_{s-1} = &\left(c_s I + a_s \frac{1 - m_s}{\delta_s} \Sigma_{r|s} \right) x_s \\ &+ \left(b_s I - a_s \frac{m_s (1 - m_s)}{\delta_s} \Sigma_{r|s} \right) y \\ &+ a_s \Sigma_{r|s} \Sigma_{r|y}^{-1} \mu_{r|y} + \sigma_s z_s, \qquad z_s \sim \mathcal{N}(0, I). \end{aligned} \tag{84}$$

Thus, once the label is frozen, the reverse update is affine in x_s and y.

D.2 Component-wise problem-adapted basis

Fix a component r. Since $\Sigma_{r|y}^{-1}$ is symmetric positive definite, there exist an orthogonal matrix U_r and positive scalars $\lambda_{r,1}, \dots, \lambda_{r,d}$ such that

$$\Sigma_{r|y}^{-1} = U_r \operatorname{diag}(\lambda_{r,1}, \dots, \lambda_{r,d}) U_r^\top. \tag{85}$$

Define the projected variables

$$\begin{aligned} x_s^{U_r} &\triangleq U_r^\top x_s, \qquad & y^{U_r} &\triangleq U_r^\top y, \\ \mu_{r|y}^{U_r} &\triangleq U_r^\top \mu_{r|y}, \qquad & z_s^{U_r} &\triangleq U_r^\top z_s. \end{aligned} \tag{86}$$

Because U_r is orthogonal and $z_s \sim \mathcal{N}(0, I)$, we also have $z_s^{U_r} \sim \mathcal{N}(0, I)$.

For each step $1 \leq s \leq S-1$, define

$$\Lambda_{r|s} \triangleq \operatorname{diag}\left(\frac{1}{\lambda_{r,1} + \frac{(1-m_s)^2}{\delta_s}}, \ldots, \frac{1}{\lambda_{r,d} + \frac{(1-m_s)^2}{\delta_s}}\right). \tag{87}$$

Then

$$\Sigma_{r|s} = \left(\Sigma_{r|y}^{-1} + \frac{(1-m_s)^2}{\delta_s} I\right)^{-1} = U_r \Lambda_{r|s} U_r^\top. \tag{88}$$

Hence every $\Sigma_{r|s}$ is diagonal in the same basis U_r.

D.3 Reverse recursion in the basis U_r

Project (84) onto U_r:

$$\begin{aligned} x_{s-1}^{U_r} = &\left(c_s I + a_s \frac{1-m_s}{\delta_s} \Lambda_{r|s}\right) x_s^{U_r} \\ &+ \left(b_s I - a_s \frac{m_s(1-m_s)}{\delta_s} \Lambda_{r|s}\right) y^{U_r} \\ &+ a_s \Lambda_{r|s} \operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d}) \mu_{r|y}^{U_r} + \sigma_s z_s^{U_r}. \end{aligned} \tag{89}$$

For $1 \leq s \leq S-1$, define the diagonal matrices

$$G_r(s) \triangleq c_s I + \frac{a_s(1-m_s)}{\delta_s} \Lambda_{r|s}, \tag{90}$$

$$N_r(s) \triangleq b_s I - \frac{a_s m_s (1-m_s)}{\delta_s} \Lambda_{r|s}, \tag{91}$$

$$M_r(s) \triangleq a_s \Lambda_{r|s} \operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d}). \tag{92}$$

Then

$$x_{s-1}^{U_r} = G_r(s) x_s^{U_r} + N_r(s) y^{U_r} + M_r(s) \mu_{r|y}^{U_r} + \sigma_s z_s^{U_r}. \tag{93}$$

Since $G_r(s)$, $N_r(s)$, and $M_r(s)$ are diagonal, the recursion decouples coordinate-wise. For coordinate k,

$$x_{s-1,k}^{U_r} = g_{r,k}(s) x_{s,k}^{U_r} + n_{r,k}(s) y_k^{U_r} + m_{r,k}(s) \mu_{r|y,k}^{U_r} + \sigma_s z_{s,k}^{U_r}, \tag{94}$$

where $g_{r,k}(s)$, $n_{r,k}(s)$, and $m_{r,k}(s)$ are the diagonal entries of $G_r(s)$, $N_r(s)$, and $M_r(s)$. For $s = S$, set

$$G_r(S) = 0, \qquad N_r(S) = m_{S-1} I, \qquad M_r(S) = (1 - m_{S-1}) I, \qquad \sigma_S^2 = \delta_{S-1}.$$

Then (93) holds for all $s = 1, \ldots, S$.

D.4 Unrolling the fixed-label recursion

Repeated substitution of (93) using the endpoint convention at $s = S$, yields, for any step $\ell \in \{0, \ldots, S-1\}$,

$$\begin{aligned} x_\ell^{U_r} = &\left[\prod_{s'=\ell+1}^{S} G_r(s')\right] x_S^{U_r} + \sum_{i=\ell+1}^{S} \left(\prod_{j=\ell+1}^{i-1} G_r(j)\right) N_r(i)\, y^{U_r} \\ &+ \sum_{i=\ell+1}^{S} \left(\prod_{j=\ell+1}^{i-1} G_r(j)\right) M_r(i)\, \mu_{r|y}^{U_r} + \sum_{i=\ell+1}^{S} \left(\prod_{j=\ell+1}^{i-1} G_r(j)\right) \sigma_i z_i^{U_r}. \end{aligned} \tag{95}$$

Here the empty product is understood as the identity. In BBDM inference, the terminal latent is initialized from the conditioning signal, so $x_S = y$ and therefore $x_S^{U_r} = y^{U_r}$. Setting $\ell = 0$ in (95) gives

$$\hat{x}_0^{U_r,(r)} = D_1^{(r)} y^{U_r} + D_2^{(r)} \mu_{r|y}^{U_r} + \sum_{i=1}^{S} \left(\prod_{j=1}^{i-1} G_r(j) \right) \sigma_i z_i^{U_r}, \tag{96}$$

where

$$D_1^{(r)} \triangleq \prod_{s'=1}^{S} G_r(s') + \sum_{i=1}^{S} \left(\prod_{j=1}^{i-1} G_r(j) \right) N_r(i), \tag{97}$$

and

$$D_2^{(r)} \triangleq \sum_{i=1}^{S} \left(\prod_{j=1}^{i-1} G_r(j) \right) M_r(i). \tag{98}$$

Because $G_r(s)$, $N_r(s)$, and $M_r(s)$ are diagonal, both $D_1^{(r)}$ and $D_2^{(r)}$ are diagonal as well.

D.5 The selected-label Gaussian law

Conditioned on $(y, J = r)$, the only randomness in (96) comes from the reverse Gaussian innovations. Hence

$$p_{\text{Select}}(\hat{x}_0^{U_r,(r)} \mid y, J = r) = \mathcal{N}\big(\mu_{\text{BBDM}}^{U_r,(r)}(y), \Sigma_{\text{BBDM}}^{U_r,(r)}\big), \tag{99}$$

with mean

$$\mu_{\text{BBDM}}^{U_r,(r)}(y) = D_1^{(r)} y^{U_r} + D_2^{(r)} \mu_{r|y}^{U_r}, \tag{100}$$

and covariance

$$\Sigma_{\text{BBDM}}^{U_r,(r)} = \sum_{i=1}^{S} \left(\prod_{j=1}^{i-1} G_r(j) \right) \sigma_i^2 I \left(\prod_{j=1}^{i-1} G_r(j) \right)^{\top}. \tag{101}$$

Returning to the original coordinates,

$$p_{\text{Select}}(\hat{x}_0 \mid y, J = r) = \mathcal{N}\big(\hat{x}_0; \mu_{\text{BBDM}}^{(r)}(y), \Sigma_{\text{BBDM}}^{(r)}\big), \tag{102}$$

where

$$\mu_{\text{BBDM}}^{(r)}(y) = U_r \mu_{\text{BBDM}}^{U_r,(r)}(y), \qquad \Sigma_{\text{BBDM}}^{(r)} = U_r \Sigma_{\text{BBDM}}^{U_r,(r)} U_r^{\top}. \tag{103}$$

Finally, averaging over the selected label gives

$$p_{\text{Select}}(\hat{x}_0 \mid y) = \sum_{r=1}^{R} \gamma_{r|y}(y) \mathcal{N}\big(\hat{x}_0; \mu_{\text{BBDM}}^{(r)}(y), \Sigma_{\text{BBDM}}^{(r)}\big). \tag{104}$$

D.6 Mean exactness and covariance deficit

We now prove the two structural claims used in the main text. The proofs use only the fixed-label recursion derived above.

Proof of Proposition 5.4. Condition on $J = r$ and work in the basis U_r. Since the Gaussian innovations have zero mean, the conditional mean of the selected-label chain obeys the deterministic recursion

$$\bar{x}_{s-1}^{U_r} = G_r(s) \bar{x}_s^{U_r} + N_r(s) y^{U_r} + M_r(s) \mu_{r|y}^{U_r}, \tag{105}$$

where

$$\bar{x}_s^{U_r} \triangleq \mathbb{E}[x_s^{U_r} \mid y, J = r].$$

We prove by backward induction that, for every step s,

$$\bar{x}_s^{U_r} = m_s y^{U_r} + (1 - m_s)\mu_{r|y}^{U_r}. \tag{106}$$

The initialization holds at the terminal step because $x_S = y$ and $m_S \cong 1$, hence

$$\bar{x}_S^{U_r} = y^{U_r} \cong m_S y^{U_r} + (1 - m_S)\mu_{r|y}^{U_r}.$$

Now assume that (106) holds at step s. Substituting this induction hypothesis into (105) gives

$$\begin{aligned} \bar{x}_{s-1}^{U_r} &= G_r(s)\big(m_s y^{U_r} + (1 - m_s)\mu_{r|y}^{U_r}\big) + N_r(s) y^{U_r} + M_r(s)\mu_{r|y}^{U_r} \\ &= \big(G_r(s) m_s + N_r(s)\big) y^{U_r} + \big(G_r(s)(1 - m_s) + M_r(s)\big)\mu_{r|y}^{U_r}. \end{aligned} \tag{107}$$

For the endpoint $s = S$, the identities

$$G_r(s) m_s + N_r(s) = m_{s-1} I, \qquad G_r(s)(1 - m_s) + M_r(s) = (1 - m_{s-1}) I$$

follow immediately from the endpoint convention. For $1 \le s \le S - 1$, they follow from the algebra below. Using,

$$G_r(s) = c_s I + \frac{a_s(1 - m_s)}{\delta_s}\Lambda_{r|s}, \qquad N_r(s) = b_s I - \frac{a_s m_s(1 - m_s)}{\delta_s}\Lambda_{r|s},$$

we get

$$\begin{aligned} G_r(s) m_s + N_r(s) &= \left(c_s I + \frac{a_s(1 - m_s)}{\delta_s}\Lambda_{r|s}\right) m_s + b_s I - \frac{a_s m_s(1 - m_s)}{\delta_s}\Lambda_{r|s} \\ &= (c_s m_s + b_s) I + \frac{a_s m_s(1 - m_s)}{\delta_s}\Lambda_{r|s} - \frac{a_s m_s(1 - m_s)}{\delta_s}\Lambda_{r|s} \\ &= (c_s m_s + b_s) I \\ &= m_{s-1} I. \end{aligned} \tag{108}$$

The last equality follows from the definitions

$$b_s = m_{s-1} - m_s c_s, \qquad c_s = \sqrt{\frac{\delta_{s-1} - \sigma_s^2}{\delta_s}}.$$

Second, using

$$M_r(s) = a_s \Lambda_{r|s} \operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d}),$$

we have

$$\begin{aligned} G_r(s)(1 - m_s) + M_r(s) &= \left(c_s I + \frac{a_s(1 - m_s)}{\delta_s}\Lambda_{r|s}\right)(1 - m_s) + a_s \Lambda_{r|s} \operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d}) \\ &= c_s(1 - m_s) I + a_s \Lambda_{r|s}\left(\frac{(1 - m_s)^2}{\delta_s} I + \operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d})\right). \end{aligned} \tag{109}$$

By definition,

$$\Lambda_{r|s} = \left(\operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d}) + \frac{(1 - m_s)^2}{\delta_s} I\right)^{-1},$$

and therefore

$$\Lambda_{r|s}\left(\frac{(1 - m_s)^2}{\delta_s} I + \operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d})\right) = I.$$

Substituting this into (109) gives

$$G_r(s)(1 - m_s) + M_r(s) = c_s(1 - m_s) I + a_s I \quad = \big(c_s(1 - m_s) + a_s\big) I = (1 - m_{s-1}) I. \tag{110}$$

The last equality follows from the definition

$$a_s = (1 - m_{s-1}) - (1 - m_s)c_s.$$

Combining (107), (108), and (110), we obtain

$$\bar{x}_{s-1}^{U_r} = m_{s-1}y^{U_r} + (1 - m_{s-1})\mu_{r|y}^{U_r}.$$

Thus the induction is complete. At $s = 0$, since $m_0 = 0$,

$$\mathbb{E}[\hat{x}_0^{U_r,(r)} \mid y, J = r] = \mu_{r|y}^{U_r}.$$

Returning to the original coordinates gives

$$\mathbb{E}[\hat{x}_0 \mid y, J = r] = \mu_{r|y}.$$

Comparing this final mean with the deterministic part of (96) gives

$$D_1^{(r)} = 0, \qquad D_2^{(r)} = I.$$

Proof of Proposition 5.5. Condition on component r and work in the basis U_r. We compare two chains that both start from the same deterministic endpoint $x_S = y$. The first chain is the selected-label chain conditioned on $(y, J = r)$. This is the chain that uses the posterior mean $\mu_{r|s}(x_s, y)$ inside the BBDM reverse update. The second chain is the exact component-conditioned bridge reverse chain conditioned on $(y, C = r)$. At each reverse step, this exact chain uses the same BBDM reverse formula before replacing x_0 by its posterior mean. Thus, for one reverse transition, it samples x_0 from the appropriate conditional posterior and then applies the bridge update.

We first consider the terminal reverse update $s = S$. By the endpoint convention, the selected-label chain uses $\hat{x}_0 = \mu_{r|y}$. Therefore, after projection to the U_r-basis,

$$x_{S-1}^{U_r,\mathrm{sel}} = m_{S-1}y^{U_r} + (1 - m_{S-1})\mu_{r|y}^{U_r} + \sqrt{\delta_{S-1}}z_S^{U_r}.$$

Conditioned on $(y, J = r)$, the first two terms are deterministic and $z_S^{U_r} \sim \mathcal{N}(0, I)$. Hence

$$\mathrm{Cov}(x_{S-1}^{U_r,\mathrm{sel}} \mid y, J = r) = \delta_{S-1}I.$$

For the exact component-conditioned chain, the terminal update uses a posterior sample

$$x_0 \mid y, C = r \sim \mathcal{N}(\mu_{r|y}, \Sigma_{r|y}).$$

Thus

$$x_{S-1}^{U_r,\mathrm{ex}} = m_{S-1}y^{U_r} + (1 - m_{S-1})x_0^{U_r} + \sqrt{\delta_{S-1}}z_S^{U_r},$$

where $z_S^{U_r} \sim \mathcal{N}(0, I)$ is independent of $x_0^{U_r}$. Again, the term $m_{S-1}y^{U_r}$ is deterministic after conditioning on y, so

$$\mathrm{Cov}(x_{S-1}^{U_r,\mathrm{ex}} \mid y, C = r) = (1 - m_{S-1})^2 \,\mathrm{Cov}(x_0^{U_r} \mid y, C = r) + \delta_{S-1}I = (1 - m_{S-1})^2 U_r^\top \Sigma_{r|y} U_r + \delta_{S-1}I.$$

Since

$$U_r^\top \Sigma_{r|y} U_r = \mathrm{diag}\left(\frac{1}{\lambda_{r,1}}, \ldots, \frac{1}{\lambda_{r,d}}\right),$$

the covariance gap at step $S - 1$ is

$$\begin{aligned}\Delta_{S-1} &\triangleq \mathrm{Cov}(x_{S-1}^{U_r,\mathrm{ex}} \mid y, C = r) - \mathrm{Cov}(x_{S-1}^{U_r,\mathrm{sel}} \mid y, J = r) \\ &= (1 - m_{S-1})^2 \,\mathrm{diag}\left(\frac{1}{\lambda_{r,1}}, \ldots, \frac{1}{\lambda_{r,d}}\right) \succeq 0.\end{aligned}$$

We now show that this covariance ordering propagates through every interior reverse step. Fix $1 \le s \le S-1$. For the selected-label chain, we already have the projected recursion

$$x_{s-1}^{U_r,\mathrm{sel}} = G_r(s)x_s^{U_r,\mathrm{sel}} + N_r(s)y^{U_r} + M_r(s)\mu_{r|y}^{U_r} + \sigma_s z_s^{U_r}.$$

