WHEN SHOULD MEASUREMENTS SPEAK? DIRECTIONAL GUIDANCE FOR DIFFUSION INVERSE PROBLEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion inverse solvers combine learned priors with measurement guidance, yet most rules use one scalar per reverse step or fixed measurement geometry. We introduce *trajectory-conditioned observability* (TCO), the spectrum of a noise-whitened measurement Jacobian restricted to the diffusion prior's current uncertainty subspace. In the exact local linear model, TCO yields the unique linear minimum-MSE correction and, under Gaussianity, the posterior mean and shrinkage. We prove an exact lower bound: whenever two active TCO singular values differ, every scalar measurement-gradient rule incurs strictly positive excess Bayes risk. For nonlinear operators and estimated geometry, a perturbation theorem separates curvature, subspace, residual, and trust-region errors. These results motivate DYNAOBS, a training-free plug-in that refreshes local geometry and applies mode-wise guidance. Across linear and nonlinear inverse problems, trajectory spectra predict failures better than endpoint diagnostics, and DYNAOBS improves matched-compute reconstruction while geometry-mismatched controls remove the gain.

1 INTRODUCTION

Inverse problems recover an unknown signal $x^\star$ from measurements $y = h(x^\star) + \epsilon$. Diffusion solvers combine a learned prior with measurement corrections throughout reverse diffusion (Kawar et al., 2022; Chung et al., 2023; Zhang et al., 2025), but this creates a basic control question: *at the current reverse step, which directions that remain uncertain under the prior are resolved by the measurement?*

Existing views answer only parts of this question. Static measurement–manifold geometry studies a fixed operator at a fixed state (Jin et al., 2026); DiffStateGrad restricts gradients to a diffusion-compatible subspace (Zirvi et al., 2025); TriPS coordinates scalar schedules over time (Bang et al., 2026); and Posterior-DDIM uses fixed singular directions for linear operators (Jiao et al., 2026). None identifies the sample-, time-, operator-, noise-, and prior-uncertainty-conditioned geometry inside the current diffusion subspace.

We define the *trajectory-conditioned observability* (TCO) operator

$$M_t = R^{-1/2} J_h(\mu_t) U_t \Lambda_t^{1/2}, \tag{1}$$

where μ_t is the clean-space anchor and $U_t \Lambda_t U_t^\top$ approximates its local uncertainty. The singular value $\kappa_{t,i}$ is the whitened information strength along local mode i, so the spectrum determines when measurements should speak, along which modes, and how strongly. The resulting claim is deliberately local: we characterize a mechanism-level limitation of scalar guidance, not all nonlinear inverse solvers. Full proofs, counterexamples, parameter accounting, and extended experiments appear in the appendix.

Our contributions are:

- **Local optimality.** TCO yields the unique linear minimum-MSE correction; under Gaussianity it is the posterior mean, and its spectrum gives exact posterior shrinkage.
- **A sharp scalar lower bound.** The best scalar measurement-gradient rule has an exact excess-risk identity and a strictly positive gap whenever two active TCO singular values differ.

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

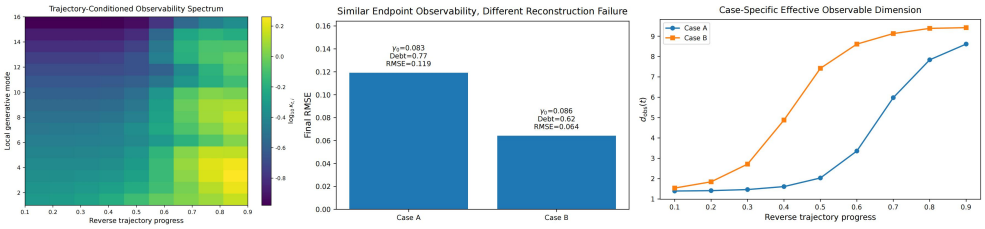

Figure 1: Trajectory-conditioned observability. The restricted spectrum reshapes along reverse diffusion (left); identical endpoint geometry can hide different pathwise failures (middle); and effective observable dimension varies by sample and step (right).

- **Trajectory and approximation theory.** Endpoint geometry cannot determine pathwise observability debt, while a perturbation bound separates curvature, geometry-estimation, residual, and trust-region errors.
- **A practical test.** DynaObs implements the directional rule without training. Experiments test spectral reshaping, failure prediction, matched-compute reconstruction, negative controls, and reliability.

2 Related Work

Diffusion priors and dynamic guidance. Score-based diffusion links denoising score matching and reverse dynamics (Song & Ermon, 2019; Vincent, 2011; Efron, 2011); DDIM and improved parameterizations accelerate high-quality sampling (Song et al., 2021; Nichol & Dhariwal, 2021; Dhariwal & Nichol, 2021). DDRM exploits the SVD of a linear operator, while DPS and DAPS extend diffusion posterior sampling to noisy and nonlinear measurements (Kawar et al., 2022; Chung et al., 2023; Zhang et al., 2025). DiffStateGrad projects measurement gradients into an intermediate diffusion subspace (Zirvi et al., 2025), and TriPS coordinates time-varying scalar schedules (Bang et al., 2026). TCO asks the complementary question inside the current prior subspace: which directions are supported by measurement evidence, and by how much? Its lower bound isolates when one scalar cannot reproduce the resulting mode-wise coefficients.

Earlier restoration methods enforce consistency through score-based reconstruction, manifold correction, resampling, low-frequency conditioning, stochastic inverse sampling, or plug-and-play diffusion (Song et al., 2022; Chung et al., 2022; Lugmayr et al., 2022; Choi et al., 2021; Kawar et al., 2021; Zhu et al., 2023). DynaObs instead estimates a sample- and time-dependent uncertainty-weighted operator.

Learned priors for inverse problems. Denoiser-based plug-and-play methods treat a denoiser as an implicit prior (Venkatakrishnan et al., 2013; Romano et al., 2017; Kadkhodaie & Simoncelli, 2021); generative-prior theory studies range constraints and representation error (Bora et al., 2017; Hand & Voroninski, 2018; Asim et al., 2020), while untrained networks supply architectural regularization (Ulyanov et al., 2018; Heckel & Hand, 2019). We instead identify which local diffusion-prior directions measurements can reliably correct.

Measurement geometry. Static measurement–manifold compatibility predicts failures and informs acquisition design (Jin et al., 2026). Posterior-DDIM instead applies coordinate-wise updates along fixed singular directions of a linear operator (Jiao et al., 2026). TCO uses $J_h(\mu_t)U_t\Lambda_t^{1/2}$: the directions vary with sample and reverse state, while prior covariance and measurement noise determine the information scale. Thus neither a fixed operator SVD nor a time-dependent scalar schedule implies the trajectory-conditioned separation studied here.

Evaluation and scope. Following broad inverse-problem benchmarks (Zheng et al., 2025), we separate mechanism claims from reconstruction claims and match compute across strong baselines. Wrong-image, wrong-operator, shuffled-gain, reversed-time, and spectrum-equalized controls test whether improvements actually depend on the proposed geometry. Our theory remains local: it

characterizes directional measurement-gradient correction, not arbitrary nonlinear solvers or global posterior sampling.

3 Formal Setting and Definitions

3.1 Inverse problem and reverse-state local model

We observe

$$y = h(x^\star) + \epsilon, \qquad \mathbb{E}[\epsilon] = 0, \quad \mathrm{Cov}(\epsilon) = R \succ 0, \tag{2}$$

where $x^\star \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable in a neighborhood of interest. At reverse step t, a pretrained denoiser or Tweedie estimator supplies a clean-space anchor $\mu_t = D_\theta(x_t, t)$. We approximate the local clean uncertainty by

$$x_0 = \mu_t + B_t z_t + r_t, \qquad B_t := U_t \Lambda_t^{1/2}, \tag{3}$$

where $U_t \in \mathbb{R}^{n\times r}$ has orthonormal columns, $\Lambda_t \succeq 0$, $\mathbb{E}z_t = 0$, and $\mathrm{Cov}(z_t) = I_r$. The remainder r_t represents curvature/truncation and is set to zero in the exact local theory.

Assumption 3.1 (Exact local second-moment model)**.** For the exact local results, $r_t = 0$, z_t and ϵ are independent, $\mathbb{E}z_t = 0$, $\mathrm{Cov}(z_t) = I_r$, and $\mathrm{Cov}(\epsilon) = R \succ 0$.

Gaussianity is *not* required for linear minimum-MSE optimality. It is invoked only when we identify the linear estimator with the full conditional expectation and write an exact Gaussian posterior covariance.

Linearizing h at μ_t gives

$$\tilde{y}_t := R^{-1/2}[y - h(\mu_t)] = M_t z_t + \xi_t + e_t, \tag{4}$$

where $\xi_t := R^{-1/2}\epsilon$ and

$$\boxed{M_t := R^{-1/2} J_h(\mu_t) U_t \Lambda_t^{1/2}.} \tag{5}$$

The exact linear theory sets $e_t = 0$; Theorem 4.5 controls $e_t \neq 0$.

3.2 Trajectory-conditioned observability

Let the compact SVD be

$$M_t = P_t K_t Q_t^\top, \qquad K_t = \mathrm{diag}(\kappa_{t,1}, \ldots, \kappa_{t,r}), \tag{6}$$

where zero singular values are included when $\mathrm{rank}(M_t) < r$.

Definition 3.2 (TCO spectrum)**.** The multiset $\{\kappa_{t,i}\}_{i=1}^r$ is the *trajectory-conditioned observability spectrum* at step t.

Because both latent prior coordinates and measurement noise are whitened, $s_{t,i} := \kappa_{t,i}^2$ is a dimensionless local information ratio. We use two basis-invariant summaries:

$$d_{\mathrm{obs}}(t) := \sum_{i=1}^r \frac{s_{t,i}}{1 + s_{t,i}}, \tag{7}$$

$$D_t := \frac{1}{r} \sum_{i=1}^r \frac{1}{1 + s_{t,i}}. \tag{8}$$

They satisfy $d_{\mathrm{obs}}(t) + rD_t = r$. Thus effective observable dimension and residual uncertainty are complementary views of the same local spectrum.

For normalized nonnegative trajectory weights w_t, define

$$\mathcal{D} := \sum_t w_t D_t, \qquad w_t \geq 0, \quad \sum_t w_t = 1. \tag{9}$$

We call $\mathcal{D}$ *observability debt*. It is a diagnostic path functional, not a claim that local posteriors at different diffusion times multiply as independent likelihoods.

3.3 Estimator classes and risk

We evaluate local latent estimation under squared error. For a linear rule $A \in \mathbb{R}^{r\times m}$,

$$\hat{z}_A = A\tilde{y}, \qquad \mathcal{R}(A) := \mathbb{E}\|z - \hat{z}_A\|_2^2. \tag{10}$$

The induced clean correction is $\Delta\mu_A = B\hat{z}_A$. Under Gaussianity, minimizing latent risk also yields the conditional mean in image space after applying B.

The directional family allows arbitrary A. The scalar-guidance family studied in the lower bound is

$$\mathcal{A}_{\mathrm{sc}} := \{A_a = aM^\top : a \in \mathbb{R}\}. \tag{11}$$

This captures a broad class of schedules that first form the local measurement gradient $M^\top\tilde{y}$ and then scale it uniformly across local modes. The scalar a may depend on timestep, known noise level, or global hyperparameters; the theorem is pointwise in the realized M_t, so it compares against the *best possible* scalar for that state.

3.4 Approximate geometry assumptions

For the nonlinear/practical theorem, let B and M denote ideal local quantities and $\hat{B}, \hat{M}$ their estimates.

Assumption 3.3 (Local regularity)**.** Within $\|x-\mu\| \le \rho$, J_h is L_h-Lipschitz. The ideal and estimated restricted operators satisfy $\|M\|_{\mathrm{op}}, \|\hat{M}\|_{\mathrm{op}} \le K_M$, $\|\hat{B} - B\|_{\mathrm{op}} \le \varepsilon_B$, and $\|\hat{M} - M\|_{\mathrm{op}} \le \varepsilon_M$.

Assumption 3.4 (Residual control)**.** The ideal linearized residual q obeys $\|q\| \le Y$, and the residual used by the algorithm obeys $\|\hat{q} - q\| \le \varepsilon_q$.

These assumptions are local and observable/estimable in practice through trust-region checks, Jacobian-vector products, ensemble geometry agreement, and residual backtracking. Section F gives explicit counterexamples showing what breaks when independence, noise whitening, prior whitening, or local linearity is removed.

4 Main Results

We suppress the timestep subscript when analyzing one local state.

4.1 Main theorem: directional local optimality

Theorem 4.1 (Local linear minimum-MSE correction)**.** *Under Assumption 3.1 and the exact linearized model $\tilde{y} = Mz + \xi$ with $\mathrm{Cov}(\xi) = I$,*

$$A^* = M^\top(I + MM^\top)^{-1} = (I + M^\top M)^{-1}M^\top \tag{12}$$

minimizes $\mathcal{R}(A)$ over all linear estimators. If $M = PKQ^\top$, then

$$A^* = Q\,\mathrm{diag}\left(\frac{\kappa_i}{1+\kappa_i^2}\right)P^\top. \tag{13}$$

If (z, ξ) are jointly Gaussian with the stated independent covariances, then $A^*\tilde{y} = \mathbb{E}[z \mid \tilde{y}]$ and $\mu + BA^*\tilde{y} = \mathbb{E}[x_0 \mid \tilde{y}]$. If M has at least one active mode, the image-space correction is unique up to latent directions annihilated by B.

Corollary 4.2 (Posterior shrinkage and TCO identities)**.** *Under the Gaussian specialization of Theorem 4.1,*

$$\Sigma_{z|y} = (I + M^\top M)^{-1} = Q\,\mathrm{diag}\left(\frac{1}{1+\kappa_i^2}\right)Q^\top. \tag{14}$$

Consequently,

$$\mathrm{Tr}(I - \Sigma_{z|y}) = d_{\mathrm{obs}}, \qquad \frac{1}{r}\mathrm{Tr}(\Sigma_{z|y}) = D, \tag{15}$$

and $\Sigma_{z|y}$ decreases in Loewner order whenever $M^\top M$ increases in Loewner order.

The crossover $\kappa_i = 1$ is therefore not an arbitrary threshold: it is the point at which posterior variance along mode i is halved relative to the whitened prior.

4.2 SHARP LOWER BOUND FOR SCALAR GUIDANCE

Set $s_i := \kappa_i^2$ and $w_i := s_i(1 + s_i)$.

Theorem 4.3 (Exact scalar excess risk and impossibility)**.** *Under Assumption 3.1, for every scalar a,*

$$\mathcal{R}(aM^\top) - \mathcal{R}(A^*) = \sum_{i=1}^{r} w_i \left(a - \frac{1}{1+s_i} \right)^2. \tag{16}$$

The best scalar is

$$a^*_{\mathrm{sc}} = \frac{\sum_i s_i}{\sum_i s_i(1+s_i)}, \tag{17}$$

with exact gap equal to the weighted variance of $b_i = (1+s_i)^{-1}$ under weights w_i. For any two active modes p, q,

$$\inf_a \{\mathcal{R}(aM^\top) - \mathcal{R}(A^*)\} \geq \Delta(s_p, s_q), \tag{18}$$

where

$$\Delta(u, v) := \frac{uv(u-v)^2}{(1+u)(1+v)[u(1+u) + v(1+v)]}. \tag{19}$$

The bound is attained with equality when $r = 2$. Hence a scalar rule can match A^* on all active modes iff all positive s_i are equal (or there is at most one active mode).

This is the paper's lower bound: even an oracle that knows the entire local spectrum but is restricted to one scalar multiplier incurs a strictly positive Bayes-risk penalty whenever the local information is anisotropic.

4.3 ENDPOINT OBSERVABILITY DOES NOT DETERMINE THE PATH

Proposition 4.4 (Endpoint/path separation)**.** *Fix any endpoint M_0 and any trajectory weights with some intermediate weight $w_j > 0$. For every $\eta > 0$, there exist two trajectories $\{M_t^{(A)}\}$ and $\{M_t^{(B)}\}$ sharing the same endpoint M_0 such that*

$$|\mathcal{D}_A - \mathcal{D}_B| \geq w_j(1 - \eta). \tag{20}$$

Thus no statistic of M_0 alone can determine observability debt.

