EXPOSE, THEN COMMIT: DIFFUSION POSTERIOR SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion posterior sampling solves inverse problems by combining a pretrained diffusion prior with measurement-consistency guidance. However, at high noise levels, clean-estimate errors can compromise full-band guidance, while stochastic refinement can modify details weakly constrained by the measurements. We argue that posterior sampling should coordinate measurement-frequency exposure during refinement with correction commitment afterward. Based on this principle, we propose a posterior continuation framework that constructs a family of intermediate posteriors whose noise-conditioned likelihoods emphasize low-frequency residuals and gradually return to full-band consistency. We instantiate this framework with a stabilized sampler combining a diffusion predictor, frequency-weighted Langevin refinement, and a Haar-domain commitment rule that fully adopts coarse corrections while progressively admitting detail corrections relative to the predictor. Across super-resolution, inpainting, and deblurring, our method achieves competitive-to-state-of-the-art restoration performance, including PSNR gains exceeding 5 dB on motion deblurring against the strongest evaluated baselines.

1 INTRODUCTION

Diffusion models (Ho et al., 2020; Song et al., 2021a;b; 2023) have emerged as powerful priors for solving challenging inverse problems (Kawar et al., 2022; Wang et al., 2023) without task-specific training, by combining pretrained diffusion priors with measurement-consistency guidance during sampling. Despite their empirical success, diffusion posterior sampling methods often struggle to recover fine details under severe degradations such as strong blur and heavy downsampling. These failures can persist even when the forward model is known exactly, indicating that measurement consistency alone does not ensure reliable corrections throughout sampling. A key observation is that the reliability of intermediate clean estimates varies with both the diffusion noise level and signal frequency (Falck et al., 2025; Thaker et al., 2025). In early and intermediate stages, high-frequency components often have low signal-to-noise ratios, whereas low-frequency structure tends to be more stable and preserve global geometry. Consequently, uniform measurement weighting need not reflect these noise-dependent reliability differences. Moreover, stochastic refinement can modify details that are weakly constrained by the measurements, even when high-frequency residuals are down-weighted. This motivates considering both the guidance used to produce a correction and the extent to which that correction should be carried forward.

Early influential works reflect related intuitions from complementary angles, including ILVR (Choi et al., 2021) for low-frequency conditioning, DDRM (Kawar et al., 2022) for efficient linear inverse sampling via operator spectral factorization, DDNM (Wang et al., 2023) for zero-shot restoration through a null-space formulation, and DPS (Chung et al., 2023b) for general inverse problems using approximate posterior score guidance.

A rich line of work further modifies either the posterior correction mechanism or the sampling trajectory. Optimization-based methods strengthen proximal interpretations of sampling, such as DiffPIR (Zhu et al., 2023) and DPnP (Xu & Chi, 2024). Particle-based methods reduce approximation error and better represent multi-modality, such as FPS (Dou & Song, 2024) and DCDP (Janati et al., 2024). PostEdit (Tian et al., 2025) directly optimizes the estimated clean data to enable global corrections. These developments broaden the use of diffusion priors under measurement constraints. However, stronger or more global corrections do not by themselves specify how coarse

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.

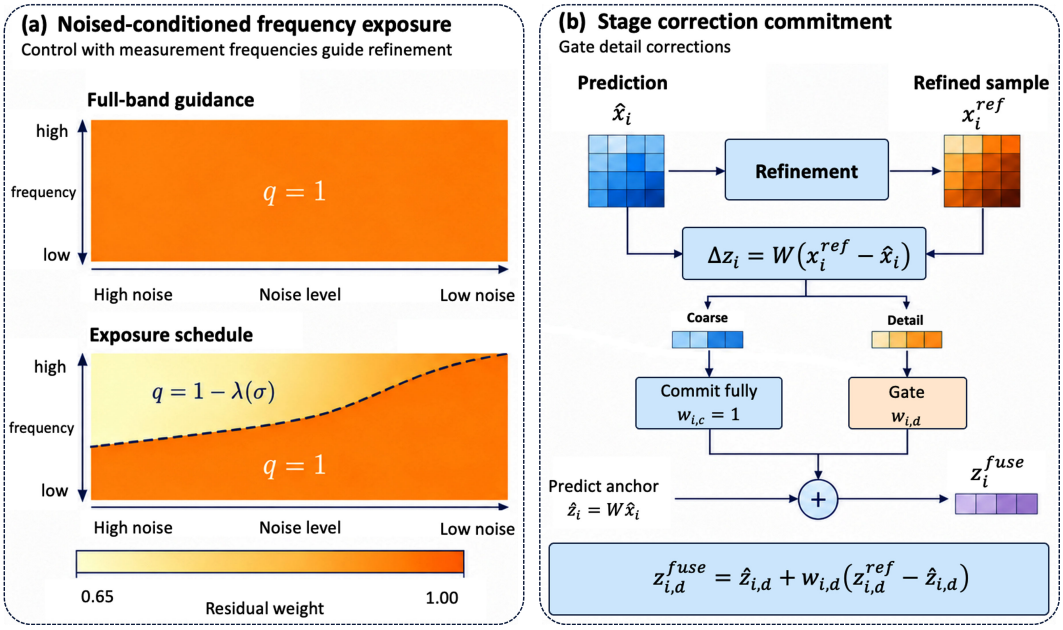

Figure 1: **Noise-conditioned exposure and staged correction commitment.** (a) Frequency exposure down-weights out-of-band measurement residuals and gradually returns to full-band weighting as noise decreases. Colors indicate residual weights and the dashed curve marks the scheduled cutoff. (b) Haar commitment fully adopts coarse corrections and gates detail corrections relative to the predictor, using an exposure-linked schedule.

and detail components should be treated as the clean estimate evolves. Against this backdrop, recent methods offer complementary perspectives. DAPS (Zhang et al., 2025) separates diffusion prediction, measurement-conditioned refinement, and re-noising to enable larger posterior moves beyond local reverse-time updates. This decoupled trajectory allows measurement corrections to be applied directly to clean estimates. Frequency-aware designs introduce spectral structure into sampling. Frequency-domain moving-average sampling (MASF) (Qian et al., 2024) averages clean predictions across sampling steps to damp unstable high-frequency fluctuations, while FGPS (Thaker et al., 2025) progressively incorporates measurement frequencies along a sampling schedule. These approaches motivate a distinction between measurement exposure and correction commitment.

As illustrated in Fig. 1, exposure determines how measurement frequencies contribute to refinement (a), whereas commitment determines how corrections enter the next sampling state (b). In particular, down-weighting high-frequency residuals changes the refinement objective, but does not directly restrict all detail changes introduced by stochastic refinement. Weakly constrained directions can still vary, and passing the entire refined estimate to the next stage adopts these variations together with coarse corrections. This raises a separate question from selecting measurement frequencies or averaging past predictions: how much of each refined component should be adopted relative to the current diffusion prediction? In this work, we propose a posterior continuation framework that coordinates noise-conditioned frequency exposure with staged detail commitment. Instead of enforcing full-band measurement consistency uniformly throughout sampling, we construct a family of intermediate posteriors whose noise-conditioned likelihoods emphasize low-frequency residuals and gradually return to full-band consistency as the noise decreases. A noise-dependent mixture of full-band and band-limited measurement residuals implements this transition through a changing cutoff and mixing weight. We instantiate this framework as a stabilized sampler that combines diffusion prediction, frequency-weighted Langevin refinement, and re-noising. At each noise level, a diffusion predictor first estimates a clean image, and frequency-weighted Langevin refinement corrects it under the corresponding intermediate posterior. We then apply a Haar-domain commitment rule that fully adopts coarse corrections while progressively admitting detail corrections relative to the predictor. The commitment operator acts on corrections, rather than filtering the predictor itself or averaging predictions across previous sampling steps. Its detail gate is linked to the exposure schedule without introducing an additional likelihood term, and the fused estimate is re-noised to obtain the next sampling state. Across super-resolution, inpainting, and deblurring, our method

achieves competitive-to-state-of-the-art restoration performance, with particularly large gains on ill-conditioned operators such as motion blur. Our main contributions are summarized as follows.

- We introduce a posterior-continuation view of diffusion inverse problems that distinguishes measurement-frequency exposure from correction commitment and coordinates these controls across the refinement objective and the subsequent sampling-state update.
- We characterize score-induced likelihood-gradient mismatch under linear-Gaussian models, showing how frequency weighting attenuates high-frequency error contributions outside the exposed passband when the forward operator and mask share a Fourier basis.
- We design a practical sampler that couples frequency-weighted Langevin refinement with predictor-relative Haar commitment, fully adopting coarse corrections and progressively admitting detail corrections through an exposure-linked gate.
- Across super-resolution, inpainting, and deblurring, our method achieves competitive-to-state-of-the-art restoration performance, including PSNR gains exceeding 5 dB on motion deblurring against the strongest evaluated baselines on both FFHQ and ImageNet.

2 PRELIMINARY

We consider an inverse problem that aims to sample the clean signal $\boldsymbol{x}_0 \in \mathbb{R}^d$ from the posterior $p(\boldsymbol{x}_0 \mid \boldsymbol{y})$, given a degraded measurement $\boldsymbol{y} \in \mathbb{R}^n$:

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{x}_0) + \boldsymbol{n}, \quad \boldsymbol{n} \sim \mathcal{N}\left(\mathbf{0}, \sigma_y^2 \boldsymbol{I}\right) \tag{1}$$

Log density terms $\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t)$ are applied to solve the inverse problem, and we can derive $\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{y})$ by Bayes' rule under posterior score guidance

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{y}) = \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t), \tag{2}$$

where the first term on the right-hand side is provided by the pre-trained diffusion model, whereas the second term is typically intractable. The measurement $\boldsymbol{y}$ induces an observation-dependent constraint that guides the sampling trajectory toward reconstructions consistent with the input, thereby complementing the prior $p(\boldsymbol{x})$ encoded by the diffusion model. To obtain an explicit and tractable surrogate for the second term, the existing method DPS Chung et al. (2023b) adopts the following approximation to enable tractable guidance

$$p(\boldsymbol{y}|\boldsymbol{x}_t) = \mathbb{E}_{\boldsymbol{x}_0 \sim p(\boldsymbol{x}_0|\boldsymbol{x}_t)}[p(\boldsymbol{y}|\boldsymbol{x}_0)] \approx p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0), \;\; \hat{\boldsymbol{x}}_0 = \mathbb{E}_{\boldsymbol{x}_0 \sim p(\boldsymbol{x}_0|\boldsymbol{x}_t)}[\boldsymbol{x}_0]. \tag{3}$$

The posterior mean $\hat{\boldsymbol{x}}_0$ can be estimated by the pretrained diffusion model, for example via Tweedie's formula under VP-SDE Song et al. (2021b) or DDIM-style denoising Song et al. (2021a). This yields a tractable approximation to $p(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y})$ under the posterior score guidance in Eq. 2.

3 METHODS

3.1 THEORETICAL ANALYSIS AND METHOD OVERVIEW

We use image restoration as a running example to motivate our method. Posterior sampling for restoration must recover both coarse structure and fine details, but the likelihood term in Eq. 2 is typically evaluated through an estimated clean image $\hat{\boldsymbol{x}}_0$ rather than the unknown clean signal. At high noise levels, this estimate can contain non-negligible score-induced errors. These errors do not imply unconditional monotone amplification for every score model and forward operator. Instead, they induce a noise-dependent sensitivity pathway. Measurement-driven corrections computed from an inaccurate $\hat{\boldsymbol{x}}_0$ can therefore be misaligned with the local posterior geometry. Consequently, full-band guidance can overemphasize unreliable high-frequency corrections during early refinement, motivating noise-conditioned weighting that prioritizes low-frequency measurement residuals before returning to full-band consistency.

We formalize this observation in Appendix A.5. Theorem A.3 first gives a worst-case sensitivity bound for the likelihood-gradient error. Theorem A.4 then provides an exact spectral characterization: measured score-error components enter the likelihood-gradient mismatch through a $\sigma^2 \mathcal{A}^\top \mathcal{A}$ factor.

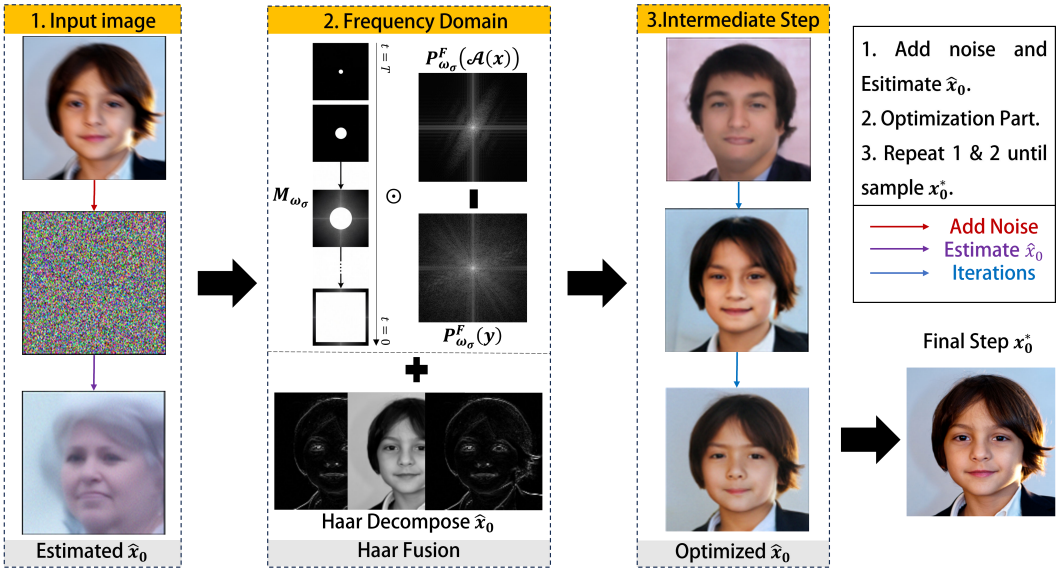

Figure 2: **Method overview.** At each noise level, a pretrained diffusion solver first predicts a clean estimate $\hat{\boldsymbol{x}}_{0,i}$. We refine it with a noise-conditioned frequency-exposure likelihood that emphasizes low-frequency residuals and gradually returns to full-band consistency. Haar-domain commitment then fully adopts coarse corrections while gating detail corrections relative to the predictor.

Consequently, under a non-degeneracy condition on the measured projection of the score error, the corresponding projected mismatch component grows with the diffusion noise level. Corollary A.5 further shows that frequency truncation removes the mismatch contributed by unexposed high-frequency bands. This directly motivates our noise-conditioned frequency exposure: at high noise, the likelihood should primarily emphasize low-frequency measurement components, and as the noise decreases, the passband is gradually expanded toward full-band consistency. Finally, Theorem A.6 shows that such gradient errors can break descent even under otherwise valid step sizes, while Corollary A.7 shows that band-limited likelihoods do not increase curvature bounds relative to full-band likelihoods, informing step-size selection during high-noise posterior refinement.

