POSTERIOR TRANSPORT MATCHING FOR INVERSE PROBLEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Posterior sampling must preserve uncertainty while remaining consistent with the observation. We introduce Posterior Transport Matching (PTM), a training-free framework that pairs numerical transport maps with stochastic input updates, termed rematching. The key is to supply each map with an input distribution that it sends to the posterior. We characterize the minimum-energy isotropic diffusion realizing these compatible laws and identify its objective with path-space relative entropy. Exact probability-flow maps admit Gaussian rematching. For computation, entropy projection yields an objective for the Gaussian transition mean, combining an expected potential with a squared-distance penalty. Denoised observation energy approximates the potential; an analytic approximation requires only the observation operator. A single bound on output KL separates potential approximation from finite computation. Controlled experiments show how the appropriate transition depends on the map. Across eight FFHQ inverse problems, PTM achieves the highest reported PSNR across eight FFHQ inverse problems and favorable quality–cost tradeoffs, with further evaluation on ImageNet.

1 INTRODUCTION

Pretrained diffusion models provide image priors that can be conditioned on observations without retraining (Song et al., 2021; Chung et al., 2023; Kawar et al., 2022). Likelihood guidance, conditional moments, and decomposed sampling make this principle effective across inverse problems (Song et al., 2023a; Boys et al., 2024; Zhang et al., 2025). Posterior sampling also requires transporting uncertainty. An accurate clean estimate at one noise level does not by itself determine the distribution of finite-step outputs.

Flow matching specifies a probability path through its velocity field (Lipman et al., 2023); consistency models, flow maps, and MeanFlow describe finite transport through endpoints or interval-average velocities (Song et al., 2023b; Boffi et al., 2025; Geng et al., 2025). Tilted flows and transition samplers provide further ways to condition this transport (Potaptchik et al., 2026; Holderrieth et al., 2026). For inverse problems with a frozen scalar-noise denoiser, endpoint evaluation has two computational requirements: conditioning changes the denoising geometry, and density transport introduces a volume correction. Local denoiser responses provide inexpensive numerical maps. Their use raises a distributional question: which input law should the map receive?

Our central observation is that posterior preservation constrains the map and its input law jointly. A compatible input law allows a numerical map to represent the posterior even when it differs from the probability-flow map. *Posterior Transport Matching (PTM)* therefore treats the stochastic update between map evaluations as a matching problem. *Rematching* evolves the inputs through isotropic diffusion while their mapped outputs retain the posterior. We characterize this compatible evolution and derive a finite transition objective from relative entropy. Its Gaussian approximation optimizes a center around the computed endpoint and adds independent noise, using a frozen denoiser and the observation operator. Our contributions are:

- **Map–transition compatibility.** We formulate a cross-level matching condition that preserves the posterior for general families of numerical maps. Exact probability-flow transport recovers Gaussian rematching as a special case.

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

- **Minimum-energy rematching.** We characterize compatible diffusion through a variational objective and establish its path-space KL representation. For smooth invertible maps, stochastic composition incorporates volume correction without a separate determinant evaluation.
- **Computable transitions and error control.** We derive an objective for Gaussian transitions, approximate its potential by denoised observation energy, and obtain an analytic approximation using only the observation operator. A single bound controls the resulting output KL. Experiments evaluate distributional accuracy in controlled problems and reconstruction quality and cost on FFHQ and ImageNet.

Our local conditioning calculation uses conditional moments (Boys et al., 2024), and DAPS provides a related clean-state and noising composition (Zhang et al., 2025). The variational characterization builds on entropic Action Matching (Neklyudov et al., 2023). PTM uses these tools to specify compatibility for the complete numerical map and analyze its stochastic transition; Appendix G gives further comparisons.

2 Background and Motivation: Posterior Transport

For an observation $y = A(Z) + \eta$ with $Z \sim p_0$ on $\mathbb{R}^n$ and independent $\eta \sim \mathcal{N}(0, R)$, $R \succ 0$, the Bayesian target is

$$\pi^y(z) = Z_y^{-1} p_0(z) e^{-E(z)}, \qquad E(z) = \tfrac{1}{2}\|A(z) - y\|^2_{R^{-1}}, \qquad q_\tau^y = \mathcal{H}_\tau \pi^y. \tag{1}$$

Here $Z_y \in (0, \infty)$ and $\mathcal{H}_\tau$ convolves with $\mathcal{N}(0, \tau I)$; τ denotes noise variance throughout. The frozen prior denoiser $\mathsf{D}_\tau(x)$ approximates $\mathbb{E}_{p_0}[Z \mid Z + \sqrt{\tau}\xi = x]$, where $\xi \sim \mathcal{N}(0, I)$ is independent of Z. We seek a sampler for π^y using this frozen prior (Stuart, 2010).

Pretrained denoisers support plug-and-play reconstruction (Venkatakrishnan et al., 2013; Zhu et al., 2023), spectral or null-space updates (Kawar et al., 2022; Wang et al., 2023), and likelihood guidance (Chung et al., 2023; Song et al., 2023a). Conditional-moment and variational methods refine local posterior estimates (Boys et al., 2024; Peng et al., 2024; Mardani et al., 2024); sequential Monte Carlo and decomposed samplers operate through stochastic transitions (Wu et al., 2023; Dou & Song, 2024; Coeurdoux et al., 2024; Xu & Chi, 2024; Zhang et al., 2025). We study the distributional condition linking an endpoint calculation to the transition that supplies its next input.

Local conditioning does not determine finite transport. Tweedie's identity gives the posterior velocity through $m_\tau^y(x) = \mathbb{E}[Z \mid X_\tau = x, y]$, where $X_\tau = Z + \sqrt{\tau}\xi$ (Efron, 2011; Song et al., 2021). The probability-flow map F_τ transports the full law: $(F_\tau)_{\#} q_\tau^y = \pi^y$ (Song et al., 2021; Boffi et al., 2025). Here $F_{\#}q$ denotes the distribution of $F(X)$ for $X \sim q$. The distinction already appears for $\pi^y = \mathcal{N}(0, v)$:

$$m_\tau^y(x) = \frac{v}{v+\tau} x, \qquad F_\tau(x) = \sqrt{\frac{v}{v+\tau}}\, x. \tag{2}$$

For the same input $X_\tau \sim \mathcal{N}(0, v + \tau)$, the mean has output variance $v^2/(v+\tau)$, while the transport map has variance v. Finite transport also depends on how the conditional mean changes away from the source. Consistency models, flow maps, and interval-average velocities describe this transport (Song et al., 2023b; Boffi et al., 2025; Geng et al., 2025).

Computing endpoints from a frozen denoiser. Trajectory integration requires repeated velocity and conditioning evaluations. Direct consistency calculations require map derivatives, including a spatial volume correction. Covariance-based alternatives impose additional structure: $\tau J_x \mathsf{D}_\tau = \mathrm{Cov}[Z \mid X_\tau = x]$ holds for an exact conditional-mean denoiser, while a learned denoiser's Jacobian need not be symmetric or positive semidefinite (Boys et al., 2024). This limits its direct use as a covariance or covariance square root. Local conditional moments provide inexpensive numerical endpoints whose transport error can be analyzed.

The input law is part of the sampler. In the Gaussian example, changing the input variance can compensate for the mean map's contraction. This suggests fixing a computationally accessible map and constructing the input law that it sends to the posterior. Rematching supplies this law at the

next noise level. The target remains π^y while its representation in noisy coordinates adapts to the numerical map.

3 METHOD

PTM alternates a numerical transport map with rematching. Given x at variance τ, the map returns $z = \Psi_\tau(x)$; rematching constructs its next noisy input. The theory specifies posterior-preserving compatibility between these operations. For computation, we use Gaussian transitions and approximate their matching potential by denoised observation energy. At the next variance τ', optimize the Gaussian center and add independent noise:

$$\mathcal{L}_{\tau',z}(u) = \mathbb{E}_\xi E\Big(\mathsf{D}_{\tau'}(u+\sqrt{\tau'}\xi)\Big) + \tfrac{\|u-z\|^2}{2\tau'} \simeq \underbrace{E(u) + \tfrac{\|u-z\|^2}{2\tau'}}_{V_{\tau',z}(u)} + \text{const},$$
$$\widehat{u} = \textsc{Minimize}(\mathcal{L}_{\tau',z}, z; B), \qquad x' = \widehat{u} + \sqrt{\tau'}\eta. \tag{3}$$

Here $\xi, \eta \sim \mathcal{N}(0, I)$, with the sampling noise η independent of all draws used in optimization. The quadratic term is the KL divergence between Gaussian transitions with covariance $\tau' I$ and centers u and z. The analytic objective $V_{\tau',z}$ removes denoiser evaluations; for affine observations, the omitted remainder is given in Equation (15). Section 3.3 derives the objective and separately bounds errors from potential approximation and finite optimization. The budget B controls finite optimization; $B = 0$ leaves $\widehat{u} = z$ and gives Gaussian rematching.

Algorithm 1 implements both evaluations. We next construct the map, characterize compatible transitions, and derive the finite objective.

Algorithm 1 Posterior Transport Matching: single-particle sampler

Input: D, A, y, R, variances $\tau_0 > \cdots > \tau_K > 0$, optimization budgets B_k, sample count M, and sampled or analytic evaluation; $E(u) = \frac{1}{2}\|A(u) - y\|^2_{R^{-1}}$.

function TRANSPORT(x, τ)
  $d \leftarrow \mathsf{D}_\tau(x)$
  $H \leftarrow J_A(d)$
  $r \leftarrow y - A(d)$
  $\widehat{C}v := \tau J_x \mathsf{D}_\tau(x) v$
  $\mathcal{B}v := Rv + H\widehat{C}(H^\top v)$
  $w \leftarrow \mathrm{GMRES}(\mathcal{B}, r)$
  $p \leftarrow \widehat{C}H^\top w$
  $g \leftarrow -H^\top R^{-1} r$
  $h \leftarrow \frac{\|g\|^2}{\|Hg\|^2_{R^{-1}} + \tau^{-1}\|g\|^2}$ ($h = 0$ if $g = 0$)
  $\mathcal{U} \leftarrow (d + p, d + p/4, d + p/16, d + p/64, d - hg, d - hg/4)$
  $e_d \leftarrow E(d)$
  for $u \in \mathcal{U}$ in order **do**
    if $E(u) < e_d$ **then**
      return u
    end if
  end for
  return d
end function

$x \sim \mathcal{N}(0, \tau_0 I)$
for $k = 0, \ldots, K-1$ **do**
  $\tau \leftarrow \tau_k, \quad \tau' \leftarrow \tau_{k+1}$
  $z \leftarrow$ TRANSPORT(x, τ)
  $u \leftarrow z$ ▷ Gaussian rematching if $B_k = 0$
  if $B_k > 0$ **then**
    if sampled evaluation **then**
      $\xi_1, \ldots, \xi_M \overset{\text{iid}}{\sim} \mathcal{N}(0, I)$
      $\ell(u) := \frac{1}{M}\sum_{j=1}^M E(\mathsf{D}_{\tau'}(u + \sqrt{\tau'}\xi_j))$
    else ▷ analytic approximation
      $\ell(u) := E(u)$
    end if
    $\widehat{\mathcal{L}}(u) := \ell(u) + \|u - z\|^2/(2\tau')$
    $u \leftarrow \textsc{Minimize}(\widehat{\mathcal{L}}, z; B_k)$
  end if
  $\eta \sim \mathcal{N}(0, I)$ ▷ fresh, independent noise
  $x \leftarrow u + \sqrt{\tau'}\eta$
end for
return $\mathsf{D}_{\tau_K}(x)$

3.1 TRANSPORT MAPS FROM A FROZEN DENOISER

The clean conditional distribution $\rho^y_\tau(z \mid x) \propto p_0(z)\exp[-\|z - x\|^2/(2\tau) - E(z)]$ determines the posterior velocity through its mean m^y_τ: $v^y_\tau(x) = (x - m^y_\tau(x))/(2\tau) = -\frac{1}{2}\nabla_x \log q^y_\tau(x)$. The probability-flow map F_τ is constant along the corresponding characteristics and transports $q^y_\tau = \mathcal{H}_\tau \pi^y$ to π^y (Song et al., 2021; Boffi et al., 2025; Geng et al., 2025). For a smooth, orientation-

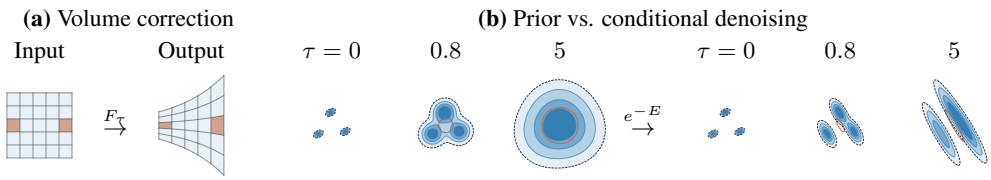


Figure 1: **Two obstacles to computing transport maps.** (a) Equal source cells acquire unequal endpoint volumes. (b) Here conditioning makes the denoising kernel anisotropic (brown outlines). Blue densities show the same prior convolved with each kernel.

preserving flow map, write $J_\tau = \nabla_x F_\tau$ and $F_0 = \mathrm{Id}$. Consistency and change of variables give

$$\partial_\tau F_\tau = \tfrac{1}{2} J_\tau \nabla_x \log q_\tau^y, \qquad q_\tau^y = (\pi^y \circ F_\tau) \det J_\tau. \tag{4}$$

Eliminating the noisy density exposes the two contributions to endpoint transport:

$$\partial_\tau F_\tau = \underbrace{\tfrac{1}{2} J_\tau J_\tau^\top [(\nabla \log p_0 - \nabla E) \circ F_\tau]}_{\text{posterior score transport}} + \underbrace{\tfrac{1}{2} J_\tau \nabla_x \log \det J_\tau}_{\text{volume correction}}. \tag{5}$$

The volume term depends on second spatial derivatives of the map (Figure 1a). The regularity assumptions and associated Doob h-transform are given in Appendices B and B.1 (Chopin et al., 2023).

Conditioning poses a separate computational requirement. For the prior clean conditional ρ_τ^0 and a characteristic with $X_\tau = x$,

$$m_\tau^y(x) = \frac{\mathbb{E}_{\rho_\tau^0(\cdot|x)}[Z e^{-E(Z)}]}{\mathbb{E}_{\rho_\tau^0(\cdot|x)}[e^{-E(Z)}]}, \qquad F_\tau(x) = m_\tau^y(x) - \int_0^\tau \mathcal{D}_s^y m_s^y(X_s)\, \mathrm{d}s, \tag{6}$$

where $\mathcal{D}_s^y = \partial_s + v_s^y \cdot \nabla_x$ (Appendix B.3). A scalar-noise prior denoiser gives the *unweighted* mean. The factor e^{-E} changes the conditional geometry: even a linear observation generally requires an anisotropic denoising kernel, while nonlinear observations also introduce nonquadratic energy (Figure 1b). The integral converts the source conditional mean into a transport endpoint by accounting for its change along the characteristic.

We choose a tractable numerical map Ψ_τ without assuming that it exactly transports the Gaussian-smoothed posterior q_τ^y. Section 3.2 determines the input law required by this map. Its local proposal uses denoiser values and Jacobian products to approximate the conditional mean:

$$\begin{aligned} d = \mathsf{D}_\tau(x), \qquad \widehat{C} v = \tau J_x \mathsf{D}_\tau(x) v, \qquad H = J_A(d), \\ (R + H\widehat{C}H^\top) w = y - A(d), \qquad \widehat{m}_\tau(x) = d + \widehat{C} H^\top w. \end{aligned} \tag{7}$$

For an exact denoiser, $\widehat{C} = \mathrm{Cov}[Z \mid X_\tau = x]$ (Boys et al., 2024). Gaussian prior conditionals, affine observations, and an exact solve give the exact conditional mean at x. We apply $\widehat{C}$ as a response operator using Jacobian–vector products and truncated GMRES (Saad & Schultz, 1986); no covariance factorization is required. Statistical linearization of nonlinear observations is described in Appendix E.2.

Freezing the conditional mean $\widehat{m}_\tau(x)$ at x while retaining the variance dependence gives the characteristic

$$\widehat{X}_r = \sqrt{r/\tau}\, x + (1 - \sqrt{r/\tau}) \widehat{m}_\tau(x), \qquad \widehat{v}_\tau^y(x) = \frac{x - \widehat{m}_\tau(x)}{2\tau}, \qquad \widehat{X}_0 = \widehat{m}_\tau(x). \tag{8}$$

The numerical map selects this proposal, a damped proposal, or a gradient step on E using Algorithm 1's energy test; it returns d if none decreases the energy. Thus $E(\Psi_\tau(x)) \le E(d)$ using the same denoiser evaluation. The map error comprises local conditioning error, the characteristic remainder, and the change introduced by the energy test.

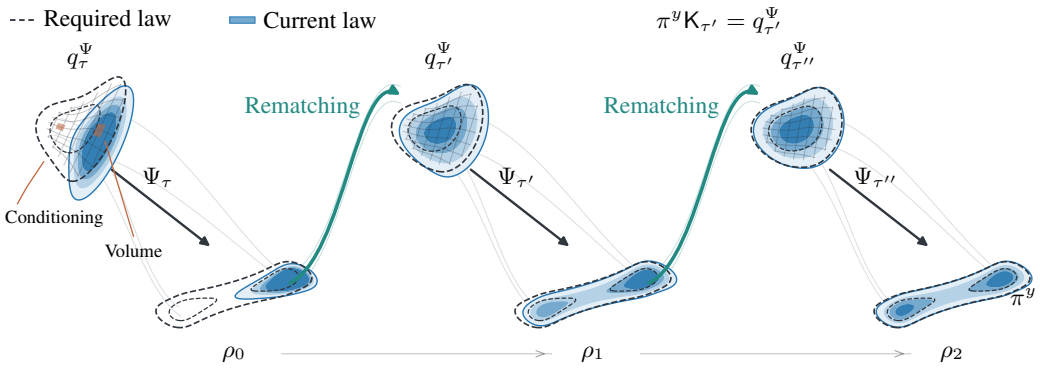


Figure 2: **Rematching supplies a compatible input law.** Dashed contours show the input laws required by the maps and their common posterior target. Rematching accounts jointly for conditioning and volume through these laws. The schematic shows the blue endpoint distributions approaching π^y under compatible, contractive transitions (Appendix D.3).

3.2 Posterior Preservation by Rematching

Fix a family of measurable maps Ψ_s. A compatible input law satisfies $(\Psi_s)_\# q^\Psi_s = \pi^y$ and therefore accounts for both conditioning geometry and volume change (Figure 2). Assume these laws form a weakly continuous curve in s. An invertible map fixes $q^\Psi_s = (\Psi_s^{-1})_\# \pi^y$; a noninvertible map requires a specified compatible curve. Rematching supplies the next map's input law through kernels satisfying

$$\pi^y \mathsf{K}_s = q^\Psi_s, \qquad (\Psi_s)_\# q^\Psi_s = \pi^y \quad \Longrightarrow \quad q^\Psi_\tau \xrightarrow{\Psi_\tau} \pi^y \xrightarrow{\mathsf{K}_{\tau'}} q^\Psi_{\tau'} \xrightarrow{\Psi_{\tau'}} \pi^y. \tag{9}$$

For example, the Gaussian mean map in Equation (2) becomes exact in distribution when supplied with $q^m_\tau = \mathcal{N}(0, (v+\tau)^2/v)$. The same mean map can therefore produce the correct posterior when its input distribution changes accordingly.

To realize these input laws, use an isotropic diffusion and minimize its control energy. For $\phi \in C^\infty_c((0,T) \times \mathbb{R}^n)$, define the entropic action objective (Neklyudov et al., 2023)

$$\begin{aligned} \mathcal{J}_{q^\Psi}(\phi) &= \int_0^T \mathbb{E}_{q^\Psi_s}\left[\tfrac{1}{2}\|\nabla\phi_s\|^2 + \partial_s\phi_s + \tfrac{1}{2}\Delta\phi_s\right] \mathrm{d}s, \\ b^* &= \arg\min_{\nabla\phi} \mathcal{J}_{q^\Psi}(\phi), \qquad \mathrm{d}X_s = b^*_s(X_s)\,\mathrm{d}s + \mathrm{d}W_s. \end{aligned} \tag{10}$$

Here s is diffusion time within a rematching transition. The minimization is over the $L^2(\mathrm{d}s\, q^\Psi_s)$ closure of the potential gradients. Stationarity against every smooth, compactly supported variation η gives

$$\begin{aligned} &\int_0^T \mathbb{E}_{q^\Psi_s}\left[\partial_s\eta_s + b^*_s \cdot \nabla\eta_s + \tfrac{1}{2}\Delta\eta_s\right] \mathrm{d}s = 0 \\ &\qquad\qquad \Longleftrightarrow \qquad \partial_s q^\Psi_s = -\nabla\cdot(q^\Psi_s b^*_s) + \tfrac{1}{2}\Delta q^\Psi_s. \end{aligned} \tag{11}$$

Stationarity enforces the Fokker–Planck equation, while the quadratic term selects the minimum-energy drift.

