DAPS++: Rethinking Diffusion Inverse Problems with Decoupled Posterior Annealing

Hao Chen, Renzheng Zhang, and Scott S. Howard*

University of Notre Dame, Notre Dame, IN 46556, USA
{hchen27, rzhang4, showard}@nd.edu

Abstract. From a Bayesian perspective, score-based diffusion solves inverse problems through joint inference, embedding the likelihood with the prior to guide the sampling process. However, this formulation fails to explain its practical behavior: the prior offers limited guidance, while reconstruction is largely driven by the measurement-consistency term, leading to an inference process that is effectively decoupled from the diffusion dynamics. We show that the diffusion prior in these solvers functions primarily as a warm initializer that places estimates near the data manifold, while reconstruction is driven almost entirely by measurement consistency. Based on this observation, we introduce **DAPS++**, which fully decouples diffusion-based initialization from likelihood-driven refinement, allowing the likelihood term to guide inference more directly while maintaining numerical stability and providing insight into why unified diffusion trajectories remain effective in practice. By requiring fewer function evaluations (NFEs) and measurement-optimization steps, **DAPS++** achieves high computational efficiency and robust reconstruction performance across diverse image restoration tasks.

1 Introduction

Solving ill-posed inverse problems is fundamental in science and engineering, particularly in scientific imaging—such as astrophotography [23,25] and biomedical imaging [8,12,16,24,31]—with broad applications in real-world systems. These problems aim to reconstruct the original signal $\mathbf{x}_0$ from observations $\mathbf{y}$ acquired under various restoration or measurement conditions, assuming a known forward model [34]. A common strategy is to formulate the inverse problem within a Bayesian framework [6], where $p(\mathbf{x}_0|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}_0)p(\mathbf{x}_0)$. Here, the likelihood $p(\mathbf{y}|\mathbf{x}_0)$ characterizes the measurement process, while the prior $p(\mathbf{x}_0)$ encodes structural assumptions about the underlying signal.

Classical Bayesian inverse problem formulations often rely on Tikhonov, total variation (TV), or wavelet-based priors to constrain the solution space and stabilize reconstruction [13, 33]. Modern generative models [15, 26, 28], particularly score-based diffusion methods [17, 29, 30, 32], provide far more expressive high-dimensional priors and have become widely adopted in Bayesian inverse

* Corresponding author.

problem frameworks [1, 3, 11, 41]. These methods incorporate measurement information into the sampling dynamics by injecting an approximate likelihood gradient $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$ at each diffusion timestep. By solving stochastic differential equations (SDEs) or their deterministic ODE counterparts [21, 22], diffusion models transform an initial noise sample $\mathbf{x}_T$ toward the data manifold $\mathbf{x}_0$, conditioned on the observation $\mathbf{y}$.

More recently, decoupled formulations such as DCDP [20] and DAPS [41] explicitly separate the diffusion prior $p(\mathbf{x}_0)$ from the data-consistency term $p(\mathbf{y}|\mathbf{x}_0)$ at the algorithmic level. In these methods, diffusion updates and measurement updates alternate: each data-consistency step moves the estimate toward the observation, and the subsequent re-noising step restores the appropriate annealed noise level. As the noise level decreases, the resulting time-marginal distribution $\pi(\mathbf{x}_t|\mathbf{y})$ is assumed to approach the posterior $p(\mathbf{x}_0|\mathbf{y})$. Although implemented in a decoupled manner, the overall framework remains nominally Bayesian, since each marginal still depends on the prior gradient $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$.

However, this expected trajectory rarely holds in practice. The gradient of data likelihood gradient often dominates the update dynamics, causing each intermediate distribution $\pi(\mathbf{x}_t|\mathbf{y})$ to act as an isolated approximation of the posterior rather than part of a coherent sequence of marginals. As a result, the path from $\mathbf{x}_T$ to $\mathbf{x}_0$ becomes a set of loosely connected estimates rather than a consistent posterior evolution. This reveals that a strictly Bayesian interpretation is insufficient for explaining the empirical behavior of diffusion-based inverse problem solvers. This reveals that the diffusion prior and the data-consistency update serve fundamentally different roles: the diffusion stage provides a structured initialization that constrains the feasible solution space to lie near the data manifold, while the likelihood gradient refines this initialization to satisfy the measurement. The two stages are thus naturally decoupled, corresponding to prior-driven initialization followed by likelihood-driven correction.