Conditioned on $(y, J=r)$, the quantities y^{U_r} and $\mu_{r|y}^{U_r}$ are deterministic. Therefore taking covariance gives

$$\mathrm{Cov}(x_{s-1}^{U_r,\mathrm{sel}} \mid y, J=r) = G_r(s)\,\mathrm{Cov}(x_s^{U_r,\mathrm{sel}} \mid y, J=r)G_r(s)^\top + \sigma_s^2 I.$$

For the exact component-conditioned chain, one reverse transition has the form

$$x_{s-1} = a_s x_0 + b_s y + c_s x_s + \sigma_s z_s.$$

After projection to the U_r-basis,

$$x_{s-1}^{U_r,\mathrm{ex}} = a_s x_0^{U_r} + b_s y^{U_r} + c_s x_s^{U_r,\mathrm{ex}} + \sigma_s z_s^{U_r}.$$

For $1 \le s \le S-1$, the conditional posterior of x_0 is

$$x_0 \mid x_s, y, C=r \sim \mathcal{N}(\mu_{r|s}(x_s,y), \Sigma_{r|s}),$$

and in the U_r-basis,

$$U_r^\top \Sigma_{r|s} U_r = \Lambda_{r|s}.$$

Moreover,

$$U_r^\top \mu_{r|s}(x_s,y) = \Lambda_{r|s}\left(\mathrm{diag}(\lambda_{r,1},\dots,\lambda_{r,d})\mu_{r|y}^{U_r} + \frac{1-m_s}{\delta_s}(x_s^{U_r} - m_s y^{U_r})\right).$$

Therefore, conditioned on $x_s^{U_r,\mathrm{ex}}$, y, and $C=r$,

$$\begin{aligned}
&\mathbb{E}[x_{s-1}^{U_r,\mathrm{ex}} \mid x_s^{U_r,\mathrm{ex}}, y, C=r] \\
&= a_s U_r^\top \mu_{r|s}(x_s^{\mathrm{ex}}, y) + b_s y^{U_r} + c_s x_s^{U_r,\mathrm{ex}} \\
&= \left(c_s I + \frac{a_s(1-m_s)}{\delta_s}\Lambda_{r|s}\right) x_s^{U_r,\mathrm{ex}} + \left(b_s I - \frac{a_s m_s(1-m_s)}{\delta_s}\Lambda_{r|s}\right) y^{U_r} \\
&\qquad + a_s \Lambda_{r|s}\,\mathrm{diag}(\lambda_{r,1},\dots,\lambda_{r,d})\mu_{r|y}^{U_r} \\
&= G_r(s)x_s^{U_r,\mathrm{ex}} + N_r(s)y^{U_r} + M_r(s)\mu_{r|y}^{U_r}.
\end{aligned}$$

Thus the exact chain and the selected-label chain have the same affine conditional mean. The difference is in the conditional covariance. Given $x_s^{U_r,\mathrm{ex}}$, y, and $C=r$, the only random terms in

$$x_{s-1}^{U_r,\mathrm{ex}} = a_s x_0^{U_r} + b_s y^{U_r} + c_s x_s^{U_r,\mathrm{ex}} + \sigma_s z_s^{U_r}$$

are $x_0^{U_r}$ and $z_s^{U_r}$. These two are independent, and

$$\mathrm{Cov}(x_0^{U_r} \mid x_s^{U_r,\mathrm{ex}}, y, C=r) = \Lambda_{r|s}.$$

Hence

$$\mathrm{Cov}(x_{s-1}^{U_r,\mathrm{ex}} \mid x_s^{U_r,\mathrm{ex}}, y, C=r) = a_s^2\Lambda_{r|s} + \sigma_s^2 I.$$

Now apply the law of total covariance:

$$\begin{aligned}
&\mathrm{Cov}(x_{s-1}^{U_r,\mathrm{ex}} \mid y, C=r) \\
&= \mathrm{Cov}\Big(\mathbb{E}[x_{s-1}^{U_r,\mathrm{ex}} \mid x_s^{U_r,\mathrm{ex}}, y, C=r] \mid y, C=r\Big) \\
&\qquad + \mathbb{E}\Big[\mathrm{Cov}(x_{s-1}^{U_r,\mathrm{ex}} \mid x_s^{U_r,\mathrm{ex}}, y, C=r) \mid y, C=r\Big].
\end{aligned}$$

Using the conditional mean and conditional covariance computed above gives

$$\mathrm{Cov}(x_{s-1}^{U_r,\mathrm{ex}} \mid y, C=r) = G_r(s)\,\mathrm{Cov}(x_s^{U_r,\mathrm{ex}} \mid y, C=r)G_r(s)^\top + \sigma_s^2 I + a_s^2 \Lambda_{r|s}.$$

We can now subtract the selected covariance recursion from the exact covariance recursion. For $s = 0, \ldots, S-1$, define

$$\Delta_s \triangleq \mathrm{Cov}(x_s^{U_r,\mathrm{ex}} \mid y, C=r) - \mathrm{Cov}(x_s^{U_r,\mathrm{sel}} \mid y, J=r).$$

For $1 \le s \le S-1$, the subtraction gives

$$\Delta_{s-1} = G_r(s)\Delta_s G_r(s)^\top + a_s^2 \Lambda_{r|s}.$$

If $\Delta_s \succeq 0$, then

$$G_r(s)\Delta_s G_r(s)^\top \succeq 0,$$

and since $\Lambda_{r|s} \succeq 0$, also

$$a_s^2 \Lambda_{r|s} \succeq 0.$$

Therefore

$$\Delta_s \succeq 0 \quad \Longrightarrow \quad \Delta_{s-1} \succeq 0.$$

We already proved that $\Delta_{S-1} \succeq 0$. Applying the implication successively for $s = S-1, S-2, \ldots, 1$ yields

$$\Delta_s \succeq 0, \qquad s = 0, \ldots, S-1.$$

It remains only to identify what $\Delta_0 \succeq 0$ says. The exact component-conditioned chain is the reverse-time conditional chain of the same bridge model after conditioning on y and $C = r$. Therefore its clean endpoint has marginal

$$x_0 \mid y, C=r \sim \mathcal{N}(\mu_{r|y}, \Sigma_{r|y}).$$

Consequently,

$$\mathrm{Cov}(x_0^{U_r,\mathrm{ex}} \mid y, C=r) = U_r^\top \Sigma_{r|y} U_r.$$

At the same time, the selected-label chain endpoint is exactly $\hat{x}_0^{U_r,(r)}$, whose covariance conditioned on $(y, J=r)$ is

$$\mathrm{Cov}(\hat{x}_0^{U_r,(r)} \mid y, J=r) = U_r^\top \Sigma_{\mathrm{BBDM}}^{(r)} U_r.$$

Thus $\Delta_0 \succeq 0$ gives

$$U_r^\top \Sigma_{r|y} U_r - U_r^\top \Sigma_{\mathrm{BBDM}}^{(r)} U_r \succeq 0.$$

Multiplying by U_r and $U_r^\top$ gives

$$\Sigma_{r|y} - \Sigma_{\mathrm{BBDM}}^{(r)} \succeq 0.$$

Finally, both covariances are diagonal in the U_r-basis:

$$U_r^\top \Sigma_{r|y} U_r = \mathrm{diag}\left(\frac{1}{\lambda_{r,1}}, \ldots, \frac{1}{\lambda_{r,d}}\right),$$

and

$$U_r^\top \Sigma_{\mathrm{BBDM}}^{(r)} U_r = \mathrm{diag}\big(\sigma_{\mathrm{BBDM},r,1}^2, \ldots, \sigma_{\mathrm{BBDM},r,d}^2\big).$$

The positive-semidefinite comparison therefore implies, coordinate by coordinate,

$$0 \le \sigma_{\mathrm{BBDM},r,k}^2 \le \frac{1}{\lambda_{r,k}}, \qquad k = 1, \ldots, d.$$

This proves the proposition.

E Schedule objectives and bounded schedule family

This appendix derives the two schedule objectives used in Section 6 and makes explicit the coupling choices behind the objectives. It also explains why the scheduling parameterization that is offered is valid under the bridge constraints.

E.1 True and selected laws and the matched-index coupling

From Appendix B, the true measurement posterior is

$$p(x_0 \mid y) = \sum_{r=1}^{R} \gamma_{r|y}(y)\, \mathcal{N}\big(x_0; \mu_{r|y}, \Sigma_{r|y}\big). \tag{111}$$

The selected-label law is

$$p_{\text{Select}}(\hat{x}_0 \mid y) = \sum_{r=1}^{R} \gamma_{r|y}(y) \mathcal{N}\big(\hat{x}_0; \mu^{(r)}_{\text{BBDM}}(y), \Sigma^{(r)}_{\text{BBDM}}\big). \tag{112}$$

Thus the two laws have the same mixture weights. We couple them by first drawing a common index

$$R_y \sim \text{Categorical}(\gamma_{1|y}(y), \ldots, \gamma_{R|y}(y)),$$

and then, conditioned on $R_y = r$, coupling the two Gaussian components. This gives the valid transport plan. Therefore, from the definition of the W2 distance:

$$\begin{aligned} &W_2^2(p(x_0 \mid y), p_{\text{Select}}(\hat{x}_0 \mid y)) \\ &\quad \le \sum_{r=1}^{R} \gamma_{r|y}(y) W_2^2\Big(\mathcal{N}(\mu_{r|y}, \Sigma_{r|y}), \mathcal{N}(\mu^{(r)}_{\text{BBDM}}(y), \Sigma^{(r)}_{\text{BBDM}})\Big). \end{aligned} \tag{113}$$

The inequality may be strict because the optimal coupling between the two full mixtures is allowed to split mass across different component labels.

E.2 Gaussian W_2 formula and cancellation of the mean term

For two Gaussian laws

$$P_1 = \mathcal{N}(\mu_1, \Sigma_1), \qquad P_2 = \mathcal{N}(\mu_2, \Sigma_2),$$

the squared 2-Wasserstein distance has the standard closed form (Dowson & Landau, 1982; Peyré & Cuturi, 2019):

$$W_2^2(P_1, P_2) = \|\mu_1 - \mu_2\|_2^2 + \operatorname{tr}\left(\Sigma_1 + \Sigma_2 - 2\left(\Sigma_2^{1/2}\Sigma_1\Sigma_2^{1/2}\right)^{1/2}\right). \tag{114}$$

For the matched component r, Proposition 5.4 gives

$$\mu^{(r)}_{\text{BBDM}}(y) = \mu_{r|y}, \qquad \|\mu_{r|y} - \mu^{(r)}_{\text{BBDM}}(y)\|_2^2 = 0. \tag{115}$$

It remains only to compare the covariances. In the basis U_r,

$$U_r^\top \Sigma_{r|y} U_r = \operatorname{diag}\left(\frac{1}{\lambda_{r,1}}, \ldots, \frac{1}{\lambda_{r,d}}\right), \tag{116}$$

whereas

$$U_r^\top \Sigma^{(r)}_{\text{BBDM}} U_r = \operatorname{diag}\big(\sigma^2_{\text{BBDM},r,1}, \ldots, \sigma^2_{\text{BBDM},r,d}\big). \tag{117}$$

Because the two covariance matrices are diagonal in the same basis, the covariance part of (114) reduces coordinate-wise. Hence

$$W_2^2\Big(\mathcal{N}(\mu_{r|y}, \Sigma_{r|y}), \mathcal{N}(\mu^{(r)}_{\text{BBDM}}(y), \Sigma^{(r)}_{\text{BBDM}})\Big) = \sum_{k=1}^{d} \left(\sigma_{\text{BBDM},r,k} - \frac{1}{\sqrt{\lambda_{r,k}}}\right)^2. \tag{118}$$

E.3 The Wasserstein upper-bound schedule objective

Average (118) over the marginal law of y and use

$$\gamma_{r|y}(y)p(y) = \pi_r p(y \mid C = r).$$

Since the covariance term does not depend on the realized value of y, the matched-component Wasserstein upper-bound objective reduces to

$$J_{\mathrm{MOG}}^{W_2} = \sum_{r=1}^{R} \pi_r \sum_{k=1}^{d} \left(\sigma_{\mathrm{BBDM},r,k} - \frac{1}{\sqrt{\lambda_{r,k}}} \right)^2 . \tag{119}$$

This is the expression stated in Corollary 6.4.

E.4 The expected-MSE schedule objective and its coupling

Fix y. First draw a label

$$R_y \sim \gamma_{\cdot|y}(y).$$

Then, conditioned on $R_y = r$, draw

$$X \sim \mathcal{N}(\mu_{r|y}, \Sigma_{r|y}), \qquad \hat{X} \sim \mathcal{N}(\mu_{\mathrm{BBDM}}^{(r)}(y), \Sigma_{\mathrm{BBDM}}^{(r)}),$$

independently. Thus the two samples use the same mixture component, but are independent inside that component.

For a fixed matched component r,

$$\begin{aligned} \mathbb{E}\big[\|X - \hat{X}\|_2^2 \mid y, R_y = r\big] &= \mathbb{E}\big[\|(X - \mu_{r|y}) - (\hat{X} - \mu_{\mathrm{BBDM}}^{(r)}(y)) \\ &\qquad + (\mu_{r|y} - \mu_{\mathrm{BBDM}}^{(r)}(y))\|_2^2\big] \\ &= \mathrm{tr}(\Sigma_{r|y}) + \mathrm{tr}(\Sigma_{\mathrm{BBDM}}^{(r)}) + \|\mu_{r|y} - \mu_{\mathrm{BBDM}}^{(r)}(y)\|_2^2. \end{aligned} \tag{120}$$

The cross terms vanish because the two centered Gaussian residuals have zero mean and are sampled independently. By Proposition 5.4,

$$\mu_{\mathrm{BBDM}}^{(r)}(y) = \mu_{r|y},$$

so the mean term in (120) cancels. Hence

$$\mathbb{E}\big[\|X - \hat{X}\|_2^2 \mid y, R_y = r\big] = \mathrm{tr}(\Sigma_{r|y}) + \mathrm{tr}(\Sigma_{\mathrm{BBDM}}^{(r)}). \tag{121}$$

In the basis U_r,

$$U_r^\top \Sigma_{r|y} U_r = \mathrm{diag}\left(\frac{1}{\lambda_{r,1}}, \ldots, \frac{1}{\lambda_{r,d}} \right),$$

and

$$U_r^\top \Sigma_{\mathrm{BBDM}}^{(r)} U_r = \mathrm{diag}\big(\sigma_{\mathrm{BBDM},r,1}^2, \ldots, \sigma_{\mathrm{BBDM},r,d}^2\big) .$$

Since the trace is invariant under orthogonal changes of basis,

$$\mathbb{E}\big[\|X - \hat{X}\|_2^2 \mid y, R_y = r\big] = \sum_{k=1}^{d} \left(\frac{1}{\lambda_{r,k}} + \sigma_{\mathrm{BBDM},r,k}^2 \right). \tag{122}$$

Again Averaging (122) over the marginal law of y and using,

$$\gamma_{r|y}(y)p(y) = \pi_r p(y \mid C = r).$$

we get,

$$J_{\mathrm{MOG}}^{\mathrm{MSE}} = \sum_{r=1}^{R} \pi_r \sum_{k=1}^{d} \left(\sigma_{\mathrm{BBDM},r,k}^2 + \frac{1}{\lambda_{r,k}} \right). \tag{123}$$

It is important to distinguish this matched-label product coupling from an independent-label coupling. Suppose instead that, for fixed y, the true posterior sample uses a label C and the selected-label reconstruction uses a label J, where

$$C \sim \gamma_{\cdot|y}(y), \qquad J \sim \gamma_{\cdot|y}(y),$$

independently. Conditioned on $C = r$ and $J = j$,

$$\begin{aligned} \mathbb{E}\big[\|X - \hat{X}\|_2^2 \mid y, C = r, J = j\big] &= \mathrm{tr}(\Sigma_{r|y}) + \mathrm{tr}(\Sigma_{\mathrm{BBDM}}^{(j)}) \\ &\quad + \|\mu_{r|y} - \mu_{\mathrm{BBDM}}^{(j)}(y)\|_2^2. \end{aligned} \tag{124}$$

Using again $\mu_{\mathrm{BBDM}}^{(j)}(y) = \mu_{j|y}$, and averaging over the independent labels, gives

$$\begin{aligned} \mathbb{E}\|X - \hat{X}\|_2^2 = &\sum_{r=1}^{R} \gamma_{r|y}(y)\, \mathrm{tr}(\Sigma_{r|y}) + \sum_{j=1}^{R} \gamma_{j|y}(y)\, \mathrm{tr}(\Sigma_{\mathrm{BBDM}}^{(j)}) \\ &+ \sum_{r=1}^{R} \sum_{j=1}^{R} \gamma_{r|y}(y) \gamma_{j|y}(y) \|\mu_{r|y} - \mu_{j|y}\|_2^2. \end{aligned} \tag{125}$$

The matched-label MSE and the independent-label MSE generally have different absolute values, because the independent-label version includes an additional between-component mean mismatch, but they lead to the same schedule minimizer.