Theorem 1 (Minimum-energy transport matching). *If $\inf_\phi \mathcal{J}_{q^\Psi}(\phi) > -\infty$, there is a unique minimizing gradient b^* in the completed gradient space. Started from q^Ψ_0, its unit-diffusion SDE admits a weak solution with $X_s \sim q^\Psi_s$ and $\Psi_s(X_s) \sim \pi^y$, using the least mean-square drift among finite-energy controls with these marginals. If the compatible laws are $q^\Psi_s = \mathcal{H}_s\pi^y$, then $b^* = 0$.*

The boundedness condition expresses finite-energy realizability of the prescribed compatible curve; neither invertibility nor a Gaussian target is required. Taking $\Psi_0 = \mathrm{Id}$ gives the kernels in Equation (9). The compatibility condition applies to the complete numerical map, including its finite solve and acceptance test. Appendix C.1 proves the general reference-diffusion form using weak Fokker–Planck duality and diffusion superposition (Trevisan, 2016).

Conditioning changes the required input law while the diffusion coefficient remains isotropic. For smooth invertible maps, Itô–Wentzell composition carries volume correction without a separate determinant evaluation (Appendix C.3) (Karatzas & Shreve, 1991; Bethencourt de Léon et al., 2020).

The matching objective also quantifies departures from compatible evolution. Fix the map and compatible law curve across candidate drifts. Let $\mathbb{P}^*$ be the minimum-energy path law from Theorem 1, and $\mathbb{P}^\phi$ the unit-diffusion path law with drift $\nabla\phi$ and the same initial distribution.

Theorem 2 (Path-space KL representation). *Under the finite-entropy Girsanov conditions of Appendix C.2,*

$$\mathcal{J}_{q^\Psi}(\phi) = \mathrm{KL}(\mathbb{P}^* \| \mathbb{P}^\phi) + \mathrm{const}, \tag{12}$$

where the constant is independent of ϕ. The objectives have identical minimizers and, for differentiable admissible parameterizations, identical gradients.

Data processing through the output map bounds KL from the posterior to the generated law by this path-space gap. The proof is in Appendix C.2 (Léonard, 2012).

3.3 Computing Rematching

For a finite update at variance $t = \tau'$, we constrain the output of the next map and leave intermediate marginals free. Starting from the endpoint law ρ, take the reference transition $K_0(z, \cdot) = \mathcal{N}(z, tI)$. Minimizing $\mathbb{E}_\rho \,\mathrm{KL}(K(z,\cdot)\|K_0(z,\cdot))$ subject to the mapped-output constraint gives

$$K_*(z, \mathrm{d}x) = Z_*(z)^{-1} e^{-U_*(x)} K_0(z, \mathrm{d}x), \qquad (\Psi_t)_\#(\rho K_*) = \pi^y. \tag{13}$$

The matching potential U_* reweights Gaussian inputs so that applying Ψ_t returns the posterior. It depends jointly on the incoming law ρ and the next map Ψ_t. Appendix D.1 derives this entropy projection under finite-entropy feasibility and dual attainment (Léonard, 2014), together with its Doob-transform diffusion representation.

Center optimization. Restrict the computed transition to $G_u = \mathcal{N}(u, tI)$. Its conditional relative entropy is

$$\mathrm{KL}(G_u \| K_*(z,\cdot)) = \underbrace{\mathbb{E}_\xi U_*(u + \sqrt{t}\xi) + \frac{\|u - z\|^2}{2t}}_{\mathcal{L}^*_{t,z}(u)} + \log Z_*(z). \tag{14}$$

Because $Z_*(z)$ is independent of u, minimizing $\mathcal{L}^*_{t,z}$ minimizes conditional transition KL. This changes the transition mean while retaining covariance tI; the squared-distance penalty follows from the Gaussian reference. For computation, we denoise each candidate noisy input and evaluate its observation energy, approximating U_* by $U_t(x) = E(\mathsf{D}_t(x))$ in Equation (3). Equation (16) separates this potential approximation from finite optimization. Each objective sample requires one denoiser forward and a vector–Jacobian product for its gradient.

Analytic evaluation. For affine $A(u) = Hu + b$, put $r(x) = \mathsf{D}_t(x) - x$ and $\widetilde{X} = u + \sqrt{t}\xi$. The denoiser objective decomposes exactly as

$$\begin{aligned} \mathcal{L}_{t,z}(u) &= V_{t,z}(u) + \tfrac{t}{2}\,\mathrm{tr}(H^\top R^{-1} H) + \mathcal{R}_t(u), \\ \mathcal{R}_t(u) &= \mathbb{E}\Big[\nabla E(\widetilde{X})^\top r(\widetilde{X}) + \tfrac{1}{2}\|Hr(\widetilde{X})\|^2_{R^{-1}}\Big]. \end{aligned} \tag{15}$$

The trace term is independent of u; omitting $\mathcal{R}_t$ therefore leaves $V_{t,z}(u) = E(u) + \|u - z\|^2/(2t)$, which uses only the observation operator. If $H\mathsf{D}_t(x) = Hx$ for all x, the remainder vanishes and both objectives have identical minimizers. More generally, $\nabla\mathcal{R}_t$ measures the response discarded

by analytic evaluation; nonlinear observations also contribute the Gaussian averaging of E. We minimize sampled $\mathcal{L}_{t,z}$ or analytic $V_{t,z}$ with a fixed iteration budget (Liu & Nocedal, 1989). Draws remain fixed within a solve; the next state uses independent noise. Zero optimization keeps $u = z$ and gives Gaussian rematching.

Accuracy of the complete update. Let $\widehat{g}_z$ be the numerical gradient at $\widehat{u}(z)$, with error at most e_z relative to $\nabla\mathcal{L}_{t,z}$, and write $\beta_z = \|\mathbb{E}_{G_{\widehat{u}(z)}}\nabla(U_t - U_*)\|$. If $\|\nabla^2 U_*\|_{\mathrm{op}} \le L_*$ and $tL_* < 1$, then

$$\mathrm{KL}((\Psi_t)_\#(\rho G_{\widehat{u}})\|\pi^y) \le \frac{nt^2L_*^2 + t\mathbb{E}_\rho(\|\widehat{g}_z\| + e_z + \beta_z)^2}{2(1 - tL_*)}. \tag{16}$$

The first term accounts for the fixed Gaussian covariance; the second combines the optimization residual, objective evaluation, and potential approximation. Analytic evaluation includes $\nabla\mathcal{R}_t(\widehat{u})$ in e_z; for sampled solves, $\mathbb{E}_\rho$ also averages over the optimization draws. Appendix D.1 proves the bound by conditional log-Sobolev and data-processing inequalities. For an already compatible Gaussian transition, $U_* = 0$ and $u = z$ recover exact preservation. Appendix D.3 relates repeated transitions to contraction and invariance error.

4 Experiments

4.1 From the Matching Objective to Posterior Distributions

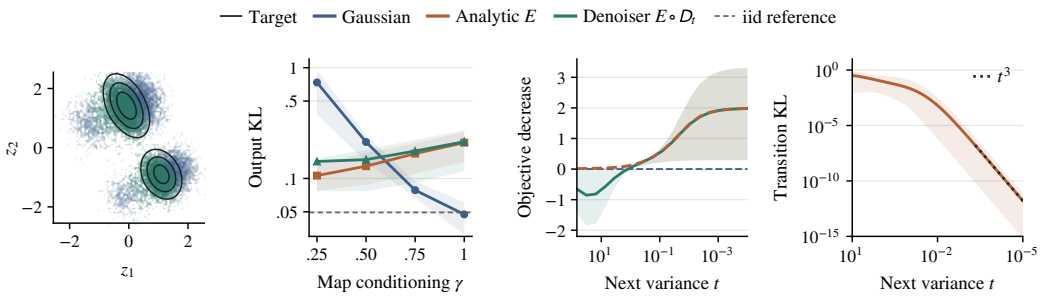
We compare Gaussian and non-Gaussian priors using the same known probability-flow map construction and 10- and 20-level schedules from Gaussian noise. The target uses full conditioning; maps use likelihood weight γ.

Matching distributions. At $\gamma = .25$, both corrections improve all eight pairs in each prior family (Table 1). Figure 3a shows Denoiser rematching concentrating the output around the target modes. The full non-Gaussian sweep in Figure 3b shows when correction is needed: at $\gamma = 1$, Gaussian rematching reaches the iid reference level (.049), and further correction increases KL.

Table 1: **Output KL at $\gamma = .25$.** Eight problem–schedule pairs and two seeds per prior family.

Prior	Gaussian	Analytic	Denoiser
Gaussian	0.542	0.082	0.082
Mixture	0.737	0.106	0.142

Computable objectives. Figure 3c fixes the incoming law and next map in eight Gaussian problems: $E(\mathsf{D}_t)$ reduces ideal transition KL at every tested $t \le 1$ for $\gamma = .25$, but can increase it at larger variance. Figure 3d compares analytic and denoiser minimizers, recovering the $O(t^3)$ transition-KL rate at small variance. Appendix F gives the controls and reference kernels.



(a) Output distributions (b) Map conditioning (c) Ideal objective (d) Analytic evaluation

Figure 3: **Rematching objectives and output distributions.** (a) Paired non-Gaussian outputs at $\gamma = .25$, after 20 levels; Gaussian and Denoiser KL are 1.525 and .287. (b) Non-Gaussian output KL versus map conditioning; mean and interquartile range over eight problem–schedule pairs, averaging two seeds. (c) Ideal-objective decrease at a common incoming law; mean and interquartile range over eight Gaussian problems. (d) Analytic-versus-denoiser transition KL; shading spans all eight problems.

Table 2: **Reconstruction quality.** Means over 100 images per task; bold/underlined: best/second. FFHQ DAPS/SITCOM use direct reruns; other baseline entries are published. Phase: best-of-four baselines; rotation-aligned ImageNet PTM. –: unreported. Full costs: Table 4.

	FFHQ								ImageNet							
Method	SR	GB	MB	RI	BI	NL	PR	HDR	SR	GB	MB	RI	BI	NL	PR	HDR
PSNR↑																
DPS	25.86	25.87	24.52	25.46	22.51	23.39	17.64	22.73	21.13	20.31	18.96	23.52	18.94	22.49	16.81	19.23
DDRM	26.58	24.93	–	–	22.26	–	–	–	22.62	21.26	–	–	18.63	–	–	–
DDNM	28.03	28.20	–	29.91	24.47	–	–	–	23.96	**28.06**	–	**31.16**	21.64	–	–	–
DiffPIR	26.64	27.36	26.57	–	–	–	–	–	23.18	22.80	24.01	–	–	–	–	–
FPS-SMC	28.42	26.54	27.39	28.21	24.86	–	–	–	24.82	23.91	24.52	24.52	22.16	–	–	–
DAPS	29.38	29.62	31.80	30.77	23.84	28.77	30.59	27.66	25.89	26.15	27.86	28.44	21.43	27.73	**25.78**	26.30
SITCOM	30.56	29.90	28.93	32.38	24.47	29.23	29.03	27.98	26.35	27.40	28.65	29.60	21.88	28.78	25.45	26.97
PTM	**30.85**	**31.00**	**33.78**	**34.83**	**26.23**	**31.64**	**33.04**	**31.31**	**26.87**	27.11	**29.71**	30.76	**22.24**	**30.36**	25.57	**29.63**
SSIM↑																
DPS	.753	.764	.801	.823	.792	.623	.441	.591	.489	.598	.629	.745	.722	.591	.427	.582
DDRM	.782	.732	–	–	.801	–	–	–	.521	.564	–	–	.733	–	–	–
DDNM	.795	.804	–	.817	.837	–	–	–	.604	.703	–	.841	.748	–	–	–
DiffPIR	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
FPS-SMC	.813	.773	.826	.823	.823	–	–	–	.703	.601	.647	.701	.726	–	–	–
DAPS	.782	.786	.841	.801	.750	.780	.808	.836	.694	.684	.766	.775	.725	.724	.743	.717
SITCOM	.863	.845	.832	.854	.730	.790	.822	.811	**.812**	**.854**	**.876**	**.915**	.742	.832	**.808**	.821
PTM	**.870**	**.869**	**.918**	**.941**	**.876**	**.886**	**.910**	**.908**	.737	.739	.828	.878	**.806**	**.846**	.642	**.896**
LPIPS↓																
DPS	.269	.219	.246	.203	.209	.278	.410	.264	.361	.397	.423	.297	.257	.306	.447	.503
DDRM	.282	.239	–	–	.207	–	–	–	.324	.443	–	–	.254	–	–	–
DDNM	.197	.216	–	.121	.235	–	–	–	.475	.278	–	.191	.319	–	–	–
DiffPIR	.260	.236	.255	–	–	–	–	–	.371	.355	.366	–	–	–	–	–
FPS-SMC	.204	.253	.227	.261	.146	–	–	–	.313	.387	.326	.316	.208	–	–	–
DAPS	.193	.179	.131	.159	.183	.177	.160	.164	.276	.253	.196	.135	.214	.169	.254	.175
SITCOM	**.153**	.172	.198	.115	.188	.185	.194	.174	**.232**	.236	.189	.127	.214	.160	**.246**	.167
PTM	.157	**.150**	**.088**	**.067**	**.124**	**.136**	**.107**	**.114**	.243	**.228**	**.163**	**.102**	**.181**	**.157**	.282	**.122**
FID↓																
DPS	81.1	79.8	65.2	69.2	61.3	91.3	104.5	112.8	106.3	116.4	137.8	87.5	126.5	101.4	197.5	146.2
DDRM	79.2	92.4	–	–	78.6	–	–	–	103.8	146.9	–	–	116.4	–	–	–
DDNM	64.6	57.8	–	44.4	46.6	–	–	–	98.6	81.4	–	63.8	**104.0**	–	–	–
DiffPIR	65.8	59.6	65.8	–	–	–	–	–	106.3	93.4	94.6	–	–	–	–	–
FPS-SMC	**49.2**	67.5	48.3	61.2	48.3	–	–	–	97.5	91.7	87.4	79.1	111.6	–	–	–
DAPS	56.6	**53.5**	37.5	35.7	38.4	51.7	**36.5**	41.2	**83.6**	**75.7**	61.8	54.2	109.8	59.9	**82.7**	64.2
SITCOM	60.4	70.4	79.0	27.6	56.3	**50.0**	84.2	55.9	–	–	–	–	–	–	–	–
PTM	61.3	62.7	**31.6**	**22.9**	**31.9**	51.2	39.8	**30.8**	95.2	98.9	**44.7**	**23.6**	104.2	**55.2**	142.6	**32.4**

4.2 Image Reconstruction and Efficiency

Each FFHQ and ImageNet task uses 100 images with observation-noise standard deviation .05. Settings are fixed per task; ImageNet budget adjustments use separate development images. Tasks comprise $4\times$ super-resolution (SR), Gaussian blur (GB), motion blur (MB), random inpainting (RI), box inpainting (BI), nonlinear blur (NL), phase retrieval (PR), and HDR. We report effective denoiser and observation evaluation counts, $D_{\text{eff}} = D_{\text{fwd}} + 2D_{\text{jvp}} + 2D_{\text{vjp}}$ and $A_{\text{eff}} = A_{\text{fwd}} + 2A_{\text{jvp}} + 2A_{\text{vjp}}$, including all particles and inner solves.

PTM leads all eight FFHQ tasks in PSNR and five on ImageNet (Table 2). Table 3 compares the FFHQ reruns task by task: PTM uses fewer denoiser and observation evaluations than DAPS in all eight tasks. Full costs and evaluation protocols appear in Table 4 and Appendix F.

Figure 4 compares complete 100-image cohorts as the PTM schedule varies from six to twenty levels. Six levels use $D_{\text{eff}} = 76$ and reach 33.30 dB for motion blur and 34.16 dB for random inpainting; further levels improve PSNR, while LPIPS is lowest at ten levels.

Table 3: **FFHQ costs by task.** Mean effective evaluations over 100 images. All methods and forward counts: Table 4.

	Denoiser: D_{eff}								Observation: A_{eff}							
Method	SR	GB	MB	RI	BI	NL	PR	HDR	SR	GB	MB	RI	BI	NL	PR	HDR
DAPS	1k	1k	1k	1k	1k	4k	16k	4k	60k	60k	60k	60k	60k	240k	480k	120k
SITCOM	1352	1439	1434	1517	1594	1660	5590	1716	1349	1437	1431	1574	1651	1657	5580	1712
PTM	176	196	200	200	832	875	2956	999	753	1173	1351	737	2174	11554	90995	11948

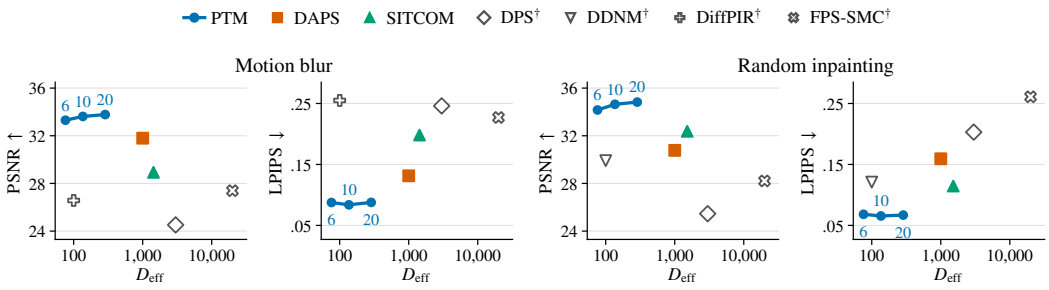


Figure 4: **Quality–cost tradeoffs on 100 FFHQ images.** Curves connect 6, 10, and 20 PTM levels. DAPS/SITCOM use direct reruns; † denotes published results.

Figure 5 shows clean estimates from two selected reconstructions at fractions of each method's native budget.

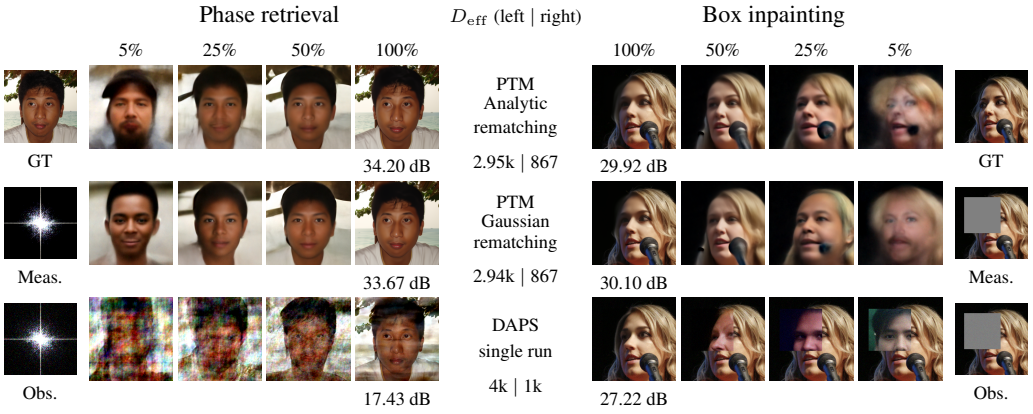


Figure 5: **Clean estimates throughout sampling.** Fractions use each method's D_{eff}. Phase follows a surviving contributor before merging; Appendix F.3 specifies the schedules and selection.

5 Conclusion

PTM formulates posterior sampling through the compatibility of numerical maps and stochastic transitions. Minimum-energy diffusion characterizes compatible evolution, and entropy projection gives a finite objective that can be approximated using a frozen denoiser. On FFHQ, the sampler improves PSNR over the reported baselines and reduces both evaluation counts relative to DAPS, with further evaluation on ImageNet. The finite update retains potential-approximation error, explicitly quantified in Equation (16) under its curvature conditions. Image benchmarks assess reconstruction with particle selection and averaging; posterior calibration for learned image priors remains an important direction for distributional evaluation.