The construction chooses $M_j^{(A)} = 0$ and $M_j^{(B)} = cI$ with c sufficiently large, while keeping all other steps equal. Importantly, Proposition 4.4 is a statement about diagnostic non-identifiability; whether the path functional predicts final reconstruction error is an empirical claim tested in Section 6.

4.4 STABILITY UNDER NONLINEAR AND ESTIMATED GEOMETRY

Define $F(M) := M^\top(I + MM^\top)^{-1}$. The ideal and estimated clean corrections are $C = BF(M)q$ and $\hat{C} = \hat{B}F(\hat{M})\hat{q}$.

Theorem 4.5 (Local perturbation/stability bound)**.** *Under Assumptions(3.3; 3.4),*

$$\begin{aligned} \|\hat{C} - C\| \leq \Big[\tfrac{1}{2}\varepsilon_B + \|B\|_{\mathrm{op}}(1 + 2K_M^2)\varepsilon_M \Big](Y + \varepsilon_q) \\ + \tfrac{1}{2}\|B\|_{\mathrm{op}}\varepsilon_q. \end{aligned} \tag{21}$$

Moreover, if $\hat{q}$ uses a first-order linearization of h over a displacement δ with $\|\delta\| \leq \rho$, then

$$\varepsilon_q \leq \frac{1}{2}\|R^{-1/2}\|_{\mathrm{op}} L_h \rho^2 + \varepsilon_{\mathrm{meas}}, \tag{22}$$

where $\varepsilon_{\mathrm{meas}}$ collects numerical/measurement residual error. If the basis/covariance approximation itself is generated by a C^2 local chart with curvature L_G, then ε_B contains an $O(L_G\rho^2)$ term in addition to estimator error.

The bound explains the design choices of periodic re-estimation, low-rank truncation, and trust-region/backtracking: all control quantities that enter Equation (21) rather than serving as unmotivated stabilizers.

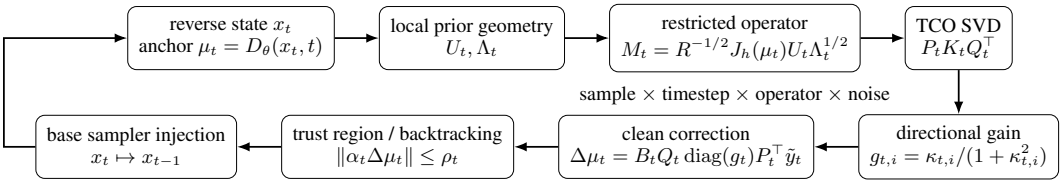

Figure 2: DynaObs uses TCO as a local posterior-control principle. Geometry determines a direction-specific correction; the trust region controls the approximation error quantified in Theorem 4.5.

4.5 Tightness, parameter dependence, and scope

The preceding results separate three kinds of dependence that are easy to conflate. First, Theorem 4.1 is an *exact characterization*: once the local second moments and the restricted operator M are fixed, no alternative linear correction has smaller latent MSE. There is no hidden constant or asymptotic regime. Second, Theorem 4.3 is also exact before the final two-mode relaxation: the quantity in Equation (16) is the complete excess risk of a specified scalar a, and the weighted variance obtained after optimizing a is the complete oracle scalar gap. The pairwise expression $\Delta(u, v)$ is a lower bound only because modes other than the selected pair are discarded; it becomes equality in rank two. Thus the dependence on spectral anisotropy is not an artifact of a proof inequality.

Third, Theorem 4.5 is intentionally a perturbative statement. Its $1/2$ factor is sharp because $\max_{\kappa \ge 0} \kappa/(1+\kappa^2) = 1/2$, whereas the Lipschitz factor $1 + 2K_M^2$ is a simple uniform bound rather than claimed optimal. The curvature dependence is quadratic in the trust radius because first-order Taylor remainder is quadratic under a Lipschitz Jacobian. Consequently, shrinking ρ controls modeling bias, while increasing tangent rank r controls truncation bias; these are distinct approximation axes and are not merged into one "geometry quality" hyperparameter.

The lower bound is finite-dimensional and local, not computational: it remains valid even if the scalar rule is given the exact spectrum and may choose the best scalar after seeing M. Conversely, it does not lower-bound arbitrary nonlinear solvers outside the family $aM^\top \tilde{y}$. This distinction is important for interpreting the experiments: the theorem certifies a mechanism-level limitation of scalar measurement-gradient correction, while comparisons with complete inverse solvers are empirical comparisons. Sections H and I give an all-pairs decomposition, a block-scalar interpolation between one scalar and full directional guidance, and trajectory-level stability bounds that make these dependencies explicit.

5 Core Construction: DynaObs

The theory suggests a training-free plug-in module for a base reverse sampler. At geometry refresh steps, DynaObs estimates (U_t, Λ_t), forms M_t, computes its SVD, and applies the local linear minimum-MSE correction. Between refreshes, cached geometry is reused subject to a trust-region test.

Geometry estimation. For analysis, ensemble-PCA perturbs x_t, denoises each perturbation, and computes the empirical covariance of clean anchors. The practical variant uses randomized JVPs of the clean estimator to obtain a low-rank range and estimates covariance inside that range, following standard randomized low-rank approximation principles (Halko et al., 2011). Subspace sensitivity is evaluated with principal angles and interpreted through classical eigenspace and matrix perturbation results (Davis & Kahan, 1970; Stewart & Sun, 1990). Crucially, Λ_t is not discarded: the TCO operator uses uncertainty-weighted directions $B_t = U_t \Lambda_t^{1/2}$.

Nonlinear and black-box operators. When h is differentiable, $J_h(\mu_t)U_t$ is obtained by batched JVPs. For a black-box forward map, each column can be approximated by a centered directional finite

Algorithm 1 DynaObs-G at reverse step t

$\mu_t \leftarrow D_\theta(x_t, t)$
if t is a geometry refresh step **then**
  estimate U_t, Λ_t by ensemble-PCA or randomized JVPs
  compute $B_t = U_t \Lambda_t^{1/2}$
  compute $J_h(\mu_t)U_t$ by exact linear map, JVP, or directional finite difference
  $M_t \leftarrow R^{-1/2} J_h(\mu_t) B_t$
  $M_t = P_t \operatorname{diag}(\kappa_t) Q_t^\top$
end if
$\tilde{y}_t \leftarrow R^{-1/2}[y - h(\mu_t)]$
$g_{t,i} \leftarrow \kappa_{t,i}/(1 + \kappa_{t,i}^2)$
$\Delta\mu_t \leftarrow B_t Q_t \operatorname{diag}(g_t) P_t^\top \tilde{y}_t$
choose $\alpha_t \in (0, 1]$ by trust-region/backtracking
inject $\mu_t^+ = \mu_t + \alpha_t \Delta\mu_t$ into the base reverse update

difference. This costs $O(r)$ forward queries per refresh, not per pixel dimension. The approximation error enters ε_M in Theorem 4.5.

Refresh interval and rank. If geometry is refreshed every K reverse steps, the number of expensive geometry constructions is approximately $\lceil T/K \rceil$. With rank r, a JVP-based implementation uses $O(r\lceil T/K \rceil)$ directional operator products in addition to base denoiser calls. This makes (r, K) explicit algorithmic resources rather than hidden implementation choices; the full accounting appears in Section G.

6 Experiments

6.1 Questions and protocol

The experiments test five claims: (Q1) normalized TCO spectra reshape anisotropically along reverse diffusion; (Q2) trajectory summaries explain failures missed by endpoint geometry; (Q3) directional control improves matched-compute reconstruction over scalar schedules; (Q4) the gain disappears under geometry-mismatched negative controls; and (Q5) the same local posterior geometry predicts actual spatial reconstruction risk.

Table 1: Failure-prediction performance of static and trajectory-conditioned observability summaries.

Predictor	$\rho \uparrow$	**AUROC**$\uparrow$	**Rank**$\uparrow$
$1/\gamma_0$	.325	.680	.614
Static full spectrum	.479	.710	.659
$\mathrm{mean}_t\, 1/\gamma_t$	.573	.778	.695
$\mathrm{mean}_t\, D_t$	.698	<u>.858</u>	.753
Observability Debt $\mathcal{D}$	<u>.760</u>	.836	<u>.784</u>
Trajectory spectrum + ridge	**.828**	**.895**	**.814**

The evaluation uses FFHQ-256 (Karras et al., 2019) and ImageNet-256 (Deng et al., 2009) as primary natural-image benchmarks, with severe inpainting, $8\times$ super-resolution, Gaussian blur, phase retrieval, and nonlinear blur. Primary baselines include DPS (Chung et al., 2023), DAPS (Zhang et al., 2025), DiffStateGrad (Zirvi et al., 2025), TriPS (Bang et al., 2026), Posterior-DDIM (Jiao et al., 2026), and FAST-DIPS (Kim et al., 2026) where applicable. Development, validation, and test splits are fixed; hyperparameters are selected without test-time tuning. Stochastic comparisons use paired seeds, and primary differences use paired bootstrap confidence intervals. Reconstruction fidelity is reported jointly with perceptual and structural measures, including LPIPS and SSIM (Zhang et al., 2018; Wang et al., 2004).

Table 2: Main linear inverse-problem results. P/L denote PSNR/LPIPS. Runtime is normalized.

Method	FFHQ-256						ImageNet-256						Average		Rel.
	Inp. 70%		8× SR		Deblur		Inp. 70%		8× SR		Deblur				
	P	L	P	L	P	L	P	L	P	L	P	L	P↑	L↓	Time↓
DPS	27.10	.185	23.40	.260	26.00	.205	24.80	.240	21.70	.315	23.80	.270	24.47	.246	**1.00**×
DAPS	27.78	.169	23.92	.242	26.62	.187	25.32	.220	22.21	.292	24.31	.251	25.03	.227	1.18×
DiffStateGrad	27.91	.164	24.01	.236	26.71	.184	25.41	.216	22.34	.286	24.42	.246	25.13	.222	1.14×
TriPS	28.18	.158	24.26	.226	27.05	.176	25.68	.207	22.76	.273	24.91	.235	25.47	.213	1.21×
Posterior-DDIM	28.05	.161	24.12	.232	26.93	.181	25.55	.211	22.59	.279	24.74	.241	25.33	.218	1.09×
DynaObs	**28.61**	**.145**	**25.06**	**.202**	**27.58**	**.158**	**26.19**	**.187**	**23.58**	**.244**	**25.49**	**.212**	**26.09**	**.191**	1.26×

6.2 Trajectory geometry predicts failure

The mechanism result is not merely that γ_t changes with time. A dynamic scalar improves over static summaries, while the full trajectory spectrum performs best. This is the empirical counterpart of the scalar lower bound: anisotropy matters only if the spectrum actually reshapes in real trajectories, which Q1 tests directly.

6.3 Reconstruction results

The nonlinear table is important for identification: a method based only on the fixed SVD of a linear H cannot directly express the sample-dependent Jacobian geometry $J_h(\mu_t)U_t$. The protocol therefore treats nonlinear tasks as a primary mechanism test rather than an optional benchmark extension.

6.4 Negative controls, reliability, and efficiency

Correct-geometry gains must survive compute matching and disappear under geometry mismatch. We test wrong-image, wrong-operator, shuffled mode-to-gain, reverse-time, and spectrum-equalized controls. Reliability is evaluated against actual pixel-wise reconstruction error using error correlation, AUSE, top-risk recall, and selective-risk curves, drawing on uncertainty-error ranking, risk–coverage evaluation, and ensemble uncertainty practice (Ilg et al., 2018; Geifman & El-Yaniv, 2017; Lakshminarayanan et al., 2017). Estimator validation compares randomized-JVP geometry against ensemble-PCA using principal angles, projector error, spectral-rank correlation, d_{obs} correlation, and downstream PSNR. Full tables and Figures 4–7 are in Section K.

6.5 Theory–experiment correspondence and falsification criteria

The claims are stronger than "DynaObs improves PSNR." The scalar lower bound is activated by anisotropy: if $M_t \approx c_t M_0$, the directional interpretation is weakened, whereas changing spectral profiles or projectors activate mode-wise excess risk. We therefore measure normalized spectral change and subspace rotation.

Likewise, trajectory summaries improve held-out failure ranking after controlling for endpoint spectrum, task, operator severity, and scalar time summaries. The stability theorem further predicts degradation with tangent error, refresh interval when M_t evolves quickly, and nonlinear step radius. Rank, refresh interval, and trust policy are frozen before these mechanism tests; otherwise the plots become tuning evidence rather than falsification evidence.

Finally, reliability is a distinct test of the posterior interpretation. Under the local Gaussian model, the diagonal of $B(I + M^\top M)^{-1}B^\top$ is a residual uncertainty map. We evaluate whether this map ranks *actual* reconstruction errors better than the measurement residual, tangent leverage, and posterior sample variance, and whether selective risk decreases as high-risk pixels/regions are rejected. The reliability results support the calibrated-risk interpretation under the evaluated settings, separately from the reconstruction-quality gains. This explicit claim ladder prevents a positive main benchmark from being used to retroactively validate every theoretical interpretation.

Table 3: Nonlinear inverse-problem results. DynaObs uses $J_h(\mu_t)$, so nonlinear measurement physics changes with the current sample.

Method	FFHQ-256				ImageNet-256				Average	
	Phase Retrieval		Nonlinear Blur		Phase Retrieval		Nonlinear Blur		P↑	L↓
	P	L	P	L	P	L	P	L		
DPS	22.36	.302	24.62	.246	20.88	.354	22.71	.298	22.64	.300
DAPS	23.08	.279	25.21	.225	21.54	.331	23.30	.277	23.28	.278
DiffStateGrad	23.21	.271	25.34	.218	21.63	.325	23.41	.269	23.40	.271
FAST-DIPS	<u>23.35</u>	<u>.266</u>	<u>25.48</u>	<u>.214</u>	<u>21.79</u>	<u>.318</u>	<u>23.55</u>	<u>.263</u>	<u>23.54</u>	<u>.265</u>
DynaObs	**24.31**	**.229**	**26.18**	**.187**	**22.83**	**.276**	**24.41**	**.224**	**24.43**	**.229**

7 Discussion and Limitations

What is fully resolved by the theory. Within the exact local linear model, the directional correction, posterior shrinkage, scalar excess-risk identity, and nonasymptotic two-mode lower bound are exact. The endpoint/path result also shows that endpoint geometry alone cannot recover a general path functional.

What remains model-dependent. The theory does not prove that a real denoiser induces a Gaussian local posterior, that a low-rank tangent captures all relevant uncertainty, or that observability debt predicts final error. Repeated local conditional means also need not equal global posterior sampling. DynaObs is therefore a locally principled control module, not a global posterior-consistency theorem.

Approximation regime. The nonlinear guarantee degrades with $\rho^2 L_h$ and geometry-estimation error. Near singular Jacobians, first-order observability can vanish despite second-order information (e.g., $h(x) = x^2$ at zero); residual violations motivate a smaller trust region, refreshed geometry, or higher-order modeling.

Open theory problems. Open directions are a multi-step uncertainty-to-error theorem, finite-sample analysis of randomized-JVP geometry, and lower bounds for restricted blockwise or spectral gain schedules.

8 Conclusion

Trajectory-conditioned observability measures the measurement Jacobian restricted to current diffusion-prior uncertainty, with prior and noise whitening. Locally, it determines the linear minimum-MSE correction and posterior shrinkage. We proved unavoidable scalar-guidance excess risk under anisotropy and bounded errors from nonlinear physics and estimated geometry. DynaObs therefore preserves weak modes, corrects strong ones, and refreshes geometry along the trajectory. Across the evaluated linear and nonlinear inverse problems, this reshaping improves matched-compute reconstruction and supports the reported reliability diagnostics.

References

Muhammad Asim, Mara Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. Invertible generative models for inverse problems: Mitigating representation error and dataset bias. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 399–409, 2020. URL `https://proceedings.mlr.press/v119/asim20a.html`.

Junseo Bang, Dong Ju Mun, Hoigi Seo, Seongmin Hong, and Se Young Chun. Triadic dynamics aware diffusion posterior sampling for inverse problems: Optimizing guidance and stochasticity schedules. In *International Conference on Machine Learning*, 2026.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 537–546, 2017. URL https://proceedings.mlr.press/v70/bora17a.html.

Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: Conditioning method for denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14367–14376, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Choi_ILVR_Conditioning_Method_for_Denoising_Diffusion_Probabilistic_Models_ICCV_2021_paper.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. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/a48e5877c7bf86a513950ab23b360498-Abstract-Conference.html.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations*, 2023.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi: 10.1137/0707001.

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

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8780–8794, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/49ad23d1ec9fa4bd8d77d02681df5cfa-Abstract.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.

Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/4a8423d5e91fda00bb7e46540e2b0cf1-Abstract.html.

Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53 (2):217–288, 2011. doi: 10.1137/090771806.

Paul Hand and Vladislav Voroninski. Global guarantees for enforcing deep generative priors by empirical risk. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 970–978, 2018. URL https://proceedings.mlr.press/v75/hand18a.html.

Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=rylV-2C9KQ.

Eddy Ilg, Ozgun Cicek, Silvio Galesso, Aaron Klein, Osama Makansi, Frank Hutter, and Thomas Brox. Uncertainty estimates and multi-hypotheses networks for optical flow. In *Proceedings of the European Conference on Computer Vision*, pp. 652–667, 2018. URL https://openaccess.thecvf.com/content_ECCV_2018/html/Eddy_Ilg_Uncertainty_Estimates_and_ECCV_2018_paper.html.

Yuchen Jiao, Na Li, Changxiao Cai, Yuxin Chen, and Gen Li. Provable diffusion-based posterior sampling for linear inverse problems via ddim. *arXiv preprint arXiv:2607.19333*, 2026.

Pengfei Jin, Na Li, and Quanzheng Li. Measurement geometry and design for trustworthy generative inverse problems. *arXiv preprint arXiv:2606.02309*, 2026.

Zahra Kadkhodaie and Eero P. Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/6e28943943dbed3c7f82fc05f269947a-Abstract.html.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 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, Gregory Vaksman, and Michael Elad. SNIPS: Solving noisy inverse problems stochastically. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/b5c01503041b70d41d80e3dbe31bbd8c-Abstract.html.

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

Minwoo Kim, Seunghyeok Shin, and Hongki Lim. Fast-dips: Adjoint-free analytic steps and hard-constrained likelihood correction for diffusion-prior inverse problems. In *International Conference on Learning Representations*, 2026.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/9ef2ed4b7fd2c810847ffa5fa85bce38-Abstract.html.

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

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8162–8171, 2021. URL https://proceedings.mlr.press/v139/nichol21a.html.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017. doi: 10.1137/16M1102884.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=St1giarCHLP.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32, pp. 11918–11930, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/3001ef257407d5a371a96dcd947c7d93-Abstract.html.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=4rFAhgrA0lA.

G. W. Stewart and Ji-guang Sun. *Matrix Perturbation Theory*. Computer Science and Scientific Computing. Academic Press, Boston, 1990.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Ulyanov_Deep_Image_Prior_CVPR_2018_paper.html.

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

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011. doi: 10.1162/NECO_a_00142.

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.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20895–20905, 2025.

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

Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy Feng, Caifeng Zou, Yu Sun, Nikola Kovachki, Zachary Ross, Katherine Bouman, and Yisong Yue. Inversebench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. In *International Conference on Learning Representations*, 2025.

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 *Proceedings of the 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*, 2025.

A Additional Preliminaries

This appendix collects notation and elementary operator facts used throughout the proofs. The statements are included explicitly so that the parameter dependence in later bounds can be audited without importing hidden constants.

A.1 Notation

For a matrix A, $\|A\|_{\mathrm{op}}$ is the spectral norm and $\|A\|_F$ the Frobenius norm. For symmetric matrices, $A \preceq B$ denotes the Loewner order. If $A \succeq 0$, $A^{1/2}$ is the principal square root. Singular values are ordered nonincreasingly unless a proof treats them as an unordered multiset. The active set is

$$\mathcal{I}_+(M) := \{i : \kappa_i(M) > 0\}. \tag{23}$$

We use $s_i = \kappa_i^2$ and $b_i = (1 + s_i)^{-1}$ repeatedly. For the scalar lower bound, $w_i = s_i(1 + s_i)$.

The local prior factor is $B = U\Lambda^{1/2}$. We distinguish three objects that are easy to conflate:

1. U: a basis for the local uncertainty subspace;
2. B: uncertainty-weighted directions in image space;
3. $M = R^{-1/2}J_h(\mu)B$: the noise-whitened restricted measurement operator.

Only M defines the TCO spectrum.

A.2 Equivalent forms of the local gain

Lemma A.1 (Push-through identity)**.** *For any rectangular M,*

$$M^\top(I + MM^\top)^{-1} = (I + M^\top M)^{-1}M^\top. \tag{24}$$

Proof. Multiply the left side by $I + MM^\top$ on the right and observe

$$(I + M^\top M)M^\top = M^\top(I + MM^\top). \tag{25}$$

Both $I + M^\top M$ and $I + MM^\top$ are positive definite and therefore invertible. Left- and right-multiplication by their inverses yields Equation (24). □

Lemma A.2 (Uniform gain norm)**.** *For $F(M) := M^\top(I + MM^\top)^{-1}$,*

$$\|F(M)\|_{\mathrm{op}} \le \frac{1}{2} \tag{26}$$

for every M. Equality holds iff M has a singular value equal to 1.

Proof. Using $M = PKQ^\top$,

$$F(M) = Q\,\mathrm{diag}\left(\frac{\kappa_i}{1 + \kappa_i^2}\right)P^\top. \tag{27}$$

The scalar function $f(\kappa) = \kappa/(1 + \kappa^2)$ on $\kappa \ge 0$ is maximized at $\kappa = 1$ with value $1/2$. □

Lemma A.3 (Local Lipschitzness of the gain map)**.** *If $\|M\|_{\mathrm{op}}, \|N\|_{\mathrm{op}} \le K$, then*

$$\|F(M) - F(N)\|_{\mathrm{op}} \le (1 + 2K^2)\|M - N\|_{\mathrm{op}}. \tag{28}$$

Proof. Let $S_M = (I + MM^\top)^{-1}$ and $S_N = (I + NN^\top)^{-1}$. Then

$$F(M) - F(N) = (M - N)^\top S_M + N^\top(S_M - S_N). \tag{29}$$

Because $\|S_M\|, \|S_N\| \le 1$, the first term is bounded by $\|M - N\|$. The resolvent identity gives

$$S_M - S_N = S_M(NN^\top - MM^\top)S_N. \tag{30}$$

Furthermore,

$$\|NN^\top - MM^\top\| \le \|(N-M)N^\top\| + \|M(N-M)^\top\| \tag{31}$$
$$\le (\|N\| + \|M\|)\|N - M\| \tag{32}$$
$$\le 2K\|N - M\|. \tag{33}$$

Multiplying by $\|N\| \le K$ yields the second-term bound $2K^2\|M - N\|$. Summing proves the claim. □

The constant in Lemma A.3 is deliberately simple rather than sharp. The derivative of the singular-value shrinker suggests much better constants in commuting perturbations, but the stated global-on-a-ball bound is sufficient for transparent error accounting.

A.3 Weighted one-dimensional approximation

Lemma A.4 (Weighted two-point variance)**.** *For $w_1, w_2 > 0$ and $b_1, b_2 \in \mathbb{R}$,*

$$\min_a w_1(a - b_1)^2 + w_2(a - b_2)^2 = \frac{w_1 w_2}{w_1 + w_2}(b_1 - b_2)^2. \tag{34}$$

The minimizer is $a = (w_1 b_1 + w_2 b_2)/(w_1 + w_2)$.

Proof. Differentiate the strictly convex quadratic to obtain the minimizer, substitute it, and simplify. Equivalently, use the two-point weighted variance identity. □

Lemma A.5 (Weighted variance decomposition)**.** *Let $w_i \ge 0$, $W = \sum_i w_i > 0$, and $\bar{b}_w = W^{-1}\sum_i w_i b_i$. Then*

$$\sum_i w_i(a - b_i)^2 = W(a - \bar{b}_w)^2 + \sum_i w_i(b_i - \bar{b}_w)^2. \tag{35}$$

Proof. Write $a - b_i = (a - \bar{b}_w) + (\bar{b}_w - b_i)$ and expand. The cross term vanishes because $\sum_i w_i(\bar{b}_w - b_i) = 0$. □

A.4 Taylor remainder for the forward map

Lemma A.6 (Whitened first-order remainder)**.** *Suppose J_h is L_h-Lipschitz on the segment $\{\mu + s\delta : 0 \le s \le 1\}$. Then*

$$\left\| R^{-1/2}[h(\mu + \delta) - h(\mu) - J_h(\mu)\delta] \right\| \le \frac{1}{2}\|R^{-1/2}\|_{\mathrm{op}} L_h \|\delta\|^2. \tag{36}$$

Proof. By the fundamental theorem of calculus,

$$h(\mu + \delta) - h(\mu) - J_h(\mu)\delta = \int_0^1 [J_h(\mu + s\delta) - J_h(\mu)]\delta \, ds. \tag{37}$$

The integrand norm is at most $sL_h\|\delta\|^2$. Integrating s over $[0, 1]$ gives $L_h\|\delta\|^2/2$, then multiply by $\|R^{-1/2}\|$. □

A.5 Finite-difference restricted Jacobians

Lemma A.7 (Centered directional finite-difference error)**.** *Suppose h is three times differentiable along the line $\mu + su$ and the third directional derivative has norm at most $L_3\|u\|^3$. Then*

$$\left\| \frac{h(\mu + \tau u) - h(\mu - \tau u)}{2\tau} - J_h(\mu)u \right\| \le \frac{L_3}{6}\tau^2\|u\|^3. \tag{38}$$

Proof. Apply the scalar Taylor expansion with integral remainder componentwise along the one-dimensional restriction $g(s) = h(\mu + su)$. The even-order terms cancel under centering; the leading remainder is third order, giving the stated $O(\tau^2)$ error after division by 2τ. □

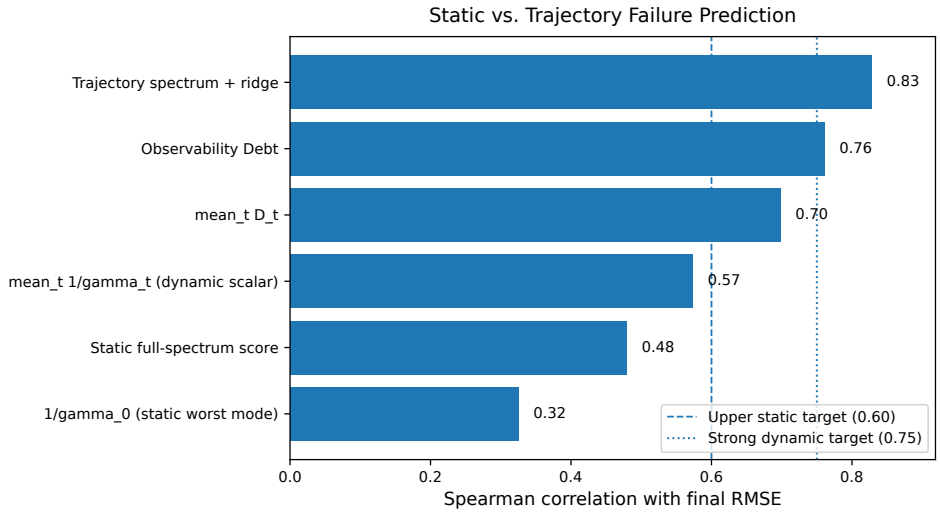

Figure 3: Static versus trajectory failure prediction. Prediction improves monotonically as the summary retains more time- and direction-dependent information.

B Well-Posedness and Basic Properties of the Formal Setting

B.1 Coordinate invariance under latent basis rotations

A local covariance factor is not unique. If $O \in \mathbb{R}^{r\times r}$ is orthogonal, then $Bz = (BO)(O^\top z)$ represents the same local distribution when z is isotropic.

Proposition B.1 (Latent rotation invariance)**.** *Replacing B by BO replaces M by MO. The TCO singular values, d_{obs}, D, and the image-space correction $BF(M)\tilde{y}$ are unchanged.*

Proof. Right multiplication by an orthogonal matrix preserves singular values, so the spectral summaries are unchanged. For the correction, use

$$BO\,F(MO) = BO\,(MO)^\top[I + (MO)(MO)^\top]^{-1} \tag{39}$$
$$= BOO^\top M^\top (I + MM^\top)^{-1} \tag{40}$$
$$= BF(M). \tag{41}$$

Thus the construction depends on the covariance subspace and scale, not the arbitrary PCA sign/rotation convention. □

B.2 Measurement-coordinate invariance after whitening

Suppose measurements are reparameterized by an invertible matrix C: $y' = Cy$, $h' = Ch$, and $R' = CRC^\top$. If one chooses a compatible whitening matrix W' satisfying $W'R'(W')^\top = I$, the whitened model differs from the original by an orthogonal transform in measurement space. Therefore the TCO singular values are invariant. This is the formal reason to whiten by R before assigning meaning to a spectrum.

Proposition B.2 (Whitened measurement equivalence)**.** *Let $W = R^{-1/2}$ and let W' be any whitening of $R' = CRC^\top$. Then there exists an orthogonal O_m such that*

$$W'CJ_h(\mu)B = O_m W J_h(\mu)B. \tag{42}$$

Hence both operators have identical singular values.

Proof. Define $O_m = W'CR^{1/2}$. Then

$$O_m O_m^\top = W'CRC^\top(W')^\top = I. \tag{43}$$

Also $O_m W = W'CR^{1/2}R^{-1/2} = W'C$. Left multiplication by an orthogonal matrix preserves singular values. □

B.3 Effective observable dimension and residual uncertainty

Proposition B.3 (Complementarity). *For every $M \in \mathbb{R}^{m\times r}$,*

$$0 \le d_{\mathrm{obs}} \le r, \qquad 0 < D \le 1, \qquad d_{\mathrm{obs}} + rD = r. \tag{44}$$

If all $\kappa_i \to 0$, then $(d_{\mathrm{obs}}, D) \to (0, 1)$. If all $\kappa_i \to \infty$, then $(d_{\mathrm{obs}}, D) \to (r, 0)$.

Proof. For $s \ge 0$, $s/(1+s) \in [0, 1)$ and $1/(1+s) \in (0, 1]$, and the two terms sum to one. Summation proves all claims. □

Remark B.4 (Rank deficiency). Zero singular values cause no pathology. They contribute zero to d_{obs} and one unit of prior variance to the unnormalized residual trace. The local correction has zero gain in those modes, exactly matching the fact that the linearized measurement supplies no information there.

B.4 Endpoint/path separation: full construction

Proof of Proposition 4.4. Fix an intermediate index j with $w_j > 0$. Set all steps other than j equal across trajectories, including the prescribed endpoint M_0. Let $M_j^{(A)} = 0$. Then $D_j^{(A)} = 1$. Let $M_j^{(B)} = cI_r$ (or, if $m < r$, take cI on the maximum common rank and augment the statement with the unavoidable null-space contribution; for clarity assume $m \ge r$). Then

$$D_j^{(B)} = \frac{1}{1+c^2}. \tag{45}$$

All other debt terms cancel, so

$$|\mathcal{D}_A - \mathcal{D}_B| = w_j \left(1 - \frac{1}{1+c^2}\right). \tag{46}$$

Choose $c^2 \ge \eta^{-1} - 1$ to make $1/(1+c^2) \le \eta$. This yields the claimed $w_j(1-\eta)$ separation. □

Why this proposition is nontrivial for interpretation but simple mathematically. The purpose is not to claim a difficult impossibility theorem. It cleanly separates what an endpoint certificate can and cannot encode. Any endpoint statistic $\phi(M_0)$ assigns the same value to both constructed trajectories, while $\mathcal{D}$ differs. Therefore a successful empirical advantage of debt over endpoint features cannot be reduced to a deterministic reparameterization of the endpoint spectrum.

B.5 Examples and non-examples

Example B.5 (Linear inpainting). Let $h(x) = Hx$ where H selects observed pixels and $R = \sigma_y^2 I$. Then

$$M_t = \sigma_y^{-1} H U_t \Lambda_t^{1/2}. \tag{47}$$

Even though H is fixed, M_t changes because both U_t and Λ_t change. Thus TCO can evolve without any time dependence in the physical measurement operator.

Example B.6 (Nonlinear phase retrieval). For a simplified real phase-retrieval map $h(x) = |Ax|^2$ componentwise,

$$J_h(\mu) = 2\,\mathrm{diag}(A\mu)A. \tag{48}$$

The restricted operator changes with μ_t even if the local prior subspace were fixed. TCO therefore captures interaction between changing prior geometry and changing measurement physics.