E.5 Formulation of the optimization problem and bounded parameter box

The schedule-design problem is to choose

$$\Theta = \{(m_s, \delta_s)\}_{s=1}^{S}$$

so as to minimize one of the two criteria,

$$\min_{\Theta} \mathcal{J}(\Theta), \qquad \mathcal{J} \in \{J_{\mathrm{MOG}}^{W_2}, J_{\mathrm{MOG}}^{\mathrm{MSE}}\}. \tag{126}$$

Because direct optimization over the discrete schedule values is inconvenient, we use the four-parameter family

$$m_s = 1 - (1 - \tau_s^{\alpha})^{\beta}, \qquad \delta_s = c\,[4 m_s (1 - m_s)]^{\gamma}, \qquad \tau_s = \frac{s}{S}. \tag{127}$$

A simple sufficient bridge-valid parameter region for this family is

$$\alpha > 0, \qquad \beta > 0, \qquad c > 0, \qquad 0 < \gamma \le 2.$$

In the experiments and schedule search we use the bounded regular subset

$$\begin{aligned} \alpha &\in [1, 2], \quad & \beta &\in [1, 2], \\ c &\in [0.2, 2], \quad & \gamma &\in [0.2, 2]. \end{aligned} \tag{128}$$

The bridge constraints are

1. $m_0 = 0$, $m_S = 1$, $\delta_0 = 0$, $\delta_S = 0$;
2. $0 < m_1 < \cdots < m_{S-1} < 1$;

3. $\delta_s \geq 0$;
4. $\delta_s - \delta_{s-1}\frac{(1-m_s)^2}{(1-m_{s-1})^2} \geq 0$.

At $\tau_0 = 0$, (127) gives $m_0 = 0$, hence $\delta_0 = 0$. At $\tau_S = 1$, it gives $m_S = 1$, hence $\delta_S = 0$. Moreover, for $\tau \in (0, 1)$,

$$\frac{dm}{d\tau} = \alpha\beta\,\tau^{\alpha-1}(1-\tau^{\alpha})^{\beta-1} > 0$$

whenever $\alpha > 0$ and $\beta > 0$. Hence m_s is strictly increasing and lies in $(0, 1)$ for all interior steps. Since $4m_s(1 - m_s) > 0$ there, we also have $\delta_s > 0$ for all interior steps whenever $c > 0$ and $\gamma > 0$.

It remains to verify the forward-transition constraint. The endpoint cases $s = 1$ and $s = S$ are automatic from $\delta_0 = 0$ and $(1 - m_S)^2 = 0$. For the remaining interior transitions, the constraint is equivalent to

$$\frac{\delta_s}{(1-m_s)^2} \geq \frac{\delta_{s-1}}{(1-m_{s-1})^2}.$$

Define

$$g(m) = \frac{\delta}{(1-m)^2} = c\,4^{\gamma}m^{\gamma}(1-m)^{\gamma-2}.$$

Differentiating with respect to m gives

$$g'(m) = c\,4^{\gamma}m^{\gamma-1}(1-m)^{\gamma-3}\big[\gamma - 2m(\gamma-1)\big].$$

For $m \in (0, 1)$, all factors outside the bracket are positive. If $0 < \gamma \leq 1$, the bracket is nonnegative for all $m \in (0, 1)$. If $\gamma > 1$, its minimum over $(0, 1)$ is approached as $m \to 1$, where it tends to $2 - \gamma$. Hence $0 < \gamma \leq 2$ guarantees $g'(m) \geq 0$ for all $m \in (0, 1)$. Since m_s is increasing, this implies $g(m_s) \geq g(m_{s-1})$, which gives the forward-transition validity condition.

Thus the family (127) satisfies the Brownian-bridge constraints throughout the sufficient validity region

$$\alpha > 0, \qquad \beta > 0, \qquad c > 0, \qquad 0 < \gamma \leq 2.$$

The box in (128) is a bounded, regular subset of this region. The bounds $\alpha, \beta \in [1, 2]$ are not imposed by the discrete bridge constraints themselves. They avoid endpoint-singular slopes of the interpolation curve $m(\tau)$ and keep the search in a moderate range. The bounds on c and γ keep the bridge variance non-degenerate, while the upper bound $\gamma \leq 2$ preserves the sufficient transition condition above.

The need for a bounded search box is also motivated by the behavior of the schedule objectives. By Proposition 5.5, each selected-label standard deviation satisfies

$$0 \leq \sigma_{\mathrm{BBDM},r,k} \leq \frac{1}{\sqrt{\lambda_{r,k}}}.$$

The MSE objective decreases as $\sigma_{\mathrm{BBDM},r,k}$ moves toward 0, whereas the Wasserstein objective decreases as $\sigma_{\mathrm{BBDM},r,k}$ moves toward $1/\sqrt{\lambda_{r,k}}$. Driving the sampler close to either boundary can make the bridge parameters increasingly extreme, while the objective improvements exhibit diminishing returns once the standard deviations are already close to the preferred boundary. The bounded box keeps the schedules in a moderate, non-degenerate range while still allowing low-loss solutions for both objectives.

F MoG-Free Schedule Heuristics from a Variance Bound

The schedule objectives derived in Corollaries 6.4 and 6.5 are explicit once the MoG posterior precisions $\lambda_{r,k}$ are known. However, fitting a reliable high-dimensional MoG is generally impractical for image-scale data. For example, images of size $256 \times 256 \times 3$ live in dimension 196,608, and estimating mixture covariances in those dimensions is extremely demanding.

This appendix derives simple schedule-parameter heuristics that do not require fitting the full MoG.

F.1 A bridge precision scale

Define, for $s = 1, \dots, S-1$,

$$\rho_s \triangleq \frac{(1-m_s)^2}{\delta_s}. \tag{129}$$

This quantity is the scalar added to the component posterior precision $\lambda_{r,k}$ in the covariance

$$\frac{1}{\lambda_{r,k} + \frac{(1-m_s)^2}{\delta_s}} = \frac{1}{\lambda_{r,k} + \rho_s}.$$

We also use the endpoint conventions

$$\rho_0 = +\infty, \qquad \rho_S = 0.$$

The bridge transition constraint from Appendix E is

$$\delta_s - \delta_{s-1}\frac{(1-m_s)^2}{(1-m_{s-1})^2} \geq 0. \tag{130}$$

Using

$$\delta_s = \frac{(1-m_s)^2}{\rho_s}, \qquad \delta_{s-1} = \frac{(1-m_{s-1})^2}{\rho_{s-1}},$$

we get

$$\begin{aligned}
\delta_s - \delta_{s-1}\frac{(1-m_s)^2}{(1-m_{s-1})^2} &= \frac{(1-m_s)^2}{\rho_s} - \frac{(1-m_{s-1})^2}{\rho_{s-1}}\frac{(1-m_s)^2}{(1-m_{s-1})^2} \\
&= \frac{(1-m_s)^2}{\rho_s} - \frac{(1-m_s)^2}{\rho_{s-1}} \\
&= (1-m_s)^2\left(\frac{1}{\rho_s} - \frac{1}{\rho_{s-1}}\right).
\end{aligned} \tag{131}$$

Therefore the bridge constraint is equivalent to

$$\rho_s \leq \rho_{s-1}. \tag{132}$$

Thus a valid bridge corresponds to a decreasing sequence

$$+\infty = \rho_0 \geq \rho_1 \geq \rho_2 \geq \cdots \geq \rho_{S-1} > 0, \qquad \rho_S = 0.$$

For the bounded schedule family,

$$\delta_s = c[4m_s(1-m_s)]^\gamma.$$

Substituting this into (129) gives

$$\rho_s = \frac{(1-m_s)^2}{c[4m_s(1-m_s)]^\gamma} = \frac{1}{c4^\gamma} m_s^{-\gamma}(1-m_s)^{2-\gamma}. \tag{133}$$

For $0 < \gamma \leq 2$, this is decreasing as a function of m_s. Indeed,

$$\frac{\partial}{\partial m}\log\left[\frac{1}{c4^\gamma} m^{-\gamma}(1-m)^{2-\gamma}\right] = -\frac{\gamma}{m} - \frac{2-\gamma}{1-m} < 0. \tag{134}$$

F.2 Exact selected-label variance in the ρ scale

We now rewrite the selected-label BBDM variance from Proposition 5.5 in terms of ρ_s.

First, the BBDM bridge increment is

$$\begin{aligned}\delta_{s|s-1} &= \delta_s - \delta_{s-1}\frac{(1-m_s)^2}{(1-m_{s-1})^2} \\ &= (1-m_s)^2\left(\frac{1}{\rho_s} - \frac{1}{\rho_{s-1}}\right).\end{aligned} \tag{135}$$

Therefore

$$\begin{aligned}\sigma_s^2 &= \frac{\delta_{s|s-1}\delta_{s-1}}{\delta_s} \\ &= \frac{(1-m_s)^2\left(\frac{1}{\rho_s} - \frac{1}{\rho_{s-1}}\right)\cdot\frac{(1-m_{s-1})^2}{\rho_{s-1}}}{\frac{(1-m_s)^2}{\rho_s}} \\ &= (1-m_{s-1})^2\frac{\rho_{s-1}-\rho_s}{\rho_{s-1}^2}.\end{aligned} \tag{136}$$

Next, compute c_s. Since

$$c_s = \sqrt{\frac{\delta_{s-1}-\sigma_s^2}{\delta_s}},$$

we first simplify the numerator:

$$\begin{aligned}\delta_{s-1}-\sigma_s^2 &= \frac{(1-m_{s-1})^2}{\rho_{s-1}} - (1-m_{s-1})^2\frac{\rho_{s-1}-\rho_s}{\rho_{s-1}^2} \\ &= (1-m_{s-1})^2\left[\frac{1}{\rho_{s-1}} - \frac{\rho_{s-1}-\rho_s}{\rho_{s-1}^2}\right] \\ &= (1-m_{s-1})^2\frac{\rho_s}{\rho_{s-1}^2}.\end{aligned} \tag{137}$$

Hence

$$\begin{aligned}c_s &= \sqrt{\frac{(1-m_{s-1})^2\frac{\rho_s}{\rho_{s-1}^2}}{\frac{(1-m_s)^2}{\rho_s}}} \\ &= \frac{1-m_{s-1}}{1-m_s}\frac{\rho_s}{\rho_{s-1}}.\end{aligned} \tag{138}$$

Then

$$\begin{aligned}a_s &= (1-m_{s-1}) - (1-m_s)c_s \\ &= (1-m_{s-1}) - (1-m_s)\frac{1-m_{s-1}}{1-m_s}\frac{\rho_s}{\rho_{s-1}} \\ &= (1-m_{s-1})\frac{\rho_{s-1}-\rho_s}{\rho_{s-1}}.\end{aligned} \tag{139}$$

The scalar diagonal entry of $G_r(s)$ is

$$g_{r,k}(s) = c_s + \frac{a_s(1-m_s)}{\delta_s}\frac{1}{\lambda_{r,k}+\rho_s}.$$

Using

$$\delta_s = \frac{(1-m_s)^2}{\rho_s},$$

we have

$$\frac{a_s(1-m_s)}{\delta_s} = \frac{a_s\rho_s}{1-m_s}.$$

Therefore

$$\begin{aligned} g_{r,k}(s) &= \frac{1-m_{s-1}}{1-m_s}\frac{\rho_s}{\rho_{s-1}} + \frac{(1-m_{s-1})\frac{\rho_{s-1}-\rho_s}{\rho_{s-1}}\cdot\rho_s}{(1-m_s)(\lambda_{r,k}+\rho_s)} \\ &= \frac{1-m_{s-1}}{1-m_s}\frac{\rho_s}{\rho_{s-1}}\left[1+\frac{\rho_{s-1}-\rho_s}{\lambda_{r,k}+\rho_s}\right] \\ &= \frac{1-m_{s-1}}{1-m_s}\frac{\rho_s}{\rho_{s-1}}\frac{\lambda_{r,k}+\rho_{s-1}}{\lambda_{r,k}+\rho_s}. \end{aligned} \tag{140}$$

Now telescope the product in Proposition 5.5. For $i \geq 2$,

$$\begin{aligned} \prod_{j=1}^{i-1} g_{r,k}(j)^2 &= \left[\prod_{j=1}^{i-1}\frac{1-m_{j-1}}{1-m_j}\frac{\rho_j}{\rho_{j-1}}\frac{\lambda_{r,k}+\rho_{j-1}}{\lambda_{r,k}+\rho_j}\right]^2 \\ &= \left[\frac{1}{1-m_{i-1}}\cdot\frac{\rho_{i-1}}{\rho_0}\cdot\frac{\lambda_{r,k}+\rho_0}{\lambda_{r,k}+\rho_{i-1}}\right]^2. \end{aligned} \tag{141}$$

Using the endpoint convention $\rho_0 = +\infty$,

$$\frac{\rho_{i-1}}{\rho_0}\frac{\lambda_{r,k}+\rho_0}{\lambda_{r,k}+\rho_{i-1}} \longrightarrow \frac{\rho_{i-1}}{\lambda_{r,k}+\rho_{i-1}}.$$

Hence

$$\prod_{j=1}^{i-1} g_{r,k}(j)^2 = \frac{\rho_{i-1}^2}{(1-m_{i-1})^2(\lambda_{r,k}+\rho_{i-1})^2}. \tag{142}$$

Multiplying (142) with (136) gives

$$\begin{aligned} \sigma_i^2\prod_{j=1}^{i-1} g_{r,k}(j)^2 &= (1-m_{i-1})^2\frac{\rho_{i-1}-\rho_i}{\rho_{i-1}^2}\cdot\frac{\rho_{i-1}^2}{(1-m_{i-1})^2(\lambda_{r,k}+\rho_{i-1})^2} \\ &= \frac{\rho_{i-1}-\rho_i}{(\lambda_{r,k}+\rho_{i-1})^2}. \end{aligned} \tag{143}$$

The $i = 1$ term vanishes under the convention $\rho_0 = +\infty$. Therefore

$$\sigma^2_{\mathrm{BBDM},r,k} = \sum_{i=2}^{S}\frac{\rho_{i-1}-\rho_i}{(\lambda_{r,k}+\rho_{i-1})^2}. \tag{144}$$

This identity is exact inside the selected-label model.

F.3 MSE upper bound and edge heuristic

By Corollary 6.5,

$$J^{\mathrm{MSE}}_{\mathrm{MOG}} = \sum_{r=1}^{R}\pi_r\sum_{k=1}^{d}\left(\frac{1}{\lambda_{r,k}} + \sigma^2_{\mathrm{BBDM},r,k}\right).$$

The first term is independent of the schedule, so the schedule-dependent part is only

$$\sum_{r=1}^{R} \pi_r \sum_{k=1}^{d} \sigma^2_{\mathrm{BBDM},r,k}.$$

For each fixed pair (r, k), the function

$$\rho \mapsto \frac{1}{(\lambda_{r,k} + \rho)^2}$$

is decreasing in ρ. From (144),

$$\sigma^2_{\mathrm{BBDM},r,k} = \sum_{i=2}^{S} (\rho_{i-1} - \rho_i) \frac{1}{(\lambda_{r,k} + \rho_{i-1})^2}.$$

On the interval $[\rho_i, \rho_{i-1}]$, the endpoint value at ρ_{i-1} is the smallest value. Hence

$$(\rho_{i-1} - \rho_i) \frac{1}{(\lambda_{r,k} + \rho_{i-1})^2} \leq \int_{\rho_i}^{\rho_{i-1}} \frac{d\rho}{(\lambda_{r,k} + \rho)^2}. \tag{145}$$

Summing over $i = 2, \ldots, S$ gives

$$\begin{aligned} \sigma^2_{\mathrm{BBDM},r,k} &\leq \int_0^{\rho_1} \frac{d\rho}{(\lambda_{r,k} + \rho)^2} \\ &= \frac{1}{\lambda_{r,k}} - \frac{1}{\lambda_{r,k} + \rho_1} \end{aligned} \tag{146}$$

Thus the MSE upper bound is minimized by making ρ_1 as small as possible.

For the bounded schedule family,

$$\rho_1 = \frac{1}{c4^\gamma} m_1^{-\gamma} (1 - m_1)^{2-\gamma}, \qquad m_1 = 1 - (1 - S^{-\alpha})^\beta.$$

We now inspect the parameter directions.

First,

$$\frac{\partial \log \rho_1}{\partial c} = -\frac{1}{c} < 0,$$

so increasing c decreases ρ_1.

Second,

$$\frac{\partial \log \rho_1}{\partial m_1} = -\frac{\gamma}{m_1} - \frac{2 - \gamma}{1 - m_1} < 0$$

for $0 < \gamma \leq 2$. Therefore increasing m_1 decreases ρ_1.