AI USE STATEMENT

In this work, we used generative AI tools to assist with implementing methods, developing mathematical arguments, drafting proofs, translation, and providing feedback on experimental design. We also used these tools to suggest experimental parameters, search for and summarize relevant literature, edit the manuscript for clarity and readability, and format references. We have reviewed all AI-assisted work. Two authors checked the AI-assisted proofs for correctness. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the assistance of generative AI.

ETHICS STATEMENT

This work studies image reconstruction using existing benchmark datasets and pretrained generative models. Reconstructions can reflect biases in the pretrained priors and introduce plausible details that are not supported by the observations. They should therefore not be treated as verified evidence of the underlying scene or a person's identity. Applications involving identifiable individuals require particular attention to privacy and appropriate use of the source data.

REPRODUCIBILITY STATEMENT

The appendices provide the assumptions and proofs supporting our theoretical results. Appendix E describes the numerical implementation, and Appendix F documents the experimental settings and evaluation protocol. We plan to release the implementation, experiment configurations, and evaluation scripts publicly.

REFERENCES

Ismail Alkhouri, Shijun Liang, Cheng-Han Huang, Jimmy Dai, Qing Qu, Saiprasad Ravishankar, and Rongrong Wang. SITCOM: Step-wise triple-consistent diffusion sampling for inverse problems. In *International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 1128–1158. PMLR, 2025. URL https://proceedings.mlr.press/v267/alkhouri25b.html.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer, 2014. doi: 10.1007/978-3-319-00227-9.

Aythami Bethencourt de Léon, Darryl D. Holm, Erwin Luesink, and So Takao. Implications of Kunita–Itô–Wentzell formula for k-forms in stochastic fluid dynamics. *Journal of Nonlinear Science*, 30(4):1421–1454, 2020. doi: 10.1007/s00332-020-09613-0. URL https://link.springer.com/article/10.1007/s00332-020-09613-0.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, first edition, 2006. ISBN 9780387310732. URL https://link.springer.com/book/9780387310732.

Nicholas M. Boffi, Michael S. Albergo, and Eric Vanden-Eijnden. Flow map matching with stochastic interpolants: A mathematical framework for consistency models. *Transactions on Machine Learning Research*, 2025. URL https://openreview.net/forum?id=cqDH0e6ak2.

Vladimir I. Bogachev, Nicolai V. Krylov, Michael Röckner, and Stanislav V. Shaposhnikov. *Fokker–Planck–Kolmogorov Equations*, volume 207 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2015. doi: 10.1090/surv/207. URL https://www.ams.org/books/surv/207/.

Benjamin Boys, Mark Girolami, Jakiw Pidstrigach, Sebastian Reich, Alan Mosca, and O. Deniz Akyildiz. Tweedie moment projected diffusions for inverse problems. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=4unJi0qrTE.

Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2011. doi: 10.1007/978-0-387-70914-7.

Gabriel Cardoso, Yazid Janati, Sylvain Le Corff, and Eric Moulines. Monte Carlo guided denoising diffusion models for Bayesian linear inverse problems. In *International Conference on Learning Representations*, pp. 44001–44037, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/bfb3c1db0deda449ff4027f77dd48704-Abstract-Conference.html.

Nicolas Chopin, Andras Fulop, Jeremy Heng, and Alexandre H. Thiery. Computational Doob h-transforms for online filtering of discretely observed diffusions. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 5904–5923. PMLR, 2023. URL https://proceedings.mlr.press/v202/chopin23a.html.

Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. In *Advances in Neural Information Processing Systems*, volume 35, pp. 25683–25696, 2022. doi: 10.52202/068431-1862. URL https://papers.neurips.cc/paper_files/paper/2022/hash/a48e5877c7bf86a513950ab23b360498-Abstract-Conference.html.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *International Conference on Learning Representations*, pp. 38922–38949, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/a8eb51c394d59ffc3324afa672fd24dc-Abstract-Conference.html.

Florentin Coeurdoux, Nicolas Dobigeon, and Pierre Chainais. Plug-and-play split Gibbs sampler: Embedding deep generative priors in Bayesian inference. *IEEE Transactions on Image Processing*, 33:3496–3507, 2024. doi: 10.1109/TIP.2024.3404338. URL https://doi.org/10.1109/TIP.2024.3404338.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *International Conference on Learning Representations*, pp. 12069–12095, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/333bbde5ed6397cd9bc5bac552b3b7a6-Abstract-Conference.html.

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

Julius Erbach, Dominik Narnhofer, Andreas Dombos, Bernt Schiele, Jan Eric Lenssen, and Konrad Schindler. Solving inverse problems with FLAIR. In *Advances in Neural Information Processing Systems*, volume 38, pp. 136667–136704, 2025. doi: 10.52202/085713-4560. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/c7ae6e9659f0c99582c2e8214ba0b413-Abstract-Conference.html.

Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. In *Advances in Neural Information Processing Systems*, volume 38, pp. 75460–75482, 2025. doi: 10.52202/085713-2534. URL https://papers.neurips.cc/paper_files/paper/2025/hash/6d13e085b79d454da5910e4ca82a3d9d-Abstract-Conference.html.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/8a1d694707eb0fefe65871369074926d-Abstract.html.

Peter Holderrieth, Uriel Singer, Tommi Jaakkola, Ricky T. Q. Chen, Yaron Lipman, and Brian Karrer. GLASS flows: Efficient inference for reward alignment of flow and diffusion models. In *International Conference on Learning Representations*, pp. 106232–106266, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ad6ac52e02e14e0f2ddc0d4587ab47f1-Abstract-Conference.html.

Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, second edition, 1991. doi: 10.1007/978-1-4612-0949-2.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4401–4410, 2019. URL https://openaccess.thecvf.com/content_CVPR_2019/html/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.html.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23593–23606, 2022. doi: 10.52202/068431-1714. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/95504595b6169131b6ed6cd72eb05616-Abstract-Conference.html.

Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. FlowDPS: Flow-driven posterior sampling for inverse problems. In *IEEE/CVF International Conference on Computer Vision*, pp. 12328–12337, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Kim_FlowDPS__Flow-Driven_Posterior_Sampling_for_Inverse_Problems_ICCV_2025_paper.html.

Bo'az Klartag and Or Ordentlich. The strong data processing inequality under the heat flow. *IEEE Transactions on Information Theory*, 71(5):3317–3333, 2025. doi: 10.1109/TIT.2025.3548961. URL https://doi.org/10.1109/TIT.2025.3548961.

Hiroshi Kunita. *Stochastic Flows and Stochastic Differential Equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 1990. ISBN 9780521350501.

Christian Léonard. Stochastic derivatives and generalized h-transforms of Markov processes. *arXiv preprint arXiv:1102.3172*, 2011. URL https://arxiv.org/abs/1102.3172.

Christian Léonard. Girsanov theory under a finite entropy condition. In *Séminaire de Probabilités XLIV*, volume 2046 of *Lecture Notes in Mathematics*, pp. 429–465. Springer, 2012. doi: 10.1007/978-3-642-27461-9_20. URL https://link.springer.com/chapter/10.1007/978-3-642-27461-9_20.

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. doi: 10.3934/dcds.2014.34.1533.

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

Dong C. Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45:503–528, 1989. doi: 10.1007/BF01589116.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting using denoising diffusion probabilistic models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11461–11471, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Lugmayr_RePaint_Inpainting_Using_Denoising_Diffusion_Probabilistic_Models_CVPR_2022_paper.html.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *International Conference on Learning Representations*, pp. 28027–28053, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/7711026951f3dc3afb2be427bf98bc73-Abstract-Conference.html.

Kirill Neklyudov, Rob Brekelmans, Daniel Severo, and Alireza Makhzani. Action matching: Learning stochastic dynamics from samples. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25858–25889. PMLR, 2023. URL https://proceedings.mlr.press/v202/neklyudov23a.html.

Xinyu Peng, Ziyang Zheng, Wenrui Dai, Nuoqian Xiao, Chenglin Li, Junni Zou, and Hongkai Xiong. Improving diffusion models for inverse problems using optimal posterior covariance. In *International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 40347–40370. PMLR, 2024. URL https://proceedings.mlr.press/v235/peng24h.html.

Peter Potaptchik, Cheuk-Kit Lee, and Michael S. Albergo. Tilt matching for scalable sampling and fine-tuning. In *International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=dQA4Gjt4KU.

François Rozet, Gérôme Andry, François Lanusse, and Gilles Louppe. Learning diffusion priors from observations by expectation maximization. In *Advances in Neural Information Processing Systems*, volume 37, pp. 87647–87682, 2024. doi: 10.52202/079017-2783. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/9f94298bac4668db4dc77ddb0a244301-Abstract-Conference.html.

Youcef Saad and Martin H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986. doi: 10.1137/0907058.

Simo Särkkä and Lennart Svensson. *Bayesian Filtering and Smoothing*. Cambridge University Press, second edition, 2023. doi: 10.1017/9781108917407. URL https://www.cambridge.org/core/books/bayesian-filtering-and-smoothing/F88740E8D25010CF3119A5CA379FA37A.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023a. URL https://openreview.net/forum?id=9_gsMA8MRKQ.

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.

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

Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010. doi: 10.1017/S0962492910000061.

Dario Trevisan. Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electronic Journal of Probability*, 21(22):1–41, 2016. doi: 10.1214/16-EJP4453. URL https://doi.org/10.1214/16-EJP4453.

Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *IEEE Global Conference on Signal and Information Processing*, pp. 945–948, 2013. doi: 10.1109/GlobalSIP.2013.6737048. URL https://doi.org/10.1109/GlobalSIP.2013.6737048.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=mRieQgMtNTQ.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

Luhuan Wu, Brian L. Trippe, Christian A. Naesseth, David M. Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 31372–31403, 2023. doi: 10.52202/075280-1363. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/63e8bc7bbf1cfea36d1d1b6538aecce5-Abstract-Conference.html.

Xingyu Xu and Yuejie Chi. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. In *Advances in Neural Information Processing Systems*, volume 37, pp. 36148–36184, 2024. doi: 10.52202/079017-1140. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/3fa2d2b637122007845a2fbb7c21453b-Abstract-Conference.html.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20895–20905, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Zhang_Improving_Diffusion_Inverse_Problem_Solving_with_Decoupled_Noise_Annealing_CVPR_2025_paper.html.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1219–1229, 2023. URL https://openaccess.thecvf.com/content/CVPR2023W/NTIRE/html/Zhu_Denoising_Diffusion_Models_for_Plug-and-Play_Image_Restoration_CVPRW_2023_paper.html.

Rayhan Zirvi, Bahareh Tolooshams, and Anima Anandkumar. Diffusion state-guided projected gradient for inverse problems. In *International Conference on Learning Representations*, pp. 31217–31242, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/4d5b7f7a0603b75e39ffebbf95e3b7ee-Abstract-Conference.html.

A Conditional Moments and Posterior Geometry

Fix conditioning information y, a prior density p_0, and a positive conditioning factor L_y with $Z_y = \int p_0(z)L_y(z)\,\mathrm{d}z \in (0,\infty)$. The energy notation in the main text uses $L_y \propto e^{-E}$; the proportionality constant cancels from normalized targets and conditional means. For the observation model, L_y is the measurement likelihood. Write $\pi^y = p_0 L_y/Z_y$, $p_\tau = \mathcal{H}_\tau p_0$, and $q_\tau^y = \mathcal{H}_\tau \pi^y$, where $\mathcal{H}_\tau$ denotes convolution with $\gamma_\tau = \mathcal{N}(0,\tau I)$. All differentiations below assume integrable derivatives and finite conditional moments; $\tau > 0$ is a variance throughout.

A.1 Conditional mean and covariance identities

For $X_\tau = Z + \sqrt{\tau}\xi$ with independent $\xi \sim \mathcal{N}(0,I)$, the prior conditional density is

$$\rho_\tau^0(z \mid x) = \frac{p_0(z)\gamma_\tau(x-z)}{p_\tau(x)}, \qquad d_\tau(x) = \int z\rho_\tau^0(z \mid x)\,\mathrm{d}z. \tag{A.1}$$

The following Tweedie identities (Efron, 2011; Boys et al., 2024) follow by differentiating the Gaussian kernel and its normalization:

$$\begin{aligned}
\nabla_x p_\tau(x) &= \int \frac{z-x}{\tau} p_0(z)\gamma_\tau(x-z)\,\mathrm{d}z,\\
\nabla_x \log p_\tau(x) &= \frac{d_\tau(x)-x}{\tau},\\
\nabla_x \log \rho_\tau^0(z \mid x) &= \frac{z-x}{\tau} - \frac{d_\tau(x)-x}{\tau} = \frac{z-d_\tau(x)}{\tau},\\
J_x d_\tau(x) &= \frac{1}{\tau}\int z(z-d_\tau(x))^\top \rho_\tau^0(z \mid x)\,\mathrm{d}z\\
&= \frac{1}{\tau}\int (z-d_\tau(x))(z-d_\tau(x))^\top \rho_\tau^0(z \mid x)\,\mathrm{d}z = \frac{1}{\tau}\operatorname{Cov}(Z \mid X_\tau = x).
\end{aligned} \tag{A.2}$$

The second equality in the covariance calculation uses $\int (z-d_\tau)\rho_\tau^0\,\mathrm{d}z = 0$. Replacing p_0 by π^y gives the corresponding posterior identities. Bayes' rule then yields

$$\begin{aligned}
h_\tau^y(x) &= \int L_y(z)\rho_\tau^0(z \mid x)\,\mathrm{d}z, \qquad q_\tau^y(x) = Z_y^{-1} p_\tau(x) h_\tau^y(x),\\
m_\tau^y(x) &= \frac{\int zL_y(z)\rho_\tau^0(z \mid x)\,\mathrm{d}z}{h_\tau^y(x)},\\
\nabla_x \log h_\tau^y(x) &= \frac{m_\tau^y(x)-d_\tau(x)}{\tau},\\
\partial_\tau q_\tau^y &= \tfrac{1}{2}\Delta q_\tau^y = -\nabla\cdot(q_\tau^y v_\tau^y), \qquad v_\tau^y(x) = \frac{x-m_\tau^y(x)}{2\tau}.
\end{aligned} \tag{A.3}$$

Thus a prior denoiser supplies the unweighted conditional mean, whereas posterior conditioning requires likelihood-weighted conditional expectations.

A.2 The numerical conditional-mean proposal

For an exact prior denoiser, let $d = \mathsf{D}_\tau(x)$, $C = \tau J_x \mathsf{D}_\tau(x)$, $H = J_A(d)$, and $b = y - A(d)$. The Gaussian moment approximation and affine observation model replace the conditional law by $\mathcal{N}(d,C)$ and the observation by $A(d) + H(Z-d)$. For $C \succ 0$ and $R \succ 0$, write $v = z-d$ and $B = R + HCH^\top$. Expanding and completing the quadratic gives the standard Gaussian conditioning identities (Bishop, 2006, Section 2.3):

$$\begin{aligned}
\Phi(v) &= \tfrac{1}{2}v^\top C^{-1}v + \tfrac{1}{2}(b-Hv)^\top R^{-1}(b-Hv)\\
&= \tfrac{1}{2}v^\top(C^{-1}+H^\top R^{-1}H)v - v^\top H^\top R^{-1}b + \tfrac{1}{2}b^\top R^{-1}b,\\
C^y &= (C^{-1}+H^\top R^{-1}H)^{-1} = C - CH^\top B^{-1}HC,\\
\widehat{m}_\tau - d &= C^y H^\top R^{-1}b = CH^\top B^{-1}b.
\end{aligned} \tag{A.4}$$

The formulas extend to $C \succeq 0$ by replacing C with $C + \varepsilon I$ and taking $\varepsilon \downarrow 0$. These conditional-moment identities (Boys et al., 2024; Rozet et al., 2024) supply the source posterior velocity. Holding the source mean fixed gives the characteristic proposal $\widehat{X}_0$ in Equation (8); the energy acceptance test then selects the numerical output Ψ_τ (Appendix E). For a learned denoiser, the numerical response is $\widehat{C} = \tau J_x \mathsf{D}_\tau$, as distinguished in the main text. It need not be symmetric or positive semidefinite, so a Gaussian interpretation additionally requires those properties. The numerical linear response remains well defined when $\widehat{B} = R + H\widehat{C}H^\top$ is invertible. For a truncated iterate w_k, the residual $r_k = b - \widehat{B}w_k$ gives

$$\begin{aligned} w_* - w_k &= \widehat{B}^{-1} r_k, \\ \widehat{m}_* - \widehat{m}_k &= \widehat{C}H^\top \widehat{B}^{-1} r_k, \\ \|\widehat{m}_* - \widehat{m}_k\| &\le \|\widehat{C}H^\top \widehat{B}^{-1}\| \, \|r_k\|. \end{aligned} \tag{A.5}$$

Jacobian products evaluate $\widehat{B}w = Rw + H(\widehat{C}(H^\top w))$ without forming the response matrix. The residual therefore controls the linear-solve contribution to the conditioned-velocity error. Section 3.1 distinguishes the moment and observation approximations; Appendix B.3 derives the characteristic remainder.

B Transport Equations and Conditioning

Throughout this appendix, y is fixed and all densities are with respect to Lebesgue measure on $\mathbb{R}^d$. We work on $[0, T]$ with a strictly positive, sufficiently smooth posterior $\pi = \pi^y$ of finite second moment. We assume $q_\tau = \mathcal{H}_\tau \pi$ and its score generate a unique nonexplosive C^2 diffeomorphic probability flow; the maps and densities have all additional derivatives used below. For the density arguments we assume integrable derivatives and decay sufficient for integration by parts and uniqueness in the class of probability solutions of the heat and Fokker–Planck equations (Bogachev et al., 2015). For example, suitable spatial Lipschitz and growth bounds on the score give sufficient stochastic well-posedness conditions on finite intervals. These assumptions concern the exact mathematical model; they are not assertions about an arbitrary learned denoiser or a singular clean-data distribution.

B.1 Probability-flow dynamics and the induced Doob factor

Bayes' rule gives $q_\tau(x) = p_\tau(x) h^y_\tau(x)/Z_y$. If $T_{r|\tau}$ is the reverse transition of the prior heat process, its conditional version is

$$T^y_{r|\tau}(x, \mathrm{d}z) = \frac{h^y_r(z)}{h^y_\tau(x)} T_{r|\tau}(x, \mathrm{d}z), \qquad r < \tau. \tag{B.1}$$

The tower property gives normalization. The heat equation can be written as

$$\partial_\tau q_\tau = -\nabla_x \cdot (q_\tau v^y_\tau), \qquad v^y_\tau = -\tfrac{1}{2} \nabla_x \log q_\tau. \tag{B.2}$$

Let Φ_τ be its forward flow and $F_\tau = \Phi_\tau^{-1}$. Then $F_\tau(\Phi_\tau(z)) = z$, giving map consistency and $(F_\tau)_\# q_\tau = \pi$. Differentiating the density identity in Equation (4) gives $\nabla \log q_\tau = J_\tau^\top [(\nabla \log \pi) \circ F_\tau] + \nabla \log \det J_\tau$, hence

$$\partial_\tau F_\tau = \tfrac{1}{2} J_\tau J_\tau^\top [(\nabla \log \pi) \circ F_\tau] + \tfrac{1}{2} J_\tau \nabla_x \log \det J_\tau. \tag{B.3}$$

The converse, including the induced-density equation and uniqueness, follows from the next lemma. Along the characteristic $\dot{X}_s = -J_s(X_s)^{-1} \partial_s F_s(X_s)$ ending at $X_\tau = x$, the clean endpoint $z = F_s(X_s)$ is constant. Integrating the velocity yields

$$z = F_\tau(x) = x + \left(\frac{1}{2} \int_0^\tau J_s(X_s)^\top \, \mathrm{d}s \right) \nabla \log \pi(z) + \frac{1}{2} \int_0^\tau \nabla_x \log \det J_s(X_s) \, \mathrm{d}s. \tag{B.4}$$

This separates score transport from volume correction without discarding the map dependence along the characteristic.

The following lemma relates the transport equation to Doob conditioning (Léonard, 2011; Chopin et al., 2023) and identifies the induced posterior transport. Write $p_\tau = \mathcal{H}_\tau p_0$ and $Z_y = \int p_0 L_y \in (0, \infty)$.

Lemma B.1 (Doob h-transform induced by the probability-flow map)**.** *Assume p_0 and $\pi^y = Z_y^{-1} p_0 L_y$ are positive smooth densities, with the regularity, orientation-preserving invertibility, and uniqueness assumptions of Appendix B. A solution of Equation* (5) *with $F_0 = \mathrm{Id}$ induces*

$$\begin{aligned} h_\tau^y &= \frac{Z_y(\pi^y \circ F_\tau)\det J_\tau}{p_\tau}, \\ \partial_\tau h_\tau^y &= \tfrac{1}{2}\Delta h_\tau^y + \nabla \log p_\tau \cdot \nabla h_\tau^y, \qquad h_0^y = L_y. \end{aligned} \tag{B.5}$$

Consequently, $h_\tau^y = \mathcal{H}_\tau(p_0 L_y)/p_\tau$ and $(F_\tau)_\# \mathcal{H}_\tau \pi^y = \pi^y$.

