# GEOMETRY-AWARE DIFFUSION GUIDANCE VIA CURVATURE-ADAPTIVE TUBULAR CORRECTION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Gradient-guided diffusion samplers provide flexible priors for inverse problems and conditional generation, but strong guidance can move the sampling trajectory into regions where the learned score is poorly supported. Existing tangent-projection strategies limit first-order departure from an iso-density surface, yet discard potentially useful normal motion and overlook the second-order departure induced by tangent motion on a curved surface. We introduce curvature-adaptive tubular correction (CAT), a training-free plugin that regulates both effects within a shared, noise-dependent geometric budget. CAT decomposes the guidance gradient into normal and tangent components, charges normal displacement at first order and tangent displacement according to directional curvature, and obtains their jointly optimal magnitudes from a one-dimensional dual equation. Armijo backtracking calibrates the resulting finite step against the actual guidance objective, while matrix-free directional derivatives avoid constructing the full score Jacobian. We establish local guarantees for the tubular approximation, uniqueness of the correction, and sufficient objective decrease. Across seven inverse problems on FFHQ and ImageNet, CAT improves the evaluated pixel- and latent-space host samplers, with particularly consistent gains in perceptual metrics. It also improves black hole reconstruction on InverseBench. These results support curvature-aware tubular control as a reusable mechanism for stabilizing diffusion guidance.

## 1 INTRODUCTION

Diffusion models provide reusable priors for recovering images from incomplete, blurred, or nonlinear measurements without task-specific retraining (Ho et al., 2020; Song et al., 2021; Kawar et al., 2022; Wang et al., 2023). This flexibility has enabled applications ranging from PDE-constrained and full-waveform inversion to accelerated MRI and volumetric reconstruction (Jiang et al., 2025; Peng et al., 2026; Min & Ma, 2026; Chung & Ye, 2022; Chung et al., 2023b). Gradient-guided posterior samplers exploit such priors by inserting a differentiable measurement objective into reverse diffusion (Chung et al., 2023a), but must enforce observations without driving samples away from regions well represented by the learned score.

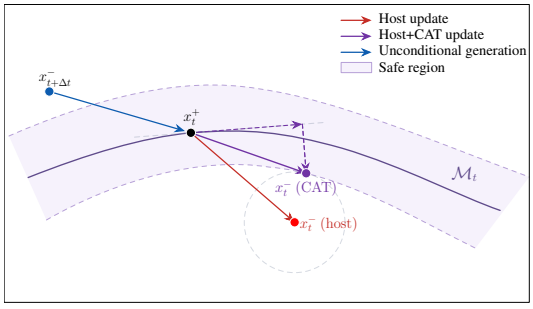



Figure 1: CAT within one reverse-diffusion step. The host update can leave the safe region, whereas CAT uses curvature-adaptive normal–tangent allocation to produce $x_t^-$ inside it.

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.

This challenge is geometric as well as numerical. Because score training concentrates on high-probability noisy-data regions, aggressive measurement updates can enter poorly supervised regions and accumulate score error. Manifold-constrained methods mitigate this risk by restricting updates (Chung et al., 2022; He et al., 2024), while score-based projection removes the score-parallel component to obtain a locally tangent direction (Hamidi & Yang, 2025). Exact tangency, however, discards potentially useful normal guidance and controls surface departure only to first order: on a curved surface, a finite tangent step still induces second-order normal displacement. Thus, neither uniform gradient scaling nor hard tangent projection controls both effects.

We introduce *curvature-adaptive tubular correction* (CAT), a training-free plugin that permits guidance within a noise-dependent tube around the local iso-density surface. CAT decomposes the guidance gradient into normal and tangent components, charging normal displacement at first order and tangent displacement at second order according to directional curvature. A convex allocation problem determines their unique jointly optimal magnitudes through a scalar dual equation, while Armijo backtracking calibrates the finite step against the actual guidance objective. CAT evaluates directional curvature with a matrix-free derivative and operates in the host solver's working variable, making the same correction applicable in pixel and latent spaces. Figure 1 illustrates this geometry.

Across seven inverse problems on FFHQ and ImageNet, CAT consistently improves the evaluated pixel- and latent-space hosts, particularly in perceptual quality, and remains robust to guidance strength, tubular radius, and measurement noise. Ablations verify that second-order tubular control and finite-step calibration are complementary. CAT also transfers to scientific imaging, improving black hole reconstruction.

## 2 Preliminaries

### 2.1 Diffusion Models

Diffusion models represent the data prior through a Gaussian perturbation process $x_t = \mu_t x_0 + \sigma_t \epsilon$, where $x_0, x_t \in \mathbb{R}^D$ and $\epsilon \sim \mathcal{N}(0, I_D)$, and learn a score model $s_\theta(x,t) \approx \nabla_x \log p_t(x)$ to reverse this process (Ho et al., 2020; Song et al., 2021). For the continuous-time forward SDE $dx = f(x,t)dt + g(t)dw_t$, the reverse drift is $f(x,t) - g(t)^2 \nabla_x \log p_t(x)$; replacing the exact score with $s_\theta$ yields a practical generative sampler.

### 2.2 Diffusion Posterior Sampling

Diffusion posterior sampling (DPS) recovers $x_0$ from measurements $y = A(x_0) + \xi$, where $A : \mathbb{R}^D \to \mathbb{R}^m$ is a known differentiable operator and $\xi \sim \mathcal{N}(0, \sigma_y^2 I_m)$. It augments the pretrained prior score with measurement guidance through $\nabla_{x_t} \log p_t(x_t \mid y) = \nabla_{x_t} \log p_t(x_t) + \nabla_{x_t} \log p_t(y \mid x_t)$ (Chung et al., 2023a). Because the conditional likelihood at time $t$ is generally intractable, DPS evaluates it at the Tweedie estimate, $p_t(y \mid x_t) \approx p(y \mid \widehat{x}_{0,\theta}(x_t, t))$. Under Gaussian measurement noise, this yields the differentiable guidance objective

$$L_t(x) = \frac{1}{2\sigma_y^2} \left\| A(\widehat{x}_{0,\theta}(x,t)) - y \right\|_2^2 . \tag{1}$$

At each reverse time, DPS first produces a prior anchor and then applies a measurement-gradient correction:

$$x_t^+ = \Psi^{\text{prior}}_{t \leftarrow t+\Delta t}(x^-_{t+\Delta t}), \qquad q_t = \nabla_x L_t(x_t^+), \qquad x_t^- = x_t^+ - \alpha_t q_t, \tag{2}$$

where $\alpha_t > 0$ controls the guidance strength. Standard DPS scales every component of $q_t$ uniformly, without distinguishing first-order normal departure from curvature-induced tangent departure. CAT introduces this geometric distinction while retaining the host sampler's reverse transition.

## 3 Our Method

### 3.1 Motivation

Strong measurement guidance can move diffusion trajectories beyond regions well represented during score training, leading to accumulated prior error and visible texture or color artifacts. The

concentration result of Chung et al. (2022) characterizes the noisy-data region underlying this behavior.

**Proposition 1** (Concentration of noisy data (Chung et al., 2022)). *Suppose the clean distribution is uniform on a locally linear manifold $\mathcal{M} \subset \mathbb{R}^D$ of intrinsic dimension $d_{\mathcal{M}} < D$. Under $x_t = \mu_t x_0 + \sigma_t \epsilon$, the noisy marginal concentrates around the $(D-1)$-dimensional shell*

$$\mathcal{S}_t = \left\{ x \in \mathbb{R}^D : \operatorname{dist}(x, \mu_t \mathcal{M}) = \sigma_t \sqrt{D - d_{\mathcal{M}}} \right\}. \tag{3}$$

Because score training draws noisy states from $p_t$, supervision is concentrated near $\mathcal{S}_t$, while aggressive guidance can carry an iterate into less reliable regions and amplify score error over subsequent reverse steps. CAT directly addresses this failure mode by regulating each correction within a noise-dependent tube around the local prior geometry. Unlike exact manifold projection, this tubular design preserves useful measurement motion while controlling both normal displacement and curvature-induced tangent departure.

### 3.2 Curvature-adaptive tubular (CAT) correction

For inverse problems, CAT targets gradient-based posterior samplers and requires an explicit measurement loss $L_t$ that is differentiable with respect to the working variable $x_t$, providing the guidance gradient $q_t = \nabla L_t(x_t)$ in either pixel or latent space. Here, $D$ denotes the dimension of the current working space. We present CAT on top of DPS, with $x_t$ denoting the prior anchor $x_t^+$ in (2) and $\alpha_t > 0$ the host guidance step size. The same construction transfers directly to other gradient-guided samplers.

CAT uses the iso-density surface through the current state to describe local prior geometry. Let $\phi_t = \log p_t$, $c_t := \phi_t(x_t) = \log p_t(x_t)$, and $\mathcal{M}_t := \{x \in \mathbb{R}^D : \phi_t(x) = c_t\}$. The surface $\mathcal{M}_t$ is updated at every reverse step. When $\phi_t$ is $C^3$ with nonzero gradient near $x_t$, its local component is a regular hypersurface oriented by the score-induced unit normal $\nu_t(z) := \nabla\phi_t(z)/\|\nabla\phi_t(z)\|_2$. Let $\mathcal{U}_t$ be a tubular neighborhood with a unique closest-point projection onto $\mathcal{M}_t$. Its signed distance, oriented by $\nu_t$, is

$$r(x, \mathcal{M}_t) := \begin{cases} +\operatorname{dist}(x, \mathcal{M}_t), & x \text{ lies on the side pointed to by } \nu_t, \\ -\operatorname{dist}(x, \mathcal{M}_t), & x \text{ lies on the opposite side,} \end{cases} \tag{4}$$

where $\operatorname{dist}(x, \mathcal{M}_t) = \inf_{z \in \mathcal{M}_t} \|x - z\|_2$, so $|r(x, \mathcal{M}_t)| = \operatorname{dist}(x, \mathcal{M}_t)$. For $z \in \mathcal{M}_t$, we have $r(z, \mathcal{M}_t) = 0$ and $\nabla_x r(z, \mathcal{M}_t) = \nu_t(z)$. Assuming $r$ is $C^3$ on $\mathcal{U}_t$, the local score geometry at $x_t$ is

$$\begin{aligned} &s_t = \nabla\phi_t(x_t), \quad g_t = \|s_t\|_2, \quad H_t = \nabla^2\phi_t(x_t), \\ &\nu_t := \nu_t(x_t) = \frac{s_t}{g_t}, \quad P_{N,t} = \nu_t \nu_t^\top, \quad P_{T,t} = I - P_{N,t}, \end{aligned} \tag{5}$$

where $P_{N,t}$ and $P_{T,t}$ project onto the score-normal and tangent spaces, respectively, and $T_{x_t}\mathcal{M}_t = \{v : s_t^\top v = 0\}$. CAT permits guidance within the noise-adaptive tube $\mathcal{T}_t(R_t) := \{x \in \mathcal{U}_t : |r(x, \mathcal{M}_t)| \le R_t\}$ with $R_t = \rho\sigma_t$, where $\rho > 0$ controls the geometric budget. Scaling the radius with $\sigma_t$ follows the width of the noisy distribution and progressively tightens guidance during denoising.

**Theorem 2** (Second-order tubular expansion). *Assume the preceding local regularity conditions and $\sup_{z \in \mathcal{U}_t} \|D^3 r(z, \mathcal{M}_t)\|_{\mathrm{op}} \le M_3$. The local curvature matrix is*

$$B_t = \frac{1}{g_t} P_{T,t} H_t P_{T,t}. \tag{6}$$

*For every sufficiently local displacement $d = d_N + d_T$, where $d_N = P_{N,t} d$ and $d_T = P_{T,t} d$,*

$$r(x_t + d, \mathcal{M}_t) = \nu_t^\top d_N + \frac{1}{2} d_T^\top B_t d_T + R_3(d), \qquad |R_3(d)| \le \frac{M_3}{6} \|d\|_2^3. \tag{7}$$

Theorem 2 reveals the key asymmetry exploited by CAT: normal motion consumes tube thickness at first order, whereas tangent motion is charged at second order according to directional curvature. For component magnitudes $r_N$ and $r_T$ along a unit tangent direction $u$, let $K_t = |u^\top B_t u|$. The resulting

non-canceling budget $r_N + \frac{1}{2}K_t r_T^2 \le R_t$ conservatively controls the second-order signed-distance model without allowing the two components to cancel. Appendix B.1 provides the derivation.