Building on this perspective, we introduce DAPS++. Leveraging the principles of Decoupled Annealing Posterior Sampling (DAPS), our method provides a more efficient and interpretable formulation of diffusion-based Bayesian inference using diffusion models that operate directly in pixel space. Empirically, DAPS++ delivers strong performance across a broad range of inverse problems, including both linear and nonlinear image restoration, while offering substantially faster sampling than existing approaches. Compared with DAPS and other diffusion-based inverse problem methods, DAPS++ achieves comparable or superior reconstruction quality while reducing sampling time and neural function evaluations (NFEs) by approximately 90%on both the FFHQ validation set and the ImageNet test dataset.

2 Related Work

2.1 Diffusion Models

Score-based diffusion models [15, 17, 29, 30, 32] learn the data distribution $p(\mathbf{x}_0)$ and its gradient, generating samples by progressively denoising Gaussian cor-

rupted data. A sequence of isotropic Gaussian perturbations $\mathcal{N}(0, \sigma_t^2 \mathbf{I})$ is applied to the data, inducing a family of noise-conditioned distributions $p(\mathbf{x}; \sigma_t)$. The model is trained to estimate the gradient of the log-density as a *score function* across time steps $t \in [0, T]$, or equivalently across decreasing noise levels σ_t from $\sigma_T = \sigma\text{max}$ to $\sigma_0 = 0$. When $\sigma_{\max}$ is sufficiently large, the terminal distribution approaches pure Gaussian noise, $\mathcal{N}(0, \sigma_{\max}^2 \mathbf{I})$.

During sampling, the learned model approximates the true score of $p_t(\mathbf{x}; \sigma_t)$ via $s_\theta(\mathbf{x}, \sigma_t) \approx \nabla_{\mathbf{x}} \log p_t(\mathbf{x}; \sigma_t)$. Starting from pure noise $\mathbf{x}_T \sim \mathcal{N}(0, \sigma_{\max}^2 \mathbf{I})$, the generative process follows the reverse stochastic differential equation (SDE) [21, 29]

$$d\mathbf{x}_t = -2\, \dot{\sigma}_t\, \sigma_t\, \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t; \sigma_t)\, dt + \sqrt{2\, \dot{\sigma}_t\, \sigma_t}\, d\mathbf{w}_t, \tag{1}$$

where $\dot{\sigma}_t$ is the derivative of the noise schedule, $d\mathbf{w}_t$ is a standard Wiener process, and the formulation follows the variance-exploding (VE) or EDM parameterization.

2.2 Inverse Problems with Diffusion Models

A general inverse problem seeks to recover the unknown signal $\mathbf{x}_0$ from noisy ill-posed measurements $\mathbf{y}$. Given a forward operator $\mathcal{A}$, the general measurement model is $\mathbf{y} = \mathcal{A}(\mathbf{x}_0) + \mathbf{n}$, where $\mathbf{n}$ is additive Gaussian noise $\mathcal{N}(0, \gamma^2 \mathbf{I})$. Under a Bayesian formulation, the posterior satisfies $p(\mathbf{x}_0|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}_0) p(\mathbf{x}_0)$, with the likelihood term $p(\mathbf{y}|\mathbf{x}_0) = \mathcal{N}(\mathcal{A}(\mathbf{x}_0), \gamma^2 \mathbf{I})$.

To incorporate the diffusion prior into inverse problems, a standard approach is to form a joint SDE by adding the data-consistency term to Eq. (1), giving

$$\begin{aligned} d\mathbf{x}_t = &-2\, \dot{\sigma}_t\, \sigma_t\, \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t; \sigma_t)\, dt \\ &- 2\, \dot{\sigma}_t\, \sigma_t\, \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t; \sigma_t)\, dt \\ &+ \sqrt{2\, \dot{\sigma}_t\, \sigma_t}\, d\mathbf{w}_t \end{aligned} \tag{2}$$

which combines the diffusion prior drift with a measurement-consistency term.