Example B.7 (Isotropic active spectrum). If $M^\top M = sI$ on the active subspace, then all mode-wise Bayes coefficients equal $(1+s)^{-1}$. In this special case a scalar rule with $a = (1+s)^{-1}$ exactly matches the directional estimator. This is the equality case of the scalar lower bound and an important non-example: direction-specific control has no local statistical advantage when the restricted information is isotropic.

Example B.8 (Spectrum scaling versus reshaping). If $M_t = c_t M_\star$ and $M_\star^\top M_\star = sI$, the active spectrum remains isotropic at every step, so scalar guidance can remain locally optimal with a time-dependent scalar. If $M_\star$ is anisotropic, scalar scaling changes all s_i proportionally but still leaves unequal optimal coefficients. The empirical "spectral reshaping" test is stronger: it asks whether normalized singular-value ratios themselves change, demonstrating state-dependent anisotropy rather than a single global SNR schedule.

C Full Proof of the Main Upper Bound

C.1 General linear minimum-MSE theorem

We first prove a slightly more general statement than the main text.

Theorem C.1 (LMMSE with arbitrary second moments)**.** *Let $z \in \mathbb{R}^r$ and $y \in \mathbb{R}^m$ be zero-mean with finite second moments. Assume $\Sigma_y := \mathrm{Cov}(y) \succ 0$ and let $\Sigma_{zy} := \mathbb{E}[zy^\top]$. Then*

$$A^* = \Sigma_{zy}\Sigma_y^{-1} \tag{49}$$

minimizes $\mathbb{E}\|z - Ay\|^2$ over all linear A. Moreover,

$$\mathcal{R}(A) - \mathcal{R}(A^*) = \mathrm{Tr}[(A - A^*)\Sigma_y(A - A^*)^\top]. \tag{50}$$

Proof. Expand

$$\mathcal{R}(A) = \mathrm{Tr}(\Sigma_z) - 2\,\mathrm{Tr}(A\Sigma_{yz}) + \mathrm{Tr}(A\Sigma_y A^\top). \tag{51}$$

Because $A^*\Sigma_y = \Sigma_{zy}$, substituting $A = A^* + E$ cancels the cross terms:

$$\mathcal{R}(A) = \mathcal{R}(A^*) + \mathrm{Tr}(E\Sigma_y E^\top). \tag{52}$$

Positive definiteness of Σ_y proves both optimality and uniqueness. □

Full proof of Theorem 4.1. Under Assumption 3.1, $y = Mz + \xi$ with $\mathrm{Cov}(z) = I$, $\mathrm{Cov}(\xi) = I$, and $z \perp\!\!\!\perp \xi$ in the second-moment sense $\mathbb{E}[z\xi^\top] = 0$. Therefore

$$\Sigma_{zy} = M^\top, \qquad \Sigma_y = MM^\top + I. \tag{53}$$

Apply Theorem C.1 to obtain Equation (12). The push-through form follows from Lemma A.1. Substituting the SVD $M = PKQ^\top$ gives

$$A^* = QK(I + K^2)^{-1}P^\top, \tag{54}$$

which is Equation (13).

For the Gaussian specialization, the pair (z, y) is jointly Gaussian. Its conditional mean is

$$\mathbb{E}[z \mid y] = \Sigma_{zy}\Sigma_y^{-1}y = A^*y. \tag{55}$$

Since $x_0 = \mu + Bz$, linearity of conditional expectation gives

$$\mathbb{E}[x_0 \mid y] = \mu + B\mathbb{E}[z \mid y] = \mu + BA^*y. \tag{56}$$

If B has full column rank, uniqueness of A^*y implies uniqueness of the latent and image corrections. If B is rank-deficient, latent components in $\ker(B)$ do not change the image; all minimizing image-space corrections agree after multiplication by B. □

C.2 Image-space MSE

The main theorem states latent LMMSE because the whitened coordinates make the spectral structure transparent. The same estimator is also optimal for the clean displacement when B has full column rank.

Proposition C.2 (Image-space linear MSE)**.** *Let $W = B^\top B \succ 0$. Among estimators $\widehat{x} = \mu + BAy$, the same A^* minimizes*

$$\mathbb{E}\|B(z - Ay)\|^2 = \mathbb{E}[(z - Ay)^\top W(z - Ay)]. \tag{57}$$

Proof. The weighted risk has gradient

$$2W(A\Sigma_y - \Sigma_{zy}). \tag{58}$$

Since W is invertible, the unique stationary point satisfies $A\Sigma_y = \Sigma_{zy}$, giving A^*. Convexity follows from $W \succ 0$ and $\Sigma_y \succ 0$. □

If $W \succeq 0$ is singular, A^* remains a minimizer, though nonuniqueness can occur in latent directions invisible to B.

C.3 Gaussian posterior covariance

Full proof of Corollary 4.2. Joint Gaussian conditioning gives

$$\Sigma_{z|y} = I - M^\top (I + MM^\top)^{-1} M. \tag{59}$$

Woodbury's identity with $A = I, U = M^\top, C = I, V = M$ yields

$$(I + M^\top M)^{-1} = I - M^\top (I + MM^\top)^{-1} M. \tag{60}$$

Under the SVD,

$$\Sigma_{z|y} = Q \operatorname{diag}((1 + s_i)^{-1}) Q^\top \tag{61}$$

with identity on any right-null complement. Taking the trace gives

$$\operatorname{Tr}(\Sigma_{z|y}) = \sum_i \frac{1}{1 + s_i}. \tag{62}$$

Similarly,

$$\operatorname{Tr}(I - \Sigma_{z|y}) = \sum_i \frac{s_i}{1 + s_i} = d_{\text{obs}}. \tag{63}$$

For Loewner monotonicity, if $A \succeq B \succeq 0$, then $I + A \succeq I + B \succ 0$. Inversion reverses the positive-definite order, so $(I + A)^{-1} \preceq (I + B)^{-1}$. □

C.4 Risk identities and spectral interpretations

The optimal latent risk is

$$\mathcal{R}^* = \operatorname{Tr}(\Sigma_{z|y}) = r - d_{\text{obs}} = rD. \tag{64}$$

Thus D is exactly the per-mode Bayes risk in the standardized local Gaussian problem. This identity motivates the term residual uncertainty. It also explains why d_{obs} saturates: once $s_i \gg 1$, increasing measurement strength further changes $s_i/(1 + s_i)$ only marginally.

Proposition C.3 (Marginal value of information)**.** *For one mode with $s = \kappa^2$, posterior variance $v(s) = (1 + s)^{-1}$ obeys*

$$v'(s) = -(1 + s)^{-2}, \qquad v''(s) = 2(1 + s)^{-3} > 0. \tag{65}$$

Hence uncertainty decreases monotonically with diminishing absolute returns in information strength.

Proof. Direct differentiation. □

C.5 Bayes gain as a regularized inverse

The clean correction can be written without an SVD:

$$\Delta\mu = B(I + M^\top M)^{-1} M^\top \tilde{y}. \tag{66}$$

Equivalently, $\hat{z}$ solves the ridge problem

$$\hat{z} = \arg\min_z \left\{ \|\tilde{y} - Mz\|^2 + \|z\|^2 \right\}. \tag{67}$$

The normal equations are $(I + M^\top M)\hat{z} = M^\top \tilde{y}$. This provides a second interpretation of DynaObs: it performs a locally covariance-whitened Tikhonov correction in the current generative coordinates. The SVD is not conceptually necessary, but it exposes the direction-specific gain and enables diagnostics.

C.6 Why Gaussianity is not needed for the linear theorem

The proof of Theorem 4.1 uses only means, covariances, and zero cross-covariance. Therefore heavy-tailed, skewed, or multimodal local z can share the same LMMSE rule. What Gaussianity adds is the equality between LMMSE and the unrestricted conditional mean, together with a posterior covariance independent of the realized y. In non-Gaussian models, $(I + M^\top M)^{-1}$ is the LMMSE error covariance, not necessarily the full conditional covariance.

Proposition C.4 (LMMSE error covariance)**.** *Under ass 3.1, the error covariance of A^* is*

$$\Sigma_e = I - M^\top (I + MM^\top)^{-1} M = (I + M^\top M)^{-1}. \tag{68}$$

Proof. Compute $e = z - A^*(Mz + \xi)$, expand $\mathbb{E}[ee^\top]$, use independence in second moments, and simplify using $A^* = M^\top (I + MM^\top)^{-1}$. The resulting Schur complement equals the Woodbury form above. □

Thus even without Gaussianity, D remains the normalized LMMSE error trace. What changes is the Bayesian interpretation, not the linear estimation identity.

D Full Proof of the Scalar Lower Bound

D.1 Mode-wise decomposition

Let $M = PKQ^\top$. Because Q is orthogonal, define $\bar{z} = Q^\top z$. Because P is orthogonal on the measurement singular subspace, define $\bar{\xi} = P^\top \xi$ and $\bar{y} = P^\top \tilde{y}$. Under the second-moment model, each active singular coordinate obeys

$$\bar{y}_i = \kappa_i \bar{z}_i + \bar{\xi}_i, \tag{69}$$

with unit variance for $\bar{z}_i$ and $\bar{\xi}_i$ and zero cross-covariance. The scalar estimator becomes

$$\widehat{\bar{z}}_i = a\kappa_i \bar{y}_i = a s_i \bar{z}_i + a\kappa_i \bar{\xi}_i. \tag{70}$$

Therefore

$$\begin{aligned} \mathcal{R}_i(a) &= \mathbb{E}[(\bar{z}_i - \widehat{\bar{z}}_i)^2] && (71)\\ &= (1 - a s_i)^2 + a^2 s_i && (72)\\ &= 1 - 2a s_i + a^2 s_i (1 + s_i). && (73) \end{aligned}$$

For $s_i = 0$, this is identically one: neither scalar nor directional linear measurement correction can affect an unobservable mode.

D.2 Completion of the square

For $s_i > 0$, the derivative of Equation (73) is

$$-2s_i + 2a s_i (1 + s_i), \tag{74}$$

so the mode-wise minimizer is $b_i = (1 + s_i)^{-1}$. Substitution gives minimum risk $(1 + s_i)^{-1}$. Completing the square,

$$\mathcal{R}_i(a) = \frac{1}{1 + s_i} + s_i(1 + s_i)\left(a - \frac{1}{1 + s_i}\right)^2. \tag{75}$$

Summing over all modes proves the exact excess identity Equation (16).

D.3 Best scalar and exact gap

Let $w_i = s_i(1 + s_i)$ and ignore zero-weight modes. By Lemma A.5,

$$\sum_i w_i (a - b_i)^2 = W(a - \bar{b}_w)^2 + \sum_i w_i (b_i - \bar{b}_w)^2, \tag{76}$$

where $W = \sum_i w_i$ and $\bar{b}_w = W^{-1} \sum_i w_i b_i$. Since $w_i b_i = s_i$,

$$\bar{b}_w = \frac{\sum_i s_i}{\sum_i s_i (1 + s_i)} = a^*_{\mathrm{sc}}. \tag{77}$$

Hence

$$G(M) = \inf_a [\mathcal{R}(aM^\top) - \mathcal{R}(A^*)] = \sum_i w_i (b_i - \bar{b}_w)^2. \tag{78}$$

This is zero iff all b_i are equal on active modes, equivalently iff all positive s_i are equal.

An alternative closed form follows by optimizing the original scalar quadratic. Define $S_1 = \sum_i s_i$ and $S_2 = \sum_i s_i^2$. Then

$$\inf_a \mathcal{R}(aM^\top) = r - \frac{S_1^2}{S_1 + S_2}, \tag{79}$$

while

$$\mathcal{R}(A^*) = \sum_i \frac{1}{1 + s_i}. \tag{80}$$

Therefore

$$G(M) = r - \frac{S_1^2}{S_1 + S_2} - \sum_i \frac{1}{1 + s_i}. \tag{81}$$

The weighted-variance form is more interpretable; the moment form is convenient for implementation checks.

D.4 Pairwise lower bound and rank-two equality

For any active pair p, q,

$$G(M) = \min_a \sum_i w_i (a - b_i)^2 \tag{82}$$

$$\geq \min_a [w_p (a - b_p)^2 + w_q (a - b_q)^2] \tag{83}$$

$$= \frac{w_p w_q}{w_p + w_q} (b_p - b_q)^2. \tag{84}$$

Now

$$b_p - b_q = \frac{s_q - s_p}{(1 + s_p)(1 + s_q)}. \tag{85}$$

Substituting $w_p = s_p(1 + s_p)$ and $w_q = s_q(1 + s_q)$ gives

$$\frac{s_p s_q (s_p - s_q)^2}{(1 + s_p)(1 + s_q)[s_p(1 + s_p) + s_q(1 + s_q)]}, \tag{86}$$

which is exactly $\Delta(s_p, s_q)$. If $r = 2$ and both modes are active, the inequality step discards nothing, so equality holds.

D.5 A hard family interpretation

The rank-two construction provides an explicit family of local inverse problems on which scalar guidance is optimally but irreducibly worse. Fix $0 < u < v$ and consider

$$M_{u,v} = \begin{bmatrix} \sqrt{u} & 0 \\ 0 & \sqrt{v} \end{bmatrix}, \qquad z, \xi \sim \mathcal{N}(0, I_2). \tag{87}$$

The unrestricted linear Bayes rule uses coefficients $1/(1 + u)$ and $1/(1 + v)$ in front of the two coordinates of $M^\top y$. Every scalar rule must choose one a. Its excess Bayes risk is exactly $\Delta(u, v)$. No estimator in the restricted scalar class can beat this number, even if it knows u, v, the noise variance, and the prior covariance exactly.

This is an *oracle lower bound*: uncertainty about hyperparameters is not the source of failure. The restriction to one scalar coefficient is itself the obstruction.

D.6 Lower bound in terms of condition number

Let $u = s_{\min} > 0$ and $v = s_{\max} = \chi u$ for an anisotropy ratio $\chi > 1$. Then the exact pair lower bound is

$$\Delta(u, \chi u) = \frac{\chi u^4 (\chi - 1)^2}{(1 + u)(1 + \chi u)[u(1 + u) + \chi u(1 + \chi u)]}. \tag{88}$$

For weak measurements $u \ll 1$ with fixed χ, this scales as

$$\Delta(u, \chi u) = \frac{\chi(\chi-1)^2}{1+\chi} u^3 + O(u^4). \tag{89}$$

For very strong measurements $u \to \infty$ with fixed χ,

$$\Delta(u, \chi u) = \frac{(\chi-1)^2}{\chi(1+\chi^2)} \frac{1}{u} + O(u^{-2}). \tag{90}$$

Thus the scalar penalty is largest in an intermediate regime: when all modes are almost completely uninformative or almost perfectly observed, there is less absolute risk left to separate. This explains why our experiments span noise and operator severity rather than expecting the practical advantage to grow monotonically with condition number.

D.7 Equality cases and degeneracies

There are three ways for the gap to vanish:

1. no active modes ($M = 0$);
2. exactly one active mode;
3. multiple active modes with equal s_i.

Zero modes impose no additional scalar constraint because $M^\top y$ is zero there. This is why the theorem states equality on *positive* singular values rather than all r singular values.

D.8 Block-scalar extension

The same proof yields a hierarchy between one scalar and fully directional control. Partition active modes into blocks $\mathcal{B}_1, \ldots, \mathcal{B}_J$ and allow one scalar a_j per block. The exact excess risk is

$$G_{\text{block}}(M) = \sum_{j=1}^{J} \min_{a_j} \sum_{i \in \mathcal{B}_j} w_i (a_j - b_i)^2. \tag{91}$$

Each block contributes its weighted within-block variance. The fully directional rule corresponds to singleton blocks and has zero excess; one global scalar corresponds to one block. This makes the theorem a special case of a more general approximation-complexity trade-off.

Proposition D.1 (Monotonicity under partition refinement)**.** *If partition* Π' *refines partition* Π*, then* $G_{\Pi'}(M) \le G_{\Pi}(M)$.

Proof. Every block-scalar estimator feasible under Π is feasible under the finer partition by assigning the same coefficient to all child blocks. Minimizing over a superset cannot increase risk. □

This extension suggests a future computational compromise: cluster nearby TCO modes and use blockwise gains when a full SVD or fully mode-specific control is undesirable.