Proof. We derive the induced-density equation for a general smooth orientation-preserving map with $F_0 = \mathrm{Id}$, retaining its PDE residual before specializing to an exact solution. Define the noisy density and unnormalized Doob factor induced by the map as

$$q_\tau^F(x) = \pi^y(F_\tau(x))\det J_\tau(x), \qquad h_\tau^F(x) = \frac{Z_y q_\tau^F(x)}{p_\tau(x)} = \frac{p_0(F_\tau(x))L_y(F_\tau(x))\det J_\tau(x)}{p_\tau(x)}. \tag{B.6}$$

The density identity holds by definition; it does not yet assert that q_τ^F is the posterior heat family. To establish that assertion, write e_τ for the residual of the closed map PDE:

$$e_\tau = \partial_\tau F_\tau - \tfrac{1}{2} J_\tau \nabla \log q_\tau^F, \qquad \nu_\tau = -J_\tau^{-1}\partial_\tau F_\tau = -\tfrac{1}{2}\nabla \log q_\tau^F - J_\tau^{-1} e_\tau. \tag{B.7}$$

Differentiating q_τ^F gives

$$\begin{aligned} \partial_\tau q_\tau^F &= \det J_\tau \, \nabla \pi^y(F_\tau) \cdot \partial_\tau F_\tau + q_\tau^F \,\mathrm{tr}(J_\tau^{-1}\partial_\tau J_\tau) \\ &= \nabla \cdot \left(q_\tau^F J_\tau^{-1}\partial_\tau F_\tau\right). \end{aligned} \tag{B.8}$$

To verify the divergence identity, write $j = \det J$ and $\mathrm{cof}\, J = jJ^{-\top}$. The Piola identity $\sum_k \partial_{x_k}(\mathrm{cof}\, J)_{ik} = 0$ and $\sum_k (\mathrm{cof}\, J)_{ik} J_{\ell k} = j\delta_{i\ell}$ give

$$\begin{aligned} \nabla_x \cdot (q^F J^{-1}\partial_\tau F) &= \sum_{i,k} \partial_{x_k}\left[\pi(F)(\mathrm{cof}\, J)_{ik}\partial_\tau F_i\right] \\ &= j\sum_i (\partial_{z_i}\pi)(F)\partial_\tau F_i + \pi(F)\sum_{i,k}(\mathrm{cof}\, J)_{ik}\partial_\tau J_{ik} \\ &= j\nabla\pi(F)\cdot\partial_\tau F + q^F \,\mathrm{tr}(J^{-1}\partial_\tau J). \end{aligned} \tag{B.9}$$

Thus q_τ^F obeys the continuity equation for $\nu_\tau = -J_\tau^{-1}\partial_\tau F_\tau$. Consequently,

$$\partial_\tau q_\tau^F = -\nabla\cdot(q_\tau^F \nu_\tau) = \tfrac{1}{2}\Delta q_\tau^F + \nabla\cdot(q_\tau^F J_\tau^{-1} e_\tau). \tag{B.10}$$

Using $p_\tau h_\tau^F = Z_y q_\tau^F$ and $\partial_\tau p_\tau = \frac{1}{2}\Delta p_\tau$ gives the explicit relation between map and Doob defects:

$$\partial_\tau h_\tau^F - \tfrac{1}{2}\Delta h_\tau^F - \nabla \log p_\tau \cdot \nabla h_\tau^F = \frac{1}{p_\tau}\nabla\cdot\left(p_\tau h_\tau^F J_\tau^{-1} e_\tau\right), \qquad h_0^F = L_y. \tag{B.11}$$

In particular, $e_\tau = 0$ implies both $q_\tau^F = \mathcal{H}_\tau \pi^y$ and $h_\tau^F = \mathcal{H}_\tau(p_0 L_y)/p_\tau = h_\tau^y$ by heat uniqueness. For an exact map, the induced-density calculation reduces to

$$\partial_\tau q_\tau^y = \nabla\cdot\left(q_\tau^y J_\tau^{-1}\partial_\tau F_\tau\right) = \tfrac{1}{2}\Delta q_\tau^y, \qquad q_0^y = \pi^y. \tag{B.12}$$

Expanding $\partial_\tau(p_\tau h_\tau^y) = \frac{1}{2}\Delta(p_\tau h_\tau^y)$ and using the prior diffusion equation proves Equation (B.5) and completes the proof of Lemma B.1. □

Thus a solution of Equation (5), with its stated boundary and target, already incorporates the exact Doob evolution equation. The converse requires the probability-flow choice: a vanishing scalar Doob defect alone allows divergence-free modifications of $q_\tau^F J_\tau^{-1} e_\tau$, and does not determine the endpoint coupling. Moreover, small e_τ alone need not make its spatial divergence small without additional derivative control. This identity is an exact structural relation, not an error guarantee for an arbitrary learned or discretized map.

B.2 Conditioning geometry for general measurements

For the observation model in Equation (1), Bayes' rule gives $\rho_\tau^y(z \mid x) \propto p_0(z)\exp[-V_\tau^y(z;x)]$, where

$$V_\tau^y(z;x) = \frac{\|z-x\|^2}{2\tau} + \frac{1}{2}\|A(z)-y\|^2_{R^{-1}}. \tag{B.13}$$

When $A : \mathbb{R}^d \to \mathbb{R}^k$ is twice differentiable, its exact curvature is

$$\nabla_z^2 V_\tau^y = \tau^{-1} I + J_A(z)^\top R^{-1} J_A(z) + \sum_{j=1}^{k} [R^{-1}(A(z)-y)]_j \nabla_z^2 A_j(z), \tag{B.14}$$

with $J_A = \nabla_z A$. The last term retains the nonlinear measurement curvature; dropping it is generally an approximation. The likelihood-modified corruption potential need not be quadratic or have isotropic curvature.

For linear A, completing the square gives the exact specialization

$$\begin{aligned} \Xi_\tau &= (\tau^{-1} I + A^\top R^{-1} A)^{-1}, \qquad b_\tau = \Xi_\tau(\tau^{-1}x + A^\top R^{-1} y), \\ \rho_\tau^y(z \mid x) &\propto p_0(z)\mathcal{N}(z; b_\tau, \Xi_\tau), \qquad m_\tau^y(x) = \mathsf{D}_{\Xi_\tau}(b_\tau). \end{aligned} \tag{B.15}$$

Here $\mathsf{D}_C(b) = \mathbb{E}_{p_0}[Z \mid Z + \xi = b]$ for independent $\xi \sim \mathcal{N}(0, C)$; no Gaussian assumption on p_0 is used. The scalar-variance denoiser is $\mathsf{D}_{\tau I} = \mathsf{D}_\tau$. If λ_i are the eigenvalues of $A^\top R^{-1} A$, the eigenvalues of Ξ_τ are $\tau/(1+\tau\lambda_i)$. Thus Ξ_τ is a scalar multiple of I exactly when all λ_i coincide. This gives the linear illustration in Figure 1.

B.3 The characteristic and its numerical approximation

This calculation specializes the transport equation to the analytic characteristic used for numerical transport. The energy acceptance test and optional rematching correction are retained as separate displacements; they are not assumed to satisfy the characteristic equation. Fix a source (x, τ) and let X_s solve the exact posterior characteristic with $X_\tau = x$. Write $m_s = m_s^y$ and $v_s = v_s^y$ in this subsection. Assume a unique smooth characteristic, the clean limits $X_0 = F_\tau(x)$ and $m_0(z) = z$, and the integrability of the displayed material derivatives. Bounds extending to zero additionally require a finite bound on $\mathcal{D}_s^y m_s(X_s)$ down to zero; local bounds on intervals bounded away from zero do not imply this condition.

Along the exact characteristic, Tweedie's identity gives $m_s(X_s) = X_s - 2s v_s(X_s)$. Differentiating every factor gives

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s} m_s(X_s) &= \partial_s m_s(X_s) + J_x m_s(X_s) v_s(X_s) = \mathcal{D}_s^y m_s(X_s) \\ &= v_s(X_s) - 2v_s(X_s) - 2s\mathcal{D}_s^y v_s(X_s) \\ &= -v_s(X_s) - 2s\mathcal{D}_s^y v_s(X_s). \end{aligned} \tag{B.16}$$

Substitution into Equation (6), or direct integration with $m_0(X_0) = X_0$, yields

$$F_\tau(x) = m_\tau(x) - \int_0^\tau \mathcal{D}_s^y m_s(X_s)\,\mathrm{d}s, \qquad \mathcal{C}_\tau^y(x) = -\int_0^\tau \mathcal{D}_s^y m_s(X_s)\,\mathrm{d}s. \tag{B.17}$$

This identity characterizes the truncation remainder of a map that freezes the source conditional mean. Its derivation uses the exact posterior characteristic and is independent of how the source mean is estimated.

Freeze the computed mean at the source, $\widehat{m} = \widehat{m}_\tau(x)$. The analytic surrogate from Equation (8) is

$$\frac{\mathrm{d}\widehat{X}_s}{\mathrm{d}s} = \frac{\widehat{X}_s - \widehat{m}}{2s}, \qquad \widehat{X}_r = \sqrt{\frac{r}{\tau}}\,x + \left(1 - \sqrt{\frac{r}{\tau}}\right)\widehat{m}, \qquad 0 < r \le \tau. \tag{B.18}$$

In particular, $\widehat{X}_0 = \widehat{m}$ by continuity. Substitution of this surrogate directly into the transport equation makes the cancellation explicit:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}F_s(\widehat{X}_s) &= (\partial_s F_s)(\widehat{X}_s) + J_s(\widehat{X}_s)\frac{\widehat{X}_s - \widehat{m}}{2s} \\ &= J_s(\widehat{X}_s)\left[-\frac{\widehat{X}_s - m_s(\widehat{X}_s)}{2s} + \frac{\widehat{X}_s - \widehat{m}}{2s}\right] \\ &= \frac{J_s(\widehat{X}_s)}{2s}\big(m_s(\widehat{X}_s) - \widehat{m}\big).\end{aligned} \tag{B.19}$$

The common state terms cancel; variation of the conditional mean and its source approximation remain. The Jacobian here is that of the exact probability-flow map, not the prior-denoiser Jacobian.

The exact characteristic has a similar integrating-factor identity, but with its mean evaluated along the path:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{X_s}{\sqrt{s}}\right) &= \frac{X_s - m_s(X_s)}{2s^{3/2}} - \frac{X_s}{2s^{3/2}} = -\frac{m_s(X_s)}{2s^{3/2}}, \\ X_r &= \sqrt{\frac{r}{\tau}}\,x + \frac{\sqrt{r}}{2}\int_r^\tau \frac{m_s(X_s)}{s^{3/2}}\,\mathrm{d}s.\end{aligned} \tag{B.20}$$

Let $e_\tau = \widehat{m}_\tau(x) - m_\tau(x)$ for this calculation. Subtracting the two expressions and using $m_s(X_s) - m_\tau(x) = -\int_s^\tau \mathcal{D}_u^y m_u(X_u)\,\mathrm{d}u$ gives

$$\widehat{X}_r - X_r = \left(1 - \sqrt{\frac{r}{\tau}}\right)e_\tau + \frac{\sqrt{r}}{2}\int_r^\tau \frac{1}{s^{3/2}}\left[\int_s^\tau \mathcal{D}_u^y m_u(X_u)\,\mathrm{d}u\right]\mathrm{d}s. \tag{B.21}$$

Equivalently, interchanging the two integrals gives the single weighted remainder

$$\widehat{X}_r - X_r = \left(1 - \sqrt{\frac{r}{\tau}}\right)e_\tau + \int_r^\tau \left(1 - \sqrt{\frac{r}{u}}\right)\mathcal{D}_u^y m_u(X_u)\,\mathrm{d}u. \tag{B.22}$$

For $L_{r,\tau}(x) = \sup_{r\le u\le\tau}\|\mathcal{D}_u^y m_u(X_u)\| < \infty$, the integral coefficient is exactly

$$\frac{\sqrt{r}}{2}\int_r^\tau \frac{\tau - s}{s^{3/2}}\,\mathrm{d}s = \frac{\sqrt{r}}{2}\left(\frac{2\tau}{\sqrt{r}} + 2\sqrt{r} - 4\sqrt{\tau}\right) = (\sqrt{\tau} - \sqrt{r})^2. \tag{B.23}$$

Consequently,

$$\|\widehat{X}_r - X_r\| \le \left(1 - \sqrt{\frac{r}{\tau}}\right)\|e_\tau\| + (\sqrt{\tau} - \sqrt{r})^2 L_{r,\tau}(x). \tag{B.24}$$

For $r = \tau - h$ with $0 < h \le \tau/2$, the two coefficients obey

$$1 - \sqrt{1 - h/\tau} = \frac{h/\tau}{1 + \sqrt{1 - h/\tau}}, \qquad (\sqrt{\tau} - \sqrt{\tau - h})^2 = \frac{h^2}{(\sqrt{\tau} + \sqrt{\tau - h})^2}. \tag{B.25}$$

Thus the segment error is $O(h\|e_\tau\|/\tau + h^2 L_{\tau-h,\tau}/\tau)$. This establishes the local truncation order on positive-variance segments. At the clean endpoint, if $L_{0,\tau} < \infty$, the same bound instead gives

$$\|\widehat{m}_\tau(x) - F_\tau(x)\| \le \|e_\tau\| + \tau L_{0,\tau}(x). \tag{B.26}$$

The full map has a remainder controlled by its source variance τ; the outer rematching increment is a separate quantity.

For the accepted map, the additional displacement is $g_\tau(x) = \Psi_\tau(x) - \widehat{m}_\tau(x)$. Thus $\|\Psi_\tau(x) - F_\tau(x)\| \le \|e_\tau\| + \|g_\tau\| + \tau L_{0,\tau}$ under the same assumptions. This is an endpoint approximation bound; rematching changes the input law and is analyzed separately in Appendix D.1.

C Posterior Transport Matching

We prove the two main theorems, then describe volume transport under smooth changes of coordinates.

C.1 THEOREM 1: GENERAL DIFFUSION FORM AND PROOF

Let $T > 0$, let $\pi = \pi^y$ be any Borel probability measure on $\mathbb{R}^d$, and let $\Psi : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ be Borel measurable. Fix a narrowly continuous probability curve $q = (q_s)_{0 \le s \le T}$ with $(\Psi_s)_{\#} q_s = \pi$ for every s. For a bimeasurable bijection, $q_s = (\Psi_s^{-1})_{\#}\pi$; otherwise a compatible curve is part of the specification. Write $\mathrm{m}(\mathrm{d}s, \mathrm{d}x) = \mathrm{d}s\, q_s(\mathrm{d}x)$ and let $\| \cdot \|_q$ denote the $L^2(\mathrm{m})$ norm.

Choose Borel reference coefficients $b^0 : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d \times r}$, with

$$a = \sigma\sigma^\top, \qquad L_s^0 \phi = b_s^0 \cdot \nabla\phi + \tfrac{1}{2} a_s : \nabla^2 \phi, \qquad \int \big(\|b^0\| + \operatorname{tr} a\big)\, \mathrm{dm} < \infty. \tag{C.1}$$

The matrix a may depend on time and position and may be singular; no uniform ellipticity is assumed. It is a prescribed diffusion covariance, not a denoiser Jacobian. For $\phi \in C_c^\infty((0,T) \times \mathbb{R}^d)$, define

$$\begin{aligned} \ell_{q,L^0}(\phi) &= -\int (\partial_s \phi + L_s^0 \phi)\, \mathrm{dm}, \\ \mathsf{G}_q^\sigma &= \overline{\{\sigma^\top \nabla\phi : \phi \in C_c^\infty((0,T) \times \mathbb{R}^d)\}}^{L^2(\mathrm{m};\mathbb{R}^r)}, \\ \mathcal{J}_{q,L^0}(\phi) &= \tfrac{1}{2}\|\sigma^\top \nabla\phi\|_q^2 - \ell_{q,L^0}(\phi), \qquad \mathcal{A}_{L^0}(q) = \sup_\phi \{2\ell_{q,L^0}(\phi) - \|\sigma^\top \nabla\phi\|_q^2\}. \end{aligned} \tag{C.2}$$

Thus rematching measures control energy in the directions supplied by the reference diffusion. The main text takes $b^0 = 0$ and $\sigma = I$. In that case, the functional agrees with entropic Action Matching (Neklyudov et al., 2023, Proposition 3.2); compact temporal support removes its boundary terms. Here the prescribed marginals are the input laws compatible with the chosen maps, and the proof below allows a general reference diffusion.

Theorem 1 (Minimum-energy transport matching; general diffusion form). *Under the preceding assumptions, $\mathcal{A}_{L^0}(q) < \infty$ if and only if q admits a weak controlled realization*

$$\mathrm{d}X_s = \big[b_s^0(X_s) + \sigma_s(X_s)\beta_s\big]\mathrm{d}s + \sigma_s(X_s)\mathrm{d}W_s, \qquad X_s \sim q_s, \qquad \mathbb{E}\int_0^T \|\beta_s\|^2 \mathrm{d}s < \infty. \tag{C.3}$$

When these equivalent finite-action conditions hold, the matching objective extends to G_q^σ and has a unique minimizer h^* there. A realization with $\beta_s = h_s^*(X_s)$ attains

$$\inf_\beta \mathbb{E}\int_0^T \|\beta_s\|^2 \mathrm{d}s = \|h^*\|_q^2 = \mathcal{A}_{L^0}(q) = -2 \inf_\phi \mathcal{J}_{q,L^0}(\phi), \qquad \Psi_s(X_s) \sim \pi. \tag{C.4}$$

The infimum over controls ranges over all progressively measurable controls with the prescribed marginals, including controls that depend on the full past. Moreover, $\mathcal{A}_{L^0}(q) = 0$ exactly when q solves the reference Fokker–Planck equation $\partial_s q_s = (L_s^0)^* q_s$.

The theorem assumes neither Gaussian marginals nor Gaussian transition kernels, and requires no differentiability or invertibility of Ψ. Uniqueness concerns the minimizing control in $L^2(\mathrm{m})$, not the stochastic path law or a classical potential. For a singular diffusion, finiteness of the action also tests whether the available control directions can realize the chosen input-law curve.

Proof. For every real c, finiteness of the action bounds $2c\ell_{q,L^0}(\phi) - c^2\|\sigma^\top \nabla\phi\|_q^2$. Optimizing over c gives

$$|\ell_{q,L^0}(\phi)|^2 \le \mathcal{A}_{L^0}(q)\|\sigma^\top \nabla\phi\|_q^2. \tag{C.5}$$

When the norm vanishes, the same bound forces $\ell_{q,L^0}(\phi) = 0$. The functional is therefore well defined on diffusion-weighted gradients and extends continuously to G_q^σ. The Riesz representation theorem (Brezis, 2011) supplies a unique $h^* \in \mathsf{G}_q^\sigma$ such that

$$\ell_{q,L^0}(\phi) = \langle h^*, \sigma^\top \nabla\phi \rangle_q. \tag{C.6}$$

Completing the square and using density of the test gradients yields

$$\begin{aligned} \mathcal{J}_{q,L^0}(\phi) &= \tfrac{1}{2}\|\sigma^\top \nabla\phi - h^*\|_q^2 - \tfrac{1}{2}\|h^*\|_q^2, \\ \mathcal{A}_{L^0}(q) &= \|h^*\|_q^2, \qquad \inf_\phi \mathcal{J}_{q,L^0}(\phi) = -\tfrac{1}{2}\mathcal{A}_{L^0}(q). \end{aligned} \tag{C.7}$$

The minimum is attained in the completed space; smooth test potentials need only approximate its control in norm.

