Noise-Level KL Rates for Multi-Marginal Schrödinger Bridge Surrogates

Hui Chen[1*] **Shen Xu**[2*] **Vikas Singh**[1]

https://github.com/vsingh-group/slice2vol-bridge

Abstract

It is common to use 2D diffusion priors for problems where the underlying object is 3D (e.g., medical imaging). One uses a 2D diffusion model to reweight a prior over 3D objects, and then vary the strength of this reweighting with an EDM noise schedule. It works well empirically. But we know very little about what it converges to and how fast. Here, our reference will be the multi-marginal Schrödinger bridge $P^\star$. It matches the prescribed 2D marginals for all N rendered slices while minimizing KL divergence from the 3D prior. Computing the bridge is intractable, so we study how closely this reweighted prior (or annealed surrogate) Π_τ approximates it. Our main idea is that the discrepancy between these two high-dimensional laws can be controlled through fluctuations of a single scalar correction field. This field separates the correction for dependence across slices from the mismatch in the annealed weights. We reveal slices one at a time after adding Gaussian noise. This separates how each observation changes our prediction of the correction field from the uncertainty remaining after all observations. Suppose the slice corrections have Lipschitz constants $O(\sigma(\tau))$, with uniform bounds on dependence and conditional concentration. Matching the added Gaussian noise to the EDM level $\sigma(\tau)$ then yields $\mathrm{KL}(\Pi_\tau \| P^\star) = O(N\sigma(\tau)^2)$ and $\|\Pi_\tau - P^\star\|_{\mathrm{TV}} = O(\sqrt{N}\,\sigma(\tau))$. For a fixed bridge problem, this noise-alignment condition requires the zero-noise surrogate to coincide with the Schrödinger bridge, and the bounds quantify the resulting noise-level convergence.

1 Introduction

A frequent strategy in modern generative modeling is to reuse a diffusion prior trained on lower dimensional data (e.g., 2D) to guide the generation or reconstruction of a 3D object. We can represent the underlying object ignoring its dimensionality for the moment, by some latent variable θ. One motivating example is brain imaging: 2D diffusion priors are used as priors for 3D MRI or CT reconstruction through slice renderings. This avoids the cost of training a full 3D diffusion model [18, 45, 28, 61, 72]. The same underlying issue appears more broadly if a high-dimensional object is constrained only via many lower-dimensional observations or projections or renderings.

Using the lower-dimensional information to model the full object requires a way to combine the information across its renderings. One way to do so is to start with a separate prior on the high-dimensional object. This prior describes which full objects are plausible before one makes use of the lower-dimensional information. We draw samples from this prior and render a slice at a fixed position which gives a distribution over slices. This is the *prior-induced marginal*. For each sample, we then compare the plausibility of *each* rendered slice under the 2D diffusion model and the prior-induced

[1]University of Wisconsin–Madison. [2]Peking University.
*Equal contribution. Correspondence to: Shen Xu <xs-psych@stu.pku.edu.cn>, Hui Chen <hchen795@wisc.edu>.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

marginal. The resultant score measures how strongly the 2D model favors the slice beyond what the object prior already predicts. These rendering-based scores are combined to reweight, or tilt, the object prior. The strength of the tilt is annealed using a noise schedule such as EDM [40]. We refer to the resulting annealed law as the "surrogate". Related approaches [61, 72, 17] have shown that empirically this general approach works well. But a basic question is:

> What distribution is this surrogate approximating and does the approximation error scale with the noise level in a controlled way?

Before we can check the surrogate's approximation error, we first need a principled reference law over the high-dimensional object. The 2D diffusion prior provides *target distributions* for the individual rendered slices. These prescribed slice marginals tell us what each slice should look like. Unfortunately, they do not determine how the slices "fit together" inside a single latent object. Indeed, many high-dimensional laws can have the same collection of rendered marginals. But the prior on the high dimensional objects can help with the ambiguity: it describes which high-dimensional couplings are plausible *before* the slice constraints are imposed. We therefore seek a law that *matches all prescribed slice marginals* while *minimizing KL divergence from the object prior*. This is the **multi-marginal Schrödinger bridge** associated with the prior and the slice marginals [14, 7, 15].

Why surrogate can be different. Both constructions use the same prior and target slice marginals. The rendered slices are all functions of the same latent object. So, reweighting the law to match one marginal can certainly change the marginal laws of the other slices. The bridge chooses its slice scores jointly to enforce all marginals. The surrogate, on the other hand, uses the prescribed scores with annealed weights, *without* jointly enforcing all the marginal constraints. In our analysis, we will separate the correction for interactions between slice constraints from the mismatch in weighting, then bound their combined contribution to the discrepancy.

1.1 The target and the computable surrogate

Let Θ denote the latent object space and let p_0 be a prior on Θ. For slice positions $s_1, \dots, s_N$, we write $S_i : \Theta \to \mathcal{X}$ for the ith measurable renderer, where $\mathcal{X}$ is the space of rendered observations. In a brain-imaging example, $\theta \in \Theta$ can parameterize a 3D volume. The "rendering" is simple: $S_i(\theta)$ will extract a 2D slice at position or slice s_i. As we mentioned above, drawing some object/sample from p_0 and applying S_i gives a distribution over slices at this position. We denote this *prior-induced marginal* by $\mu_i^0 = (S_i)_{\#} p_0$. The target marginal at the same slice position is denoted by q_{s_i}. It describes the desired distribution of rendered slices, given by the 2D diffusion prior. We assume q_{s_i} and μ_i^0 are mutually absolutely continuous. Now, fix a finite, strictly positive Borel density ratio $r_i = dq_{s_i}/d\mu_i^0$ and set $\ell_i = \log r_i$. We use $\ell_i(S_i(\theta))$ as the *slice-level log score*. It is high when the rendered slice is *more plausible* under the target marginal (than under the prior-induced marginal).

We combine the slice log scores using weights $w_i > 0$ with $\sum_i w_i = 1$. The surrogate reweights the prior by these scores, with strength controlled by the EDM-inspired factor $\beta(\tau) = (\sigma(\tau)^2 + \sigma_{\mathrm{data}}^2)^{-1}$:

$$\frac{d\Pi_\tau}{dp_0}(\theta) \propto \exp\left(\beta(\tau) \sum_{i=1}^{N} w_i \, \ell_i(S_i(\theta))\right). \tag{1}$$

For a fixed τ, this defines a law over latent objects. The exponential factor will favor objects whose rendered slices score highly (compared to the prior-induced marginals). Here $\sigma(\tau) > 0$ is the noise level and $\sigma_{\mathrm{data}} > 0$ is a fixed data scale.

The bridge instead enforces every prescribed slice marginal:

$$P^\star \in \arg\min_P \mathrm{KL}(P \| p_0) \quad \text{subject to} \quad (S_i)_{\#} P = q_{s_i}, \qquad i = 1, \dots, N. \tag{2}$$

The constraints essentially fix the slice marginals. Among laws which satisfy them, the KL objective picks the one closest to the object prior. This is the static Schrödinger bridge problem [10, 34, 65, 15]. We assume the surrogate normalizer is finite and the bridge admits a minimizer with finite KL divergence. The precise assumptions are given later in §2. Both laws use the same prior, renderers, and target marginals. The bridge jointly chooses how each slice contributes to reweighting the prior (so all marginal constraints hold). But the surrogate uses the prescribed slice log-ratios with annealed weights.

Running example. In brain imaging, θ can be the parameters of an implicit neural representation (INR) of a 3D volume [74, 31, 55, 76]. As mentioned above, the renderer $S_i(\theta)$ extracts a slice at position s_i. The target marginal q_{s_i} is the desired distribution of slices at that position, from the 2D diffusion prior. We can handle other measurable renderers but we will avoid digression. In tomography, for example, S_i can be a projection operator such as a discrete Radon transform [51].

Next, we state the reduction that makes the bridge–surrogate gap analyzable.

1.2 Analysis overview and main results

The first step is to understand the exact bridge–surrogate gap. Under the assumptions in §2, we fix finite functions $\varphi_i^\star$, called *dual potentials*, representing the bridge as $dP^\star/dp_0 \propto \exp\big(\sum_{i=1}^N \varphi_i^\star(S_i)\big)$ [23, 7, 53]. For this choice of potentials, the difference between the surrogate and bridge terms for slice i is $g_{i,\tau} = \beta(\tau)w_i\ell_i - \varphi_i^\star$. Summing these differences gives the correction field $h_\tau(\theta) = \sum_{i=1}^N g_{i,\tau}(S_i(\theta))$. Comparing the two densities then gives the tilt identity

$$\frac{dP^\star}{d\Pi_\tau} \propto e^{-h_\tau}. \tag{3}$$

In §2, we use the slice constraints to split h_τ into two pieces: **(i)** a coupling correction for enforcing all marginals jointly and **(ii)** some mismatch from the annealed coefficients $\beta(\tau)w_i$.

From identity to a bound. The identity (3) reduces the problem to fluctuations of h_τ under Π_τ. An additive constant in h_τ is absorbed by normalization. The field h_τ depends on *all* rendered slices, which can be dependent because they come from the same latent object θ. We reveal noisy slices one at a time, adding Gaussian noise at an auxiliary scale $\xi > 0$. These successive observations define the *Gaussian-smoothed filtration*. We assume *Wasserstein screening*, which bounds how changing one noisy observation affects the conditional laws of the slices. The *screening profile* is the sequence of these bounds, indexed by separation between slice indices. We also assume conditional moment-generating-function (MGF) bounds to control uncertainty at each step and after all noisy slices have been revealed. Building on smoothing and concentration tools [32, 24, 49, 75], these assumptions and Lipschitz regularity of the slice corrections yield a sub-Gaussian bound for the centered correction field under Π_τ, with variance proxy

$$v_\tau = r_\tau^2 + \sum_{i=1}^N s_{i,\tau}^2. \tag{4}$$

The term r_τ^2 controls uncertainty in h_τ remaining after all noisy slices have been observed. The terms $s_{i,\tau}^2$ control fluctuations in predicting h_τ as each noisy slice is revealed. In §3, we will make both these terms explicit.

Informal Theorem 1 (Reduction and concentration). Under the assumptions in §2, the bridge satisfies $dP^\star/d\Pi_\tau \propto e^{-h_\tau}$. If, in addition, the hypotheses of Theorem 3.1 hold for a smoothing scale $\xi > 0$, then for all $\lambda \in \mathbb{R}$,

$$\log \mathbb{E}_{\Pi_\tau} \exp(\lambda(h_\tau - \mathbb{E}_{\Pi_\tau} h_\tau)) \le \frac{\lambda^2 v_\tau}{2}, \tag{5}$$

with $v_\tau = r_\tau^2 + \sum_i s_{i,\tau}^2$. Together with the tilt identity, this gives $\mathrm{KL}(\Pi_\tau \| P^\star) \le v_\tau/2$ and $\|\Pi_\tau - P^\star\|_{\mathrm{TV}} \le \frac{1}{2}\sqrt{v_\tau}$ (see Theorem 4.1).

Informal Theorem 2 (Noise-level rate). Set the smoothing scale equal to the EDM noise level, $\xi = \sigma(\tau)$. Suppose the hypotheses of Informal Theorem 1 hold throughout a range of noise levels approaching zero. Assume the screening profile is summable with a common bound, and the latent and posterior concentration controls remain uniformly bounded over this range. Let $L_{i,\tau}$ bound the Lipschitz constants of the chosen extensions of $g_{i,\tau}$ to the ambient Euclidean space. The additional *noise-alignment* condition requires these corrections to become less sensitive as the noise decreases. If $\max_{1\le i\le N} L_{i,\tau} \le c_L\sigma(\tau)$ for a constant c_L independent of τ, then as $\sigma(\tau) \downarrow 0$,

$$\mathrm{KL}(\Pi_\tau \| P^\star) = O\big(N\sigma(\tau)^2\big), \qquad \|\Pi_\tau - P^\star\|_{\mathrm{TV}} = O\Big(\sqrt{N}\,\sigma(\tau)\Big). \tag{6}$$

For the fixed bridge problem considered here, this alignment condition forces the limiting slice corrections to be constant, so the zero-noise surrogate coincides with $P^\star$. Our constants depend only on the screening, concentration, and alignment bounds. The dependence on N is uniform whenever these bounds are independent of N. The precise conditions and constants are given in §4.

Our **contributions** are as follows. **(a) Bridge–surrogate identity.** We derive an exponential-tilt representation that reduces the distributional comparison to the scalar correction field h_τ. **(b) Smoothed concentration decomposition.** We introduce a Gaussian-smoothed filtration and obtain the proxy $v_\tau = r_\tau^2 + \sum_i s_{i,\tau}^2$, separating posterior uncertainty from fluctuations as noisy slices are revealed. **(c) Noise-level convergence.** Under uniform screening and concentration controls, if we match the smoothing scale to the EDM noise level and assume $L_{i,\tau} = O(\sigma(\tau))$, the surrogate satisfies $\mathrm{KL}(\Pi_\tau \| P^\star) = O(N\sigma(\tau)^2)$. We will develop this analysis in order: §2 establishes the tilt identity, §3 proves the concentration bound and §4 derives the KL and noise-level rates.

2 The bridge is a tilt of the surrogate

We specified two laws on the same latent space Θ: the bridge $P^\star$ enforces the slice marginals exactly, while the surrogate Π_τ tilts the prior p_0 by aggregated slice scores. We already know how Π_τ reweights p_0. To compare it with $P^\star$, let us work out the bridge's reweighting. Under the assumptions below [23, 66, 7], we express the bridge as an exponential tilt of the surrogate. Their comparison then reduces to controlling fluctuations of a scalar correction field.

Setup. We use the notation from §1. Four conditions are used here: **(i)** the bridge problem is feasible and admits a minimizer $P^\star$ with finite KL divergence, **(ii)** the slice density ratios $r_i = dq_{s_i}/d\mu_i^0$ are represented by finite, strictly positive Borel functions, with $\ell_i = \log r_i$, **(iii)** the surrogate normalizer is finite, so that Π_τ is a probability law, and **(iv)** the bridge admits the additive *dual-potential representation* [46], with finite measurable potentials $\varphi_i^\star$ and $Z_\star \in (0, \infty)$:

$$\frac{dP^\star}{dp_0}(\theta) = \frac{1}{Z_\star} \exp\left(\sum_{i=1}^N \varphi_i^\star(S_i(\theta)) \right). \tag{7}$$

This exponential form is standard for KL minimization under marginal constraints [22, 21, 46, 59]. The appendix gives the conditions on measurability, conditional laws, and representation used below.

The dual representation [46, 59, 54, 47] gives the bridge a sum of slice terms in the exponent, as in the surrogate. The surrogate density in (1) uses the observable slice log-ratios ℓ_i, while the bridge density in (7) involves the unknown coupled potentials $\varphi_i^\star$. The marginal constraints will help us link up these terms.

What one slice tells us. If we look at the bridge through one renderer S_i, we can use the constraint $(S_i)_\# P^\star = q_{s_i}$. For a μ_i^0-a.e. rendered slice x, consider the objects satisfying $S_i(\theta) = x$. This set is the *fiber* over x. On this set, the ith potential is fixed at $\varphi_i^\star(x)$, while the other slice potentials can vary. We average their combined exponential weight under the conditional prior to obtain the factor $\mathcal{C}_i(x)$ below.

Proposition 2.1 (Slice marginal factorization)**.** *Fix $i \in \{1, \dots, N\}$, and let $p_{0,i}(d\theta \mid x)$ be a regular conditional law of $\theta \sim p_0$ given $S_i(\theta) = x$. Define*

$$\mathcal{C}_i(x) = \int_\Theta \exp\left(\sum_{j \neq i} \varphi_j^\star(S_j(\theta)) \right) p_{0,i}(d\theta \mid x). \tag{8}$$

Then, for each i,

$$q_{s_i}(dx) = \frac{1}{Z_\star} e^{\varphi_i^\star(x)} \mathcal{C}_i(x)\, \mu_i^0(dx). \tag{9}$$

Therefore, with $f_i = \log \mathcal{C}_i$ on the full-μ_i^0-measure set where $\mathcal{C}_i \in (0, \infty)$,

$$\varphi_i^\star(x) = \ell_i(x) - f_i(x) + \log Z_\star \qquad \mu_i^0\text{-a.e.} \tag{10}$$

So, up to the constant $\log Z_\star$, the bridge potential is $\ell_i - f_i$. The term f_i is how the other slice potentials impact this expression.

Proof sketch. The key step is to compute the marginal of the bridge through a single renderer S_i. We test $(S_i)_{\#}P^\star$ against a bounded function and substitute the dual-potential density (7). Conditioning p_0 on $S_i(\theta) = x$ fixes the ith exponential factor at $e^{\varphi_i^\star(x)}$. It comes outside the conditional integral, while the other factors remain inside and form $\mathcal{C}_i(x)$. This gives (9). Comparing the slice density with $r_i = dq_{s_i}/d\mu_i^0$ and taking logarithms gives (10). See Appendix C. □

Proposition 2.1 collects the contributions from all other slice potentials into the single factor $\mathcal{C}_i(x) = e^{f_i(x)}$. We can now compare the slice terms in the bridge and the surrogate directly.

Notice that $\ell_i = \log r_i$ in (10) is the same slice log-ratio that the surrogate uses, with coefficient $\beta(\tau)w_i$. What the bridge also needs is the correction f_i. To compute it, we must integrate over latent objects that share the same rendered slice. This is why it cannot generally be obtained from a slice-wise model alone. We can now see the gap in (10): the surrogate uses $\beta(\tau)w_i\ell_i$, while the bridge uses $\ell_i - f_i$ up to normalization.

Putting the slice corrections together. For fixed τ, let us compare these terms by subtracting $\ell_i - f_i$ from $\beta(\tau)w_i\ell_i$. This gives the slice-wise correction

$$g_{i,\tau}(x) = f_i(x) - \big(1 - \beta(\tau)w_i\big)\ell_i(x). \tag{11}$$

We then add these corrections over all slices to obtain the correction field

$$h_\tau(\theta) = \sum_{i=1}^N g_{i,\tau}(S_i(\theta)) = \sum_{i=1}^N f_i(S_i(\theta)) - \sum_{i=1}^N \big(1 - \beta(\tau)w_i\big)\ell_i(S_i(\theta)). \tag{12}$$

The first expression is emphasizing that h_τ is a sum of slice-level corrections. The second expression separates the two sources of discrepancy: the hidden coupling correction f_i and the mismatch between the bridge coefficient 1 and the annealed surrogate coefficient $\beta(\tau)w_i$.