We next allocate the shared budget optimally between the normal and tangent guidance components. Decompose $q_{N,t} = P_{N,t}q_t$ and $q_{T,t} = P_{T,t}q_t$, with magnitudes $a_t = \|q_{N,t}\|_2$ and $b_t = \|q_{T,t}\|_2$. For each nonzero component, define $e_{N,t} = q_{N,t}/a_t$ and $e_{T,t} = q_{T,t}/b_t$; vanishing components receive zero displacement. We seek a descent correction $d_t = -r_{N,t}e_{N,t} - r_{T,t}e_{T,t}$ and measure tangent cost by the directional curvature $K_t = |e_{T,t}^\top B_t e_{T,t}|$, with $K_t = 0$ when $b_t = 0$. The optimal magnitudes solve

$$\begin{aligned} \min_{r_{N,t}, r_{T,t} \ge 0} \quad & -a_t r_{N,t} - b_t r_{T,t} + \frac{r_{N,t}^2 + r_{T,t}^2}{2\alpha_t} \\ \text{subject to} \quad & r_{N,t} + \frac{1}{2}K_t r_{T,t}^2 \le R_t. \end{aligned} \tag{8}$$

The objective stays close to the host update while maximizing first-order measurement descent. Without the tube constraint, its solution $(r_{N,t}, r_{T,t}) = (\alpha_t a_t, \alpha_t b_t)$ exactly recovers the original isotropic guidance step.

**Theorem 3** (Unique curvature-adaptive step)**.** *For $\alpha_t, R_t > 0$ and $a_t, b_t, K_t \ge 0$,* (8) *has a unique global minimizer given by $r_{N,t}(\lambda_t) = \alpha_t(a_t - \lambda_t)_+$ and $r_{T,t}(\lambda_t) = \alpha_t b_t/(1 + \alpha_t\lambda_t K_t)$, where $\lambda_t = 0$ if $\alpha_t a_t + \frac{1}{2}K_t(\alpha_t b_t)^2 \le R_t$. Otherwise, $\lambda_t$ is the unique positive root of*

$$F_t(\lambda_t) := \alpha_t(a_t - \lambda_t)_+ + \frac{K_t}{2}\left(\frac{\alpha_t b_t}{1 + \alpha_t\lambda_t K_t}\right)^2 = R_t. \tag{9}$$

The scalar equation is inexpensive to solve independently for each sample using bracketing and bisection. The resulting component-wise step sizes are

$$\eta_{N,t} = \begin{cases} \alpha_t\left(1 - \frac{\lambda_t}{a_t}\right)_+, & a_t > 0, \\ 0, & a_t = 0, \end{cases} \qquad \eta_{T,t} = \begin{cases} \frac{\alpha_t}{1+\alpha_t\lambda_t K_t}, & b_t > 0, \\ 0, & b_t = 0. \end{cases} \tag{10}$$

giving $d_t = -\eta_{N,t}q_{N,t} - \eta_{T,t}q_{T,t}$. Thus, CAT adapts the normal component to its first-order tube cost and the tangent component to its directional curvature through a single scalar multiplier. Appendix B.2 provides the proof.

**Theorem 4** (Measurement descent and Armijo acceptance)**.** *Suppose $\nabla L_t$ is $L_g$-Lipschitz along the update segment, and let $d_t^*$ solve* (8)*. If $0 < \alpha_t \le 1/L_g$, then $L_t(x_t + d_t^*) \le L_t(x_t)$, with strict inequality when $q_t \ne 0$. For any $\alpha_t > 0$ and $q_t \ne 0$, $d_t^*$ remains a strict descent direction, and Armijo backtracking with $c, \beta \in (0,1)$ terminates at a scale $\gamma_t \in \{1, \beta, \beta^2, \ldots\}$ satisfying*

$$L_t(x_t + \gamma_t d_t^*) \le L_t(x_t) + c\gamma_t q_t^\top d_t^* < L_t(x_t). \tag{11}$$

*Moreover, the accepted displacement remains second-order feasible: $\gamma_t r_{N,t}^* + \frac{1}{2}K_t\gamma_t^2(r_{T,t}^*)^2 \le R_t$.*

Theorem 4 couples geometric control with reliable finite-step optimization. Rather than estimating the trajectory-dependent constant $L_g$, CAT applies Armijo backtracking to each curvature-adaptive proposal. The accepted update achieves sufficient decrease in the actual guidance objective while preserving second-order tubular feasibility. The theorem concerns backtracking without a fixed trial cap; our implementation tests at most $J_{\max}$ candidate scales and rejects the guidance update if all fail, as detailed in Appendix C. To reduce the cost of repeated objective evaluations, we additionally introduce an Armijo period $p$: backtracking is performed once every $p$ reverse steps, and the most recently accepted scale $\gamma_t$ is reused at the intervening steps. Increasing $p$ therefore amortizes the line-search overhead and provides a controllable efficiency–quality trade-off, which we evaluate in Section 4.4. Appendix B.3 gives the proof of the per-step acceptance result.

Combining curvature-adaptive allocation with finite-step calibration, CAT replaces the host's isotropic correction by

$$x_t^- = x_t + \gamma_t d_t = x_t - \gamma_t\left(\eta_{N,t}q_{N,t} + \eta_{T,t}q_{T,t}\right). \tag{12}$$

Here the component-wise step sizes $\eta_{N,t}$ and $\eta_{T,t}$ are defined in (10), while the Armijo scale $\gamma_t$ is selected according to (11). CAT therefore preserves the host prior transition and adapts only its guidance correction.

In practice, CAT uses the learned score $\widehat{s}_t = s_\theta(x_t, t)$ to form $\widehat{\nu}_t = \widehat{s}_t / \|\widehat{s}_t\|_2$ and directly decomposes the guidance as $\widehat{q}_{N,t} = \widehat{\nu}_t(\widehat{\nu}_t^\top q_t)$ and $\widehat{q}_{T,t} = q_t - \widehat{q}_{N,t}$. CAT evaluates only the directional curvature needed by the update. For $\widehat{b}_t = \|\widehat{q}_{T,t}\|_2 > 0$, define

$$u_t = \mathrm{stopgrad}\left(\frac{\widehat{q}_{T,t}}{\widehat{b}_t}\right), \qquad v_t = D_{x_t}\left(u_t^\top \widehat{s}_t\right), \qquad \widehat{K}_t = \frac{|u_t^\top v_t|}{\|\widehat{s}_t\|_2}. \tag{13}$$

Here $D_{x_t}$ denotes reverse-mode differentiation with respect to $x_t$. Because $u_t$ is stop-gradient, this evaluates the required quadratic form in one backward pass using only $D$-dimensional auxiliary vectors. If $\widehat{b}_t = 0$, we set $\widehat{K}_t = 0$ and skip the pass.

The complete CAT procedure decomposes the guidance gradient using the learned score, evaluates (13), solves the scalar dual equation, and calibrates the proposal by Armijo backtracking. A complete implementation summary is provided in Algorithm 1 in Appendix C.

# 4 Experiment

## 4.1 Main Results

**Evaluation protocol.** We evaluate on 100 images from FFHQ and 100 images from ImageNet at the native resolution of $256 \times 256$. The benchmark covers random-mask and box inpainting, super-resolution, motion and Gaussian deblurring, high-dynamic-range (HDR) reconstruction, and nonlinear deblurring. Complete operator and measurement-noise settings are provided in Appendix C. We report PSNR, SSIM, LPIPS, and FID; higher PSNR and SSIM are better, whereas lower LPIPS and FID are better.

**Methods and plugin protocol.** We therefore attach CAT correction to multiple host samplers and report the resulting pairs separately: DPS (Chung et al., 2023a) and DPS+CAT in pixel space, and PSLD (Rout et al., 2023)/PSLD+CAT and ReSample (Song et al., 2024)/ReSample+CAT in latent space. This design tests whether the proposed geometric correction transfers across solver families. We additionally include DiffState-enhanced variants (Zirvi et al., 2025) of the same hosts, together with the pixel-space baselines DPS, DDRM (Kawar et al., 2022), DDNM (Wang et al., 2023), DAPS (Zhang et al., 2025), DCDP (Li et al., 2026), and DiffRGD (Liao et al., 2026). The algorithm hyperparameter settings used in these experiments are provided in Appendix C.

**Quantitative comparison.** Table 1 summarizes all seven tasks on both FFHQ and ImageNet. On FFHQ, integrating CAT consistently improves the underlying solvers across all seven inverse problems, with particularly strong reductions in LPIPS and FID, indicating better perceptual fidelity. These gains extend to ImageNet, where DPS+CAT achieves the strongest perceptual results among the pixel-space methods across all reported tasks. ReSample+CAT delivers the strongest overall latent-space results on the five linear tasks. The consistent improvements across DPS, PSLD, and ReSample demonstrate that CAT serves as an effective plug-and-play geometric correction rather than a solver-specific alternative. Figure 2 illustrates these gains on super-resolution and nonlinear deblurring: DPS+CAT better preserves facial structure and fine details while suppressing local artifacts. Figure 7 in Appendix E provides the complete qualitative comparison across six representative tasks.

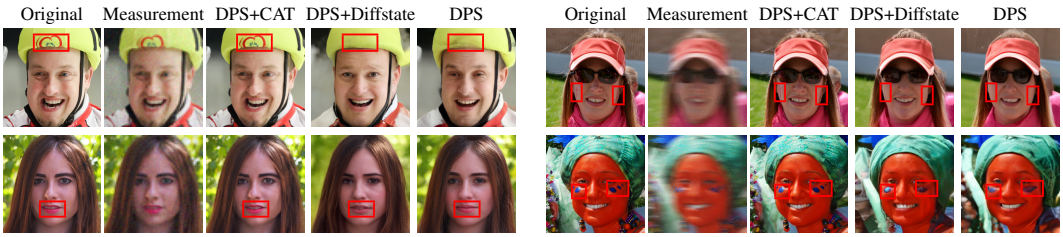
## 4.2 Robustness Analysis

We evaluate the robustness of CAT along three sources of sensitivity: the host guidance step size, the tubular radius, and the measurement-noise level. All sweeps use FFHQ motion deblurring and aggregate 100 images at each setting.

**Robustness to the guidance step size.** The guidance step size controls the strength of the measurement-consistency updates applied by the host sampler. Although stronger guidance can improve data fidelity, an excessively large step may push the iterate away from regions well modeled by the learned prior. To assess this sensitivity, we sweep the PSLD guidance weight $\omega \in \{0.1, 0.5, 1, 2, 5\}$ on FFHQ motion deblurring, holding the latent data-consistency weight fixed at $\gamma_{\mathrm{PSLD}} = 0.01$ and the measurement-noise level fixed at $\sigma_y = 0.05$. All other hyperparameters

Table 1: Main results on FFHQ and ImageNet using 100 images per dataset. Higher PSNR/SSIM and lower LPIPS/FID are better. **Bold** denotes the best result within each task and type.