Choose a Borel representative of h^*. The Riesz identity is precisely weak probability conservation:

$$\begin{aligned} 0 &= \int \left[\partial_s \phi + (b^0 + \sigma h^*) \cdot \nabla \phi + \tfrac{1}{2} a : \nabla^2 \phi\right] \mathrm{dm}, \\ \partial_s q_s &= -\nabla \cdot \left(q_s (b^0_s + \sigma_s h^*_s)\right) + \tfrac{1}{2} \sum_{i,j} \partial_{ij} (a_{s,ij} q_s). \end{aligned} \tag{C.8}$$

The second line is distributional; no density is needed. With $\bar{b} = b^0 + \sigma h^*$, Cauchy–Schwarz gives the required coefficient integrability:

$$\int \|\bar{b}\| \,\mathrm{dm} \le \int \|b^0\| \,\mathrm{dm} + \left(\int \operatorname{tr} a \,\mathrm{dm}\right)^{1/2} \mathcal{A}_{L^0}(q)^{1/2} < \infty. \tag{C.9}$$

Together with narrow continuity and $\int \operatorname{tr} a \,\mathrm{dm} < \infty$, this verifies the hypotheses of the diffusion superposition principle (Trevisan, 2016, Theorem 2.5). It produces a continuous-path law with marginals q_s solving the martingale problem for $\bar{b} \cdot \nabla + \frac{1}{2} a : \nabla^2$. Indeed, for $f \in C_c^\infty(\mathbb{R}^d)$,

$$M_s^f = f(X_s) - f(X_0) - \int_0^s \left(\bar{b}_u \cdot \nabla f + \tfrac{1}{2} a_u : \nabla^2 f\right)(X_u) \mathrm{d}u \tag{C.10}$$

is a martingale. Localizing the coordinate and quadratic coordinate functions identifies a continuous local martingale

$$M_s = X_s - X_0 - \int_0^s \bar{b}_u(X_u) \mathrm{d}u, \qquad \mathrm{d}\langle M \rangle_s = a_s(X_s) \mathrm{d}s. \tag{C.11}$$

This also yields the prescribed factor σ when it is singular. Let $\sigma^\dagger$ denote its Moore–Penrose inverse, evaluated along X, and enlarge the space by an independent r-dimensional Brownian motion B. On this product enlargement, M remains a local martingale and $[M, B] = 0$. With $P_s = \sigma_s^\dagger \sigma_s$, define

$$W_s = \int_0^s \sigma_u^\dagger \,\mathrm{d}M_u + \int_0^s (I_r - P_u) \,\mathrm{d}B_u. \tag{C.12}$$

The stochastic integrals are well defined because P_u is an orthogonal projection and

$$\sigma^\dagger a (\sigma^\dagger)^\top = P, \qquad \mathrm{d}\langle W \rangle_s = [P_s + (I_r - P_s)] \mathrm{d}s = I_r \mathrm{d}s. \tag{C.13}$$

Lévy's characterization (Karatzas & Shreve, 1991) makes W Brownian, while $\sigma(I_r - P) = 0$ and the zero quadratic variation of $M - \int \sigma \sigma^\dagger \mathrm{d}M$ give $M = \int \sigma \mathrm{d}W$. Hence the superposition law realizes Equation (C.3) with $\beta_s = h^*_s(X_s)$. Its endpoint law is

$$\mathrm{Law}(\Psi_s(X_s)) = (\Psi_s)_\# q_s = \pi \qquad \text{for every } s \in [0, T]. \tag{C.14}$$

No uniqueness of the stochastic realization has been used.

Conversely, let any adapted finite-energy control β realize the same marginals. Itô's formula, the compact support of the test potential, and the preceding integrability estimates give

$$\ell_{q, L^0}(\phi) = \mathbb{E} \int_0^T \beta_s \cdot \sigma_s(X_s)^\top \nabla \phi_s(X_s) \,\mathrm{d}s. \tag{C.15}$$

The pointwise inequality $2v \cdot w - \|w\|^2 \le \|v\|^2$ implies

$$\mathcal{A}_{L^0}(q) \le \mathbb{E} \int_0^T \|\beta_s\|^2 \mathrm{d}s < \infty. \tag{C.16}$$

Approximate h^* by the test gradients in this identity. Convergence in $L^2(\mathrm{m})$ is convergence in $L^2(\mathrm{d}s \,\mathrm{d}\mathbb{P})$ after evaluation at X_s, since $X_s \sim q_s$. Consequently

$$\mathbb{E} \int_0^T \beta_s \cdot h^*_s(X_s) \mathrm{d}s = \|h^*\|_q^2, \qquad \mathbb{E} \int_0^T \|\beta_s\|^2 \mathrm{d}s = \mathcal{A}_{L^0}(q) + \mathbb{E} \int_0^T \|\beta_s - h^*_s(X_s)\|^2 \mathrm{d}s. \tag{C.17}$$

This proves optimality among history-dependent controls as well as state controls. For $h \in L^2(\mathrm{m}; \mathbb{R}^r)$, q solves the controlled Fokker–Planck equation exactly when $\nabla \cdot (q\sigma(h - h^*)) = 0$ distributionally; such deviations are orthogonal to G_q^σ.

Finally, zero action is equivalent to $\ell_{q, L^0}(\phi) = 0$ for every test potential, which is exactly $\partial_s q_s = (L_s^0)^* q_s$. If that reference equation is unique for q_0, zero additional control characterizes its unique marginal curve. Otherwise the statement remains the existence of a reference-diffusion realization of the specified curve. □

Recovery of the main-text theorem. Set $b^0 = 0$ and $\sigma = I$. Then $\mathsf{G}_q^\sigma = \mathsf{G}_q$, $h^* = b^*$, and

$$\ell_q(\phi) = -\int (\partial_s\phi + \tfrac{1}{2}\Delta\phi)\mathrm{dm} = \langle b^*, \nabla\phi\rangle_q, \qquad \mathcal{A}(q) = -2\inf_\phi \mathcal{J}_q(\phi) < \infty. \tag{C.18}$$

These are the notation and objective used in the main text and in the following path-entropy and smooth-coordinate calculations. The optimal control is the least mean-square drift. Zero action now implies the heat equation. Testing against localized Fourier modes, then removing the localization, gives

$$\partial_s\widehat{q}_s(\omega) = -\tfrac{1}{2}\|\omega\|^2\widehat{q}_s(\omega), \qquad \widehat{q}_s(\omega) = e^{-s\|\omega\|^2/2}\widehat{q}_0(\omega). \tag{C.19}$$

Thus $q_s = \mathcal{H}_s q_0$ if and only if $b^* = 0$. For probability-flow maps with $q_0 = \pi$, the transition from a clean endpoint is

$$X_{s'} \mid X_0 = z \sim \mathcal{N}(z, s'I), \qquad (\Psi_{s'})_\#(\mathcal{H}_{s'}\pi) = \pi. \tag{C.20}$$

This is Gaussian rematching. Each outer transition restarts at its clean endpoint, so s' is the next prescribed variance even for a decreasing outer schedule. For general b^0 and σ, the reference transition need not be Gaussian; its covariance rate alone does not specify its finite-time law.

C.2 Theorem 2: statement and proof

Use the unit-diffusion specialization of Theorem 1, with compatible laws $q_s = q_s^\Psi$, gradient space G_q, and minimizing drift b^*. Write $\mathbb{P}^*$ for its law on $C([0,T];\mathbb{R}^d)$, with marginals q_s. Let $\widehat{b} = \nabla\phi \in \mathsf{G}_q$ be a candidate gradient, interpreted by completion when needed, and let $\mathbb{P}^\phi$ be a weak solution with drift $\widehat{b}$, unit diffusion, and initial law μ_0 fixed across candidates. The objective is extended to gradients as in Appendix C.1. For $\phi \in C^{1,2}$ with finite endpoint expectations and $\int_0^T \mathbb{E}_{q_s}(|\partial_s\phi_s| + |\Delta\phi_s| + \|\nabla\phi_s\|^2)\mathrm{d}s < \infty$, the extension includes the endpoint terms:

$$\mathcal{J}_q(\phi) = \int_0^T \mathbb{E}_{q_s}\left[\tfrac{1}{2}\|\nabla\phi_s\|^2 + \partial_s\phi_s + \tfrac{1}{2}\Delta\phi_s\right]\mathrm{d}s - \mathbb{E}_{q_T}\phi_T + \mathbb{E}_{q_0}\phi_0. \tag{C.21}$$

Indeed, testing the forward equation gives the part linear in ϕ as $-\int_0^T \mathbb{E}_{q_s}[b_s^*\cdot\nabla\phi_s]\mathrm{d}s$. The boundary terms vanish for the compactly supported potentials in Equation (10).

Assume $q_0 \ll \mu_0$ with finite $\mathrm{KL}(q_0\|\mu_0)$, nonexplosion of both processes, and uniqueness in law for the martingale problem with drift b^* and initial law q_0. Set $r_0 = \mathrm{d}q_0/\mathrm{d}\mu_0$ and $e = b^* - \widehat{b}$. Assume $e_s(X_s)$ is locally square-integrable under $\mathbb{P}^\phi$ and that the weighted stochastic exponential

$$L_s = r_0(X_0)\exp\left\{\int_0^s e_r(X_r)\cdot\mathrm{d}W_r^\phi - \tfrac{1}{2}\int_0^s \|e_r(X_r)\|^2\mathrm{d}r\right\} \tag{C.22}$$

is a true martingale of mean one under $\mathbb{P}^\phi$. Finally assume $\mathbb{E}_{\mathbb{P}^*}\int_0^T \|e_s(X_s)\|^2\mathrm{d}s < \infty$. These are explicit sufficient Girsanov and entropy conditions; finite action in Theorem 1 alone does not imply them (Léonard, 2012).

Theorem 2 (Path-space KL representation; full statement). *Under the preceding assumptions,*

$$\mathcal{J}_q(\phi) = \mathrm{KL}(\mathbb{P}^*\|\mathbb{P}^\phi) + \inf\mathcal{J}_q - \mathrm{KL}(q_0\|\mu_0). \tag{C.23}$$

The objectives have identical minimizers on any admissible candidate family and identical parameter gradients wherever differentiable. For $\mu_0 = q_0$, the additive constant is $\inf\mathcal{J}_q$.

Proof. The Riesz identity of Theorem 1 gives, in the completed gradient space,

$$\begin{aligned}\mathcal{J}_q(\phi) &= \tfrac{1}{2}\|\widehat{b}\|_q^2 - \langle b^*, \widehat{b}\rangle_q \\ &= \tfrac{1}{2}\|\widehat{b} - b^*\|_q^2 - \tfrac{1}{2}\|b^*\|_q^2, \\ \mathcal{J}_q(\phi) - \inf\mathcal{J}_q &= \tfrac{1}{2}\int_0^T \mathbb{E}_{q_s}\|\widehat{b}_s - b_s^*\|^2\mathrm{d}s.\end{aligned} \tag{C.24}$$

The stochastic change of measure with density L_T changes the initial law to q_0 and the drift to b^*. Reference uniqueness therefore identifies that law as $\mathbb{P}^*$, so $\mathrm{d}\mathbb{P}^*/\mathrm{d}\mathbb{P}^\phi = L_T$. Under $\mathbb{P}^*$, $W_s^* = W_s^\phi - \int_0^s e_r(X_r)\mathrm{d}r$ is Brownian motion. Substitution in the logarithm of Equation (C.22) yields

$$\begin{aligned} \log \frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}^\phi} &= \log r_0(X_0) + \int_0^T e_s(X_s)\cdot \mathrm{d}W_s^* + \tfrac{1}{2}\int_0^T \|e_s(X_s)\|^2 \mathrm{d}s, \\ \mathrm{KL}(\mathbb{P}^*\|\mathbb{P}^\phi) &= \mathrm{KL}(q_0\|\mu_0) + \tfrac{1}{2}\int_0^T \mathbb{E}_{q_s}\|b_s^* - \widehat{b}_s\|^2 \mathrm{d}s. \end{aligned} \tag{C.25}$$

The stochastic integral has zero expectation by its square integrability under $\mathbb{P}^*$. Combining with Equation (C.24) proves the stated identity. The map, required law curve, and initial distribution are held fixed across candidates; the remaining terms are therefore independent of ϕ. This proves the minimizer and gradient equivalence. □

The posterior consequence follows by applying a common output map. Let $\mathcal{R}(\omega) = \Psi_T(\omega_T)$ and write $\nu^\phi = \mathcal{R}_\# \mathbb{P}^\phi$. Compatibility gives $\mathcal{R}_\#\mathbb{P}^* = \pi^y$. Euclidean and continuous-path spaces are standard Borel, so regular conditional path laws given this output exist. Disintegrating the path likelihood ratio through $z = \mathcal{R}(\omega)$ gives

$$\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}^\phi}(\omega) = \frac{\mathrm{d}\pi^y}{\mathrm{d}\nu^\phi}(z)\frac{\mathrm{d}\mathbb{P}^*(\cdot \mid \mathcal{R}=z)}{\mathrm{d}\mathbb{P}^\phi(\cdot \mid \mathcal{R}=z)}(\omega), \quad \mathbb{P}^*\text{-a.e.} \tag{C.26}$$

Integrating the logarithm proves the stronger decomposition

$$\begin{aligned} \mathrm{KL}(q_0\|\mu_0) + \mathcal{J}_q(\phi) - \inf \mathcal{J}_q = {} & \mathrm{KL}(\pi^y\|\nu^\phi) \\ & + \int \mathrm{KL}\big(\mathbb{P}^*(\cdot \mid \mathcal{R}=z) \,\big\|\, \mathbb{P}^\phi(\cdot \mid \mathcal{R}=z)\big)\, \pi^y(\mathrm{d}z). \end{aligned} \tag{C.27}$$

The conditional entropy term is nonnegative, giving

$$\mathrm{KL}(\pi^y\|\nu^\phi) \le \mathrm{KL}(q_0\|\mu_0) + \mathcal{J}_q(\phi) - \inf \mathcal{J}_q. \tag{C.28}$$

It measures information in the stochastic path that is not determined by its clean endpoint; no inverse of the transport map is used.

With $\mu_0 = q_0$, the earlier equal-initialization form is recovered:

$$\mathrm{KL}\big(\pi^y\|(\Psi_T)_\#\rho_T^\phi\big) \le \mathcal{J}_q(\phi) - \inf \mathcal{J}_q. \tag{C.29}$$

The result identifies the variational drift objective with path-law matching for the prescribed compatible curve. Appendix D.1 separately analyzes the finite Gaussian transition and its potential approximation.

C.3 Stochastic composition and volume transport

Let Ψ_s be a C^1-in-time, C^3-in-space family of orientation-preserving diffeomorphisms, and let π be a positive smooth density. Assume nonexplosion, integrability of the derivatives below, and boundary decay sufficient for integration by parts. Write $J_s = \nabla_x \Psi_s$ and $j_s = \det J_s > 0$. The required input density and its transport velocity are

$$q_s = (\pi \circ \Psi_s) j_s, \qquad w_s = -J_s^{-1}\partial_s \Psi_s, \qquad \partial_s q_s + \nabla\cdot(q_s w_s) = 0. \tag{C.30}$$

Consequently the drift $\widetilde{b}_s = w_s + \nabla \log q_s/2$ is compatible:

$$-\nabla\cdot(q_s\widetilde{b}_s) + \tfrac{1}{2}\Delta q_s = -\nabla\cdot(q_s w_s) = \partial_s q_s. \tag{C.31}$$

If it has finite energy, its projection onto the completed gradient space gives Theorem 1's minimum-energy control. A discarded component has zero q_s-weighted divergence and does not change these marginals.

For the deterministic map family used here, the Kunita–Itô–Wentzell formula reduces componentwise to the time-dependent Itô rule (Kunita, 1990; Bethencourt de Léon et al., 2020). For $f \in C^{1,2}$ and $\mathrm{d}X_s = b_s(X_s)\mathrm{d}s + \mathrm{d}W_s$,

$$\mathrm{d}f_s(X_s) = \big(\partial_s f_s + b_s\cdot\nabla_x f_s + \tfrac{1}{2}\Delta_x f_s\big)(X_s)\mathrm{d}s + \nabla_x f_s(X_s)\cdot \mathrm{d}W_s. \tag{C.32}$$

The map has no stochastic field component, so there is no additional field–flow cross-variation term.

Apply the scalar specialization componentwise to $Z_s = \Psi_s(X_s)$, where $\mathrm{d}X_s = \widetilde{b}_s(X_s)\mathrm{d}s + \mathrm{d}W_s$. The time-transport terms cancel:

$$\begin{aligned}\mathrm{d}Z_s &= \left[\partial_s\Psi_s + J_s\widetilde{b}_s + \tfrac{1}{2}\Delta_x\Psi_s\right](X_s)\mathrm{d}s + J_s(X_s)\mathrm{d}W_s \\ &= \tfrac{1}{2}\left[J_sJ_s^\top(\nabla\log\pi)\circ\Psi_s + J_s\nabla_x\log j_s + \Delta_x\Psi_s\right](X_s)\mathrm{d}s + J_s(X_s)\mathrm{d}W_s.\end{aligned} \tag{C.33}$$

The score identity used here follows by differentiating $\log q_s = \log\pi\circ\Psi_s + \log j_s$. For this calculation only, define $Q_s(\Psi_s(x)) = J_s(x)J_s(x)^\top$ and $(\operatorname{div} Q)_i = \sum_j \partial_{z_j}Q_{ij}$. Since $\partial_{z_j} = \sum_k (J_s^{-1})_{kj}\partial_{x_k}$,

$$\begin{aligned}(\operatorname{div}_z Q_s)_i\circ\Psi_s &= \sum_{j,k,\ell}(J_s^{-1})_{kj}\partial_{x_k}\left[(J_s)_{i\ell}(J_s)_{j\ell}\right] \\ &= \sum_\ell \partial_{x_\ell}(J_s)_{i\ell} + \sum_\ell (J_s)_{i\ell}\operatorname{tr}(J_s^{-1}\partial_{x_\ell}J_s) \\ &= \Delta_x\Psi_s^i + (J_s\nabla_x\log j_s)_i.\end{aligned} \tag{C.34}$$

The second equality uses symmetry of each component Hessian, $\partial_{x_k}(J_s)_{j\ell} = \partial_{x_\ell}(J_s)_{jk}$; the third uses Jacobi's determinant identity. No symmetry of J_s is required. Thus the mapped covariance is Q_s, and its drift is $(Q_s\nabla\log\pi + \operatorname{div} Q_s)/2$. Substituting these coefficients into the forward equation gives

$$\begin{aligned}\partial_s\nu_s &= -\tfrac{1}{2}\nabla_z\cdot[\nu_s(Q_s\nabla\log\pi + \operatorname{div}Q_s)] + \tfrac{1}{2}\sum_{i,j}\partial_{z_iz_j}\big((Q_s)_{ij}\nu_s\big), \\ \partial_s\nu_s|_{\nu_s=\pi} &= -\tfrac{1}{2}\nabla_z\cdot(Q_s\nabla\pi + \pi\operatorname{div}Q_s) + \tfrac{1}{2}\nabla_z\cdot(Q_s\nabla\pi + \pi\operatorname{div}Q_s) = 0.\end{aligned} \tag{C.35}$$

The volume derivative and Itô response jointly form the divergence of the mapped covariance. Composing the source diffusion with the map includes both; omitting a nonzero contribution generally changes the stationary law. Isotropic diffusion coefficients do not require isotropic marginal densities.

For a probability-flow map F_s, $q_s = \mathcal{H}_s\pi$ and $w_s = -\nabla\log q_s/2$, so $\widetilde{b}_s = 0$. With independent $\xi\sim\mathcal{N}(0,I)$ and Gaussian density γ_t, the finite transition satisfies

$$\begin{aligned}\int\pi(z)\mathbb{E}_\xi f(F_t(z+\sqrt{t}\xi))\,\mathrm{d}z &= \int f(F_t(x))\left[\int\pi(z)\gamma_t(x-z)\,\mathrm{d}z\right]\mathrm{d}x \\ &= \int f(F_t(x))q_t(x)\,\mathrm{d}x = \int f(w)\pi(w)\,\mathrm{d}w.\end{aligned} \tag{C.36}$$

This finite identity uses no small-step expansion. For numerical maps, the compatible drift and terminal potential can change; Appendix D.1 retains those approximation errors explicitly.