In practice, measurement consistency is defined in terms of the clean likelihood $p(\mathbf{y}|\mathbf{x}_0)$; as a result, the noised likelihood $p(\mathbf{y}|\mathbf{x}_t; \sigma_t)$ is typically inaccessible or ill-defined. Several approaches approximate this quantity [2, 27] to embed measurement guidance directly into the diffusion process. Among these, Diffusion Posterior Sampling (DPS) [3] employs Tweedie's formula to estimate $p(\mathbf{y}|\mathbf{x}_t) \approx p(\mathbf{y}|\mathbf{x}_0 = \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t])$, achieving strong empirical performance. However, recent studies show that DPS behaves more like a maximum a posteriori (MAP) estimator than a true posterior sampler [38, 40]. This has prompted data-consistency frameworks with Plug-and-Play (PnP) priors [20, 37, 39] that incorporate explicit likelihood optimization into diffusion-based solvers. These PnP-based methods use diffusion models as learned denoising priors within an explicit data-consistency loop, rather than depending on measurement-guided updates along the diffusion manifold. Nevertheless, approaches that minimize $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{x}_t; \sigma_t)$ often exhibit instability due to overfitting. Another line of work uses SVD-based projections and projection-driven sampling [4, 19, 35] to

enforce measurement constraints, while recent decoupled frameworks such as DAPS [41] reformulate Eq. (2) into a two-stage structure that preserves the time-marginal distribution.

However, similar to DPS, such decoupled methods deviate from posterior sampling as defined in Eq. (2), because they implicitly assume that the time-marginal distributions remain consistent with the true posterior given $\hat{\mathbf{x}}_0(\mathbf{x}_t)$, which does not hold in practice. In reality, as we show below from Sec. 3.4, the two stages operate as separate generation and optimization procedures rather than components of a unified posterior sampler, limiting both the theoretical validity and computational efficiency of the framework. This observation motivates reinterpreting diffusion-based inverse problem solvers primarily as methods for optimizing $p(\mathbf{y}|\mathbf{x}_0)$ within a diffusion-generated solution space, offering a more accurate perspective on diffusion-based inference.

3 Method

3.1 The Diffusion Prior as a Warm Initializer

We begin by analyzing the interaction between the prior and likelihood gradients to clarify the functional role of the diffusion prior in inverse problem solving.

To quantify the directional relationship between the two terms, we examine the empirical inner product

$$A_t = \Big\langle \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t; \sigma_t),\, \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t; \sigma_t) \Big\rangle, \tag{3}$$

where the two factors denote the likelihood and prior score at noise level σ_t, respectively. Under high-noise conditions, this inner product is empirically close to zero, indicating that the two gradients are effectively orthogonal. This orthogonality helps explain why methods such as DPS can combine both terms without destructive interference (see Sec. 3.4).

To assess the relative strength of the two components, we define the gradient ratio

$$\kappa_t := \frac{\|\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t; \sigma_t)\|}{\|\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t; \sigma_t)\|}. \tag{4}$$

Under the Gaussian noise model $p(\mathbf{y}|\mathbf{x}_0) = \mathcal{N}(\mathcal{A}(\mathbf{x}_0), \gamma^2\mathbf{I})$, the likelihood gradient at noise level σ_t can be approximated as

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{x}_t; \sigma_t) \simeq -\frac{1}{\gamma^2} \nabla_{\mathbf{x}_t} \left\|\mathbf{y} - \mathcal{A}\big(\hat{\mathbf{x}}_0(\mathbf{x}_t)\big)\right\|^2, \tag{5}$$

where $\mathcal{A}(\cdot)$ is the measurement operator and $\hat{\mathbf{x}}_0(\mathbf{x}_t)$ denotes the denoised estimate associated with state $\mathbf{x}_t$.