| Task | Type | Method | FFHQ | | | | ImageNet | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | PSNR ↑ | SSIM ↑ | LPIPS ↓ | FID ↓ | PSNR ↑ | SSIM ↑ | LPIPS ↓ | FID ↓ |
| Super-resolution 4× | *Pixel* | DAPS | 28.07 | 0.749 | 0.2773 | 76.15 | 25.54 | 0.642 | 0.3759 | 113.16 |
| | | DCDP | 26.55 | 0.765 | 0.3074 | 103.17 | 25.80 | 0.681 | 0.3808 | 128.53 |
| | | DDNM | **28.80** | **0.822** | 0.2435 | 80.38 | **26.84** | **0.743** | 0.3083 | 102.82 |
| | | DDRM | 28.47 | 0.814 | 0.2511 | 84.87 | 26.43 | 0.730 | 0.3175 | 99.85 |
| | | DiffRGD | 27.49 | 0.778 | 0.2266 | 64.60 | 23.96 | 0.625 | 0.3687 | 121.27 |
| | | DPS | 24.01 | 0.672 | 0.3176 | 88.23 | 22.09 | 0.542 | 0.4486 | 166.28 |
| | | DPS+Diffstate | 22.67 | 0.629 | 0.3475 | 90.20 | 20.85 | 0.494 | 0.4933 | 192.38 |
| | | DPS+CAT (ours) | 27.58 | 0.780 | **0.2132** | **54.63** | 25.67 | 0.691 | **0.2977** | **82.44** |
| | *Latent* | PSLD | 23.91 | 0.610 | 0.3973 | 134.91 | 22.62 | 0.531 | 0.4974 | 265.51 |
| | | PSLD+DiffState | 24.53 | 0.649 | 0.3444 | 94.91 | 21.20 | 0.474 | 0.5525 | 299.54 |
| | | PSLD+CAT (ours) | **27.16** | **0.754** | **0.2497** | **71.01** | 23.43 | 0.575 | 0.4487 | 233.00 |
| | | ReSample | 23.25 | 0.585 | 0.4635 | 164.56 | 24.42 | 0.637 | 0.3618 | 150.49 |
| | | ReSample+Diffstate | 21.31 | 0.436 | 0.5914 | 291.91 | 23.36 | 0.570 | 0.4467 | 235.14 |
| | | ReSample+CAT (ours) | 26.80 | 0.720 | 0.3719 | 124.63 | **25.41** | **0.694** | **0.3538** | **138.28** |
| Inpainting (box) | *Pixel* | DAPS | 24.46 | 0.752 | 0.2128 | 54.73 | **20.94** | 0.711 | 0.2754 | 97.58 |
| | | DCDP | 22.87 | 0.750 | 0.3009 | 103.14 | 19.93 | 0.757 | 0.2694 | 122.64 |
| | | DDNM | **24.51** | **0.860** | 0.1553 | 48.08 | 20.74 | **0.802** | 0.2172 | 99.90 |
| | | DDRM | 22.18 | 0.818 | 0.2117 | 72.35 | 19.14 | 0.759 | 0.2685 | 121.00 |
| | | DiffRGD | 23.33 | 0.839 | 0.1664 | 47.21 | 19.30 | 0.707 | 0.3359 | 154.77 |
| | | DPS | 23.60 | 0.801 | 0.2201 | 65.92 | 19.97 | 0.695 | 0.3295 | 121.47 |
| | | DPS+Diffstate | 23.23 | 0.736 | 0.2626 | 72.40 | 19.74 | 0.626 | 0.3818 | 137.10 |
| | | DPS+CAT (ours) | 24.35 | 0.849 | **0.1437** | **38.88** | 20.01 | 0.786 | **0.1769** | **64.99** |
| | *Latent* | PSLD | 23.50 | 0.857 | 0.1163 | 41.16 | 19.71 | 0.804 | 0.1956 | 120.25 |
| | | PSLD+DiffState | 23.78 | 0.862 | 0.1140 | 40.54 | 19.74 | 0.804 | 0.2005 | 125.97 |
| | | PSLD+CAT (ours) | **24.25** | 0.876 | **0.0983** | **33.38** | 20.19 | 0.814 | 0.1888 | 115.24 |
| | | ReSample | 20.63 | 0.838 | 0.1444 | 80.55 | 20.79 | 0.822 | 0.1769 | 108.18 |
| | | ReSample+Diffstate | 20.66 | 0.793 | 0.1919 | 141.88 | 19.86 | 0.811 | **0.1766** | 108.59 |
| | | ReSample+CAT (ours) | 23.83 | **0.876** | 0.1079 | 40.42 | **21.10** | **0.827** | 0.1782 | **106.21** |
| Inpainting (random) | *Pixel* | DAPS | 29.44 | 0.784 | 0.2144 | 52.17 | 27.35 | 0.725 | 0.2516 | 48.76 |
| | | DCDP | 27.17 | 0.786 | 0.2838 | 98.72 | 26.67 | 0.743 | 0.3111 | 88.60 |
| | | DDNM | 29.60 | 0.862 | 0.1845 | 59.98 | 27.32 | 0.785 | 0.2394 | 67.16 |
| | | DDRM | 25.65 | 0.768 | 0.2791 | 109.90 | 23.24 | 0.633 | 0.4105 | 170.01 |
| | | DiffRGD | **30.88** | **0.889** | 0.1329 | 32.88 | 28.71 | 0.822 | 0.2415 | 72.51 |
| | | DPS | 28.08 | 0.804 | 0.2362 | 72.72 | 26.17 | 0.707 | 0.3328 | 105.54 |
| | | DPS+Diffstate | 26.24 | 0.740 | 0.2768 | 80.76 | 24.41 | 0.633 | 0.3962 | 142.86 |
| | | DPS+CAT (ours) | 30.69 | 0.883 | **0.1260** | **29.56** | **29.22** | **0.841** | **0.1781** | **39.52** |
| | *Latent* | PSLD | 23.86 | 0.651 | 0.4148 | 148.66 | 24.93 | 0.699 | 0.4098 | 188.71 |
| | | PSLD+DiffState | 22.64 | 0.600 | 0.4572 | 161.32 | 23.56 | 0.647 | 0.4622 | 245.96 |
| | | PSLD+CAT (ours) | 29.85 | 0.858 | 0.2117 | 69.04 | 25.31 | 0.722 | 0.3842 | 176.52 |
| | | ReSample | 29.61 | 0.849 | 0.1911 | 55.54 | 27.38 | 0.802 | 0.2667 | 85.86 |
| | | ReSample+Diffstate | 29.14 | 0.833 | 0.2338 | 86.92 | 27.03 | 0.776 | 0.3091 | 110.79 |
| | | ReSample+CAT (ours) | **30.23** | **0.866** | **0.1725** | **51.99** | **27.46** | **0.805** | **0.2650** | **85.22** |
| Gaussian deblurring | *Pixel* | DAPS | 28.23 | 0.752 | 0.2687 | 72.33 | 25.95 | 0.662 | 0.3578 | 108.40 |
| | | DCDP | 28.17 | 0.790 | 0.2650 | 81.75 | 25.89 | 0.673 | 0.4067 | 149.88 |
| | | DDNM | **29.19** | **0.828** | 0.2707 | 95.79 | 26.77 | 0.733 | 0.3347 | 115.95 |
| | | DDRM | 28.66 | 0.817 | 0.2499 | 83.63 | **26.82** | **0.735** | 0.3168 | 105.26 |
| | | DiffRGD | 26.46 | 0.750 | 0.2259 | 64.63 | 22.62 | 0.568 | 0.4158 | 141.71 |
| | | DPS | 25.15 | 0.698 | 0.2719 | 73.70 | 22.42 | 0.547 | 0.3834 | 119.95 |
| | | DPS+Diffstate | 24.53 | 0.678 | 0.2890 | 76.71 | 22.50 | 0.547 | 0.3951 | 136.01 |
| | | DPS+CAT (ours) | 27.20 | 0.764 | **0.2064** | **50.69** | 24.86 | 0.657 | **0.2736** | **74.27** |
| | *Latent* | PSLD | 22.96 | 0.561 | 0.3875 | 119.22 | 20.76 | 0.464 | 0.5297 | 270.49 |
| | | PSLD+DiffState | 22.51 | 0.549 | 0.3949 | 109.64 | 20.77 | 0.429 | 0.5379 | 277.55 |
| | | PSLD+CAT (ours) | 25.59 | 0.706 | 0.3141 | 97.68 | 22.01 | 0.519 | 0.5165 | 275.79 |
| | | ReSample | 25.88 | 0.643 | 0.3274 | 94.59 | 24.65 | 0.640 | 0.4351 | 186.59 |
| | | ReSample+Diffstate | 24.45 | 0.565 | 0.3857 | 112.11 | 24.26 | 0.621 | 0.4542 | 203.70 |
| | | ReSample+CAT (ours) | **27.91** | **0.779** | **0.2460** | **75.52** | **24.85** | **0.652** | **0.3981** | **164.85** |
| Motion deblurring | *Pixel* | DAPS | **30.28** | 0.811 | 0.2093 | 56.99 | 28.62 | 0.768 | 0.2487 | 54.31 |
| | | DCDP | 29.53 | 0.808 | 0.2185 | 57.84 | 27.27 | 0.704 | 0.2961 | 72.03 |
| | | DiffRGD | 28.81 | 0.816 | 0.1913 | 53.84 | 23.32 | 0.597 | 0.4284 | 155.59 |
| | | DPS | 26.31 | 0.740 | 0.2532 | 70.15 | 24.55 | 0.643 | 0.3479 | 102.60 |
| | | DPS+Diffstate | 25.33 | 0.706 | 0.2773 | 75.78 | 23.55 | 0.597 | 0.3864 | 128.44 |
| | | DPS+CAT (ours) | 29.80 | **0.833** | **0.1634** | **39.25** | **28.76** | **0.800** | **0.2072** | **48.09** |
| | *Latent* | PSLD | 23.79 | 0.606 | 0.3706 | 117.02 | 21.68 | 0.488 | 0.5275 | 280.95 |
| | | PSLD+DiffState | 23.40 | 0.601 | 0.3695 | 100.68 | 20.33 | 0.433 | 0.5752 | 303.47 |
| | | PSLD+CAT (ours) | 27.05 | 0.751 | 0.2670 | 76.61 | 22.77 | 0.558 | 0.4825 | 253.68 |
| | | ReSample | 28.19 | 0.742 | 0.2589 | 74.53 | 25.34 | 0.675 | 0.4004 | 173.04 |
| | | ReSample+Diffstate | 28.04 | 0.746 | 0.2670 | 85.76 | 24.94 | 0.652 | 0.4246 | 190.48 |
| | | ReSample+CAT (ours) | **29.23** | **0.806** | **0.2109** | **66.22** | **25.45** | **0.684** | **0.3597** | **140.46** |
| HDR | *Pixel* | DAPS | 26.98 | 0.828 | 0.1977 | 53.10 | **26.12** | 0.827 | 0.2051 | 45.60 |
| | | DCDP | **27.09** | **0.842** | 0.2222 | 61.53 | 26.11 | **0.831** | 0.2126 | 50.16 |
| | | DPS | 23.59 | 0.721 | 0.2938 | 82.85 | 17.12 | 0.460 | 0.5375 | 184.88 |
| | | DPS+Diffstate | 24.00 | 0.717 | 0.2931 | 78.95 | 20.34 | 0.525 | 0.5038 | 186.90 |
| | | DPS+CAT (ours) | 25.91 | 0.819 | **0.1664** | **42.80** | 25.93 | 0.822 | **0.1763** | **41.82** |
| | *Latent* | ReSample | 24.98 | 0.769 | 0.2877 | 91.27 | – | – | – | – |
| | | ReSample+Diffstate | **25.67** | 0.757 | 0.2927 | 94.89 | – | – | – | – |
| | | ReSample+CAT (ours) | 25.61 | **0.794** | **0.2561** | **81.15** | – | – | – | – |
| Nonlinear deblurring | *Pixel* | DAPS | 26.25 | 0.710 | 0.2809 | 80.99 | **25.16** | 0.661 | 0.3429 | 119.91 |
| | | DCDP | 24.22 | 0.710 | 0.3182 | 101.08 | 22.17 | 0.531 | 0.4619 | 176.27 |
| | | DPS | 21.97 | 0.627 | 0.3219 | 85.68 | 20.05 | 0.502 | 0.4320 | 169.87 |
| | | DPS+Diffstate | 21.57 | 0.607 | 0.3336 | 83.03 | 18.78 | 0.428 | 0.5075 | 194.37 |
| | | DPS+CAT (ours) | **26.30** | **0.762** | **0.2197** | **55.17** | 24.95 | **0.690** | **0.2583** | **73.02** |
| | *Latent* | ReSample | 26.75 | 0.742 | 0.2790 | 81.67 | – | – | – | – |
| | | ReSample+Diffstate | 25.96 | 0.712 | 0.3191 | 108.60 | – | – | – | – |
| | | ReSample+CAT (ours) | **27.00** | **0.750** | **0.2666** | **74.39** | – | – | – | – |




**(a)** Super-resolution ($\times 4$). **(b)** Nonlinear deblurring.

Figure 2: Qualitative comparison of DPS-based methods on FFHQ for super-resolution and nonlinear deblurring. Each panel shows two examples with the same original image and measurement. DPS+CAT denotes DPS equipped with CAT, while DPS+Diffstate is the corresponding DiffState baseline.

remain unchanged. We evaluate PSLD, PSLD+DiffState, and PSLD+CAT on the same set of 100 images; Figure 3 reports PSNR, SSIM, LPIPS, and FID across the sweep.

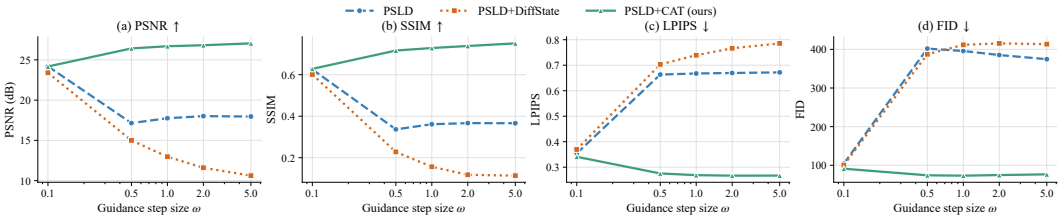



Figure 3: Robustness to the guidance step size $\omega$ on FFHQ motion deblurring. Each point aggregates 100 images.