This differs from the definition in §1 only by an additive constant. For the same fixed dual potentials, these corrections agree μ_i^0-a.e. with those in §1 plus $\log Z_\star$. Thus the fields differ by $N \log Z_\star$ Π_τ-a.s. These constants cancel in normalized tilts and centered fields. The Lipschitz constants are also unchanged if the functions are chosen by these same constant shifts. Now that we have h_τ in hand, we can write the normalized density ratio.

Theorem 2.2 (Exponential-tilt identity)**.** *Under the standing conditions above,*

$$\frac{dP^\star}{d\Pi_\tau}(\theta) = \frac{e^{-h_\tau(\theta)}}{\mathbb{E}_{\Pi_\tau}[e^{-h_\tau}]} \qquad \Pi_\tau\text{-a.s.} \tag{13}$$

Proof sketch. Proposition 2.1 now lets us compare the two laws directly. Substituting $\varphi_i^\star(S_i) = \ell_i(S_i) - f_i(S_i) + \log Z_\star$ into the ratio $(dP^\star/dp_0)/(d\Pi_\tau/dp_0)$ gives an exponent equal to $-h_\tau$ plus a constant. The constant disappears when we normalize under Π_τ, which gives (13). See Appendix C. □

What the identity tells us. We have used two ingredients. The dual representation gives the bridge a sum of slice terms in the exponent, as in the surrogate. The marginal constraints then express those terms through ℓ_i and f_i. Theorem 2.2 puts them together in the single correction field h_τ.

The term $\sum_i f_i(S_i(\theta))$ is the coupling correction. It comes from the fiber factors $\mathcal{C}_i = e^{f_i}$ and accounts for enforcing all slice marginals simultaneously. The surrogate combines the slice log-ratios without explicitly including these fiber corrections. The second term, $-\sum_i(1 - \beta(\tau)w_i)\ell_i(S_i(\theta))$, is the tempering mismatch. Here we are comparing the coefficients of ℓ_i. The bridge uses coefficient 1 in (10), while the surrogate uses $\beta(\tau)w_i$. Since $\beta(\tau) = (\sigma(\tau)^2 + \sigma_{\mathrm{data}}^2)^{-1}$, the EDM noise schedule affects the correction field through this term. The rate theorem asks that the *combined* slice correction $g_{i,\tau}$ be noise-aligned, meaning that its Lipschitz constant is $O(\sigma(\tau))$. Here it is the combined correction that is relevant. The condition allows the two pieces to cancel. Together with the screening and concentration assumptions, this lets us control fluctuations of h_τ and obtain the noise-level rate in §4.

We now have an exact expression for the density ratio. To turn it into a KL bound, we need to control fluctuations of h_τ under Π_τ. The next section does this under its additional assumptions, using a Gaussian-smoothed filtration to prove a sub-Gaussian bound for the centered field. The resulting variance proxy will give the KL and noise-level rates in §4.

3 The correction field is sub-Gaussian

The tilt identity helped reduce the comparison between $P^\star$ and Π_τ to the single scalar correction field h_τ. The next question is how much this field can fluctuate when $\theta \sim \Pi_\tau$. Assuming that h_τ is sub-Gaussian under Π_τ would be mathematically sufficient to turn the tilt identity into a KL bound. But this alone would not tell us how tightly h_τ concentrates (or if the bound improves at all as the noise level decreases). We will instead obtain the bound from the slice corrections and their dependence.

The main issue we must handle here is dependence. The field h_τ is a function of the rendered slices $Y_i = S_i(\theta)$. But these slices are all functions of the same latent object. Revealing one slice will indeed change what we know about the others. So, our proof proceeds by revealing noisy slices one at a time and tracking how the conditional prediction of h_τ changes.

Revealing the slices. Let $\xi > 0$ be a smoothing scale. For now, ξ is free. In §4, we set $\xi = \sigma(\tau)$ to obtain the EDM noise-level rate. We will view the rendered slices as vectors, with $\mathcal{X} \subseteq \mathbb{R}^d$, and let $\Gamma_\xi = \mathcal{N}(0, \xi^2 I_d)$. If we draw $\theta \sim \Pi_\tau$ and N Gaussian noise vectors independently, it gives the law $\bar{Q} = \Pi_\tau \otimes \Gamma_\xi^{\otimes N}$ on the product space $\bar{\Omega} = \Theta \times (\mathbb{R}^d)^N$. Now, write

$$Y_i = S_i(\theta), \qquad \widetilde{Y}_i = Y_i + \zeta_i, \qquad \zeta_i \sim \Gamma_\xi, \tag{14}$$

with the ζ_i being independent. The information available after each noisy observation is described by the smoothed filtration

$$\widetilde{\mathcal{F}}_i = \sigma(\widetilde{Y}_1, \dots, \widetilde{Y}_i), \qquad i = 0, \dots, N. \tag{15}$$

Here $\widetilde{\mathcal{F}}_i$ contains what we know about θ after observing the first i noisy slices ($\widetilde{\mathcal{F}}_0$ is trivial). We add this Gaussian noise for the proof, so its scale ξ can be checked through the concentration bound. Below, expectations are under $\bar{Q}$ (unless a law is specified). For functions of θ alone, they agree with expectations under Π_τ.

We need three types of bounds: **(i) Lipschitz corrections (Assumption 5).** Each slice correction $g_{i,\tau}$ has an $L_{i,\tau}$-Lipschitz extension to $\mathbb{R}^d$. We fix these extensions throughout the argument. **(ii) Smoothed Wasserstein screening (Assumption 6).** The screening profile is a nonnegative sequence $\widetilde{\gamma}(r)$, $r = 0, 1, \dots$, with $\widetilde{\gamma}(0) = 1$. With the first $i-1$ noisy slices fixed, perturbing the ith noisy slice $\widetilde{Y}_i$ changes the conditional law of Y_j in Wasserstein distance W_1 by at most $\widetilde{\gamma}(|j-i|)$ times the perturbation size. **(iii) Specific conditional concentration (Assumptions 7–8).** We assume conditional moment-generating-function (MGF) bounds for the particular functions used to control the prediction of h_τ and the uncertainty left after all observations. The latent bounds have scales $\kappa_{i,\tau}$, and the posterior bound has constant C_{post}. The specific kernel and conditional MGF statements are given in Appendix D.

How one observation will affect the field. With the earlier noisy slices fixed, the conditional prediction of h_τ is Lipschitz in $\widetilde{Y}_i$ with bound

$$a_{i,\tau} = \sum_{j=1}^{N} L_{j,\tau} \widetilde{\gamma}(|j-i|). \tag{16}$$

Let us briefly interpret this quantity. If the ith noisy slice $\widetilde{Y}_i$ changes, then our prediction of the *entire* correction field can change through every slice j. The factor $L_{j,\tau}$ bounds the sensitivity of the correction at slice j. And $\widetilde{\gamma}(|j-i|)$ bounds the influence of the noisy observation on that slice's conditional law. Summing their product over j bounds the total influence on the prediction of h_τ. Notice that the sum includes the earlier slices also. Those observations were noisy, so a new observation can still change what we infer about an earlier clean slice.

There are two points for us to keep in mind. First, screening is imposed after Gaussian noise has been added [32, 52]. This lets us work with noisy conditional predictions while keeping track of the scale ξ. Second, our conditional MGF assumptions relate to the particular functions needed in the proof. A suitable conditional T_1 transportation inequality would give such bounds for every Lipschitz function [2, 48, 4].

What is remaining still? Let us follow our prediction of h_τ as slices are revealed. Each observation changes this prediction, and there may still be an error after all N noisy slices have been seen. We

can write the fluctuation as

$$h_\tau - \mathbb{E}h_\tau = R_\tau + \sum_{i=1}^{N} \widetilde{D}_i, \quad R_\tau = h_\tau - \mathbb{E}[h_\tau \mid \widetilde{\mathcal{F}}_N], \quad \widetilde{D}_i = \mathbb{E}[h_\tau \mid \widetilde{\mathcal{F}}_i] - \mathbb{E}[h_\tau \mid \widetilde{\mathcal{F}}_{i-1}]. \quad (17)$$

Here R_τ is the prediction error after all noisy slices are observed. Each $\widetilde{D}_i$ is the change in prediction when the ith noisy slice is revealed. These are the increments of the Doob martingale $\mathbb{E}[h_\tau \mid \widetilde{\mathcal{F}}_i]$. Adding the increments cancels the intermediate predictions, and R_τ completes the decomposition.

Let us first bound the error left after the last observation. The posterior concentration assumption gives the residual proxy

$$r_\tau^2 = C_{\mathrm{post}} \xi^2 \sum_{i=1}^{N} L_{i,\tau}^2. \quad (18)$$

The factor ξ^2 in the middle is the squared smoothing scale. To see the other factor, write the correction in slice coordinates as $G_\tau(y_1, \ldots, y_N) = \sum_i g_{i,\tau}(y_i)$. Under the Euclidean product metric, its squared Lipschitz constant is at most $\sum_i L_{i,\tau}^2$.

Now consider one change in the prediction. With the first $i-1$ noisy slices fixed, screening makes the prediction after observing $\widetilde{Y}_i$ an $a_{i,\tau}$-Lipschitz function of that observation. The observation $\widetilde{Y}_i = Y_i + \zeta_i$ has two sources of randomness. The latent conditional MGF bound contributes the scale $\kappa_{i,\tau}^2$, and the Gaussian bound contributes ξ^2. Thus the ith increment has proxy

$$s_{i,\tau}^2 = a_{i,\tau}^2(\kappa_{i,\tau}^2 + \xi^2). \quad (19)$$

We can now combine the conditional bounds for the residual and the prediction changes. The total variance proxy is

$$v_\tau = r_\tau^2 + \sum_{i=1}^{N} s_{i,\tau}^2. \quad (20)$$

The next theorem states the resultant concentration bound.

Theorem 3.1 (Correction-field concentration)**.** *Under the above assumptions, for every $\lambda \in \mathbb{R}$,*

$$\log \mathbb{E}_{\Pi_\tau} \exp(\lambda(h_\tau - \mathbb{E}_{\Pi_\tau} h_\tau)) \le \frac{\lambda^2 v_\tau}{2}. \quad (21)$$

Proof sketch. We use the decomposition (17) under $\bar{Q}$. Posterior concentration controls R_τ and gives (18). For each increment, screening gives the Lipschitz bound $a_{i,\tau}$. The Gaussian and latent conditional MGF bounds then give (19). To combine them, first we condition on $\widetilde{\mathcal{F}}_N$. The sum of the prediction changes is then fixed, and the posterior bound controls R_τ. We then iterate the conditional bounds for the changes along the filtration. Since h_τ depends only on θ, the resulting MGF bound is also the bound under Π_τ in (21). See Appendix D. □

Theorem 3.1 gives the concentration bound we need to turn the tilt identity into a KL estimate. The formula for v_τ tells us which quantities must remain controlled with the noise level changing. In the next section, we first work with (21) and turn it into a KL bound. We then set $\xi = \sigma(\tau)$ and apply noise alignment together with uniform screening and latent/posterior concentration bounds. The explicit ξ^2 factor in the residual becomes $\sigma(\tau)^2$ and these controls give the noise-level rate in §4.

4 The KL bound and the noise-level rate

We now have the two ingredients needed to compare the laws. First, in §2 we wrote the bridge as an exponential tilt of the surrogate, with correction field h_τ. And then, §3 showed that the centered field is sub-Gaussian under the surrogate law, with proxy v_τ. We can now put these together to obtain the KL and total-variation bounds mentioned in §1.

From tilt identity to KL. The tilt identity is normalized under Π_τ. This makes the KL divergence from the surrogate to the bridge simple to write out. Let

$$X_\tau = h_\tau - \mathbb{E}_{\Pi_\tau} h_\tau. \quad (22)$$

The mean contributes only a constant to the exponent. So, centering leaves the normalized tilt unchanged, and Theorem 2.2 gives

$$\frac{dP^\star}{d\Pi_\tau} = \frac{e^{-X_\tau}}{\mathbb{E}_{\Pi_\tau}[e^{-X_\tau}]}. \tag{23}$$

To compute $\mathrm{KL}(\Pi_\tau \| P^\star)$, we take the log of the reciprocal ratio and average under Π_τ. The term X_τ has mean zero, leaving

$$\mathrm{KL}(\Pi_\tau \| P^\star) = \log \mathbb{E}_{\Pi_\tau}\left[e^{-X_\tau}\right]. \tag{24}$$

The expectation is under the surrogate. This explains the direction of this KL bound. The total variation (TV) bound is symmetric and controls differences in the probabilities assigned to any measurable event.

Theorem 4.1 (Surrogate-to-bridge KL bound)**.** *Under the hypotheses of Theorems 3.1 and 2.2, using the same correction functions $g_{i,\tau}$,*

$$\mathrm{KL}(\Pi_\tau \| P^\star) \le \frac{v_\tau}{2}, \qquad \|\Pi_\tau - P^\star\|_{\mathrm{TV}} \le \frac{1}{2}\sqrt{v_\tau}. \tag{25}$$

Proof sketch. In (24), the exponent is $-X_\tau$, so we apply the concentration bound at $\lambda = -1$. Theorem 3.1 gives $\mathrm{KL}(\Pi_\tau \| P^\star) \le v_\tau/2$. Pinsker's inequality gives $\|\mu - \nu\|_{\mathrm{TV}} \le \sqrt{\mathrm{KL}(\mu\|\nu)/2}$ for any pair (μ, ν) [19]. Taking $\mu = \Pi_\tau$ and $\nu = P^\star$ gives the TV bound. □

The KL bound is now written with one quantity, v_τ. Let us bring back its definition from §3:

$$v_\tau = r_\tau^2 + \sum_{i=1}^N s_{i,\tau}^2, \qquad r_\tau^2 = C_{\mathrm{post}}\xi^2 \sum_{i=1}^N L_{i,\tau}^2, \qquad s_{i,\tau}^2 = a_{i,\tau}^2(\kappa_{i,\tau}^2 + \xi^2). \tag{26}$$

Matching the two noise scales. The surrogate depends on the EDM noise level $\sigma(\tau)$, while the proof introduced the auxiliary scale ξ. We now set $\xi = \sigma(\tau)$ and examine how the other factors in v_τ behave as the noise decreases.

Conditions for the noise-level rate. We need four conditions to hold uniformly as the noise decreases. First, the conditional latent concentration scales must remain bounded:

$$\max_{1\le i\le N} \kappa_{i,\tau} \le K_{\mathrm{data}}. \tag{27}$$

These are the scales used in the increment argument after conditioning on the earlier noisy slices. And second, the selected corrections $g_{i,\tau}$ must become less sensitive at the noise scale:

$$\max_{1\le i\le N} L_{i,\tau} \le c_L\sigma(\tau). \tag{28}$$

This is the noise-alignment condition. Recall that the selected correction agrees μ_i^0-a.e. with $\beta(\tau)w_i\ell_i - \varphi_i^\star$ up to an additive constant. The bound therefore controls how sensitive this difference is to a change in the slice, uniformly in i. This is an additional compatibility condition between the bridge and the surrogate. As $\sigma(\tau) \downarrow 0$, we have $\beta(\tau) \to \sigma_{\mathrm{data}}^{-2}$. For the fixed bridge problem considered here, the Lipschitz bound forces the limiting correction to be constant. In other words, $\sigma_{\mathrm{data}}^{-2}w_i\ell_i - \varphi_i^\star$ must be constant μ_i^0-a.e. for each i. Consequently, the zero-noise surrogate Π_0, defined by $\beta_0 = \sigma_{\mathrm{data}}^{-2}$, coincides with $P^\star$. This is a condition on the bridge problem itself (cannot be ensured by accurate marginal score estimates alone). The fixed-noise bound in Theorem 4.1 holds without this contraction. Third, the screening profile must be summable:

$$\sum_{r=1}^{\infty} \widetilde{\gamma}(r) \le \widetilde{\Gamma}_1 < \infty. \tag{29}$$

Finally, the posterior concentration constant must remain bounded, $C_{\mathrm{post}} \le C_{\mathrm{post}}^{\max}$.

Corollary 4.2 (Noise-level rate)**.** *Set $\xi = \sigma(\tau)$. Assume the hypotheses of Theorem 4.1 hold, together with* (27)–(29) *and $C_{\mathrm{post}} \le C_{\mathrm{post}}^{\max}$, with $K_{\mathrm{data}}, c_L, \widetilde{\Gamma}_1$, and $C_{\mathrm{post}}^{\max}$ independent of τ over the noise levels considered. Then*

$$\mathrm{KL}(\Pi_\tau \| P^\star) \le \frac{N\sigma(\tau)^2}{2}\left[c_L^2(1 + 2\widetilde{\Gamma}_1)^2\left(K_{\mathrm{data}}^2 + \sigma(\tau)^2\right) + C_{\mathrm{post}}^{\max}c_L^2\sigma(\tau)^2\right]. \tag{30}$$

So, as $\sigma(\tau) \downarrow 0$,

$$\mathrm{KL}(\Pi_\tau \| P^\star) = O\big(N\sigma(\tau)^2\big)\,, \qquad \|\Pi_\tau - P^\star\|_{\mathrm{TV}} = O\Big(\sqrt{N}\,\sigma(\tau)\Big)\,. \tag{31}$$

The dependence on N is uniform when the four bounding constants are also independent of N.

Proof sketch. Noise alignment bounds every $L_{j,\tau}$ by $c_L\sigma(\tau)$. In the remaining profile sum, the self term is $\widetilde{\gamma}(0) = 1$, and each positive index separation appears at most twice. Thus

$$a_{i,\tau} = \sum_{j=1}^{N} L_{j,\tau}\widetilde{\gamma}(|j-i|) \le c_L\sigma(\tau)(1 + 2\widetilde{\Gamma}_1). \tag{32}$$

For the increments, this gives $\sum_i s_{i,\tau}^2 \le N c_L^2\sigma(\tau)^2(1+2\widetilde{\Gamma}_1)^2(K_{\mathrm{data}}^2 + \sigma(\tau)^2)$. For the residual, we get $r_\tau^2 \le NC_{\mathrm{post}}^{\max}c_L^2\sigma(\tau)^4$. Substituting these estimates into $v_\tau = r_\tau^2 + \sum_i s_{i,\tau}^2$ and applying Theorem 4.1 gives (30). The total-variation rate follows from (25). □

Where is the rate coming from? The constants $\kappa_{i,\tau}$ and $L_{i,\tau}$ influence the bound in different ways. The $\kappa_{i,\tau}$ are conditional concentration scales used in the increment argument. Condition (27) keeps them bounded, but they need not vanish as the noise decreases. The $L_{i,\tau}$ bound the sensitivity of the corrections $g_{i,\tau}$. Noise alignment requires these sensitivities to shrink at least as fast as $\sigma(\tau)$.