As shown in Fig. 1(a), $\kappa_t \gg 1$ throughout the diffusion process. A Lipschitz-bound analysis in the Appendix (Sec. 6.1) formalizes this: under standard regularity assumptions, $\kappa_t \gtrsim \frac{\sigma^+_{\min}(\mathcal{A})}{\gamma^2 C}\sigma_t|\mathbf{r}_t|$, which grows with the noise-to-measurement

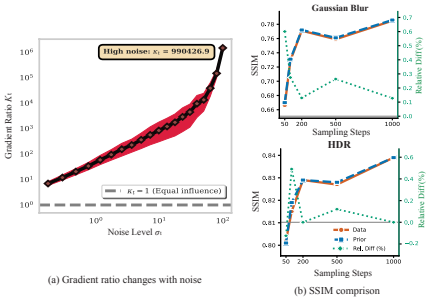

Fig. 1: (a) Evolution of gradient ratio κ_t vs. noise level during a Gaussian-blur iteration, showing data-consistency gradient dominates throughout. (b) Reconstruction quality of DAPS when driven by the **prior** (time-marginal) vs. the **data likelihood** term only across annealing steps, showing that data likelihood alone accounts for nearly all reconstruction quality.

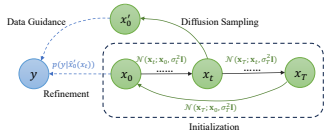

Fig. 2: Diagram of the **DAPS++** framework. **Stage 1: Initialization** constructs the constrained optimization space $p(\mathbf{x}_0)$. **Stage 2: Refinement** optimizes under measurement guidance. The two stages are fully decoupled, with noise re-injected between steps.

ratio σ_t/γ. This does not imply that the diffusion prior is unnecessary; instead, it is essential for placing the initial estimate near the data manifold. The analysis shows that once initialization is achieved, the prior gradient contributes negligibly to subsequent updates: it is small in magnitude ($\kappa_t \gg 1$) and approximately orthogonal to the likelihood gradient ($A_t \approx 0$), meaning it neither accelerates nor redirects the measurement-driven correction. As confirmed in Fig. 1(b), removing the prior term from the time-marginal update and relying solely on the data-consistency term produces nearly identical reconstruction quality. This demonstrates that the continuous prior guidance used in existing methods can be effectively replaced by a one-time diffusion-based initialization.

3.2 Complete Stage Separation: Initialization and Refinement

The observation above suggests that decoupled noise-annealing approaches do not strictly follow the time-marginal distribution, especially when initialized from high-noise conditions. Nevertheless, they produce strong reconstructions empirically because the diffusion model provides a high-quality initialization that the likelihood gradient can then refine. This motivates a complete separation between the two stages, as illustrated in Fig. 2.

Stage 1: Diffusion Initialization. We start from $\mathbf{x}_T \sim \mathcal{N}(0, \sigma_{\max}^2 \mathbf{I})$ and run the reverse diffusion process without likelihood guidance. A key design question is what form the initialization should take. In the EDM noise schedule, the probability flow ODE is

$$d\mathbf{x} = -\dot{\sigma}\sigma \nabla_{\mathbf{x}} \log p(\mathbf{x}; \sigma)\, dt, \tag{6}$$

and a single Euler step from $\mathbf{x}_t$ recovers exactly Tweedie's formula:

$$\mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t] \;\approx\; \mathbf{x}_t \;+\; \sigma_t^2\, s_\theta(\mathbf{x}_t, \sigma_t), \tag{7}$$

where $s_\theta(\mathbf{x}_t, \sigma_t)$ denotes the learned score function. As a first-order mean estimator, Tweedie's formula produces an initialization near the data manifold that is smooth and fast to compute. Higher-order ODE solvers follow the diffusion trajectory more faithfully but do not improve Stage 2 convergence: what matters for refinement is proximity to the manifold, not precision of the sampling path, since the likelihood gradient handles the remaining correction. Figure S1 in the Supplementary Material validates this by comparing initialization strategies across different sampling method settings, confirming that a single Tweedie step at an appropriate σ provides a more effective starting point than ODE-based alternatives.