The quantitative and qualitative results consistently show that CAT substantially reduces sensitivity to the guidance step size. As the guidance becomes more aggressive, PSLD and PSLD+DiffState degrade sharply across the evaluation metrics, whereas PSLD+CAT maintains strong reconstruction quality over a broad operating range. Figure 8 in Appendix E further reveals pronounced structural distortions and color artifacts in the comparison methods at large step sizes, while PSLD+CAT preserves coherent facial geometry and a natural appearance.

**Robustness to the tubular radius.** The coefficient $\rho$ controls the noise-dependent tubular budget $R_t = \rho\sigma_t$: a small value restricts the correction to a narrow neighborhood of the local prior geometry, whereas a larger value permits more measurement-driven motion. We sweep $\rho \in \{0.1, 0.5, 1, 1.5, 2\}$ for PSLD+CAT on FFHQ motion deblurring while retaining the fixed host settings and measurement-noise level $\sigma_y = 0.05$. Every radius is evaluated on the same 100 images. Figure 4 reports the resulting metric curves.

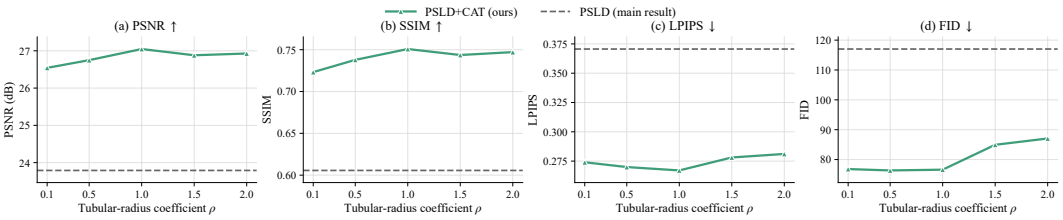



Figure 4: Robustness of PSLD+CAT to the tubular-radius coefficient $\rho$ on FFHQ motion deblurring. Each point aggregates 100 images, and the horizontal dashed line marks the PSLD result reported in the main experiment.

The radius sweep confirms that PSLD+CAT is insensitive to the choice of $\rho$ over a broad operating range. Across all tested radii, PSLD+CAT consistently outperforms the PSLD result from the main experiment, shown by the horizontal dashed line in Figure 4. The distortion metrics remain stable across the sweep, while Figure 9 in Appendix E shows that PSLD+CAT preserves global facial structure more faithfully than PSLD across all tested radii. Overall, PSLD+CAT maintains reliable reconstruction quality without careful tuning of $\rho$, demonstrating robust behavior across a practical range of tubular radii.

**Robustness to measurement noise.** Measurement noise makes data-consistency guidance less reliable by perturbing the observations used throughout posterior sampling. We evaluate this sensitivity by varying $\sigma_y \in \{0, 0.01, 0.05, 0.1\}$ while fixing the motion kernel and all solver hyperparameters. We compare DAPS, DCDP, DiffRGD, DPS, DPS+Diffstate, and DPS+CAT on the same 100 FFHQ images. Figure 5 reports one curve per method for each metric.

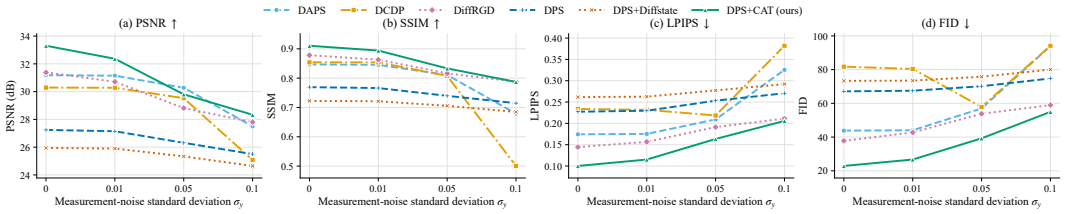



Figure 5: Robustness to the measurement-noise standard deviation $\sigma_y$ on FFHQ motion deblurring. Each point aggregates 100 images.

The noise sweep demonstrates the strong robustness of DPS+CAT to measurement corruption. As the observations become increasingly unreliable, DPS+CAT consistently delivers the best perceptual quality while retaining highly competitive distortion metrics, indicating that CAT remains effective even under substantial measurement perturbations. Figure 10 in Appendix E reinforces this advantage: DPS+CAT faithfully preserves facial boundaries and high-contrast details such as eyeglass frames, while avoiding the noise amplification and oversmoothing observed in several baselines. Together, these results show that CAT provides stable, high-quality reconstructions across a broad range of noise levels without relying on narrowly tuned measurement conditions.

### 4.3 Ablation Studies

We evaluate six diagnostic variants of CAT on FFHQ, using DPS for nonlinear deblurring and PSLD for motion deblurring. Host applies strong guidance without CAT, whereas Host+Armijo adds only CAT's backtracking line search. First-order CAT disables the curvature term by setting $\hat{K}_t = 0$, Tangent-only removes the normal guidance component, and CAT w/o Armijo disables backtracking by fixing $\gamma_t = 1$. Full CAT retains both the second-order tubular constraint and adaptive line search. The measurement-noise level is fixed at $\sigma_y = 0.05$ throughout.

Table 2: Ablation of CAT on FFHQ using DPS for nonlinear deblurring and PSLD for motion deblurring. Each host and variant are evaluated on 100 images. "1st" and "2nd" denote first- and second-order tube models, respectively. "Projection" denotes tangent hyperplane projection ($r_N = 0$), and "N/T" denotes normal/tangent decomposition. Best results are shown in bold.

| Variant | N/T | Curv. | Tube | Armijo | DPS: Nonlinear deblurring | | | | PSLD: Motion deblurring | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | PSNR ↑ | SSIM ↑ | LPIPS ↓ | FID ↓ | PSNR ↑ | SSIM ↑ | LPIPS ↓ | FID ↓ |
| Host | No | No | – | No | 17.06 | 0.326 | 0.5698 | 214.46 | 17.96 | 0.366 | 0.6719 | 374.57 |
| Host+Armijo | No | No | – | Yes | 24.08 | 0.696 | 0.2751 | 71.71 | 25.74 | 0.695 | 0.3482 | 123.15 |
| First-order CAT | Yes | No | 1st | Yes | 24.26 | 0.700 | 0.2718 | 73.44 | 25.80 | 0.699 | 0.3471 | 121.86 |
| Tangent-only | Yes | No | Projection | Yes | 24.15 | 0.696 | 0.2731 | 75.45 | 25.61 | 0.688 | 0.3504 | 125.81 |
| CAT w/o Armijo | Yes | Yes | 2nd | No | 17.33 | 0.409 | 0.4555 | 192.73 | 19.18 | 0.413 | 0.6243 | 323.10 |
| **Full CAT** | Yes | Yes | 2nd | Yes | **26.30** | **0.762** | **0.2197** | **55.17** | **27.05** | **0.751** | **0.2670** | **76.61** |

**Component ablation.** Armijo backtracking substantially improves both host solvers, but does not explain the full gain of CAT. Full CAT achieves the best overall performance for both hosts, with consistent and non-negligible improvements over Host+Armijo. Conversely, CAT w/o Armijo provides only partial gains, showing that geometric control and finite-step loss calibration play complementary roles. Figure 11 in Appendix E supports this conclusion: Full CAT suppresses the severe artifacts remaining in the partial variants and recovers coherent facial structure.

**Role of second-order geometry.** First-order CAT remains close to Host+Armijo, whereas Full CAT consistently performs better across both hosts. This gap indicates that constraining only first-order normal displacement is insufficient and that modeling curvature-induced tangent departure is important in practice. Tangent-only also trails Full CAT, suggesting that exact tangent projection discards useful measurement motion, while CAT benefits from retaining a controlled normal component.

### 4.4 Efficiency Analysis

Both the number of reverse sampling steps and the Armijo period provide practical controls over CAT's quality–efficiency trade-off. Using fewer reverse steps or performing Armijo searches less frequently reduces runtime while retaining a clear reconstruction advantage over the DPS host, whereas allocating more computation yields the strongest overall quality. Both controls incur only negligible additional memory overhead, making CAT adaptable to different inference budgets without materially increasing memory demand. Appendix D.1 provides the complete protocol and quality–cost sweep.

### 4.5 Extensions

**Scientific inverse problems.** We evaluate whether CAT transfers beyond natural-image restoration by applying it to black hole imaging in InverseBench (Zheng et al., 2025). Across both observation regimes, DPS+CAT achieves the strongest image-space reconstruction quality among the compared methods and recovers sharper, more coherent emission structures than its host. The complete protocol, quantitative comparison, and qualitative results are provided in Appendix D.2.

## 5 Conclusion

This paper introduced CAT, a training-free correction for stabilizing gradient-guided diffusion. CAT allocates a noise-dependent tubular budget between first-order normal displacement and curvature-induced second-order tangent departure, retaining useful normal guidance while limiting movement beyond locally reliable regions of the prior. The allocation has a unique scalar-dual solution, and Armijo backtracking calibrates each finite step against the guidance objective. Its matrix-free implementation applies in both pixel and latent spaces.

Across seven inverse problems on FFHQ and ImageNet, CAT improved the evaluated DPS, PSLD, and ReSample hosts, particularly in perceptual quality. Ablations and robustness sweeps support the complementary roles of curvature-aware control and finite-step calibration. CAT also improved black hole reconstruction over DPS. Extending CAT beyond local geometric control toward globally reliable posterior sampling will be an important direction for future work.

### AI use statement

In this work, we used generative AI tools for assisting in the writing of proofs, proposing or refining hypotheses, helping develop theoretical models or conceptual frameworks, implementing methods, assisting with translation and supporting qualitative and thematic data analysis. We have not used generative AI tools for formulating mathematical claims, providing critical ingredients for proving mathematical claims, designing or providing feedback on research methodology or experiments and interpreting results. Generating synthetic data sets and cleaning or reformatting datasets are not applicable to this work. Additionally, we used generative AI tools for creating or modifying scientific figures or images, drafting parts of a research paper, summarizing or analysing existing literature and editing a research paper to improve readability. We have reviewed all AI-assisted work. All AI-assisted text, code, and figures were manually reviewed and verified by the authors for correctness

and consistency with the reported results. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

### Reproducibility statement

Details supporting the reproducibility of our results, including the algorithms, experimental settings, pretrained model provenance, and hyperparameter configurations, are provided in Appendix C.

## References


Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022.

Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *Advances in neural information processing systems*, 35:25683–25696, 2022.

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

Hyungjin Chung, Dohoon Ryu, Michael T McCann, Marc L Klasky, and Jong Chul Ye. Solving 3d inverse problems using pre-trained 2d diffusion models. In *2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 22542–22551. IEEE, 2023b.

Hyungjin Chung, Jong Chul Ye, Peyman Milanfar, and Mauricio Delbracio. Prompt-tuning latent diffusion models for inverse problems. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 8941–8967. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/chung24b.html.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Shayan Mohajer Hamidi and En-Hui Yang. Score-based manifold projection for diffusion-based inverse problems. In *2025 IEEE International Symposium on Information Theory (ISIT)*, pp. 1–6. IEEE, 2025.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, Zico Kolter, Ruslan Salakhutdinov, et al. Manifold preserving guided diffusion. In *International Conference on Learning Representations*, volume 2024, pp. 44819–44850, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Enze Jiang, Jishen Peng, Zheng Ma, and Xiong-Bin Yan. Ode-dps: Ode-based diffusion posterior sampling for linear inverse problems in partial differential equation. *Journal of Scientific Computing*, 102(3):69, 2025.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in neural information processing systems*, 35:23593–23606, 2022.

Xiang Li, Soo Min Kwon, Shijun Liang, Ismail Alkhouri, Saiprasad Ravishankar, and Qing Qu. Dcdp: Decoupled data consistency via diffusion purification for inverse problems. In *2026 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–15. IEEE, 2026.

Jia-Wei Liao, Li-Xuan Peng, Mei-Heng Yueh, Min Sun, Cheng-Fu Chou, and Jun-Cheng Chen. Diffrgd: An inference-time diffusion guidance through riemannian gradient descent. *arXiv preprint arXiv:2606.28417*, 2026.

Chen Min and Zheng Ma. Decoupled latent optimization of diffusion models for full waveform inversion. *arXiv preprint arXiv:2606.14139*, 2026.