D Finite Rematching and Posterior Accuracy

D.1 Compatibility and the finite optimization objective

We now identify an objective whose minimization recovers compatibility for a fixed numerical transport map, and quantify the approximation made by the finite Gaussian update in Algorithm 1. Fix an incoming endpoint law ρ, a variance $t = \tau' > 0$, a measurable next map Ψ, and $K_0(z,\mathrm{d}x) = \mathcal{N}(z,tI)(\mathrm{d}x)$. Let $R_0(\mathrm{d}z,\mathrm{d}x) = \rho(\mathrm{d}z)K_0(z,\mathrm{d}x)$. The entropy projection keeps the incoming law fixed while imposing the mapped output:

$$\inf_R\left\{\mathrm{KL}(R\|R_0) : R_Z = \rho,\quad \Psi_\#R_X = \pi^y\right\}. \tag{D.1}$$

This is a constrained Schrödinger problem (Léonard, 2014). Assume finite-entropy feasibility and attainment by a finite dual potential f_*, with the following normalizers and expectations finite; feasibility alone need not imply dual attainment. For a candidate potential f, define

$$\begin{aligned}Z_f(z) &= \int e^{f(\Psi(x))}K_0(z,\mathrm{d}x), \qquad K_f(z,\mathrm{d}x) = \frac{e^{f(\Psi(x))}}{Z_f(z)}K_0(z,\mathrm{d}x), \\ \mathcal{L}(f) &= \mathbb{E}_\rho\log Z_f(Z) - \mathbb{E}_{\pi^y}f, \qquad R_f = \rho K_f,\quad \nu_f = \Psi_\#(R_f)_X.\end{aligned} \tag{D.2}$$

Normalization separately for every z preserves $R_Z = \rho$. For admissible bounded test functions h, differentiation under the integral gives

$$D\mathcal{L}(f)[h] = \mathbb{E}_{\nu_f} h - \mathbb{E}_{\pi^y} h, \qquad D^2\mathcal{L}(f)[h,h] = \mathbb{E}_\rho \operatorname{Var}_{K_f(Z,\cdot)}[h(\Psi(X))] \geq 0. \tag{D.3}$$

Consequently stationarity against all such tests is precisely $\nu_f = \pi^y$. For any feasible R, the change of measure yields

$$\mathrm{KL}(R\|R_0) = \mathrm{KL}(R\|R_{f_*}) - \mathcal{L}(f_*), \tag{D.4}$$

which proves primal optimality of R_{f_*} when $\nu_{f_*} = \pi^y$. Moreover,

$$\begin{aligned} \mathrm{KL}(R_{f_*}\|R_f) &= \mathbb{E}_{\pi^y}(f_* - f) + \mathbb{E}_\rho(\log Z_f - \log Z_{f_*}) = \mathcal{L}(f) - \mathcal{L}(f_*), \\ \mathrm{KL}(\pi^y\|\nu_f) &\leq \mathcal{L}(f) - \mathcal{L}(f_*). \end{aligned} \tag{D.5}$$

Thus the dual objective closes finite-transition compatibility in reverse KL. It constrains the final mapped law; matching an entire prescribed curve q_s^Ψ remains the stronger requirement of Theorem 1. For a positive terminal tilt, assume the Gaussian convolutions below are finite and define a function $h \in C^{1,2}([0,t)\times\mathbb{R}^d)$. The same conditional kernel has a diffusion representation:

$$\begin{aligned} h_s(x) &= \mathcal{H}_{t-s}[e^{f\circ\Psi}](x), \qquad \mathrm{d}X_s = \nabla \log h_s(X_s)\mathrm{d}s + \mathrm{d}W_s, \\ X_0 &= z, \qquad \mathrm{Law}(X_t \mid X_0 = z) = K_f(z,\cdot). \end{aligned} \tag{D.6}$$

Here $\partial_s h_s + \Delta h_s/2 = 0$. Under Brownian motion started at z, $h_s(X_s)/h_0(z)$ is the conditional expectation of $e^{f(\Psi(X_t))}/h_0(z)$, hence a mean-one martingale defining the tilted path law. The finite objective below restricts this diffusion transition to Gaussian centers and approximates its terminal potential.

Posterior accuracy of a finite transition. The finite Gaussian objective has an exact conditional-relative-entropy interpretation. Write $U_* = -f_* \circ \Psi$ and, for any terminal energy U,

$$\begin{aligned} K_U(z,\mathrm{d}x) &= Z_U(z)^{-1} e^{-U(x)} K_0(z,\mathrm{d}x), \\ G_u(z,\cdot) &= \mathcal{N}(u(z), tI), \qquad \mathcal{L}^U_{t,z}(u) = \mathbb{E}_\xi U(u + \sqrt{t}\xi) + \frac{\|u-z\|^2}{2t}. \end{aligned} \tag{D.7}$$

Assume $0 < Z_U(z) < \infty$ and finite displayed expectations for ρ-almost every z. Expansion of the log-density ratio gives

$$\begin{aligned} \mathrm{KL}(G_u(z,\cdot)\|K_U(z,\cdot)) &= \mathrm{KL}(G_u(z,\cdot)\|K_0(z,\cdot)) + \mathbb{E}_{G_u} U + \log Z_U(z) \\ &= \mathcal{L}^U_{t,z}(u) + \log Z_U(z). \end{aligned} \tag{D.8}$$

Thus objective differences are exactly differences of conditional transition KL. The quadratic penalty also equals the least Brownian control energy for this Gaussian center:

$$\inf_{\substack{\mathrm{d}X_s = b_s\mathrm{d}s + \mathrm{d}W_s,\ X_0 = z \\ \mathrm{Law}(X_t) = \mathcal{N}(u,tI)}} \frac{1}{2}\mathbb{E}\int_0^t \|b_s\|^2 \mathrm{d}s = \frac{\|u-z\|^2}{2t}. \tag{D.9}$$

Indeed, $u - z = \mathbb{E}\int_0^t b_s \mathrm{d}s$ gives the lower bound by Cauchy–Schwarz, and the constant drift $b_s = (u-z)/t$ attains it.

For the compatible potential U_*, apply the entropy chain rule to the joint laws $\rho(\mathrm{d}z)G_u(z,\mathrm{d}x)$ and $\rho(\mathrm{d}z)K_*(z,\mathrm{d}x)$, then map their second coordinate by Ψ:

$$\begin{aligned} \mathrm{KL}(\Psi_\#(\rho G_u)\|\pi^y) &\leq \mathbb{E}_\rho \,\mathrm{KL}(G_u(Z,\cdot)\|K_*(Z,\cdot)) \\ &= \mathbb{E}_\rho\left[\mathcal{L}^{U_*}_{t,Z}(u(Z)) + \log Z_*(Z)\right]. \end{aligned} \tag{D.10}$$

Minimizing over Gaussian centers minimizes this upper bound. The optimal value includes the Gaussian-family restriction error; it need not be zero. For $U = E \circ \mathsf{D}_t$, Equation (D.8) gives the computed objective's exact local KL interpretation, while the following result retains the difference from U_* explicitly.

Proposition 1 (Finite rematching accuracy)**.** *Fix a probability measure ρ on $\mathbb{R}^n$ and a measurable map $\Psi : \mathbb{R}^n \to \mathbb{R}^n$. Assume a compatible kernel $K_*(z, \mathrm{d}x) = Z_*(z)^{-1} e^{-U_*(x)} \mathcal{N}(x; z, tI)\mathrm{d}x$ with $\Psi_\#(\rho K_*) = \pi^y$. Let $U_* \in C^2$ satisfy $\|\nabla^2 U_*\|_{\mathrm{op}} \le L_*$ globally and $tL_* < 1$. Let $U \in C^1$, let $\widehat{u}$ be measurable, and put $G_z = \mathcal{N}(\widehat{u}(z), tI)$. Assume differentiation under the Gaussian expectation is valid, ∇U is integrable under G_z, and the residuals below are square-integrable under ρ:*

$$a_z = \nabla \mathcal{L}^U_{t,z}(\widehat{u}(z)), \qquad b_z = \mathbb{E}_{G_z}(\nabla U - \nabla U_*), \qquad \beta_z = \|b_z\|. \tag{D.11}$$

Then

$$\begin{aligned} \mathrm{KL}(\Psi_\#(\rho G)\|\pi^y) &\le \frac{t}{2(1 - tL_*)} \mathbb{E}_\rho \left[\mathrm{Var}_{G_Z}(\nabla U_*) + \|a_Z - b_Z\|^2 \right] \\ &\le \frac{nt^2 L_*^2 + t\mathbb{E}_\rho(\|a_Z\| + \beta_Z)^2}{2(1 - tL_*)}, \end{aligned} \tag{D.12}$$

where the variance of a vector is the expected squared norm of its centered value.

Proof. For a fixed z, the negative log-density of $K_*(z, \cdot)$ has Hessian

$$\nabla^2 \left(U_*(x) + \frac{\|x - z\|^2}{2t} \right) = \nabla^2 U_*(x) + t^{-1} I \succeq (t^{-1} - L_*) I. \tag{D.13}$$

Set $m = t^{-1} - L_* > 0$. The kernel therefore satisfies a log-Sobolev inequality with constant at most m^{-1} (Bakry et al., 2014). The Gaussian center gives

$$\begin{aligned} \nabla_x \log \frac{G_z(x)}{K_*(z, x)} &= -\frac{x - \widehat{u}}{t} + \frac{x - z}{t} + \nabla U_*(x) = \frac{\widehat{u} - z}{t} + \nabla U_*(x), \\ a_z &= \frac{\widehat{u} - z}{t} + \mathbb{E}_{G_z} \nabla U. \end{aligned} \tag{D.14}$$

Subtracting the mean compatible force leaves a centered fluctuation:

$$\nabla_x \log \frac{G_z(x)}{K_*(z, x)} = \nabla U_*(x) - \mathbb{E}_{G_z} \nabla U_* + a_z - b_z. \tag{D.15}$$

The cross term integrates to zero. Consequently,

$$\begin{aligned} \mathrm{KL}(G_z \| K_*(z, \cdot)) &\le \frac{1}{2m} \mathbb{E}_{G_z} \left\| \nabla_x \log \frac{G_z}{K_*(z, \cdot)} \right\|^2 \\ &= \frac{1}{2m} \left[\mathrm{Var}_{G_z}(\nabla U_*) + \|a_z - b_z\|^2 \right]. \end{aligned} \tag{D.16}$$

For independent $X, X' \sim G_z$, Lipschitz continuity of ∇U_* yields

$$\begin{aligned} \mathrm{Var}_{G_z}(\nabla U_*) &= \tfrac{1}{2} \mathbb{E} \|\nabla U_*(X) - \nabla U_*(X')\|^2 \\ &\le \tfrac{1}{2} L_*^2 \mathbb{E} \|X - X'\|^2 = ntL_*^2. \end{aligned} \tag{D.17}$$

Integrating the conditional bound under ρ, using $\|a_z - b_z\| \le \|a_z\| + \beta_z$, and applying the common next map gives

$$\begin{aligned} \mathrm{KL}(\Psi_\#(\rho G)\|\pi^y) &\le \mathbb{E}_\rho \, \mathrm{KL}(G_Z \| K_*(Z, \cdot)) \\ &\le \frac{nt^2 L_*^2 + t\mathbb{E}_\rho(\|a_Z\| + \beta_Z)^2}{2(1 - tL_*)}. \end{aligned} \tag{D.18}$$

This proves Equation (D.12). The bound applies to measurable Ψ and uses only the curvature of the compatible potential. □

For a center computed using auxiliary draws $\omega \sim \lambda(\cdot \mid z)$, apply the conditional calculation to each $G_{z,\omega}$. Compare the joint laws $\rho(\mathrm{d}z)\lambda(\mathrm{d}\omega \mid z)G_{z,\omega}(\mathrm{d}x)$ and $\rho(\mathrm{d}z)\lambda(\mathrm{d}\omega \mid z)K_*(z, \mathrm{d}x)$. Forgetting ω and applying Ψ gives the same bound averaged over both z and ω. The innovation added after

optimizing the center is independent of ω. For a numerical gradient $\widehat{g}_{z,\omega}$ satisfying $\|\widehat{g}_{z,\omega} - a_{z,\omega}\| \le e_{z,\omega}$,

$$\|a_{z,\omega} - b_{z,\omega}\| \le \|\widehat{g}_{z,\omega}\| + e_{z,\omega} + \beta_{z,\omega}. \tag{D.19}$$

Substitution in Equation (D.12) proves Equation (16), with expectations also over the optimization draws when present.

Equation (D.12) separates three approximations. The centered force fluctuation is the error from retaining covariance tI; a_z is finite optimization error; β_z is terminal-potential error. For a family with $tL_* \to 0$, $nt^2L_*^2 \to 0$, and $t\mathbb{E}_\rho(\|a_Z\| + \beta_Z)^2 \to 0$, the bound gives posterior convergence of the complete step. The implementation uses $U = E \circ \mathsf{D}_t$; analytic evaluation adds its response error to the finite objective residual.

If the reference kernel already obeys $\Psi_\#(\rho K_0) = \pi^y$, it is feasible with zero relative entropy and is the unique entropy projection. One may take $U_* = 0$. For $U = 0$, the objective reduces to $\|u - z\|^2/(2t)$, so $u = z$ and $G_z = K_0(z, \cdot)$. All terms in the posterior bound vanish. In particular, $\rho = \pi^y$ and $(F_t)_\# \mathcal{H}_t \pi^y = \pi^y$ recover Gaussian rematching from Theorem 1.

D.2 Finite rematching objective and analytic evaluation

Fix an endpoint z, next variance $t = \tau' > 0$, and denoiser $D = \mathsf{D}_t$. Equation (D.8) gives the Gaussian objective $\mathcal{L}_{t,z}(u) = \mathbb{E}_\xi E(D(u + \sqrt{t}\xi)) + \|u - z\|^2/(2t)$. A computed center induces the transition

$$\mathrm{d}X_s = (u - z)\mathrm{d}s/t + \mathrm{d}W_s, \qquad X_0 = z, \qquad X_t \sim \mathcal{N}(u, tI). \tag{D.20}$$

The denoiser and observation model are fixed throughout the center solve.

Suppose first that $A(w) = Hw + b$ and $E(w) = \|Hw + b - y\|^2_{R^{-1}}/2$, with $R \succ 0$. Set $K = H^\top R^{-1} H$, $\widetilde{X} = u + \sqrt{t}\xi$, and $r(w) = D(w) - w$. The quadratic energy satisfies the exact identity

$$E(D(w)) = E(w) + \nabla E(w)^\top r(w) + \tfrac{1}{2} r(w)^T K r(w). \tag{D.21}$$

Since $\mathbb{E}(\widetilde{X} - u) = 0$ and $\mathrm{Cov}(\widetilde{X}) = tI$,

$$\begin{aligned} \mathbb{E}E(\widetilde{X}) &= E(u) + \frac{t}{2} \operatorname{tr} K, \\ \mathcal{L}_{t,z}(u) &= V_{t,z}(u) + \frac{t}{2} \operatorname{tr} K + \mathcal{R}_t(u), \\ \mathcal{R}_t(u) &= \mathbb{E}\left[\nabla E(\widetilde{X})^\top r(\widetilde{X}) + \tfrac{1}{2} r(\widetilde{X})^T K r(\widetilde{X})\right]. \end{aligned} \tag{D.22}$$

Thus $HD(w) = Hw$ suffices for exact agreement up to a center-independent constant. The denoiser may still change unobserved coordinates. Without this response condition, increasing the Monte Carlo sample count recovers the expectation but retains $\mathcal{R}_t$.

Assume next that r is differentiable and that differentiation under the Gaussian expectation is justified by an integrable dominating function on the relevant set of centers. Using $\nabla^2 E = K$ and the symmetry of K gives

$$\nabla \mathcal{R}_t(u) = \mathbb{E}\left[K r(\widetilde{X}) + J r(\widetilde{X})^\top \nabla E(\widetilde{X}) + J r(\widetilde{X})^T K r(\widetilde{X})\right]. \tag{D.23}$$

Both the denoiser displacement and its observed derivative response enter the approximation. For nonlinear observations, define instead

$$\mathcal{R}_t(u) = \mathbb{E}E(D(\widetilde{X})) - E(u). \tag{D.24}$$

It includes the denoiser response and the Gaussian averaging of the energy itself. For $E \in C^4$ with bounded fourth derivatives, the latter is $\mathbb{E}E(\widetilde{X}) = E(u) + t\Delta E(u)/2 + O(t^2)$; the Laplacian term now depends on u. An affine expansion of the observation gives a quadratic local model, with its linearization error retained separately. Smooth expansions require localization when the observation contains clipping or other nonsmooth operations.

For affine observations, $\nabla^2 V_{t,z} = K + t^{-1}I \succeq t^{-1}I$. Let u_V minimize $V_{t,z}$, and let u_* be a stationary point of $\mathcal{L}_{t,z}$ with finite $\nabla\mathcal{R}_t(u_*)$. Strong monotonicity and stationarity imply

$$\begin{aligned} t^{-1}\|u_* - u_V\|^2 &\le \langle \nabla V_{t,z}(u_*) - \nabla V_{t,z}(u_V), u_* - u_V\rangle \\ &= -\langle \nabla\mathcal{R}_t(u_*), u_* - u_V\rangle, \\ \|u_* - u_V\| &\le t\|\nabla\mathcal{R}_t(u_*)\|. \end{aligned} \tag{D.25}$$

For a finite analytic solve $\widehat{u}$, set $a = \nabla V_{t,z}(\widehat{u})$ and $\delta = \|\nabla\mathcal{R}_t(u_*)\|$. The same argument bounds $\|\widehat{u} - u_V\| \le t\|a\|$. Consequently,

$$\begin{aligned} \|\widehat{u} - u_*\| &\le t(\|a\| + \delta), \\ \mathrm{KL}(\mathcal{N}(\widehat{u}, tI)\|\mathcal{N}(u_*, tI)) &= \frac{\|\widehat{u} - u_*\|^2}{2t} \le \frac{t}{2}(\|a\| + \delta)^2. \end{aligned} \tag{D.26}$$

Data processing preserves this upper bound after any common measurable next map. For a non-convex energy, the same argument uses a positive strong-monotonicity constant on the relevant segments; global existence or uniqueness is then an additional assumption.

The small-variance order follows directly from Equation (D.23). Write $r(w) = ts_t(w)$ and suppose, along the stationary centers under consideration, that

$$\sup_{0<t<t_0} \mathbb{E}\left[\|Ks_t(\widetilde{X})\| + \|Js_t(\widetilde{X})^\top \nabla E(\widetilde{X})\| + t\|Js_t(\widetilde{X})^T Ks_t(\widetilde{X})\|\right] < \infty. \tag{D.27}$$

Then $\delta = O(t)$, so exact analytic optimization gives center discrepancy $O(t^2)$ and Gaussian-transition KL $O(t^3)$. These are conditional approximation orders, not consequences of a small noise label alone. They do not require a positive-semidefinite denoiser Jacobian.

For stochastic evaluation, the draws used in $\widehat{\mathcal{L}}$ are fixed during optimization and independent of the noise injected after the solve.

$$\widehat{\mathcal{L}}_{t,z}(u) = \frac{1}{M}\sum_{j=1}^{M} E(D(u + \sqrt{t}\xi_j)) + \frac{\|u - z\|^2}{2t}. \tag{D.28}$$

At a center $\widehat{u}$ computed using these draws, its population residual decomposes exactly as

$$\nabla\mathcal{L}_{t,z}(\widehat{u}) = \nabla\widehat{\mathcal{L}}_{t,z}(\widehat{u}) + \left[\nabla\mathcal{L}_{t,z}(\widehat{u}) - \nabla\widehat{\mathcal{L}}_{t,z}(\widehat{u})\right]. \tag{D.29}$$

The bracket is the Monte Carlo gradient error at the selected center; pointwise unbiasedness at a fixed center does not remove this adaptive discrepancy. Conditional on those draws, the actual transition is Gaussian about the computed center; averaging them gives a mixture kernel. For a common mixing distribution λ and conditional comparison kernels P_ω, Q_ω, the entropy chain rule gives

$$\mathrm{KL}\left(\int P_\omega\lambda(\mathrm{d}\omega)\,\middle\|\,\int Q_\omega\lambda(\mathrm{d}\omega)\right) \le \int \mathrm{KL}(P_\omega\|Q_\omega)\lambda(\mathrm{d}\omega). \tag{D.30}$$

This separates Monte Carlo and finite-solver effects from the analytic-response approximation. The gradient discrepancy enters e_z in Equation (16); the potential approximation enters β_z. Gaussian rematching sets the optimization budget to zero.