Now

$$m_1 = 1 - (1 - S^{-\alpha})^\beta.$$

For β,

$$\frac{\partial m_1}{\partial \beta} = -(1 - S^{-\alpha})^\beta \log(1 - S^{-\alpha}) > 0,$$

because $0 < 1 - S^{-\alpha} < 1$. Thus increasing β increases m_1 and therefore decreases ρ_1.

For α,

$$\frac{\partial m_1}{\partial \alpha} = -\beta (1 - S^{-\alpha})^{\beta - 1} (\log S) S^{-\alpha} < 0.$$

Thus decreasing α increases m_1 and therefore decreases ρ_1.

Finally,

$$\frac{\partial \log \rho_1}{\partial \gamma} = -\log(4m_1(1-m_1)).$$

Since $4m_1(1-m_1) \leq 1$, we have

$$-\log(4m_1(1-m_1)) \geq 0.$$

Thus decreasing γ decreases ρ_1.

Consequently, over the bounded box

$$\alpha \in [1,2], \qquad \beta \in [1,2], \qquad c \in [0.2,2], \qquad \gamma \in [0.2,2],$$

the MSE upper bound is minimized by the edge choice

$$\boxed{(\alpha, \beta, c, \gamma) = (1, 2, 2, 0.2).} \tag{147}$$

This conclusion does not require the numerical values of the $\lambda_{r,k}$'s.

F.4 W2 heuristic: use the opposite edge

By Corollary 6.4, the selected-label Wasserstein objective is

$$J_{\mathrm{MOG}}^{W_2} = \sum_{r=1}^{R} \pi_r \sum_{k=1}^{d} \left(\sigma_{\mathrm{BBDM},r,k} - \frac{1}{\sqrt{\lambda_{r,k}}} \right)^2 .$$

By Proposition 5.5,

$$0 \leq \sigma_{\mathrm{BBDM},r,k} \leq \frac{1}{\sqrt{\lambda_{r,k}}}.$$

Therefore, coordinate-wise, the Wasserstein objective is reduced when $\sigma_{\mathrm{BBDM},r,k}$ is increased toward the posterior standard deviation.

The bound in (146) gives

$$\sigma^2_{\mathrm{BBDM},r,k} \leq \frac{1}{\lambda_{r,k}} - \frac{1}{\lambda_{r,k} + \rho_1}$$

The right-hand side is increasing in ρ_1. Hence, as a simple MoG-free W2 heuristic, we take the opposite direction from the MSE rule: instead of minimizing ρ_1, we choose the parameters that increase ρ_1.

Thus, over the bounded box, the W2-oriented heuristic reverses the MSE edge:

$$\alpha \uparrow, \qquad \beta \downarrow, \qquad c \downarrow, \qquad \gamma \uparrow .$$

Equivalently, the edge candidate is

$$\boxed{(\alpha, \beta, c, \gamma) = (2, 1, 0.2, 2).} \tag{148}$$

This is not a certificate of a universal W2 optimum, but a practical schedule heuristic: it increases the upper bound on the sampler variance, which is preferred for matching the posterior variance.

G Selected-chain self-consistency

This appendix states and proves a self-consistency result supporting the selected-label approximation. The result shows that, in a linearly separated shared-covariance MoG regime, freezing the correct component label is self-consistent: along the selected-label chain, the exact MoG responsibilities continue to favor the same label with high probability.

Assumption G.1. *All mixture covariances are shared:*

$$\Sigma_r = \Sigma \succ 0 \qquad \text{for all } r \in \{1, \dots, R\}.$$

Theorem G.2. *Define the pairwise measurement separation*

$$M_{rj} \triangleq \big(H(\mu_r - \mu_j)\big)^\top \big(H\Sigma H^\top + \sigma_y^2 I\big)^{-1} \big(H(\mu_r - \mu_j)\big). \tag{149}$$

Fix a true component $r^\star$. Suppose that R and S are fixed, $\pi_r \geq \pi_{\min} > 0$, and for some $\alpha > 0$,

$$M_{r^\star j} \geq \alpha d \qquad \text{for all } j \neq r^\star. \tag{150}$$

Generate y under $C = r^\star$, and run the selected-label chain with the label forced to $J = r^\star$, producing X_s^{sel}. Let $\mathcal{E}_{r^\star,\varepsilon}^{\text{sel}}$ be the event that

$$\gamma_{r^\star|y}(y) \geq 1 - \varepsilon,$$
$$\gamma_{r^\star|s}(X_s^{\text{sel}}, y) \geq 1 - \varepsilon \quad \forall s = 1, \dots, S-1.$$

For every fixed $\varepsilon \in (0, 1/2)$, there exist $c, K > 0$ and $d_0 \in \mathbb{N}$, independent of d, such that for all $d \geq d_0$,

$$\mathbb{P}\big(\mathcal{E}_{r^\star,\varepsilon}^{\text{sel}} \mid C = r^\star\big) \geq 1 - KSe^{-cd}. \tag{151}$$

Since $J \sim \gamma_{\cdot|y}(y)$, also

$$\mathbb{P}(J = r^\star \mid C = r^\star) \geq 1 - \varepsilon - KSe^{-cd}.$$

Proof outline

The proof has two steps. First, we prove a bridge-marginal one-hot concentration result. Second, we transfer the same concentration mechanism to the selected-label chain, using the fact that the frozen surrogate is mean exact and covariance deficient compared to the target posterior. We also record a deterministic local perturbation bound between the exact MoG update and the frozen-label update.

Shared-covariance model

Under Assumption G.1, all components share the same covariance. Thus,

$$x_0 \mid (C = r) \sim \mathcal{N}(\mu_r, \Sigma), \qquad y \mid x_0 \sim \mathcal{N}(Hx_0, \sigma_y^2 I).$$

The measurement-conditioned posterior covariance becomes component-independent:

$$\Sigma_{|y}^{-1} = \Sigma^{-1} + \frac{1}{\sigma_y^2} H^\top H. \tag{152}$$

The corresponding posterior mean for component r is

$$\mu_{r|y} = \Sigma_{|y} \left(\Sigma^{-1}\mu_r + \frac{1}{\sigma_y^2} H^\top y \right). \tag{153}$$

Therefore, for every pair $r \neq j$,

$$\mu_{r|y} - \mu_{j|y} = \Sigma_{|y}\Sigma^{-1}(\mu_r - \mu_j), \tag{154}$$

which is independent of the realization of y.

The conditional bridge law under a fixed component

Under the shared-covariance assumption, the conditional law of the bridge variable x_s given $(y, C = r)$ also has a covariance that does not depend on r. Indeed,

$$x_s \mid (y, C = r) \sim \mathcal{N}\big(u_{r,s}(y), V_s\big), \tag{155}$$

where

$$u_{r,s}(y) = (1 - m_s)\mu_{r|y} + m_s y, \tag{156}$$

and

$$V_s = (1 - m_s)^2 \Sigma_{|y} + \delta_s I. \tag{157}$$

Using (154), the difference between the component-conditioned bridge means is

$$u_{r,s}(y) - u_{j,s}(y) = (1 - m_s)\Sigma_{|y}\Sigma^{-1}(\mu_r - \mu_j), \tag{158}$$

which is again independent of the realization of y.

G.3 Two separation quantities

The proof will be expressed through two pairwise separations.

Measurement separation. For each pair of components $r \neq j$, define

$$M_{rj} \triangleq \big(H(\mu_r - \mu_j)\big)^\top \big(H\Sigma H^\top + \sigma_y^2 I\big)^{-1} \big(H(\mu_r - \mu_j)\big). \tag{159}$$

Equivalently, if we define

$$A \triangleq H^\top \big(H\Sigma H^\top + \sigma_y^2 I\big)^{-1} H,$$

then

$$M_{rj} = (\mu_r - \mu_j)^\top A(\mu_r - \mu_j). \tag{160}$$

Bridge separation at step s. For each pair $r \neq j$, define

$$B_{rj,s} \triangleq \big(u_{r,s}(y) - u_{j,s}(y)\big)^\top V_s^{-1} \big(u_{r,s}(y) - u_{j,s}(y)\big). \tag{161}$$

By (158), this quantity is deterministic and does not depend on the realization of y. It can be written as

$$B_{rj,s} = (1 - m_s)^2 (\mu_r - \mu_j)^\top \Sigma^{-1} \Sigma_{|y}\, V_s^{-1}\, \Sigma_{|y} \Sigma^{-1} (\mu_r - \mu_j). \tag{162}$$

G.4 Pairwise log-posterior ratios

We now study the pairwise log-ratio between the correct component and an incorrect one.

Lemma 1 For $r \neq j$, define

$$L_{rj}^{(y)}(y) \triangleq \log \frac{\gamma_{r|y}(y)}{\gamma_{j|y}(y)}. \tag{163}$$

Then, under the event $C = r$,

$$L_{rj}^{(y)}(y) \sim \mathcal{N}\left(\log \frac{\pi_r}{\pi_j} + \frac{1}{2} M_{rj},\ M_{rj}\right). \tag{164}$$

Proof. Under the shared-covariance model,

$$y \mid (C = r) \sim \mathcal{N}(H\mu_r,\ H\Sigma H^\top + \sigma_y^2 I).$$

Set

$$S_y \triangleq H\Sigma H^\top + \sigma_y^2 I, \qquad m_r \triangleq H\mu_r, \qquad m_j \triangleq H\mu_j.$$

Using the formula for $\gamma_{r|y}(y)$, we obtain:

$$\begin{aligned} L_{rj}^{(y)}(y) &= \log \frac{\pi_r \mathcal{N}(y; m_r, S_y)}{\pi_j \mathcal{N}(y; m_j, S_y)} \\ &= \log \frac{\pi_r \exp\left(-\frac{1}{2}(y - m_r)^\top S_y^{-1}(y - m_r)\right)}{\pi_j \exp\left(-\frac{1}{2}(y - m_j)^\top S_y^{-1}(y - m_j)\right)} \\ &= \log \frac{\pi_r}{\pi_j} - \frac{1}{2}(y - m_r)^\top S_y^{-1}(y - m_r) + \frac{1}{2}(y - m_j)^\top S_y^{-1}(y - m_j). \end{aligned} \tag{165}$$

Expanding the quadratic forms and using the symmetry of S_y^{-1} (which implies $y^\top S_y^{-1} m_r = m_r^\top S_y^{-1} y$), the $y^\top S_y^{-1} y$ terms cancel out:

$$\begin{aligned} L_{rj}^{(y)}(y) &= \log \frac{\pi_r}{\pi_j} - \frac{1}{2}\left(-2m_r^\top S_y^{-1} y + m_r^\top S_y^{-1} m_r\right) + \frac{1}{2}\left(-2m_j^\top S_y^{-1} y + m_j^\top S_y^{-1} m_j\right) \\ &= \log \frac{\pi_r}{\pi_j} + (m_r - m_j)^\top S_y^{-1} y - \frac{1}{2} m_r^\top S_y^{-1} m_r + \frac{1}{2} m_j^\top S_y^{-1} m_j. \end{aligned} \tag{166}$$

Finally, we can factor the remaining terms by noting that $(m_r - m_j)^\top S_y^{-1}(m_r + m_j) = m_r^\top S_y^{-1} m_r - m_j^\top S_y^{-1} m_j$ because the cross-terms cancel. This yields:

$$L_{rj}^{(y)}(y) = \log \frac{\pi_r}{\pi_j} + (m_r - m_j)^\top S_y^{-1}\left(y - \frac{1}{2}(m_r + m_j)\right). \tag{167}$$

If $C = r$, then $y = m_r + S_y^{1/2}\xi$ with $\xi \sim \mathcal{N}(0, I)$. Substituting into (167) gives

$$L_{rj}^{(y)}(y) = \log \frac{\pi_r}{\pi_j} + \frac{1}{2}(m_r - m_j)^\top S_y^{-1}(m_r - m_j) + \left(S_y^{-1/2}(m_r - m_j)\right)^\top \xi.$$

The last term is Gaussian with mean 0 and variance $(m_r - m_j)^\top S_y^{-1}(m_r - m_j) = M_{rj}$. This proves (164).

Lemma 2 For $r \neq j$ and each step s, define

$$Q_{rj,s}(x_s, y) \triangleq \log \frac{p(x_s \mid y, C = r)}{p(x_s \mid y, C = j)}. \tag{168}$$

Then, conditioned on $(y, C = r)$,

$$Q_{rj,s}(x_s, y) \sim \mathcal{N}\left(\frac{1}{2}B_{rj,s},\ B_{rj,s}\right). \tag{169}$$

Proof. From (155),

$$x_s \mid (y, C = r) \sim \mathcal{N}(u_{r,s}(y), V_s), \qquad x_s \mid (y, C = j) \sim \mathcal{N}(u_{j,s}(y), V_s).$$

Since both Gaussians have the same covariance V_s, using similar steps to Lemma 1,

$$Q_{rj,s}(x_s, y) = \left(u_{r,s}(y) - u_{j,s}(y)\right)^\top V_s^{-1}\left(x_s - \frac{1}{2}(u_{r,s}(y) + u_{j,s}(y))\right). \tag{170}$$

If $C = r$, then

$$x_s = u_{r,s}(y) + V_s^{1/2}\zeta_s, \qquad \zeta_s \sim \mathcal{N}(0, I).$$

Substituting into (170) yields

$$Q_{rj,s}(x_s, y) = \frac{1}{2}\big(u_{r,s}(y) - u_{j,s}(y)\big)^\top V_s^{-1}\big(u_{r,s}(y) - u_{j,s}(y)\big) + \big(V_s^{-1/2}(u_{r,s}(y) - u_{j,s}(y))\big)^\top \zeta_s.$$

The first term is $\frac{1}{2}B_{rj,s}$ and the second term is centered Gaussian with variance $B_{rj,s}$. This proves (169).

Lemma 3 For $r \neq j$, define

$$L_{rj,s}(x_s, y) \triangleq \log \frac{\gamma_{r|s}(x_s, y)}{\gamma_{j|s}(x_s, y)}. \tag{171}$$

Then

$$L_{rj,s}(x_s, y) = L_{rj}^{(y)}(y) + Q_{rj,s}(x_s, y). \tag{172}$$

Proof. By Bayes' rule,

$$\gamma_{r|s}(x_s, y) \propto \gamma_{r|y}(y)\, p(x_s \mid y, C = r).$$

Taking the ratio between components r and j gives

$$\frac{\gamma_{r|s}(x_s, y)}{\gamma_{j|s}(x_s, y)} = \frac{\gamma_{r|y}(y)}{\gamma_{j|y}(y)} \cdot \frac{p(x_s \mid y, C = r)}{p(x_s \mid y, C = j)}.$$

Taking logarithms yields (172).

G.5 Turning pairwise margins into one-hot responsibilities

The next elementary lemma converts a lower bound on all pairwise log-ratios into a lower bound on the winning responsibility.

Lemma 4 Suppose that for some index r and some threshold $T \geq 0$,

$$\log \frac{\gamma_r}{\gamma_j} \geq T \qquad \text{for all } j \neq r. \tag{173}$$

Then

$$\gamma_r \geq \frac{1}{1 + (R-1)e^{-T}}. \tag{174}$$

Proof. From (173),

$$\gamma_j \leq e^{-T}\gamma_r \qquad \text{for all } j \neq r.$$

Therefore,

$$1 = \gamma_r + \sum_{j \neq r} \gamma_j \leq \gamma_r\big(1 + (R-1)e^{-T}\big),$$

which implies (174).

G.6 Main theorem: the same component stays dominant across all noise levels

We now state the main theorem. Its role is to make the concentration mechanism completely explicit: once the measurement term is large enough, the bridge fluctuations are not strong enough to overturn the winning component.

Theorem 1 Assume the shared-covariance model (G.1). Fix a true component $r^\star$ and suppose that the sample is generated under $C = r^\star$.

Let $T_0 > T_1 > 0$, and define the event

$$\mathcal{E}(T_0, T_1) \triangleq \left\{ \begin{array}{ll} L^{(y)}_{r^\star j}(y) \geq T_0 & \text{for all } j \neq r^\star, \\ L_{r^\star j,s}(x_s, y) \geq T_1 & \text{for all } j \neq r^\star \text{ and all } s = 1, \ldots, S-1 \end{array} \right\}. \tag{175}$$

Then

$$\mathbb{P}\big(\mathcal{E}(T_0, T_1)^c \mid C = r^\star\big) \leq \sum_{j \neq r^\star} \exp\left(-\frac{\left[\log \frac{\pi_{r^\star}}{\pi_j} + \frac{1}{2} M_{r^\star j} - T_0\right]_+^2}{2M_{r^\star j}} \right) + (S-1)(R-1)e^{-(T_0 - T_1)}. \tag{176}$$

Moreover, on the event $\mathcal{E}(T_0, T_1)$, the following hold simultaneously:

1. the measurement posterior is concentrated on the true component:

$$\gamma_{r^\star|y}(y) \geq \frac{1}{1 + (R-1)e^{-T_0}}; \tag{177}$$

2. for every noise level $s = 1, \ldots, S-1$, the stepwise responsibilities are concentrated on the same component:

$$\gamma_{r^\star|s}(x_s, y) \geq \frac{1}{1 + (R-1)e^{-T_1}}; \tag{178}$$

3. the same component $r^\star$ is strictly dominant for all $s = 1, \ldots, S-1$.