Jishen Peng, Enze Jiang, Zheng Ma, and Xiongbin Yan. Robust physics-guided diffusion for full-waveform inversion. *arXiv preprint arXiv:2603.16393*, 2026.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 10674–10685. IEEE, 2022.

Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, and Sanjay Shakkottai. Solving linear inverse problems provably via posterior sampling with latent diffusion models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=XKBFdYwfRo.

Litu Rout, Yujia Chen, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. Beyond first-order tweedie: Solving inverse problems using latent diffusion. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9472–9481. IEEE, 2024.

Bowen Song, Soo Min Kwon, Zecheng Zhang, Xinyu Hu, Qing Qu, and Liyue Shen. Solving inverse problems with latent diffusion models via hard data consistency. In *International Conference on Learning Representations*, volume 2024, pp. 7624–7654, 2024.

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

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

Xingyu Xu and Yuejie Chi. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=SLnsoaY4u1.

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

Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy Feng, Caifeng Zou, Yu Sun, Nikola Kovachki, Zachary Ross, et al. Inversebench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. In *International Conference on Learning Representations*, volume 2025, pp. 90912–90940, 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 *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1219–1229. IEEE, 2023.

Rayhan Zirvi, Bahareh Tolooshams, and Anima Anandkumar. Diffusion state-guided projected gradient for inverse problems. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=kRBQwlkFSP.

## A Related Work

**Training-free diffusion priors for inverse problems.** A broad class of methods repurposes an unconditional diffusion model as a task-independent image prior and enforces measurements only at inference time. For linear operators, DDRM exploits a spectral decomposition of the forward model, whereas DDNM separates its range and null spaces to impose data consistency during reverse diffusion (Kawar et al., 2022; Wang et al., 2023). DPS instead differentiates a likelihood surrogate

evaluated at the Tweedie estimate, which accommodates noisy nonlinear operators without task-specific training (Chung et al., 2023a); DiffPIR places a diffusion denoiser inside a plug-and-play restoration scheme (Zhu et al., 2023). More recent approaches improve the surrounding sampling mechanism: DPnP alternates a likelihood-only proximal sampler with a diffusion-prior sampler and provides robustness guarantees (Xu & Chi, 2024), while DAPS decouples successive noise levels to permit larger transitions and correct errors accumulated earlier in the trajectory (Zhang et al., 2025). These works primarily determine how the likelihood and diffusion prior form a conditional sampler. Our work addresses a complementary question: once a differentiable measurement gradient is available, how should its normal and tangent components be scaled so that the correction remains in a locally reliable region of the learned prior? Consequently, our correction can be attached to gradient-guided solvers without replacing their underlying reverse transition.

**Geometry-preserving diffusion guidance.** The closest line of work controls conditional updates using the geometry of the diffusion prior. MCG identifies accumulated off-manifold error as a failure mode of diffusion inverse solvers and introduces a manifold-constrained gradient correction (Chung et al., 2022). MPGD develops training-free on-manifold guidance using pretrained autoencoders and shows that a corresponding shortcut preserves the latent manifold for latent diffusion models (He et al., 2024). Score-based manifold projection directly treats the score as the normal to an iso-density surface and removes the score-parallel component of the measurement gradient, thereby producing a first-order tangent update (Hamidi & Yang, 2025). DiffState projects guidance onto a low-rank subspace constructed from intermediate diffusion states (Zirvi et al., 2025), while DiffRGD formulates guidance as Riemannian gradient descent on a distribution-induced manifold (Liao et al., 2026). These approaches stabilize guidance through subspace or manifold geometry. Exact manifold or tangent constraints can discard useful normal motion, while a finite tangent step can still depart from a curved surface at second order. We instead admit both components inside a signed-distance tube: normal motion consumes the budget at first order, tangent motion consumes it through directional curvature, and a scalar dual equation selects their unique jointly optimal scaling. This tubular formulation therefore relaxes exact projection while still preventing unrestricted movement into regions where the learned score is poorly supported.

**Inverse problems with latent diffusion models.** Latent diffusion priors reduce generative cost but make measurement consistency more delicate because the forward operator is composed with a nonlinear decoder. ReSample addresses this difficulty by solving a hard data-consistency subproblem and resampling the resulting estimate back to the noisy latent manifold (Song et al., 2024). P2L jointly adapts the text prompt, latent code, and pixel reconstruction through alternating optimization (Chung et al., 2024), whereas STSL constructs a tractable surrogate for second-order Tweedie information to improve latent posterior sampling (Rout et al., 2024). These methods redesign data consistency or the posterior approximation specifically for latent diffusion. In contrast, our construction is representation-agnostic: the decoder is included in the measurement objective, while the score geometry, gradient decomposition, curvature, and tubular step are all computed with respect to the current working variable. The same correction can thus operate in pixel or latent space without changing its geometric derivation.

# B Proofs of the Main Results

## B.1 Proof of Theorem 2

Write $r_t(x) := r(x, \mathcal{M}_t)$. At the anchor,

$$\phi_t(x_t) - c_t = 0, \qquad D\phi_t(x_t) = s_t^\top, \qquad \operatorname{rank} D\phi_t(x_t) = 1. \tag{14}$$

The regular-level-set theorem therefore gives

$$\dim \mathcal{M}_t = D - 1, \qquad T_{x_t}\mathcal{M}_t \subseteq \ker(s_t^\top). \tag{15}$$

Indeed, for any $C^1$ curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}_t$ with $\gamma(0) = x_t$,

$$0 = \frac{d}{d\tau}\big(\phi_t(\gamma(\tau)) - c_t\big)\Big|_{\tau=0} = s_t^\top \gamma'(0). \tag{16}$$

Since

$$\dim T_{x_t}\mathcal{M}_t = D - 1 = \dim \ker(s_t^\top), \tag{17}$$

(15) is an equality:

$$T_{x_t}\mathcal{M}_t = \ker(s_t^\top) = \{v \in \mathbb{R}^D : s_t^\top v = 0\}. \tag{18}$$

Moreover,

$$\|\nu_t\|_2 = 1, \qquad P_{N,t}^2 = P_{N,t}, \qquad P_{T,t}^2 = P_{T,t}, \qquad P_{N,t}P_{T,t} = 0, \qquad P_{N,t} + P_{T,t} = I. \tag{19}$$

Extend the score and score-norm notation from (5) to nearby points by

$$s_t(x) := \nabla\phi_t(x), \qquad g_t(x) := \|s_t(x)\|_2, \qquad \nu_t(x) = \frac{s_t(x)}{g_t(x)}, \tag{20}$$

so that $s_t(x_t) = s_t$ and $g_t(x_t) = g_t$. For every $v \in \mathbb{R}^D$,

$$Ds_t(x_t)[v] = H_t v. \tag{21}$$

Differentiating $g_t(x)^2 = s_t(x)^\top s_t(x)$ gives

$$\begin{aligned} 2g_t(x_t)Dg_t(x_t)[v] &= 2s_t(x_t)^\top Ds_t(x_t)[v], \\ Dg_t(x_t)[v] &= \frac{s_t^\top H_t v}{g_t} = \nu_t^\top H_t v. \end{aligned} \tag{22}$$

Hence

$$\begin{aligned} D\nu_t(x_t)[v] &= \frac{Ds_t(x_t)[v]}{g_t} - \frac{s_t}{g_t^2}Dg_t(x_t)[v] \\ &= \frac{1}{g_t}\left(H_t v - \nu_t\nu_t^\top H_t v\right) \\ &= \frac{1}{g_t}P_{T,t}H_t v. \end{aligned} \tag{23}$$

Projecting the argument and the result onto the tangent space yields

$$P_{T,t}D\nu_t(x_t)P_{T,t} = \frac{1}{g_t}P_{T,t}H_tP_{T,t} = B_t. \tag{24}$$

Thus, for $u = P_{T,t}u$ and $\|u\|_2 = 1$,

$$\kappa_t(u) = u^\top D\nu_t(x_t)[u] = u^\top B_t u = \frac{u^\top H_t u}{g_t}, \tag{25}$$

which proves (6).

On $\mathcal{M}_t$ and throughout its tubular neighborhood,

$$r_t|_{\mathcal{M}_t} = 0, \qquad \nabla r_t(z) = \nu_t(z) \quad (z \in \mathcal{M}_t), \qquad \|\nabla r_t(x)\|_2^2 = 1. \tag{26}$$

For every $w \in \mathbb{R}^D$,

$$\begin{aligned} 0 &= \frac{1}{2}D\left(\|\nabla r_t(x)\|_2^2\right)[w] \\ &= w^\top\nabla^2 r_t(x)\nabla r_t(x). \end{aligned} \tag{27}$$

Since this holds for all $w$,

$$\nabla^2 r_t(x)\nabla r_t(x) = 0, \qquad \nabla^2 r_t(x_t)\nu_t = 0, \qquad \nu_t^\top\nabla^2 r_t(x_t) = 0. \tag{28}$$

For $v \in T_{x_t}\mathcal{M}_t$, take $\gamma(0) = x_t$ and $\gamma'(0) = v$ in (16). Differentiating $\nabla r_t(\gamma(\tau)) = \nu_t(\gamma(\tau))$ at $\tau = 0$ gives

$$\nabla^2 r_t(x_t)v = D\nu_t(x_t)[v]. \tag{29}$$

Equations (24), (28), and (19) imply

$$\begin{aligned} \nabla^2 r_t(x_t) &= (P_{N,t} + P_{T,t})\nabla^2 r_t(x_t)(P_{N,t} + P_{T,t}) \\ &= P_{T,t}\nabla^2 r_t(x_t)P_{T,t} \quad \text{by (28)} \\ &= \frac{1}{g_t}P_{T,t}H_tP_{T,t} = B_t. \end{aligned} \tag{30}$$

For a displacement $d$ whose segment $\{x_t + \tau d : 0 \le \tau \le 1\}$ lies in $\mathcal{U}_t$, define

$$\psi_d(\tau) := r_t(x_t + \tau d). \tag{31}$$

Then

$$\begin{aligned} \psi_d(0) &= 0, \\ \psi_d'(0) &= \nabla r_t(x_t)^\top d = \nu_t^\top (d_N + d_T) = \nu_t^\top d_N, \\ \psi_d''(0) &= d^\top \nabla^2 r_t(x_t) d = d^\top B_t d = d_T^\top B_t d_T, \\ \psi_d'''(\tau) &= D^3 r_t(x_t + \tau d)[d, d, d]. \end{aligned} \tag{32}$$

Taylor's formula with integral remainder gives

$$\begin{aligned} r_t(x_t + d) &= \psi_d(1) \\ &= \psi_d(0) + \psi_d'(0) + \frac{1}{2}\psi_d''(0) + R_3(d) \\ &= \nu_t^\top d_N + \frac{1}{2} d_T^\top B_t d_T + R_3(d), \end{aligned} \tag{33}$$

where

$$R_3(d) = \frac{1}{2}\int_0^1 (1-\tau)^2 D^3 r_t(x_t + \tau d)[d, d, d]\, d\tau. \tag{34}$$

Therefore,

$$\begin{aligned} |R_3(d)| &\le \frac{1}{2}\int_0^1 (1-\tau)^2 \|D^3 r_t(x_t + \tau d)\|_{\mathrm{op}} \|d\|_2^3\, d\tau \\ &\le \frac{M_3}{2}\|d\|_2^3 \int_0^1 (1-\tau)^2\, d\tau \\ &= \frac{M_3}{6}\|d\|_2^3. \end{aligned} \tag{35}$$

Equations (33) and (35) prove (7).

To obtain the scalar budget, let $r_N = \|d_N\|_2$, write $d_T = r_T u$ with $r_T = \|d_T\|_2$ and $u$ a unit tangent vector, and set $K_t = |u^\top B_t u|$ (with $K_t = 0$ when $d_T = 0$). By the triangle inequality, (33) gives

$$\begin{aligned} |r_t(x_t + d)| &\le |\nu_t^\top d_N| + \frac{1}{2}|d_T^\top B_t d_T| + |R_3(d)| \\ &\le r_N + \frac{1}{2} K_t r_T^2 + \frac{M_3}{6}\|d\|_2^3. \end{aligned}$$

Thus, imposing $r_N + \frac{1}{2}K_t r_T^2 \le R_t$ yields the conservative constraint on the second-order signed-distance model used in (8).