D.3 One-step matching and contraction

Write $\pi = \pi^y$ and fix a Markov kernel P that does not depend on the law being evolved. For a transport map Ψ and a clean-to-noisy kernel K,

$$\mathsf{P}(z, B) = \int \mathbf{1}_B(\Psi(x))\mathsf{K}(z, \mathrm{d}x), \qquad \nu = \pi\mathsf{P} = (\Psi)_\#(\pi\mathsf{K}), \qquad \rho^+ = \rho\mathsf{P}. \tag{D.31}$$

Thus $\pi\mathsf{K} = q^\Psi$ and $\Psi_\# q^\Psi = \pi$ imply $\nu = \pi$. For a sequence of endpoint families, the cross-level kernel is

$$\mathsf{T}_{\tau\to\tau'}(x, B) = \mathsf{K}_{\tau'}(\Psi_\tau(x), B). \tag{D.32}$$

Testing against an arbitrary bounded measurable function f gives

$$\begin{aligned}\int f(x')\,(q_\tau^\Psi \mathsf{T}_{\tau\to\tau'})(\mathrm{d}x') &= \int\!\!\int f(x')\mathsf{K}_{\tau'}(\Psi_\tau(x),\mathrm{d}x')q_\tau^\Psi(\mathrm{d}x)\\ &= \int\!\!\int f(x')\mathsf{K}_{\tau'}(z,\mathrm{d}x')\pi(\mathrm{d}z) = \int f(x')q_{\tau'}^\Psi(\mathrm{d}x').\end{aligned} \tag{D.33}$$

Applying $\Psi_{\tau'}$ gives the posterior output law in Equation (9). This argument is independent of whether $\tau' < \tau$; each $\mathsf{K}_{\tau'}$ is a clean-to-noisy transition with its own increasing internal clock.

Proposition 2 (One-step posterior error)**.** *For a fixed Markov kernel* P*, assume the displayed relative entropies and log-ratio expectation are finite, and* $\rho^+ \ll \nu \ll \pi$*. Then*

$$\begin{aligned}\mathrm{KL}(\rho^+\|\pi) &= \mathrm{KL}(\rho\|\pi) - \ell_{\mathsf{P}}(\rho) + \mathbb{E}_{\rho^+}\log\frac{\nu}{\pi},\\ \ell_{\mathsf{P}}(\rho) &= \mathrm{KL}(\rho\|\pi) - \mathrm{KL}(\rho\mathsf{P}\|\pi\mathsf{P}) \geq 0.\end{aligned} \tag{D.34}$$

Proof of Proposition 2. Assume finite input KL, the absolute continuity stated in the proposition, and integrability of the displayed log ratios. Define the common-kernel joint laws

$$P(\mathrm{d}z,\mathrm{d}z') = \rho(\mathrm{d}z)\mathsf{P}(z,\mathrm{d}z'), \qquad Q(\mathrm{d}z,\mathrm{d}z') = \pi(\mathrm{d}z)\mathsf{P}(z,\mathrm{d}z'). \tag{D.35}$$

If $r = \mathrm{d}\rho/\mathrm{d}\pi$, then $\mathrm{d}P/\mathrm{d}Q(z,z') = r(z)$, whence $\mathrm{KL}(P\|Q) = \mathrm{KL}(\rho\|\pi)$. Disintegrating with respect to the output and factoring the likelihood ratio in that order gives

$$\begin{aligned}\frac{\mathrm{d}P}{\mathrm{d}Q}(z,z') &= \frac{\mathrm{d}\rho^+}{\mathrm{d}\nu}(z')\frac{\mathrm{d}P(\cdot\mid z')}{\mathrm{d}Q(\cdot\mid z')}(z),\\ \mathrm{KL}(\rho\|\pi) &= \mathrm{KL}(\rho^+\|\nu) + \int \mathrm{KL}(P(\cdot\mid z')\|Q(\cdot\mid z'))\rho^+(\mathrm{d}z').\end{aligned} \tag{D.36}$$

The last integral is precisely $\ell_{\mathsf{P}}(\rho) \geq 0$. The output density-ratio identity then yields

$$\begin{aligned}\mathrm{KL}(\rho^+\|\pi) &= \int \log\left(\frac{\mathrm{d}\rho^+}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\pi}\right)\mathrm{d}\rho^+\\ &= \mathrm{KL}(\rho\|\pi) - \ell_{\mathsf{P}}(\rho) + \mathbb{E}_{\rho^+}\log\frac{\mathrm{d}\nu}{\mathrm{d}\pi}.\end{aligned} \tag{D.37}$$

This is Equation (D.34). For compatible transitions $\nu = \pi$, so the last term vanishes and data processing gives posterior KL nonincrease. For approximate transitions, the last term is a signed invariance error averaged under the actual output law. At $\rho = \pi$, it equals $\mathrm{KL}(\nu\|\pi)$; for other inputs it need not be nonnegative. □

Gaussian rematching and finite matching accuracy. Let $\mathsf{C}_k(z,\mathrm{d}u)$ denote the law of the computed center given an incoming endpoint z, including any independent draws used in optimization. A deterministic center $c_k(z)$ gives $\mathsf{C}_k(z,\mathrm{d}u) = \delta_{c_k(z)}(\mathrm{d}u)$. For $t_k = \tau_{k+1} > 0$, the actual finite transition is

$$\mathsf{P}_k(z,B) = \int\!\!\int \mathbf{1}_B(\Psi_{\tau_{k+1}}(x))\,\mathcal{N}(x;u,t_kI)\,\mathrm{d}x\,\mathsf{C}_k(z,\mathrm{d}u). \tag{D.38}$$

The center kernel and next map include the prescribed finite solves and selection rules and are fixed independently of the evolving input law. For Algorithm 1, $\rho_0 = (\Psi_{\tau_0})_\#\mathcal{N}(0,\tau_0 I)$ is the first endpoint law. In the last transition, replace Ψ_{τ_K} in Equation (D.38) by the actual output D_{τ_K}. All data-processing arguments hold for this measurable map as well; its invariance defect is retained. The Gaussian innovation is independent of the center computation. Write $\rho_{k+1} = \rho_k\mathsf{P}_k$, $\nu_k = \pi\mathsf{P}_k$, and $\chi_k = \pi\mathsf{C}_k$. Assume χ_k has log-Sobolev constant $0 < C_k < \infty$ in the convention

$$\mathrm{Ent}_{\chi_k}(f^2) \leq 2C_k\int\|\nabla f\|^2\,\mathrm{d}\chi_k, \qquad \mathrm{Ent}_\chi(g) = \int g\log\frac{g}{\int g\,\mathrm{d}\chi}\,\mathrm{d}\chi. \tag{D.39}$$

For finite $\mathrm{KL}(\rho_k\|\pi)$, data processing at the center and endpoint, with Gaussian-channel contraction (Klartag & Ordentlich, 2025, Theorem 1.7) between them, gives

$$\begin{aligned}\mathrm{KL}(\rho_{k+1}\|\nu_k) &= \mathrm{KL}\big((\Psi_{\tau_{k+1}})_{\#}\mathcal{H}_{t_k}(\rho_k\mathsf{C}_k)\,\big\|\,(\Psi_{\tau_{k+1}})_{\#}\mathcal{H}_{t_k}\chi_k\big)\\ &\le \mathrm{KL}(\mathcal{H}_{t_k}(\rho_k\mathsf{C}_k)\,\|\,\mathcal{H}_{t_k}\chi_k)\\ &\le \frac{C_k}{C_k+t_k}\,\mathrm{KL}(\rho_k\mathsf{C}_k\|\chi_k) \le \frac{C_k}{C_k+t_k}\,\mathrm{KL}(\rho_k\|\pi).\end{aligned} \tag{D.40}$$

The LSI is a property of the corrected-center law; it need not hold for an arbitrary finite solver. For a deterministic center, a sufficient condition is that π has constant C_π and c_k is L_k-Lipschitz; the chain rule then gives $C_k \le L_k^2 C_\pi$. No smoothness of the next endpoint is required for the data-processing inequality.

Suppose additionally that $\rho_{k+1} \ll \nu_k \ll \pi$ and the log-density-ratio expectations below are finite. Set

$$d_k = \mathrm{KL}(\rho_k\|\pi), \qquad \alpha_k = \frac{C_k}{C_k+t_k}, \qquad \delta_k = \mathbb{E}_{\rho_{k+1}}\log\frac{\mathrm{d}\nu_k}{\mathrm{d}\pi}. \tag{D.41}$$

Factoring the output density ratio and applying Equation (D.40) yields

$$d_{k+1} = \mathrm{KL}(\rho_{k+1}\|\nu_k) + \delta_k \le \alpha_k d_k + \delta_k. \tag{D.42}$$

Repeated substitution proves the varying-level estimate

$$d_N \le \left(\prod_{j=0}^{N-1}\alpha_j\right) d_0 + \sum_{k=0}^{N-1}\left(\prod_{j=k+1}^{N-1}\alpha_j\right)\delta_k, \tag{D.43}$$

where an empty product equals one. If $\alpha_k \le \alpha < 1$ and $\delta_k \le \varepsilon$ for a common $\varepsilon \ge 0$, then $d_k \ge 0$ gives $d_{k+1} \le \alpha d_k + \varepsilon$. Summing this geometric recurrence gives

$$\mathrm{KL}(\rho_N\|\pi^y) \le \underbrace{\alpha^N\,\mathrm{KL}(\rho_0\|\pi^y)}_{\text{initialization}} + \underbrace{\frac{1-\alpha^N}{1-\alpha}\,\varepsilon}_{\text{finite matching accuracy}}. \tag{D.44}$$

This retains both initialization error and the accumulated finite-transition defect. For matched transitions $\nu_k = \pi$, every δ_k is zero. For a persistent upper bound on invariance error, it gives $\limsup_{N\to\infty} d_N \le \varepsilon/(1-\alpha)$. The signed term δ_k depends on the complete output law; minimizing the center energy alone does not estimate it. Without a uniform contraction factor, Equation (D.43) still applies. The initialization term vanishes when $\prod_{j<N}\alpha_j \to 0$, equivalently $\sum_{j<N}\log(1+t_j/C_j) \to \infty$ for positive C_j. If the weighted sum of invariance defects also tends to zero, the bound certifies posterior convergence. Decreasing noise levels alone need not provide enough accumulated contraction.

For history-dependent or interacting-particle algorithms, the common kernel is defined on the corresponding augmented state; averaging selected endpoints is a further output operation. The single-particle calculation does not establish posterior accuracy of those benchmark output rules.

E Numerical Implementation

The map evaluation constructs $\Psi_\tau(x)$ from one denoising site per level. Rematching then computes the center of the next Gaussian transition. Stochastic objective evaluation uses additional denoiser calls inside that computation; the analytic approximation uses only the observation operator. All formulas below use variance τ; code that indexes the denoiser by standard deviation uses $\sigma = \sqrt{\tau}$.

E.1 Characteristic approximation and acceptance test

Appendix B.3 derives the integrating factor, its residual in the exact PDE, and the resulting endpoint error. The implementation accepts or rejects this proposal by evaluating the actual measurement energy, rather than assuming that the linearized proposal improves it. Let $d = \mathsf{D}_\tau(x)$, $E(u) =$

$\|A(u) - y\|^2_{R^{-1}}/2$, and $g = J_A(d)^\top R^{-1}(A(d) - y)$. Besides the moment displacement from Equation (7), the scalar fallback is obtained by restricting the local quadratic to $-g$:

$$\begin{aligned} \phi(t) &= \tfrac{1}{2}\|A(d) - y - tJ_A(d)g\|^2_{R^{-1}} + \frac{t^2}{2\tau}\|g\|^2, \\ \phi'(t) &= -\|g\|^2 + t\left(\|J_A(d)g\|^2_{R^{-1}} + \tau^{-1}\|g\|^2\right), \\ t_* &= \frac{\|g\|^2}{\|J_A(d)g\|^2_{R^{-1}} + \tau^{-1}\|g\|^2}, \qquad \delta^{\mathrm{sc}} = -t_* g. \end{aligned} \tag{E.1}$$

Set $t_* = 0$ if $g = 0$. Writing $p = \widehat{m}_\tau(x) - d$, the ordered candidate list in Equation (E.3) is

$$(u_1, \ldots, u_6) = \left(d + p,\ d + \tfrac{1}{4}p,\ d + \tfrac{1}{16}p,\ d + \tfrac{1}{64}p,\ d + \delta^{\mathrm{sc}},\ d + \tfrac{1}{4}\delta^{\mathrm{sc}}\right), \qquad u_7 = d. \tag{E.2}$$

The selected endpoint is

$$\Psi_\tau(x) = u_{j_*}, \qquad j_* = \min\big(\{j \le 6 : E(u_j) < E(d)\} \cup \{7\}\big). \tag{E.3}$$

Candidates are checked in this order against the strict inequality $E(u_j) < E(d)$; accepted states are not evaluated again. If all candidates fail, the output is d. This outer acceptance rule is separate from the line search inside the optional L-BFGS center correction (Liu & Nocedal, 1989). The resulting map satisfies $E(\Psi_\tau(x)) \le E(d)$ using the same denoiser input. The acceptance displacement is included in the map approximation bound of Appendix B.3.

Algorithm 1 at the start of Section 3 specifies the single-particle update. The rematching objective and Gaussian increment both use the next variance τ_{k+1}. Stochastic evaluation fixes its independent Gaussian draws during each solve and draws fresh sampling noise after the solve. Analytic evaluation minimizes $V_{\tau_{k+1},z}$; setting the optimization budget to zero keeps the center z. The last level returns $\mathsf{D}_{\tau_K}(x_K)$ without correction or additional noise. Appendix F specifies particle schedules and returned reconstructions. Appendix D.2 derives the finite objective and its analytic approximation. For a center $\widehat{u}$, the realized constant control is $(\widehat{u} - z)/t$. Writing $r = \nabla \mathcal{L}^U_{t,z}(\widehat{u})$ makes its implicit form explicit:

$$\widehat{u} = z - t\mathbb{E}_\xi \nabla U(\widehat{u} + \sqrt{t}\xi) + tr, \qquad X' = \widehat{u} + \sqrt{t}\xi'. \tag{E.4}$$

Here ξ' is independent of the draws used to compute $\widehat{u}$; the residual includes finite optimization and, when relevant, Monte Carlo error.

E.2 Statistical approximation of the observation response

Statistical linear regression represents a nonlinear observation by an affine response and residual covariance (Särkkä & Svensson, 2023, Section 9.4). The HDR configuration replaces the observation response in the same conditional-moment solve by averages over four operator probes, keeping the denoiser at the source (x, τ). For two antithetic pairs $e_1, -e_1, e_2, -e_2$, define

$$\begin{aligned} \overline{A} &= \tfrac{1}{4}\sum_{j=1}^{2}\big[A(d + e_j) + A(d - e_j)\big], \\ \overline{H} &= \tfrac{1}{4}\sum_{j=1}^{2}\big[J_A(d + e_j) + J_A(d - e_j)\big], \\ b_j &= \tfrac{1}{2}\big[A(d + e_j) + A(d - e_j)\big] - \overline{A}, \qquad \widehat{\Omega} = \operatorname{diag}\left(\tfrac{1}{2}\sum_{j=1}^{2} b_j \odot b_j\right). \end{aligned} \tag{E.5}$$

The numerical response is then

$$(R + \widehat{\Omega} + \overline{H}\widehat{C}\overline{H}^\top)w = y - \overline{A}, \qquad \widehat{m}_\tau(x) = d + \widehat{C}\overline{H}^\top w. \tag{E.6}$$

The symbol $\odot$ denotes componentwise multiplication. The even probe responses provide a diagonal estimate of the nonlinear variation unresolved by the affine response; this is a numerical approximation to observation uncertainty. For affine A, antithetic symmetry gives $\overline{A} = A(d)$, $\overline{H} = J_A$, and $\widehat{\Omega} = 0$ exactly. Thus the same endpoint construction covers both observation approximations.

The implemented probe width is obtained without additional denoiser calls. With n signal coordinates, define

$$c = \mathrm{clip}\left(\tau - \frac{\|x - d\|^2}{n},\, 10^{-4}\tau,\, \tau\right). \tag{E.7}$$

Two stratified standard-normal probes are scaled by $\sqrt{c}$ and paired with their negatives. This width is a state-dependent approximation; the conditional-variance identity holds after averaging over the prior noisy law, not pointwise. The benchmark activates these operator probes through its nonlinearity diagnostic; otherwise it uses the local Jacobian response. The characteristic integration in Equation (8) is identical in both cases.

F Benchmark Protocol

The benchmark compares reconstruction quality on FFHQ inverse problems using a shared pretrained prior (Karras et al., 2019; Zhang et al., 2025). PTM and DAPS use the same saved observations. SITCOM uses the same image cohort and task definitions with independent observation realizations. Each reported cohort uses one fixed configuration and a saved inclusion manifest.

DPS, DDRM, DDNM, DiffPIR, and FPS-SMC reproduce Table 1 of Zhang et al. (2025), evaluated on 100 validation images per dataset with observation-noise standard deviation .05. ImageNet DAPS uses that source; ImageNet SITCOM reproduces Table 1 of Alkhouri et al. (2025). These published aggregates follow their source evaluation protocols. Unreported entries remain blank. The DAPS source evaluates normalized images using PIQ, with replace-pooling enabled for LPIPS; phase retrieval selects the best of four samples. FID (Heusel et al., 2017) uses the 100 reconstructed and reference images in each reported cohort. For direct runs, 8-bit reconstructions and their paired references are resized to 299×299 by bilinear interpolation, normalized with ImageNet statistics, and encoded by torchvision's pretrained Inception-v3 with its classifier removed; PIQ computes FID from these features, following the released DAPS evaluator. The published SITCOM table reports running time as its fourth metric, not FID; we retain only its PSNR, SSIM, and LPIPS. FFHQ SITCOM FID is computed from the last-step outputs of the direct reruns.

For published baseline costs, we translate the released schedules into the same evaluation convention. DPS uses 1,000 forward/backward pairs (four complete runs for phase retrieval); DDRM, DDNM, and DiffPIR use 20, 100, and 100 denoiser forwards. FPS-SMC uses 20 particles and 1,000 time indices; initialization occupies the first index and the released loop performs 999 neural updates per particle, giving 19,980 denoiser forwards. For the cited configurations, DDRM uses 82 spectral-factor applications (one $U^\top$, 21 V, and 60 $V^\top$ calls), DDNM uses 100 nullspace corrections, and DiffPIR uses 99 nonterminal observation solves. These counts exclude operator setup and are not A_{eff} equivalents. FPS-SMC's complete observation count is unreported. ImageNet DAPS costs follow its published schedules, which coincide with the FFHQ schedules counted below. SITCOM's published ImageNet results do not include denoiser/observation evaluation ledgers; those costs are unreported.

PTM and DAPS use floating-point reconstructions, clipped to the prescribed image range only for PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018) evaluation. Ground truth is excluded from PTM sampling and final particle selection. FFHQ configuration comparisons use development images that are included in the reported 100-image cohort; the new solver settings were screened on the first 16 images before completion on the remaining 84. Baseline output selection is specified below. SITCOM metrics use its last sampling step, rather than the intermediate step with the highest ground-truth PSNR. Its saved reconstructions are evaluated with the same metric implementations; its preprocessed reference images differ from PTM's by at most one 8-bit intensity level.

For each run, record denoiser and observation forwards, Jacobian–vector products, and vector–Jacobian products separately. We report

$$D_{\mathrm{eff}} = D_{\mathrm{fwd}} + 2D_{\mathrm{jvp}} + 2D_{\mathrm{vjp}}, \qquad A_{\mathrm{eff}} = A_{\mathrm{fwd}} + 2A_{\mathrm{jvp}} + 2A_{\mathrm{vjp}}. \tag{F.1}$$

Each Jacobian–vector or vector–Jacobian product is charged as two forward-equivalent evaluations. All particles, inner objective evaluations, and line searches are included. Diagnostic probes, measurement generation, and final metric evaluation are counted separately from sampling. Operator-specific spectral or pseudoinverse operations are documented separately.

Table 4: **Denoiser and observation costs.** Forward and effective evaluations occupy separate rows and include every particle and solve. SPEC., PROJ., and SOLVE count spectral-factor applications, projections, and solves, excluding setup; they are not equivalent autodiff evaluations. A tag without a number has no reported count; – denotes unreported or inapplicable.