To address these issues, we introduce a posterior-continuation sampler coordinating noise-conditioned frequency exposure with a post-refinement multi-resolution commitment rule. Let $\{\tau_i\}_{i=0}^N$ denote the outer reverse-time grid and $\sigma_i = \sigma(\tau_i)$ the corresponding noise levels, with $\sigma_N > \cdots > \sigma_0 = 0$. We iterate from $i = N$ to 1. Starting from $x_{\tau_N} \sim \mathcal{N}(\mathbf{0}, \sigma_N^2 \boldsymbol{I})$, we first use a pretrained diffusion solver to obtain a clean estimate $\hat{\boldsymbol{x}}_{0,i}$ at noise level σ_i. We then refine this estimate under a noise-conditioned likelihood that emphasizes low-frequency measurement residuals at high diffusion noise levels and gradually transitions to full-band consistency. The cutoff ω_i expands as σ_i decreases, so the sampler moves from coarse measurement correction to full-band consistency. Finally, because frequency-weighted likelihoods can leave detail coefficients weakly constrained, we apply a Haar-domain commitment step: coarse corrections are committed fully, whereas detail corrections are progressively admitted relative to the current diffusion predictor. The procedure is summarized in Alg. 1 and Fig. 2, which illustrate the algorithmic flow and module interactions.

In the following two sections, Sec. 3.2 details the noise-conditioned frequency-exposure likelihood used to refine $\hat{\boldsymbol{x}}_{0,i}$, and Sec. 3.3 describes the Haar-domain commitment rule.

3.2 Noise-Conditioned Frequency Exposure as Likelihood Design

We now define the noise-conditioned likelihood used to refine $\hat{\boldsymbol{x}}_{0,i}$. Given an observation $\boldsymbol{y}$ and a forward operator $\mathcal{A}$, the goal is to emphasize measurement consistency on low-frequency components using noise-conditioned weighting during the early sampling stages, while gradually returning to full-band consistency as the noise decreases. To selectively enforce measurement consistency within a controllable frequency band, we define the (normalized) 2D discrete Fourier transform for a

multi-channel tensor $\boldsymbol{z} \in \mathbb{R}^{C\times H\times W}$ by

$$\mathcal{F}(\boldsymbol{z})_{c,u,v} = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} \boldsymbol{z}_{c,h,w} \exp\left(-2\pi i \left(\frac{uh}{H} + \frac{vw}{W}\right)\right). \tag{4}$$

Let $\mathcal{J}$ denote the frequency-shift operator that centers the zero-frequency component. We define coordinates (u, v) on the shifted frequency grid and let r_{Nyq} be the Nyquist radius of the observable residual grid. For a normalized cutoff $\omega \in [0, 1]$, the radial low-pass mask is

$$M_\omega(u, v) = \mathbf{1}\left(\frac{\sqrt{u^2 + v^2}}{r_{\mathrm{Nyq}}} < \omega\right), \tag{5}$$

with the mask broadcast across channels. We then define the masked, shifted spectrum

$$P^F_\omega(\boldsymbol{z}) := \mathcal{J}\big(\mathcal{F}(\boldsymbol{z})\big) \odot M_\omega. \tag{6}$$

where $\odot$ denotes elementwise multiplication. Throughout, $\|\cdot\|_2^2$ applied to complex spectra denotes the sum of squared magnitudes, i.e., $\|Z\|_2^2 = \sum_{c,u,v} |Z_{c,u,v}|^2$. With this normalization, Parseval's identity implies that $\|P^F_\omega(\boldsymbol{r})\|_2^2$ measures the band-limited residual energy.

The frequency mask is applied in the observable residual space. For operators whose outputs remain image-like, such as blur and masking, $P^F_{\omega_i}$ acts directly on $\mathcal{A}(\boldsymbol{x}) - \boldsymbol{y}$. For resolution-changing operators, such as super-resolution, the same construction is applied on the corresponding measurement grid. The band-limited measurement loss is

$$\mathcal{L}_{\mathrm{freq}}(\boldsymbol{x}, \boldsymbol{y}, \omega_i) = \big\|P^F_{\omega_i}\big(\mathcal{A}(\boldsymbol{x}) - \boldsymbol{y}\big)\big\|_2^2. \tag{7}$$

The band-limited term measures measurement consistency within the passband specified by ω_i. Let σ_i denote the current noise level on the sampling trajectory. We schedule a cutoff $\omega_i = \omega(\sigma_i)$ and a mixing weight $\lambda_i = \lambda(\sigma_i) \in [0, 1]$. The frequency-guided measurement objective is

$$\mathcal{L}_i(\boldsymbol{x}; \boldsymbol{y}) := (1 - \lambda_i)\, \|\mathcal{A}(\boldsymbol{x}) - \boldsymbol{y}\|_2^2 + \lambda_i\, \mathcal{L}_{\mathrm{freq}}(\boldsymbol{x}, \boldsymbol{y}, \omega_i). \tag{8}$$

Eq. 8 defines a noise-dependent likelihood that interpolates between band-limited and full-band measurement consistency. Early in the trajectory, a positive λ_i and a small ω_i attenuate out-of-band high-frequency residuals. As σ_i decreases, ω_i expands and λ_i decays, recovering full-band consistency near the terminal stage. Corollary A.8 shows that, when the operator and mask share a Fourier basis, the mixed objective weights out-of-passband score-induced error contributions by $1 - \lambda_i$. This characterization motivates noise-conditioned frequency exposure in the refinement objective rather than uniform full-band weighting throughout sampling.

We then formulate an intermediate posterior density whose log-gradient will be used to refine the estimated $\hat{\boldsymbol{x}}_{0,i}$ from the PF-ODE at noise level σ_i. Specifically, using the frequency-guided measurement objective $\mathcal{L}_i(\cdot\,;\, \boldsymbol{y})$ in Eq. 8, we define the intermediate target distribution that combines measurement consistency with a Gaussian anchor centered at $\hat{\boldsymbol{x}}_{0,i}$

$$\pi_i(\boldsymbol{x}_0) \;\propto\; \exp\left(-\frac{1}{2\beta_i^2}\, \mathcal{L}_i(\boldsymbol{x}_0; \boldsymbol{y})\right) \cdot \mathcal{N}\big(\boldsymbol{x}_0; \hat{\boldsymbol{x}}_{0,i}, \sigma_i^2 \boldsymbol{I}\big), \tag{9}$$

where $\beta_i > 0$ is a temperature coefficient controlling the strength of the measurement term. To maintain stochastic exploration under the intermediate posterior, we employ Langevin dynamics Welling & Teh (2011) and perform T refinement steps (indexed by ℓ) at each outer step i. Starting from $\boldsymbol{x}^{(0)}_{0,i} = \hat{\boldsymbol{x}}_{0,i}$, the Langevin refinement update is written as

$$\boldsymbol{x}^{(\ell+1)}_{0,i} = \boldsymbol{x}^{(\ell)}_{0,i} + \eta_{i,\ell}\, \nabla_{\boldsymbol{x}_0} \log \pi_i\Big(\boldsymbol{x}^{(\ell)}_{0,i}\Big) + \sqrt{2\eta_{i,\ell}}\, \boldsymbol{\xi}_{i,\ell}, \tag{10}$$

where $\eta_{i,\ell} > 0$ is the Langevin step size at inner iteration ℓ of outer step i, $\boldsymbol{\xi}_{i,\ell} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, and we take $\boldsymbol{x}^{\mathrm{ref}}_{0,i} = \boldsymbol{x}^{(T)}_{0,i}$ as the refined estimate. In practice, $\eta_{i,\ell}$ can be reduced as β_i decreases to maintain stable updates. All implementation details and hyperparameter settings are provided in Appendix A.4.

3.3 POST-REFINEMENT HAAR-DOMAIN DETAIL COMMITMENT

The frequency-exposure refinement in Sec. 3.2 is designed to reduce high-noise drift by emphasizing low-frequency measurement residuals in the likelihood. However, frequency-weighted likelihoods can still leave detail coefficients weakly constrained. Under the idealized low-pass model in Lemma A.11, the data term has vanishing gradient along high-frequency subspaces, so the corresponding posterior directions are dominated by the Gaussian anchor around $\hat{\boldsymbol{x}}_{0,i}$. When the noise level is large, the refinement process in Eq. 10 can therefore inject high-variance, operator-dependent detail components, including null-space or near-null-space artifacts. This motivates Haar fusion as an explicit commitment mechanism: coarse corrections are committed fully, while detail corrections are adopted progressively relative to the predictor under the linked exposure schedule.

Corollary A.12 further explains this effect. When the likelihood provides little information about Haar-detail components, their conditional posterior is dominated by the Gaussian anchor centered at $\hat{\boldsymbol{x}}_{0,i}$ after being mapped into the Haar domain, i.e.,

$$\boldsymbol{z}_{\mathrm{d}} \mid (\boldsymbol{y}, \boldsymbol{z}_{\mathrm{c}}) \approx \mathcal{N}(\hat{\boldsymbol{z}}_{\mathrm{d},i}, \sigma_i^2 \boldsymbol{I}).$$

At high noise levels, this posterior yields refined high-frequency coefficients with high variance around their predicted values. Let $\hat{\boldsymbol{x}}_{0,i}$ denote the image estimated by PF-ODE and $\boldsymbol{x}_{0,i}^{\mathrm{ref}}$ the refined estimate produced by the Langevin update in Eq. 10. We use an orthonormal Haar transform W. The coarse block c denotes the low-pass Haar coefficients, and the detail block d collects the remaining high-pass coefficients across orientations and levels. We apply W to both $\hat{\boldsymbol{x}}_{0,i}$ and $\boldsymbol{x}_{0,i}^{\mathrm{ref}}$:

$$\hat{\boldsymbol{z}}_i = W\hat{\boldsymbol{x}}_{0,i}, \qquad \boldsymbol{z}_i^{\mathrm{ref}} = W\boldsymbol{x}_{0,i}^{\mathrm{ref}}. \tag{11}$$

A Haar-domain commitment step that fuses coefficients bandwise is performed as

$$\boldsymbol{z}_{i,\mathrm{c}}^{\mathrm{fuse}} = \boldsymbol{z}_{i,\mathrm{c}}^{\mathrm{ref}}, \qquad \boldsymbol{z}_{i,\mathrm{d}}^{\mathrm{fuse}} = \hat{\boldsymbol{z}}_{i,\mathrm{d}} + w_{i,\mathrm{d}}\big(\boldsymbol{z}_{i,\mathrm{d}}^{\mathrm{ref}} - \hat{\boldsymbol{z}}_{i,\mathrm{d}}\big), \tag{12}$$

followed by $\boldsymbol{x}_{0,i}^{\mathrm{fuse}} = W^\top \boldsymbol{z}_i^{\mathrm{fuse}}$. Importantly, Haar fusion acts on predictor-relative corrections after Langevin refinement. It neither changes the likelihood gradient nor directly filters the predicted image. Coarse corrections are committed fully. Detail corrections are adopted progressively using the exposure-linked gate $w_{i,\mathrm{d}}$ defined by

$$w_{i,\mathrm{d}} = \underbrace{\Big(d_s + (d_e - d_s)\rho_i^{\gamma}\Big)}_{\mathrm{unlock}(i)} \cdot \underbrace{(1-\lambda_i)}_{\text{gated schedule}}, \qquad \rho_i = \frac{N-i}{N-1}. \tag{13}$$

The factor $1-\lambda_i$ links post-refinement detail adoption to the exposure schedule. The factor $\mathrm{unlock}(i)$ increases monotonically from d_s to d_e, with $\gamma > 0$ controlling its curvature. Since $0 \le d_s \le d_e \le 1$ and $\lambda_i \in [0,1]$, we have $w_{i,d} \in [0,1]$.

Finally, we obtain the next state by re-noising the fused estimate according to the noise schedule

$$\boldsymbol{x}_{\tau_{i-1}} \sim \mathcal{N}\big(\boldsymbol{x}_{0,i}^{\mathrm{fuse}}, \sigma_{i-1}^2 \boldsymbol{I}\big). \tag{14}$$

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

We use the pretrained FFHQ diffusion model from Chung et al. (2023b) and the pretrained ImageNet diffusion model from Dhariwal & Nichol (2021). We evaluate four representative linear inverse problems: $4\times$ bicubic super-resolution, motion deblurring, Gaussian deblurring, and random inpainting. Following Chung et al. (2023b); Zhang et al. (2025), we synthesize degraded observations using 61×61 blur kernels for deblurring and 70% random masks for inpainting. All experiments target 256×256 outputs and are evaluated by PSNR, SSIM, and LPIPS. We set the measurement noise level to $\sigma_y = 0.05$ in Eq. 1 and, unless otherwise specified, evaluate on 100 validation images from each dataset. Additional implementation details and hyperparameters are provided in Appendix A.4.

We compare with representative baselines, including posterior-sampling methods such as DAPS Zhang et al. (2025), FGPS Thaker et al. (2025), FPS Dou & Song (2024), DCDP Janati et al. (2024),



Figure 3: **Qualitative comparisons on FFHQ and ImageNet.** The first four rows show FFHQ results, and the last four rows show ImageNet results. Our method better preserves structure and details while suppressing ringing and spurious high-frequency artifacts under challenging degradations.

DPS Chung et al. (2023b), and ReSample Song et al. (2024), as well as diffusion-prior solvers such as DDNM Wang et al. (2023) and DDRM Kawar et al. (2022). We use official implementations whenever available and otherwise follow the settings reported in Zhang et al. (2025). Across methods, we align the measurement operators, noise levels, output resolution, and metric implementation. Since official sampler configurations of listed baselines differ in score evaluations, inner optimization steps, solvers, and backbones, we report computational cost alongside restoration quality in Tab. 1. Unless otherwise stated, our method uses a single default hyperparameter configuration across datasets and degradation types, without per-image or per-baseline tuning.

4.2 Main Results

We report quantitative results across four restoration tasks on both FFHQ and ImageNet in Tab. 1. Overall, our method achieves comparable or superior performance across datasets and degradation types, with particularly large gains on $4\times$ super-resolution, motion deblurring, and random inpainting. Although several baselines are competitive on ImageNet Gaussian deblurring, their advantages do not consistently transfer to other degradations, whereas our method maintains strong overall performance

Table 1: **Quantitative results on four linear inverse problems.** The best and second-best results are marked in ***bold*** and *underlined*, respectively. Runtime (s/image), peak memory (GB), and diffusion NFE are reported in the rightmost columns.