Now let us look at the term $a_{i,\tau}^2\kappa_{i,\tau}^2$ in the increment proxy. Noise alignment/summable screening give $a_{i,\tau} = O(\sigma(\tau))$. With bounded $\kappa_{i,\tau}$, this gives an $O(\sigma(\tau)^2)$ bound per slice, and so $O(N\sigma(\tau)^2)$ after summing. The Gaussian part of the increment bound is $O(N\sigma(\tau)^4)$. The residual has the same $O(N\sigma(\tau)^4)$ upper bound (its explicit factor $\xi^2 = \sigma(\tau)^2$ multiplies $\sum_i L_{i,\tau}^2 = O(N\sigma(\tau)^2)$). So, the estimate combines an $O(N\sigma(\tau)^2)$ increment bound with an $O(N\sigma(\tau)^4)$ posterior-residual bound.

For a fixed problem the rate can be faster. The condition in (28) makes the limiting correction to be constant, so $g_{i,\tau}$ is a constant minus $\frac{\sigma(\tau)^2}{\sigma_{\mathrm{data}}^2(\sigma(\tau)^2+\sigma_{\mathrm{data}}^2)}w_i\ell_i$, and $L_{i,\tau} = O(\sigma(\tau)^2)$. If the concentration bounds hold at this scale, Theorem 4.1 gives $\mathrm{KL}(\Pi_\tau\|P^\star) = O(N\sigma(\tau)^4)$.

A Gaussian example with a nonzero gap. Let us see these conditions in a case where we can compute the gap. Write $\sigma = \sigma(\tau) > 0$ and index the example by σ. Let $\Theta = \mathbb{R}^4$, $S_i(\theta) = \theta_i$, $p_0 = \bigotimes_{i=1}^4 \mathcal{N}(c_i, 1)$, and $q_{s_i} = \mathcal{N}(-c_i, 1)$ for a fixed nonzero $c \in \mathbb{R}^4$. Take $w_i = 1/4$, $\sigma_{\mathrm{data}} = 1/2$, and $\xi = \sigma$. The independent structure makes the bridge explicit: $P^\star = \bigotimes_{i=1}^4 q_{s_i}$. Our choices give $\beta(\tau)w_i = 1/(1+4\sigma^2)$, which approaches the bridge coefficient 1.

The slice corrections have Lipschitz constants $L_{i,\sigma} = 8|c_i|\sigma^2/(1+4\sigma^2) \le 2|c_i|\sigma$, so we can take $c_L = 2\max_i|c_i|$. We also have $\widetilde{\Gamma}_1 = 0$ and $K_{\mathrm{data}} = C_{\mathrm{post}}^{\max} = 1$. Thus the hypotheses of Corollary 4.2 hold uniformly in σ, and the exact gap is

$$\mathrm{KL}(\Pi_\sigma\|P^\star) = \frac{32\sigma^4}{(1+4\sigma^2)^2}\sum_{i=1}^4 c_i^2 > 0 \qquad \text{for every } \sigma > 0.$$

As $\sigma \downarrow 0$, the exact gap is of order σ^4 as explained above. In this example, taking $\kappa_{i,\sigma} = 1$, the bound of Theorem 4.1 is $(1+2\sigma^2)$ times the exact gap, so it is tight as $\sigma \downarrow 0$. Also see Appendix E.1.

Practical implications. The four controls in Corollary 4.2 can guide the choice of a noise level for a fixed bridge problem satisfying these hypotheses. If the constants are certified, we can compare the bound $v_\tau/2$ with the desired KL tolerance. With estimates of these constants for a given model, we can use the bound to assess how closely the surrogate approximates the bridge.

While our focus is theoretical, we have also implemented a sampler motivated by the surrogate (Algorithm 1), which performs favorably relative to alternatives on 3D brain MRI (Appendix A).

5 Related work

EDM parameterization. Score-based and denoising diffusion models learn noise-conditioned scores that drive stochastic or deterministic samplers [37, 67]. We use EDM [40] and track the

effect of the noise schedule through our EDM-inspired tempering factor $\beta(\tau)$. Our results analyze the static annealed law induced by inserting lower-dimensional density-ratio information into a high-dimensional prior, separately from the sampling dynamics.

Lifting 2D diffusion priors to higher dimensions. Many proposals use 2D diffusion models as priors for 3D generation or reconstruction. DreamFusion introduced score distillation sampling for text-to-3D [61], and Score Jacobian Chaining backpropagates 2D scores through differentiable renderers [72]. In medical imaging, score-based priors are used for accelerated MRI [16], and 2D or slice-based priors guide 3D MRI and CT reconstruction [18, 28]. Our setup is similar to renderer-based lifting [61, 72]. We study the annealed density-ratio surrogate as a probability law and ask how close it is to the bridge that enforces all prescribed slice marginals.

Schrödinger bridges/Entropic projections. The bridge $P^\star$ is a projection of the prior onto slice marginal constraints. Under suitable regularity, the dual-potential representation comes from classical results on I-projections under linear constraints [22], with OT interpretations developed in the Schrödinger-problem literature [46] and multi-marginal extensions in [35]. We keep this representation as a standing assumption in §2. Other work uses Schrödinger bridges as sampling devices [7, 69]. Here, the bridge serves as the variational reference law (for comparing with annealed surrogate).

Convergence theory for diffusion samplers. A separate line of work analyzes convergence of reverse-time SDE samplers under assumptions on the score [13, 1, 6]. Work on diffusion-based posterior sampling for inverse problems [17, 18] includes particle-based methods with asymptotic guarantees for linear inverse problems [27]. These analyses concern sampler output laws relative to their prescribed targets. Our analysis compares the static annealed surrogate Π_τ with the corresponding multi-marginal bridge $P^\star$.

Concentration tools. The analysis builds on standard transport-cost inequalities and their connection to sub-Gaussian concentration, especially T_1-type inequalities [2, 48, 33]. Our concentration assumptions are conditional MGF bounds for the specific functionals used in the proof, and suitable conditional T_1 inequalities imply these bounds. The structural reduction is what lets us apply these tools here. We first reduce the bridge–surrogate comparison to the correction field h_τ. We then reveal Gaussian-perturbed slices one at a time and use screening to control their influence on this field. Together with the conditional MGF bounds, this gives the variance proxy $v_\tau = r_\tau^2 + \sum_i s_{i,\tau}^2$.

Implementation. We use an Instant-NGP-style implicit neural representation [50]. The analysis treats θ abstractly as the input to measurable renderers S_i.

6 Conclusions

Diffusion priors trained in one representation are often used in problems whose objects of interest live in a larger space. The annealed slice-tilted construction is a convenient and meaningful representative of ideas used in practice, but which law it actually targets, and how its approximation quality scales with the noise schedule, had not been made precise. We used the multi-marginal Schrödinger bridge defined by the prior and the prescribed slice marginals as a principled reference law. We showed how the gap between the two is governed by a single scalar correction field. Establishing sub-Gaussian control of this field gives the noise-level upper bound under noise-alignment and the stated uniform controls, with constants expressed in terms of model-specific structural quantities.

We make a brief comment on the linear scaling in N because it may seem that more slices should bring the surrogate closer to the bridge. But note that the bound is not measuring how much the slice information determines the 3D object. Instead, it is measuring the gap between two constructions that both use the same slice information. Each new slice adds its own coupling correction f_i and its own tempering mismatch term to h_τ, so the variance proxy v_τ accumulates across slices. When the bounding constants in Corollary 4.2 remain uniform over a family of bridge problems as N grows and $\sigma(\tau)$ decreases, the joint refinement $N\sigma(\tau)^2 \to 0$ brings the surrogate close to the bridge. This uniformity depends on the slice corrections and the dependence across slices in the chosen family.

Open questions. For a fixed problem, the order $\sigma(\tau)^4$ is attained by the Gaussian example. The sharp dependence on N remains open however. Extending our analysis from the static surrogate Π_τ to discretized reverse-time samplers, and incorporating ratio-estimation error in ℓ_i, would be steps toward guarantees accounting for sampling and estimation errors. We expect the same reduction strategy – entropy projection, exponential tilt, smoothed-filtration concentration – to apply beyond the slice setting, in posterior sampling for inverse problems and in guided generation.

Acknowledgments and Disclosure of Funding

The authors are grateful to Sihyeon Kim for discussions and input in the early stage of the project, and to Aaron Chen and Doug Dean III for discussions on extending this work.

Authors declare competing-interests/financial-interests disclosures as follows: Singh (Google DeepMind).

References

[1] Joe Benton, Valentin De Bortoli, Arnaud Doucet, and George Deligiannidis. Nearly d-linear convergence bounds for diffusion models via stochastic localization. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=r5njV3BsuD.

[2] S. G. Bobkov and F. Götze. Exponential Integrability and Transportation Cost Related to Logarithmic Sobolev Inequalities. *Journal of Functional Analysis*, 163(1):1–28, 1999. ISSN 0022-1236. doi: https://doi.org/10.1006/jfan.1998.3326. URL https://www.sciencedirect.com/science/article/pii/S0022123698933262.

[3] Vladimir I. Bogachev. *Conditional measures and conditional*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-34514-5. doi: 10.1007/978-3-540-34514-5_10. URL https://doi.org/10.1007/978-3-540-34514-5_10.

[4] François Bolley and Cédric Villani. Weighted csiszár-kullback-pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des sciences de Toulouse : Mathématiques*, 14(3):331–352, 2005. URL http://eudml.org/doc/73650.

[5] Christer Borell. The brunn-minkowski inequality in gauss space. *Inventiones mathematicae*, 30: 207–216, 1975. URL http://eudml.org/doc/142349.

[6] Valentin De Bortoli. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=MhK5aXo3gB.

[7] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=9BnCwiXB0ty.

[8] Jonathan M. Borwein and Adrian S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples*. CMS Books in Mathematics. Springer, New York, 2 edition, 2006. ISBN 978-0-387-31256-9. URL https://doi.org/10.1007/978-0-387-31256-9.

[9] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, February 2013. ISBN 978-0-19-953525-5. URL https://doi.org/10.1093/acprof:oso/9780199535255.001.0001.

[10] Rob Brekelmans and Kirill Neklyudov. On schrödinger bridge matching and expectation maximization. In *NeurIPS 2023 Workshop Optimal Transport and Machine Learning*, 2023. URL https://openreview.net/forum?id=Bd4DTPzOGO.

[11] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer, New York, 2011. ISBN 978-0-387-70913-0. URL https://doi.org/10.1007/978-0-387-70914-7.

[12] Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis, 2019. URL https://arxiv.org/abs/1904.00625.

[13] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=zyLVMgsZ0U_.

[14] Yu Chen, Wei Deng, Shikai Fang, Fengpei Li, Nicole Tianjiao Yang, Yikai Zhang, Kashif Rasul, Shandian Zhe, Anderson Schneider, and Yuriy Nevmyvaka. Provably convergent schrödinger bridge with applications to probabilistic time series imputation. *CoRR*, abs/2305.07247, 2023. URL https://doi.org/10.48550/arXiv.2305.07247.

[15] Rahul Choudhary and Hanbaek Lyu. Linear convergence of sinkhorn's algorithm for generalized static schrödinger bridge. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 10741–10767. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/choudhary25a.html.

[16] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated MRI. *Medical Image Analysis*, 80:102479, 2022. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2022.102479. URL https://www.sciencedirect.com/science/article/pii/S1361841522001268.

[17] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

[18] Hyungjin Chung, Dohoon Ryu, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Solving 3d inverse problems using pre-trained 2d diffusion models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, page 22542–22551. IEEE, 2023. doi: 10.1109/cvpr52729.2023.02159. URL http://dx.doi.org/10.1109/CVPR52729.2023.02159.

[19] Thomas M. Cover and Joy A. Thomas. *Inequalities in Information Theory*, chapter 17, pages 657–687. John Wiley Sons, Ltd, 2005. ISBN 9780471748823. doi: https://doi.org/10.1002/047174882X.ch17. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.ch17.

[20] Imre Csiszár and František Matúš. Information projections revisited. *IEEE Transactions on Information Theory*, 49(6):1474–1490, 2003. doi: 10.1109/TIT.2003.810633. URL https://doi.org/10.1109/TIT.2003.810633.

[21] Imre Csiszár and František Matúš. Information closures of exponential families. In *Proceedings of the 2004 IEEE International Symposium on Information Theory*, page 30. IEEE, 2004. URL https://doi.org/10.1109/ISIT.2004.1365065.

[22] I. Csiszár. i-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975. ISSN 00911798, 2168894X. URL http://www.jstor.org/stable/2959270.

[23] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf.

[24] Yunzi Ding and Jonathan Niles-Weed. Asymptotics of smoothed wasserstein distances in the small noise regime. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=z2cG3k8xa3C.

[25] Hacène Djellout, Arnaud Guillin, and Liming Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *The Annals of Probability*, 32 (3B):2702–2732, 2004. doi: 10.1214/009117904000000531. URL https://doi.org/10.1214/009117904000000531.

[26] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. i. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975. URL https://doi.org/10.1002/cpa.3160280102.

[27] Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=tplXNcHZs1.

[28] Chenhe Du, Qing Wu, Xuanyu Tian, Jingyi Yu, Hongjiang Wei, and Yuyao Zhang. Improving 2d diffusion models for 3d medical imaging with inter-slice consistent stochasticity. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=R5ETdN6ifA.

[29] R. M. Dudley. *Real Analysis and Probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2 edition, 2002. ISBN 978-0-521-00754-2. doi: 10.1017/CBO9780511755347. URL https://doi.org/10.1017/CBO9780511755347. Revised reprint of the 1989 original.

[30] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. URL https://doi.org/10.1198/jasa.2011.tm11181.

[31] Wei Fang, Yuxing Tang, Heng Guo, Mingze Yuan, Tony C.W. Mok, Ke Yan, Jiawen Yao, Xin Chen, Zaiyi Liu, Le Lu, Ling Zhang, and Minfeng Xu. Cycleinr: Cycle implicit neural representation for arbitrary-scale volumetric super-resolution of medical data. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11631–11641, 2024. URL https://ieeexplore.ieee.org/document/10656686.

[32] Ziv Goldfeld and Kristjan Greenewald. Gaussian-smoothed optimal transport: Metric structure and statistical efficiency. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 3327–3337. PMLR, 26–28 Aug 2020. URL https://proceedings.mlr.press/v108/goldfeld20a.html.

[33] Nathael Gozlan and Christian Léonard. Transport inequalities. a survey, 2010. URL https://arxiv.org/abs/1003.3852.

[34] Nikita Gushchin, Sergei Kholkin, Evgeny Burnaev, and Alexander Korotin. Light and optimal schrödinger bridge matching. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 17100–17122. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/gushchin24a.html.

[35] Isabel Haasler, Axel Ringh, Yongxin Chen, and Johan Karlsson. Multimarginal Optimal Transport with a Tree-Structured Cost and the Schrödinger Bridge Problem, 2021. URL https://doi.org/10.1137/20M1320195.

[36] Peter Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics. Academic Press, New York, 1980. ISBN 978-0-12-319350-6. URL https://shop.elsevier.com/books/martingale-limit-theory-and-its-application/hall/978-0-12-319350-6.

[37] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

[38] Olav Kallenberg. *Foundations of Modern Probability*, volume 99 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 3 edition, 2021. ISBN 978-3-030-61870-4. doi: 10.1007/978-3-030-61871-1. URL https://doi.org/10.1007/978-3-030-61871-1.

[39] L. V. Kantorovich and G. S. Rubinstein. On a space of completely additive functions. *Vestnik Leningradskogo Universiteta*, 13(7):52–59, 1958. In Russian.

[40] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=k7FuTOWMOc7.

[41] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *ICLR (Poster)*, 2015. URL http://dblp.uni-trier.de/db/conf/iclr/iclr2015.html#KingmaB14.

[42] M. Kirszbraun. Über die zusammenziehende und Lipschitzsche transformationen. *Fundamenta Mathematicae*, 22(1):77–108, 1934. URL https://www.semanticscholar.org/paper/Ãœber-die-zusammenziehende-und-Lipschitzsche-Kirszbraun/2e50a837da4883efa0cdfa7b261e2db5835e0d4d.

[43] Leonid (Aryeh) Kontorovich and Kavita Ramanan. Concentration inequalities for dependent random variables via the martingale method. *The Annals of Probability*, 36(6):2126 – 2158, 2008. doi: 10.1214/07-AOP384. URL https://doi.org/10.1214/07-AOP384.

[44] Michel Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001. ISBN 978-0-8218-3792-4. URL https://doi.org/10.1090/surv/089.

[45] Suhyeon Lee, Hyungjin Chung, Minyoung Park, Jonghyuk Park, Wi-Sun Ryu, and Jong Chul Ye. Improving 3d imaging with pre-trained perpendicular 2d diffusion models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10676–10686, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/papers/Lee_Improving_3D_Imaging_with_Pre-Trained_Perpendicular_2D_Diffusion_Models_ICCV_2023_paper.pdf.

[46] Christian Léonard. A survey of the schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014. ISSN 1078-0947. doi: 10.3934/dcds.2014.34.1533. URL https://www.aimsciences.org/article/id/d5bcf817-901d-4104-b7da-eade7847c53e.

[47] Simone Di Marino and Augusto Gerolin. An optimal transport approach for the schrödinger bridge problem and convergence of sinkhorn algorithm. *Journal of Scientific Computing*, 85(2):27, 2020. doi: 10.1007/s10915-020-01325-7. URL https://doi.org/10.1007/s10915-020-01325-7.

[48] K. Marton. Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration. *The Annals of Probability*, 24(2):857 – 866, 1996. doi: 10.1214/aop/1039639365. URL https://doi.org/10.1214/aop/1039639365.

[49] Gonzalo Mena and Jonathan Niles-Weed. *Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem*. Curran Associates Inc., Red Hook, NY, USA, 2019. URL https://dl.acm.org/doi/abs/10.5555/3454287.3454695.

[50] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4), July 2022. ISSN 0730-0301. doi: 10.1145/3528223.3530127. URL https://doi.org/10.1145/3528223.3530127.

[51] F. Natterer. *The Radon Transform and Related Transforms*, pages 9–53. Vieweg+Teubner Verlag, Wiesbaden, 1986. ISBN 978-3-663-01409-6. doi: 10.1007/978-3-663-01409-6_2. URL https://doi.org/10.1007/978-3-663-01409-6_2.