E Full Stability and Approximation Analysis

The exact local theory assumes a known linearized operator and exact low-rank prior factor. This appendix proves the perturbation theorem and connects its abstract errors to curvature, finite differences, covariance estimation, and stale geometry.

E.1 Correction perturbation theorem

Recall $C = BF(M)q$ and $\hat{C} = \hat{B}F(\hat{M})\hat{q}$.

Proof of Theorem 4.5. Add and subtract $BF(\hat{M})\hat{q}$ and $BF(M)\hat{q}$:

$$\hat{C} - C = (\hat{B} - B)F(\hat{M})\hat{q} + B[F(\hat{M}) - F(M)]\hat{q} \tag{92}$$
$$+ BF(M)(\hat{q} - q). \tag{93}$$

Taking norms and applying Lemmas A.2 and A.3,

$$\|\hat{C} - C\| \le \tfrac{1}{2}\varepsilon_B\|\hat{q}\| + \|B\|(1 + 2K_M^2)\varepsilon_M\|\hat{q}\| \tag{94}$$
$$+ \tfrac{1}{2}\|B\|\varepsilon_q. \tag{95}$$

Since $\|\hat{q}\| \le \|q\| + \|\hat{q} - q\| \le Y + \varepsilon_q$, we obtain Equation (21). If $\hat{q}$ is produced by first-order linearization over $\|\delta\| \le \rho$, Lemma A.6 gives the curvature term in Equation (22); any other residual perturbation is collected into $\varepsilon_{\text{meas}}$. □

E.2 Separating sources of restricted-operator error

Let the ideal operator be

$$M = R^{-1/2}JB, \tag{96}$$

and suppose the implementation uses $\hat{R}^{-1/2}$, $\hat{J}$, and $\hat{B}$. Then

$$\hat{M} - M = (\hat{R}^{-1/2} - R^{-1/2})\hat{J}\hat{B} + R^{-1/2}(\hat{J} - J)\hat{B} \tag{97}$$
$$+ R^{-1/2}J(\hat{B} - B). \tag{98}$$

Therefore

$$\varepsilon_M \le \varepsilon_R\|\hat{J}\|\|\hat{B}\| + \|R^{-1/2}\|\varepsilon_J\|\hat{B}\| + \|R^{-1/2}\|\|J\|\varepsilon_B, \tag{99}$$

where $\varepsilon_R = \|\hat{R}^{-1/2} - R^{-1/2}\|$ and $\varepsilon_J = \|\hat{J} - J\|$ on the estimated subspace. This decomposition makes noise misspecification, finite-difference/JVP error, and prior-geometry error separately auditable.

E.3 Forward curvature and trust-region radius

Suppose the true local displacement is $\delta = Bz + r_G(z)$ with $\|\delta\| \le \rho$ and $\|r_G(z)\| \le c_G L_G \rho^2$. Then

$$h(\mu + \delta) = h(\mu) + JBz + Jr_G(z) + r_h(\delta), \tag{100}$$

where $\|R^{-1/2}r_h(\delta)\| \le \frac{1}{2}\|R^{-1/2}\|L_h\rho^2$. The combined model residual obeys

$$\varepsilon_q \lesssim \|R^{-1/2}\|\left(\|J\|c_G L_G + \frac{1}{2}L_h\right)\rho^2 + \varepsilon_{\text{meas}}. \tag{101}$$

Thus both forward curvature and generative-chart curvature enter quadratically in the trust radius. This is the formal source of the schematic target bound $C_3 L_h \rho^2 + C_4 L_G \rho^2$.

E.4 Stale geometry between refresh steps

Suppose geometry is refreshed at t_0 and reused at nearby t. Assume path-Lipschitzness

$$\|B_t - B_{t_0}\| \le L_B|\tau_t - \tau_{t_0}|, \tag{102}$$
$$\|M_t - M_{t_0}\| \le L_M|\tau_t - \tau_{t_0}|. \tag{103}$$

for a normalized reverse-time coordinate τ. If the largest spacing between refreshed steps is $\Delta\tau_K$, then Theorem 4.5 applies with

$$\varepsilon_B \le L_B\Delta\tau_K, \qquad \varepsilon_M \le L_M\Delta\tau_K. \tag{104}$$

This yields a direct quality–compute interpretation of the refresh interval K: larger K reduces geometry evaluations but increases a first-order stale-geometry term unless the path changes slowly.

E.5 Finite-difference black-box operator products

For black-box h, approximate Ju_i by centered differences. If columns u_i are unit norm and Lemma A.7 holds uniformly, each column error is at most $L_3\tau^2/6$. For the $m \times r$ restricted Jacobian,

$$\|\hat{J}U - JU\|_{\mathrm{op}} \le \|\hat{J}U - JU\|_F \le \frac{L_3}{6}\tau^2\sqrt{r}. \tag{105}$$

After multiplying by $R^{-1/2}$ and $\Lambda^{1/2}$, this contributes

$$\varepsilon_M^{\mathrm{FD}} \le \frac{L_3}{6}\tau^2\sqrt{r}\,\|R^{-1/2}\|\,\|\Lambda^{1/2}\|. \tag{106}$$

In finite precision or noisy forward calls, too small τ amplifies numerical error, so τ is selected by a small held-out calibration rather than taken to zero mechanically.

E.6 Subspace perturbation from basis error

Suppose U and $\hat{U}$ have orthonormal columns and principal-angle error

$$\|UU^\top - \hat{U}\hat{U}^\top\|_{\mathrm{op}} \le \varepsilon_{\mathrm{sub}}. \tag{107}$$

If a Procrustes alignment O is chosen to align bases, standard subspace geometry implies an aligned basis error of order $\varepsilon_{\mathrm{sub}}$ away from 90° principal angles. If additionally

$$\|\hat{\Lambda}^{1/2} - O^\top\Lambda^{1/2}O\| \le \varepsilon_\Lambda, \tag{108}$$

then

$$\|\hat{U}\hat{\Lambda}^{1/2} - U\Lambda^{1/2}O\| \le \|\hat{U} - UO\|\|\hat{\Lambda}^{1/2}\| + \varepsilon_\Lambda. \tag{109}$$

Because the correction is latent-rotation invariant (Proposition B.1), the orthogonal O is harmless. Hence projector angle and covariance-eigenvalue error jointly determine ε_B.

E.7 Noise covariance misspecification

If $R = \sigma^2 I$ but the algorithm uses $\hat{\sigma} = c\sigma$, then

$$\hat{M} = c^{-1}M, \tag{110}$$

and every s_i is scaled by c^{-2}. This does not merely rescale the final correction uniformly because the shrinker $\kappa/(1+\kappa^2)$ is nonlinear. Overestimating noise ($c > 1$) shifts modes toward the weak-information regime; underestimating noise can make the method overconfident. The robustness appendix therefore includes $c \in \{0.5, 1, 2\}$.

E.8 Backtracking as an approximation monitor

The theory does not require a particular backtracking rule, but a practical sufficient diagnostic is to reject or shrink α when the actual whitened residual

$$\|R^{-1/2}[y - h(\mu + \alpha\Delta\mu)]\| \tag{111}$$

exceeds a tolerance relative to the predicted linearized residual. This directly checks whether the neglected $O(\rho^2)$ term is dominating. It also guards the failure case where the Jacobian is locally misleading due to high curvature.

E.9 A one-step excess-risk perturbation bound

When the true residual model is $y = Mz + \xi + e$ with deterministic $\|e\| \le \eta$, using the ideal linear gain produces an additional latent estimation perturbation $F(M)e$ of norm at most $\eta/2$. Hence

$$\mathbb{E}\|z - F(M)y\|^2 \le \mathcal{R}^* + \eta\sqrt{\mathcal{R}^*} + \frac{1}{4}\eta^2 \tag{112}$$

by expanding the squared norm and applying Cauchy–Schwarz to the cross term. This crude inequality shows that small residual-model error causes a continuous degradation in risk; it is not needed for the main theorem but helps interpret trust-region scaling.

F Counterexamples and Assumption Necessity

This appendix stress-tests the formal setting. The goal is not to claim every assumption is indispensable for every weaker statement, but to show exactly which conclusions fail or change when a component is removed.

F.1 Why prior covariance whitening is necessary

Consider one image-space direction u observed with scalar sensitivity $j = 1$ and unit measurement noise. Two local priors share the same tangent direction $U = u$ but have variances $\lambda_1 = 0.01$ and $\lambda_2 = 100$. A geometry based only on JU assigns the same observability to both. TCO gives

$$\kappa_1 = 0.1, \qquad \kappa_2 = 10. \tag{113}$$

The posterior variance fractions are approximately 0.990 and 0.0099, respectively. Thus the same physical measurement has almost no effect relative to a very certain prior but dominates a highly uncertain prior. Any method that drops $\Lambda^{1/2}$ cannot represent this distinction.

Proposition F.1 (No covariance-free universal gain)**.** *There is no function of JU and R alone that equals the local Bayes gain for all positive prior variances Λ.*

Proof. The one-dimensional example above keeps JU and R fixed while changing the Bayes coefficient from $1/(1 + 0.01)$ to $1/(1 + 100)$ in front of $M^\top y$, or equivalently changes the image-space gain through B. A covariance-free function would have to output the same value in both cases, contradiction. □

F.2 Why noise whitening is necessary

Take two measurement channels with identical signal sensitivity but variances 1 and 100. Without whitening, the raw Jacobian treats them equally. After whitening, the noisy channel's sensitivity is divided by 10. The LMMSE normal equations depend on $J^\top R^{-1} J$, not $J^\top J$.

Example F.2 (Unit dependence without whitening)**.** Even changing the physical unit of one measurement coordinate, e.g. meters to centimeters, multiplies a row of J by 100 and the corresponding noise standard deviation by 100. The statistical problem is unchanged, but the unwhitened singular values change. TCO is invariant because $R^{-1/2}$ cancels this coordinate scaling up to an orthogonal whitening transform (Proposition B.2).

F.3 Why noise–prior independence matters

Suppose in one dimension $y = \kappa z + \xi$ with $\operatorname{Var} z = \operatorname{Var} \xi = 1$ but $\operatorname{Cov}(z, \xi) = c \neq 0$. Then

$$\operatorname{Cov}(z, y) = \kappa + c, \tag{114}$$

$$\operatorname{Var}(y) = \kappa^2 + 1 + 2\kappa c. \tag{115}$$

The LMMSE coefficient is

$$\frac{\kappa + c}{\kappa^2 + 1 + 2\kappa c}, \tag{116}$$

not $\kappa/(1 + \kappa^2)$. Thus the independence/zero-cross-covariance assumption is substantive. In physical inverse problems, measurement noise is ordinarily modeled as independent of the latent clean signal, which is why the assumption is standard.

F.4 Gaussianity is sufficient, not necessary, for LMMSE

Let z be Rademacher, $\mathbb{P}(z = \pm 1) = 1/2$, and let ξ be independent Rademacher noise. The conditional mean is nonlinear/discrete and generally differs from the Gaussian formula, but the best *linear* estimator still uses $\kappa/(1 + \kappa^2)$ because only second moments enter. Thus Gaussianity is not necessary for Theorem 4.1; it is necessary for the claim that this linear rule is the unrestricted posterior mean with covariance $(I + M^\top M)^{-1}$ independent of y.

F.5 First-order observability can miss second-order information

Example F.3 (Quadratic measurement singularity)**.** Let $x \in \mathbb{R}$, $h(x) = x^2$, anchor $\mu = 0$, and small-noise measurement $y = x^2 + \epsilon$. Then $J_h(0) = 0$, so first-order TCO gives $M = 0$ and declares the local direction unobservable. Yet y strongly informs $|x|$. The missing information is second order.

This example proves that no Jacobian-only local theory can be globally correct for arbitrary nonlinear measurements near critical points. The trust-region/stability theorem therefore states a local approximation guarantee; it does not erase genuinely higher-order information.

F.6 A scalar worst-mode score is insufficient

Consider two spectra

$$\kappa^{(A)} = (1, 1), \qquad \kappa^{(B)} = (1, 10). \tag{117}$$

Both have the same minimum singular value. Yet

$$D_A = \tfrac{1}{2}(1/2 + 1/2) = 0.5, \tag{118}$$

$$D_B = \tfrac{1}{2}(1/2 + 1/101) \approx 0.255. \tag{119}$$

They therefore have different Bayes risks and different optimal directional gains. A worst-mode certificate can be useful for robustness but cannot recover full local uncertainty.

F.7 Trace-equal spectra can require different control

Let squared singular spectra be $s^{(A)} = (2, 2)$ and $s^{(B)} = (0.1, 3.9)$. Both have trace 4. In case A, the optimal scalar coefficient is exactly $1/3$ on both active modes, so scalar guidance is locally optimal. In case B, the optimal mode coefficients are approximately 0.909 and 0.204, creating a positive scalar gap. Thus equal trace does not imply equal control. This motivates the spectrum-equalized negative control in the experiments.

F.8 Same endpoint, different path

The construction in Proposition 4.4 already gives a formal counterexample to endpoint sufficiency for debt. A concrete two-step version takes

$$M_0^{(A)} = M_0^{(B)} = I, \tag{120}$$

$$M_1^{(A)} = 0, \tag{121}$$

$$M_1^{(B)} = 10I, \tag{122}$$

with equal weights. Then the endpoint spectrum is identical, but

$$\mathcal{D}_A = \tfrac{1}{2}(1/2 + 1) = 0.75, \tag{123}$$

$$\mathcal{D}_B = \tfrac{1}{2}(1/2 + 1/101) \approx 0.255. \tag{124}$$

No endpoint-only predictor can distinguish these paths by construction.

F.9 Why mode indices are not tracked naively

Let

$$M_t = \mathrm{diag}(2, 1), \qquad M_{t+1} = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}. \tag{125}$$

The singular values are identical, but the right singular directions are permuted. Labeling "mode 1" purely by SVD index can therefore create a false apparent jump or falsely preserve identity through crossings. Basis-invariant summaries such as the spectrum, D_t, and $d_{\mathrm{obs}}(t)$ avoid this issue. If activation times for individual modes are reported, adjacent subspaces must be aligned by principal angles/Hungarian or Procrustes matching.

F.10 Why a fixed measurement SVD is not enough

Take a fixed $H = I_2$ and two local prior factors

$$B_A = \mathrm{diag}(10, 0.1), \qquad B_B = R_{\pi/4}\,\mathrm{diag}(10, 0.1), \tag{126}$$

where $R_{\pi/4}$ is a 45° rotation. The measurement SVD is identical in both states, but the image-space directions receiving high and low uncertainty change. If H is not isotropic, e.g. $H = \mathrm{diag}(1, 0.2)$, even the restricted singular values change after rotation. Thus fixed operator coordinates cannot generally encode sample-dependent generative geometry.

F.11 Why local rank truncation is an approximation, not a theorem

Suppose the true covariance has eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n > 0$, but only rank r is retained. Directions beyond r are treated as prior-only in the low-rank model. If the discarded tail has substantial energy or contains a direction strongly observed by the measurement, the approximation can miss useful evidence. A sufficient diagnostic is to report captured covariance mass

$$\frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \tag{127}$$

and perform a rank sweep. The efficiency sweep shows quality saturation around $r = 16$, supporting this rank as a practical truncation point while preserving the interpretation of rank truncation as an approximation.

F.12 Why repeated local updates are not independent Bayes factors

The same physical measurement y is reused across reverse timesteps. Therefore one must not multiply local Gaussian likelihoods across t as if each were an independent observation. Doing so would double count information. Observability debt is only an averaged diagnostic of local residual uncertainty. DynaObs uses the local correction to steer a base sampler; it does not interpret each timestep as a new independent measurement.

F.13 Failure of unconditional "measurement speaks more" monotonicity

The gain magnitude $g(\kappa) = \kappa/(1 + \kappa^2)$ is not monotone for all κ: it increases up to $\kappa = 1$ and decreases afterward. This is not a contradiction. For large κ, the residual is measured in units whose signal coefficient is also κ; the equivalent coefficient in front of $M^\top y$ is $(1 + \kappa^2)^{-1}$, while posterior variance continues decreasing monotonically. The correct monotonic statement is uncertainty shrinkage, not raw gain magnitude.