Dataset	Method	4× Super-resolution			Motion Deblurring			Gaussian Deblurring			Inpainting (random)			Efficiency		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	Time	Mem.	NFE
FFHQ	DAPS	29.35	0.782	0.193	29.66	0.847	0.157	29.19	0.817	**0.165**	30.72	0.800	0.159	80	2.28	1000
	DCDP	28.66	0.807	0.178	25.08	0.512	0.364	27.50	0.699	0.304	30.69	0.842	0.142	**2**	4.93	181
	DDNM	28.03	0.795	0.197	31.12	0.871	0.152	28.20	0.804	0.216	29.91	0.817	0.121	5	4.70	100
	DDRM	26.58	0.782	0.282	31.54	0.853	0.173	24.93	0.732	0.239	28.46	0.822	0.183	5	4.77	**20**
	DPS	25.86	0.753	0.269	24.52	0.801	0.246	25.87	0.764	0.219	25.46	0.823	0.203	65	9.44	1000
	FGPS	-	-	-	21.70	0.574	0.294	22.01	0.601	0.267	-	-	-	110	4.65	1000
	FPS	28.42	0.813	0.204	27.39	0.826	0.227	26.54	0.773	0.253	28.21	0.823	0.261	100	21.56	1000
	ReSample	23.29	0.594	0.392	27.41	0.823	0.198	26.39	0.714	0.255	29.61	0.746	0.140	780	6.77	500
	Ours	**31.88**	**0.890**	**0.090**	**36.69**	**0.940**	**0.054**	**29.99**	**0.821**	0.175	**32.86**	**0.909**	**0.066**	130	**2.23**	1000
ImageNet	DAPS	25.47	0.637	0.294	27.86	0.766	0.196	25.89	0.658	0.268	28.44	0.775	0.135	160	4.68	1000
	DCDP	25.17	0.688	0.213	-	-	-	26.08	**0.727**	**0.195**	20.55	0.864	0.101	**5**	12.55	181
	DDNM	23.61	0.658	0.428	28.41	0.791	0.230	**28.06**	0.703	0.278	23.64	0.821	0.107	15	6.37	100
	DDRM	24.13	0.633	0.304	28.48	0.784	0.182	26.33	0.713	0.243	26.34	0.764	0.181	**5**	6.44	**20**
	DPS	21.13	0.489	0.361	18.96	0.629	0.423	20.31	0.598	0.397	23.52	0.745	0.297	200	9.44	1000
	FGPS	-	-	-	21.87	0.560	0.288	23.53	0.583	0.305	-	-	-	375	11.08	1000
	FPS	24.82	0.703	0.313	24.52	0.647	0.326	23.91	0.601	0.387	24.52	0.701	0.316	150*	57.66	1000
	ReSample	22.61	0.576	0.370	26.94	0.738	0.227	25.97	0.703	0.254	27.50	0.756	0.143	780	6.77	500
	Ours	**27.67**	**0.765**	**0.170**	**34.70**	**0.935**	0.155	26.02	0.672	0.283	**28.60**	**0.887**	**0.042**	200	**4.56**	1000

*Measured on an RTX Pro 6000 Blackwell due to A6000 memory limits, a GPU approximately 2.5× faster than the A6000.

across diverse linear inverse problems. The "-" entries denote settings that are not natively supported by official implementations, and are omitted to avoid unreliable re-implementations. For the listed baselines, we adopt the default settings specified in their corresponding papers and open-source implementations, with a brief summary provided in Appendix A.4.

Additional box-inpainting, reduced-budget, robustness, high-resolution, and sensitivity results are provided in Appendix A.3, further evaluating the method across diverse settings and computational budgets, with Tabs. 8 and 9 demonstrating a favorable quality–efficiency trade-off. Fig. 3 shows improved structural preservation and detail recovery, with reduced ringing and spurious high-frequency artifacts on both datasets. Trajectory visualizations in Appendix A.3.1 further illustrate the coarse-to-fine behavior induced by the expanding frequency-exposure schedule.

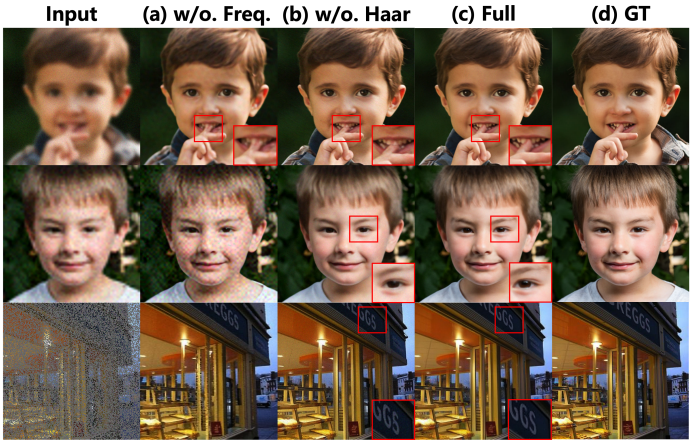

Figure 4: **Visual ablation of frequency continuation and Haar commitment.** We visualize two ablation variants: column (a) replaces noise-conditioned frequency continuation with full-band guidance, and column (b) removes the Haar fusion step in Eq. 12. The first two rows are evaluated on FFHQ, and the third row is evaluated on ImageNet.

4.3 Ablation Studies

We ablate the proposed posterior sampling framework to isolate the contribution of each design choice and to verify the mechanistic claims in Sec. 3.2 and Sec. 3.3. Specifically, we examine fre-

quency continuation, Haar commitment, and alternatives to the Haar-based fusion design. Parameter sensitivity is reported in Appendix A.3.4.

Frequency continuation and Haar commitment. We first ablate the two core components, as visualized in Fig. 4. We remove frequency continuation by setting $\lambda_i = 0$ only in the measurement objective of Eq. 8, yielding full-band guidance. The commitment weights $w_{i,\mathrm{d}}$ remain identical to those of the full model at every outer step and are not recomputed using the modified likelihood weight. We remove Haar commitment by directly outputting $\boldsymbol{x}_0^{(T)}$ from Eq. 10, without the coarse-detail fusion in Eq. 12. Columns (a) and (b) in Fig. 4 visualize these two removals, respectively. Full-band guidance tends to introduce drift and operator-sensitive artifacts at high noise, consistent with Theorem A.6 and Corollary A.7, while removing Haar fusion degrades detail quality, matching Lemma A.11 and Corollary A.12. These variants separate the two roles: frequency continuation controls trusted measurement directions, whereas Haar commitment controls reliable detail adoption. These observations are further supported by the quantitative results in Tab. 2.

Table 2: **Quantitative ablation of frequency continuation and Haar commitment.** We compare the full model with variants that remove frequency continuation or Haar commitment on FFHQ.

Method	Super-resolution			Motion Deblurring			Gaussian Deblurring			Inpainting		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
w/o. Freq	27.17	0.655	0.342	29.46	0.723	0.204	24.15	<u>0.774</u>	**0.160**	24.67	0.784	0.163
w/o. Haar	<u>29.60</u>	<u>0.804</u>	<u>0.186</u>	<u>32.78</u>	0.879	0.132	<u>24.54</u>	0.745	0.177	<u>29.58</u>	<u>0.841</u>	0.140
Full	**31.88**	**0.890**	**0.090**	**36.69**	**0.940**	**0.054**	**29.99**	**0.821**	<u>0.175</u>	**32.86**	**0.909**	**0.066**

Haar basis and frequency-reweighting alternatives. We further examine the design choice of Haar fusion and whether it can be replaced by simpler frequency-domain reweighting. We use motion deblurring as a representative ill-conditioned setting because it is particularly sensitive to high-frequency artifacts and operator-induced drift. Tab. 3 shows that Haar achieves the best PSNR and SSIM on both datasets and the best overall trade-off among the tested Daubechies-2 and Biorthogonal 4.4 bases in this setting, with qualitative comparisons provided in Appendix A.3.5 and Fig. 11. The table also includes a frequency-dependent measurement-reweighting alternative without Haar fusion, which performs substantially worse than the full model. This suggests that pre-update spectral weighting alone is insufficient to replace post-refinement detail commitment.

Table 3: **Ablation studies of wavelet bases and frequency-reweighted updates** on motion deblurring. The first block compares different wavelet bases for the commitment module, while the second block tests replacing Haar commitment with frequency-dependent measurement-consistency updates.

Ablation	Method	FFHQ			ImageNet		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Wavelet basis	Daubechies-2	26.82	<u>0.890</u>	<u>0.119</u>	24.26	<u>0.856</u>	**0.142**
	Biorthogonal 4.4	<u>30.59</u>	0.876	0.151	<u>27.64</u>	0.811	0.186
Freq. reweighting	Freq.-weighted update w/o Haar	14.88	0.179	0.583	15.02	0.244	0.482
Full	Haar + Freq.	**36.69**	**0.940**	**0.054**	**34.70**	**0.935**	<u>0.155</u>

5 Conclusion

We presented a posterior continuation framework for diffusion inverse problems that coordinates noise-conditioned frequency exposure with staged detail commitment. The central principle is to coordinate measurement guidance during refinement with correction adoption afterward. Frequency exposure defines a sequence of intermediate posteriors that emphasize low-frequency residuals at high noise and gradually return to full-band measurement consistency as the noise decreases. To realize this principle, we introduced a sampler that combines diffusion prediction, frequency-weighted Langevin refinement, and a post-refinement Haar-domain commitment rule. Experiments on super-resolution, inpainting, and Gaussian and motion deblurring demonstrate strong restoration performance across datasets and operators, with particularly large gains on ill-conditioned motion blur. Ablations further support the coupled design of frequency exposure and Haar commitment, showing that restoration benefits from combining noise-conditioned measurement guidance with post-refinement detail commitment.

AI USE STATEMENT

In this work, we used generative AI tools to assist with language editing of the manuscript, including correcting grammar, refining phrasing, and improving clarity and readability. The authors reviewed all AI-assisted revisions to ensure that they preserved the intended scientific meaning and did not introduce unsupported claims or alter the reported results. We take full responsibility for the final content of this work, including all text, claims, and artifacts produced with the assistance of generative AI.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we provide a detailed description of the proposed method and its algorithmic procedure in Section 3. The datasets, experimental settings, and evaluation protocols are described in Section 4. Additional implementation details, including hyperparameter settings and algorithmic configurations, are provided in Appendix A.4. Together, these descriptions document the procedures and configurations used to obtain the reported results and support independent reproduction of our experiments.

REFERENCES

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 1982.

Riccardo Barbano, Alexander Denker, Hyungjin Chung, Tae-Hoon Roh, Simon Arrdige, Peter Maass, Bangti Jin, and Jong Chul Ye. Steerable conditional diffusion for out-of-distribution adaptation in imaging inverse problems. *CoRR*, 2023.

Johnathan M. Bardsley, Antti Solonen, Heikki Haario, and Marko Laine. Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. *SIAM J. Sci. Comput.*, 2014.

Michael Betancourt and Mark Girolami. Hamiltonian monte carlo for hierarchical models. *Current trends in Bayesian methodology with applications*, 2015.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. Compressed sensing using generative models. In *ICML*, 2017.

Stephen P. Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 2011.

Benjamin Boys, Mark Girolami, Jakiw Pidstrigach, Sebastian Reich, Alan Mosca, and Ömer Deniz Akyildiz. Tweedie moment projected diffusions for inverse problems. *Trans. Mach. Learn. Res.*, 2024.

Emmanuel J. Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Commun. ACM*, 2012.

Emmanuel J Candès, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 2006.

Gabriel Cardoso, Yazid Janati El Idrissi, Sylvain Le Corff, and Eric Moulines. Monte carlo guided denoising diffusion models for bayesian linear inverse problems. In *ICLR*, 2024.

Stanley H. Chan, Xiran Wang, and Omar A. Elgendy. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Trans. Computational Imaging*, 2017.

Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: conditioning method for denoising diffusion probabilistic models. In *ICCV*, 2021.

Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *CVPR*, 2023a.

Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *ICLR*, 2023b.

Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G. Dimakis, and Mauricio Delbracio. A survey on diffusion models for inverse problems. *CoRR*, 2024.

Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, 2021.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *ICLR*, 2024.

Fabian Falck, Teodora Pandeva, Kiarash Zahirnia, Rachel Lawrence, Richard E. Turner, Edward Meeds, Javier Zazo, and Sushrut Karmalkar. A fourier space perspective on diffusion models. *CoRR*, 2025.

Zhengyang Geng, Yiyang Lu, Zongze Wu, Eli Shechtman, J. Zico Kolter, and Kaiming He. Improved mean flows: On the challenges of fastforward generative models. In *CVPR*, 2026.

Davis Gilton, Gregory Ongie, and Rebecca Willett. Deep equilibrium architectures for inverse problems in imaging. *IEEE Trans. Computational Imaging*, 2021.

Dan Hendrycks and Thomas G. Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. In *ICLR*, 2019.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *CoRR*, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

Yazid Janati, Badr Moufad, Alain Durmus, Eric Moulines, and Jimmy Olsson. Divide-and-conquer posterior sampling for denoising diffusion priors. In *NeurIPS*, 2024.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *NeurIPS*, 2022.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *NeurIPS*, 2022.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *NeurIPS*, 2022.

Andreas Lugmayr, Martin Danelljan, Andrés Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022.

Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.*, 2021.

Yurui Qian, Qi Cai, Yingwei Pan, Yehao Li, Ting Yao, Qibin Sun, and Tao Mei. Boosting diffusion models with moving average sampling in frequency domain. In *CVPR*, 2024.

Ron Raphaeli, Sean Man, and Michael Elad. SILO: solving inverse problems with latent operators. In *ICCV*, 2025.

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 *ICLR*, 2024.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021a.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021b.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *ICML*, 2023.

Darshan Thaker, Abhishek Goyal, and René Vidal. Frequency-guided posterior sampling for diffusion-based image restoration. In *ICCV*, 2025.

Feng Tian, Yixuan Li, Yichao Yan, Shanyan Guan, Yanhao Ge, and Xiaokang Yang. Postedit: Posterior sampling for efficient zero-shot image editing. In *ICLR*, 2025.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Comput.*, 2011.

Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *ICLR*, 2023.

Max Welling and Yee Whye Teh. Bayesian learning via stochastic gradient langevin dynamics. In *ICML*, 2011.

Xingyu Xu and Yuejie Chi. Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. In *NeurIPS*, 2024.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *CVPR*, 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 *CVPR*, 2023.

A Appendix

A.1 Related Work

Recent posterior sampling methods for inverse problems increasingly adopt diffusion and score based generative models as expressive implicit priors. Foundational results connect reverse time diffusion to score estimation and denoising, including Reverse time diffusion equation models Anderson (1982) and A connection between score matching and denoising autoencoders Vincent (2011). Modern diffusion backbones include Denoising diffusion probabilistic models Ho et al. (2020) and implicit sampling via Denoising diffusion implicit models Song et al. (2021a), with a continuous time formulation in Score based generative modeling through stochastic differential equations Song et al. (2021b). Sampling quality and efficiency further benefit from EDM Karras et al. (2022), fast ODE solvers such as DPM-Solver Lu et al. (2022), and guidance mechanisms such as Classifier-free diffusion guidance Ho & Salimans (2022). These advances make diffusion priors practical plug in components for Bayesian posterior inference under ill posed degradations.