### B.2 Proof of Theorem 3

For notational brevity, set

$$\begin{aligned} m_t(r_N, r_T) &:= -a_t r_N - b_t r_T + \frac{r_N^2 + r_T^2}{2\alpha_t}, \\ c_t(r_N, r_T) &:= r_N + \frac{1}{2} K_t r_T^2 - R_t, \\ \mathcal{C}_t &:= \{(r_N, r_T) \in \mathbb{R}_+^2 : c_t(r_N, r_T) \le 0\}. \end{aligned} \tag{36}$$

Their Hessians are

$$\nabla^2 m_t = \frac{1}{\alpha_t} I_2 \succ 0, \qquad \nabla^2 c_t = \begin{bmatrix} 0 & 0 \\ 0 & K_t \end{bmatrix} \succeq 0. \tag{37}$$

Thus $m_t$ is strictly convex and $\mathcal{C}_t$ is convex. Moreover,

$$m_t(r_N, r_T) = \frac{1}{2\alpha_t}\left\| \begin{bmatrix} r_N \\ r_T \end{bmatrix} - \alpha_t \begin{bmatrix} a_t \\ b_t \end{bmatrix} \right\|_2^2 - \frac{\alpha_t}{2}(a_t^2 + b_t^2), \tag{38}$$

so

$$\lim_{\|(r_N,r_T)\|_2\to\infty} m_t(r_N, r_T) = +\infty. \tag{39}$$

The set $\mathcal{C}_t$ is nonempty because $(0, 0) \in \mathcal{C}_t$, and it is closed because $c_t$ is continuous. For $K_t > 0$, choose

$$0 < \epsilon < \min\left\{\frac{R_t}{2}, \sqrt{\frac{R_t}{K_t}}\right\}; \tag{40}$$

then

$$\epsilon + \frac{1}{2}K_t\epsilon^2 < \frac{R_t}{2} + \frac{R_t}{2} = R_t. \tag{41}$$

For $K_t = 0$, any $0 < \epsilon < R_t$ gives the same strict inequality. Consequently,

$$c_t(\epsilon, \epsilon) < 0, \qquad -\epsilon < 0, \qquad -\epsilon < 0. \tag{42}$$

For multipliers $\lambda_t, \mu_N, \mu_T \geq 0$, define

$$\begin{aligned}\mathcal{L} = & -a_t r_N - b_t r_T + \frac{r_N^2 + r_T^2}{2\alpha_t} \\ & + \lambda_t\left(r_N + \frac{1}{2}K_t r_T^2 - R_t\right) - \mu_N r_N - \mu_T r_T.\end{aligned} \tag{43}$$

Slater's condition holds, so equations (37)–(42) imply

$$\begin{aligned}&\arg\min_{(r_N,r_T)\in\mathcal{C}_t} m_t(r_N, r_T) = \{(r_{N,t}^*, r_{T,t}^*)\}, \\ &\min_{(r_N,r_T)\in\mathcal{C}_t} m_t(r_N, r_T) = \max_{\lambda_t,\mu_N,\mu_T\geq 0}\ \inf_{(r_N,r_T)\in\mathbb{R}^2} \mathcal{L}(r_N, r_T, \lambda_t, \mu_N, \mu_T), \\ &\quad (r_N, r_T) \text{ is optimal} \iff (r_N, r_T, \lambda_t, \mu_N, \mu_T) \text{ satisfies the KKT system.}\end{aligned} \tag{44}$$

The complete KKT system is

$$\begin{cases} -a_t + \dfrac{r_N}{\alpha_t} + \lambda_t - \mu_N = 0, \\ -b_t + \dfrac{r_T}{\alpha_t} + \lambda_t K_t r_T - \mu_T = 0, \\ r_N + \dfrac{1}{2}K_t r_T^2 - R_t \leq 0, \\ r_N,\ r_T,\ \lambda_t,\ \mu_N,\ \mu_T \geq 0, \\ \lambda_t\left(r_N + \dfrac{1}{2}K_t r_T^2 - R_t\right) = 0, \\ \mu_N r_N = \mu_T r_T = 0. \end{cases} \tag{45}$$

The normal stationarity and complementarity conditions give

$$\begin{cases} r_N > 0 \implies \mu_N = 0,\ r_N = \alpha_t(a_t - \lambda_t),\ \lambda_t < a_t, \\ r_N = 0 \implies \mu_N = \lambda_t - a_t \geq 0,\ \lambda_t \geq a_t. \end{cases} \tag{46}$$

Hence

$$r_N(\lambda_t) = \alpha_t(a_t - \lambda_t)_+. \tag{47}$$

If $b_t > 0$ and $r_T = 0$, tangent stationarity would require

$$-b_t - \mu_T = 0 \quad \implies \quad \mu_T = -b_t < 0, \tag{48}$$

contradicting dual feasibility. Therefore $r_T > 0$, $\mu_T = 0$, and

$$\left(\frac{1}{\alpha_t} + \lambda_t K_t\right) r_T = b_t \quad \implies \quad r_T(\lambda_t) = \frac{\alpha_t b_t}{1 + \alpha_t\lambda_t K_t}. \tag{49}$$

If $b_t = 0$, tangent stationarity and complementarity give

$$r_T\left[\left(\frac{1}{\alpha_t}+\lambda_t K_t\right) r_T - \mu_T\right] = \left(\frac{1}{\alpha_t}+\lambda_t K_t\right) r_T^2 = 0 \quad \Longrightarrow \quad r_T = 0, \tag{50}$$

which is also (49). Equations (47) and (49) therefore prove the claimed closed-form expressions for $r_{N,t}(\lambda_t)$ and $r_{T,t}(\lambda_t)$.

At $\lambda_t = 0$,

$$(r_N(0), r_T(0)) = (\alpha_t a_t, \alpha_t b_t), \tag{51}$$

and its constraint value is

$$F_t(0) - R_t = \alpha_t a_t + \frac{1}{2} K_t (\alpha_t b_t)^2 - R_t. \tag{52}$$

If $\alpha_t a_t + \frac{1}{2} K_t (\alpha_t b_t)^2 \le R_t$, then $\lambda_t = 0$ together with (47)–(49) satisfies (45), and hence gives the unique optimum.

If this condition fails, then $\lambda_t \neq 0$ by (52). Thus $\lambda_t > 0$, and complementarity enforces

$$r_N(\lambda_t) + \frac{1}{2} K_t r_T(\lambda_t)^2 = R_t, \tag{53}$$

which is exactly (9).

For $\lambda \ge 0$, define

$$F_t(\lambda) = \alpha_t (a_t - \lambda)_+ + \frac{K_t \alpha_t^2 b_t^2}{2(1+\alpha_t \lambda K_t)^2}. \tag{54}$$

It satisfies

$$F_t(0) > R_t > 0, \qquad \lim_{\lambda \to \infty} F_t(\lambda) = 0 < R_t. \tag{55}$$

On $\lambda < a_t$,

$$F_t'(\lambda) = -\alpha_t - \frac{K_t^2 \alpha_t^3 b_t^2}{(1+\alpha_t \lambda K_t)^3} < 0. \tag{56}$$

On $\lambda > a_t$,

$$F_t'(\lambda) = -\frac{K_t^2 \alpha_t^3 b_t^2}{(1+\alpha_t \lambda K_t)^3} \le 0. \tag{57}$$

Continuity and (55) give existence of a root. For $0 \le \lambda_1 < \lambda_2$ with $F_t(\lambda_1) > 0$,

$$\begin{cases} \lambda_1 < a_t & \Longrightarrow\ F_t(\lambda_2) < F_t(\lambda_1), \\ \lambda_1 \ge a_t & \Longrightarrow\ K_t b_t > 0 \Longrightarrow\ F_t(\lambda_2) < F_t(\lambda_1). \end{cases} \tag{58}$$

Therefore two distinct roots at the positive level $R_t$ cannot exist, so the nonnegative solution of (9) is unique.

### B.3 Proof of Theorem 4

For any $d$ whose update segment is in the $L_g$-smooth neighborhood,

$$\begin{aligned} L_t(x_t + d) - L_t(x_t) &= \int_0^1 \nabla L_t(x_t + \tau d)^\top d \, d\tau \\ &= q_t^\top d + \int_0^1 \left[\nabla L_t(x_t + \tau d) - \nabla L_t(x_t)\right]^\top d \, d\tau \\ &\le q_t^\top d + \int_0^1 \left\|\nabla L_t(x_t + \tau d) - \nabla L_t(x_t)\right\|_2 \|d\|_2 \, d\tau \\ &\le q_t^\top d + \int_0^1 L_g \tau \|d\|_2^2 \, d\tau \\ &= q_t^\top d + \frac{L_g}{2} \|d\|_2^2. \end{aligned} \tag{59}$$

Let

$$d_t^* = -r_{N,t}^* e_{N,t} - r_{T,t}^* e_{T,t}. \tag{60}$$

Orthogonality of the normal and tangent components gives

$$\begin{aligned} q_t^\top d_t^* &= (q_{N,t} + q_{T,t})^\top (-r_{N,t}^* e_{N,t} - r_{T,t}^* e_{T,t}) \\ &= -a_t r_{N,t}^* - b_t r_{T,t}^*, \\ \|d_t^*\|_2^2 &= (r_{N,t}^*)^2 + (r_{T,t}^*)^2. \end{aligned} \tag{61}$$

Since $(0,0)$ is feasible,

$$m_t(r_{N,t}^*, r_{T,t}^*) \le m_t(0,0) = 0. \tag{62}$$

If $a_t > 0$, choose

$$0 < \epsilon_N < \min\{R_t, 2\alpha_t a_t\}. \tag{63}$$

Then $(\epsilon_N, 0)$ is feasible and

$$m_t(\epsilon_N, 0) = -a_t \epsilon_N + \frac{\epsilon_N^2}{2\alpha_t} = -\epsilon_N \left( a_t - \frac{\epsilon_N}{2\alpha_t} \right) < 0. \tag{64}$$

If $a_t = 0$ and $b_t > 0$, select

$$0 < \epsilon_T < \begin{cases} \min\left\{ 2\alpha_t b_t, \sqrt{2R_t/K_t} \right\}, & K_t > 0, \\ 2\alpha_t b_t, & K_t = 0. \end{cases} \tag{65}$$

Then $(0, \epsilon_T)$ is feasible and

$$m_t(0, \epsilon_T) = -b_t \epsilon_T + \frac{\epsilon_T^2}{2\alpha_t} = -\epsilon_T \left( b_t - \frac{\epsilon_T}{2\alpha_t} \right) < 0. \tag{66}$$

Thus

$$q_t \neq 0 \quad \Longrightarrow \quad a_t > 0 \text{ or } b_t > 0 \quad \Longrightarrow \quad m_t(r_{N,t}^*, r_{T,t}^*) < 0. \tag{67}$$

Using (61),

$$m_t(r_{N,t}^*, r_{T,t}^*) = q_t^\top d_t^* + \frac{1}{2\alpha_t} \|d_t^*\|_2^2. \tag{68}$$

For $0 < \alpha_t \le 1/L_g$,

$$\begin{aligned} L_t(x_t + d_t^*) - L_t(x_t) &\le q_t^\top d_t^* + \frac{L_g}{2} \|d_t^*\|_2^2 \\ &\le q_t^\top d_t^* + \frac{1}{2\alpha_t} \|d_t^*\|_2^2 \\ &= m_t(r_{N,t}^*, r_{T,t}^*) \\ &\le 0. \end{aligned} \tag{69}$$

Equations (67) and (69) make the final inequality strict when $q_t \neq 0$, proving the stated full-step decrease.

For arbitrary $\alpha_t > 0$, (67) and (68) imply

$$q_t^\top d_t^* < -\frac{1}{2\alpha_t} \|d_t^*\|_2^2 < 0 \qquad (q_t \neq 0). \tag{70}$$

For $\gamma \in (0, 1]$, apply (59) to $\gamma d_t^*$:

$$L_t(x_t + \gamma d_t^*) - L_t(x_t) \le \gamma q_t^\top d_t^* + \frac{L_g}{2} \gamma^2 \|d_t^*\|_2^2. \tag{71}$$

The Armijo inequality follows whenever

$$\begin{aligned} \gamma q_t^\top d_t^* + \frac{L_g}{2} \gamma^2 \|d_t^*\|_2^2 &\le c \gamma q_t^\top d_t^* \\ \iff \qquad \gamma &\le \overline{\gamma}_t := \frac{2(1-c)(-q_t^\top d_t^*)}{L_g \|d_t^*\|_2^2}. \end{aligned} \tag{72}$$

Define

$$\gamma_{\mathrm{req}} := \min\{1, \overline{\gamma}_t\}, \qquad j_\star := \max\left\{0, \left\lceil \frac{\log \gamma_{\mathrm{req}}}{\log \beta} \right\rceil \right\}. \tag{73}$$

Because $0 < \beta < 1$,

$$0 < \beta^{j_\star} \le \gamma_{\mathrm{req}} \le \overline{\gamma}_t. \tag{74}$$

Hence Armijo backtracking accepts no later than iteration $j_\star$, and

$$\begin{aligned} L_t(x_t + \beta^{j_\star} d_t^*) &\le L_t(x_t) + c\beta^{j_\star} q_t^\top d_t^* \\ &< L_t(x_t), \end{aligned} \tag{75}$$

which proves (11).