Proof. We split the proof into two parts.

Part 1: concentration of the measurement posterior.

Define

$$\mathcal{E}_y(T_0) \triangleq \left\{ L^{(y)}_{r^\star j}(y) \geq T_0 \quad \text{for all } j \neq r^\star \right\}. \tag{179}$$

By Lemma 1, for each $j \neq r^\star$,

$$L^{(y)}_{r^\star j}(y) \mid (C = r^\star) \sim \mathcal{N}\left(\log \frac{\pi_{r^\star}}{\pi_j} + \frac{1}{2} M_{r^\star j},\ M_{r^\star j} \right).$$

Let

$$\Delta_{r^\star j}(T_0) \triangleq \left[\log \frac{\pi_{r^\star}}{\pi_j} + \frac{1}{2} M_{r^\star j} - T_0 \right]_+.$$

If $\Delta_{r^\star j}(T_0) = 0$, then the trivial bound

$$\mathbb{P}\left(L^{(y)}_{r^\star j}(y) < T_0 \mid C = r^\star \right) \leq 1$$

already yields

$$\mathbb{P}\left(L^{(y)}_{r^\star j}(y) < T_0 \mid C = r^\star \right) \leq \exp\left(-\frac{\Delta_{r^\star j}(T_0)^2}{2M_{r^\star j}} \right).$$

If instead $\Delta_{r^\star j}(T_0) > 0$, then T_0 lies below the Gaussian mean, and the standard Gaussian tail bound $\Phi(-u) \leq e^{-u^2/2}$ gives

$$\mathbb{P}\left(L^{(y)}_{r^\star j}(y) < T_0 \mid C = r^\star \right) \leq \exp\left(-\frac{\Delta_{r^\star j}(T_0)^2}{2M_{r^\star j}} \right).$$

Applying a union bound over all $j \neq r^\star$ gives

$$\mathbb{P}\big(\mathcal{E}_y(T_0)^c \mid C = r^\star\big) \leq \sum_{j \neq r^\star} \exp\left(- \frac{\left[\log \frac{\pi_{r^\star}}{\pi_j} + \frac{1}{2} M_{r^\star j} - T_0\right]_+^2}{2M_{r^\star j}} \right). \tag{180}$$

On the event $\mathcal{E}_y(T_0)$, we have

$$\log \frac{\gamma_{r^\star|y}(y)}{\gamma_{j|y}(y)} \geq T_0 \qquad \text{for all } j \neq r^\star,$$

so Lemma 4 implies (177).

Part 2: stability across all noise levels.

Fix a pair (j, s) with $j \neq r^\star$ and $s \in \{1, \ldots, S-1\}$. On the event $\mathcal{E}_y(T_0)$, if

$$L_{r^\star j,s}(x_s, y) < T_1,$$

then by Lemma 3,

$$Q_{r^\star j,s}(x_s, y) = L_{r^\star j,s}(x_s, y) - L^{(y)}_{r^\star j}(y) < -(T_0 - T_1).$$

Therefore,

$$\mathbb{P}(L_{r^\star j,s}(x_s, y) < T_1 \mid C = r^\star, \mathcal{E}_y(T_0), y) \leq \mathbb{P}(Q_{r^\star j,s}(x_s, y) < -(T_0 - T_1) \mid C = r^\star, y)\,. \tag{181}$$

By Lemma 2,

$$Q_{r^\star j,s}(x_s, y) \mid (C = r^\star, y) \sim \mathcal{N}\left(\frac{1}{2} B_{r^\star j,s},\, B_{r^\star j,s}\right).$$

If $B_{r^\star j,s} = 0$, then $Q_{r^\star j,s}(x_s, y) = 0$ almost surely, so the probability on the right-hand side of (181) is 0, and the desired bound is trivial. Otherwise, $B_{r^\star j,s} > 0$, and

$$\mathbb{P}(Q_{r^\star j,s}(x_s, y) < -(T_0 - T_1) \mid C = r^\star, y) \leq \exp\left(- \frac{\big(T_0 - T_1 + \frac{1}{2} B_{r^\star j,s}\big)^2}{2B_{r^\star j,s}} \right). \tag{182}$$

Set $a \triangleq T_0 - T_1 > 0$ and $B \triangleq B_{r^\star j,s}$. Then

$$\frac{(a + \frac{1}{2}B)^2}{2B} = \frac{a^2}{2B} + \frac{a}{2} + \frac{B}{8} \geq a,$$

because

$$\frac{a^2}{2B} + \frac{B}{8} \geq \frac{a}{2}$$

by the arithmetic-geometric mean inequality. Therefore

$$\mathbb{P}(L_{r^\star j,s}(x_s, y) < T_1 \mid C = r^\star, \mathcal{E}_y(T_0), y) \leq e^{-(T_0 - T_1)}. \tag{183}$$

This bound is uniform in y, so after averaging over y it remains valid:

$$\mathbb{P}(L_{r^\star j,s}(x_s, y) < T_1 \mid C = r^\star, \mathcal{E}_y(T_0)) \leq e^{-(T_0 - T_1)}. \tag{184}$$

Applying a union bound over all $j \neq r^\star$ and all $s = 1, \ldots, S-1$,

$$\begin{aligned} &\mathbb{P}(\exists\, j \neq r^\star, \exists\, s \in \{1, \ldots, S-1\} : L_{r^\star j,s}(x_s, y) < T_1 \mid C = r^\star, \mathcal{E}_y(T_0)) \\ &\quad \leq (S-1)(R-1) e^{-(T_0 - T_1)}. \end{aligned} \tag{185}$$

Combining (180) and (185) proves (176).

Finally, on the event $\mathcal{E}(T_0, T_1)$, we have

$$\log \frac{\gamma_{r^\star|s}(x_s, y)}{\gamma_{j|s}(x_s, y)} \geq T_1 \qquad \text{for all } j \neq r^\star \text{ and all } s = 1, \ldots, S-1.$$

Applying Lemma 4 at each noise level yields (178). Since $T_1 > 0$, these pairwise log-ratios are strictly positive, hence $\gamma_{r^\star|s}(x_s, y) > \gamma_{j|s}(x_s, y)$ for every $j \neq r^\star$ and every $s = 1, \ldots, S-1$, so the same component is strictly dominant across all noise levels.

G.7 A high-dimensional corollary

The following corollary reformulates the previous theorem into the statement used in the main text: when the measurement separation is proportional to the dimension, the failure probability decays exponentially in d.

Corollary 1 Assume the setting of Theorem 1, and suppose that there exists a constant $\alpha > 0$ such that

$$M_{rj} \geq \alpha d \qquad \text{for all } r \neq j. \tag{186}$$

Assume also that the mixture weights are uniformly bounded below:

$$\pi_r \geq \pi_{\min} > 0 \qquad \text{for all } r. \tag{187}$$

Fix R, S, and $\varepsilon \in (0, 1/2)$, and define

$$T_\varepsilon \triangleq \log \frac{(R-1)(1-\varepsilon)}{\varepsilon}. \tag{188}$$

Then there exist constants $c > 0$, $K > 0$, and $d_0 \in \mathbb{N}$ depending only on α, $\pi_{\min}$, ε, and R, such that for every $d \geq d_0$,

$$\begin{aligned} &\mathbb{P}\Big(\gamma_{r^\star|y}(y) \geq 1-\varepsilon,\ \gamma_{r^\star|s}(x_s, y) \geq 1-\varepsilon,\ \text{and } r^\star \text{ is dominant } \forall s = 1, \ldots, S-1 \ \Big|\ C = r^\star\Big) \\ &\qquad \geq 1 - KSe^{-cd}. \end{aligned} \tag{189}$$

Proof. Choose

$$T_1 = T_\varepsilon, \qquad T_0 = \frac{\alpha d}{4}. \tag{190}$$

By Lemma 4, the condition

$$\log \frac{\gamma_{r^\star|s}}{\gamma_{j|s}} \geq T_\varepsilon \quad \forall j \neq r^\star$$

implies

$$\gamma_{r^\star|s} \geq \frac{1}{1 + (R-1)e^{-T_\varepsilon}} = 1 - \varepsilon.$$

The same implication also holds for $\gamma_{r^\star|y}(y)$.

It therefore suffices to control the failure probability in (176). Since $M_{r^\star j} \geq \alpha d$ and $\pi_{r^\star}/\pi_j \geq \pi_{\min}$, for sufficiently large d we have

$$\log \frac{\pi_{r^\star}}{\pi_j} + \frac{1}{2} M_{r^\star j} - T_0 \geq \log \pi_{\min} + \frac{1}{4} M_{r^\star j} \geq \frac{1}{8} M_{r^\star j}.$$

Therefore the first term in (176) is at most

$$\sum_{j \neq r^\star} \exp\left(-\frac{M_{r^\star j}}{128}\right) \leq (R-1)e^{-\alpha d/128}.$$

Similarly, for sufficiently large d,

$$T_0 - T_1 = \frac{\alpha d}{4} - T_\varepsilon \geq \frac{\alpha d}{8},$$

so the second term in (176) is at most

$$(S-1)(R-1)e^{-\alpha d/8}.$$

Combining the two bounds yields

$$\mathbb{P}\big(\mathcal{E}(T_0, T_1)^c \mid C = r^\star\big) \leq (R-1)e^{-\alpha d/128} + (S-1)(R-1)e^{-\alpha d/8},$$

which implies (189) for suitable constants c, K, d_0.

G.8 Random means imply linear measurement separation

The high-dimensional corollary is driven by the pairwise measurement separation

$$M_{rj} = (\mu_r - \mu_j)^\top A(\mu_r - \mu_j).$$

$$A = H^\top (H\Sigma H^\top + \sigma_y^2 I)^{-1} H.$$

Since A is symmetric positive semidefinite, we can write its eigendecomposition as

$$A = \sum_{k=1}^{d} \eta_k v_k v_k^\top, \qquad \eta_k \geq 0.$$

Assume that A has a linear number of informative directions: there exists $I \subset \{1, \dots, d\}$ such that

$$|I| \geq \rho d \qquad \text{and} \qquad \eta_k \geq \eta_{\min} > 0 \quad \text{for all } k \in I. \tag{191}$$

Then

$$A \succeq \eta_{\min} \sum_{k \in I} v_k v_k^\top.$$

Proposition 1 Suppose the component means are drawn independently in a fixed orthonormal coordinate system:

$$(\mu_r)_\ell \overset{\text{i.i.d.}}{\sim} \text{Unif}[-1, 1], \qquad r = 1, \dots, R, \quad \ell = 1, \dots, d.$$

Assume (191). Then for every fixed pair $r \neq j$, there exist constants $\alpha, c > 0$, depending only on ρ and $\eta_{\min}$, such that for all sufficiently large d,

$$\mathbb{P}(M_{rj} \leq \alpha d) \leq e^{-cd}. \tag{192}$$

Proof. Fix a pair $r \neq j$, and define

$$\Delta_{rj} \triangleq \mu_r - \mu_j.$$

Since the coordinates of μ_r and μ_j are drawn independently from $\text{Unif}[-1, 1]$, the coordinates of Δ_{rj} are independent. For each coordinate ℓ,

$$(\Delta_{rj})_\ell = (\mu_r)_\ell - (\mu_j)_\ell.$$

Thus $(\Delta_{rj})_\ell$ is centered and bounded by 2 in absolute value. Moreover,

$$\begin{aligned}
\mathbb{E}[(\Delta_{rj})_\ell^2] &= \mathbb{E}[(U - V)^2] \\
&= \mathbb{E}[U^2] + \mathbb{E}[V^2] - 2\mathbb{E}[U]\mathbb{E}[V] \\
&= \frac{1}{3} + \frac{1}{3} = \frac{2}{3},
\end{aligned}$$

where

$$U, V \overset{\text{i.i.d.}}{\sim} \text{Unif}[-1, 1].$$

Therefore

$$\mathbb{E}[\Delta_{rj}] = 0, \qquad \text{Cov}(\Delta_{rj}) = \frac{2}{3} I.$$

Let

$$P_I = \sum_{k \in I} v_k v_k^\top.$$

We first record the elementary properties of P_I directly from this formula. Since the vectors $v_1, \ldots, v_d$ form an orthonormal basis,

$$v_k^\top v_\ell = \begin{cases} 1, & k = \ell, \\ 0, & k \neq \ell. \end{cases}$$

First,

$$P_I^\top = \left(\sum_{k \in I} v_k v_k^\top \right)^\top = \sum_{k \in I} v_k v_k^\top = P_I.$$

Second,

$$\begin{aligned} P_I^2 &= \left(\sum_{k \in I} v_k v_k^\top \right) \left(\sum_{\ell \in I} v_\ell v_\ell^\top \right) \\ &= \sum_{k \in I} \sum_{\ell \in I} v_k (v_k^\top v_\ell) v_\ell^\top \\ &= \sum_{k \in I} v_k v_k^\top = P_I. \end{aligned}$$

Thus P_I is symmetric and satisfies $P_I^2 = P_I$.

Since

$$A = \sum_{k=1}^{d} \eta_k v_k v_k^\top,$$

we have, for every vector u,

$$u^\top (A - \eta_{\min} P_I) u = \sum_{k \in I} (\eta_k - \eta_{\min})(v_k^\top u)^2 + \sum_{k \notin I} \eta_k (v_k^\top u)^2.$$

Every term on the right-hand side is nonnegative, because $\eta_k \geq \eta_{\min}$ for $k \in I$ and $\eta_k \geq 0$ for all k. Therefore

$$A \succeq \eta_{\min} P_I.$$

Applying this with $u = \Delta_{rj}$ gives

$$M_{rj} = \Delta_{rj}^\top A \Delta_{rj} \geq \eta_{\min} \Delta_{rj}^\top P_I \Delta_{rj}. \tag{193}$$

We now study

$$Q \triangleq \Delta_{rj}^\top P_I \Delta_{rj}.$$

Using $P_I^\top = P_I$ and $P_I^2 = P_I$,

$$\|P_I \Delta_{rj}\|_2^2 = (P_I \Delta_{rj})^\top (P_I \Delta_{rj}) = \Delta_{rj}^\top P_I^\top P_I \Delta_{rj} = \Delta_{rj}^\top P_I^2 \Delta_{rj} = \Delta_{rj}^\top P_I \Delta_{rj} = Q.$$

Next, compute the expectation of Q. For any deterministic matrix B and any centered random vector X,

$$X^\top B X = \text{tr}(X^\top B X) = \text{tr}(B X X^\top).$$

Therefore

$$\mathbb{E}[X^\top BX] = \operatorname{tr}\big(B\,\mathbb{E}[XX^\top]\big) = \operatorname{tr}(B\operatorname{Cov}(X))\,.$$

Using this with $X = \Delta_{rj}$ and $B = P_I$, we get

$$\begin{aligned}\mathbb{E}[Q] &= \operatorname{tr}(P_I \operatorname{Cov}(\Delta_{rj}))\\ &= \operatorname{tr}\left(P_I \frac{2}{3} I\right)\\ &= \frac{2}{3}\operatorname{tr}(P_I).\end{aligned}$$

Now

$$\operatorname{tr}(P_I) = \operatorname{tr}\left(\sum_{k\in I} v_k v_k^\top\right) = \sum_{k\in I}\operatorname{tr}(v_k v_k^\top).$$

For each k,

$$\operatorname{tr}(v_k v_k^\top) = v_k^\top v_k = 1.$$

Hence

$$\operatorname{tr}(P_I) = |I|,$$

and therefore

$$\mathbb{E}[Q] = \frac{2}{3}|I|. \tag{194}$$

We now control the probability that Q is much smaller than its mean. We use the Hanson–Wright inequality (Rudelson & Vershynin, 2013) in the following standard form: if X has independent, centered, bounded coordinates and B is deterministic, then there exists a universal constant $c_{\mathrm{HW}} > 0$ such that, for every $t > 0$,

$$\mathbb{P}\big(|X^\top BX - \mathbb{E}[X^\top BX]| \ge t\big) \le 2\exp\left[-c_{\mathrm{HW}}\min\left(\frac{t^2}{\|B\|_F^2}, \frac{t}{\|B\|_{\mathrm{op}}}\right)\right]. \tag{195}$$

Here

$$\|B\|_{\mathrm{op}} = \sup_{\|x\|_2=1}\|Bx\|_2$$

and

$$\|B\|_F^2 = \operatorname{tr}(B^\top B).$$

We apply this with

$$X = \Delta_{rj}, \qquad B = P_I.$$

The assumptions apply because the coordinates of Δ_{rj} are independent, centered, and bounded by 2.