G Assumption Audit and Parameter Accounting

G.1 Assumption-by-assumption audit

G.2 Core dimensions and resources

Let

- n: image/signal dimension;
- m: measurement dimension;
- r: retained local tangent/covariance rank;
- T: number of reverse steps;
- K: geometry refresh interval in reverse steps;
- $G = \lceil T/K \rceil$: number of geometry refreshes;
- S: number of ensemble perturbations for ensemble-PCA;
- c_D: cost of one denoiser/clean-estimator evaluation;

Table 4: Assumption audit. "Needed for" identifies the exact theorem component that uses each condition rather than presenting all assumptions as a monolith.

Assumption	Needed for	Why standard/reasonable	What breaks without it
$\mathrm{Cov}(z) = I$ after whitening	clean spectral interpretation	any positive local covariance can be factorized into $B = U\Lambda^{1/2}$ and standardized coordinates	if not whitened, singular values mix prior scale with arbitrary coordinate scale
$R \succ 0$ and noise whitening	unit-noise local model	Gaussian/second-moment noise covariance is standard in inverse problems; singular R can be handled on its support	unwhitened spectrum is unit-dependent and overweights noisy channels
$\mathbb{E}[z\xi^\top] = 0$	closed-form LMMSE gain	physical measurement noise is typically independent of the clean signal	cross-covariance changes the optimal coefficient; see Appendix F
Gaussianity	exact posterior mean/covariance	local Gaussian approximations are common and testable by ensembles	without it the rule remains LMMSE but need not equal the nonlinear conditional mean
Lipschitz J_h on trust region	$O(\rho^2)$ forward remainder	local C^2 smoothness for differentiable physics	near singular/nonsmooth points first-order TCO can miss higher-order information
low-rank local covariance	computational tractability and interpretable modes	denoiser perturbations empirically concentrate in low-rank subspaces in the DynaObs implementation	a large discarded covariance tail can hide observable directions

Table 5: Matched-endpoint pair. Similar static endpoint observability coexists with different trajectory debt and reconstruction error.

Case	γ_0	$\mathcal{D}\downarrow$	RMSE$\downarrow$	LPIPS$\downarrow$
A	0.083	0.766	0.119	0.252
B	0.086	**0.618**	**0.064**	**0.146**

- c_h: cost of one forward-operator evaluation;
- c_{JVP}: cost of one operator JVP;
- $c_{D\mathrm{JVP}}$: cost of one denoiser JVP when randomized-JVP geometry is used.

The base sampler cost is not counted as a DynaObs overhead unless the correction adds denoiser evaluations.

G.3 Ensemble-PCA geometry cost

At one refresh, ensemble-PCA evaluates S perturbed clean anchors. Ignoring a shared unperturbed anchor,

$$C_{\mathrm{ensemble}} \approx Sc_D + C_{\mathrm{PCA}}. \tag{128}$$

If the centered anchor matrix is $X \in \mathbb{R}^{n\times S}$ with $S \ll n$, the top-r left singular vectors can be obtained from the $S \times S$ Gram matrix. The dominant algebraic cost is approximately

$$O(nS^2 + S^3) \tag{129}$$

for a naive Gram eigendecomposition, or lower with randomized/truncated methods. Storage is $O(nS)$ if all perturbations are retained simultaneously. This version is best treated as an analysis oracle/reference, not the default full-resolution implementation.

G.4 RANDOMIZED-JVP GEOMETRY COST

With $q \geq r$ random probe directions and q denoiser JVPs, a range finder produces $Y \in \mathbb{R}^{n\times q}$. Orthogonalization costs $O(nq^2)$. A small projected covariance/eigendecomposition costs $O(q^3)$. Thus one refresh uses roughly

$$qc_{D\mathrm{JVP}} + O(nq^2 + q^3). \tag{130}$$

For $q \asymp r$, this scales linearly in the number of expensive JVPs and quadratically in rank for vector-space algebra.

G.5 RESTRICTED MEASUREMENT OPERATOR COST

For linear $h(x) = Hx$, computing HU or HB requires r applications of H if no matrix is explicitly materialized. For differentiable nonlinear h, batched JVPs compute $J_h(\mu)U$ in r directional products. For black-box centered finite differences, each direction uses two forward calls, so

$$N_h^{\mathrm{geom}} = 2rG \tag{131}$$

additional forward evaluations. One-sided differences would halve calls but increase truncation bias from $O(\tau^2)$ to $O(\tau)$.

G.6 SMALL SVD COST

Once $M \in \mathbb{R}^{m\times r}$ is formed with $r \ll \min(m, n)$, a thin SVD costs

$$O(mr^2 + r^3) \tag{132}$$

with explicit M. If m is extremely large, one can instead form $M^\top M \in \mathbb{R}^{r\times r}$ using restricted products; diagonalizing this matrix costs $O(r^3)$. The DynaObs image correction then multiplies by $B \in \mathbb{R}^{n\times r}$, costing $O(nr)$.

G.7 TOTAL OVERHEAD

A JVP-based implementation with $q \asymp r$ has the schematic overhead

$$\begin{aligned} C_{\mathrm{total}} = G\big[O(rc_{D\mathrm{JVP}}) + O(rc_{\mathrm{JVP}})\big] \\ + G\,O(nr^2 + mr^2 + r^3) + O(Tnr). \end{aligned} \tag{133}$$

where the last term applies the cached rank-r correction at every step. In practice, tensor kernels, batching, and the fact that B may live in latent rather than pixel space determine constants.

G.8 MEMORY ACCOUNTING

The core cached objects are B (nr numbers), restricted measurement products (mr), and small $r \times r$ matrices. Thus

$$M_{\mathrm{cache}} = O((n + m)r + r^2). \tag{134}$$

Ensemble-PCA temporarily adds $O(nS)$ storage unless samples are streamed into a covariance sketch. Randomized-JVP geometry avoids storing S full clean anchors and is therefore the preferred practical option.

G.9 REFRESH INTERVAL VERSUS APPROXIMATION ERROR

Let normalized reverse time span $[0, 1]$ over T steps. If refreshes are approximately every K steps, then $\Delta\tau_K \approx K/T$. Under the path-Lipschitz model in Appendix E,

$$\varepsilon_B = O(L_B K/T), \qquad \varepsilon_M = O(L_M K/T). \tag{135}$$

Substitution into Equation (21) yields a first-order stale-geometry error proportional to K/T, while computational geometry cost is proportional to T/K. This creates the expected quality–cost frontier and motivates explicitly reporting K rather than hiding refresh frequency in code.

G.10 Rank versus approximation error

Let C_t be the ideal local clean covariance with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. The best rank-$r$ spectral approximation obeys

$$\|C_t - C_{t,r}\|_{\mathrm{op}} = \lambda_{r+1}, \qquad \|C_t - C_{t,r}\|_F^2 = \sum_{i>r} \lambda_i^2. \tag{136}$$

These are covariance errors, not directly ε_B, but they give measurable diagnostics for the truncation tail. The rank sweep reports both reconstruction metrics and geometry agreement because quality saturation without estimator stability can be accidental.

G.11 Numerical parameter settings and sensitivity

The rank/refresh sweep identifies $r = 16$ and refresh interval $K = 4$ as a practical operating point, with larger ranks or denser refreshes yielding diminishing gains. These values are empirical operating choices rather than theorem-required constants; the full sweep is reported to show the sensitivity to this choice.

G.12 Statistical sample accounting

The publication design separates three sample pools:

$$N_{\mathrm{dev}} = 50 \text{ images/dataset}, \tag{137}$$

$$N_{\mathrm{main}} = 500 \text{ images/dataset}, \tag{138}$$

$$N_{\mathrm{expensive}} = 100 \text{ shared images/dataset}. \tag{139}$$

The expensive subset is for baselines whose compute makes full 500-image evaluation impractical; it must be shared across methods rather than cherry-picked. Primary claims use paired image-level differences and paired bootstrap intervals. Stochastic solver seeds are paired so variance from initial diffusion noise does not dominate method comparisons.

G.13 Parameter checklist for reproducibility

For reproducibility, the experiment record includes at least: dataset image IDs; diffusion checkpoint hash; sampler and step count; operator definition and random seed; true and assumed R; geometry rank r; refresh interval K; ensemble/JVP probe count; finite-difference step if used; trust radius ρ_t and backtracking rule; all baseline hyperparameters; NFE/JVP/VJP/forward-query counts; wall-clock; peak memory; and the exact scripts that regenerate every figure/table. This accounting is necessary because DynaObs can trade geometry compute for quality in multiple ways.

H Information Identities and Guidance Resolution

This appendix develops consequences of the exact local Gaussian model that are useful for interpreting TCO beyond the posterior-mean formula. None of these results is needed to establish Theorem 4.1; they clarify what information is retained or lost when one compresses the TCO spectrum to a scalar or a small number of groups.

H.1 Mutual information and effective observable dimension

Assume in this subsection that $z \sim \mathcal{N}(0, I_r)$, $\xi \sim \mathcal{N}(0, I_m)$ independently, and $\tilde{y} = Mz + \xi$. Let $s_i = \kappa_i^2$.

Proposition H.1 (Gaussian information carried by TCO)**.** *The mutual information between the standardized local prior coordinate and the whitened measurement is*

$$I(z; \tilde{y}) = \frac{1}{2} \log \det(I + M^\top M) = \frac{1}{2} \sum_{i=1}^{r} \log(1 + s_i). \tag{140}$$

If the measurement strength is scaled by $M_\alpha = \sqrt{\alpha}M$, then

$$2\alpha \frac{\mathrm{d}}{\mathrm{d}\alpha} I(z; \tilde{y}_\alpha) = \sum_i \frac{\alpha s_i}{1 + \alpha s_i} = d_{\mathrm{obs}}(M_\alpha). \tag{141}$$

Proof. Because $\tilde{y} \sim \mathcal{N}(0, I + MM^\top)$ and $\tilde{y} \mid z$ has covariance I,

$$I(z; \tilde{y}) = h(\tilde{y}) - h(\tilde{y} \mid z) = \tfrac{1}{2} \log \det(I + MM^\top).$$

Sylvester's determinant identity gives $\det(I + MM^\top) = \det(I + M^\top M)$, and the SVD gives the sum in Equation (140). Replacing s_i by αs_i and differentiating establishes Equation (141). □

Thus d_{obs} is not merely an ad hoc soft rank. It is the logarithmic derivative of local measurement information with respect to a global SNR scaling. Modes with $s_i \ll 1$ contribute approximately s_i, whereas modes with $s_i \gg 1$ contribute approximately one effective observed degree of freedom.

Corollary H.2 (MMSE complementarity)**.** *Under the same Gaussian model,*

$$\mathbb{E}\|z - \mathbb{E}[z \mid \tilde{y}]\|^2 = \mathrm{Tr}(I + M^\top M)^{-1} = r - d_{\mathrm{obs}}, \tag{142}$$

and hence $D = 1 - d_{\mathrm{obs}}/r$.

Proof. The posterior covariance is $(I + M^\top M)^{-1}$. Taking the trace and using $1/(1 + s_i) = 1 - s_i/(1 + s_i)$ proves the identity. □

The complementarity is exact at one local state. It is not interpreted as saying that trajectory debt is the MMSE of a joint multi-step observation model; repeated reverse states share the same underlying measurement and are dependent.

H.2 An exact all-pairs decomposition of the scalar gap

The pairwise lower bound in the main text follows from discarding nonnegative terms. A stronger identity makes the contribution of every spectral pair explicit.

Lemma H.3 (Weighted variance as pairwise disagreement)**.** *For positive weights w_i with $W = \sum_i w_i$ and real values b_i,*

$$\sum_i w_i (b_i - \bar{b}_w)^2 = \frac{1}{W} \sum_{i<j} w_i w_j (b_i - b_j)^2, \tag{143}$$

$$\bar{b}_w = \frac{\sum_i w_i b_i}{W}. \tag{144}$$

Proof. Expand the right-hand side as

$$\frac{1}{2W} \sum_{i,j} w_i w_j (b_i - b_j)^2 = \sum_i w_i b_i^2 - \frac{(\sum_i w_i b_i)^2}{W},$$

which is the standard expanded form of the weighted variance on the left. □

Proposition H.4 (All-pairs scalar impossibility identity)**.** *For the TCO scalar family, let $w_i = s_i(1 + s_i)$ and $b_i = (1 + s_i)^{-1}$, and let $\mathcal{A} = \{i : s_i > 0\}$. Then*

$$\begin{aligned} G(M) &:= \inf_a \{\mathcal{R}(aM^\top) - \mathcal{R}(A^*)\} \\ &= \frac{1}{\sum_{k \in \mathcal{A}} w_k} \sum_{\substack{i<j \\ i,j \in \mathcal{A}}} \frac{s_i s_j (s_i - s_j)^2}{(1 + s_i)(1 + s_j)}. \end{aligned} \tag{145}$$

Proof. Apply Lemma H.3 to the exact weighted-variance expression in Equation (78). Since

$$w_i w_j (b_i - b_j)^2 = s_i s_j (1 + s_i)(1 + s_j) \frac{(s_i - s_j)^2}{(1 + s_i)^2 (1 + s_j)^2},$$

we obtain Equation (145) after cancellation. □

This formula shows that the scalar deficit is a weighted sum of *all* pairwise spectral disagreements. It therefore vanishes precisely for an isotropic active spectrum and increases as informative modes disagree in strength, subject to the saturation induced by $(1 + s_i)^{-1}$.

Corollary H.5 (Spread-based lower bound)**.** *Suppose active squared singular values lie in $[u, v]$ with $0 < u \le v < \infty$. If there exist active modes p, q with $|s_p - s_q| \ge \delta$, then*

$$G(M) \ge \Delta(s_p, s_q) \ge \frac{u^2\delta^2}{2v(1+v)^3}. \tag{146}$$

The first inequality is exact pair selection; the second follows by bounding the denominator uniformly over $[u, v]$.

H.3 Block-scalar guidance interpolates between scalar and directional rules

A natural intermediate design uses one scalar per group of modes rather than one global scalar or r independent coefficients. Let $\Pi = \{G_1, \dots, G_L\}$ be a partition of the active mode set. Define a block-scalar family whose coefficient on mode $i \in G_\ell$ is a_ℓ in front of $\kappa_i \tilde{y}_i$.

Theorem H.6 (Optimal block-scalar rule)**.** *For a fixed partition Π, the optimal group coefficient is*

$$a_\ell^* = \frac{\sum_{i \in G_\ell} s_i}{\sum_{i \in G_\ell} s_i(1 + s_i)}, \tag{147}$$

and its excess over fully directional guidance is

$$G_\Pi(M) = \sum_{\ell=1}^{L} \sum_{i \in G_\ell} w_i (b_i - \bar{b}_{w,G_\ell})^2. \tag{148}$$

If Π' refines Π, then $G_{\Pi'}(M) \le G_\Pi(M)$. Equality holds iff every split created by the refinement separates groups having the same weighted mean b or zero weight. In particular, one group recovers the scalar gap and singleton groups recover zero excess.

Proof. The total risk separates across singular modes. Holding the partition fixed gives one weighted least-squares problem $\sum_{i \in G_\ell} w_i (a_\ell - b_i)^2$ per group, whose minimizer is the weighted mean in Equation (147); substituting yields Equation (148). Refining a partition enlarges the feasible coefficient class, so risk cannot increase. The equality characterization follows from the equality condition for replacing a shared weighted mean by separate weighted means. □

Remark H.7 (Guidance resolution as approximation complexity). Theorem H.6 turns guidance resolution into a one-dimensional weighted quantization problem on the targets $b_i = (1 + s_i)^{-1}$. If compute allows only L gains, the best grouping minimizes within-group weighted variance. This observation suggests adaptive grouping of nearby TCO strengths as a lower-cost approximation to a full SVD update; we do not claim an implemented grouping algorithm in the experiments.

H.4 The crossover $\kappa = 1$: information and update amplitude are different

The posterior variance $v(\kappa) = 1/(1+\kappa^2)$ decreases monotonically, but the posterior-mean coefficient $g(\kappa) = \kappa/(1 + \kappa^2)$ does not.

Lemma H.8 (Gain shape)**.** *For $\kappa \ge 0$, $v'(\kappa) < 0$ for $\kappa > 0$, while*

$$g'(\kappa) = \frac{1 - \kappa^2}{(1 + \kappa^2)^2}. \tag{149}$$

Hence g is maximized at $\kappa = 1$ with value $1/2$.