To balance efficiency and fidelity, DAPS++ introduces a noise threshold $\bar{\sigma}$: for $\sigma_t > \bar{\sigma}$, Tweedie denoising (Eq. (7)) provides fast initialization; for $\sigma_t \leq \bar{\sigma}$, a single higher-order ODE step (RK4) preserves fine details at minimal additional cost. The resulting $\hat{\mathbf{x}}_0$ captures the global structure of the signal and lies near the high-density region of $p(\mathbf{x}_0)$. Stage 2 requires only that $\hat{\mathbf{x}}_0$ is close to the data manifold; the specific initialization path is irrelevant, ensuring complete decoupling between the two stages.

Stage 2: Likelihood-Driven Refinement. Given the initialization $\hat{\mathbf{x}}_0$, we perform MCMC sampling driven solely by the measurement likelihood. Under the Gaussian measurement model, the likelihood gradient is globally Lipschitz and the log-density is strongly concave, giving the unadjusted Langevin algorithm (ULA) well-established convergence guarantees that depend on the specific type of regularization [5, 10]. Following ULA, the update rule is

$$\begin{aligned}\mathbf{x}_0^{(j+1)} = \mathbf{x}_0^{(j)} + \eta\Big(&\nabla_{\mathbf{x}_0^{(j)}} \log p(\mathbf{y} \mid \mathbf{x}_0^{(j)}) \\ &+ \nabla_{\mathbf{x}_0^{(j)}} \log p(\mathbf{x}_0^{(j)})\Big) + \sqrt{2\eta}\,\boldsymbol{\epsilon}_j,\end{aligned} \tag{8}$$

where η is the step size and $\boldsymbol{\epsilon}_j \sim \mathcal{N}(0, \mathbf{I})$. In standard MCMC, the prior gradient $\nabla_{\mathbf{x}_0} \log p(\mathbf{x}_0)$ serves as a regularization term. However, because the diffusion initialization places $\hat{\mathbf{x}}_0$ in a region where $\|\nabla \log p(\mathbf{x}_0)\| = O(\varepsilon)$, this contribution is numerically negligible. Once we abandon the time-marginal requirement entirely, convergence is governed purely by the geometry of the likelihood, and far fewer steps suffice. We therefore omit the prior gradient and obtain a likelihood-only refinement:

$$\mathbf{x}_0^{(j+1)} = \mathbf{x}_0^{(j)} - \frac{\eta}{\gamma^2}\, \nabla_{\mathbf{x}_0^{(j)}} \left\| \mathbf{y} - \mathcal{A}\Big(\mathbf{x}_0^{(j)}\Big) \right\|^2 + \sqrt{2\eta}\,\boldsymbol{\epsilon}_j. \tag{9}$$

This refinement drives the estimate toward the measurement-consistent solution but reduces variability in noise-corrupted or unobserved components. To restore this variability, each refinement cycle is followed by a re-noising step. This controlled stochasticity prevents overfitting and allows the next cycle to explore

Algorithm 1 DAPS++ Inference with Decoupled Initialization-Refinement

Require: Measurement $\mathbf{y}$, operator $\mathcal{A}$, effective noise variance γ^2; diffusion score s_θ, noise schedule $\{\sigma_t\}$, threshold $\bar{\sigma}$, refinement steps J, step sizes $\{\eta_j\}$, total cycles K.
Ensure: Final reconstruction $\hat{\mathbf{x}}_0$.