We now compute the two norms of P_I. First,

$$\|P_I\|_F^2 = \operatorname{tr}(P_I^\top P_I) = \operatorname{tr}(P_I^2) = \operatorname{tr}(P_I) = |I|.$$

Second, for any vector x,

$$P_I x = \sum_{k\in I} v_k(v_k^\top x).$$

Using orthonormality,

$$\|P_I x\|_2^2 = \sum_{k\in I}(v_k^\top x)^2 \le \sum_{k=1}^{d}(v_k^\top x)^2 = \|x\|_2^2.$$

Thus, for every x with $\|x\|_2 = 1$,

$$\|P_I x\|_2 \le 1,$$

so

$$\|P_I\|_{\mathrm{op}} \le 1.$$

If I is nonempty, choosing $x = v_k$ for some $k \in I$ gives $P_I x = x$, so in fact $\|P_I\|_{\mathrm{op}} = 1$.

By (194),

$$\mathbb{E}[Q] = \frac{2}{3}|I|.$$

The event

$$Q \leq \frac{1}{3}|I|$$

implies

$$\mathbb{E}[Q] - Q \geq \frac{2}{3}|I| - \frac{1}{3}|I| = \frac{1}{3}|I|.$$

Therefore

$$\left\{ Q \leq \frac{1}{3}|I| \right\} \subseteq \left\{ |Q - \mathbb{E}[Q]| \geq \frac{1}{3}|I| \right\}.$$

Using (195) with $t = |I|/3$, $B = P_I$, and the norm bounds above,

$$\begin{aligned}
\mathbb{P}\left(Q \leq \frac{1}{3}|I| \right) &\leq \mathbb{P}\left(|Q - \mathbb{E}[Q]| \geq \frac{1}{3}|I| \right) \\
&\leq 2\exp\left[-c_{\mathrm{HW}} \min\left(\frac{(|I|/3)^2}{|I|}, \frac{|I|/3}{1} \right) \right] \\
&= 2\exp\left[-c_{\mathrm{HW}} \min\left(\frac{|I|}{9}, \frac{|I|}{3} \right) \right].
\end{aligned}$$

Absorbing the numerical constants into a new universal constant $c_0 > 0$, we obtain

$$\mathbb{P}\left(Q \leq \frac{1}{3}|I| \right) \leq 2e^{-c_0|I|}. \tag{196}$$

Since $|I| \geq \rho d$, the event

$$Q \leq \frac{\rho d}{3}$$

is contained in the event

$$Q \leq \frac{|I|}{3}.$$

Therefore

$$\mathbb{P}\left(Q \leq \frac{\rho d}{3} \right) \leq 2e^{-c_0|I|} \leq 2e^{-c_0\rho d}.$$

Finally, by (193),

$$M_{rj} \geq \eta_{\min} Q.$$

Hence, if

$$M_{rj} \leq \frac{\rho\eta_{\min}}{3} d,$$

then necessarily

$$Q \leq \frac{\rho d}{3}.$$

Thus

$$\mathbb{P}\left(M_{rj} \leq \frac{\rho\eta_{\min}}{3} d \right) \leq \mathbb{P}\left(Q \leq \frac{\rho d}{3} \right) \leq 2e^{-c_0\rho d}.$$

For sufficiently large d, the prefactor 2 can be absorbed into the exponential by reducing the exponent constant. Thus the claim holds with

$$\alpha = \frac{\rho\eta_{\min}}{3}.$$

All-pairs consequence. If R is fixed, then a union bound gives

$$\mathbb{P}\Big(\min_{r\neq j} M_{rj} \leq \alpha d\Big) \leq \binom{R}{2} e^{-cd}. \tag{197}$$

Hence, for a fixed number of components R, the constant prefactor $\binom{R}{2}$ is asymptotically dominated by the exponential decay in d, guaranteeing that the linear-separation condition holds with overwhelming probability in high dimensions.

G.9 Transfer to the selected-label chain and local perturbation

We now transfer the bridge-marginal concentration bound to the selected-label chain.

Selected-chain transfer. Fix the true component $r^\star$, condition on y, and run the selected-label chain with the label forced to $J = r^\star$. The proof of Proposition 5.4 gives, at every step,

$$\mathbb{E}[X_s^{\mathrm{sel}} \mid y, J = r^\star] = u_{r^\star,s}(y). \tag{198}$$

Similarly, the covariance recursion used in the proof of Proposition 5.5 gives the intermediate-time domination

$$\mathrm{Cov}(X_s^{\mathrm{sel}} \mid y, J = r^\star) \preceq V_s, \qquad s = 1, \ldots, S-1. \tag{199}$$

Thus, relative to the component-conditioned bridge marginal

$$x_s \mid (y, C = r^\star) \sim \mathcal{N}(u_{r^\star,s}(y), V_s),$$

the selected chain has the same mean and no larger covariance.

Now fix $j \neq r^\star$ and $s \in \{1, \ldots, S-1\}$. By (170), the pairwise bridge log-likelihood ratio

$$Q_{r^\star j,s}(x, y) = \log \frac{p(x \mid y, C = r^\star)}{p(x \mid y, C = j)}$$

is affine in x. Therefore, applying this same affine functional to X_s^{sel}, equations (198)–(199) imply that

$$\mathbb{E}[Q_{r^\star j,s}(X_s^{\mathrm{sel}}, y) \mid y, J = r^\star] = \frac{1}{2} B_{r^\star j,s}, \tag{200}$$

and

$$\mathrm{Var}[Q_{r^\star j,s}(X_s^{\mathrm{sel}}, y) \mid y, J = r^\star] \leq B_{r^\star j,s}. \tag{201}$$

Since the selected-chain margin has the same mean and smaller variance, the Gaussian tail bound used in Part 2 of Theorem 1 applies with x_s replaced by X_s^{sel}. Hence, on the event

$$L_{r^\star j}^{(y)}(y) \geq T_0 \qquad \text{for all } j \neq r^\star,$$

we get, exactly as in (183),

$$\mathbb{P}\big(L_{r^\star j,s}(X_s^{\mathrm{sel}}, y) < T_1 \mid C = r^\star, \mathcal{E}_y(T_0), y\big) \leq e^{-(T_0 - T_1)}. \tag{202}$$

A union bound over $j \neq r^\star$ and $s = 1, \ldots, S-1$, together with the measurement part of Theorem 1, gives the selected-chain analogue of (176). Choosing

$$T_1 = T_\varepsilon, \qquad T_0 = \frac{\alpha d}{4},$$

as in Corollary 1 yields

$$\mathbb{P}\big(\mathcal{E}_{r^\star,\varepsilon}^{\mathrm{sel}} \mid C = r^\star\big) \geq 1 - KSe^{-cd}. \tag{203}$$

which proves Theorem G.2.

It remains to justify the sampled-label statement. Since $J \sim \gamma_{\cdot|y}(y)$,

$$\mathbb{P}(J \neq r^\star \mid C = r^\star) = \mathbb{E}[1 - \gamma_{r^\star|y}(y) \mid C = r^\star].$$

On the event $\gamma_{r^\star|y}(y) \geq 1 - \varepsilon$, the integrand is at most ε, and otherwise it is at most 1. Therefore,

$$\mathbb{P}(J \neq r^\star \mid C = r^\star) \leq \varepsilon + \mathbb{P}(\gamma_{r^\star|y}(y) < 1 - \varepsilon \mid C = r^\star) \leq \varepsilon + KSe^{-cd}.$$

Equivalently,

$$\mathbb{P}(J = r^\star \mid C = r^\star) \geq 1 - \varepsilon - KSe^{-cd}.$$

Local exact-versus-frozen perturbation bound. Finally, we record the deterministic one-step consequence of high responsibility. Fix a component r, a step s, and a state (x, y). If

$$\gamma_{r|s}(x, y) \geq 1 - \varepsilon,$$

then

$$\begin{aligned} \left\|\hat{x}_{0,s}^{\mathrm{MMSE}}(x, y) - \mu_{r|s}(x, y)\right\|_2 &= \left\|\sum_{j \neq r} \gamma_{j|s}(x, y)\big(\mu_{j|s}(x, y) - \mu_{r|s}(x, y)\big)\right\|_2 \\ &\leq \varepsilon \max_{j \neq r} \left\|\mu_{j|s}(x, y) - \mu_{r|s}(x, y)\right\|_2 . \end{aligned} \tag{204}$$

Coupling the exact MoG update and the frozen-label update with the same Gaussian innovation z_s, the shared terms $b_s y + c_s x + \sigma_s z_s$ cancel, and therefore

$$\begin{aligned} &\left\|\big(a_s \hat{x}_{0,s}^{\mathrm{MMSE}}(x, y) + b_s y + c_s x + \sigma_s z_s\big) - \big(a_s \mu_{r|s}(x, y) + b_s y + c_s x + \sigma_s z_s\big)\right\|_2 \\ &\quad \leq |a_s| \varepsilon \max_{j \neq r} \left\|\mu_{j|s}(x, y) - \mu_{r|s}(x, y)\right\|_2 . \end{aligned} \tag{205}$$

Thus, along states where the frozen label has high responsibility, one exact MoG step and one frozen-label step are close. A full pathwise tracking theorem would require additional stability assumptions controlling the accumulation of these local errors, so we validate the whole-chain discrepancy empirically.

H MoG analysis of conditional DDIM with an oracle posterior denoiser

This appendix shows that the selected-label MoG analysis is not specific to the Brownian-bridge machinery. We repeat the same pipeline for a standard non-bridge conditional DDIM sampler. The forward process is the usual data-to-noise diffusion marginal, while the observation y is incorporated only through the denoiser.

H.1 Conditional DDIM model and notation

In contrast to BBDM, DDIM does not use a bridge endpoint at y. Its forward marginal is

$$q_{\mathrm{D}}(x_s \mid x_0) = \mathcal{N}\big(x_s; \sqrt{\bar{\alpha}_s}\, x_0, (1-\bar{\alpha}_s)I\big)\,, \tag{206}$$

where $s \in \{0, \dots, S\}$, $\bar{\alpha}_0 = 1$, and $\bar{\alpha}_S$ is close to zero.
The measurement model and MoG prior are unchanged:

$$y \mid x_0 \sim \mathcal{N}(Hx_0, \sigma_y^2 I), \qquad p(x_0) = \sum_{r=1}^{R} \pi_r \mathcal{N}(x_0; \mu_r, \Sigma_r). \tag{207}$$

Given an estimate $\hat{x}_0(x_s, y)$, the deterministic DDIM update can be written in the affine form

$$x_{s-1} = a_s^{\mathrm{D}} x_s + b_s^{\mathrm{D}} \hat{x}_0(x_s, y), \tag{208}$$

where

$$a_s^{\mathrm{D}} = \frac{\sqrt{1-\bar{\alpha}_{s-1}}}{\sqrt{1-\bar{\alpha}_s}}, \tag{209}$$

$$b_s^{\mathrm{D}} = \sqrt{\bar{\alpha}_{s-1}} - \frac{\sqrt{\bar{\alpha}_s}\sqrt{1-\bar{\alpha}_{s-1}}}{\sqrt{1-\bar{\alpha}_s}}. \tag{210}$$

The sampler is initialized as in standard DDIM with

$$x_S \sim \mathcal{N}(0, I), \tag{211}$$

independently of y.

H.2 Exact posterior and oracle conditional DDIM denoiser

We now compute the posterior $p(x_0 \mid x_s, y)$ under the DDIM forward marginal. The measurement-only posterior within a fixed component is identical to the one used in the BBDM analysis:

$$\Sigma_{r|y}^{-1} = \Sigma_r^{-1} + \frac{1}{\sigma_y^2} H^\top H, \qquad \mu_{r|y} = \Sigma_{r|y}\left(\Sigma_r^{-1}\mu_r + \frac{1}{\sigma_y^2} H^\top y\right). \tag{212}$$

Thus,

$$x_0 \mid (y, C = r) \sim \mathcal{N}(\mu_{r|y}, \Sigma_{r|y}). \tag{213}$$

Conditioned on $(y, C = r)$, the DDIM latent x_s is an affine Gaussian observation of x_0:

$$x_s = \sqrt{\bar{\alpha}_s}\, x_0 + \sqrt{1-\bar{\alpha}_s}\, \epsilon.$$

Applying the Gaussian conditioning identity from Appendix A, with $A = \sqrt{\bar{\alpha}_s} I$ and $R = (1-\bar{\alpha}_s)I$, gives

$$p_{\mathrm{D}}(x_0 \mid x_s, y, C = r) = \mathcal{N}\Big(x_0; \mu_{r|s}^{\mathrm{D}}(x_s, y), \Sigma_{r|s}^{\mathrm{D}}\Big)\,, \tag{214}$$

where

$$\Sigma_{r|s}^{\mathrm{D}} = \left(\Sigma_{r|y}^{-1} + \frac{\bar{\alpha}_s}{1-\bar{\alpha}_s} I\right)^{-1}, \tag{215}$$

$$\mu^{\mathrm{D}}_{r|s}(x_s, y) = \Sigma^{\mathrm{D}}_{r|s}\left(\Sigma^{-1}_{r|y}\mu_{r|y} + \frac{\sqrt{\bar{\alpha}_s}}{1-\bar{\alpha}_s}x_s\right). \tag{216}$$

Equivalently, using (212),

$$\mu^{\mathrm{D}}_{r|s}(x_s, y) = \Sigma^{\mathrm{D}}_{r|s}\left(\Sigma^{-1}_{r}\mu_{r} + \frac{1}{\sigma_y^2}H^\top y + \frac{\sqrt{\bar{\alpha}_s}}{1-\bar{\alpha}_s}x_s\right). \tag{217}$$

The component probabilities are also obtained exactly. First,

$$\gamma_{r|y}(y) = \frac{\pi_r \mathcal{N}\big(y; H\mu_r, H\Sigma_r H^\top + \sigma_y^2 I\big)}{\sum_{j=1}^{R}\pi_j \mathcal{N}\big(y; H\mu_j, H\Sigma_j H^\top + \sigma_y^2 I\big)}. \tag{218}$$

Next, conditioned on $(y, C = r)$,

$$p_{\mathrm{D}}(x_s \mid y, C = r) = \mathcal{N}\big(x_s; \sqrt{\bar{\alpha}_s}\,\mu_{r|y}, \bar{\alpha}_s\Sigma_{r|y} + (1-\bar{\alpha}_s)I\big). \tag{219}$$

Therefore the DDIM responsibilities are

$$\gamma^{\mathrm{D}}_{r|s}(x_s, y) \triangleq \mathbb{P}(C = r \mid x_s, y) = \frac{\gamma_{r|y}(y)p_{\mathrm{D}}(x_s \mid y, C = r)}{\sum_{j=1}^{R}\gamma_{j|y}(y)p_{\mathrm{D}}(x_s \mid y, C = j)}. \tag{220}$$

Combining the component posteriors and responsibilities yields

$$p_{\mathrm{D}}(x_0 \mid x_s, y) = \sum_{r=1}^{R}\gamma^{\mathrm{D}}_{r|s}(x_s, y)\mathcal{N}\Big(x_0; \mu^{\mathrm{D}}_{r|s}(x_s, y), \Sigma^{\mathrm{D}}_{r|s}\Big). \tag{221}$$

Hence the oracle conditional DDIM denoiser is

$$\hat{x}^{\mathrm{D}}_{0,s}(x_s, y) = \mathbb{E}[x_0 \mid x_s, y] = \sum_{r=1}^{R}\gamma^{\mathrm{D}}_{r|s}(x_s, y)\mu^{\mathrm{D}}_{r|s}(x_s, y). \tag{222}$$

H.3 Exact conditional DDIM reverse dynamics and loss of global affinity

Substituting (222) into the DDIM update (208) gives

$$x_{s-1} = a^{\mathrm{D}}_s x_s + b^{\mathrm{D}}_s \sum_{r=1}^{R}\gamma^{\mathrm{D}}_{r|s}(x_s, y)\mu^{\mathrm{D}}_{r|s}(x_s, y). \tag{223}$$

Using (216), this becomes

$$\begin{aligned} x_{s-1} = &\left(a^{\mathrm{D}}_s I + b^{\mathrm{D}}_s \frac{\sqrt{\bar{\alpha}_s}}{1-\bar{\alpha}_s}\sum_{r=1}^{R}\gamma^{\mathrm{D}}_{r|s}(x_s, y)\Sigma^{\mathrm{D}}_{r|s}\right)x_s \\ &+ b^{\mathrm{D}}_s \sum_{r=1}^{R}\gamma^{\mathrm{D}}_{r|s}(x_s, y)\Sigma^{\mathrm{D}}_{r|s}\Sigma^{-1}_{r|y}\mu_{r|y}. \end{aligned} \tag{224}$$

The structural obstruction is identical to the one found for BBDM. If $R = 1$, then $\gamma^{\mathrm{D}}_{1|s} \equiv 1$ and the reverse update is globally affine. If $R > 1$, the coefficients in (224) depend on the current state through $\gamma^{\mathrm{D}}_{r|s}(x_s, y)$. Therefore, the exact MoG conditional DDIM chain cannot be unrolled by simply composing one fixed sequence of linear maps.