Prior to diffusion priors, inverse problems were largely addressed by explicit regularization and convex recovery, exemplified by Stable signal recovery from incomplete and inaccurate measurements Candès et al. (2006) and Exact matrix completion via convex optimization Candès & Recht (2012). Learned generative priors then enabled model based recovery with neural generators, including Compressed sensing using generative models Bora et al. (2017) and high fidelity GAN backbones such as A style based generator architecture for generative adversarial networks Karras et al. (2019). Plug and play optimization further linked model based solvers with learned denoisers, with convergence oriented instantiations such as Plug and play ADMM for image restoration Fixed point convergence and applications Chan et al. (2017) grounded in ADMM Boyd et al. (2011). Related directions include algorithm unrolling Monga et al. (2021) and posterior sampling toolkits beyond MAP, such as SGLD Welling & Teh (2011), HMC Betancourt & Girolami (2015), and Randomize then optimize Bardsley et al. (2014). This trajectory motivates diffusion priors as a stronger learned prior that still interfaces naturally with likelihood constrained inference.

Building on these foundations, diffusion based posterior sampling directly targets Bayesian inference with diffusion priors, including DPS Chung et al. (2023b). Sequential viewpoints formalize diffusion inverse sampling through filtering style inference Dou & Song (2024), while other efforts improve likelihood incorporation and posterior score surrogates, including Monte Carlo guided variants Cardoso et al. (2024) and Tweedie moment–based projected dynamics Boys et al. (2024). To address local correction induced by small reverse time steps, decoupling strategies such as DAPS Zhang et al. (2025) enable larger and more non local posterior moves. In parallel, projection and data consistency based diffusion restorers remain influential, including DDRM Kawar et al. (2022) and operator aware null space formulations such as DDNM Wang et al. (2023). For conditional editing and inpainting, representative mechanisms include RePaint Lugmayr et al. (2022) and ILVR Choi et al. (2021).

Beyond standard degradations, robustness and distribution shift are increasingly emphasized. Steerable conditional diffusion explicitly targets OOD inverse problems Barbano et al. (2023), aligning with broader robustness benchmarks Hendrycks & Dietterich (2019). Equilibrium formulations provide an alternative lens via deep equilibrium architectures Gilton et al. (2021). Recent surveys consolidate diffusion inverse methods and taxonomies Daras et al. (2024).

A.2 Pseudo-code of The Algorithm

We provide the corresponding Pseudo-code for our method in Alg. 1.

A.3 Additional Experimental Results

A.3.1 Trajectory Visualization of Noise-Conditioned Frequency Exposure

To better understand the sampling dynamics, we visualize the intermediate trajectory induced by the noise-conditioned frequency exposure schedule. Fig. 5(a) shows the masked Fourier magnitude maps at different stages, where the exposed bandwidth gradually expands as the diffusion noise decreases. Fig. 5(b) presents the corresponding spatial-domain reconstructions. At early stages, only low-frequency measurement directions are exposed, so the reconstruction mainly captures coarse layout,

Algorithm 1 Noise-Conditioned Posterior Continuation

Require: Diffusion model ε_θ, clean-estimate solver f, measurement $(\boldsymbol{y}, \mathcal{A})$, number of outer sampling steps N, sampling times $\{\tau_i\}_{i=0}^N$, noise levels $\{\sigma_i\}_{i=0}^N$ with $\sigma_i = \sigma(\tau_i)$ and $\sigma_N > \cdots > \sigma_0 = 0$, exposure schedules $\omega(\sigma)$ and $\lambda(\sigma)$, temperature schedule $\{\beta_i\}_{i=1}^N$, Langevin step sizes $\{\eta_{i,\ell}\}$, Langevin steps T, Haar transform W, Haar gate parameters d_s, d_e, γ.
Initialize $\boldsymbol{x}_{\tau_N} \sim \mathcal{N}(\mathbf{0}, \sigma_N^2 \boldsymbol{I})$.
for $i = N$ to 1 **do**
  $\hat{\boldsymbol{x}}_{0,i} \leftarrow f(\boldsymbol{x}_{\tau_i}, \varepsilon_\theta, \sigma_i)$
  // Diffusion predictor. Estimate the clean image at the current noise level.
  $\omega_i \leftarrow \omega(\sigma_i)$, $\lambda_i \leftarrow \lambda(\sigma_i)$
  $\mathcal{L}_i(\boldsymbol{x}; \boldsymbol{y}) \leftarrow (1-\lambda_i)\|\mathcal{A}(\boldsymbol{x}) - \boldsymbol{y}\|_2^2 + \lambda_i \|P_{\omega_i}^F(\mathcal{A}(\boldsymbol{x}) - \boldsymbol{y})\|_2^2$
  $\pi_i(\boldsymbol{x}_0) \propto \exp\left(-\frac{1}{2\beta_i^2}\mathcal{L}_i(\boldsymbol{x}_0; \boldsymbol{y})\right) \mathcal{N}(\boldsymbol{x}_0; \hat{\boldsymbol{x}}_{0,i}, \sigma_i^2 \boldsymbol{I})$
  // Frequency exposure. Unreliable high-frequency residuals are attenuated.
  $\boldsymbol{x}_{0,i}^{(0)} \leftarrow \hat{\boldsymbol{x}}_{0,i}$
  for $\ell = 0$ to $T-1$ **do**
    Sample $\boldsymbol{\xi}_{i,\ell} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$
    $\boldsymbol{x}_{0,i}^{(\ell+1)} \leftarrow \boldsymbol{x}_{0,i}^{(\ell)} + \eta_{i,\ell} \nabla_{\boldsymbol{x}_0} \log \pi_i(\boldsymbol{x}_{0,i}^{(\ell)}) + \sqrt{2\eta_{i,\ell}}\boldsymbol{\xi}_{i,\ell}$
  end for
  $\boldsymbol{x}_{0,i}^{\mathrm{ref}} \leftarrow \boldsymbol{x}_{0,i}^{(T)}$
  // Langevin refinement under the exposed likelihood.
  $\hat{\boldsymbol{z}}_i \leftarrow W\hat{\boldsymbol{x}}_{0,i}$, $\boldsymbol{z}_i^{\mathrm{ref}} \leftarrow W\boldsymbol{x}_{0,i}^{\mathrm{ref}}$
  $\rho_i \leftarrow (N-i)/(N-1)$
  $w_{i,\mathrm{c}} \leftarrow 1$, $w_{i,\mathrm{d}} \leftarrow [d_s + (d_e - d_s)\rho_i^\gamma](1-\lambda_i)$
  $\boldsymbol{z}_{i,b}^{\mathrm{fuse}} \leftarrow (1-w_{i,b})\hat{\boldsymbol{z}}_{i,b} + w_{i,b}\boldsymbol{z}_{i,b}^{\mathrm{ref}}, \quad b \in \{\mathrm{c}, \mathrm{d}\}$
  $\boldsymbol{x}_{0,i}^{\mathrm{fuse}} \leftarrow W^\top \boldsymbol{z}_i^{\mathrm{fuse}}$
  // Post-refinement Haar commitment. Coarse corrections are committed, details are gated.
  $\boldsymbol{x}_{\tau_{i-1}} \sim \mathcal{N}(\boldsymbol{x}_{0,i}^{\mathrm{fuse}}, \sigma_{i-1}^2 \boldsymbol{I})$
  // Re-noise the committed estimate for the next reverse step.
end for
return $\boldsymbol{x}_{0,1}^{\mathrm{fuse}}$

silhouette, and global illumination. As the cutoff frequency increases, higher-frequency components are progressively introduced, leading to sharper edges, textures, and fine-scale details. Fig. 5(c) further visualizes the sampling trajectory and shows that the reconstruction evolves consistently with the frequency-exposure schedule.

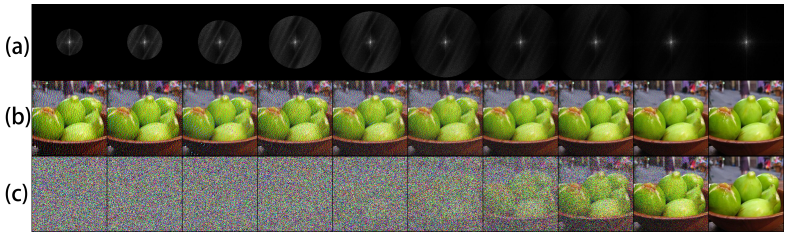

Figure 5: **Example visualization of noise-conditioned frequency exposure on ImageNet.** (a) Masked Fourier magnitude maps induced by the cutoff-frequency schedule. (b) Spatial-domain reconstructions along the progressive frequency-exposure trajectory. (c) Corresponding sampling trajectory.

We also track quantitative metrics along the sampling process in Fig. 6 and Fig. 7. Fig. 6 reports the metric evolution for FFHQ motion deblurring, while Fig. 7 shows the corresponding results for FFHQ Gaussian deblurring. In both figures, the orange curves report metrics computed from the intermediate estimates produced by the refinement update in Eq. 10, while the red/green curves report metrics computed from the denoised estimates after each solver step. Overall, PSNR and

SSIM improve while LPIPS decreases along the trajectory, indicating that the proposed sampler progressively enhances both measurement consistency and perceptual quality. These results further support the intended coarse-to-fine behavior of our continuation process: early updates stabilize reliable low-frequency structure, while later updates commit increasingly fine details as they become more identifiable.

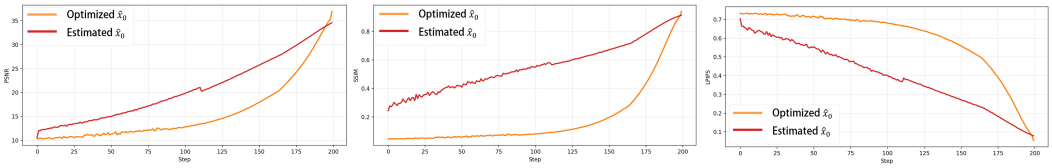

Figure 6: **Metric evolution during sampling for FFHQ motion deblurring.** The orange curves report metrics computed from the intermediate estimates produced by the refinement update in Eq. 10, while the red curves report metrics computed from the denoised estimates after each solver step.

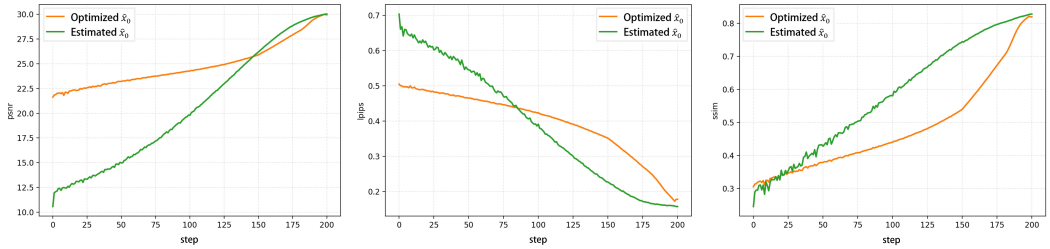

Figure 7: **Metric evolution during sampling for FFHQ Gaussian deblurring.** The orange curves report metrics computed from the intermediate estimates produced by the refinement update in Eq. 10, while the green curves report metrics computed from the denoised estimates after each solver step.

A.3.2 Additional Results on Box Inpainting

We further evaluate our method on box inpainting, where the missing region is spatially contiguous. As shown in Tab. 4, our method remains competitive and achieves the best LPIPS on both FFHQ and ImageNet. It also obtains the best SSIM on FFHQ and stays close to the strongest baselines in PSNR and SSIM on ImageNet. The gains are smaller than in deblurring or super-resolution because the box mask only constrains visible context, leaving the missing region unobserved. Since masking mixes frequencies in the Fourier domain, the low-frequency residual mainly enforces visible-region consistency and boundary compatibility rather than revealing the true masked structure. Thus, Haar fusion cannot create missing semantics, and it primarily prevents premature commitment to unreliable details, as qualitatively illustrated in Fig. 8.

Table 4: **Quantitative results on box inpainting.** We report results on FFHQ and ImageNet at 256×256. The best and second-best results are marked in ***bold*** and *underlined*, respectively.

Method	FFHQ			ImageNet		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
DAPS	24.07	0.814	0.133	21.43	0.725	0.214
DCDP	23.89	0.760	0.163	-	-	-
DDNM	24.47	<u>0.837</u>	0.235	21.64	**0.748**	0.319
DDRM	22.26	0.801	0.207	18.63	0.733	0.254
DPS	22.51	0.792	0.209	18.94	0.722	0.257
FPS	**24.86**	0.823	0.146	**22.16**	0.726	<u>0.208</u>
ReSample	20.06	0.749	0.184	18.29	0.631	0.262
Ours	<u>24.57</u>	**0.838**	**0.126**	<u>21.58</u>	<u>0.739</u>	**0.207**

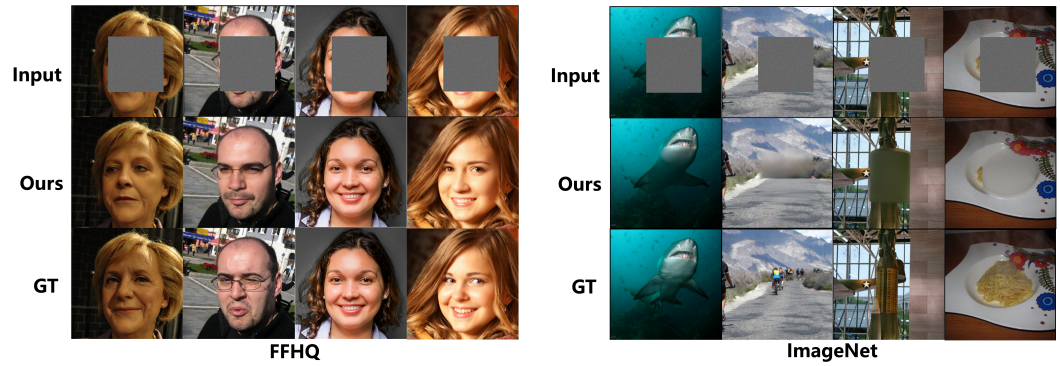

Figure 8: **Qualitative examples of box inpainting.** Box masks leave the missing region unobserved, making semantic recovery more challenging than deblurring or super-resolution.

A.3.3 Robustness, Scalability, and Boundary Conditions

We further evaluate the robustness, scalability, and boundary conditions of our design under real-world blur, 512×512 latent-space restoration, and operator mismatch. These experiments clarify when the proposed continuation remains effective and where its assumptions become limiting. Overall, our method is most effective when the assumed likelihood is approximately correct and reliable measurements follow a coarse-to-fine structure.

Real-world blur. For real-world blur, we follow BlindDPS Chung et al. (2023a). We provide qualitative examples under real-world blur in Fig. 9. The results show that our method remains robust beyond matched synthetic degradations and still outperforms the compared baseline.