[52] Sloan Nietert, Ziv Goldfeld, and Kengo Kato. Smooth p-wasserstein distance: Structure, empirical approximation, and statistical applications. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8172–8183. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/nietert21a.html.

[53] Maxence Noble, Valentin De Bortoli, Arnaud Doucet, and Alain Durmus. Tree-based diffusion schrödinger bridge with applications to wasserstein barycenters. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=H2SuXHbFIn.

[54] Marcel Nutz and Johannes Wiesel. Entropic optimal transport: convergence of potentials. *Probability Theory and Related Fields*, 184(1):401–424, October 2022. ISSN 1432-2064. doi: 10.1007/s00440-021-01096-8. URL https://doi.org/10.1007/s00440-021-01096-8.

[55] Sourav Pal, Harshavardhan Adepu, Clinton Wang, Polina Golland, and Vikas Singh. Implicit representations via operator learning. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=2W3KUAaZgO.

[56] K. R. Parthasarathy. *Probability Measures on Metric Spaces*. Academic Press, 1967.

[57] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf.

[58] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner. Alzheimer's disease neuroimaging initiative (adni). *Neurology*, 74(3):201–209, 2010. doi: 10.1212/WNL.0b013e3181cb3e25. URL https://www.neurology.org/doi/abs/10.1212/WNL.0b013e3181cb3e25.

[59] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Found. Trends Mach. Learn.*, 11(5–6):355–607, February 2019. ISSN 1935-8237. doi: 10.1561/2200000073. URL https://doi.org/10.1561/2200000073.

[60] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, 1964. ISBN 978-0-8162-6804-7. Translated and annotated by Amiel Feinstein.

[61] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. DreamFusion: Text-to-3D using 2D diffusion. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=FjNys5c7VyY.

[62] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[63] Paul-Marie Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *The Annals of Probability*, 28(1):416 – 461, 2000. doi: 10.1214/aop/1019160125. URL https://doi.org/10.1214/aop/1019160125.

[64] Erwin Schrödinger. Über die umkehrung der naturgesetze. *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, pages 144–153, 1931.

[65] Yunyi Shen, Renato Berlinghieri, and Tamara Broderick. Multi-marginal schrödinger bridges with iterative reference refinement. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025. URL https://openreview.net/forum?id=VcwZ3gtYFY.

[66] Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=qy07OHsJT5.

[67] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

[68] V. N. Sudakov and B. S. Tsirelson. Extremal properties of half-spaces for spherically invariant measures. *Journal of Soviet Mathematics*, 9:9–18, 1978. URL https://doi.org/10.1007/BF01086099.

[69] Francisco Vargas, Pierre Thodoroff, Austen Lamacraft, and Neil Lawrence. Correction: Vargas et al. solving schrödinger bridges via maximum likelihood. entropy 2021, 23, 1134. *Entropy*, 25(2), 2023. ISSN 1099-4300. doi: 10.3390/e25020289. URL https://www.mdpi.com/1099-4300/25/2/289.

[70] Cédric Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003. URL https://bookstore.ams.org/gsm-58.

[71] Cédric Villani. The founding fathers of optimal transport. In *Optimal Transport: Old and New*, pages 29–37. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-540-71050-9. doi: 10.1007/978-3-540-71050-9_3. URL https://doi.org/10.1007/978-3-540-71050-9_3.

[72] Haochen Wang, Xiaodan Du, Jiahao Li, Raymond A. Yeh, and Greg Shakhnarovich. Score jacobian chaining: Lifting pretrained 2d diffusion models for 3d generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12619–12629, 2023. URL https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10203874.

[73] Max A. Woodbury. Inverting modified matrices. Technical Report Memorandum Report 42, Statistical Research Group, Princeton University, 1950.

[74] Junshen Xu, Daniel Moyer, Borjan Gagoski, Juan Eugenio Iglesias, P. Ellen Grant, Polina Golland, and Elfar Adalsteinsson. Nesvor: Implicit neural representation for slice-to-volume reconstruction in mri. *IEEE Transactions on Medical Imaging*, 42(6):1707–1719, 2023. doi: 10.1109/TMI.2023.3236216. URL https://doi.org/10.1109/TMI.2023.3236216.

[75] Rui Ray Zhang, Xingwu Liu, Yuyi Wang, and Liwei Wang. *McDiarmid-type inequalities for graph-dependent variables and stability bounds*. Curran Associates Inc., Red Hook, NY, USA, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/658bbbdef9415ba5e2ff857f1146ba6e-Abstract.html.

[76] Zhenxuan Zhang, Lipei Zhang, Yanqi Cheng, Zi Wang, Fanwen Wang, Haosen Zhang, Yue Yang, Yinzhe Wu, Jiahao Huang, Angelica I. Aviles-Rivero, Zhifan Gao, Guang Yang, and Peter J. Lally. From coarse to continuous: Progressive refinement implicit neural representation for motion-robust anisotropic mri reconstruction. *IEEE Transactions on Image Processing*, 35:3550–3565, 2026. doi: 10.1109/TIP.2026.3678052. URL https://doi.org/10.1109/TIP.2026.3678052.

A Case study: 3D brain MRI from 2D slice priors

We work through the brain-imaging example from §1. The aim is to see how the surrogate score $\nabla_\theta \log \Pi_\tau$ drives a usable sampler when the learned objects are 2D slice scores and the unknown is a 3D volume. §A.1 covers the sampler: latent parameterization, slice-ratio score approximation, annealed noise schedule, and mini-batched renderer gradients. §A.2 gives the ADNI setup, baselines, and evaluation protocol, and §A.3 reports the results.

A.1 Sampling algorithm

Algorithm 1 is the annealed slice-ratio Langevin sampler used in the case study. Differentiating the surrogate density gives the formal score:

$$\nabla_\theta \log \Pi_\tau(\theta) \;=\; \nabla_\theta \log p_0(\theta) \;+\; \beta(\tau) \sum_{i=1}^{N} w_i \left[DS_i(\theta)\right]^\top \nabla_x \ell_i\big(S_i(\theta)\big), \tag{33}$$

The expression has two parts: the prior score on the 3D representation and the sum of slice-level guidance terms transported back through the renderers. The implementation follows this decomposition, approximating the slice ratio guidance with Gaussian-smoothed scores evaluated at noisy

renderings. In each annealing stage we render a set of slices, add the same Gaussian noise level used by the 2D denoiser, estimate a target-minus-prior slice score, backpropagate those slice forces to the 3D parameters, and take a Langevin step. The details below are included because they are the places where the theory meets numerical choices:

(a) the prior score is evaluated in a low-dimensional latent coordinate system rather than in the raw implicit-network parameter space;

(b) the slice-ratio score is estimated by combining a frozen 2D EDM denoiser with a local Gaussian approximation to the prior-rendered slice law;

(c) the 2D denoiser is evaluated along the EDM noise schedule;

(d) the sum over slices is evaluated with an unbiased mini-batch estimator.

Algorithm 1 Annealed slice-ratio Langevin sampler (ASRL)

Require: Prior p_0 on Θ; renderers S_i; slice-ratio scores $\nabla_x \ell_i$; weights w_i; schedule $(\sigma_k, \beta_k)_{k=0}^{K-1}$; step sizes $(\eta_k)_{k=0}^{K-1}$.

1: Initialize $\theta_0 \sim p_0$ or from a prior-centered warm start.
2: **for** $k = 0, 1, \dots, K-1$ **do**
3: Render slices $x_i \leftarrow S_i(\theta_k)$ for $i = 1, \dots, N$.
4: Add analytical noise $y_i \leftarrow x_i + \sigma_k \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, I)$.
5: Evaluate the slice-ratio score $r_i \leftarrow \nabla_x \log q_i^{\sigma_k}(y_i) - \nabla_x \log \mu_i^{0,\sigma_k}(y_i)$.
6: Compute the approximate surrogate drift

$$G_k \leftarrow \nabla_\theta \log p_0(\theta_k) + \beta_k \sum_{i=1}^{N} w_i \big[DS_i(\theta_k)\big]^\top r_i.$$

7: Update $\theta_{k+1} \leftarrow \theta_k + \eta_k G_k + \sqrt{2\eta_k}\, \xi_k, \quad \xi_k \sim \mathcal{N}(0, I)$.
8: **end for**
9: **return** θ_K, or aggregate draws from independent chains.

Volume parameterization. The 3D object θ is represented by the parameters of an implicit neural volume [50]. The renderers S_i query this volume at coronal positions and return 2D slices. Sampling directly in the full parameter space is possible in principle, but it is a poor numerical choice: the prior score is not available in closed form, and the raw parameterization contains many directions that barely affect the rendered anatomy. We therefore run the chain in a lower-dimensional coordinate system tied to a prior-centered reconstruction. Let $\theta^\star$ be a MAP reconstruction under the volume prior, and let $U \in \mathbb{R}^{P \times r}$ be an orthogonal basis for the latent directions kept by the sampler. We write

$$\theta(z) = \theta^\star + Uz, \qquad z \sim \mathcal{N}(0, \lambda_z I_r). \tag{34}$$

The prior score in these coordinates is the linear term $-z/\lambda_z$, and the renderer Jacobian becomes $[DS_i(\theta(z))]\, U$. This change of variables is only a computational restriction to the latent subspace; the law on volumes is the pushforward of the z-space chain under $z \mapsto \theta(z)$. This parameterization specifies the space in which the sampler is implemented, while the theoretical bounds in Corollary 4.2 concern the annealed surrogate under its stated assumptions.

Slice-ratio score estimation. The sampler needs the score of a density ratio, not only the score of the learned slice distribution. We write

$$r_i(y, \sigma_k) = \nabla_x \log q_i^{\sigma_k}(y) - \nabla_x \log \mu_i^{0,\sigma_k}(y), \tag{35}$$

and approximate the two terms separately. The first term is the score of the noise-conditioned target slice law. It is supplied by the frozen 2D EDM denoiser $D_i(y, \sigma)$ [40] through the Tweedie identity [30],

$$\nabla_x \log q_i^{\sigma}(y) \;\approx\; \frac{D_i(y, \sigma) - y}{\sigma^2}. \tag{36}$$

The second term is the score of the prior-rendered slice law after the same Gaussian smoothing. This term prevents the sampler from treating high prior probability slices as evidence in favor of the target;

Table 1: Hyperparameters used for Algorithm 1 in the ADNI case study.

Hyperparameter	Value
Latent dimension r	512
Prior variance λ_z	1.0
Langevin steps per noise level	1,000
Noise schedule $\sigma_{\max}$	80
Noise schedule $\sigma_{\min}$	0.002
Schedule exponent ρ	7
Slice mini-batch $\lvert\mathcal{I}_k\rvert$	16
Slice weights w_i	uniform $(1/N)$

it is the subtraction that makes the drift a slice-ratio drift rather than a plain slice-prior drift. We approximate it locally by a Gaussian centered at the rendered MAP reconstruction:

$$\nabla_x \log \mu_i^{0,\sigma}(y) \;\approx\; -C_{i,\sigma}^{-1}\big(y - S_i(\theta^\star)\big). \tag{37}$$

Here $C_{i,\sigma}$ combines the latent prior geometry $U\lambda_z U^\top$, propagated through the renderer Jacobian at $\theta^\star$, with isotropic Gaussian noise at scale σ. We apply the Woodbury identity [73] to avoid inverting a full slice covariance: the cost is reduced from $O(d^3)$ to $O(r^2 d)$, where d is the number of pixels in the rendered slice and r is the latent dimension. This local Gaussian approximation incorporates the prior and renderer geometry while avoiding a second learned score model for prior-rendered slices.

Analytical noise and the smoothed filtration. The noise $\sigma_k \varepsilon_i$ on line 4 of Algorithm 1 sets the noise level at which the 2D EDM denoiser is evaluated. It is distinct from the Gaussian variables ζ_i of Section 3, which define the noisy slices $\widetilde{Y}_i = S_i(\theta) + \zeta_i$ only as a device for the concentration proof. Corollary 4.2 sets the analytical smoothing scale ξ equal to the EDM noise level $\sigma(\tau)$, and the implementation uses the same schedule for the denoiser. This shared schedule does not by itself give a guarantee for the finite-step sampler. This alignment is useful in practice because high-noise stages move the volume using broad anatomical information, while low-noise stages sharpen the slice details without changing the structure of the drift.

Mini-batched slice gradients. A full step would backpropagate through every rendered slice. Adjacent positions in a dense coronal stack carry highly correlated guidance, so a small mini-batch captures essentially the same drift signal. At step k, we sample a mini-batch $\mathcal{I}_k \subset \{1, \dots, N\}$ uniformly without replacement and replace the slice sum on line 6 by the unbiased estimator

$$\widehat{G}_k \;=\; \nabla \log p_0(\theta_k) + \frac{N}{|\mathcal{I}_k|}\,\beta_k \sum_{i \in \mathcal{I}_k} w_i \left[DS_i(\theta_k)\right]^\top r_i. \tag{38}$$

The Jacobian-vector product $[DS_i(\theta_k)]^\top r_i$ is computed by automatic differentiation through the implicit representation. The resulting Monte Carlo noise is mean-zero conditional on θ_k. In finite-step Langevin dynamics it still affects mixing and discretization error, so we keep the mini-batch size fixed across methods and report it in Table 1. In the experiments, a batch of 16 slices gave a good balance between stable anatomy-level motion and affordable backpropagation.

Hyperparameters. Table 1 gives the values used in the case study. The schedule constants $\sigma_{\max}$, $\sigma_{\min}$, and ρ follow the EDM defaults of Karras et al. [40]; the latent dimension r, prior variance λ_z, and mini-batch size $|\mathcal{I}_k|$ were chosen by standard practice for the Instant-NGP backbone [50] and were not tuned against the evaluation metrics.

A.2 Case study setup

Data. We use the T1-weighted structural MRI scans from the ADNI1 phase of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`https://adni.loni.usc.edu`) [58], downloaded on April 1, 2025. Led by Principal Investigator Michael W. Weiner, MD, ADNI began in 2003 as a public-private partnership aimed at tracking the progression of mild cognitive impairment and early Alzheimer's disease using imaging, biological markers, and clinical assessments. Each volume is registered to Talairach/MNI305 space and resampled to a $256 \times 256 \times 256$ isotropic grid. The 2D EDM denoiser is trained on coronal slices from this corpus with continuous slice-position

conditioning. After training, the denoiser is frozen, and all methods in the comparison use the same slice model.

Baselines. We compare complete sampling pipelines built around the same trained 2D EDM slice prior. Our method additionally uses a 3D implicit representation and a prior-centered latent subspace, while TPDM uses the perpendicular-plane denoiser required by its method.

(a) Independent samples each coronal slice independently from the 2D prior and stacks the results.

(b) ISCS [28] samples from the 2D prior with slerp-interpolated noise correlation across slice positions before stacking the slices into a volume.

(c) TPDM [45] combines two pre-trained 2D diffusion models on perpendicular planes. Our implementation uses the slice denoiser above together with a second denoiser trained on the perpendicular plane.

Evaluation. We evaluate each method along three axes.

- **(i) Image fidelity.** FID and KID computed on MedicalNet ResNet-10 features [12] over $1{,}000$ generated volumes per method. MedicalNet provides a 3D-aware feature extractor whose statistics are more domain-relevant than ImageNet features.
- **(ii) 3D spatial coherence.** Adjacent-slice structural similarity ($\mathrm{SSIM}_{\parallel}$), perceptual distance ($\mathrm{LPIPS}_{\parallel}$), and mutual information ($\mathrm{MI}_{\parallel}$) between consecutive coronal slices, together with the Total Variation (TV) anisotropy ratio $\rho_{\mathrm{TV}} = \mathrm{TV}_{\mathrm{cor}}/\mathrm{TV}_{\mathrm{ax}}$. Real volumetric data has $\rho_{\mathrm{TV}} \approx 1$; values $\ll 1$ are the spectral signature of slice-stacking with weakly coupled noise. Adjacent-slice statistics are computed within coronal slices 32–220 to avoid boundary padding; SSIM is sampled over 30 positions per volume, LPIPS over 20, and MI over 10.
- **(iii) Qualitative cross-check.** Side-by-side displays of representative axial, sagittal, and coronal slices from one volume per method, used as a visual check that the numerical comparisons in (i) and (ii) reflect structural differences that are also visible to a human reader.

Compute. All experiments were run on a single NVIDIA A100 (40 GB), implemented in PyTorch [57]. The 2D EDM denoiser was trained with Adam [41] and frozen for all sampling experiments; the same GPU was used to run Algorithm 1 and the baselines.

A.3 Results

Image fidelity and adjacent-slice coherence. Table 2 gives the numerical results. Our method achieves the lowest FID (1.31 vs. 2.12 for TPDM, the next best) and the lowest KID (0.0098). On the adjacent-slice coherence side, $\mathrm{SSIM}_{\parallel}$ (0.944) and $\mathrm{LPIPS}_{\parallel}$ (0.021) match the real-data reference ($0.946/0.022$) within measurement variability, while every baseline departs from the reference on at least one of these two metrics. TPDM matches the real $\mathrm{LPIPS}_{\parallel}$ (0.025) but its $\mathrm{SSIM}_{\parallel}$ collapses to 0.791, a profile in which adjacent slices look locally plausible yet share little structural information. Independent shows the most extreme failure mode: its ρ_{TV} drops to 0.37, with axial total variation roughly three times the lateral, which is exactly the spectral signature predicted for unrelated slices stacked into a volume. For mutual information $\mathrm{MI}_{\parallel}$, our method gives 0.883, compared with 0.760 for real data, 0.570 for TPDM, 0.666 for ISCS, and 0.349 for Independent. The higher value indicates stronger adjacent-slice dependence than in the real-data reference. We interpret this measure alongside $\mathrm{SSIM}_{\parallel}$, $\mathrm{LPIPS}_{\parallel}$, and the cross-sections in Figure 1.

Qualitative cross-check. Figure 1 shows representative coronal, sagittal, and axial slices from one generated volume per method. The four samplers all produce coronal cross-sections that read as plausible brain MRI. Differences between them are most readable in the sagittal and axial directions, where slice-based synthesis tends to leave its spatial signature. The Independent baseline exhibits the strongest banding across slices in these views, consistent with the elevated $\mathrm{TV}_{\mathrm{ax}}$ and reduced $\mathrm{MI}_{\parallel}$ reported in Table 2. ISCS and TPDM produce visibly smoother sagittal and axial cross-sections, with residual inter-slice discontinuities most apparent in ISCS. The volumes drawn by Algorithm 1 are visually consistent with coherent 3D anatomy in all three planes.