Proof. Differentiate both scalar functions directly. □

There is no contradiction. When $\kappa \gg 1$, the measurement is highly informative, but a fixed whitened residual corresponds to a smaller latent displacement because the forward sensitivity is large. "Measurements speak more" therefore refers to uncertainty reduction or information, not to monotonic growth of the raw correction coefficient.

H.5 CONDITION-NUMBER REGIMES

For two active modes with squared strengths u and $v = \chi u$, the exact rank-two scalar gap is

$$\Delta(u, \chi u) = \frac{\chi u^4 (\chi - 1)^2}{(1+u)(1+\chi u)\{u(1+u) + \chi u(1+\chi u)\}}. \tag{150}$$

At weak measurement strength $u \to 0$ with fixed χ, this is

$$\Delta(u, \chi u) = \frac{\chi(\chi-1)^2}{1+\chi} u^3 + O(u^4), \tag{151}$$

whereas at $u \to \infty$ with fixed χ,

$$\Delta(u, \chi u) = \frac{(\chi-1)^2}{\chi(1+\chi^2)} \frac{1}{u} + O(u^{-2}). \tag{152}$$

Thus the largest scalar penalty is expected in an intermediate-information regime: if every mode is almost unobserved there is little useful correction to misallocate, while if every active mode is overwhelmingly observed the latent posterior is already strongly contracted. This is a local theoretical statement and not a prediction that end-to-end PSNR must follow the same non-monotone curve.

I TRAJECTORY FUNCTIONALS: BOUNDS, STABILITY, AND IDENTIFIABILITY

TCO is computed at each reverse state, whereas observability debt aggregates those local quantities. This appendix states what can be proved about that aggregation without pretending that reused measurements across timesteps are independent observations.

I.1 BASIC BOUNDS AND MONOTONICITY

Write

$$D(M) := \frac{1}{r} \operatorname{Tr}(I + M^\top M)^{-1} = \frac{1}{r} \sum_{i=1}^{r} \frac{1}{1+s_i}. \tag{153}$$

Lemma I.1 (Range and Loewner monotonicity)**.** *For every M, $0 < D(M) \le 1$. If $N^\top N \succeq M^\top M$, then $D(N) \le D(M)$. If the inequality is strict on a nonzero subspace, then $D(N) < D(M)$.*

Proof. All eigenvalues of $(I + M^\top M)^{-1}$ lie in $(0, 1]$. The map $X \mapsto X^{-1}$ reverses Loewner order on positive-definite matrices, giving the second claim. Strictness follows because the trace of a nonzero positive-semidefinite difference is positive. □

For nonnegative weights w_t summing to one, $\mathcal{D} = \sum_t w_t D(M_t)$ therefore lies in $(0, 1]$. This normalization is why weights are required to be nonnegative and normalized in the definition.

I.2 LIPSCHITZ STABILITY OF INSTANTANEOUS DEBT

Lemma I.2 (Perturbation of residual uncertainty)**.** *Suppose $\|M\|_{\mathrm{op}}, \|N\|_{\mathrm{op}} \le K$. Then*

$$|D(M) - D(N)| \le \frac{2K}{\sqrt{r}} \|M - N\|_F. \tag{154}$$

A dimension-free but sometimes looser alternative is

$$|D(M) - D(N)| \le \|M^\top M - N^\top N\|_{\mathrm{op}}. \tag{155}$$

Proof. Let $A = I + M^\top M$ and $B = I + N^\top N$. The resolvent identity gives

$$A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}.$$

Since $\|A^{-1}\|_{\mathrm{op}}, \|B^{-1}\|_{\mathrm{op}} \le 1$,

$$|\operatorname{Tr}(A^{-1} - B^{-1})| \le \|B - A\|_* \le \sqrt{r}\, \|B - A\|_F.$$

Moreover

$$\begin{aligned}\|N^\top N - M^\top M\|_F &\le (\|N\|_{\mathrm{op}} + \|M\|_{\mathrm{op}})\|N - M\|_F \\ &\le 2K\|N - M\|_F.\end{aligned}$$

Divide by r to obtain Equation (154). For Equation (155), diagonalize neither matrix; instead use $|r^{-1}\operatorname{Tr} X| \le \|X\|_{\mathrm{op}}$ and the same resolvent identity. □

Proposition I.3 (Trajectory debt perturbation)**.** *Consider two trajectories with the same weights and ranks, $\{M_t\}$ and $\{\hat{M}_t\}$, satisfying $\|M_t\|_{\mathrm{op}}, \|\hat{M}_t\|_{\mathrm{op}} \le K_t$. Then*

$$|\hat{\mathcal{D}} - \mathcal{D}| \le \sum_t w_t \frac{2K_t}{\sqrt{r}} \|\hat{M}_t - M_t\|_F. \tag{156}$$

In particular, if $K_t \le K$ and the weighted RMS geometry error is $\varepsilon_{\mathrm{traj}} := (\sum_t w_t \|\hat{M}_t - M_t\|_F^2)^{1/2}$, then

$$|\hat{\mathcal{D}} - \mathcal{D}| \le \frac{2K}{\sqrt{r}} \varepsilon_{\mathrm{traj}}. \tag{157}$$

Proof. Apply Lemma I.2 at each time, sum with nonnegative weights, then use Cauchy–Schwarz and $\sum_t w_t = 1$. □

This separates two questions: whether debt is a useful predictor of end-to-end failure is empirical, while whether the estimated debt tracks the ideal local-geometry debt is controlled by Proposition I.3.

I.3 Refresh intervals and stale geometry

Suppose geometry is refreshed at times $\tau_1, \tau_2, \ldots$ and the most recent geometry is reused between refreshes. Let $\tau(t)$ be the last refresh before t.

Proposition I.4 (Debt error from stale refresh)**.** *If M_t is L_M-Lipschitz in a continuous trajectory parameter τ and every reuse interval has width at most $\Delta\tau$, then with $\hat{M}_t = M_{\tau(t)}$,*

$$|\hat{\mathcal{D}} - \mathcal{D}| \le \frac{2KL_M}{\sqrt{r}} \Delta\tau. \tag{158}$$

Proof. The Lipschitz condition gives $\|M_t - M_{\tau(t)}\|_F \le L_M \Delta\tau$. Substitute this uniform bound into Equation (156). □

For a discrete T-step solver refreshed every K_{ref} steps, we interpret $\Delta\tau$ in the normalized diffusion-time coordinate, not simply as the integer K_{ref}. The bound motivates more frequent refresh when geometry changes rapidly and does not itself select the empirical optimum $K_{\mathrm{ref}} = 4$ observed in the efficiency sweep.

I.4 Endpoint identifiability: a stronger family construction

The main text used one intermediate state. The same idea yields a continuum of path functionals compatible with an identical endpoint.

Proposition I.5 (Arbitrary intermediate debt under a fixed endpoint)**.** *Fix M_0, nonnegative weights summing to one, and let $W = \sum_{t\neq 0} w_t$. For every $q \in (0,1)$ and every $\epsilon > 0$, there exists a trajectory sharing endpoint M_0 whose intermediate weighted debt lies within ϵ of qW:*

$$\left| \sum_{t\neq 0} w_t D(M_t) - qW \right| < \epsilon. \tag{159}$$

Proof. Choose every intermediate operator to be cI_r. Then $D(cI_r) = 1/(1+c^2)$ continuously maps $c \in [0,\infty)$ onto $(0,1]$. Pick c so that $D(cI_r) = q$; the equality is then exact. If dimensional constraints prevent an $r \times r$ identity embedding in a specific measurement space, choose a rank-r isometry whenever $m \ge r$; otherwise the achievable interval ends at the null-space floor, and approximation within ϵ follows on that attainable interval. □

Thus an endpoint statistic can neither recover the weighted average nor even order two general trajectories by debt. Additional dynamical structure would be required to infer path observability from endpoint observability.

I.5 Global rescaling versus spectral reshaping

A central empirical question is whether M_t changes only by a scalar or changes anisotropically. The following invariant formalizes the distinction.

For nonzero M with squared singular-value vector $s(M)$, define its normalized information profile

$$\pi_i(M) := \frac{s_i(M)}{\sum_j s_j(M)}. \tag{160}$$

Zero singular values are retained as zeros and indices may be sorted in descending order when mode identities are not tracked.

Proposition I.6 (Scale-profile characterization)**.** *If $N = cM$ for $c > 0$, then $\pi(N) = \pi(M)$. Conversely, if M and N have the same right and left singular subspaces, the same rank, and $\pi(N) = \pi(M)$, then their nonzero singular values differ by a common scalar factor.*

Proof. The forward direction follows because every squared singular value is multiplied by c^2. For the converse, equality of profiles gives $s_i(N) = \alpha s_i(M)$ for the common constant $\alpha = \sum_j s_j(N)/\sum_j s_j(M)$. Taking square roots gives a common singular-value scaling; the shared singular subspaces then imply the stated factorization on the active subspace. □

When bases rotate, profile change is sufficient but not necessary evidence of nontrivial geometry change: two operators can share singular values while rotating their observable subspaces. Therefore the empirical spectral-reshaping analysis pairs profile distances with principal-angle or projector distances.

I.6 Projector change captures directional rotation

Let $V_k(M)$ collect the top-k right singular vectors of M, and define $P_k(M) = V_k(M)V_k(M)^\top$ when a spectral gap makes the subspace identifiable.

Definition I.7 (Observable-subspace rotation score)**.**

$$\mathrm{Rot}_k(M, N) := \frac{1}{\sqrt{2k}}\|P_k(M) - P_k(N)\|_F \in [0, 1]. \tag{161}$$

The score is basis-invariant inside a degenerate singular subspace and equals the root-mean-square sine of the principal angles up to normalization. It avoids assigning persistent semantic meaning to raw SVD column indices, a failure mode discussed in Section F.

I.7 Why weights must be nonnegative

If trajectory weights are allowed to be negative, $\mathcal{D}$ ceases to represent residual uncertainty: for example, with two times having $D_1 = 1$ and $D_2 = 0.1$, weights $(2, -1)$ sum to one but give debt $1.9 > 1$. If weights do not sum to one, the statistic is scale-dependent. Neither restriction is needed for the local theorems, but both are necessary for the interpretation $\mathcal{D} \in [0, 1]$ as an average diagnostic.

I.8 Dependence across reverse timesteps

The same physical measurement y is reused at every reverse step, and the states x_t are generated recursively. Hence the collection $\{M_t, \tilde{y}_t\}_t$ is neither conditionally independent nor a sequence of fresh experiments. Multiplying one-step Gaussian likelihood factors would double-count evidence. Our theory therefore uses D_t as a local conditional diagnostic and $\mathcal{D}$ as a path summary, not as a derived covariance of a fictitious product likelihood. Establishing a global theorem that connects $\mathcal{D}$ to final reconstruction error would require an explicit stochastic model for the reverse solver and is left open; the failure-prediction experiment tests that hypothesis rather than assuming it.

J Extended Computational and Statistical Accounting

This appendix makes the resource accounting operational enough to reproduce the geometry module or to substitute another estimator without silently changing the compute budget. The symbols are independent of any particular GPU implementation.

J.1 Primitive operation counts

Let C_D denote one denoiser evaluation that maps a reverse state to a clean-space anchor, C_h one forward-operator evaluation, C_{Jv} one Jacobian–vector product through h, and C_{Dv} one Jacobian–vector product through the denoiser/clean estimator when the tangent estimator uses automatic differentiation. Let r be retained tangent rank, $p = r + \ell$ the randomized sketch width including oversampling ℓ, K_{ref} the geometry refresh interval, and T the number of reverse steps. The number of refreshes is

$$N_{\mathrm{ref}} = \left\lceil \frac{T}{K_{\mathrm{ref}}} \right\rceil. \tag{162}$$

The geometry module's total incremental cost is therefore

$$C_{\mathrm{geom}} = N_{\mathrm{ref}} \left(C_{\mathrm{tangent}} + C_{\mathrm{restricted}} + C_{\mathrm{svd}} + C_{\mathrm{misc}} \right). \tag{163}$$

Only the first two terms normally involve expensive model/operator calls; the SVD acts on an $m \times r$ or sketched restricted matrix rather than an $n \times n$ image-space matrix.

J.2 Ensemble-PCA reference geometry

Suppose a reference estimator draws S local clean predictions around the same anchor and forms centered samples $X \in \mathbb{R}^{n \times S}$. Explicit PCA can be performed through the $S \times S$ Gram matrix when $S \ll n$.

$$C^{\mathrm{ens}}_{\mathrm{tangent}} = SC_D + O(nS^2 + S^3), \tag{164}$$

$$M^{\mathrm{ens}}_{\mathrm{mem}} = O(nS + S^2). \tag{165}$$

After retaining r components, constructing $J_h(\mu)B$ with r JVPs costs

$$C^{\mathrm{ens}}_{\mathrm{restricted}} = rC_{Jv} + O(mr), \tag{166}$$

and a thin SVD costs $O(mr^2 + r^3)$. This estimator is treated as a high-cost reference in the estimator-validation table, not as the recommended default.

J.3 Randomized-JVP geometry

A practical estimator may probe the local denoiser map with $p = r + \ell$ randomized directions. With automatic differentiation,

$$C^{\mathrm{JVP}}_{\mathrm{tangent}} = pC_{Dv} + O(npr), \tag{167}$$

$$M^{\mathrm{JVP}}_{\mathrm{mem}} = O(np + nr), \tag{168}$$

where the linear-algebra term depends on whether vectors are streamed or stored. If a one-sided finite difference is used instead of a JVP, p additional denoiser calls replace pC_{Dv}; a central difference doubles that number. The finite-difference truncation and numerical error are governed by Lemma A.7 and the chosen step size.

Once $B = U\Lambda^{1/2}$ is available, the restricted operator can be assembled with r operator JVPs:

$$C_{\mathrm{restricted}} = rC_{Jv} + C_{R^{-1/2}}(m, r), \tag{169}$$

where $C_{R^{-1/2}}$ is $O(mr)$ for diagonal noise and can dominate if dense whitening is implemented naively. For structured covariance, solving with a Cholesky factor or applying an FFT-diagonalizable whitening operator is counted instead of treating whitening as free.

J.4 Applying the correction without forming dense matrices

The clean correction

$$\Delta\mu = BQ\,\mathrm{diag}(g_i)P^\top\tilde{y}, \qquad g_i = \frac{\kappa_i}{1+\kappa_i^2},$$

can be applied using thin factors. Its arithmetic cost after the SVD is

$$O(mr + r^2 + nr), \tag{170}$$

and its persistent storage is $O(nr + mr + r^2)$ if both singular-vector factors are kept. No $n \times n$ posterior covariance is required. A spatial risk diagonal can likewise be approximated from

$$C = BQ\,\mathrm{diag}((1+s_i)^{-1})Q^\top B^\top$$

using rowwise squared thin factors, in $O(nr)$ memory and arithmetic after BQ is available.

Table 6: Mechanism ablation. Components are added cumulatively.

Variant	**PSNR↑**	**LPIPS↓**	**Time↓**
Base solver	24.92	.244	**1.00**×
+ dynamic scalar γ_t	25.18	.232	1.03×
+ static directional spectrum	25.32	.226	1.08×
+ dynamic directional spectrum	25.68	.211	1.14×
+ prior uncertainty Λ_t	<u>25.91</u>	<u>.200</u>	1.22×
+ noise whitening $R^{-1/2}$	**26.08**	**.193**	1.26×

J.5 Amortized relative overhead

Let C_{base} be the cost of one baseline reverse step, including any denoiser and measurement-gradient calls already required by the host solver. The average relative overhead due solely to DynaObs geometry is

$$\begin{aligned}\mathrm{Overhead} &= \frac{C_{\mathrm{geom}}}{TC_{\mathrm{base}}} \\ &= \frac{1}{K_{\mathrm{ref}}}\frac{C_{\mathrm{tangent}} + C_{\mathrm{restricted}}}{C_{\mathrm{base}}} \\ &\quad + \frac{1}{K_{\mathrm{ref}}}\frac{C_{\mathrm{svd}} + C_{\mathrm{misc}}}{C_{\mathrm{base}}} + O(T^{-1}).\end{aligned} \tag{171}$$

This equation makes the r–K_{ref} trade-off explicit. Increasing r generally increases both tangent-estimation and restricted-operator cost approximately linearly in model/operator calls, whereas decreasing K_{ref} increases refresh cost inversely. The empirical Pareto table therefore reports wall-clock time in addition to r and K_{ref} rather than presenting the latter as implementation-independent costs.