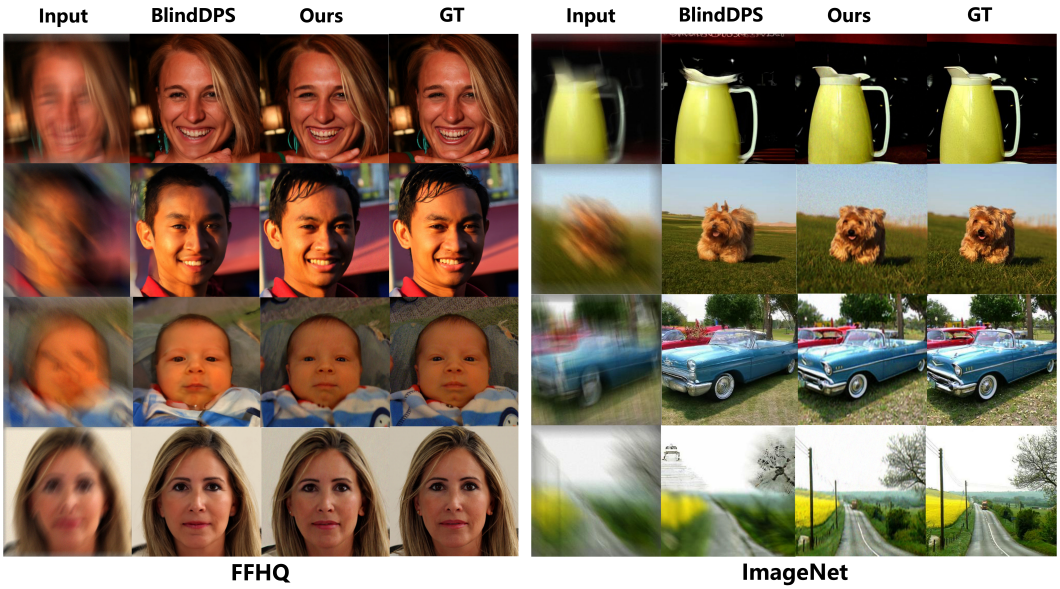

Figure 9: **Qualitative examples of real-world blur.** Our method remains robust under real-world blur and produces clearer restorations than the compared baseline.

Stable-Diffusion-based restoration at 512×512. Since different pretrained diffusion models have different native resolution preferences, we select Stable Diffusion v1.5 as a representative latent diffusion backbone for 512×512 restoration. To verify the effectiveness of our method for higher-resolution restoration in latent space, we compare it with the recent SILO method Raphaeli et al. (2025), which also adopts Stable Diffusion v1.5 and operates at 512×512 resolution. Tab. 5 and

Fig. 10 report the quantitative and qualitative comparison between SILO and our method. Compared with pixel-space restoration, latent-space posterior sampling is more sensitive to posterior geometry, decoder distortion, representation bottlenecks, and text-conditioning effects.

For box inpainting, the improvement is smaller because a contiguous mask only constrains the visible context and does not directly reveal the missing-region semantics. Nevertheless, our method remains competitive, achieving better SSIM and LPIPS than SILO under the aligned 512×512 setting.

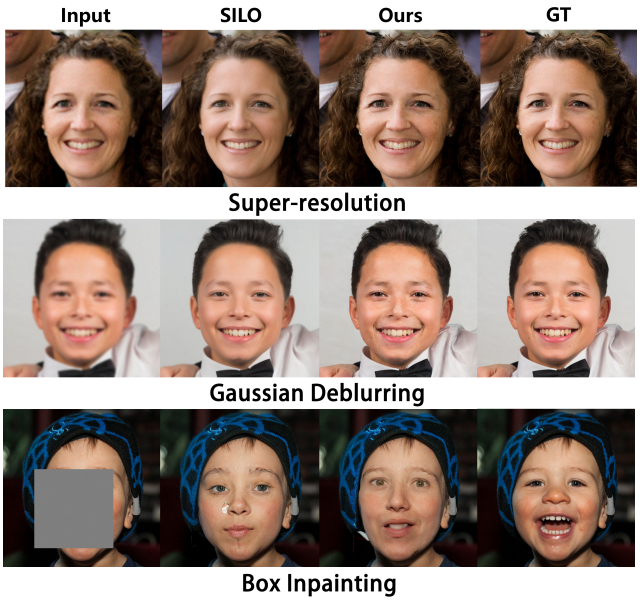

Figure 10: **Qualitative comparison with SILO on FFHQ** 512×512**.** Both methods use Stable Diffusion v1.5 as the backbone. Our method produces competitive restorations with sharper visible details in these examples.

Table 5: **Quantitative comparison with SILO on FFHQ** (512×512)**.** We report results on three restoration tasks.

Methods	Gaussian Deblurring			Super-resolution			Box Inpainting		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SILO (512×512)	<u>26.53</u>	<u>0.745</u>	<u>0.317</u>	<u>26.79</u>	<u>0.761</u>	<u>0.297</u>	**22.73**	<u>0.782</u>	<u>0.237</u>
Ours (512×512)	**27.29**	**0.762**	**0.232**	**28.04**	**0.803**	**0.177**	<u>22.62</u>	**0.785**	**0.232**

Operator mismatch. We next examine the boundary of our method under operator mismatch. The observation is generated by the true blur operator $\mathcal{A}_{\text{true}}$, with kernel size 61 and intensity 0.50, while reconstruction uses a perturbed operator $\mathcal{A}_{\text{mismatch}}$. As shown in Tab. 6, performance drops substantially when the assumed operator deviates from the true one. This is expected because the sampler targets a biased posterior induced by the misspecified likelihood rather than the posterior induced by $\mathcal{A}_{\text{true}}$. The measurement-consistency term is effectively computed as

$$\|\mathcal{A}_{\text{mismatch}}(\hat{\boldsymbol{x}}_0) - \mathcal{A}_{\text{true}}(\boldsymbol{x}_{\text{gt}})\|_2^2,$$

where $\hat{\boldsymbol{x}}_0$ and $\boldsymbol{x}_{\text{gt}}$ denote the estimated and ground-truth images. Thus, our method can stabilize sampling under a given likelihood, but it cannot correct a systematically misspecified forward model. This represents a natural failure mode when the assumed degradation operator is substantially wrong.

A.3.4 Parameter Sensitivity

We analyze the sensitivity of key sampling and fusion parameters on FFHQ motion deblurring. We first study the interaction between the continuation steps N and inner refinement steps T, which

Table 6: **Operator mismatch on FFHQ motion deblurring.** The observation is generated with the true operator of kernel size 61 and intensity 0.50, while reconstruction uses a perturbed operator.

Assumed operator	PSNR↑	SSIM↑	LPIPS↓
Kernel size 51, intensity 0.05	19.36	0.477	0.416
Kernel size 71, intensity 0.05	15.20	0.286	0.526
Kernel size 61, intensity 0.65	9.69	0.155	0.689
Kernel size 61, intensity 0.08	10.03	0.150	0.672
Kernel size 61, intensity 0.50 (Matched)	**36.69**	**0.940**	**0.054**

controls the quality–cost trade-off. We then examine the effects of the measurement noise level σ_y and representative continuation and fusion parameters.

Sampling budget N, T. Tab. 7 reports the joint sensitivity to the number of continuation steps N and inner refinement steps T. Increasing N makes the continuation trajectory finer, so the transition from low-frequency guidance to full-band consistency becomes smoother. Increasing T improves local refinement at each fixed stage. The two factors are complementary: a large T cannot fully compensate for a coarse continuation path, while a large N cannot fully compensate for insufficient local refinement. The results show that the gain saturates beyond moderate budgets and can degrade when the budget is overly large. We therefore use $N = T = 200$ as a practical quality–cost trade-off.

Table 7: **Sensitivity to sampling budget.** We report quantitative results for different combinations of the continuation steps N and inner refinement steps T of Alg. 1 on FFHQ motion deblurring.

(N, T)	PSNR↑	SSIM↑	LPIPS↓	Runtime (s/image)
$(50, 200)$	35.82	0.924	0.069	~ 35
$(100, 200)$	36.28	0.935	0.055	~ 65
$(300, 200)$	**37.08**	**0.953**	**0.039**	~ 195
$(500, 200)$	37.04	0.949	0.046	~ 280
$(200, 50)$	35.81	0.943	0.047	~ 70
$(200, 100)$	36.86	0.945	0.050	~ 110
$(200, 300)$	36.95	0.950	0.041	~ 160
$(200, 500)$	35.80	0.913	0.075	~ 215
$(300, 100)$	35.34	0.938	0.053	~ 175
$(400, 50)$	35.84	0.943	0.046	~ 200
$(500, 25)$	36.40	0.950	0.040	~ 240
$(300, 300)$	36.97	0.948	0.047	~ 225
$(400, 400)$	36.80	0.940	0.054	~ 320
$(500, 500)$	36.41	0.928	0.063	~ 425
Ours $(200, 200)$	36.69	0.940	0.054	~ 130

To further examine the low-runtime regime, we compare our method with DCDP under a similar computational budget in Tab. 8. With $N = 9$ and $T = 90$, our method runs in approximately 2 seconds per image and outperforms DCDP on FFHQ motion deblurring across all metrics. This suggests that the proposed continuation can be scaled down for efficiency while retaining competitive restoration quality.

Measurement noise level. We also evaluate different measurement noise levels σ_y. This experiment is not intended as a method-parameter ablation, since σ_y specifies the observation noise in the inverse problem. Instead, it illustrates how restoration quality changes with measurement SNR. As shown

Table 8: **Low-runtime comparison under a similar computational budget.** We compare different methods on FFHQ 256×256 under a comparable runtime budget. Our method uses a reduced sampling budget with $N = 9, T = 90$.

Method	Super-resolution			Motion Deblurring			Gaussian Deblurring			Inpainting			Time (s)
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
DCDP	28.66	0.807	0.178	25.08	0.512	0.364	**27.50**	**0.699**	**0.304**	**30.69**	**0.842**	**0.142**	∼2
Ours ($N = 9, T = 90$)	**30.61**	**0.855**	**0.104**	**27.31**	**0.603**	**0.264**	26.59	0.582	0.307	28.79	0.758	0.201	∼2

Table 9: **Low-runtime comparison under a similar computational budget.** We compare different methods on ImageNet 256×256 under a comparable runtime budget.

Method	Super-resolution			Motion Deblurring			Gaussian Deblurring			Inpainting			Time (s)
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
DCDP	25.17	0.688	0.213	–	–	–	26.08	0.727	**0.195**	20.55	**0.864**	**0.101**	∼5
Ours ($N = 9, T = 90$)	26.86	**0.741**	**0.188**	29.07	0.847	0.144	24.15	0.507	0.354	22.99	0.561	0.325	∼**4**
Ours + iMeanFlow Geng et al. (2026) ($N = 100, T = 100$)	**27.39**	0.682	0.329	**32.53**	**0.918**	**0.103**	**28.74**	**0.763**	0.259	**31.20**	0.857	0.181	∼50

in Tab. 10, performance decreases as σ_y increases, because fine-scale components become less identifiable under noisier observations. We use $\sigma_y = 0.05$ in the main experiments to follow the standard noisy inverse-problem setting and maintain a fair comparison with baselines.

Table 10: **Sensitivity to measurement noise level.** We vary the observation noise level σ_y of Eq. 1 on FFHQ motion deblurring.

σ_y	PSNR↑	SSIM↑	LPIPS↓
0.00	**38.53**	**0.971**	**0.034**
0.03	37.49	0.963	0.042
0.05 (Ours)	36.69	0.940	0.054
0.07	34.95	0.935	0.082

Other parameters. Finally, we vary representative continuation and fusion parameters one at a time while keeping all other parameters fixed. As shown in Tab. 11, reducing the low-frequency commitment weight, weakening the continuation schedule, or using an overly aggressive initial detail-unlocking value can degrade performance. In particular, increasing d_s significantly hurts restoration quality, indicating that premature detail adoption is harmful when high-frequency directions are still weakly identifiable. These results support the adopted default configuration.

Table 11: **Sensitivity to other continuation and fusion parameters.** Each row changes one parameter while keeping all remaining parameters identical to the default configuration of Tab. 12. Results are evaluated on FFHQ motion deblurring.

Settings	PSNR↑	SSIM↑	LPIPS↓
λ_i starts at 0.30 and decays to 0	35.33	**0.943**	**0.045**
$d_s = 0.3$	30.78	0.813	0.187
Ours	**36.69**	0.940	0.054

A.3.5 Additional Qualitative Results

Fig. 11 provides qualitative comparisons of different wavelet bases, complementing the quantitative ablation in Tab. 3. We further provide additional qualitative results on FFHQ and ImageNet in Fig. 12 and Fig. 13, covering motion deblurring, super-resolution, Gaussian deblurring, and random inpainting.

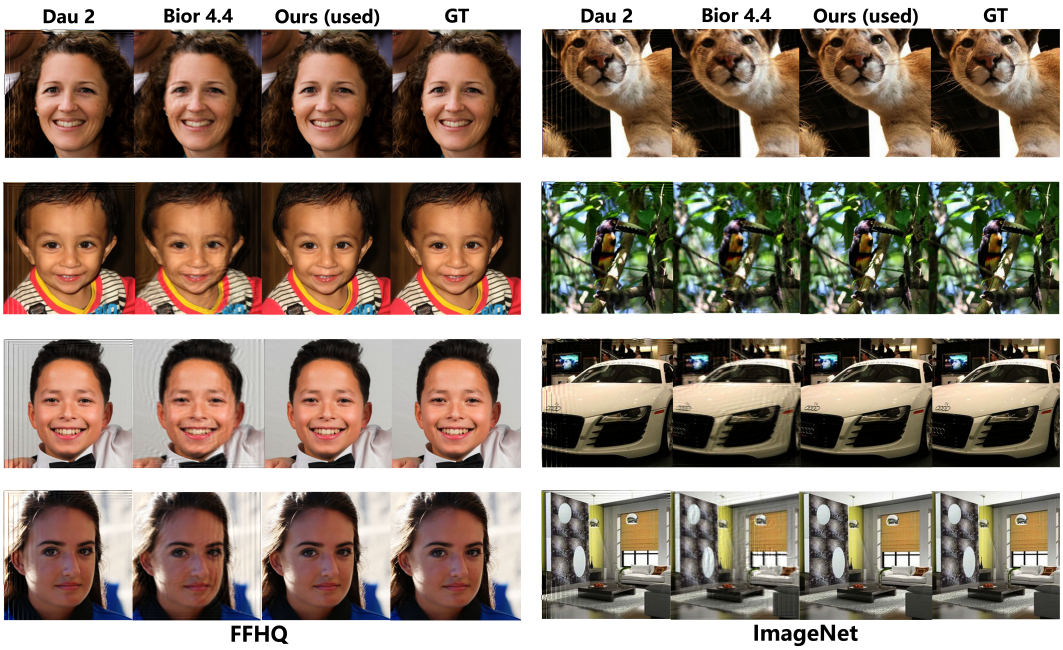

Figure 11: **Qualitative comparison of different wavelet bases.** Results are evaluated on motion deblurring at 256×256. Haar produces fewer artifacts and better preserves coarse structure and fine details among the tested bases.