Initialize $x_{\text{in}}^{(0)} \sim \mathcal{N}(\mathbf{0}, \sigma_{\max}^2 \mathbf{I})$
for $k = 1$ **to** K **do** ▷ Stage 1: Diffusion initialization
  if $\sigma_{\text{in}} > \bar{\sigma}$ **then**
    $\hat{\mathbf{x}}_0^{(k)} \leftarrow x_{\text{in}}^{(k-1)} + \sigma_{\text{in}}^2 s_\theta\big(x_{\text{in}}^{(k-1)}, \sigma_{\text{in}}\big)$
  else
    $\hat{\mathbf{x}}_0^{(k)} \leftarrow \texttt{ODESolver}\big(x_{\text{in}}^{(k-1)}, \sigma_{\text{in}}\big)$
  end if ▷ Stage 2: Likelihood-driven refinement
  $z^{(0)} \leftarrow \hat{\mathbf{x}}_0^{(k)}$
  for $j = 1$ **to** J **do**
    $\mathbf{r}^{(j-1)} \leftarrow \mathbf{y} - \mathcal{A}(z^{(j-1)})$
    $\nabla_z \mathcal{L} \leftarrow \frac{1}{\gamma^2} \texttt{VJP}_{\mathcal{A}}\big(z^{(j-1)}, \mathbf{r}^{(j-1)}\big)$
    Sample $\boldsymbol{\xi}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
    $z^{(j)} \leftarrow z^{(j-1)} + \eta_j \nabla_z \mathcal{L} + \sqrt{2\eta_j}\, \boldsymbol{\xi}_j$ ▷ ULA refinement
  end for
  $\tilde{\mathbf{x}}_0^{(k)} \leftarrow z^{(J)}$ ▷ Re-noising for next iteration
  Sample $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $x_{\text{in}}^{(k)} \leftarrow \tilde{\mathbf{x}}_0^{(k)} + \sigma_k\, \boldsymbol{\epsilon}_k$
end for
return $\hat{\mathbf{x}}_0 \leftarrow \tilde{\mathbf{x}}_0^{(K)}$

from a fresh state while maintaining measurement consistency. Unlike maximum a posteriori (MAP) estimation, our method performs stochastic sampling in the diffusion-defined latent space, exploring the posterior rather than collapsing to one mode. To improve convergence stability, we adopt an EDM-inspired annealing schedule in which the step size decays polynomially (exponent in $[-4, -7]$), concentrating more iterations in the low-noise regime where fine-scale refinement is most effective.

To provide a comprehensive structural overview of the proposed framework, we present the complete pseudocode for DAPS++ in Algorithm 1. This outlines the heterogeneous decomposition strategy, transitioning from the prior-dominant generation in Stage 1 to the likelihood-dominant refinement in Stage 2.

3.3 Initialization Quality and Convergence Analysis

To justify this decoupled strategy, we analyze the initialization quality and refinement convergence in turn, showing that once the estimate is near the data manifold, each stage needs to be performed only once.

Initialization quality. At a given noise level σ_t, the Tweedie estimate has already recovered the dominant structure of the signal from the score function: large-scale geometry, color distribution, and semantic content are established.

The measurement residual at this point is of order $O(|\mathcal{A}|\sigma_t) + O(\gamma\sqrt{d_y})$, where d_y is the measurement dimension. At early annealing stages where σ_t is large, the residual is also large, but high precision is not required since re-noising at the end of each cycle erases fine-scale corrections; at late stages where σ_t is small, the residual has shrunk to near the noise floor and only a few Langevin steps are needed. With $\bar{\sigma} = 10\gamma$, the transition from Tweedie to a single RK4 step occurs when the first-order truncation error begins to dominate the residual.

Convergence across problem types. The number of refinement steps required in Stage 2 is governed by two properties of the measurement operator in the neighborhood of $\hat{\mathbf{x}}_0$. The first is the condition number of the forward map, which controls the convergence rate of the likelihood gradient. The second is the degree of convexity of the negative log-likelihood near the initialization, which determines whether local refinement can reach the correct solution. A formal convergence bound relating these quantities to the required step count is provided in the Appendix. For well-conditioned linear operators, the likelihood is quadratic and convergence is fast: super-resolution requires only 2 steps because the downsampling operator preserves most signal energy, while Gaussian deblurring requires more steps due to its broader eigenvalue spread. For mild nonlinearities such as per-pixel tone mapping in HDR, the likelihood remains approximately convex near the initialization and convergence is comparable to the linear case (5 steps). For stronger nonlinearities such as nonlinear deblurring, the signal-dependent kernel introduces inhomogeneous curvature, requiring more iterations (50 steps). In phase retrieval, the likelihood has multiple equivalent minima; if the initialization falls in the wrong basin, local refinement cannot recover the correct solution.