H.4 Selected-label approximation for conditional DDIM

We now apply the same selected-label approximation used for BBDM. Draw

$$\mathbb{P}(J = r \mid y) = \gamma_{r|y}(y), \tag{225}$$

and keep J fixed along the DDIM reverse chain. Conditioned on $J = r$, replace the mixture denoiser by the component-conditioned posterior mean

$$\hat{x}_0(x_s, y) = \mu^{\mathrm{D}}_{r|s}(x_s, y).$$

Inserting (216) into (208) gives the fixed-label conditional DDIM recursion

$$x_{s-1} = \left(a^{\mathrm{D}}_s I + b^{\mathrm{D}}_s \frac{\sqrt{\bar{\alpha}_s}}{1 - \bar{\alpha}_s} \Sigma^{\mathrm{D}}_{r|s} \right) x_s + b^{\mathrm{D}}_s \Sigma^{\mathrm{D}}_{r|s} \Sigma^{-1}_{r|y} \mu_{r|y}. \tag{226}$$

The update is affine once the label is frozen. Unlike the BBDM update (22), there is no explicit bridge endpoint term proportional to y and no reverse innovation term. The measurement affects the recursion only through the posterior mean $\mu_{r|y}$ and precision $\Sigma^{-1}_{r|y}$.

H.5 Component-wise basis and diagonal scalar recursions

For each component r, use the same problem-adapted basis as in Section 5:

$$\Sigma^{-1}_{r|y} = U_r \operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d}) U_r^\top. \tag{227}$$

Since $\Sigma^{\mathrm{D}}_{r|s}$ is obtained by adding a scalar multiple of the identity to $\Sigma^{-1}_{r|y}$ and inverting, the same basis diagonalizes all DDIM posterior covariances:

$$\Sigma^{\mathrm{D}}_{r|s} = U_r \Lambda^{\mathrm{D}}_{r|s} U_r^\top, \tag{228}$$

where

$$\Lambda^{\mathrm{D}}_{r|s} = \operatorname{diag}\left(\frac{1}{\lambda_{r,1} + \frac{\bar{\alpha}_s}{1-\bar{\alpha}_s}}, \ldots, \frac{1}{\lambda_{r,d} + \frac{\bar{\alpha}_s}{1-\bar{\alpha}_s}} \right). \tag{229}$$

Projecting (226) onto this basis and defining

$$x^{U_r}_s = U_r^\top x_s, \qquad \mu^{U_r}_{r|y} = U_r^\top \mu_{r|y}, \qquad x^{U_r}_S = U_r^\top x_S,$$

we obtain

$$x^{U_r}_{s-1} = G^{\mathrm{D}}_r(s) x^{U_r}_s + M^{\mathrm{D}}_r(s) \mu^{U_r}_{r|y}, \tag{230}$$

with diagonal matrices

$$G^{\mathrm{D}}_r(s) = a^{\mathrm{D}}_s I + b^{\mathrm{D}}_s \frac{\sqrt{\bar{\alpha}_s}}{1 - \bar{\alpha}_s} \Lambda^{\mathrm{D}}_{r|s}, \tag{231}$$

$$M^{\mathrm{D}}_r(s) = b^{\mathrm{D}}_s \Lambda^{\mathrm{D}}_{r|s} \operatorname{diag}(\lambda_{r,1}, \ldots, \lambda_{r,d}). \tag{232}$$

Thus, conditioned on $J = r$, the DDIM chain also decouples into d scalar recursions in the basis U_r.

H.6 Unrolled reconstruction law

Starting from $x_S \sim \mathcal{N}(0, I)$, and using the convention that empty products are the identity, the recursion (230) gives

$$\hat{x}^{U_r,(r)}_{0,\mathrm{D}} = \left(\prod_{s'=1}^{S} G^{\mathrm{D}}_r(s') \right) x^{U_r}_S + D^{(r),\mathrm{D}}_2 \mu^{U_r}_{r|y}, \tag{233}$$

where

$$D_2^{(r),\mathrm{D}} = \sum_{i=1}^{S} \left(\prod_{j=1}^{i-1} G_r^{\mathrm{D}}(j) \right) M_r^{\mathrm{D}}(i). \tag{234}$$

The matrix $D_2^{(r),\mathrm{D}}$ and the product $\prod_{s'=1}^{S} G_r^{\mathrm{D}}(s')$ are diagonal. Since $x_S^{U_r} \sim \mathcal{N}(0, I)$, the fixed-label conditional DDIM output satisfies

$$p_{\mathrm{Select}}^{\mathrm{D}}(\hat{x}_0 \mid y, J = r) = \mathcal{N}\Big(\hat{x}_0; \mu_{\mathrm{DDIM}}^{(r)}(y), \Sigma_{\mathrm{DDIM}}^{(r)}\Big), \tag{235}$$

with

$$\mu_{\mathrm{DDIM}}^{(r)}(y) = U_r D_2^{(r),\mathrm{D}} \mu_{r|y}^{U_r}, \tag{236}$$

and

$$\Sigma_{\mathrm{DDIM}}^{(r)} = U_r \left[\left(\prod_{s'=1}^{S} G_r^{\mathrm{D}}(s') \right) \left(\prod_{s'=1}^{S} G_r^{\mathrm{D}}(s') \right)^{\top} \right] U_r^{\top}. \tag{237}$$

Removing the conditioning on the selected label gives the mixture law

$$p_{\mathrm{Select}}^{\mathrm{D}}(\hat{x}_0 \mid y) = \sum_{r=1}^{R} \gamma_{r|y}(y) \mathcal{N}\Big(\hat{x}_0; \mu_{\mathrm{DDIM}}^{(r)}(y), \Sigma_{\mathrm{DDIM}}^{(r)}\Big). \tag{238}$$

This is the conditional-DDIM analogue of Corollary 5.3. The weights are the same measurement-posterior weights as in the true posterior $p(x_0 \mid y)$, while the Gaussian attached to each component is induced by the DDIM reverse dynamics.

We now use the same endpoint-idealized viewpoint as in the BBDM mean-exactness analysis. In practice, DDIM schedules take $\bar{\alpha}_S$ close to zero. Approximating this terminal value by

$$\bar{\alpha}_S = 0,$$

the initialization $x_S \sim \mathcal{N}(0, I)$ satisfies

$$\mathbb{E}[x_S \mid y, J = r] = 0 = \sqrt{\bar{\alpha}_S}\, \mu_{r|y}.$$

Repeating the same backward-induction argument used for the selected-label BBDM mean-exactness result gives

$$\mathbb{E}[\hat{x}_0 \mid y, J = r] = \mu_{r|y}. \tag{239}$$

Equivalently, in the unrolled representation,

$$D_2^{(r),\mathrm{D}} = I. \tag{240}$$

Therefore, under the endpoint approximation $\bar{\alpha}_S \simeq 0$, the fixed-label conditional DDIM law simplifies to

$$p_{\mathrm{Select}}^{\mathrm{D}}(\hat{x}_0 \mid y, J = r) = \mathcal{N}\Big(\hat{x}_0; \mu_{r|y}, \Sigma_{\mathrm{DDIM}}^{(r)}\Big), \tag{241}$$

and the selected-label mixture law becomes

$$p_{\mathrm{Select}}^{\mathrm{D}}(\hat{x}_0 \mid y) = \sum_{r=1}^{R} \gamma_{r|y}(y) \mathcal{N}\Big(\hat{x}_0; \mu_{r|y}, \Sigma_{\mathrm{DDIM}}^{(r)}\Big). \tag{242}$$

Thus, after the same type of endpoint idealization used for BBDM, the selected-label conditional DDIM sampler is also component-wise mean exact. The remaining discrepancy from the true posterior is therefore in the component covariances.

H.7 Matched posterior comparison and DDIM schedule objectives

The true measurement posterior remains

$$p(x_0 \mid y) = \sum_{r=1}^{R} \gamma_{r|y}(y) \mathcal{N}(x_0; \mu_{r|y}, \Sigma_{r|y}). \tag{243}$$

Since (242) and (243) share the same component weights, the same matched-index coupling used in Section 6 gives

$$\begin{aligned} &W_2^2\big(p(x_0 \mid y), p^{\mathrm{D}}_{\mathrm{Select}}(\hat{x}_0 \mid y)\big) \\ &\leq \sum_{r=1}^{R} \gamma_{r|y}(y)\, W_2^2\Big(\mathcal{N}(\mu_{r|y}, \Sigma_{r|y}), \mathcal{N}(\mu_{r|y}, \Sigma^{(r)}_{\mathrm{DDIM}})\Big). \end{aligned} \tag{244}$$

Define

$$\sigma_{\mathrm{DDIM},r,k} = \left| \left[\prod_{s'=1}^{S} G_r^{\mathrm{D}}(s') \right]_{kk} \right|. \tag{245}$$

In the U_r basis, the true posterior covariance is

$$\mathrm{diag}\left(\frac{1}{\lambda_{r,1}}, \ldots, \frac{1}{\lambda_{r,d}} \right),$$

while the DDIM-induced covariance is

$$\mathrm{diag}\big(\sigma^2_{\mathrm{DDIM},r,1}, \ldots, \sigma^2_{\mathrm{DDIM},r,d}\big).$$

Because the means match under the endpoint approximation, the matched Gaussian Wasserstein cost contains no mean term:

$$\begin{aligned} &W_2^2\Big(\mathcal{N}(\mu_{r|y}, \Sigma_{r|y}), \mathcal{N}(\mu_{r|y}, \Sigma^{(r)}_{\mathrm{DDIM}})\Big) \\ &= \sum_{k=1}^{d} \left(\sigma_{\mathrm{DDIM},r,k} - \frac{1}{\sqrt{\lambda_{r,k}}} \right)^2. \end{aligned} \tag{246}$$

Averaging over y as in Appendix E and using

$$\mathbb{E}_{p(y)} \sum_{r=1}^{R} \gamma_{r|y}(y) f_r(y) = \sum_{r=1}^{R} \pi_r \mathbb{E}_{p(y|C=r)} f_r(y),$$

we obtain the endpoint-approximated DDIM Wasserstein schedule objective

$$J^{W_2}_{\mathrm{MOG,DDIM}} = \sum_{r=1}^{R} \pi_r \sum_{k=1}^{d} \left(\sigma_{\mathrm{DDIM},r,k} - \frac{1}{\sqrt{\lambda_{r,k}}} \right)^2. \tag{247}$$

The corresponding matched-label product-coupling expected-MSE objective is

$$J^{\mathrm{MSE}}_{\mathrm{MOG,DDIM}} = \sum_{r=1}^{R} \pi_r \sum_{k=1}^{d} \left(\sigma^2_{\mathrm{DDIM},r,k} + \frac{1}{\lambda_{r,k}} \right). \tag{248}$$

I Additional experimental figures

Figure 5 provides the full high-dimensional validation plot discussed in Section 8. Figure 6 shows the four one-dimensional MNIST schedule sweeps discussed in Section 8. Figure 7 shows representative MNIST reconstructions under increasingly rich fitted MoG priors for the deblurring inverse problem. Table 5 shows the FFHQ inverse-problem results for different numbers of sampling steps. Figures 8, 9, and 10 show qualitative FFHQ reconstructions for the blur setting $(V, \sigma_y) = (0.03, 0.10)$, the 8× super-resolution task, and the $p = 0.125$ distributed inpainting inverse problem.

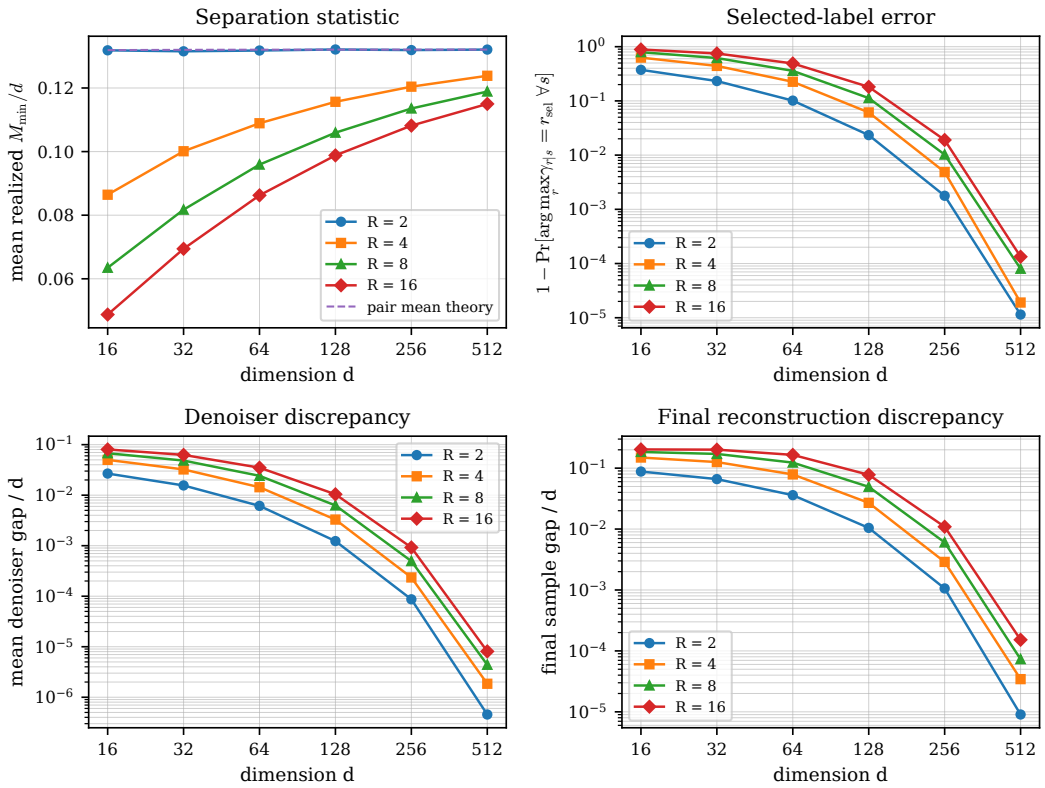

Figure 5: High-dimensional validation of the frozen-selection approximation in the controlled shared-covariance MoG setting. For each ambient dimension $d \in \{16, 32, 64, 128, 256, 512\}$ and number of components $R \in \{2, 4, 8, 16\}$, the component means are sampled independently as $\mu_{r,k} \sim \text{Unif}[-1, 1]$. The inverse problem is denoising with $H = I$ and $\sigma_y = 2$, and all components share the diagonal covariance $\Sigma = \text{diag}(\text{geomspace}(0.5, 2, d))$. The BBDM reverse chains use $S = 20$ steps. In accordance with the selected-label approximation, the frozen label is sampled from the measurement posterior, $r_{\text{sel}} \sim \text{Categorical}(\gamma_{\cdot|y}(y))$, and is then kept fixed throughout the reverse chain. The exact MoG chain and the frozen selected-label chain are run with the same Gaussian innovations. The top-left panel reports the realized minimum measurement separation $M_{\min}/d$, where $M_{\min} = \min_{r \neq j}(\mu_r - \mu_j)^\top (\Sigma + \sigma_y^2 I)^{-1}(\mu_r - \mu_j)$, together with the pairwise mean-theory reference. The remaining panels show the selected-label error, a time-averaged local denoiser discrepancy per dimension, and the final reconstruction discrepancy per dimension. The states that both denoisers are evaluated on are the states of the frozen path. The selected-label error measures whether the sampled frozen label remains the dominant responsibility along the frozen chain. As the dimension increases, the measurement posterior becomes increasingly one-hot, the sampled label stabilizes, and the exact and frozen chains become increasingly close, consistent with the responsibility-concentration mechanism predicted by the theory.