Table 12: **Parameter configuration of our method.** Default hyperparameters used in our main experiments.

Parameter	Default	Description
N	200	`num_steps`, which determines the total number of diffusion iterations in line 3 of Alg. 1.
$\sigma_{\max}$	100	Initial noise level of σ used in Alg. 1.
$\sigma_{\min}$	0.1	Final noise level of σ used in Alg. 1.
σ_y in Eq. 1	0.05	Noise level for the measurement.
Timestep	poly–7	Time-step discretization scheme, polynomial with $\rho = 7$.
T	200	Number of inner Langevin refinement steps per outer iteration in Alg. 1.
Denoising Steps	5	Number of denoising steps of the PF-ODE in line 4 of Alg. 1.
Cutoff-frequency schedule	linear	Coupled to the noise level, with start and end bounds set to 0.4 and 1.0, respectively.
λ_i in Eq. 8	cosine schedule	Starts at 0.35 and decays to 0.
β_i	$\frac{0.01}{\sqrt{2}}\left[1.6 - \left(\frac{N-i}{N+1}\right)^2\right]$	Measurement-term temperature.
$\eta_{i,\ell}$	10^{-4}	Langevin step size. This value is multiplied by $1 - \frac{N-i}{N+1}$.
$w_{i,d}$ in Eq. 12	Eq. 13	Dynamic detail-commitment weight following Eq. 13.
d_s in Eq. 13	0.05	Initial value for detail unlocking.
d_e in Eq. 13	1.0	Final value for detail unlocking.
γ in Eq. 13	10.0	Controls the curvature of the detail-unlocking schedule.

A.4 Implementation Details

Diffusion Model. We use the checkpoints based on EDM Karras et al. (2022) provided by DPS Chung et al. (2023b) for all experiments, including ffhq_10m.pt (357.1 MB) and imagenet256.pt (2.06 GB).



Figure 12: **Additional qualitative results on FFHQ** 256×256**.** We show results for (a) motion deblurring, (b) super-resolution, (c) Gaussian deblurring, and (d) random inpainting. **Recons** denotes the reconstruction generated by our method.

Hyper-parameters in Alg. 1. To facilitate reproducibility, we provide the detailed parameter configurations of our method in Tab. 12. Unless otherwise stated, we use a single default hyperparameter configuration across datasets and degradation types, without per-image or per-baseline tuning. The default configuration is selected once and then kept fixed for the main comparisons.

Configurations of Other Baselines:

DPS. We follow the official DPS configuration with a DDPM sampler with 1000 linear-schedule steps, `epsilon` prediction, learned-range variance, and clipping of the denoised estimate.

DDRM. We follow the default configuration, setting $\eta_B = 1.0$ and $\eta = 0.85$, with 20 DDIM sampling steps.

DDNM. We adopt the default configuration, setting $\eta_B = 1.0$ and $\eta = 0.85$, with 100 DDIM sampling steps.

DCDP. We use the default setting described in the paper and directly employ the official open-source implementation for all reported results.

FPS. We follow the default configuration in the paper, with $M = 20$ and $N = 1000$.

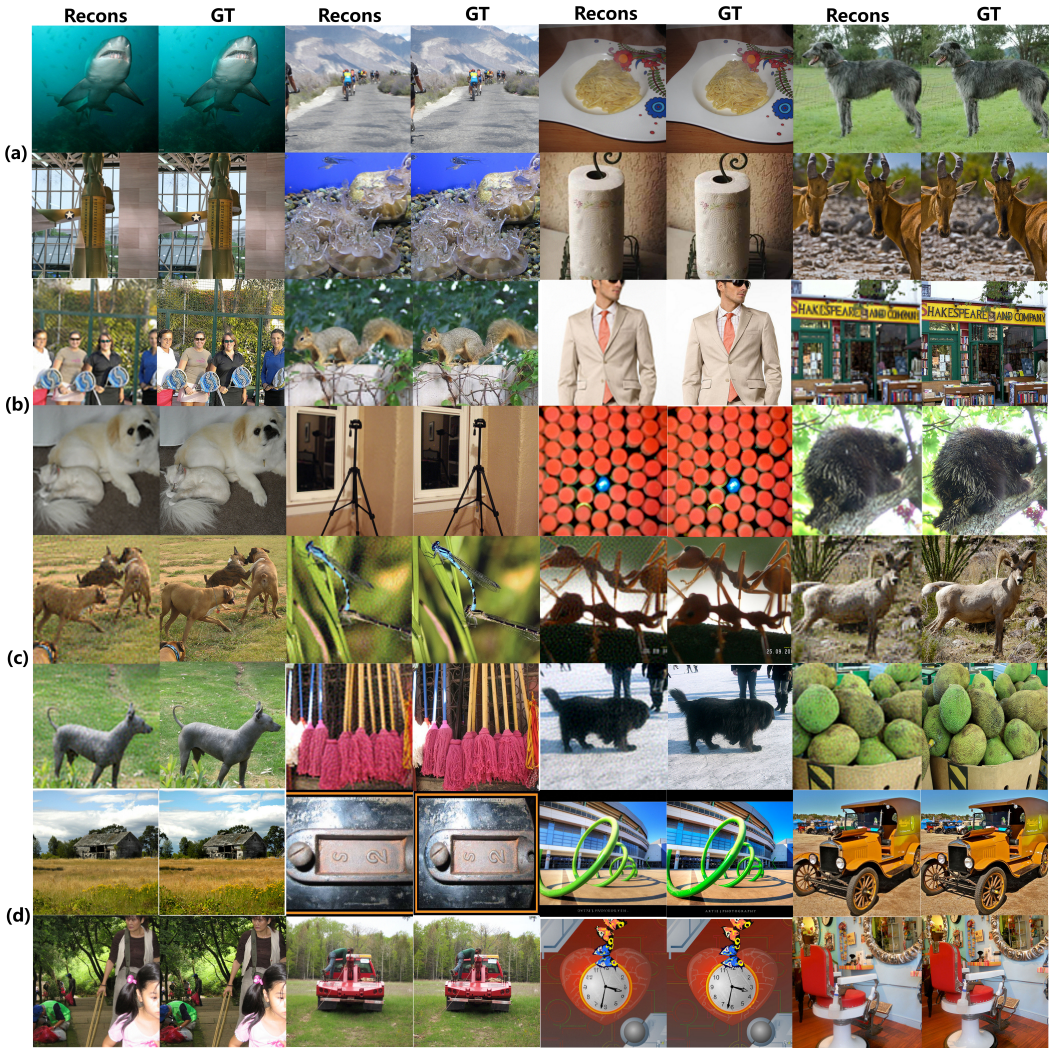

Figure 13: **Additional qualitative results on ImageNet** 256×256**.** We show results for (a) motion deblurring, (b) super-resolution, (c) Gaussian deblurring, and (d) random inpainting. **Recons** denotes the reconstruction generated by our method.

DAPS. We follow the default configuration in the paper, with $N = 200$ and 100 Langevin steps.

FGPS. We follow the default configuration in the paper, with kernel size set to 61 and intensity set to 0.5 and 3.0 for motion blur and Gaussian blur, respectively.

PSLD. We use the official implementation of PSLD with its default configurations. We use SD v1.5 for ImageNet as per its official default, providing a stronger backbone than the LDM-VQ4 used in other baselines.

ReSample. All ReSample experiments are conducted based on the official implementation from the paper and open-source code, using a DDIM sampler with 500 steps.

Device. All experiments on FFHQ 256×256 are conducted on a single NVIDIA A6000 GPU. Experiments on ImageNet 256×256 are performed on a single NVIDIA A800/A100 GPU with 80 GB of memory, or on an RTX Pro 6000 Blackwell GPU with 96 GB of memory.

A.5 THEOREM, LEMMA AND COROLLARY

We provide analysis under a standard linear-Gaussian measurement model

$$\boldsymbol{y} = \mathcal{A}\boldsymbol{x}^{\mathrm{gt}} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_y^2 \boldsymbol{I}), \tag{15}$$

where $\boldsymbol{x}^{\mathrm{gt}}$ denotes the ground-truth signal, $\mathcal{A}$ is a linear operator, and $\sigma_y > 0$ is the measurement noise level. The full-band negative log-likelihood energy is

$$\mathcal{E}(\boldsymbol{x}) = \frac{1}{2\sigma_y^2}\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2, \qquad g(\boldsymbol{x}) := \nabla\mathcal{E}(\boldsymbol{x}) = \frac{1}{\sigma_y^2}\mathcal{A}^\top(\mathcal{A}\boldsymbol{x} - \boldsymbol{y}). \tag{16}$$

For notational simplicity, the first part of the analysis uses the measurement noise scale σ_y. In the intermediate posterior of Eq. 9, β_i plays the same role as an effective likelihood temperature, so the same bounds apply after replacing σ_y by β_i.

For a spectral cutoff radius ω, let P_ω^F be the masked spectrum operator in Eq. 6. We use $\mathcal{P}_\omega$ to denote the corresponding ideal residual-space orthogonal low-pass projector used in the theoretical analysis, which satisfies $\|\mathcal{P}_\omega\|_2 \le 1$. With the unitary Fourier normalization, $\|P_\omega^F \boldsymbol{r}\|_2^2$ equals the energy of the projected residual $\|\mathcal{P}_\omega \boldsymbol{r}\|_2^2$. Thus, P_ω^F is used for the implementation-oriented spectral notation in the main method, whereas $\mathcal{P}_\omega$ is used below as the equivalent residual-space orthogonal projector.

Throughout this section, σ_i denotes the diffusion noise level at outer step i, ω_i denotes the corresponding cutoff radius, $\lambda_i \in [0, 1]$ denotes the continuation weight, $\hat{\boldsymbol{x}}_{0,i}$ denotes the PF-ODE anchor, and β_i is the temperature coefficient in Eq. 9. We write

$$\Delta_s(\boldsymbol{x}, \sigma) = s_\theta(\boldsymbol{x}, \sigma) - s^*(\boldsymbol{x}, \sigma)$$

for the score-estimation error, where s^* denotes the oracle score.

Lemma A.1 (Operator conditioning controls likelihood-gradient sensitivity)**.** *The likelihood energy $\mathcal{E}$ is L-smooth with*

$$L = \frac{\|\mathcal{A}\|_2^2}{\sigma_y^2}. \tag{17}$$

Moreover, for any $\boldsymbol{x}_1, \boldsymbol{x}_2$,

$$\|g(\boldsymbol{x}_1) - g(\boldsymbol{x}_2)\|_2 = \|\nabla\mathcal{E}(\boldsymbol{x}_1) - \nabla\mathcal{E}(\boldsymbol{x}_2)\|_2 \le \frac{\|\mathcal{A}\|_2^2}{\sigma_y^2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2. \tag{18}$$

Proof. Since

$$\mathcal{E}(\boldsymbol{x}) = \frac{1}{2\sigma_y^2}\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2,$$

we have

$$\nabla\mathcal{E}(\boldsymbol{x}) = \frac{1}{\sigma_y^2}\mathcal{A}^\top(\mathcal{A}\boldsymbol{x} - \boldsymbol{y}), \qquad \nabla^2\mathcal{E}(\boldsymbol{x}) = \frac{1}{\sigma_y^2}\mathcal{A}^\top\mathcal{A}.$$

Hence

$$\|\nabla^2\mathcal{E}(\boldsymbol{x})\|_2 = \frac{1}{\sigma_y^2}\|\mathcal{A}^\top\mathcal{A}\|_2 = \frac{\|\mathcal{A}\|_2^2}{\sigma_y^2},$$

so $\mathcal{E}$ is L-smooth. Also,

$$\nabla\mathcal{E}(\boldsymbol{x}_1) - \nabla\mathcal{E}(\boldsymbol{x}_2) = \frac{1}{\sigma_y^2}\mathcal{A}^\top\mathcal{A}(\boldsymbol{x}_1 - \boldsymbol{x}_2),$$

and taking operator norms yields the Lipschitz bound. □

To formalize the high-noise regime, consider a Tweedie-type estimator

$$\hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma) = \boldsymbol{x} + \sigma^2 s_\theta(\boldsymbol{x}, \sigma), \qquad \boldsymbol{x} = \boldsymbol{x}^{\mathrm{gt}} + \sigma\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}), \tag{19}$$

and let

$$s^*(\boldsymbol{x}, \sigma) = \nabla_{\boldsymbol{x}} \log p_\sigma(\boldsymbol{x})$$

denote the oracle score.

Lemma A.2 (Large-σ amplifies score errors into clean-estimate errors)**.** *For any* $(\boldsymbol{x}, \sigma)$*,*

$$\|\hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma) - \boldsymbol{x}_0^*(\boldsymbol{x}, \sigma)\|_2 = \sigma^2 \|s_\theta(\boldsymbol{x}, \sigma) - s^*(\boldsymbol{x}, \sigma)\|_2 , \tag{20}$$

where

$$\boldsymbol{x}_0^*(\boldsymbol{x}, \sigma) \triangleq \boldsymbol{x} + \sigma^2 s^*(\boldsymbol{x}, \sigma)$$

is the oracle Tweedie estimator. In particular, if

$$\|s_\theta(\boldsymbol{x}, \sigma) - s^*(\boldsymbol{x}, \sigma)\|_2 \le \epsilon(\sigma),$$

then

$$\|\hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma) - \boldsymbol{x}_0^*(\boldsymbol{x}, \sigma)\|_2 \le \sigma^2 \epsilon(\sigma). \tag{21}$$

Proof. By definition,

$$\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_0^* = \big(\boldsymbol{x} + \sigma^2 s_\theta\big) - \big(\boldsymbol{x} + \sigma^2 s^*\big) = \sigma^2 (s_\theta - s^*).$$

Taking norms yields the claim. □

Theorem A.3 (Gradient-mismatch sensitivity upper bound)**.** *Under the linear-Gaussian likelihood above, for any* $(\boldsymbol{x}, \sigma)$*,*

$$\|g(\hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma)) - g(\boldsymbol{x}_0^*(\boldsymbol{x}, \sigma))\|_2 \le \frac{\|\mathcal{A}\|_2^2}{\sigma_y^2} \|\hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma) - \boldsymbol{x}_0^*(\boldsymbol{x}, \sigma)\|_2 . \tag{22}$$

If additionally

$$\|s_\theta(\boldsymbol{x}, \sigma) - s^*(\boldsymbol{x}, \sigma)\|_2 \le \epsilon(\sigma),$$

then

$$\|g(\hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma)) - g(\boldsymbol{x}_0^*(\boldsymbol{x}, \sigma))\|_2 \le \frac{\|\mathcal{A}\|_2^2}{\sigma_y^2} \sigma^2 \epsilon(\sigma). \tag{23}$$

Proof. The first inequality is Lemma A.1 with $\boldsymbol{x}_1 = \hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma)$ and $\boldsymbol{x}_2 = \boldsymbol{x}_0^*(\boldsymbol{x}, \sigma)$. The second follows from Lemma A.2. □

Exact spectral form and conditional amplification. The preceding result is an upper-bound statement. We next give an exact spectral characterization in the linear-Gaussian setting, which clarifies when the σ^2 factor yields actual amplification on measured directions rather than merely a worst-case sensitivity envelope.