Prior	Method	Count	SR	GB	MB	RI	BI	NL	PR	HDR
Denoiser										
Both	DPS	D_{fwd}	1k	1k	1k	1k	1k	1k	4k	1k
		D_{eff}	3k	3k	3k	3k	3k	3k	12k	3k
	DDRM	D_{fwd}	20	20	–	–	20	–	–	–
		D_{eff}	20	20	–	–	20	–	–	–
	DDNM	D_{fwd}	100	100	–	100	100	–	–	–
		D_{eff}	100	100	–	100	100	–	–	–
	DiffPIR	D_{fwd}	100	100	100	–	–	–	–	–
		D_{eff}	100	100	100	–	–	–	–	–
	FPS-SMC	D_{fwd}	19,980	19,980	19,980	19,980	19,980	–	–	–
		D_{eff}	19,980	19,980	19,980	19,980	19,980	–	–	–
FFHQ	DAPS	D_{fwd}	1k	1k	1k	1k	1k	4k	16k	4k
		D_{eff}	1k	1k	1k	1k	1k	4k	16k	4k
	SITCOM	D_{fwd}	455	483	481	508	532	554	1,877	572
		D_{eff}	1,352	1,439	1,434	1,517	1,594	1,660	5,590	1,716
	PTM	D_{fwd}	20	20	20	20	328	236	955	331
		D_{eff}	176	196	200	200	832	875	2,956	999
ImageNet	DAPS	D_{fwd}	1k	1k	1k	1k	1k	4k	16k	4k
		D_{eff}	1k	1k	1k	1k	1k	4k	16k	4k
	SITCOM	D_{fwd}	–	–	–	–	–	–	–	–
		D_{eff}	–	–	–	–	–	–	–	–
	PTM	D_{fwd}	20	20	64	20	296	557	979	331
		D_{eff}	179	342	654	203	750	1,962	2,999	999
Observation										
Both	DPS	A_{fwd}	1k	1k	1k	1k	1k	1k	4k	1k
		A_{eff}	3k	3k	3k	3k	3k	3k	12k	3k
	DDRM	Ops.	82 SPEC.	82 SPEC.	–	–	82 SPEC.	–	–	–
	DDNM	Ops.	100 PROJ.	100 PROJ.	–	100 PROJ.	100 PROJ.	–	–	–
	DiffPIR	Ops.	99 SOLVE	99 SOLVE	99 SOLVE	–	–	–	–	–
	FPS-SMC	Ops.	SPEC.	SPEC.	SPEC.	PROJ.	PROJ.	–	–	–
FFHQ	DAPS	A_{fwd}	20k	20k	20k	20k	20k	80k	160k	40k
		A_{eff}	60k	60k	60k	60k	60k	240k	480k	120k
	SITCOM	A_{fwd}	455	483	481	528	552	554	1,877	572
		A_{eff}	1,349	1,437	1,431	1,574	1,651	1,657	5,580	1,712
	PTM	A_{fwd}	174	184	186	186	583	2,187	25,612	3,333
		A_{eff}	753	1,173	1,351	737	2,174	11,554	90,995	11,948
ImageNet	DAPS	A_{fwd}	20k	20k	20k	20k	20k	80k	160k	40k
		A_{eff}	60k	60k	60k	60k	60k	240k	480k	120k
	SITCOM	A_{fwd}	–	–	–	–	–	–	–	–
		A_{eff}	–	–	–	–	–	–	–	–
	PTM	A_{fwd}	175	257	611	187	619	3,840	25,730	3,362
		A_{eff}	763	1,624	4,619	745	2,272	20,514	97,230	12,062

Particle selection and averaging define a returned reconstruction. Their image-quality metrics are distinct from distributional posterior accuracy of the single-particle transition analyzed in the Method.

PTM configurations. Each task uses one fixed configuration for all 100 images and all reported metrics. Super-resolution, Gaussian blur, motion blur, and random inpainting share 20 levels with analytic rematching: conjugate gradients solve the likelihood-plus-distance objective, with at most

16 iterations and relative residual tolerance .05. The source solve uses at most six GMRES iterations, stops no earlier than the second iteration, and uses the same tolerance; cached Jacobian products avoid a repeated output evaluation. These solvers use automatic differentiation of the observation operator, without operator-specific spectral formulas. For ImageNet Gaussian blur, the source solve uses at most 12 GMRES iterations and relative residual tolerance .01, retaining 20 levels and the original rematching settings; all four metrics use the same 100 paired reconstructions.

Box inpainting and nonlinear blur use 32 and 40 levels. A directional predicted-decrease test allocates sampled rematching, with thresholds .1 and .2, respectively; other transitions use the analytic objective. One independent pilot draw is fixed within each sampled solve, with at most four iterations for box inpainting and 16 for nonlinear blur. The analytic budgets are 16 and 64 iterations. The threshold allocates objective evaluations by predicted decrease; accuracy is analyzed through the residual in Equation (16). Phase retrieval uses 148 levels and analytic rematching with at most 16 quasi-Newton iterations, reusing observation responses for the initial inverse metric; its inner source preconditioner uses relative residual tolerance .1. HDR retains the 148-level Gaussian-rematching configuration. All innovations are isotropic; sampled-objective pilots are independent of the committed innovation. Evaluation policies and budgets are fixed per task.

Super-resolution, Gaussian blur, motion blur, random inpainting, and nonlinear blur use a single particle. Nonlinear blur starts at standard deviation $\sqrt{\tau} = 3$; other tasks start at 12, and all end at .01. Box inpainting uses two particles for 20 levels down to $\sqrt{\tau} = 1$, then one for 12 levels. HDR uses four particles for 36 levels down to 3 and 24 more down to 1, then one for 12 levels down to .5 and 76 down to .01. At each merge, retain the mean of all candidates if its observation energy does not exceed the best candidate's energy; otherwise retain that candidate.

Phase retrieval starts with 16 particles, retains eight at $\sqrt{\tau} = 3$, and merges to one at .5, with the same 36/24/12/76 level allocation as HDR. FFHQ ranks candidates by observation energy. At the final merge, it averages candidates whose pairwise PSNR to the lowest-energy candidate is at least 20 dB, using the native $[-1, 1]$ range; it retains this mean only if its energy does not exceed that candidate's. ImageNet instead selects one candidate using the denoising-residual score below. Neither selection uses a reference image.

F.1 Direct baseline runs and budget curves

All eight DAPS tasks in Table 2 contain completed runs on FFHQ images 49000–49099. DAPS uses its released sampler and task configurations (Zhang et al., 2025): 200 levels with five prior evaluations per level for the five linear tasks, and 400 levels with ten evaluations per level for nonlinear blur, HDR, and phase retrieval. Each level uses 100 Langevin updates, or 200 for nonlinear blur. Instrumentation records every denoiser and observation forward and backward pass. Phase retrieval selects the highest ground-truth PSNR among four complete runs and charges their summed cost; one run instead gives 26.13 dB and .260 LPIPS at $D_{\text{eff}} = 4{,}000$.

SITCOM uses its released noisy-observation algorithm (Alkhouri et al., 2025), with 20 sampling steps, at most 30 inner iterations, learning rate .02, and the released discrepancy stopping rule. Phase retrieval reports the highest ground-truth PSNR among the last outputs of four seeds, 42–45, charging all four runs. Other tasks return the last output of one run. Measurement noise has standard deviation .05; random inpainting uses the benchmark mask definition and HDR uses the operator recovered from the released repository's history.

SITCOM's denoiser forwards and backwards are recorded for every image. Additional instrumentation on four images per task verifies the observation-call identities and attention-checkpoint replays. These identities are applied to the 100-image ledgers, including the extra observation forward and backward at each inpainting step. The reported denoiser cost also charges attention replay using its measured convolution/matrix-product FLOP ratio to a full denoiser forward. This is a calibrated correction to the evaluation-count convention, not an exact count of every arithmetic operation or a new instrumented run of all 800 reconstructions.

Figure 4 uses six completed PTM cohorts: two tasks, each with 100 images at six, ten, and twenty levels. Only the schedule length changes; all use one particle, six GMRES iterations, at most 16 L-BFGS iterations for the analytic rematching objective V, the saved benchmark observations, and seed 200010. The budget curves use L-BFGS for analytic rematching; the table configurations are

specified above. Every point is the terminal output of a complete run from noise. The twenty-level outputs are reused; the 400 shorter runs are independently completed, and all 600 floating-point outputs are rescored with the same evaluator. Curves connect all three PTM settings in budget order, including the twenty-level LPIPS points that are dominated by ten levels. DAPS and SITCOM points use the direct reruns above, with SITCOM's audited denoiser cost. DPS, DDNM, DiffPIR, and FPS-SMC points reproduce the FFHQ block of Table 1 in Zhang et al. (2025). Their denoiser counts are $3{,}000$, 100, 100, and $19{,}980$, respectively: DPS includes a forward and backward per step, whereas the other three use forwards alone; FPS-SMC includes all 20 particles. The latter also require observation-specific projections or linear solves; the denoiser-cost plot does not assign these operations zero cost.

F.2 Additional image prior

Table 2 evaluates all eight tasks on 100 ImageNet images at 256×256 resolution with the released unconditional ImageNet prior, sampler seed 200010, measurement seed 43, and observation-noise standard deviation $.05$. Metrics use the final returned reconstruction. Five tasks retain the FFHQ configurations, including 148-level Gaussian rematching for HDR. Motion and nonlinear blur use 64 and 80 levels; phase retrieval retains 148 levels and allows 32 quasi-Newton iterations and 64 objective evaluations. These adjustments were fixed using separate development images before the 100-image runs. Box inpainting evaluates independent particle rows separately to fit memory, preserving the objective and random draws. Published baseline means follow their source protocols and are not paired reruns.

ImageNet phase retrieval ranks particles by observation energy plus a denoising-residual score. For each candidate z, the latter sums $\|D_{s^2}(z + s\xi) - z\|^2/s^2$ over six logarithmically spaced standard deviations $s \in [.05, 2]$ and four Gaussian draws per scale. The same draws are used across candidates, and the two scores are balanced by their empirical standard deviations across candidates. This selection uses no ground truth; all 576 additional denoiser evaluations are included in the cost. The four quality metrics account for the 180° ambiguity by aligning the final reconstruction to its reference. This alignment is applied during evaluation, after particle selection.

F.3 Recorded reconstruction progress

Figure 5 restores recorded runs on FFHQ image 3 for phase retrieval and image 90 for box inpainting, with seed 200010 and paired benchmark observations. Phase image 3 was selected for a large historical reconstruction gap in the eight-image development cohort. For box inpainting, images 90, 39, and 42 were rerun and image 90 had the largest PTM–DAPS gap with analytic rematching. These selected examples illustrate progress, not average gains.

PTM uses 148 levels with a 16-to-8-to-1 particle schedule for phase and 48 levels with an eight-to-one schedule for box inpainting. The analytic center correction uses V; its ablation retains the same source map and keyed innovations but uses Gaussian rematching. DAPS uses 400 outer levels and ten prior evaluations per level for phase, and 200 outer levels and five evaluations per level for box, with 100 Langevin updates per level. The phase display is one DAPS run, without ground-truth selection across restarts.

Frames are recorded nearest 5%, 25%, 50%, and 100% of each run's measured D_{eff}, with terminal reconstructions nearest the central labels. PTM records the clean estimate before rematching; DAPS records its clean estimate after the inner update. For phase, saved pruning indices track the representative retained by the final merge: initial particle 10 for analytic rematching and particle 1 for Gaussian rematching, using zero-based indices. The analytic run averages initial particles 1, 9, 8, 13, and 10; the Gaussian run rejects averaging and retains particle 1. Earlier frames therefore follow a surviving contributor, while the final frame is the actual returned reconstruction. Box inpainting shows the first live particle before its all-particle average. Particle selection uses no ground truth, and costs include every particle.

The outer columns show the clean image, its noise-free measurement, and the noisy observation. Phase amplitudes use the released DAPS linear-contrast display convention, with shared clean-measurement statistics for both measurement images. Metrics are computed from the unmodified tensors. Recording adds no denoiser or observation evaluations.

F.4 CONTROLLED DISTRIBUTIONS

The Gaussian objective controls use zero-mean priors with covariance $C_0 = I$ or $O\,\mathrm{diag}(1, .04)O^\top$, where O rotates by .6 radians. Observations use $H = (1, .2)$ with $y = .5$, or $H = \mathrm{diag}(1, .4)$ with $y = (.5, -.3)^\top$, and $R = r^2 I$ for $r \in \{.2, 1\}$. For $K = H^\top R^{-1} H$, $h = H^\top R^{-1} y$, and likelihood weight $\gamma \in \{.25, .5, .75, 1\}$, the diagnostic map is

$$S_\gamma = (C_0^{-1} + \gamma K)^{-1}, \quad p_\gamma = \gamma S_\gamma h, \quad \Psi_\tau^\gamma(x) = p_\gamma + S_\gamma^{1/2}(S_\gamma + \tau I)^{-1/2}(x - p_\gamma). \tag{F.2}$$

All combinations use 10 or 20 geometrically spaced variances from 144 to 10^{-4} and start from $\mathcal{N}(0, 144I)$; the target has $\gamma = 1$. With $Q_t = C_0(C_0 + tI)^{-1}$, the three rematching centers are z, $(I + tK)^{-1}(z + th)$, and $(I + tQ_t^\top K Q_t)^{-1}(z + tQ_t^\top h)$. Each transition adds independent $\mathcal{N}(0, tI)$ noise at the next variance; the final map receives no further noise. Affine moment propagation gives the exact output KL. The objective comparison fixes $z = p_1 + S_1^{1/2}(1, -1)^\top$ and uses 49 logarithmically spaced variances in $[10^{-5}, 10]$ across the eight geometries. These closed-form controls isolate the map and objective choices without Monte Carlo or optimization error.

Figure 3(c) tests the denoised-energy surrogate against the exact compatible kernel. At each transition of the Gaussian-rematching trajectory, hold the incoming endpoint law $\rho = \mathcal{N}(m_\rho, C_\rho)$ and next map fixed. Its required input law is $\mathcal{N}(m_q, C_q) = (\Psi_t^\gamma)_\#^{-1}\pi^y$. The minimum-relative-entropy reference kernel is

$$K_*(z, \cdot) = \mathcal{N}\big(m_q + t^{-1}B(z - m_\rho), B\big), \qquad B + t^{-2}BC_\rho B = C_q, \quad B \succ 0. \tag{F.3}$$

We solve this positive-definite matrix equation exactly and compare Gaussian, Analytic, and Denoiser centers at the same incoming law. The plotted decrease is $\mathbb{E}_\rho[\mathrm{KL}(G_z \| K_*(z, \cdot)) - \mathrm{KL}(G_{u(z)} \| K_*(z, \cdot))]$, so positive values indicate descent of the ideal objective in Equation (14). These candidate interventions do not alter the common reference trajectory. The known posterior determines the diagnostic reference kernel only. The panel uses $\gamma = .25$ and 20 levels across all eight geometries; curves show their mean and bands their interquartile range, not a confidence interval.

For the non-Gaussian controls, both coordinates independently follow either a two-component mixture with means $(-1.3, 1.3)$, standard deviations $(.35, .45)$, and weights $(.4, .6)$, or a three-component mixture with means $(-1.8, 0, 1.8)$, standard deviations $(.3, .45, .4)$, and weights $(.25, .35, .4)$. Each prior also has a Gaussian control with the same mean and variance. We observe the first coordinate at $y = .25$ with noise standard deviation .3 or .8. The map at likelihood weight γ is the coordinatewise probability-flow map

$$\Psi_\tau^\gamma(x) = \mathrm{CDF}_{\pi_\gamma}^{-1}\big(\mathrm{CDF}_{\pi_\gamma * \mathcal{N}(0,\tau)}(x)\big), \qquad \pi_\gamma \propto p_0 e^{-\gamma E}. \tag{F.4}$$

The CDFs refer to the corresponding one-dimensional marginals; the unobserved coordinate uses the prior marginal. This construction holds the map mechanism fixed between Gaussian and non-Gaussian problems; it uses known distributions to isolate rematching, rather than approximating the map as in the image sampler.

All combinations of two prior shapes, two distribution families, two observation-noise levels, two schedules, four conditioning weights, three rematching rules, and two seeds are retained: 384 runs with 8,192 samples each. The schedules descend geometrically from standard deviation 12 to .01 in 10 or 20 levels. Analytic rematching uses its exact scalar proximal solution. Denoiser rematching uses one independent pilot draw per sample, fixed within an L-BFGS solve with at most 16 iterations and 32 objective evaluations. The transition innovation is independent of the pilot; no rematching follows the final map. The two seeds are 2609262 and 2609263.

Output KL is computed against exact target bin probabilities on a two-dimensional histogram with 64 interior bins per coordinate on $[-5, 5]$ and additional tail bins; no samples are discarded and no density smoothing is applied. The iid reference uses the same sample count and bins and is not subtracted from the reported KL. Checks with 32 and 128 bins retain the qualitative comparison. Table 1 compares both prior families at $\gamma = .25$; each entry averages eight problem–schedule pairs and two seeds. Figure 3(b) retains all four conditioning strengths for the mixture priors, with means and interquartile ranges over the eight seed-averaged pairs. The illustrated two-component problem uses observation noise .3, 20 levels, $\gamma = .25$, and the first seed; this case was fixed before running the sweep. A common 27-degree rotation is used for display only. CDF inversion errors are below 2.6×10^{-8}; an independent probability-flow ODE integration agrees with the maps to 2.5×10^{-7}, and the denoiser Jacobian agrees with the exact conditional-variance identity to 2.3×10^{-16}.

G Related work

Inverse reconstruction with pretrained priors. Plug-and-play methods separate measurement fidelity from denoising-based prior information (Venkatakrishnan et al., 2013); DiffPIR incorporates this construction into diffusion restoration (Zhu et al., 2023). DDRM exploits the spectral structure of linear observations (Kawar et al., 2022), while DDNM imposes range-space consistency and uses a diffusion prior to generate the null-space component (Wang et al., 2023). RePaint combines observation conditioning with forward–reverse resampling for inpainting (Lugmayr et al., 2022). These methods establish both pretrained-prior reconstruction and resampling as existing computational tools.

Guidance, moments, and numerical conditioning. DPS and ΠGDM approximate observation-dependent guidance from a prior denoiser (Chung et al., 2023; Song et al., 2023a). Manifold-constrained gradients and DiffStateGrad modify guidance to better respect the geometry of generated states (Chung et al., 2022; Zirvi et al., 2025). DDS combines denoising with Krylov-subspace data-consistency updates (Chung et al., 2024). TMPD uses denoiser derivatives to approximate conditional moments (Boys et al., 2024); posterior-covariance estimation further improves Gaussian conditioning (Peng et al., 2024). Our conditional-moment calculation builds on this moment-based line (Boys et al., 2024; Rozet et al., 2024). The endpoint equation identifies the additional finite-transport requirement: even exact local conditioning does not itself specify the transport map.

Posterior objectives and transition kernels. RED-Diff formulates inverse inference through a variational objective (Mardani et al., 2024); SITCOM develops constrained optimization within diffusion-based reconstruction (Alkhouri et al., 2025). For distributional inference, twisted diffusion sampling, MCGDiff, and filtering posterior sampling use sequential Monte Carlo to correct approximate proposals (Wu et al., 2023; Cardoso et al., 2024; Dou & Song, 2024). Plug-and-play split Gibbs sampling separates likelihood sampling from Bayesian denoising (Coeurdoux et al., 2024); DPnP analyzes an alternating proximal-consistency sampler and a denoising diffusion sampler (Xu & Chi, 2024). DAPS decouples successive noise levels through clean-state inference and annealing (Zhang et al., 2025). These constructions are directly related to our endpoint–rematching composition. PTM specifies compatibility between a transport map and its rematching kernel, then separates potential approximation and numerical error in the finite transition.

Flow-based inverse solvers and finite transport. FlowDPS adapts pretrained flow models by decomposing the velocity into clean and noise estimates and modifying them with likelihood guidance and stochasticity (Kim et al., 2025). FLAIR develops a variational inverse solver using flow-matching regularization (Erbach et al., 2025). Score-based SDEs connect learned scores, reverse diffusions, and probability-flow ODEs (Song et al., 2021); flow matching constructs transport velocities from conditional probability paths (Lipman et al., 2023). Consistency models, flow maps, and MeanFlow represent finite transport through endpoint consistency or average velocity (Song et al., 2023b; Boffi et al., 2025; Geng et al., 2025). These perspectives motivate our use of a posterior characteristic. We construct its numerical endpoint at inference time from a pretrained prior denoiser and the observation, and couple it to a diffusion through an explicit marginal-compatibility condition.

Action Matching recovers gradient dynamics from marginal samples; its entropic extension gives the variational drift objective for a prescribed diffusion coefficient (Neklyudov et al., 2023). PTM applies this characterization to input laws constrained by the chosen transport maps and the posterior target. The finite implementation restricts rematching to Gaussian transitions and approximates the required terminal potential using denoised observation energy. An analytic response approximation gives the observation-energy proximal step; its gradient error and finite optimization residual enter the transition analysis.

Adapting a generative prior to an observation is a likelihood tilt. Tilt Matching derives the evolution of transport velocities under reward tilting and regression objectives for learning them (Potaptchik et al., 2026). GLASS Flows recover inner flows from pretrained models to sample stochastic transition kernels without retraining (Holderrieth et al., 2026). Our formulation focuses on measurement-conditioned clean endpoints and their composition with isotropic rematching.