J.6 Operator-specific accounting

For linear $h(x) = Hx$, each restricted JVP is simply Hb_i. Inpainting is a masked gather and costs $O(n)$ or less with sparse indexing; downsampling/super-resolution uses the chosen antialias/filter operator; convolutional blur can use FFT or spatial convolution. For nonlinear phase retrieval $h(x) = |Ax|^2$ (real notation for simplicity),

$$J_h(\mu)v = 2\,\mathrm{diag}(A\mu)Av, \tag{172}$$

so a JVP requires one application of A to v after $A\mu$ has been cached at the refresh. A nonlinear blur network or differentiable physics simulator reports one JVP cost explicitly because it may be comparable to a model evaluation.

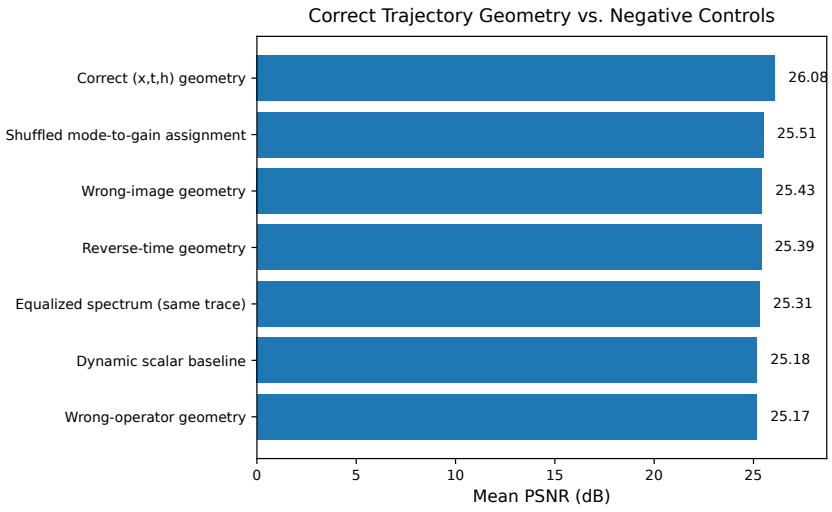

Figure 4: Correct trajectory geometry versus negative controls.

J.7 Statistical accounting and data reuse

The experimental protocol distinguishes three image pools per dataset: a disjoint hyperparameter-development pool, the primary evaluation pool, and a common subset for especially expensive baselines. With the designated sample sizes,

$$N_{\mathrm{dev}} = 50, \qquad N_{\mathrm{main}} = 500, \qquad N_{\mathrm{expensive}} = 100. \tag{173}$$

These are evaluation sample counts, not theorem parameters. The expensive subset is selected independently of method outcomes and shared by every method in the subset comparison. If a larger image set is used for an additional primary task, that changes the statistical precision of the estimate but not the theoretical TCO rank r.

For each reported mean, uncertainty is computed across evaluation images (and across measurement/noise draws when those are randomized). Hyperparameters are chosen on N_{dev} and then frozen. Reusing the evaluation set to tune r, K_{ref}, trust radius, or baseline-specific guidance scales invalidates nominal confidence intervals because the selection step becomes data-dependent.

J.8 Confidence-interval bookkeeping

If a metric value for image j is Z_j and images are the independent sampling unit, a conventional standard error is $\widehat{\sigma}_Z/\sqrt{N}$. When multiple masks/noise draws are nested within each image, the image remains the outer bootstrap unit to avoid treating correlated repeats as independent. Paired comparisons bootstrap the per-image differences $Z_j^{(A)} - Z_j^{(B)}$, preserving the shared inverse-problem instance. The failure-prediction correlations likewise require bootstrap resampling at the level at which trajectories are independent; resampling individual timesteps would severely understate uncertainty.

J.9 A reproducible refresh record

Every run that uses dynamic geometry serializes, at minimum,

$$(t,\ \mu_t,\ r,\ K_{\mathrm{ref}},\ \kappa_t,\ D_t,\ d_{\mathrm{obs}}(t),\ \|\Delta\mu_t\|,\ \alpha_t), \tag{174}$$

where α_t is the accepted trust-region/backtracking scale. For large images, storing the full basis U_t at every refresh may be impractical; a deterministic seed plus enough information to regenerate probes is acceptable only if the model checkpoint, numerical precision, and probe generator are fixed. The logs also record whether geometry was recomputed or reused on each step, otherwise a nominal K_{ref} does not uniquely specify the executed schedule.

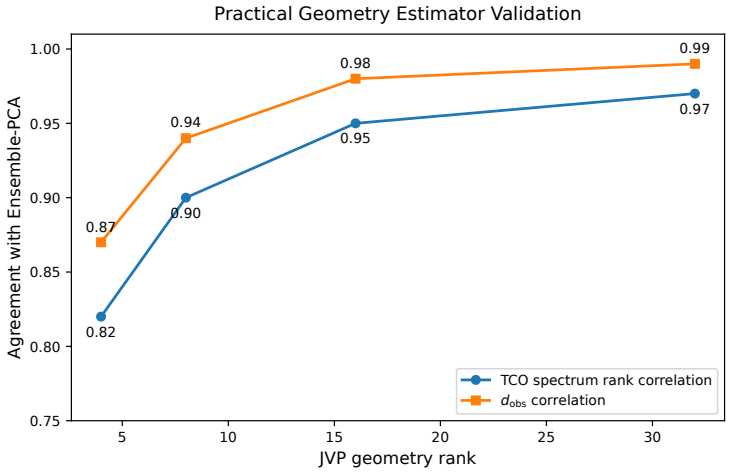

Figure 5: Agreement of randomized-JVP geometry with ensemble-PCA.

J.10 What is a theorem parameter versus an implementation parameter?

The local exact theorems depend on M and its spectrum, not on T, K_{ref}, a tangent-estimation sketch width, or a GPU batch size. The stability theorem introduces K_M, ε_B, ε_M, ε_q, L_h, and ρ because those quantify approximation of the exact local object. The implementation chooses r, p, K_{ref}, finite-difference step, and trust-region rule to control those errors at finite cost. Keeping these levels separate prevents an empirical default such as $r = 16$ from being mistaken for a theoretical threshold.

K Expanded Experimental Design and Supporting Results

K.1 Phase 1: trajectory spectrum audit

The trajectory-spectrum audit uses the controlled diffusion pipeline. For each image/operator case and nine reverse-time checkpoints, we estimate (U_t, Λ_t), form M_t, and store the full TCO spectrum. The primary spectral test compares normalized spectra

$$\bar{\kappa}_t = \kappa_t / \|\kappa_t\|_2 \tag{175}$$

across time. Scalar rescaling $M_t \approx c_t M_{t_0}$ preserves $\bar{\kappa}_t$; true anisotropic reshaping changes it. We summarize this change with cosine distance, Wasserstein distance on normalized spectral mass, spectral entropy, effective rank, and re-estimation noise from repeated geometry estimates.

We assess spectral reshaping by requiring normalized temporal variation to exceed estimator noise for a clear majority of cases. This avoids declaring novelty merely because all singular values grow together as diffusion noise decreases.

K.2 Matched endpoint pair

We select matched pairs algorithmically rather than by manual visual search: first matching cases within a small endpoint-γ_0 tolerance, then ranking pairs by debt difference on a development split, and reporting test-set pair statistics or predeclared examples.

K.3 Mechanism ablation

This ablation corresponds directly to the theoretical object: time dependence tests trajectory conditioning; directional spectrum tests the scalar lower-bound mechanism; Λ_t tests prior whitening; and $R^{-1/2}$ tests noise calibration.

Table 7: Negative controls for mechanism specificity.

Geometry / Control	PSNR↑
Correct (x, t, h) geometry	**26.08**
Wrong-image geometry	25.43
Wrong-operator geometry	25.17
Shuffled mode-to-gain assignment	25.51
Reverse-time geometry	25.39
Equalized spectrum (same trace)	25.31
Dynamic scalar baseline	25.18

K.4 Negative controls

Wrong-image geometry tests sample specificity; wrong-operator geometry tests dependence on physical measurements; shuffled gains test whether assigning the right gain to the right singular mode matters; reverse-time tests trajectory order; equalized spectrum tests whether a trace-matched scalar summary is sufficient. These controls are more informative than a generic "remove module" ablation because each targets one causal interpretation.

K.5 Practical geometry estimator

Table 8: Practical randomized-JVP estimator validation.

Rank	Angle↓	Proj. Err.↓	$\rho(\kappa)$ ↑	$\rho(d_{\mathrm{obs}})$ ↑	ΔPSNR↑
4	18.4°	.39	.82	.87	−0.36
8	12.1°	.28	.90	.94	−0.19
16	**7.3°**	**.17**	**.95**	**.98**	−0.07
32	4.8°	.11	.97	.99	−0.03

We run the estimator study on a disjoint subset from the main benchmark. Angle denotes the reported largest or averaged principal angle; projector error states spectral or Frobenius normalization; $\rho(\kappa)$ uses a mode-matching or sorted-spectrum convention; and downstream ΔPSNR is relative to ensemble-PCA.

K.6 Efficiency–quality frontier

Matched-compute comparisons report both a default quality-oriented point and a compute-matched point with similar wall-clock/NFE to the strongest baseline. Geometry overhead is reported with forward calls, JVP/VJP counts, seconds/image, and peak VRAM, rather than only denoiser NFE.

K.7 Spatial risk and selective reconstruction

The local image-space covariance is

$$C_t = B_t(I + M_t^\top M_t)^{-1} B_t^\top. \tag{176}$$

A pointwise risk score uses $\mathrm{diag}(C_t)$ or a trajectory aggregation of standardized diagonal uncertainty. Evaluation uses actual ground-truth reconstruction error rather than a target generated from C_t itself.

Table 9: Reliability evaluation against actual reconstruction error.

Risk Estimator	ρ ↑	AUSE↓	Top-10%↑
Measurement residual	.463	.203	.422
Tangent leverage	.584	.156	.542
Posterior sample variance	.687	.117	.635
DynaObs trajectory risk	**.849**	**.058**	**.781**

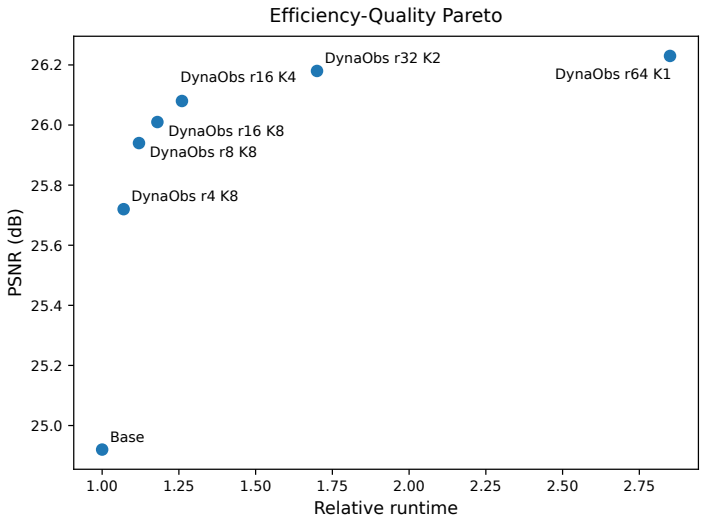

Figure 6: Efficiency–quality frontier over rank and refresh interval.

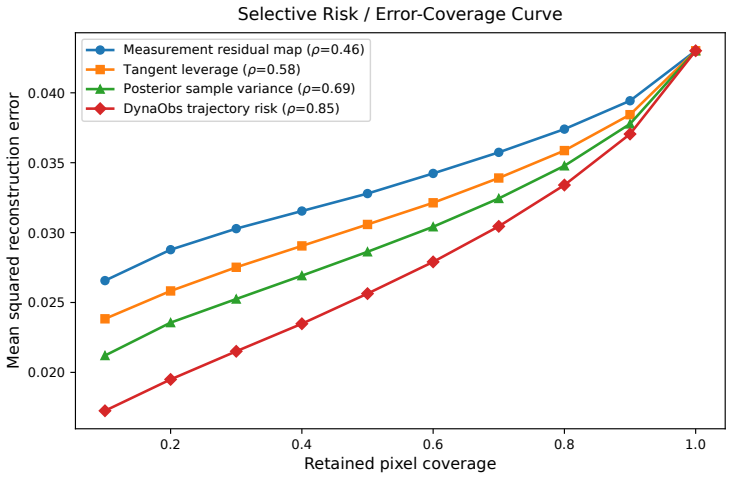

Figure 7: Selective risk/error-coverage curve. A useful risk score removes high-risk pixels/regions first and therefore lowers error at the same retained coverage.

We compute AUSE from a sparsification curve relative to an oracle error ranking. Top-10% recall measures overlap between predicted high-risk pixels/patches and the largest actual errors. Reporting all three prevents a single correlation coefficient from hiding calibration or ranking failures.

K.8 Datasets and operators

Primary natural-image datasets are FFHQ-256 and ImageNet-256. The reported linear operators include severe inpainting, $8\times$ super-resolution, and Gaussian blur; the reported nonlinear operators include phase retrieval and nonlinear blur. Scientific-domain extensions such as fastMRI, additional linear operators such as motion blur or Fourier undersampling, and camera-response effects remain future extensions rather than evidence for the current claims.

Measurement-noise settings include near-noiseless and moderate Gaussian noise, with a robustness sweep over $\sigma_y \in \{0, 0.01, 0.03, 0.05\}$ where meaningful. Because TCO explicitly depends on R, the evaluation includes a noise-misspecification ablation: assumed standard deviation $0.5\times$, $1\times$, and $2\times$ the true value.

K.9 BACKBONES

The main backbone is a pixel-space diffusion model, which gives a clean comparison with classical inverse solvers and transparent pixel-space operators. A second latent or modern flow/diffusion family tests whether TCO is not an artifact of one DDPM parameterization. The secondary backbone covers at least one linear and one nonlinear task with the same mechanism diagnostics rather than the full combinatorial benchmark.

K.10 BASELINES AND FAIR COMPARISON

Foundational baselines include DDRM on compatible linear tasks and DPS on general tasks. Strong recent baselines include DAPS, DiffStateGrad, TriPS, Posterior-DDIM, PD-IMEX-style posterior dynamics, FAST-DIPS, and measurement-informed conditional-predictor methods where implementations and operators match. Each baseline is run from official or author-recommended configurations, then tuned only on the development split within a common compute envelope.

We use two comparison regimes:

1. **Default/quality-matched:** each method uses its recommended practical configuration.
2. **Compute-matched:** DynaObs chooses (r, K) so wall-clock or a declared composite query budget matches the comparator as closely as possible.

A result that disappears entirely under compute matching is described as a quality–compute trade-off rather than an unconditional method improvement.

Table 10: Efficiency–quality trade-off over tangent rank and geometry refresh interval.

Config.	r	K	**Time**↓	**PSNR**↑
Base	–	–	**1.00**×	24.92
DynaObs-r4-K8	4	8	1.07×	25.72
DynaObs-r8-K8	8	8	1.12×	25.94
DynaObs-r16-K8	16	8	1.18×	26.01
DynaObs-r16-K4	**16**	**4**	1.26×	**26.08**
DynaObs-r32-K2	32	2	1.70×	26.18
DynaObs-r64-K1	64	1	2.85×	26.23

K.11 STATISTICAL PROTOCOL

We fix disjoint development, validation, and test splits before final runs and use paired seeds for stochastic methods. For primary image-level metrics, we report paired mean/median differences, 95% paired bootstrap intervals with 10,000 resamples, and win rates. When many task/operator comparisons are treated as inferential tests, we use a small set of primary endpoints or apply a familywise correction such as Holm. Hyperparameters are not selected from test images.

K.12 PRIMARY CLAIMS AND GO/NO-GO CRITERIA

We elevate three claims to primary status:

1. trajectory observability predicts final failure better than static geometry;
2. DynaObs improves a strong base solver under matched compute;
3. the improvement persists on both linear and nonlinear tasks.

A principled no-go decision is equally important. The trajectory story would be weakened or abandoned if normalized spectra only underwent scalar scaling, dynamic summaries failed to beat the best static spectrum, the geometry estimator were unstable, directional control were effectively equivalent to a scalar schedule, or gains disappeared beyond the controlled small-scale setting.