Theorem A.4 (Exact gradient-mismatch scaling)**.** *Consider the linear-Gaussian likelihood*

$$\mathcal{E}(\boldsymbol{x}) = \frac{1}{2\sigma_y^2} \|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2, \qquad g(\boldsymbol{x}) = \nabla \mathcal{E}(\boldsymbol{x}) = \frac{1}{\sigma_y^2} \mathcal{A}^\top (\mathcal{A}\boldsymbol{x} - \boldsymbol{y}).$$

Let

$$\hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma) = \boldsymbol{x} + \sigma^2 s_\theta(\boldsymbol{x}, \sigma), \qquad \boldsymbol{x}_0^*(\boldsymbol{x}, \sigma) = \boldsymbol{x} + \sigma^2 s^*(\boldsymbol{x}, \sigma),$$

and define the score error

$$\Delta_s(\boldsymbol{x}, \sigma) = s_\theta(\boldsymbol{x}, \sigma) - s^*(\boldsymbol{x}, \sigma).$$

Then the likelihood-gradient mismatch satisfies the exact identity

$$g(\hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma)) - g(\boldsymbol{x}_0^*(\boldsymbol{x}, \sigma)) = \frac{\sigma^2}{\sigma_y^2} \mathcal{A}^\top \mathcal{A} \Delta_s(\boldsymbol{x}, \sigma). \tag{24}$$

If

$$\mathcal{A} = U \operatorname{diag}(a_1, \ldots, a_r) V^\top$$

is a singular-value decomposition, then

$$\|g(\hat{\boldsymbol{x}}_0(\boldsymbol{x}, \sigma)) - g(\boldsymbol{x}_0^*(\boldsymbol{x}, \sigma))\|_2^2 = \frac{\sigma^4}{\sigma_y^4} \sum_{j=1}^{r} a_j^4 \left|\langle \boldsymbol{v}_j, \Delta_s(\boldsymbol{x}, \sigma)\rangle\right|^2 . \tag{25}$$

Consequently, for any measured subspace

$$\mathcal{S} = \operatorname{span}\{\boldsymbol{v}_j : j \in J\},$$

define the projected mismatch

$$e_{\mathcal{S}}(\boldsymbol{x},\sigma) := \Pi_{\mathcal{S}}\left(g(\hat{\boldsymbol{x}}_0(\boldsymbol{x},\sigma)) - g(\boldsymbol{x}_0^*(\boldsymbol{x},\sigma))\right).$$

Then

$$\|e_{\mathcal{S}}(\boldsymbol{x},\sigma)\|_2^2 = \frac{\sigma^4}{\sigma_y^4}\sum_{j\in J} a_j^4 \left|\langle \boldsymbol{v}_j, \Delta_s(\boldsymbol{x},\sigma)\rangle\right|^2. \tag{26}$$

Thus, the mismatch component on $\mathcal{S}$ scales quadratically with σ, modulated by the score-error energy on the same measured directions. In particular, for $\sigma_h > \sigma_l$, consider two states $(\boldsymbol{x}_h,\sigma_h)$ and $(\boldsymbol{x}_l,\sigma_l)$. If

$$\left(\sum_{j\in J} a_j^4 \left|\langle \boldsymbol{v}_j, \Delta_s(\boldsymbol{x}_h,\sigma_h)\rangle\right|^2\right)^{1/2} \geq \left(\frac{\sigma_l}{\sigma_h}\right)^2 \left(\sum_{j\in J} a_j^4 \left|\langle \boldsymbol{v}_j, \Delta_s(\boldsymbol{x}_l,\sigma_l)\rangle\right|^2\right)^{1/2}, \tag{27}$$

then

$$\|e_{\mathcal{S}}(\boldsymbol{x}_h,\sigma_h)\|_2 \geq \|e_{\mathcal{S}}(\boldsymbol{x}_l,\sigma_l)\|_2. \tag{28}$$

Proof. By the Tweedie form,

$$\hat{\boldsymbol{x}}_0(\boldsymbol{x},\sigma) - \boldsymbol{x}_0^*(\boldsymbol{x},\sigma) = \sigma^2\Delta_s(\boldsymbol{x},\sigma).$$

Since

$$g(\boldsymbol{x}) = \frac{1}{\sigma_y^2}\mathcal{A}^\top(\mathcal{A}\boldsymbol{x} - \boldsymbol{y}),$$

we have

$$g(\hat{\boldsymbol{x}}_0) - g(\boldsymbol{x}_0^*) = \frac{1}{\sigma_y^2}\mathcal{A}^\top\mathcal{A}(\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_0^*) = \frac{\sigma^2}{\sigma_y^2}\mathcal{A}^\top\mathcal{A}\Delta_s.$$

Using

$$\mathcal{A}^\top\mathcal{A} = V\,\mathrm{diag}(a_1^2,\ldots,a_r^2)V^\top$$

gives

$$\left\|\mathcal{A}^\top\mathcal{A}\Delta_s\right\|_2^2 = \sum_{j=1}^{r} a_j^4\left|\langle \boldsymbol{v}_j, \Delta_s\rangle\right|^2,$$

which proves the exact spectral identity. Projecting onto $\mathcal{S} = \mathrm{span}\{\boldsymbol{v}_j : j\in J\}$ restricts the same sum to $j\in J$. The monotonic comparison follows directly from the displayed projected identity and the stated condition. □

Corollary A.5 (Frequency truncation removes high-frequency mismatch)**.** *Assume $\mathcal{A}$ and the spectral mask $\mathcal{P}_\omega$ are simultaneously diagonalizable in the Fourier basis, with transfer function $a(\xi)$. Let*

$$e_{\mathrm{full}}(\sigma) = \frac{\sigma^2}{\sigma_y^2}\mathcal{A}^\top\mathcal{A}\Delta_s(\sigma),$$

and

$$e_{\omega}(\sigma) = \frac{\sigma^2}{\sigma_y^2}\mathcal{A}^\top\mathcal{P}_\omega^\top\mathcal{P}_\omega\mathcal{A}\Delta_s(\sigma).$$

Then

$$\|e_{\mathrm{full}}(\sigma)\|_2^2 - \|e_{\omega}(\sigma)\|_2^2 = \frac{\sigma^4}{\sigma_y^4}\sum_{\xi\notin\Omega_\omega} |a(\xi)|^4|\widehat{\Delta_s}(\xi,\sigma)|^2 \geq 0. \tag{29}$$

The inequality is strict whenever the score error has nonzero energy on frequencies outside the exposed passband and the operator has nonzero transfer there.

Proof. Under the simultaneous diagonalization assumption, $\mathcal{A}$ acts in the Fourier domain by multiplication with $a(\xi)$, and $\mathcal{P}_\omega$ acts by the binary mask $m_\omega(\xi)$. Hence

$$\widehat{e_{\mathrm{full}}}(\xi,\sigma) = \frac{\sigma^2}{\sigma_y^2}|a(\xi)|^2\widehat{\Delta_s}(\xi,\sigma),$$

whereas

$$\widehat{e_\omega}(\xi,\sigma) = \frac{\sigma^2}{\sigma_y^2} m_\omega(\xi)|a(\xi)|^2 \widehat{\Delta_s}(\xi,\sigma).$$

Since $m_\omega(\xi) = 1$ for $\xi \in \Omega_\omega$ and $m_\omega(\xi) = 0$ otherwise, subtracting the squared norms gives the stated identity. □

Theorem A.6 (Inexact likelihood gradients can break descent)**.** *Let $\mathcal{E}$ be L-smooth. Consider an inexact gradient step*

$$\boldsymbol{x}^+ = \boldsymbol{x} - \alpha\left(\nabla\mathcal{E}(\boldsymbol{x}) + \boldsymbol{e}\right), \tag{30}$$

where $\mathbf{e}$ is an additive gradient error and $\alpha > 0$. Then

$$\begin{aligned}\mathcal{E}\left(\boldsymbol{x}^+\right) \leq & \mathcal{E}(\boldsymbol{x}) - \alpha\left(1 - \frac{L\alpha}{2}\right)\|\nabla\mathcal{E}(\boldsymbol{x})\|_2^2 \\ & + \alpha(1+L\alpha)\left|\langle\nabla\mathcal{E}(\boldsymbol{x}), \boldsymbol{e}\rangle\right| + \frac{L\alpha^2}{2}\|\boldsymbol{e}\|_2^2.\end{aligned} \tag{31}$$

In particular, even if $\alpha \leq 1/L$, at any non-stationary point $\nabla\mathcal{E}(\boldsymbol{x}) \neq \mathbf{0}$, there exist errors $\mathbf{e}$ such that $\mathcal{E}(\boldsymbol{x}^+) > \mathcal{E}(\boldsymbol{x})$. Moreover, a sufficient condition for guaranteed decrease is

$$\alpha \leq \frac{1}{L} \qquad \textit{and} \qquad \|\boldsymbol{e}\|_2 \leq (\sqrt{5}-2)\|\nabla\mathcal{E}(\boldsymbol{x})\|_2. \tag{32}$$

Proof. By L-smoothness, for any direction $\boldsymbol{d}$,

$$\mathcal{E}(\boldsymbol{x}+\boldsymbol{d}) \leq \mathcal{E}(\boldsymbol{x}) + \langle\nabla\mathcal{E}(\boldsymbol{x}), \boldsymbol{d}\rangle + \frac{L}{2}\|\boldsymbol{d}\|_2^2.$$

Set

$$\boldsymbol{d} = -\alpha\left(\nabla\mathcal{E}(\boldsymbol{x}) + \boldsymbol{e}\right).$$

Then

$$\begin{aligned}\mathcal{E}(\boldsymbol{x}^+) \leq & \mathcal{E}(\boldsymbol{x}) - \alpha\|\nabla\mathcal{E}(\boldsymbol{x})\|_2^2 - \alpha\left\langle\nabla\mathcal{E}(\boldsymbol{x}), \boldsymbol{e}\right\rangle \\ & + \frac{L\alpha^2}{2}\left\|\nabla\mathcal{E}(\boldsymbol{x}) + \boldsymbol{e}\right\|_2^2.\end{aligned}$$

Expanding the squared norm gives

$$\begin{aligned}\mathcal{E}(\boldsymbol{x}^+) \leq & \mathcal{E}(\boldsymbol{x}) - \alpha\left(1 - \frac{L\alpha}{2}\right)\|\nabla\mathcal{E}(\boldsymbol{x})\|_2^2 \\ & - \alpha(1-L\alpha)\left\langle\nabla\mathcal{E}(\boldsymbol{x}), \boldsymbol{e}\right\rangle + \frac{L\alpha^2}{2}\|\boldsymbol{e}\|_2^2.\end{aligned}$$

Since

$$-\alpha(1-L\alpha)\left\langle\nabla\mathcal{E}(\boldsymbol{x}), \boldsymbol{e}\right\rangle \leq \alpha(1+L\alpha)\left|\langle\nabla\mathcal{E}(\boldsymbol{x}), \boldsymbol{e}\rangle\right|,$$

we obtain the stated bound.

To show that descent can fail, assume $\nabla\mathcal{E}(\boldsymbol{x}) \neq \mathbf{0}$. For any sufficiently small $t > 0$, differentiability implies

$$\mathcal{E}\left(\boldsymbol{x} + t\nabla\mathcal{E}(\boldsymbol{x})\right) > \mathcal{E}(\boldsymbol{x}).$$

Choose

$$\boldsymbol{e} = -\left(1 + \frac{t}{\alpha}\right)\nabla\mathcal{E}(\boldsymbol{x}).$$

Then

$$\boldsymbol{x}^+ = \boldsymbol{x} - \alpha\left(\nabla\mathcal{E}(\boldsymbol{x}) + \boldsymbol{e}\right) = \boldsymbol{x} + t\nabla\mathcal{E}(\boldsymbol{x}),$$

so $\mathcal{E}(\boldsymbol{x}^+) > \mathcal{E}(\boldsymbol{x})$.

For the sufficient decrease condition, assume $\alpha \leq 1/L$. Then

$$1 - \frac{L\alpha}{2} \geq \frac{1}{2}, \qquad 1 + L\alpha \leq 2, \qquad \frac{L\alpha^2}{2} \leq \frac{\alpha}{2}.$$

Using

$$|\langle\nabla\mathcal{E}(\boldsymbol{x}), \boldsymbol{e}\rangle| \leq \|\nabla\mathcal{E}(\boldsymbol{x})\|_2\|\boldsymbol{e}\|_2,$$

we obtain

$$\mathcal{E}(\boldsymbol{x}^+) - \mathcal{E}(\boldsymbol{x}) \le \alpha \|\nabla \mathcal{E}(\boldsymbol{x})\|_2^2 \left(-\frac{1}{2} + 2\rho + \frac{\rho^2}{2}\right), \qquad \rho := \frac{\|\boldsymbol{e}\|_2}{\|\nabla \mathcal{E}(\boldsymbol{x})\|_2}.$$

The quadratic

$$-\frac{1}{2} + 2\rho + \frac{\rho^2}{2} \le 0$$

holds for $\rho \in [0, \sqrt{5} - 2]$. Therefore

$$\|\boldsymbol{e}\|_2 \le (\sqrt{5} - 2)\|\nabla \mathcal{E}(\boldsymbol{x})\|_2$$

guarantees

$$\mathcal{E}(\boldsymbol{x}^+) \le \mathcal{E}(\boldsymbol{x}).$$

□

Corollary A.7 (Band-limited likelihood reduces curvature and sensitivity)**.** *Define a filtered measurement energy*

$$\mathcal{E}_\omega(\boldsymbol{x}) \triangleq \frac{1}{2\sigma_y^2} \|\mathcal{P}_\omega(\mathcal{A}\boldsymbol{x} - \boldsymbol{y})\|_2^2, \tag{33}$$

where $\mathcal{P}_\omega$ is a linear operator with $\|\mathcal{P}_\omega\|_2 \le 1$, such as the ideal orthogonal low-pass projector associated with Eq. 6. Then $\mathcal{E}_\omega$ is L_ω-smooth with

$$L_\omega = \frac{1}{\sigma_y^2} \left\|\mathcal{A}^\top \mathcal{P}_\omega^\top \mathcal{P}_\omega \mathcal{A}\right\|_2 \le \frac{\|\mathcal{A}\|_2^2}{\sigma_y^2} \|\mathcal{P}_\omega\|_2^2 \le \frac{\|\mathcal{A}\|_2^2}{\sigma_y^2} = L. \tag{34}$$

Consequently, the gradient Lipschitz constant and the smoothness-driven step-size restriction are reduced under the band-limited likelihood. This helps explain why full-band enforcement at high noise levels can be more schedule- and operator-sensitive.