Finally, for any accepted $\gamma_t \in (0, 1]$,

$$\begin{aligned} \gamma_t r_{N,t}^* + \frac{1}{2} K_t \gamma_t^2 (r_{T,t}^*)^2 &\le \gamma_t \left[ r_{N,t}^* + \frac{1}{2} K_t (r_{T,t}^*)^2 \right] \\ &\le \gamma_t R_t \le R_t, \end{aligned} \tag{76}$$

which proves the stated second-order feasibility.

## C Reproducibility and Hyperparameter Settings

**Implementation summary.** Algorithm 1 summarizes the DPS+CAT integration used in our implementation. It explicitly includes score-based gradient decomposition, matrix-free directional-curvature estimation, scalar-dual solution, and Armijo backtracking in the host's working variable. The same CAT correction is applied in the latent working variable for PSLD+CAT and ReSample+CAT; their host-specific settings are reported in Tables 4 and 5, respectively. Across all three integrations, the Armijo test is performed every $p$ reverse steps, with the most recently accepted $\gamma_t$ reused on the intervening steps. Each scheduled search tests at most $J_{\max}$ scales; if all fail, the current guidance update is rejected and the cached scale is left unchanged. For the latent variants, the PSLD weights $\gamma_{\mathrm{PSLD}}$ and $\omega$ are absorbed into its guidance objective, giving the nominal host step $\alpha_t = 1$, whereas ReSample uses $\alpha_t = s_{\mathrm{DPS}}$ for its soft latent DPS correction.

Input: state $x^-_{t+\Delta t}$, observation $y$,
Armijo trial limit $J_{\max} \geq 1$,
initial step $\alpha_t$, budget coefficient $\rho$,
Armijo period $p$, Armijo constants $c, \beta$,
cached scale $\gamma_{\text{cache}}$ (initialized to 1)
$x_t \leftarrow \text{prior_step}(x^-_{t+\Delta t})$
$q_t \leftarrow \nabla L_t(x_t)$; $\widehat{s}_t \leftarrow s_\theta(x_t, t)$
$\widehat{\nu}_t \leftarrow \widehat{s}_t / \|\widehat{s}_t\|_2$
$\widehat{q}_{N,t} \leftarrow \widehat{\nu}_t(\widehat{\nu}_t^\top q_t)$
$\widehat{q}_{T,t} \leftarrow q_t - \widehat{q}_{N,t}$
$\widehat{a}_t \leftarrow \|\widehat{q}_{N,t}\|_2$
$\widehat{b}_t \leftarrow \|\widehat{q}_{T,t}\|_2$
if $\widehat{b}_t > 0$ then
    $u_t \leftarrow \text{stopgrad}(\widehat{q}_{T,t}/\widehat{b}_t)$
    $v_t \leftarrow \text{autograd}_{x_t}(u_t^\top \widehat{s}_t)$
    $\widehat{K}_t \leftarrow |u_t^\top v_t| / \|\widehat{s}_t\|_2$
else
    $\widehat{K}_t \leftarrow 0$
end if
$R_t \leftarrow \rho\sigma_t$
if $\alpha_t\widehat{a}_t + \frac{1}{2}\widehat{K}_t(\alpha_t\widehat{b}_t)^2 \leq R_t$ then
    $\lambda_t \leftarrow 0$
else
    solve $\widehat{F}_t(\lambda_t) = R_t$ by bisection
end if
$\eta_{N,t} \leftarrow 0$; $\eta_{T,t} \leftarrow 0$
if $\widehat{a}_t > 0$ then
    $\eta_{N,t} \leftarrow \alpha_t(1 - \lambda_t/\widehat{a}_t)_+$
end if
if $\widehat{b}_t > 0$ then
    $\eta_{T,t} \leftarrow \alpha_t/(1 + \alpha_t\lambda_t\widehat{K}_t)$
end if
$d_t \leftarrow -\eta_{N,t}\widehat{q}_{N,t} - \eta_{T,t}\widehat{q}_{T,t}$
if $t$ is scheduled for Armijo by $p$ then
    $\gamma_t \leftarrow 0$
    for $j = 0, \ldots, J_{\max} - 1$ do
        $\gamma_{\text{trial}} \leftarrow \beta^j$
        if $L_t(x_t + \gamma_{\text{trial}} d_t) \leq$
            $L_t(x_t) + c\gamma_{\text{trial}} q_t^\top d_t$ then
            $\gamma_t \leftarrow \gamma_{\text{trial}}$
            $\gamma_{\text{cache}} \leftarrow \gamma_t$
            break
        end if
    end for
    // If all trials fail, keep $\gamma_t = 0$ and $\gamma_{\text{cache}}$ unchanged.
else
    $\gamma_t \leftarrow \gamma_{\text{cache}}$
end if
return $x_t^- \leftarrow x_t + \gamma_t d_t$

Algorithm 1: DPS+CAT.

**Inverse-problem operators.** The natural-image benchmark uses random-mask inpainting with a missing-pixel probability of $0.7$, box inpainting with a $128 \times 128$ mask, $4\times$ super-resolution, motion and Gaussian deblurring with $61 \times 61$ kernels and intensities $0.5$ and $3.0$, respectively, high-dynamic-range reconstruction, and nonlinear deblurring. Gaussian measurement noise with standard deviation $\sigma_y = 0.05$ is added in every task except the noise-robustness experiment, which evaluates multiple noise levels.

**Configuration reporting protocol.** Tables 3–5 summarize the hyperparameters used in the natural-image experiments reported in Table 1; Table 3 additionally reports the DPS+CAT configuration used for the scientific inverse problem in Table 7. For the inverse-problem integrations, $\rho$ controls the noise-dependent tubular budget $R_t = \rho\sigma_t$, $p$ denotes the interval between Armijo evaluations, and $N$ is the number of reverse sampling steps. In experiments that use Armijo backtracking, we set $c = 10^{-4}$ and $\beta = 0.5$, and set the maximum number of Armijo trials to $J_{\max} = 3$, testing scales $1, \beta, \beta^2$ (i.e., $1, 0.5, 0.25$). If none satisfies the Armijo condition, we set $\gamma_t = 0$ and reject the current guidance update, retaining the prior anchor. The cached scale $\gamma_{\text{cache}}$ is updated only after a successful trial; a rejected search leaves it unchanged for reuse on intervening steps. Each scheduled search restarts at scale 1. Unless otherwise reported in these tables, the hyperparameters of DPS+CAT, PSLD+CAT, and ReSample+CAT are kept identical to those of DPS (Chung et al., 2023a), PSLD (Rout et al., 2023), and ReSample (Song et al., 2024), respectively.

**Pretrained model provenance.** For the pixel-space experiments, the FFHQ prior is the pretrained diffusion model released with DPS by Chung et al. (2023a), and the ImageNet prior is the unconditional $256 \times 256$ diffusion model of Dhariwal & Nichol (2021). For the latent-space experiments, the FFHQ prior is the public LDM-VQ-4 checkpoint and the ImageNet prior is the public class-conditional LDM-VQ-8 checkpoint released by Rombach et al. (2022). The scientific extension uses the public $64 \times 64$ black-hole diffusion prior released with InverseBench (Zheng et al., 2025). All diffusion priors are used without retraining or task-specific fine-tuning.

Table 3: DPS+CAT hyperparameters used in the natural-image and scientific inverse-problem experiments. The host guidance step corresponds to `scale` for natural images and `step_size` for InverseBench. The symbols $\rho$, $p$, and $N$ denote the tubular-radius coefficient, Armijo period, and denoising-step count, respectively.

| Dataset | Inverse problem | Host step | $\rho$ | $p$ | $N$ |
|---|---|---|---|---|---|
| FFHQ | Random inpainting | 4.0 | 0.1 | 1 | 1000 |
| | Box inpainting | 2.0 | 0.1 | 1 | 1000 |
| | Super-resolution | 2.0 | 0.1 | 1 | 1000 |
| | Motion deblurring | 5.0 | 0.1 | 1 | 1000 |
| | Gaussian deblurring | 5.0 | 0.1 | 1 | 1000 |
| | HDR reconstruction | 10.0 | 0.1 | 1 | 1000 |
| | Nonlinear deblurring | 5.0 | 0.1 | 1 | 1000 |
| ImageNet | Random inpainting | 2.0 | 1.0 | 1 | 1000 |
| | Box inpainting | 4.0 | 1.0 | 1 | 1000 |
| | Super-resolution | 1.5 | 1.0 | 1 | 1000 |
| | Motion deblurring | 6.0 | 1.0 | 1 | 1000 |
| | Gaussian deblurring | 5.0 | 1.0 | 1 | 1000 |
| | HDR reconstruction | 5.0 | 0.01 | 1 | 1000 |
| | Nonlinear deblurring | 10.0 | 1.0 | 1 | 1000 |
| InverseBench | Black-hole imaging (100% obs.) | 95 | 0.0005 | 1 | 1000 |
| | Black-hole imaging (10% obs.) | 95 | 0.0005 | 1 | 1000 |

**Settings of the comparison methods.** For DPS, PSLD, ReSample, DDRM, DDNM, DAPS, DCDP, DiffRGD, DPS+Diffstate, PSLD+DiffState, and ReSample+Diffstate, we use the hyperparameter settings reported in their respective original papers whenever available and otherwise tune the unspecified hyperparameters by grid search.

# D Additional Experimental Results

## D.1 Efficiency Analysis

We analyze the computational trade-offs of CAT on FFHQ nonlinear deblurring with DPS, using the same 100 images as in the main experiment. All runs are conducted on a single NVIDIA GeForce RTX 4090 GPU. We vary two factors: the number of reverse sampling steps $N$ and the Armijo

Table 4: PSLD+CAT hyperparameters used in the main experiments. The original PSLD hyperparameters are the latent data-consistency weight $\gamma_{\mathrm{PSLD}}$ and guidance weight $\omega$; the remaining notation follows Table 3. We denote the PSLD latent data-consistency weight by $\gamma_{\mathrm{PSLD}}$ to distinguish it from the Armijo scale $\gamma_t$. PSLD is evaluated on the five linear inverse problems supported by its host implementation.

| **Dataset** | **Inverse problem** | $\boldsymbol{\gamma_{\mathrm{PSLD}}}$ | $\boldsymbol{\omega}$ | $\boldsymbol{\rho}$ | $\boldsymbol{p}$ | $\boldsymbol{N}$ |
|---|---|---|---|---|---|---|
| FFHQ | Random inpainting | 0.01 | 2.0 | 1.0 | 1 | 1000 |
| | Box inpainting | 0.01 | 0.5 | 1.0 | 1 | 1000 |
| | Super-resolution | 0.01 | 5.0 | 1.0 | 1 | 1000 |
| | Motion deblurring | 0.01 | 5.0 | 1.0 | 1 | 1000 |
| | Gaussian deblurring | 0.10 | 10.0 | 1.0 | 1 | 1000 |
| ImageNet | Random inpainting | 0.10 | 4.0 | 0.1 | 1 | 1000 |
| | Box inpainting | 0.01 | 1.0 | 0.1 | 1 | 1000 |
| | Super-resolution | 0.10 | 2.0 | 0.1 | 1 | 1000 |
| | Motion deblurring | 0.10 | 1.0 | 0.1 | 1 | 1000 |
| | Gaussian deblurring | 0.10 | 2.0 | 0.1 | 1 | 1000 |

Table 5: ReSample+CAT hyperparameters used in the main experiments. $s_{\mathrm{DPS}}$ is the scale of the soft latent DPS step, $P_{\mathrm{tt}}$ is the time-travel period, and the remaining notation follows Table 3.

| **Dataset** | **Inverse problem** | $\boldsymbol{s_{\mathrm{DPS}}}$ | $\boldsymbol{P_{\mathrm{tt}}}$ | $\boldsymbol{\rho}$ | $\boldsymbol{p}$ | $\boldsymbol{N}$ |
|---|---|---|---|---|---|---|
| FFHQ | Random inpainting | 15.0 | 10 | 0.1 | 5 | 500 |
| | Box inpainting | 5.0 | 500 | 0.1 | 5 | 500 |
| | Super-resolution | 150.0 | 500 | 0.1 | 5 | 500 |
| | Motion deblurring | 5.0 | 10 | 0.1 | 5 | 500 |
| | Gaussian deblurring | 10.0 | 10 | 0.1 | 5 | 500 |
| | HDR reconstruction | 30.0 | 500 | 0.1 | 1 | 500 |
| | Nonlinear deblurring | 1.0 | 10 | 0.1 | 1 | 500 |
| ImageNet | Random inpainting | 2.0 | 10 | 0.1 | 5 | 500 |
| | Box inpainting | 2.0 | 10 | 0.1 | 5 | 500 |
| | Super-resolution | 15.0 | 10 | 0.1 | 5 | 500 |
| | Motion deblurring | 2.5 | 10 | 0.1 | 5 | 500 |
| | Gaussian deblurring | 2.0 | 10 | 0.1 | 5 | 500 |

period $p$ introduced above. All other method settings remain fixed within each sweep. Table 6 reports reconstruction quality, GPU memory usage, and wall-clock time.