Table 2: Image fidelity and 3D spatial coherence on the ADNI case study (mean $\pm$ std over 1,000 generated volumes per method). FID and KID use MedicalNet ResNet-10 features. $\mathrm{SSIM}_{\parallel}$, $\mathrm{LPIPS}_{\parallel}$, and $\mathrm{MI}_{\parallel}$ are structural similarity, perceptual distance, and mutual information between consecutive coronal slices. $\rho_{\mathrm{TV}} = \mathrm{TV}_{\mathrm{cor}}/\mathrm{TV}_{\mathrm{ax}}$ is the ratio of coronal to axial total variation, with target value 1 (isotropic); values $\ll 1$ indicate slice-stacking artifacts. Best generated result in **bold**. $\mathrm{MI}_{\parallel}$ is interpreted relative to the real-data reference.

Method	FID↓	KID↓	$\mathrm{SSIM}_{\parallel}\uparrow$	$\mathrm{LPIPS}_{\parallel}\downarrow$	$\mathrm{MI}_{\parallel}$	$\rho_{\mathrm{TV}} \to 1$
Real (ref.)	—	—	0.946 ± 0.021	0.022 ± 0.004	0.760 ± 0.074	1.016 ± 0.028
Independent	3.04	0.030	0.800 ± 0.007	0.123 ± 0.007	0.349 ± 0.010	0.369 ± 0.005
TPDM	2.12	0.016	0.791 ± 0.002	0.025 ± 0.000	0.570 ± 0.002	1.029 ± 0.001
ISCS	2.74	0.020	0.929 ± 0.010	0.054 ± 0.010	0.666 ± 0.131	0.809 ± 0.087
Ours	**1.31**	**0.010**	$\mathbf{0.944 \pm 0.007}$	$\mathbf{0.021 \pm 0.001}$	0.883 ± 0.079	0.875 ± 0.012

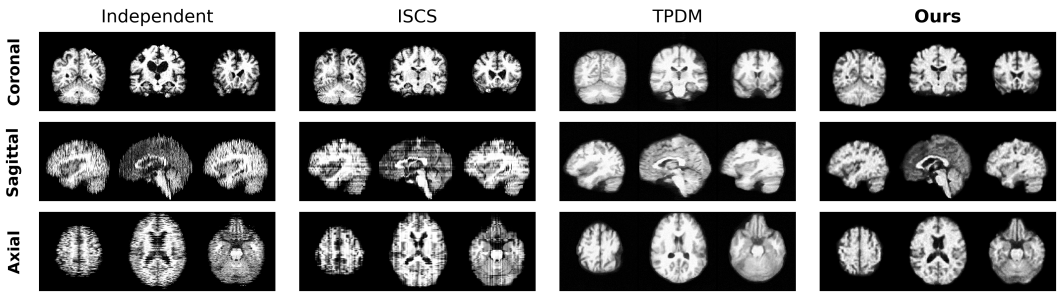

Figure 1: Representative coronal, sagittal, and axial slices from one generated volume per method (Independent, ISCS, TPDM, ours). Three slice positions per plane are shown for each method. Sagittal and axial cross-sections are smoothest for our sampler and TPDM, ISCS shows residual inter-slice discontinuities, and Independent shows banding across slices, the expected artifact when uncorrelated slice samples are stacked into a volume.

B Technical setup and measure-theoretic tools

The main text suppresses several measure-theoretic details that would interrupt the argument. We collect them here before proving the three claims used in the paper: the fiber factorization of the bridge potentials, the exponential-tilt identity, and the smoothed-filtration concentration bound. The common setting is standard Borel, so regular conditional laws and disintegrations are available; see, for example, [38, 3, 56, 29].

Measurable structure. Throughout the appendix, Θ and $\mathcal{X} \subseteq \mathbb{R}^d$ are standard Borel spaces, meaning measurable spaces equivalent to Borel subsets of complete separable metric spaces, and $S_i : \Theta \to \mathcal{X}$ are Borel measurable renderers. The prior is a probability law p_0 on Θ, and $\mu_i^0 = (S_i)_{\#}p_0$, where $T_{\#}\mu$ denotes the pushforward of a measure μ by a measurable map T. We write $\mathcal{B}(E)$ for the Borel σ-field of a space E. Statements that hold "a.s." hold almost surely with respect to the probability law currently being used; statements that hold "a.e." hold almost everywhere with respect to the measure named in the same sentence or formula. A fiber over x means the set of parameters $\{\theta : S_i(\theta) = x\}$. All Wasserstein distances on $\mathcal{X}$ use the Euclidean metric inherited from $\mathbb{R}^d$. Whenever a function on $\mathcal{X}$ is said to be Lipschitz on $\mathbb{R}^d$, we mean that the chosen Borel representative has been extended to a Borel Lipschitz function on the ambient Euclidean space. Here a representative means a concrete Borel version of an equivalence class of functions that agree except on a null set.

Definition B.1 (Probability kernels)**.** *Let E and F be standard Borel spaces. A probability kernel from E to F is a map $K : E \times \mathcal{B}(F) \to [0,1]$ such that $A \mapsto K(e,A)$ is a probability measure for each $e \in E$, and $e \mapsto K(e,A)$ is Borel measurable for each Borel set $A \subseteq F$. If $(\Omega, \mathcal{G})$ is a measurable space, a map $K : \Omega \times E \times \mathcal{B}(F) \to [0,1]$ is a kernel in (ω, e) when the same two properties hold with respect to $\mathcal{G} \otimes \mathcal{B}(E)$.*

We first state the measurability fact used whenever a conditional kernel is integrated against a slice-level test function.

Lemma B.2 (Kernel integrals are measurable)**.** *Let K be a probability kernel from $(E, \mathcal{E})$ to a standard Borel space F. If $\psi : F \to [0, \infty]$ is Borel, then*

$$e \longmapsto \int_F \psi(z)\, K(e, dz) \tag{39}$$

is $\mathcal{E}$-measurable. If ψ is signed and $\int_F |\psi(z)|K(e, dz) < \infty$ for every e, then $e \mapsto \int_F \psi(z)K(e, dz)$ is measurable as well.

Proof. For nonnegative simple functions this is the definition of a kernel. A monotone class argument gives the nonnegative Borel case. The signed case follows by writing $\psi = \psi^+ - \psi^-$ and applying the first part to the positive and negative parts. □

The next identity is used several times with test functions that depend on the noisy slices already included in the conditioning. It is the standard regular-conditional-law identity, with the dependence on ω made explicit.

Lemma B.3 (Random test functions and regular conditional laws)**.** *Let $(\Omega, \mathcal{F}, Q)$ be a probability space, let $\mathcal{G} \subseteq \mathcal{F}$, and let Z be a random element in a standard Borel space E. Let $K(\omega, dz)$ be a regular conditional law of Z given $\mathcal{G}$. If $F : \Omega \times E \to [0, \infty]$ is $\mathcal{G} \otimes \mathcal{B}(E)$-measurable, then*

$$\mathbb{E}_Q[F(\cdot, Z) \mid \mathcal{G}](\omega) = \int_E F(\omega, z)K(\omega, dz) \qquad Q\text{-a.s.} \tag{40}$$

where both sides may be $+\infty$. If $F(\cdot, Z) \in L^1(Q)$, the same identity holds for signed F with finite representatives.

Proof. For indicators $F(\omega, z) = \mathbf{1}_A(\omega)\mathbf{1}_B(z)$, where $\mathbf{1}_A$ denotes the indicator of a set A, with $A \in \mathcal{G}$ and $B \in \mathcal{B}(E)$, the identity is exactly the defining property of the regular conditional law. A monotone-class argument gives the nonnegative case. The signed integrable case follows by applying the nonnegative case to positive and negative parts. □

Screening kernels are naturally specified through bounded test functions. The correction terms are Lipschitz, hence usually unbounded. The following extension is the bridge between those two statements.

Lemma B.4 (Extension from bounded tests to integrable tests)**.** *Let $\mathcal{G} \subseteq \mathcal{F}$, let Y be a random element in a standard Borel space E, and let Z be a random element in a standard Borel space F. Suppose $K(\omega, y, dz)$ is a kernel such that, for every bounded measurable $\psi : F \to \mathbb{R}$,*

$$\mathbb{E}[\psi(Z) \mid \sigma(\mathcal{G}, Y)](\omega) = \int_F \psi(z)K(\omega, Y(\omega), dz) \qquad Q\text{-a.s.} \tag{41}$$

Let $\psi : F \to \mathbb{R}$ be Borel with $\psi(Z) \in L^1(Q)$ and suppose the kernel integral $\int_F |\psi(z)|K(\omega, Y(\omega), dz)$ is finite Q-a.s. Then (41) *also holds for this ψ.*

Proof. Apply (41) to the bounded truncations $\psi_n = (-n) \vee (\psi \wedge n)$. Since $\psi_n(Z) \to \psi(Z)$ in $L^1(Q)$ and conditional expectation is an L^1 contraction, the left-hand side converges in L^1 to $\mathbb{E}[\psi(Z) \mid \sigma(\mathcal{G}, Y)]$. On the right-hand side, the finite absolute kernel-integral assumption and dominated convergence give convergence to $\int \psi(z)K(\omega, Y(\omega), dz)$ along the chosen kernel version. Uniqueness of limits in probability identifies these two limits. □

We will use conditional expectations as L^1 limits of bounded approximations. The contraction property below is the only analytic input needed for that step.

Lemma B.5 (L^1 continuity of conditional expectation)**.** *Let $(\Omega, \mathcal{F}, Q)$ be a probability space and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-field. If $X_n, X \in L^1(Q)$ and $X_n \to X$ in $L^1(Q)$, then*

$$\mathbb{E}_Q[X_n \mid \mathcal{G}] \to \mathbb{E}_Q[X \mid \mathcal{G}] \qquad \text{in } L^1(Q). \tag{42}$$

Proof. Conditional expectation is an L^1 contraction:

$$\|\mathbb{E}_Q[X_n \mid \mathcal{G}] - \mathbb{E}_Q[X \mid \mathcal{G}]\|_1 \le \|X_n - X\|_1. \tag{43}$$

□

The bridge factorization conditions on a single rendered slice. The disintegration below supplies the conditional prior on the fiber $\{\theta : S_i(\theta) = x\}$.

Lemma B.6 (Fiber-supported disintegration)**.** *Fix i. There exists a regular conditional law $p_{0,i}(d\theta \mid x)$ of $\theta \sim p_0$ given $S_i(\theta) = x$ such that, for μ_i^0-almost every x,*

$$p_{0,i}\big(\{\theta : S_i(\theta) = x\} \mid x\big) = 1. \tag{44}$$

Consequently, for every nonnegative Borel function $H : \Theta \to [0, \infty]$ and every bounded nonnegative Borel function $\phi : \mathcal{X} \to [0, \infty)$,

$$\int_\Theta \phi(S_i(\theta)) H(\theta)\, p_0(d\theta) = \int_{\mathcal{X}} \phi(x) \left\{ \int_\Theta H(\theta)\, p_{0,i}(d\theta \mid x) \right\} \mu_i^0(dx). \tag{45}$$

The same identity holds for integrable signed H and bounded signed Borel ϕ, by decomposing into positive and negative parts.

Proof. Regular conditional laws exist because all spaces are standard Borel. The fiber-supported version follows from the fact that $S_i(\theta) = x$ almost surely under a conditional law given $S_i = x$. The integral identity is the usual disintegration formula; on the fiber the factor $\phi(S_i(\theta))$ is $\phi(x)$. □

The small algebraic step that makes the fiber factorization work is that any function of $S_i(\theta)$ is constant on the fiber $S_i(\theta) = x$.

Lemma B.7 (Fiber constant factors)**.** *With $p_{0,i}(d\theta \mid x)$ as in Lemma B.6, if $\psi : \mathcal{X} \to \mathbb{R}$ is Borel and $H : \Theta \to [0, \infty]$ is Borel, then for μ_i^0-almost every x,*

$$\int_\Theta \psi(S_i(\theta)) H(\theta)\, p_{0,i}(d\theta \mid x) = \psi(x) \int_\Theta H(\theta)\, p_{0,i}(d\theta \mid x), \tag{46}$$

whenever both sides are well-defined.

Proof. On a full-μ_i^0 set, $p_{0,i}(\cdot \mid x)$ is supported on $\{\theta : S_i(\theta) = x\}$. On this set, $\psi(S_i(\theta)) = \psi(x)$, and the claim follows by integration. □

Screening will give Wasserstein control of conditional laws. The next bound turns that Wasserstein control into a Lipschitz bound for conditional predictions; it is the Kantorovich–Rubinstein duality estimate [39, 70, 71].

Lemma B.8 (Kantorovich–Rubinstein bound)**.** *Let ν and ν' be probability laws on $\mathbb{R}^d$ with finite first moments. If $F : \mathbb{R}^d \to \mathbb{R}$ is L-Lipschitz, then*

$$\left| \int F\, d\nu - \int F\, d\nu' \right| \le L W_1(\nu, \nu'). \tag{47}$$

Proof. This is the dual representation of W_1, applied to F/L if $L > 0$. The case $L = 0$ is immediate. □

Finally, the added Gaussian coordinate contributes the explicit ξ^2 term in the variance proxy. We use the usual Gaussian concentration inequality for Lipschitz functions, conditionally on the past [44, 9, 5, 68].

Lemma B.9 (Gaussian concentration for a random Lipschitz function)**.** *Let $Z \sim \mathcal{N}(0, \xi^2 I_d)$ be independent of a σ-field $\mathcal{G}$. Let $F : \Omega \times \mathbb{R}^d \to \mathbb{R}$ be $\mathcal{G} \otimes \mathcal{B}(\mathbb{R}^d)$-measurable. Suppose $F(\omega, \cdot)$ is a-Lipschitz for every ω and $F(\cdot, Z) \in L^1$. Define*

$$\bar{F}(\omega) = \int_{\mathbb{R}^d} F(\omega, z)\, \mathcal{N}(0, \xi^2 I_d)(dz). \tag{48}$$

Then $\bar{F}$ is $\mathcal{G}$-measurable, $\bar{F} = \mathbb{E}[F(\cdot, Z) \mid \mathcal{G}]$ a.s., and for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\{\lambda(F(\cdot, Z) - \bar{F})\} \mid \mathcal{G}\right] \le \exp\left(\frac{\lambda^2 a^2 \xi^2}{2}\right) \qquad \text{a.s.} \tag{49}$$

Proof. The conditional-expectation identity follows from Lemma B.3 with the constant Gaussian kernel. Conditionally on $\mathcal{G}$, the function $z \mapsto F(\omega, z)$ is deterministic and a-Lipschitz. The standard Gaussian isoperimetric concentration bound gives (49) for each such ω. □

Remark B.10 (Finite density ratios and representatives)**.** *The main paper writes $\ell_i = \log r_i$ and later imposes Lipschitz regularity on $g_{i,\tau} = f_i - (1 - \beta(\tau)w_i)\ell_i$. These are finite-valued statements. Thus the appendix assumes that the selected density-ratio representatives $r_i = dq_{s_i}/d\mu_i^0$ are finite and strictly positive on the relevant slice support. If one assumes only $q_{s_i} \ll \mu_i^0$, then r_i may vanish on a set of positive μ_i^0-mass and ℓ_i is extended-valued; the proofs below do not cover that extended-valued case. Here $\nu \ll \mu$ means that ν is absolutely continuous with respect to μ.*

C Proofs for the bridge–surrogate identity

The structural step in Section 2 is to rewrite the bridge potentials in terms of observable slice density ratios and one hidden fiber correction. Once that rewrite is available, the tilt identity is just the ratio of the bridge density and the surrogate density.

C.1 Structural assumptions

Assumption 1 (Bridge existence)**.** *The feasible set*

$$\mathcal{C} = \{P \in \mathcal{P}(\Theta) : (S_i)_{\#}P = q_{s_i},\ i = 1, \dots, N\} \tag{50}$$

is nonempty, and the constrained problem

$$\inf_{P \in \mathcal{C}} \mathrm{KL}(P \| p_0) \tag{51}$$

admits a minimizer $P^\star$ with $\mathrm{KL}(P^\star \| p_0) < \infty$.

Assumption 2 (Finite positive grid density ratios)**.** *For each i, $q_{s_i} \sim \mu_i^0$ on the relevant slice support, where $\sim$ denotes mutual absolute continuity. We choose a finite positive Borel representative*

$$r_i = \frac{dq_{s_i}}{d\mu_i^0} : \mathcal{X} \to (0, \infty) \tag{52}$$

and set $\ell_i = \log r_i$.

Assumption 3 (Surrogate normalizability)**.** *For every τ under consideration,*

$$Z_{\Pi,\tau} = \int_\Theta \exp\left(\beta(\tau) \sum_{i=1}^N w_i \ell_i(S_i(\theta))\right) p_0(d\theta) \tag{53}$$

is finite and strictly positive. Hence

$$\frac{d\Pi_\tau}{dp_0}(\theta) = \frac{1}{Z_{\Pi,\tau}} \exp\left(\beta(\tau) \sum_{i=1}^N w_i \ell_i(S_i(\theta))\right) \tag{54}$$

defines a probability law on Θ.

The bridge is an entropy projection under linear marginal constraints. In the regular regime where the dual problem is attained, the optimizer has the additive exponential form below [22, 20, 8, 64, 46, 59, 54, 47]. We use this representation as the regularity condition for the bridge analysis.