Corollary A.8 (Mixed frequency exposure directly attenuates likelihood-gradient error)**.** *Let the frequency-exposure objective at outer step i be*

$$\mathcal{E}_i(\boldsymbol{x}) = \frac{1}{2\sigma_y^2} \left[(1 - \lambda_i)\|\mathcal{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda_i \|\mathcal{P}_{\omega_i}(\mathcal{A}\boldsymbol{x} - \boldsymbol{y})\|_2^2\right], \tag{35}$$

where $\mathcal{P}_{\omega_i}$ is the orthogonal frequency projector onto the exposed passband and $\lambda_i \in [0, 1]$. Define

$$Q_i := (1 - \lambda_i)\boldsymbol{I} + \lambda_i \mathcal{P}_{\omega_i}^\top \mathcal{P}_{\omega_i}. \tag{36}$$

Assume that the clean-estimate error induced by the score approximation satisfies

$$\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_0^* = \sigma^2 \Delta_s(\boldsymbol{x}, \sigma), \tag{37}$$

where $\boldsymbol{x}_0^*$ denotes the oracle clean estimate and $\Delta_s(\boldsymbol{x}, \sigma)$ denotes the score-estimation error. Then the likelihood-gradient mismatch satisfies

$$\boldsymbol{e}_i(\sigma) := \nabla \mathcal{E}_i(\hat{\boldsymbol{x}}_0) - \nabla \mathcal{E}_i(\boldsymbol{x}_0^*) = \frac{\sigma^2}{\sigma_y^2} \mathcal{A}^\top Q_i \mathcal{A} \Delta_s(\boldsymbol{x}, \sigma). \tag{38}$$

Moreover, if $\mathcal{A}$ and $\mathcal{P}_{\omega_i}$ are simultaneously diagonalized in the Fourier basis, with transfer function $a(\xi)$ and binary mask $m_{\omega_i}(\xi) \in \{0, 1\}$, then

$$\|\boldsymbol{e}_i(\sigma)\|_2^2 = \frac{\sigma^4}{\sigma_y^4} \sum_\xi q_i(\xi)^2 |a(\xi)|^4 |\widehat{\Delta_s}(\xi, \sigma)|^2, \tag{39}$$

where

$$q_i(\xi) = (1 - \lambda_i) + \lambda_i m_{\omega_i}(\xi). \tag{40}$$

Consequently, compared with the full-band likelihood-gradient mismatch

$$\|\boldsymbol{e}_{\text{full}}(\sigma)\|_2^2 = \frac{\sigma^4}{\sigma_y^4} \sum_\xi |a(\xi)|^4 |\widehat{\Delta_s}(\xi, \sigma)|^2, \tag{41}$$

the mixed frequency-exposure objective satisfies

$$\|\boldsymbol{e}_{\mathrm{full}}(\sigma)\|_2^2 - \|\boldsymbol{e}_i(\sigma)\|_2^2 = \frac{\sigma^4}{\sigma_y^4} \sum_{\xi} \left(1 - q_i(\xi)^2\right) |a(\xi)|^4 |\widehat{\Delta_s}(\xi, \sigma)|^2 \geq 0. \tag{42}$$

For an ideal binary low-pass mask, this reduction can be written as

$$\|\boldsymbol{e}_{\mathrm{full}}(\sigma)\|_2^2 - \|\boldsymbol{e}_i(\sigma)\|_2^2 = \frac{\sigma^4}{\sigma_y^4} \sum_{\xi \notin \Omega_{\omega_i}} \left[1 - (1 - \lambda_i)^2\right] |a(\xi)|^4 |\widehat{\Delta_s}(\xi, \sigma)|^2 \geq 0. \tag{43}$$

Thus, frequency exposure directly attenuates the likelihood-gradient error by suppressing the contribution of unexposed frequency bands. This effect is distinct from, and complementary to, the reduction of the effective curvature and the resulting relaxation of smoothness-driven step-size constraints.

Proof. The gradient of $\mathcal{E}_i$ is

$$\nabla \mathcal{E}_i(\boldsymbol{x}) = \frac{1}{\sigma_y^2} \mathcal{A}^\top \left[(1 - \lambda_i)\boldsymbol{I} + \lambda_i \mathcal{P}_{\omega_i}^\top \mathcal{P}_{\omega_i}\right] (\mathcal{A}\boldsymbol{x} - \boldsymbol{y}) = \frac{1}{\sigma_y^2} \mathcal{A}^\top Q_i (\mathcal{A}\boldsymbol{x} - \boldsymbol{y}). \tag{44}$$

Therefore,

$$\boldsymbol{e}_i(\sigma) = \nabla \mathcal{E}_i(\hat{\boldsymbol{x}}_0) - \nabla \mathcal{E}_i(\boldsymbol{x}_0^*) \tag{45}$$

$$= \frac{1}{\sigma_y^2} \mathcal{A}^\top Q_i \mathcal{A} (\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_0^*). \tag{46}$$

Using

$$\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_0^* = \sigma^2 \Delta_s(\boldsymbol{x}, \sigma)$$

gives

$$\boldsymbol{e}_i(\sigma) = \frac{\sigma^2}{\sigma_y^2} \mathcal{A}^\top Q_i \mathcal{A} \Delta_s(\boldsymbol{x}, \sigma). \tag{47}$$

When $\mathcal{A}$ and $\mathcal{P}_{\omega_i}$ are diagonalized in the Fourier basis, $\mathcal{A}$ acts by multiplication with $a(\xi)$, and Q_i acts by multiplication with

$$q_i(\xi) = (1 - \lambda_i) + \lambda_i m_{\omega_i}(\xi).$$

Hence the Fourier coefficient of $\boldsymbol{e}_i(\sigma)$ at frequency ξ is

$$\widehat{\boldsymbol{e}}_i(\xi, \sigma) = \frac{\sigma^2}{\sigma_y^2} q_i(\xi) |a(\xi)|^2 \widehat{\Delta_s}(\xi, \sigma),$$

which implies

$$\|\boldsymbol{e}_i(\sigma)\|_2^2 = \frac{\sigma^4}{\sigma_y^4} \sum_{\xi} q_i(\xi)^2 |a(\xi)|^4 |\widehat{\Delta_s}(\xi, \sigma)|^2.$$

The full-band case corresponds to $q_i(\xi) = 1$ for all ξ. Subtracting the mixed-objective error from the full-band error yields

$$\|\boldsymbol{e}_{\mathrm{full}}(\sigma)\|_2^2 - \|\boldsymbol{e}_i(\sigma)\|_2^2 = \frac{\sigma^4}{\sigma_y^4} \sum_{\xi} \left(1 - q_i(\xi)^2\right) |a(\xi)|^4 |\widehat{\Delta_s}(\xi, \sigma)|^2.$$

Since $\lambda_i \in [0, 1]$ and $m_{\omega_i}(\xi) \in \{0, 1\}$, we have $q_i(\xi) \in [0, 1]$, and therefore $1 - q_i(\xi)^2 \geq 0$.

For an ideal binary low-pass mask, $m_{\omega_i}(\xi) = 1$ for $\xi \in \Omega_{\omega_i}$ and $m_{\omega_i}(\xi) = 0$ for $\xi \notin \Omega_{\omega_i}$. Thus $q_i(\xi) = 1$ inside the exposed passband and $q_i(\xi) = 1 - \lambda_i$ outside it. The reduction therefore vanishes inside Ω_{ω_i} and equals $1 - (1 - \lambda_i)^2$ outside the passband. □

Lemma A.9 (Non-expansiveness of spectral masking)**.** *Let P_ω^F be defined in Eq. 6 with the unitary normalization of $\mathcal{F}$. Then for any residual $\boldsymbol{r}$,*

$$\|P_\omega^F(\boldsymbol{r})\|_2 \leq \|\boldsymbol{r}\|_2. \tag{48}$$

Proof. The frequency shift is a permutation and hence norm-preserving. With the chosen normalization, $\mathcal{F}$ is unitary, and Parseval's identity gives

$$\|\mathcal{F}(\boldsymbol{r})\|_2 = \|\boldsymbol{r}\|_2.$$

Multiplication by M_ω is an orthogonal projection in the Fourier domain and is therefore non-expansive. □

Theorem A.10 (Low-pass-only likelihood induces non-identifiability)**.** *Fix a cutoff radius ω and define the low-pass loss*

$$\mathcal{L}_{\text{freq}}(\boldsymbol{x}) = \left\|P_\omega^F(\mathcal{A}\boldsymbol{x} - \boldsymbol{y})\right\|_2^2. \tag{49}$$

If there exists $\boldsymbol{d} \neq \mathbf{0}$ such that

$$P_\omega^F(\mathcal{A}\boldsymbol{d}) = \mathbf{0},$$

then

$$\mathcal{L}_{\text{freq}}(\boldsymbol{x} + \boldsymbol{d}) = \mathcal{L}_{\text{freq}}(\boldsymbol{x}) \qquad \textit{for all } \boldsymbol{x}. \tag{50}$$

Hence the low-pass data term alone cannot distinguish $\boldsymbol{x}$ along the subspace $\ker(P_\omega^F\mathcal{A})$.

Proof. By linearity of P_ω^F,

$$\mathcal{L}_{\text{freq}}(\boldsymbol{x} + \boldsymbol{d}) = \left\|P_\omega^F(\mathcal{A}\boldsymbol{x} - \boldsymbol{y}) + P_\omega^F(\mathcal{A}\boldsymbol{d})\right\|_2^2.$$

If $P_\omega^F(\mathcal{A}\boldsymbol{d}) = \mathbf{0}$, the value is unchanged. □

Lemma A.11 (Low-pass data term is insensitive to Haar details under idealization)**.** *Let W denote an orthonormal Haar transform and let*

$$\boldsymbol{z} = W\boldsymbol{x}$$

be its coefficients, with

$$\boldsymbol{z} = (\boldsymbol{z}_{\text{c}}, \boldsymbol{z}_{\text{d}})$$

denoting coarse and detail blocks. Let $\mathcal{P}_{\omega_i}$ denote the ideal low-pass orthogonal projector induced by the spectral mask in Sec. 3.2 with cutoff radius ω_i. Assume that the composition

$$\mathcal{P}_{\omega_i} \circ \mathcal{A} \circ W^\top$$

annihilates Haar-detail components in the idealized early regime, i.e.,

$$\mathcal{P}_{\omega_i}\mathcal{A}W^\top(\mathbf{0}, \boldsymbol{z}_{\text{d}}) = \mathbf{0} \qquad \textit{for all } \boldsymbol{z}_{\text{d}}. \tag{51}$$

Then the low-pass residual depends only on $\boldsymbol{z}_{\text{c}}$, and

$$\nabla_{\boldsymbol{z}_{\text{d}}} \left\|\mathcal{P}_{\omega_i}\left(\mathcal{A}(W^\top\boldsymbol{z}) - \boldsymbol{y}\right)\right\|_2^2 = \mathbf{0}. \tag{52}$$

When the annihilation condition holds approximately, the gradient along the Haar-detail coordinates is correspondingly small.

Proof. By linearity,

$$W^\top(\boldsymbol{z}_{\text{c}}, \boldsymbol{z}_{\text{d}}) = W^\top(\boldsymbol{z}_{\text{c}}, \mathbf{0}) + W^\top(\mathbf{0}, \boldsymbol{z}_{\text{d}}).$$

The annihilation assumption gives

$$\mathcal{P}_{\omega_i}\mathcal{A}W^\top(\mathbf{0}, \boldsymbol{z}_{\text{d}}) = \mathbf{0}.$$

Therefore,

$$\mathcal{P}_{\omega_i}\left(\mathcal{A}W^\top(\boldsymbol{z}_{\text{c}}, \boldsymbol{z}_{\text{d}}) - \boldsymbol{y}\right) = \mathcal{P}_{\omega_i}\left(\mathcal{A}W^\top(\boldsymbol{z}_{\text{c}}, \mathbf{0}) - \boldsymbol{y}\right),$$

which is independent of $\boldsymbol{z}_{\text{d}}$. Hence the squared low-pass residual is constant along the Haar-detail coordinates, and its gradient with respect to $\boldsymbol{z}_{\text{d}}$ is zero. □

Corollary A.12 (Detail coefficients remain prior-dominated in the idealized low-pass regime)**.** *This corollary analyzes the low-pass component of the intermediate posterior and is intended to explain why detail directions are weakly constrained early in the trajectory. Consider the idealized low-pass component of the intermediate target at outer step i,*

$$\pi_i(\boldsymbol{x}_0) \propto \exp\left(-\frac{1}{2\beta_i^2}\left\|\mathcal{P}_{\omega_i}\left(\mathcal{A}(\boldsymbol{x}_0)-\boldsymbol{y}\right)\right\|_2^2\right)\cdot\mathcal{N}\left(\boldsymbol{x}_0;\hat{\boldsymbol{x}}_{0,i},\sigma_i^2\boldsymbol{I}\right). \tag{53}$$

Let

$$\boldsymbol{z}=W\boldsymbol{x}_0, \qquad \hat{\boldsymbol{z}}_i=W\hat{\boldsymbol{x}}_{0,i}.$$

Under the conditions of Lemma A.11, the conditional distribution of detail coefficients is approximately

$$\boldsymbol{z}_{\mathrm{d}} \mid (\boldsymbol{y},\boldsymbol{z}_{\mathrm{c}}) \approx \mathcal{N}\left(\hat{\boldsymbol{z}}_{\mathrm{d},i},\sigma_i^2\boldsymbol{I}\right). \tag{54}$$

In particular, when σ_i is large, the posterior variance of $\boldsymbol{z}_{\mathrm{d}}$ remains large, so Langevin refinement can introduce high-variance and operator-dependent spurious details in the detail bands.

Proof. Since W is orthonormal, the anchor Gaussian remains isotropic in Haar coordinates:

$$\mathcal{N}\left(\boldsymbol{x}_0;\hat{\boldsymbol{x}}_{0,i},\sigma_i^2\boldsymbol{I}\right)=\mathcal{N}\left(\boldsymbol{z};\hat{\boldsymbol{z}}_i,\sigma_i^2\boldsymbol{I}\right).$$

Lemma A.11 implies that the low-pass term depends only on $\boldsymbol{z}_{\mathrm{c}}$ and is constant with respect to $\boldsymbol{z}_{\mathrm{d}}$ in the idealized low-pass regime. Therefore the joint density factorizes into a term depending on $\boldsymbol{z}_{\mathrm{c}}$ and an independent Gaussian term in $\boldsymbol{z}_{\mathrm{d}}$. This gives

$$\boldsymbol{z}_{\mathrm{d}} \mid (\boldsymbol{y},\boldsymbol{z}_{\mathrm{c}}) \approx \mathcal{N}\left(\hat{\boldsymbol{z}}_{\mathrm{d},i},\sigma_i^2\boldsymbol{I}\right).$$

□