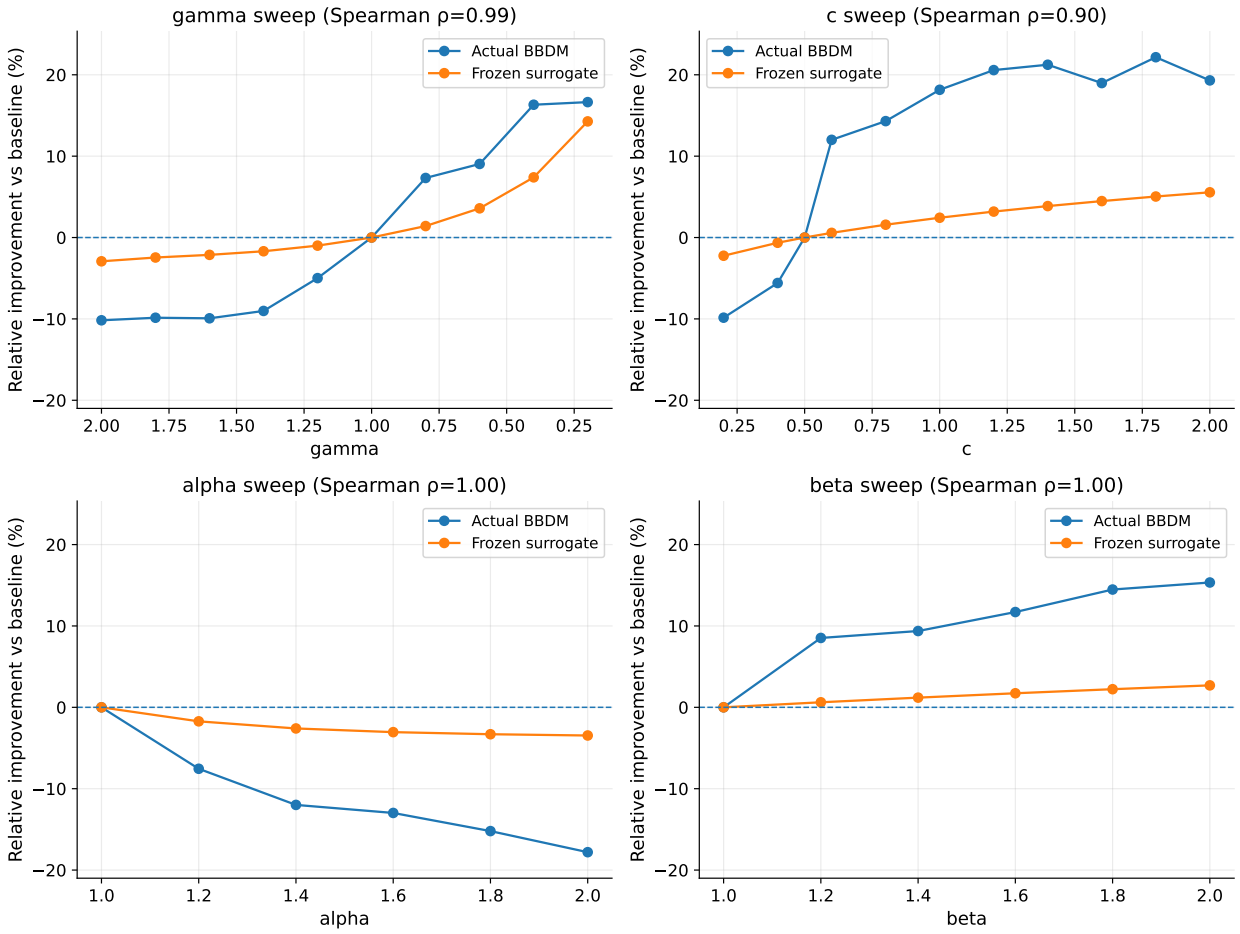

Figure 6: One-dimensional MNIST schedule sweeps for the baseline blur setting $(V, \sigma_y) = (0.10, 0.10)$. In each panel, one parameter in $(\alpha, \beta, c, \gamma)$ is varied while the other three are fixed. The curves compare relative improvement in reconstruction MSE to the ground-truth clean digit over the MNIST test set for a trained BBDM and the frozen selected-label surrogate fitted with 70 Gaussians per digit. Both methods favor the same directions: $\alpha \downarrow$, $\beta \uparrow$, $c \uparrow$, and $\gamma \downarrow$. The corresponding Spearman rank correlations between the surrogate and trained-BBDM sweeps are $\rho_\alpha = 1.00$, $\rho_\beta = 1.00$, $\rho_c = 0.90$, and $\rho_\gamma = 0.99$. The comparison tests agreement in schedule ranking and favorable direction, rather than equality of the absolute MSE values.

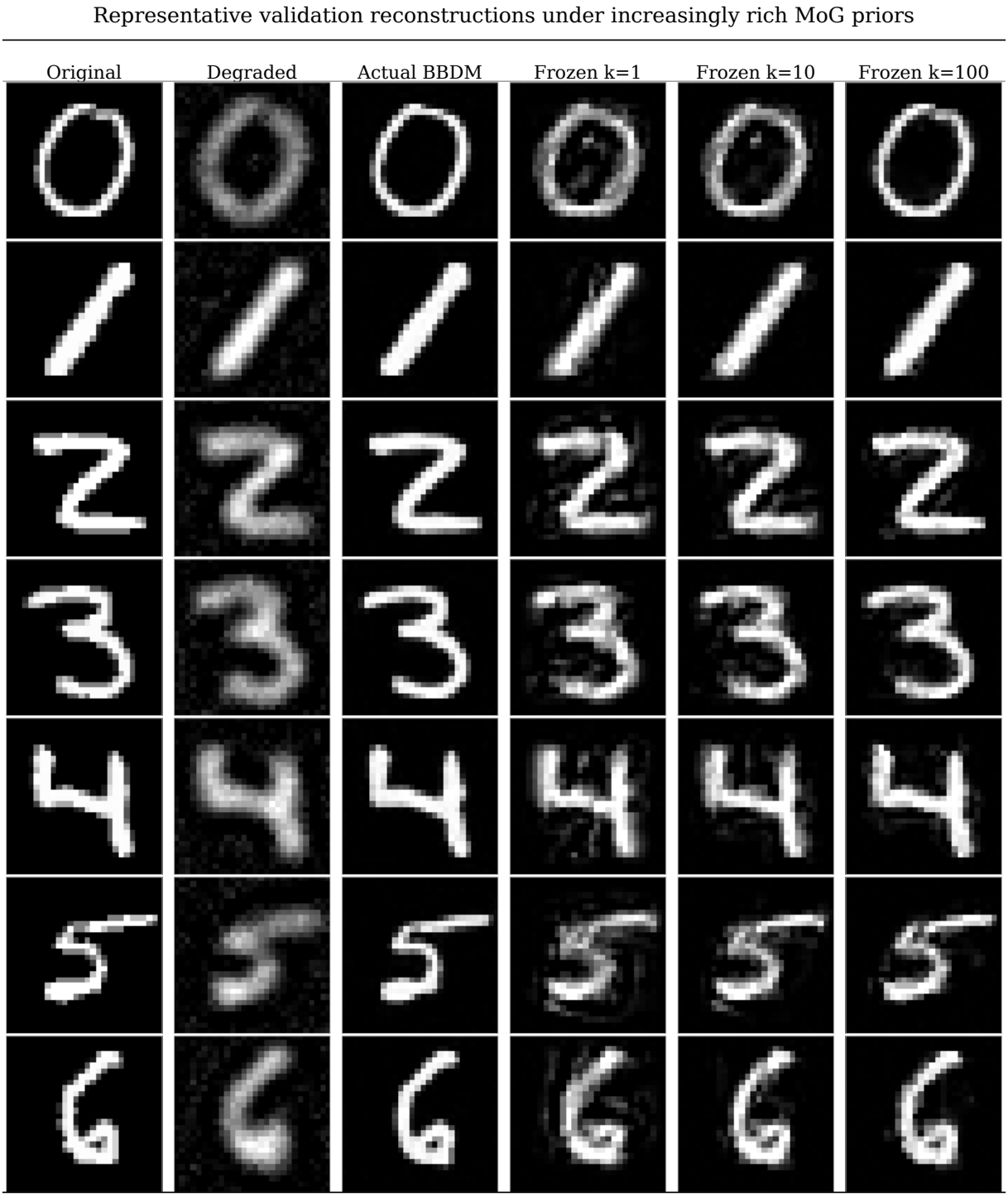

Figure 7: Effect of prior richness in the frozen selected-label surrogate for MNIST deblurring. The number of fitted Gaussian components is increased from $k = 1$ to $k = 100$ per digit while the reconstruction rule and degradation setting are kept fixed. Increasing the richness of the fitted MoG produces cleaner and more stable reconstructions, supporting the use of a sufficiently expressive prior in the surrogate schedule search. These are outputs of the mathematical surrogate, not separately trained BBDM models.

Table 5: FFHQ schedule comparison across sampling budgets and inverse-problem settings.

S	Metric	Sch.	B_1	B_2	B_3	B_4	SR_4	SR_8	I_{25}	$I_{12.5}$
20	FID↓	def.	14.544	19.785	9.938	21.876	14.520	26.046	17.609	19.742
		MSE	21.874	25.082	13.165	26.871	23.968	34.715	19.159	22.853
		W_2	**9.533**	**14.666**	**6.493**	**12.257**	**11.201**	**17.778**	**8.936**	**9.914**
	LPIPS↓	def.	0.0850	0.1447	0.0450	0.1133	0.0990	0.1820	0.0860	0.1152
		MSE	0.1041	0.1741	0.0511	0.1225	0.1279	0.2069	0.0855	0.1221
		W_2	**0.0741**	**0.1382**	**0.0366**	**0.0955**	**0.0919**	**0.1766**	**0.0725**	**0.1128**
	PSNR↑	def.	31.277	27.970	34.066	30.209	29.614	25.745	30.487	28.017
		MSE	**31.814**	**28.579**	**34.883**	**30.834**	**30.205**	**26.236**	**31.316**	**28.819**
		W_2	30.882	28.007	33.467	29.797	29.384	25.588	28.618	25.873
	SSIM↑	def.	0.878	0.795	0.923	0.857	0.853	0.743	0.876	0.828
		MSE	**0.888**	**0.807**	**0.934**	**0.868**	**0.866**	**0.758**	**0.889**	**0.844**
		W_2	0.866	0.788	0.913	0.844	0.837	0.736	0.839	0.771
50	FID↓	def.	13.223	18.109	9.143	20.148	13.515	23.221	16.364	17.789
		MSE	21.878	25.066	13.169	26.871	23.997	34.823	19.199	22.861
		W_2	**5.805**	**10.041**	**3.653**	**7.134**	**6.545**	**12.491**	**5.238**	**8.874**
	LPIPS↓	def.	0.0784	0.1351	0.0427	0.1067	0.0929	0.1715	0.0833	**0.1113**
		MSE	0.1039	0.1735	0.0511	0.1223	0.1277	0.2055	0.0855	0.1211
		W_2	**0.0606**	**0.1195**	**0.0297**	**0.0804**	**0.0771**	**0.1586**	**0.0663**	0.1186
	PSNR↑	def.	30.973	27.608	33.820	29.886	29.289	25.369	30.170	27.641
		MSE	**31.800**	**28.561**	**34.876**	**30.822**	**30.187**	**26.138**	**31.315**	**28.803**
		W_2	30.392	27.623	32.898	29.334	28.905	25.146	28.201	25.222
	SSIM↑	def.	0.873	0.786	0.920	0.851	0.841	0.731	0.871	0.819
		MSE	**0.888**	**0.807**	**0.934**	**0.869**	**0.861**	**0.756**	**0.890**	**0.845**
		W_2	0.853	0.772	0.903	0.826	0.822	0.710	0.823	0.747
100	FID↓	def.	12.369	17.133	8.656	18.597	12.800	20.845	15.347	16.390
		MSE	21.880	25.067	13.159	26.811	23.968	34.573	19.186	22.867
		W_2	**5.101**	**9.352**	**3.062**	**6.170**	**5.366**	**11.832**	**4.781**	**10.078**
	LPIPS↓	def.	0.0744	0.1309	0.0414	0.1019	0.0891	0.1653	0.0812	**0.1084**
		MSE	0.1040	0.1734	0.0511	0.1223	0.1277	0.2052	0.0855	0.1211
		W_2	**0.0572**	**0.1148**	**0.0281**	**0.0763**	**0.0729**	**0.1540**	**0.0659**	0.1229
	PSNR↑	def.	30.738	27.391	33.604	29.658	29.077	25.163	29.963	27.415
		MSE	**31.794**	**28.551**	**34.873**	**30.816**	**30.176**	**26.099**	**31.313**	**28.795**
		W_2	30.144	27.430	32.629	29.098	28.666	24.955	28.000	24.955
	SSIM↑	def.	0.869	0.780	0.917	0.846	0.836	0.722	0.867	0.813
		MSE	**0.888**	**0.807**	**0.934**	**0.869**	**0.860**	**0.756**	**0.890**	**0.845**
		W_2	0.845	0.762	0.898	0.817	0.812	0.699	0.814	0.734

Settings: B_1: blur with $(V, \sigma_y) = (0.10, 0.10)$; B_2: blur with $(V, \sigma_y) = (0.03, 0.10)$; B_3: blur with $(V, \sigma_y) = (0.30, 0.10)$; B_4: blur with $(V, \sigma_y) = (0.10, 0.20)$; SR_4: $4\times$ super-resolution with $\sigma_y = 0.10$; SR_8: $8\times$ super-resolution with $\sigma_y = 0.10$; I_{25}: inpainting with $(p, \sigma_y) = (0.25, 0.10)$; $I_{12.5}$: inpainting with $(p, \sigma_y) = (0.125, 0.10)$.

Figure 8: Qualitative FFHQ blur reconstructions for $(V, \sigma_y) = (0.03, 0.10)$. Columns show the original image, degraded observation, MSE-oriented reconstruction, default reconstruction, and W_2-oriented reconstruction.

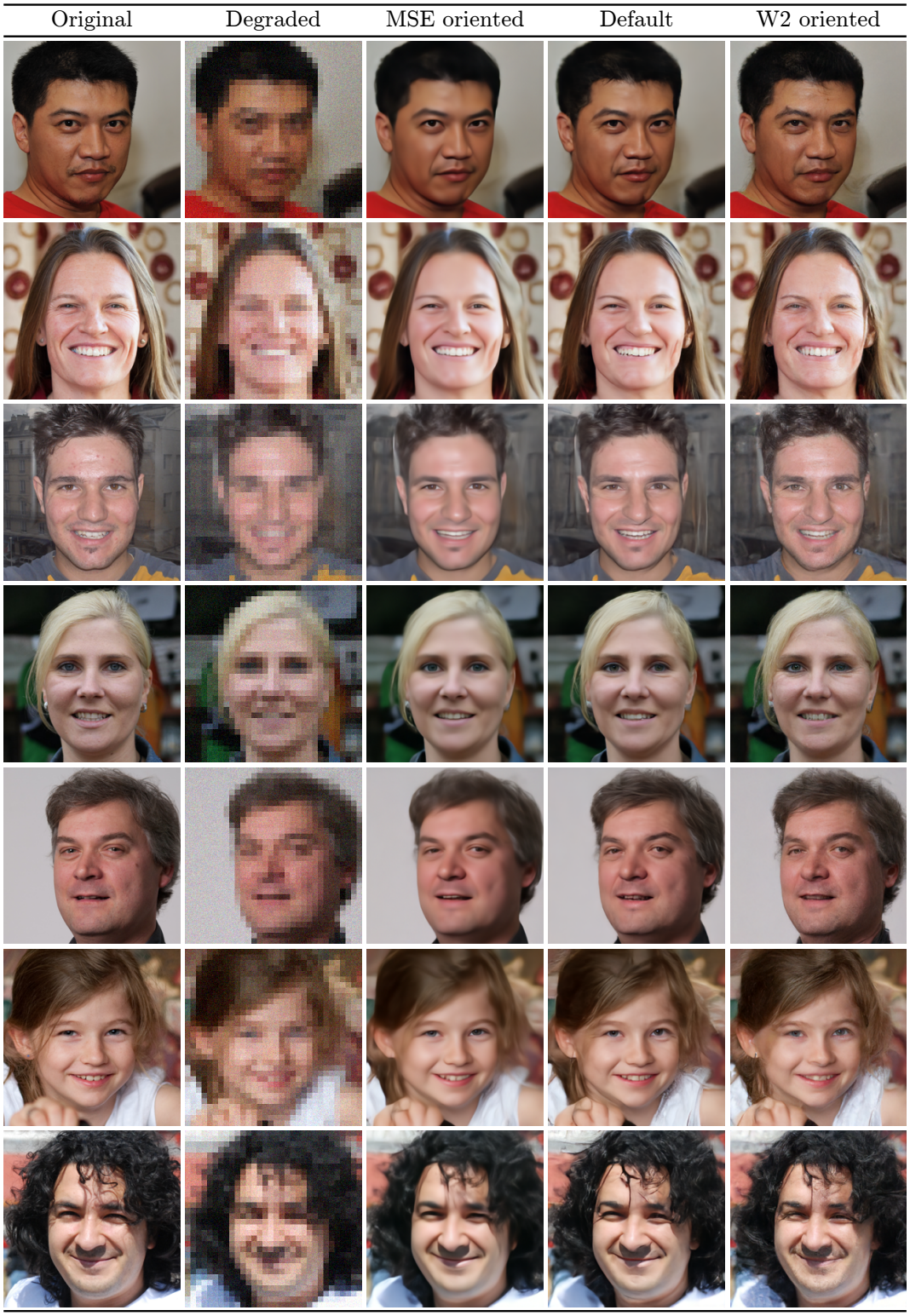

Figure 9: Qualitative FFHQ 8× super-resolution reconstructions with $\sigma_y = 0.1$. Columns show the original image, degraded observation, MSE-oriented reconstruction, default reconstruction, and W_2-oriented reconstruction.

Figure 10: Qualitative FFHQ distributed inpainting reconstructions for $p = 0.125$ and $\sigma_y = 0.1$. Columns show the original image, degraded observation, MSE-oriented reconstruction, default reconstruction, and W_2-oriented reconstruction.