Assumption 4 (Additive dual-potential representation)**.** *The bridge admits measurable finite-valued potentials $\varphi_i^\star : \mathcal{X} \to \mathbb{R}$ and a constant $Z_\star \in (0, \infty)$ such that*

$$\frac{dP^\star}{dp_0}(\theta) = \frac{1}{Z_\star} \exp\left(\sum_{i=1}^N \varphi_i^\star(S_i(\theta))\right) \qquad p_0\text{-a.s.} \tag{55}$$

Remark C.1 (Dual attainment)**.** *Assumption 1 explicitly posits a finite-KL minimizer. Finite real-valued dual potentials are the additional regularity used here. The slice factorization below follows from* (55) *and the marginal constraints.*

Remark C.2 (Density-ratio representatives)**.** *The surrogate uses $\ell_i = \log r_i$, so the proof works with finite positive representatives. Under the finite-potential representation above, the bridge density is positive p_0-a.s.; taking its marginal through S_i gives $q_{s_i} = (S_i)_{\#}P^\star \sim \mu_i^0$. This is the measure-theoretic version of the positive density-ratio convention used in the main text.*

Sufficient regimes for the dual representation. Write $R_N = (S_1, \dots, S_N)$, $\mu^0_{1:N} = (R_N)_{\#} p_0$, and $m_N = \bigotimes_{i=1}^N q_{s_i}$. A sufficient regime is that $\mathcal{X}$ is Polish and, for some $\varepsilon > 0$ and bounded Borel $c : \mathcal{X}^N \to \mathbb{R}$,

$$\mu^0_{1:N} = Z_0^{-1} e^{-c/\varepsilon} m_N, \qquad Z_0 = \int_{\mathcal{X}^N} e^{-c/\varepsilon}\, dm_N.$$

For couplings π with marginals q_{s_i}, the entropic objective $\int c\, d\pi + \varepsilon \mathrm{KL}(\pi \| m_N)$ equals $\varepsilon \mathrm{KL}(\pi \| \mu^0_{1:N}) - \varepsilon \log Z_0$. Thus [47, Theorem 4.5(ii)], with its marginals $\rho_i = q_{s_i}$, gives an optimal joint law $\pi^\star$ and potentials $u_i^\star \in L^\infty(q_{s_i})$ such that

$$\pi^\star(dx) = \exp\left(\frac{\sum_i u_i^\star(x_i) - c(x)}{\varepsilon} \right) m_N(dx),$$

$$\frac{d\pi^\star}{d\mu^0_{1:N}}(x) = Z_0 \exp\left(\sum_i \frac{u_i^\star(x_i)}{\varepsilon} \right).$$

Choose bounded Borel representatives of the $u_i^\star$. This preserves the density identity since $\mu^0_{1:N} \sim m_N$. Retaining the conditional prior along R_N gives the latent optimizer

$$\frac{dP^\star}{dp_0} = \left(\frac{d\pi^\star}{d\mu^0_{1:N}} \right) \circ R_N.$$

Indeed, this lift has joint law $\pi^\star$ and the same finite relative entropy, while every feasible P satisfies $\mathrm{KL}(P \| p_0) \geq \mathrm{KL}((R_N)_{\#} P \| \mu^0_{1:N})$ by the entropy chain rule. Consequently, (55) holds with $\varphi_i^\star = u_i^\star / \varepsilon$ and $Z_\star = Z_0^{-1}$.

For $N = 2$, [46, Theorem 2.4] gives another sufficient criterion in the common-reference-marginal regime $\mu_1^0 = \mu_2^0 = m$. Identify its R_{01} with $\mu^0_{1:2}$, its (μ_0, μ_1) with (q_{s_1}, q_{s_2}). Under all of its hypotheses (i)–(vii), its equation (23) gives $\pi^\star(dx_1 dx_2) = \alpha_1(x_1)\alpha_2(x_2)\mu^0_{1:2}(dx_1 dx_2)$ with finite positive measurable factors. Choosing Borel versions, the same lift gives $\varphi_i^\star = \log \alpha_i$ and $Z_\star = 1$. Outside these sufficient regimes, we retain Assumption 4 explicitly.

C.2 Slice marginal factorization

Fix i. The idea is to inspect the bridge density while holding the rendered slice $S_i(\theta) = x$ fixed. The remaining bridge potentials are then averaged over the prior fiber above x. Let $p_{0,i}(d\theta \mid x)$ be a fiber-supported disintegration of p_0 along S_i, as in Lemma B.6. Define

$$\mathcal{C}_i(x) = \int_\Theta \exp\left(\sum_{j \neq i} \varphi_j^\star(S_j(\theta)) \right) p_{0,i}(d\theta \mid x). \tag{56}$$

The integral is a nonnegative Borel function of x by Lemma B.2. It is finite and positive μ_i^0-a.e. because the marginal density obtained below equals the finite positive density ratio r_i.

Proposition C.3 (Slice marginal factorization)**.** *Under Assumptions 1, 2, and 4, for each i,*

$$q_{s_i}(dx) = \frac{1}{Z_\star} e^{\varphi_i^\star(x)} \mathcal{C}_i(x)\, \mu_i^0(dx). \tag{57}$$

Consequently, with $f_i = \log \mathcal{C}_i$ on the full-μ_i^0 set where $\mathcal{C}_i \in (0, \infty)$,

$$\varphi_i^\star(x) = \ell_i(x) - f_i(x) + \log Z_\star \qquad \mu_i^0\text{-a.e.} \tag{58}$$

Proof. Let $\phi : \mathcal{X} \to [0, \infty)$ be bounded and Borel. Using the density (55),

$$\int_{\mathcal{X}} \phi(x) (S_i)_{\#} P^\star(dx) = \frac{1}{Z_\star} \int_\Theta \phi(S_i(\theta)) \exp\left(\sum_{j=1}^N \varphi_j^\star(S_j(\theta)) \right) p_0(d\theta). \tag{59}$$

Disintegrate p_0 along S_i. On the fiber $S_i(\theta) = x$, the factor $\phi(S_i(\theta)) e^{\varphi_i^\star(S_i(\theta))}$ is equal to $\phi(x) e^{\varphi_i^\star(x)}$. Therefore

$$\int_{\mathcal{X}} \phi(x) (S_i)_{\#} P^\star(dx) = \int_{\mathcal{X}} \phi(x) \frac{1}{Z_\star} e^{\varphi_i^\star(x)} \mathcal{C}_i(x)\, \mu_i^0(dx). \tag{60}$$

Since equality of integrals against all bounded nonnegative Borel functions ϕ determines a probability measure, (57) follows. The bridge constraint gives $(S_i)_{\#}P^\star = q_{s_i}$, hence comparison with $r_i = dq_{s_i}/d\mu_i^0$ gives

$$r_i(x) = \frac{1}{Z_\star} e^{\varphi_i^\star(x)} \mathcal{C}_i(x) \qquad \mu_i^0\text{-a.e.} \tag{61}$$

Taking logarithms gives (58). □

Remark C.4 (Representatives)**.** *The function $f_i = \log \mathcal{C}_i$ is specified only μ_i^0-a.e. by the factorization. We choose a Borel finite-valued representative once and for all on the set where $\mathcal{C}_i \in (0, \infty)$, which has full μ_i^0-measure, and extend it measurably elsewhere. Because $(S_i)_{\#}p_0 = \mu_i^0$, modifying f_i on a μ_i^0-null set changes $f_i(S_i)$ only on a p_0-null set, and hence only on Π_τ- and $P^\star$-null sets.*

This choice becomes important in Appendix D. There we ask for Lipschitz representatives of the correction terms. The constants $L_{i,\tau}$ always refer to the representatives fixed here and in Assumption 5.

C.3 The exponential-tilt identity

We now insert the slice factorization into the density ratio $(dP^\star/dp_0)/(d\Pi_\tau/dp_0)$. This is the point where the bridge–surrogate comparison collapses to the scalar field h_τ.

For fixed τ, define

$$g_{i,\tau}(x) = f_i(x) - \big(1 - \beta(\tau) w_i\big)\ell_i(x), \tag{62}$$

and

$$h_\tau(\theta) = \sum_{i=1}^N g_{i,\tau}(S_i(\theta)) = \sum_{i=1}^N f_i(S_i(\theta)) - \sum_{i=1}^N \big(1 - \beta(\tau) w_i\big)\ell_i(S_i(\theta)). \tag{63}$$

Theorem C.5 (Exponential-tilt identity)**.** *Under Assumptions 1–4,*

$$\frac{dP^\star}{d\Pi_\tau}(\theta) = \frac{e^{-h_\tau(\theta)}}{\mathbb{E}_{\Pi_\tau}[e^{-h_\tau}]} \qquad \Pi_\tau\text{-a.s.} \tag{64}$$

Proof. The densities in (55) and (54) are finite and strictly positive p_0-a.s., so $P^\star \sim p_0 \sim \Pi_\tau$. The Radon–Nikodym chain rule therefore gives

$$\frac{dP^\star}{d\Pi_\tau}(\theta) = \frac{Z_{\Pi,\tau}}{Z_\star} \exp\left(\sum_{i=1}^N \varphi_i^\star(S_i(\theta)) - \beta(\tau) \sum_{i=1}^N w_i \ell_i(S_i(\theta)) \right) \qquad \Pi_\tau\text{-a.s.} \tag{65}$$

By Proposition C.3, after composition with S_i,

$$\varphi_i^\star(S_i) = \ell_i(S_i) - f_i(S_i) + \log Z_\star \qquad p_0\text{-a.s.} \tag{66}$$

Substituting this identity into the exponent gives

$$\sum_{i=1}^N \varphi_i^\star(S_i) - \beta(\tau) \sum_{i=1}^N w_i \ell_i(S_i) = -h_\tau + N \log Z_\star. \tag{67}$$

Thus $dP^\star/d\Pi_\tau = A_\tau e^{-h_\tau}$ for a finite positive constant A_τ. Since the Radon–Nikodym derivative integrates to one under Π_τ, $A_\tau^{-1} = \mathbb{E}_{\Pi_\tau}[e^{-h_\tau}]$, which proves (64). □

D Proofs for correction-field concentration

We now turn to the concentration bound for the correction field. The argument uses the smoothed filtration from the main paper. The central estimate is the increment bound: screening makes a conditional prediction Lipschitz, the Gaussian perturbation contributes a ξ^2 variance term, and a narrowed latent concentration assumption contributes the $\kappa_{i,\tau}^2$ term.

D.1 Smoothed product space and integrability

Fix τ and write $Q = \Pi_\tau$. Let $\xi > 0$ and $\Gamma_\xi = \mathcal{N}(0, \xi^2 I_d)$. Work on

$$\bar{\Omega} = \Theta \times (\mathbb{R}^d)^N, \qquad \bar{Q} = Q \otimes \Gamma_\xi^{\otimes N}. \tag{68}$$

Here $\otimes$ denotes product measure, so $\bar{Q}$ is obtained by drawing $\theta \sim Q$ and then adding independent Gaussian variables $\zeta_1, \dots, \zeta_N$ with law Γ_ξ. The added Gaussian coordinates are only a proof device. They play the same regularizing role as in smoothed Wasserstein distances and Gaussian-smoothed transport estimates [32, 52]. A point is denoted $(\theta, \zeta_1, \dots, \zeta_N)$. Define

$$Y_i = S_i(\theta), \qquad \widetilde{Y}_i = Y_i + \zeta_i, \qquad \widetilde{\mathcal{F}}_i = \sigma(\widetilde{Y}_1, \dots, \widetilde{Y}_i). \tag{69}$$

The variables ζ_i are independent of θ and of each other under $\bar{Q}$. Since h_τ depends only on θ, expectations of h_τ under Q and under $\bar{Q}$ are the same.

Assumption 5 (Lipschitz correction representatives)**.** *For the fixed τ, the representatives $g_{i,\tau}$ in* (62) *are chosen so that each has a Borel measurable $L_{i,\tau}$-Lipschitz extension to $\mathbb{R}^d$. We use the same symbol $g_{i,\tau}$ for the extension. Once chosen, these representatives are fixed throughout the concentration proof.*

Remark D.1 (Checking the Lipschitz condition)**.** *If the slice outputs lie in a subset $\mathcal{K} \subset \mathbb{R}^d$ and $g_{i,\tau}$ is Lipschitz on $\mathcal{K}$, a global extension may be obtained in Euclidean spaces by Kirszbraun's theorem [42, 11]. The extension fixes the representatives used in the Lipschitz and linear-growth estimates below. The correction field is evaluated on the clean slices Y_i. Gaussian averaging is applied to the conditional prediction Φ_i to form H_i.*

The concentration proof uses conditional expectations of Lipschitz functions of the slice variables. We make integrability explicit rather than hiding it in the notation. The Lipschitz assumption gives the linear-growth bound

$$|g_{i,\tau}(x)| \le |g_{i,\tau}(0)| + L_{i,\tau}\|x\|. \tag{70}$$

The integrability needed for the increment bound follows already from Assumptions 6 and 7 below. Indeed, $\widetilde{\mathcal{F}}_0$ is trivial, so the latter gives $\mathbb{E}_{\bar{Q}}\|Y_1\| < \infty$. Write $b_j = \int_{\mathcal{X}} \|z\| \widetilde{K}_{j|1}(0, dz) < \infty$, suppressing the trivial past argument. Applying (74) to bounded truncations of $\|z\|$, then using the Wasserstein bound (75) and monotone convergence, gives

$$\mathbb{E}_{\bar{Q}}\|Y_j\| \le b_j + \widetilde{\gamma}(|j-1|)\mathbb{E}_{\bar{Q}}\|\widetilde{Y}_1\| < \infty.$$

Together with (70), this implies $g_{j,\tau}(Y_j) \in L^1(\bar{Q})$ for every j, and hence $h_\tau = \sum_j g_{j,\tau}(Y_j) \in L^1(\bar{Q})$.

Define the Doob martingale and its increments by

$$\widetilde{M}_i = \mathbb{E}_{\bar{Q}}[h_\tau \mid \widetilde{\mathcal{F}}_i], \qquad \widetilde{D}_i = \widetilde{M}_i - \widetilde{M}_{i-1}, \qquad R_\tau = h_\tau - \mathbb{E}_{\bar{Q}}[h_\tau \mid \widetilde{\mathcal{F}}_N]. \tag{71}$$

This is the standard martingale-difference decomposition behind bounded difference and martingale concentration arguments [36, 9, 75]. Then

$$h_\tau - \mathbb{E}_Q[h_\tau] = R_\tau + \sum_{i=1}^{N} \widetilde{D}_i. \tag{72}$$

Indeed, $h_\tau - \mathbb{E}[h_\tau \mid \widetilde{\mathcal{F}}_N]$ plus the martingale telescoping sum from $\widetilde{M}_0$ to $\widetilde{M}_N$ equals $h_\tau - \mathbb{E}[h_\tau]$.

D.2 Smoothed screening and the increment probes

The screening assumption is stated for chosen versions of conditional kernels. This is necessary because regular conditional laws are defined only almost surely at the realized conditioning values, while the Lipschitz estimate below compares the kernels at arbitrary deterministic values u, u' of the noisy slice variable $\widetilde{Y}_i$. This is the same version-level issue that appears in concentration bounds for dependent sequences and Wasserstein mixing coefficients [48, 63, 43]; the noisy slice variables also connect the condition to the smoothed transport viewpoint of [32, 52].

Assumption 6 (Smoothed Wasserstein screening)**.** *For each $i, j \in \{1, \dots, N\}$ there exists a map*

$$\widetilde{K}_{j|i} : \bar{\Omega} \times \mathbb{R}^d \times \mathcal{B}(\mathcal{X}) \to [0, 1] \tag{73}$$

such that $A \mapsto \widetilde{K}_{j|i}(\omega, u, A)$ is a probability law on $\mathcal{X}$ and $(\omega, u) \mapsto \widetilde{K}_{j|i}(\omega, u, A)$ is $\widetilde{\mathcal{F}}_{i-1} \otimes \mathcal{B}(\mathbb{R}^d)$-measurable. For every bounded Borel $\psi : \mathcal{X} \to \mathbb{R}$,

$$\mathbb{E}_{\bar{Q}}[\psi(Y_j) \mid \sigma(\widetilde{\mathcal{F}}_{i-1}, \widetilde{Y}_i)] = \int_{\mathcal{X}} \psi(z) \widetilde{K}_{j|i}(\cdot, \widetilde{Y}_i, dz) \qquad \bar{Q}\text{-a.s.} \tag{74}$$

Moreover, there is a profile $\widetilde{\gamma} : \mathbb{N}_0 \to [0, \infty)$, where $\mathbb{N}_0 = \{0, 1, 2, \dots\}$, with $\widetilde{\gamma}(0) = 1$ such that, for each pair i, j, there is an event $\widetilde{\Omega}_{i,j} \in \widetilde{\mathcal{F}}_{i-1}$ with $\bar{Q}(\widetilde{\Omega}_{i,j}) = 1$ on which, for every $u \in \mathbb{R}^d$, the law $\widetilde{K}_{j|i}(\omega, u, \cdot)$ has finite first moment, and for every $u, u' \in \mathbb{R}^d$,

$$W_1\left(\widetilde{K}_{j|i}(\omega, u, \cdot), \widetilde{K}_{j|i}(\omega, u', \cdot)\right) \le \widetilde{\gamma}(|j - i|)\|u - u'\|. \tag{75}$$

Remark D.2 (Kernel versions)**.** *The probability-one sets in Assumption 6 are part of the screening hypothesis. They ensure that the Wasserstein estimate and the finite-first-moment condition hold simultaneously for all deterministic u, u' after the smoothed past is fixed.*

Define the screened Lipschitz profile

$$a_{i,\tau} = \sum_{j=1}^{N} L_{j,\tau} \widetilde{\gamma}(|j - i|). \tag{76}$$

The next two objects are the probes used by the martingale increment proof. For a fixed smoothed past, $\Phi_i(\omega, u)$ is a version of the conditional prediction of the whole correction field when the ith noisy slice value is set to u. Averaging Φ_i over the Gaussian noise gives H_i, the function of the latent slice Y_i to which the specific conditional MGF bound in Assumption 7 will be applied.