Table 6: Efficiency of DPS+CAT on FFHQ nonlinear deblurring over 100 images using a single NVIDIA GeForce RTX 4090 GPU. Both blocks include the same DPS baseline for reference. The upper block varies the number of reverse sampling steps $N$ with Armijo applied at every step, while the lower block fixes $N = 1000$ and varies the Armijo period $p$. Memory and time denote the recorded GPU memory usage and wall-clock runtime per image, respectively. Bold marks the best reconstruction quality within each block.

| Study | Setting | PSNR ↑ | SSIM ↑ | LPIPS ↓ | FID ↓ | Memory (MiB) | Time (s) |
|---|---|---|---|---|---|---|---|
| Sampling steps | DPS (reference) | 21.97 | 0.6269 | 0.3219 | 85.68 | 3490 | 38 |
| | $N = 200$ | 23.50 | 0.6851 | 0.2922 | 82.91 | 3610 | 26 |
| | $N = 400$ | 24.62 | 0.7152 | 0.2633 | 70.54 | 3610 | 51 |
| | $N = 600$ | 25.42 | 0.7380 | 0.2432 | 64.24 | 3610 | 78 |
| | $N = 800$ | 25.92 | 0.7525 | 0.2297 | 58.73 | 3612 | 103 |
| | $N = 1000$ | **26.30** | **0.7623** | **0.2197** | **55.17** | 3612 | 132 |
| Armijo period | DPS (reference) | 21.97 | 0.6269 | 0.3219 | 85.68 | 3490 | 38 |
| | $p = 5$ | 24.85 | 0.7292 | 0.2473 | 67.04 | 3616 | 89 |
| | $p = 3$ | 24.82 | 0.7285 | 0.2472 | 62.92 | 3614 | 97 |
| | $p = 2$ | 24.48 | 0.7161 | 0.2561 | 69.29 | 3614 | 106 |
| | $p = 1$ (Full CAT) | **26.30** | **0.7623** | **0.2197** | **55.17** | 3612 | 132 |

**Armijo frequency.** The Armijo period provides a flexible efficiency–quality trade-off: less frequent searches reduce the runtime overhead while preserving a clear quality advantage over DPS. When reconstruction quality is the priority, performing backtracking at every step yields the strongest and most consistent performance across the evaluation metrics. Overall, CAT provides a practical and adaptable enhancement whose computational cost can be adjusted without forfeiting its core reconstruction benefits.

**Number of sampling steps.** Compared with DPS, DPS+CAT improves all four reconstruction metrics at every tested step count. DPS+CAT at $N = 200$ achieves higher PSNR than DPS at $N = 1000$ in less time, indicating that the gain is not merely due to a larger computational budget. As $N$ increases, runtime scales approximately linearly while memory usage remains nearly constant, and reconstruction quality improves gradually. This smooth quality–efficiency trade-off enables CAT to adapt readily to different inference budgets without increasing memory demand, making it practical for both quality-oriented and resource-constrained settings.

### D.2 Scientific Inverse Problems

**Evaluation protocol.** We extend the evaluation beyond natural image restoration to black hole imaging in InverseBench (Zheng et al., 2025). This task reconstructs a source image from sparse interferometric observations using nonlinear closure-phase and log-closure-amplitude constraints, together with a total-flux constraint. We evaluate DPS+CAT on 100 test images at $64 \times 64$ resolution using the pretrained black-hole diffusion prior from InverseBench (Zheng et al., 2025) and observation time ratios of $100\%$ and $10\%$. The algorithm hyperparameter settings used in these experiments are provided in Appendix C. We compare against the domain-specific methods SMILI and EHT-Imaging and the diffusion-prior methods DPS, LGD, RED-Diff, PnPDM, DAPS, and DiffPIR, using their reported InverseBench results. Image quality is measured by PSNR and Blur PSNR. We also report the reduced chi-squared statistics $\widetilde{\chi}^2_{\mathrm{cp}}$ and $\widetilde{\chi}^2_{\mathrm{logca}}$ for closure phase and log closure amplitude; values near one indicate agreement at the measurement noise level.

**Quantitative comparison.** Table 7 shows that DPS+CAT consistently achieves the strongest image-space reconstruction quality across both observation-time settings. Compared with the DPS host, CAT improves both PSNR and Blur PSNR, indicating better pixel-wise fidelity and large-scale morphological agreement with the ground truth. This advantage persists under reduced observation

Table 7: Black hole imaging on InverseBench at 100% and 10% observation time. Entries are mean (standard deviation). Baseline results are taken from InverseBench (Zheng et al., 2025); DPS+CAT is evaluated on 100 test images. Bold marks the highest mean PSNR/Blur PSNR and the lowest mean reduced chi-squared within each observation setting.

| Method | Observation time ratio: 100% | | | | Observation time ratio: 10% | | | |
|---|---|---|---|---|---|---|---|---|
| | PSNR ↑ | Blur PSNR ↑ | $\widetilde{\chi}^2_{\text{cp}}$ ↓ | $\widetilde{\chi}^2_{\text{logca}}$ ↓ | PSNR ↑ | Blur PSNR ↑ | $\widetilde{\chi}^2_{\text{cp}}$ ↓ | $\widetilde{\chi}^2_{\text{logca}}$ ↓ |
| SMILI | 22.67 (3.13) | 27.79 (4.02) | 1.878 (0.952) | 17.612 (10.299) | 20.85 (2.90) | 25.24 (3.86) | 1.209 (0.169) | 21.788 (12.491) |
| EHT-Imaging | 24.28 (3.63) | 28.57 (4.52) | **1.251 (0.250)** | 1.259 (0.316) | 22.67 (3.46) | 26.66 (3.93) | **1.166 (0.156)** | **1.240 (0.205)** |
| DPS | 25.86 (3.90) | 32.94 (6.19) | 8.759 (37.784) | 5.456 (24.185) | 24.36 (3.72) | 30.79 (5.75) | 13.052 (43.087) | 6.614 (26.789) |
| LGD | 21.22 (3.64) | 26.06 (4.98) | 13.239 (17.231) | 13.233 (39.107) | 22.08 (3.75) | 27.48 (5.09) | 10.775 (21.684) | 13.375 (56.397) |
| RED-Diff | 23.77 (4.13) | 29.13 (6.22) | 1.853 (0.938) | 2.050 (2.361) | 22.53 (3.02) | 27.67 (4.53) | 2.488 (2.925) | 4.916 (13.221) |
| PnPDM | 26.07 (3.70) | 32.88 (6.02) | 1.311 (0.195) | **1.199 (0.221)** | 24.57 (3.47) | 30.80 (5.22) | 1.433 (0.417) | 1.336 (0.478) |
| DAPS | 25.60 (3.64) | 32.78 (5.68) | 1.300 (0.324) | 1.229 (0.532) | 23.99 (3.56) | 30.10 (5.13) | 1.545 (0.705) | 2.253 (9.903) |
| DiffPIR | 25.01 (4.64) | 31.86 (6.56) | 3.271 (1.623) | 2.970 (1.202) | 23.84 (3.59) | 30.04 (5.03) | 5.374 (3.733) | 5.205 (5.556) |
| DPS+CAT (ours) | **27.94 (3.95)** | **35.94 (6.17)** | 2.625 (13.114) | 1.663 (4.784) | **26.35 (4.04)** | **33.14 (5.99)** | 1.239 (0.222) | 1.264 (0.221) |

time, supporting the applicability of CAT beyond natural-image restoration to a challenging physics-based nonlinear inverse problem. Figure 6 further shows that DPS+CAT recovers a sharper, more continuous emission ring with fewer artifacts, particularly at 10% observation time.

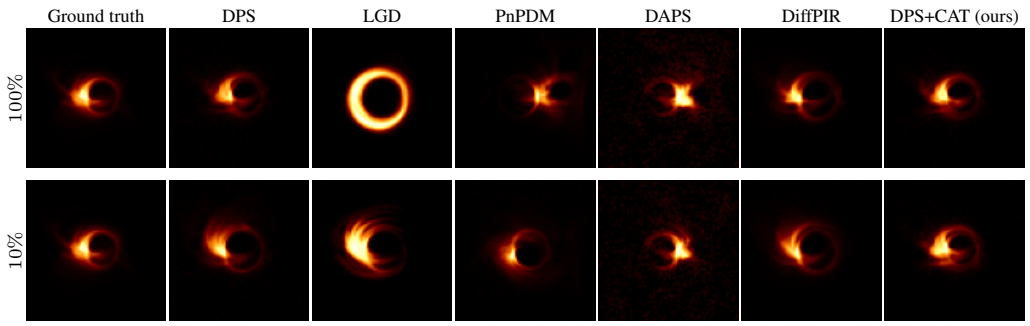



Figure 6: Qualitative black hole imaging results for the same ground-truth source from InverseBench at 100% (top) and 10% (bottom) observation time.

## E VISUALIZATION




**(a)** Random inpainting.

**(b)** Motion deblurring.

**(c)** Gaussian deblurring.

**(d)** Super-resolution ($\times 4$).

**(e)** Nonlinear deblurring.

**(f)** HDR reconstruction.

Figure 7: Complete qualitative comparison of DPS-based methods on FFHQ across six representative inverse problems. Each panel shows two examples with the same original image and measurement. DPS+CAT denotes DPS equipped with CAT, while DPS+Diffstate is the corresponding DiffState baseline.

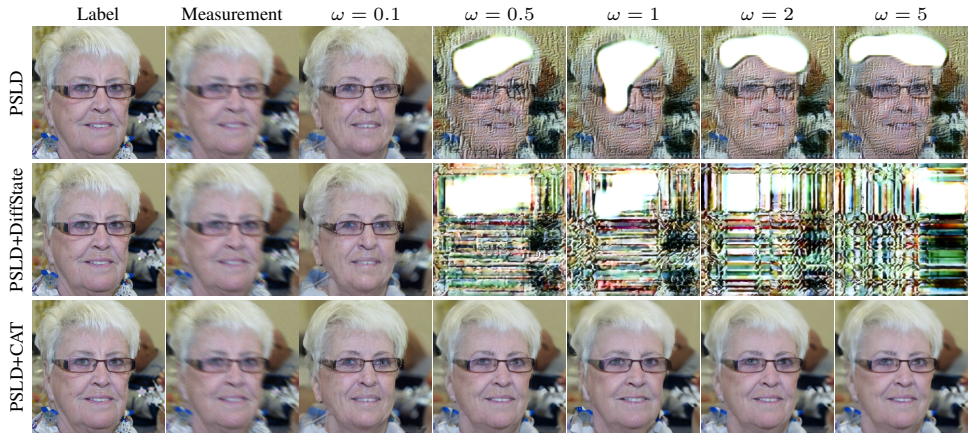



Figure 8: Qualitative robustness to the guidance step size on FFHQ motion deblurring. Rows correspond to PSLD, PSLD+DiffState, and PSLD+CAT; columns show the ground-truth label, the blurred measurement, and reconstructions obtained with five values of $\omega$. All entries use the same input image.




Figure 9: Qualitative robustness of PSLD+CAT to the tubular-radius coefficient on FFHQ motion deblurring. Columns show the ground-truth label, the blurred measurement, the PSLD reconstruction from the main experiment, and PSLD+CAT reconstructions obtained with five values of $\rho$ for the same input.




Figure 10: Qualitative robustness to measurement noise on FFHQ motion deblurring. The left image shows the ground-truth label. Rows correspond to the four noise levels, while columns show the measurement and reconstructions from each method for the same input image.




Figure 11: Qualitative component ablation on FFHQ. The first row uses DPS for nonlinear deblurring, and the second uses PSLD for motion deblurring. Each row shares the same ground truth and measurement across variants.