For the increment proof, define the conditional prediction version

$$\Phi_i(\omega, u) = \mathbf{1}_{\widetilde{\Omega}_i}(\omega) \sum_{j=1}^{N} \int_{\mathcal{X}} g_{j,\tau}(z) \widetilde{K}_{j|i}(\omega, u, dz), \qquad \widetilde{\Omega}_i = \bigcap_{j=1}^{N} \widetilde{\Omega}_{i,j}. \tag{77}$$

When the unbounded signed integral is written in this compact form, it is understood via positive and negative parts and truncation as in Lemma B.2; the finite-first-moment condition and the linear growth bound (70) make the integral finite on $\widetilde{\Omega}_i$, while the indicator defines it to be zero on the complement of that probability-one set. Here $\widetilde{\Omega}_i$ is the intersection of the events $\widetilde{\Omega}_{i,j}$ over j, and $\mathbf{1}_{\widetilde{\Omega}_i}$ is its indicator. Finally, define

$$H_i(\omega, y) = \int_{\mathbb{R}^d} \Phi_i(\omega, y + z)\, \Gamma_\xi(dz), \qquad y \in \mathcal{X}. \tag{78}$$

Lemma D.3 (Measurability and Lipschitzness of the probes)**.** *Under Assumptions 5 and 6, Φ_i is $\widetilde{\mathcal{F}}_{i-1} \otimes \mathcal{B}(\mathbb{R}^d)$-measurable, H_i is $\widetilde{\mathcal{F}}_{i-1} \otimes \mathcal{B}(\mathcal{X})$-measurable, and for every ω the maps $u \mapsto \Phi_i(\omega, u)$ and $y \mapsto H_i(\omega, y)$ are $a_{i,\tau}$-Lipschitz.*

Proof. Measurability follows from Lemma B.2, using positive and negative parts of $g_{j,\tau}$ and then truncating to obtain a finite signed version. On $\widetilde{\Omega}_i^c$, $\Phi_i(\omega, \cdot)$ is identically zero; here A^c denotes the complement of a set A. On $\widetilde{\Omega}_i$, for any $u, u' \in \mathbb{R}^d$,

$$\begin{aligned} |\Phi_i(\omega, u) - \Phi_i(\omega, u')| &\le \sum_{j=1}^{N} L_{j,\tau} W_1\left(\widetilde{K}_{j|i}(\omega, u, \cdot), \widetilde{K}_{j|i}(\omega, u', \cdot)\right) \\ &\le \sum_{j=1}^{N} L_{j,\tau} \widetilde{\gamma}(|j - i|)\|u - u'\| = a_{i,\tau}\|u - u'\|, \end{aligned} \tag{79}$$

where the first inequality is Lemma B.8. This proves the Lipschitz bound for Φ_i. The definition of H_i is a Gaussian average of $\Phi_i(\omega, \cdot)$, so

$$|H_i(\omega, y) - H_i(\omega, y')| \le \int |\Phi_i(\omega, y + z) - \Phi_i(\omega, y' + z)|\, \Gamma_\xi(dz) \le a_{i,\tau}\|y - y'\|. \tag{80}$$

The same kernel-measurability argument gives measurability of H_i. □

D.3 Narrowed conditional concentration assumptions

The proof only needs moment-generating-function bounds, abbreviated MGF bounds, for the probes it constructs, not for every Lipschitz function of every conditional law. This is why the assumptions below are stated in a narrowed form. Suitable conditional T_1 inequalities, with the corresponding deterministic bounds on their constants, imply these MGF bounds. See [2, 48, 33, 25]. The Gaussian perturbation supplies the separate conditional bound in (92). The latent and posterior MGF bounds remain explicit assumptions.

Assumption 7 (Specific latent conditional concentration)**.** *Let $\Lambda_i(\omega, \cdot)$ be a regular conditional law of Y_i given $\widetilde{\mathcal{F}}_{i-1}$. There exist constants $\kappa_{i,\tau} \in [0, \infty)$ and events $\Omega_i^{\mathrm{lat}} \in \widetilde{\mathcal{F}}_{i-1}$ with $\bar{Q}(\Omega_i^{\mathrm{lat}}) = 1$ such that, for every $\omega \in \Omega_i^{\mathrm{lat}}$,*

$$\int_{\mathcal{X}} \|y\| \, \Lambda_i(\omega, dy) < \infty, \qquad \int_{\mathcal{X}} |H_i(\omega, y)| \, \Lambda_i(\omega, dy) < \infty, \tag{81}$$

and, for every $\lambda \in \mathbb{R}$,

$$\log \int_{\mathcal{X}} \exp\left(\lambda \left[H_i(\omega, y) - \int_{\mathcal{X}} H_i(\omega, y')\Lambda_i(\omega, dy')\right]\right) \Lambda_i(\omega, dy) \le \frac{\lambda^2 a_{i,\tau}^2 \kappa_{i,\tau}^2}{2}. \tag{82}$$

Remark D.4 (Latent concentration scale)**.** *The constants $\kappa_{i,\tau}$ are concentration scales for the particular probe $H_i(\omega, \cdot)$, conditional on the smoothed past. They need not be interpreted as global concentration constants for the entire conditional law.*

Assumption 8 (Specific posterior conditional concentration)**.** *Let*

$$G_\tau(y_{1:N}) = \sum_{i=1}^{N} g_{i,\tau}(y_i), \qquad L_\tau = \left(\sum_{i=1}^{N} L_{i,\tau}^2\right)^{1/2}. \tag{83}$$

Assume $(Y_1, \ldots, Y_N)$ has finite first moment under $\bar{Q}$ for the Euclidean product metric. There exists $C_{\mathrm{post}} \in (0, \infty)$ such that the conditional exponential moment below is finite and, for every $\lambda \in \mathbb{R}$,

$$\log \mathbb{E}_{\bar{Q}}\left[\exp\left(\lambda\left(G_\tau(Y_{1:N}) - \mathbb{E}_{\bar{Q}}[G_\tau(Y_{1:N}) \mid \widetilde{\mathcal{F}}_N]\right)\right) \mid \widetilde{\mathcal{F}}_N\right] \le \frac{\lambda^2 C_{\mathrm{post}} \xi^2 L_\tau^2}{2} \qquad \bar{Q}\text{-a.s.} \tag{84}$$

Remark D.5 (Posterior residual scale)**.** *The constant C_{post} controls the residual uncertainty that remains after all noisy slices have been conditioned on. The function being concentrated is only G_τ, the correction field viewed as a function of the slice vector $Y_{1:N}$.*

D.4 Increment and residual bounds

Define

$$s_{i,\tau}^2 = a_{i,\tau}^2(\kappa_{i,\tau}^2 + \xi^2), \qquad r_\tau^2 = C_{\mathrm{post}} \xi^2 \sum_{i=1}^{N} L_{i,\tau}^2. \tag{85}$$

Proposition D.6 (Conditional sub-Gaussian increments)**.** *Under Assumptions 5, 6, and 7, for every $i = 1, \ldots, N$ and every $\lambda \in \mathbb{R}$,*

$$\mathbb{E}_{\bar{Q}}\left[e^{\lambda \widetilde{D}_i} \mid \widetilde{\mathcal{F}}_{i-1}\right] \le \exp\left(\frac{\lambda^2 s_{i,\tau}^2}{2}\right) \qquad \bar{Q}\text{-a.s.} \tag{86}$$

Proof. We begin by identifying the increment as a centered conditional prediction, then separate the part due to the Gaussian perturbation from the part due to the latent slice.

Set $\mathcal{H}_i = \sigma(\widetilde{\mathcal{F}}_{i-1}, \widetilde{Y}_i)$, which is $\widetilde{\mathcal{F}}_i$. By the defining property of the kernels $\widetilde{K}_{j|i}$ and Lemma B.4,

$$\mathbb{E}_{\bar{Q}}[g_{j,\tau}(Y_j) \mid \mathcal{H}_i] = \int g_{j,\tau}(z) \widetilde{K}_{j|i}(\cdot, \widetilde{Y}_i, dz) \qquad \bar{Q}\text{-a.s.} \tag{87}$$

Summing over j gives

$$\mathbb{E}_{\bar{Q}}[h_\tau \mid \mathcal{H}_i] = \Phi_i(\cdot, \widetilde{Y}_i) \qquad \bar{Q}\text{-a.s.} \tag{88}$$

Therefore

$$\widetilde{D}_i = \Phi_i(\cdot, \widetilde{Y}_i) - \mathbb{E}_{\bar{Q}}[\Phi_i(\cdot, \widetilde{Y}_i) \mid \widetilde{\mathcal{F}}_{i-1}] \qquad \bar{Q}\text{-a.s.} \tag{89}$$

The next step separates the Gaussian perturbation from the latent uncertainty. Since ζ_i is independent of $\sigma(\widetilde{\mathcal{F}}_{i-1}, Y_i)$ and $\widetilde{Y}_i = Y_i + \zeta_i$, conditionally on $\sigma(\widetilde{\mathcal{F}}_{i-1}, Y_i)$ the only remaining randomness in $\Phi_i(\cdot, \widetilde{Y}_i)$ is the Gaussian variable ζ_i. This is the usual Gaussian concentration step for Lipschitz functions [5, 68, 44]. The map

$$F_i(\omega, z) = \Phi_i(\omega, Y_i(\omega) + z) \tag{90}$$

is $\sigma(\widetilde{\mathcal{F}}_{i-1}, Y_i) \otimes \mathcal{B}(\mathbb{R}^d)$ measurable and $a_{i,\tau}$-Lipschitz in z. Lemma B.9 therefore gives

$$\mathbb{E}_{\bar{Q}}[\Phi_i(\cdot, \widetilde{Y}_i) \mid \widetilde{\mathcal{F}}_{i-1}, Y_i] = H_i(\cdot, Y_i) \qquad \bar{Q}\text{-a.s.} \tag{91}$$

and

$$\mathbb{E}_{\bar{Q}}\left[\exp\left(\lambda(\Phi_i(\cdot, \widetilde{Y}_i) - H_i(\cdot, Y_i))\right) \mid \widetilde{\mathcal{F}}_{i-1}, Y_i\right] \le \exp\left(\frac{\lambda^2 a_{i,\tau}^2 \xi^2}{2}\right). \tag{92}$$

Let

$$c_i = \mathbb{E}_{\bar{Q}}[H_i(\cdot, Y_i) \mid \widetilde{\mathcal{F}}_{i-1}]. \tag{93}$$

Taking conditional expectations in (91) gives $c_i = \mathbb{E}[\Phi_i(\cdot, \widetilde{Y}_i) \mid \widetilde{\mathcal{F}}_{i-1}]$ a.s. Hence, with

$$A_i = \Phi_i(\cdot, \widetilde{Y}_i) - H_i(\cdot, Y_i), \qquad B_i = H_i(\cdot, Y_i) - c_i, \tag{94}$$

we have $\widetilde{D}_i = A_i + B_i$ a.s.

By Lemma B.3 and the definition of Λ_i, $c_i(\omega) = \int H_i(\omega, y)\Lambda_i(\omega, dy)$ on a probability-one event in $\widetilde{\mathcal{F}}_{i-1}$. Assumption 7 therefore yields

$$\mathbb{E}_{\bar{Q}}[e^{\lambda B_i} \mid \widetilde{\mathcal{F}}_{i-1}] \le \exp\left(\frac{\lambda^2 a_{i,\tau}^2 \kappa_{i,\tau}^2}{2}\right) \qquad \bar{Q}\text{-a.s.} \tag{95}$$

Finally, use iterated conditional expectation. The random variable B_i is measurable with respect to $\sigma(\widetilde{\mathcal{F}}_{i-1}, Y_i)$, while (92) controls the conditional MGF of A_i after this sigma-field is fixed. This conditioning step is the same mechanism used in martingale proofs of concentration for dependent sequences [9, 43]. Hence

$$\begin{aligned}
\mathbb{E}[e^{\lambda \widetilde{D}_i} \mid \widetilde{\mathcal{F}}_{i-1}] &= \mathbb{E}\left[e^{\lambda B_i}\mathbb{E}[e^{\lambda A_i} \mid \widetilde{\mathcal{F}}_{i-1}, Y_i] \mid \widetilde{\mathcal{F}}_{i-1}\right] \\
&\le \exp\left(\frac{\lambda^2 a_{i,\tau}^2 \xi^2}{2}\right) \mathbb{E}[e^{\lambda B_i} \mid \widetilde{\mathcal{F}}_{i-1}] \\
&\le \exp\left(\frac{\lambda^2 a_{i,\tau}^2 (\kappa_{i,\tau}^2 + \xi^2)}{2}\right).
\end{aligned} \tag{96}$$

This is (86). □

Lemma D.7 (Posterior residual bound)**.** *Under Assumptions 5 and 8, for every $\lambda \in \mathbb{R}$,*

$$\log \mathbb{E}_{\bar{Q}}\left[e^{\lambda R_\tau} \mid \widetilde{\mathcal{F}}_N\right] \le \frac{\lambda^2 r_\tau^2}{2} \qquad \bar{Q}\text{-a.s.} \tag{97}$$

Proof. The residual is what remains after conditioning on the entire noisy slice vector. At that point the only input needed is the posterior concentration assumption for the slice-vector functional G_τ.

The map $G_\tau(y_{1:N}) = \sum_i g_{i,\tau}(y_i)$ is Borel and has Lipschitz at most constant $L_\tau = (\sum_i L_{i,\tau}^2)^{1/2}$ under the Euclidean product metric, by Cauchy–Schwarz. Since $R_\tau = G_\tau(Y_{1:N}) - \mathbb{E}[G_\tau(Y_{1:N}) \mid \widetilde{\mathcal{F}}_N]$, the assertion is exactly Assumption 8 with $r_\tau^2 = C_{\mathrm{post}} \xi^2 L_\tau^2$. □

D.5 The correction-field MGF

Lemma D.8 (MGF of the martingale sum)**.** *If* (86) *holds for every i, then, for all $\lambda \in \mathbb{R}$,*

$$\log \mathbb{E}_{\bar{Q}}\left[\exp\left(\lambda \sum_{i=1}^{N} \widetilde{D}_i\right)\right] \le \frac{\lambda^2}{2} \sum_{i=1}^{N} s_{i,\tau}^2. \tag{98}$$

Proof. Let $S_n = \sum_{i=1}^{n} \widetilde{D}_i$. Iterated conditioning gives

$$\mathbb{E}[e^{\lambda S_N}] = \mathbb{E}\left[e^{\lambda S_{N-1}} \mathbb{E}[e^{\lambda \widetilde{D}_N} \mid \widetilde{\mathcal{F}}_{N-1}]\right] \le \exp\left(\frac{\lambda^2 s_{N,\tau}^2}{2}\right) \mathbb{E}[e^{\lambda S_{N-1}}]. \tag{99}$$

Repeating the argument from N to 1 yields (98). This is the elementary exponential-supermartingale iteration that underlies Azuma–Hoeffding and McDiarmid-type inequalities [9, 75]. □

Theorem D.9 (Correction-field concentration)**.** *Assume $Q = \Pi_\tau$ is well defined. Assume Assumptions 5, 6, 7, and 8, with the chosen representatives and kernel versions fixed as above. Define*

$$v_\tau = r_\tau^2 + \sum_{i=1}^{N} s_{i,\tau}^2 = C_{\mathrm{post}} \xi^2 \sum_{i=1}^{N} L_{i,\tau}^2 + \sum_{i=1}^{N} a_{i,\tau}^2 (\kappa_{i,\tau}^2 + \xi^2). \tag{100}$$

Then for every $\lambda \in \mathbb{R}$,

$$\log \mathbb{E}_Q\left[\exp\left(\lambda(h_\tau - \mathbb{E}_Q h_\tau)\right)\right] \le \frac{\lambda^2 v_\tau}{2}. \tag{101}$$

Proof. The residual is conditionally sub-Gaussian given the full noisy slice vector, and the increments are conditionally sub-Gaussian along the filtration generated by the noisy slices. Iterating the increment bounds and then applying the residual bound gives the variance proxy in the theorem.

By (72), $h_\tau - \mathbb{E}_Q h_\tau = R_\tau + S_\tau$, where $S_\tau = \sum_i \widetilde{D}_i$ is $\widetilde{\mathcal{F}}_N$-measurable. Hence

$$\mathbb{E}_{\bar{Q}}[e^{\lambda(h_\tau - \mathbb{E}_Q h_\tau)}] = \mathbb{E}_{\bar{Q}}\left[e^{\lambda S_\tau} \mathbb{E}_{\bar{Q}}[e^{\lambda R_\tau} \mid \widetilde{\mathcal{F}}_N]\right]. \tag{102}$$

Use Lemma D.7 to bound the conditional residual MGF, and then Lemma D.8 for S_τ. This gives

$$\mathbb{E}_{\bar{Q}}[e^{\lambda(h_\tau - \mathbb{E}_Q h_\tau)}] \le \exp\left(\frac{\lambda^2 r_\tau^2}{2}\right) \exp\left(\frac{\lambda^2}{2} \sum_{i=1}^{N} s_{i,\tau}^2\right). \tag{103}$$

Taking logarithms proves the claim. Since h_τ depends only on θ and the θ-marginal of $\bar{Q}$ is Q, the MGF under $\bar{Q}$ equals the MGF under Q. □

E Proofs for KL and noise-level rates

It remains to convert the correction-field MGF into the discrepancy bound and then specialize the variance proxy to the EDM noise scale.

At this point the probability estimate is already contained in Theorem D.9. The KL step only uses the tilt identity and evaluates the log-Laplace transform at $\lambda = -1$, a standard entropy/log-partition calculation closely related to the Gibbs variational principle [26, 62, 19].

Theorem E.1 (Surrogate-to-bridge KL and total variation)**.** *Under the assumptions of Theorems C.5 and D.9, with the same representatives of h_τ,*

$$\mathrm{KL}(\Pi_\tau \| P^\star) \le \frac{v_\tau}{2}, \qquad \|\Pi_\tau - P^\star\|_{\mathrm{TV}} \le \frac{1}{2}\sqrt{v_\tau}. \tag{104}$$

Proof. Let $Q = \Pi_\tau$ and set $X_\tau = h_\tau - \mathbb{E}_Q h_\tau$. Centering only multiplies numerator and denominator of the tilt identity by the same constant, so

$$\frac{dP^\star}{dQ} = \frac{e^{-X_\tau}}{\mathbb{E}_Q[e^{-X_\tau}]}. \tag{105}$$

Therefore $Q \ll P^\star$ and

$$\frac{dQ}{dP^\star} = \mathbb{E}_Q[e^{-X_\tau}]e^{X_\tau}. \tag{106}$$

Taking Q-expectations of the logarithm gives

$$\mathrm{KL}(Q\|P^\star) = \log \mathbb{E}_Q[e^{-X_\tau}] + \mathbb{E}_Q[X_\tau] = \log \mathbb{E}_Q[e^{-X_\tau}]. \tag{107}$$

The concentration theorem at $\lambda = -1$ gives $\log \mathbb{E}_Q[e^{-X_\tau}] \le v_\tau/2$. Pinsker's inequality [60, 19], with $\|\mu - \nu\|_{\mathrm{TV}} = \sup_A |\mu(A) - \nu(A)|$, gives

$$\|Q - P^\star\|_{\mathrm{TV}} \le \sqrt{\frac{1}{2}\mathrm{KL}(Q\|P^\star)} \le \frac{1}{2}\sqrt{v_\tau}. \tag{108}$$

□

To obtain the noise-level statement, set the smoothing scale equal to the EDM noise level. Under the uniform bounds below, noise-alignment and summable screening give $a_{i,\tau} = O(\sigma(\tau))$, and the residual proxy is $O(N\sigma(\tau)^4)$.

Corollary E.2 (Noise-level rate)**.** *Set $\xi = \sigma(\tau) > 0$. Assume the hypotheses of Theorem E.1 hold at each noise level considered, and suppose*

$$\max_i \kappa_{i,\tau} \le K_{\mathrm{data}}, \qquad \max_i L_{i,\tau} \le c_L\sigma(\tau), \qquad \sum_{r=1}^{\infty} \widetilde{\gamma}(r) \le \widetilde{\Gamma}_1, \qquad C_{\mathrm{post}} \le C_{\mathrm{post}}^{\max}. \tag{109}$$

Here $K_{\mathrm{data}}, c_L, \widetilde{\Gamma}_1$, and $C_{\mathrm{post}}^{\max}$ are finite and independent of τ over these noise levels. Then

$$\mathrm{KL}(\Pi_\tau\|P^\star) \le \frac{N\sigma(\tau)^2}{2}\left[c_L^2(1+2\widetilde{\Gamma}_1)^2(K_{\mathrm{data}}^2 + \sigma(\tau)^2) + C_{\mathrm{post}}^{\max}c_L^2\sigma(\tau)^2\right]. \tag{110}$$

Consequently, as $\sigma(\tau) \downarrow 0$,

$$\mathrm{KL}(\Pi_\tau\|P^\star) = O(N\sigma(\tau)^2), \qquad \|\Pi_\tau - P^\star\|_{\mathrm{TV}} = O(\sqrt{N}\,\sigma(\tau)). \tag{111}$$

The displayed dependence on N is uniform when these four bounding constants are also independent of N.

Proof. By the alignment and screening assumptions,

$$a_{i,\tau} = \sum_{j=1}^{N} L_{j,\tau}\widetilde{\gamma}(|j-i|) \le c_L\sigma(\tau)\left(1 + 2\sum_{r=1}^{\infty}\widetilde{\gamma}(r)\right) \le c_L\sigma(\tau)(1+2\widetilde{\Gamma}_1). \tag{112}$$

Thus

$$\sum_{i=1}^{N} s_{i,\tau}^2 \le Nc_L^2\sigma(\tau)^2(1+2\widetilde{\Gamma}_1)^2(K_{\mathrm{data}}^2 + \sigma(\tau)^2). \tag{113}$$

The residual satisfies

$$r_\tau^2 = C_{\mathrm{post}}\sigma(\tau)^2\sum_{i=1}^{N} L_{i,\tau}^2 \le NC_{\mathrm{post}}^{\max}c_L^2\sigma(\tau)^4. \tag{114}$$

Substituting these estimates into $v_\tau = r_\tau^2 + \sum_i s_{i,\tau}^2$ and applying Theorem E.1 gives (110). The total variation rate follows from the total variation bound in Theorem E.1. □

Remark E.3 (Why this KL direction?)**.** *The expectation in the log-Laplace transform is under the base law $Q = \Pi_\tau$, which is why the direct argument gives $\mathrm{KL}(\Pi_\tau\|P^\star)$. The total-variation consequence is symmetric, so it controls the distributional gap without choosing a direction.*

E.1 Verification of the Gaussian example

We verify the example in Section 4, using the same correction representatives in the tilt and concentration arguments. Write $\sigma = \sigma(\tau) > 0$ and use σ to index the example. Let $\Theta = \mathbb{R}^4$, $\mathcal{X} = \mathbb{R}$, and $S_i(\theta) = \theta_i$. For a fixed nonzero $c \in \mathbb{R}^4$, take

$$p_0 = \bigotimes_{i=1}^{4} \mathcal{N}(c_i, 1), \qquad q_{s_i} = \mathcal{N}(-c_i, 1), \qquad i = 1, \dots, 4. \tag{115}$$

Set $w_i = 1/4$, $\sigma_{\mathrm{data}} = 1/2$, and $\xi = \sigma$. Then

$$\beta(\sigma) = \frac{1}{\sigma^2 + 1/4}, \qquad \rho_\sigma = 1 - \frac{\beta(\sigma)}{4} = \frac{4\sigma^2}{1 + 4\sigma^2}. \tag{116}$$

Bridge and surrogate. Let $P^\star = \bigotimes_{i=1}^{4} q_{s_i}$. For any feasible P with finite $\mathrm{KL}(P\|p_0)$, the prescribed Gaussian marginals make $\log(dP^\star/dp_0) = -2\sum_i c_i\theta_i$ integrable, and

$$\mathrm{KL}(P\|p_0) = \mathrm{KL}(P\|P^\star) + 2\sum_{i=1}^{4} c_i^2, \tag{117}$$

since $\mathbb{E}_P\theta_i = -c_i$. Nonnegativity of KL shows that $P^\star$ is the unique minimizer over all feasible laws, with finite objective $2\sum_i c_i^2$. Candidates with infinite KL cannot improve this value. The density ratios are positive and finite, and we may choose $\ell_i(y) = \varphi_i^\star(y) = -2c_i y$ and $Z_\star = 1$. Using the product disintegration of p_0, we obtain $\mathcal{C}_i(y) = 1$ and $f_i(y) = 0$ for every $y \in \mathbb{R}$.

Completing the square in the surrogate density gives

$$\begin{aligned} \Pi_\sigma &= \bigotimes_{i=1}^{4} \mathcal{N}(m_i, 1), \qquad m_i = -c_i + 2\rho_\sigma c_i, \\ Z_{\Pi,\sigma} &= \exp\left(-2(1-\rho_\sigma)\rho_\sigma \sum_{i=1}^{4} c_i^2\right). \end{aligned} \tag{118}$$

In particular, the normalizer is finite and positive. This verifies Assumptions 1–4. The representatives in (62) are

$$\begin{aligned} g_{i,\sigma}(y) &= 2\rho_\sigma c_i y, \\ L_{i,\sigma} &= 2\rho_\sigma|c_i| = \frac{8|c_i|\sigma^2}{1+4\sigma^2} \le 2|c_i|\sigma. \end{aligned} \tag{119}$$

They are globally Lipschitz, as required by Assumption 5. The last inequality follows from $4\sigma \le 1 + 4\sigma^2$ and holds for every $\sigma > 0$.

Screening and conditional concentration. Under $Q = \Pi_\sigma$, the coordinates Y_i are independent with laws $\mathcal{N}(m_i, 1)$. Put $\eta_\sigma = (1+\sigma^2)^{-1}$ and add the independent reveal noises with variance σ^2. Choose the Gaussian conditional kernels

$$\widetilde{K}_{j|i}(\omega, u, \cdot) = \begin{cases} \mathcal{N}\Big(m_j + \eta_\sigma(\widetilde{Y}_j(\omega) - m_j), \sigma^2\eta_\sigma\Big), & j < i, \\ \mathcal{N}\big(m_i + \eta_\sigma(u - m_i), \sigma^2\eta_\sigma\big), & j = i, \\ \mathcal{N}(m_j, 1), & j > i. \end{cases} \tag{120}$$

These versions have the measurability required by Assumption 6 and finite first moments for every noisy history and every $u \in \mathbb{R}$. Only the $j = i$ kernel depends on u; its W_1 distance at u, u' is $\eta_\sigma|u - u'| \le |u - u'|$. Thus the screening assumption holds with $\widetilde{\gamma}(0) = 1$ and $\widetilde{\gamma}(r) = 0$ for $r \ge 1$, taking all probability-one events to be the whole space. In particular, $a_{i,\sigma} = L_{i,\sigma}$.

For these kernels, the probes $\Phi_i(\omega, u)$ and $H_i(\omega, y)$ in (77)–(78) are affine in their last arguments, both with slope $2\rho_\sigma c_i\eta_\sigma$. The conditional law of Y_i given $\widetilde{\mathcal{F}}_{i-1}$ remains $\Lambda_i(\omega, \cdot) = \mathcal{N}(m_i, 1)$. Consequently, the centered log-MGF in (82) equals $\lambda^2\eta_\sigma^2 L_{i,\sigma}^2/2$ for every $\lambda \in \mathbb{R}$. This verifies Assumption 7 with $\kappa_{i,\sigma} = \eta_\sigma$.

Given all four noisy slices, the posterior covariance is $\sigma^2\eta_\sigma I_4$. Since $G_\sigma(y_{1:4}) = 2\rho_\sigma \sum_i c_i y_i$, the residual R_σ is conditionally centered Gaussian with variance $\sigma^2\eta_\sigma \sum_i L_{i,\sigma}^2$. Its conditional log-MGF is therefore $\lambda^2\sigma^2\eta_\sigma \sum_i L_{i,\sigma}^2/2$ for all $\lambda \in \mathbb{R}$, proving Assumption 8 with $C_{\mathrm{post}} = \eta_\sigma$. The Gaussian laws and affine probes also give all the required first-moment and integrability conditions. Hence all the assumptions hold jointly, and the rate constants can be chosen as

$$c_L = 2\max_i |c_i|, \qquad K_{\mathrm{data}} = 1, \qquad \widetilde{\Gamma}_1 = 0, \qquad C_{\mathrm{post}}^{\max} = 1, \tag{121}$$

independently of σ.

Discrepancy and variance proxy. The bridge and surrogate have identity covariance, so their KL divergence is

$$\begin{aligned} \mathrm{KL}(\Pi_\sigma \| P^\star) &= \frac{1}{2}\sum_{i=1}^4 (m_i + c_i)^2 = 2\rho_\sigma^2 \sum_{i=1}^4 c_i^2 \\ &= \frac{32\sigma^4}{(1+4\sigma^2)^2} \sum_{i=1}^4 c_i^2 > 0. \end{aligned} \tag{122}$$

Substituting the constants above into (100) gives

$$\begin{aligned} v_\sigma &= \left(\eta_\sigma\sigma^2 + \eta_\sigma^2 + \sigma^2\right) \sum_{i=1}^4 L_{i,\sigma}^2, \\ \frac{v_\sigma/2}{\mathrm{KL}(\Pi_\sigma \| P^\star)} &= 1 + \frac{\sigma^4(2+\sigma^2)}{(1+\sigma^2)^2} > 1. \end{aligned} \tag{123}$$

Thus the bound applies at every positive noise level with a nonzero discrepancy. As $\sigma \downarrow 0$ in this fixed-N example, the stronger contraction $L_{i,\sigma} = O(\sigma^2)$ yields an actual KL of order σ^4, consistent with the corollary's $O(\sigma^2)$ upper bound for fixed $N = 4$.

F Variance bound under the clean filtration

We give an auxiliary variance bound under the clean filtration $\mathcal{F}_i = \sigma(Y_1, \ldots, Y_i)$. The bound is logically separate from the smoothed-filtration MGF bound used in Theorem E.1; we include it because the calculation makes the role of screening explicit: it shows how Doob orthogonality and Wasserstein contraction combine to control the variance of h_τ. The calculation is a direct application of martingale orthogonality and Wasserstein contraction, in the same broad family as martingale variance and concentration arguments for dependent data [36, 43].

Fix $Q = \Pi_\tau$ and work on $(\Theta, \mathcal{B}(\Theta), Q)$. Let $Y_i = S_i(\theta)$ and $h_\tau = \sum_i g_{i,\tau}(Y_i)$.

Assumption 9 (Finite second moments)**.** *For each i, $\mathbb{E}_Q\|Y_i\|^2 < \infty$.*

Definition F.1 (Innovation size)**.** *For $i = 1, \ldots, N$, define*

$$\eta_i^2(\tau) = \mathbb{E}_Q\left[\left\|Y_i - \mathbb{E}_Q[Y_i \mid \mathcal{F}_{i-1}]\right\|^2\right]. \tag{124}$$

Assumption 10 (Clean kernel screening)**.** *For each $1 \le i \le j \le N$, there is a kernel $K_{j|i} : \Theta \times \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ such that $(\theta, y) \mapsto K_{j|i}(\theta, y, A)$ is $\mathcal{F}_{i-1} \otimes \mathcal{B}(\mathcal{X})$-measurable for every $A \in \mathcal{B}(\mathcal{X})$. For every bounded Borel ψ,*

$$\mathbb{E}_Q[\psi(Y_j) \mid \sigma(\mathcal{F}_{i-1}, Y_i)] = \int \psi(z) K_{j|i}(\cdot, Y_i, dz) \qquad Q\text{-a.s.} \tag{125}$$

There is a profile $\gamma : \mathbb{N}_0 \to [0, \infty)$ with $\gamma(0) = 1$ and

$$\Gamma_2 = \sum_{r=0}^\infty \gamma(r)^2 < \infty \tag{126}$$

such that, on an event in $\mathcal{F}_{i-1}$ with Q-probability one, the conditional laws have finite first moments and

$$W_1(K_{j|i}(\theta, y, \cdot), K_{j|i}(\theta, y', \cdot)) \le \gamma(j-i)\|y - y'\| \tag{127}$$

for every deterministic $y, y' \in \mathcal{X}$.

Lemma F.2 (Clean Doob decomposition)**.** *Assume $h_\tau \in L^2(Q)$, and let*

$$M_i = \mathbb{E}_Q[h_\tau \mid \mathcal{F}_i], \qquad D_i = M_i - M_{i-1}. \tag{128}$$

Then

$$h_\tau - \mathbb{E}_Q h_\tau = \sum_{i=1}^{N} D_i, \qquad \mathrm{Var}_Q(h_\tau) = \sum_{i=1}^{N} \mathbb{E}_Q[D_i^2]. \tag{129}$$

Proof. $(M_i)_{i=0}^N$ is an L^2 martingale. The first identity is telescoping. The second follows because martingale differences are orthogonal in L^2 [36]. □

Lemma F.3 (L^2 well-posedness)**.** *Under Assumptions 5 and 9, each $g_{i,\tau}(Y_i)$ belongs to $L^2(Q)$, and hence $h_\tau \in L^2(Q)$.*

Proof. The linear-growth bound (70) gives $|g_{i,\tau}(Y_i)| \le |g_{i,\tau}(0)| + L_{i,\tau}\|Y_i\|$, which is square integrable by Assumption 9. Since N is finite, h_τ is also square integrable. □

Proposition F.4 (Clean-filtration variance bound)**.** *Under Assumptions 5, 9, and 10,*

$$\mathrm{Var}_Q(h_\tau) \le \Gamma_2 \sum_{i=1}^{N} \eta_i^2(\tau) \sum_{j=i}^{N} L_{j,\tau}^2. \tag{130}$$

Proof. Let $U_i = \sum_{j=i}^N g_{j,\tau}(Y_j)$. Since $\sum_{j<i} g_{j,\tau}(Y_j)$ is $\mathcal{F}_{i-1}$-measurable,

$$D_i = \mathbb{E}_Q[U_i \mid \mathcal{F}_i] - \mathbb{E}_Q[U_i \mid \mathcal{F}_{i-1}]. \tag{131}$$

For this proof, define the auxiliary prediction

$$\Phi_i(\theta, y) = \sum_{j=i}^{N} \int g_{j,\tau}(z) K_{j|i}(\theta, y, dz) \tag{132}$$

on the common probability-one event in $\mathcal{F}_{i-1}$ where these integrals and the screening bounds hold for all $j \ge i$, and set it to zero on the complement. This defines an $\mathcal{F}_{i-1} \otimes \mathcal{B}(\mathcal{X})$-measurable function. Then

$$\mathbb{E}_Q[U_i \mid \mathcal{F}_i] = \Phi_i(\cdot, Y_i) \qquad Q\text{-a.s.} \tag{133}$$

On this event, the Lipschitz–Wasserstein estimate gives

$$\mathrm{Lip}(\Phi_i(\theta, \cdot)) \le \sum_{j=i}^{N} L_{j,\tau} \gamma(j-i). \tag{134}$$

Conditionally on $\mathcal{F}_{i-1}$, this is a fixed Lipschitz function. Apply the pairwise variance identity to two independent draws from the conditional law of Y_i. The preceding Lipschitz bound, followed by taking expectations, gives

$$\mathbb{E}_Q[D_i^2] \le \left(\sum_{j=i}^{N} L_{j,\tau} \gamma(j-i) \right)^2 \eta_i^2(\tau). \tag{135}$$

By Cauchy–Schwarz,

$$\left(\sum_{j=i}^{N} L_{j,\tau} \gamma(j-i) \right)^2 \le \left(\sum_{j=i}^{N} L_{j,\tau}^2 \right) \left(\sum_{j=i}^{N} \gamma(j-i)^2 \right) \le \Gamma_2 \sum_{j=i}^{N} L_{j,\tau}^2. \tag{136}$$

Summing over i and using Lemma F.2 gives (130). □

Remark F.5 (Status of the clean bound)**.** *The clean-filtration result is an exact variance estimate, not an MGF estimate. It is included only as an additional decomposition. The main KL theorem uses the smoothed-filtration concentration bound in Appendix D.*

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Section 1 state the tilt-identity reduction and the noise-level rate as the paper's main results, and these are proved in Sections 2–4 with full versions in the appendix.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 6 discusses the scope of the rate (linear-in-N scaling, open lower bound) and the gap between the static surrogate analyzed here and the dynamic samplers used in practice.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Main-text theorems are stated with proof sketches and pointers to the appendix; full assumptions and complete proofs are given in Appendices B–F.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Algorithm 1 specifies the sampler in full, Table 1 lists all hyperparameters, and Appendix A describes the ADNI setup, baselines, and evaluation protocol; ADNI is publicly accessible upon registration.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The algorithm, hyperparameters, and evaluation protocol are specified in full in Algorithm 1, Table 1, and Appendix A, and ADNI is publicly available upon registration.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Appendix A describes the ADNI preprocessing, the frozen 2D EDM denoiser, the baselines, and the evaluation protocol; Table 1 lists the schedule, latent dimension, mini-batch size, and step counts used for Algorithm 1.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Table 2 reports mean $\pm$ standard deviation for SSIM, LPIPS, MI, and the TV anisotropy ratio.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix A.2 reports the GPU type and memory, the deep learning framework, and the optimizer used for training the 2D denoiser.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

Justification: The paper's contributions are theoretical convergence rates for a surrogate construction.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [Yes]

Justification: Release of the 2D EDM denoiser checkpoint will be gated by the ADNI data use agreement; the model is a 2D slice prior for brain MRI and is not intended for clinical use, which we will state in the model card.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: ADNI [58] is used under its standard data use agreement; EDM [40], Instant-NGP [50], MedicalNet [12], and the TPDM [45] and ISCS [28] baselines are cited at point of use.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: No new assets are released at submission time.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or new research with human subjects; it uses the existing ADNI corpus.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

Justification: The paper does not conduct new research with human subjects; ADNI was collected under its own IRB-approved protocol and is used here under its standard data use agreement.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not used as a core component of the methods or experiments in this paper.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.