3.4 Connections to Existing Diffusion-Based Solvers

The two-stage formulation clarifies why DAPS++ achieves its efficiency gains through structural decoupling rather than parameter reduction. Although many diffusion-based inverse problem approaches are presented as posterior samplers, recent studies show that their updates often resemble MAP-style iterative refinements [38]. The distinction between methods lies in how tightly they couple the diffusion prior with the likelihood at each step, and what computational cost that coupling incurs.

In DPS [3], the data-consistency term requires differentiating the measurement residual $\|\mathbf{y} - \mathcal{A}(\hat{\mathbf{x}}_0(\mathbf{x}_t))\|_2^2$ through Tweedie's estimator $\hat{\mathbf{x}}_0$, propagating gradients back through the score function at every diffusion timestep. This Jacobian-vector product through the score function at each of T steps is the primary computational bottleneck of DPS, and makes it structurally incompatible with higher-order ODE solvers, where additional score evaluations per step would each require a corresponding backward pass. From Eq. (3), the prior and likelihood gradients are statistically independent, so the DPS update decomposes

as

$$\mathbf{x}'_{t-1} = \underbrace{\mathbf{x}_t + \sigma_t^2\, s_\theta(\mathbf{x}_t, \sigma_t) + \sigma_{t-1}\mathbf{z}}_{\text{initialization}} - \underbrace{\eta_t \Big[\nabla_{\mathbf{x}_0} \|\mathbf{y} - \mathcal{A}(\mathbf{x}_0)\|_2^2\Big]_{\mathbf{x}_0=\hat{\mathbf{x}}_0(\mathbf{x}_t)}}_{\text{refinement}}, \tag{10}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and η_t denotes the step size. DPS therefore performs initialization and a single likelihood correction at every timestep, but couples them so that the score network and measurement operator must be evaluated jointly at each step.

DAPS [41] breaks this coupling at the algorithmic level by alternating Tweedie estimation with Langevin dynamics. By applying this directly to $\hat{\mathbf{x}}_0$ in image space, the method iteratively updates the clean-image estimate without back-propagating through the score function. However, because DAPS re-invokes the score network at each annealing level to preserve the time-marginal distribution $p(\mathbf{x}_t|\mathbf{y})$, the diffusion prior remains an active ingredient throughout sampling. As a result, a large number of Langevin steps per annealing level are required to approximately sample from $p(\mathbf{x}_0|\mathbf{x}_t, \mathbf{y})$, which is the necessary condition for theoretically preserving the time-marginal $p(\mathbf{x}_t|\mathbf{y})$ at each level. Without sufficiently converged Langevin samples at each stage, this guarantee breaks down.

DAPS++ takes this separation to its logical conclusion: the score network is used only in Stage 1, and Stage 2 operates entirely under the likelihood with no further access to the prior, removing the time-marginal constraint and yielding the approximately 90% reduction in NFEs reported in Sec. 4.3.

4 Experiments

4.1 Experimental Setup

We evaluate our method using pixel-space pre-trained diffusion models on the FFHQ [3] and ImageNet [9] datasets. For the noise schedule, we adopt the POLY(-7) discretization from EDM [17]. To improve efficiency, $\hat{\mathbf{x}}_0(\mathbf{x}_t)$ is generated using Eq. (7) while $\sigma_t > \bar{\sigma}$, and a single RK4 update is used once $\sigma_t \leq \bar{\sigma}$. Empirically, we set $\bar{\sigma} = 10 \times \gamma$ to balance computational cost and reconstruction accuracy. The total number of neural function evaluations (NFEs) remains low because each iteration requires only one network evaluation, further reduced by an annealed step-size schedule.

We use DAPS++-50 (50 annealing steps) for FFHQ and DAPS++-100 (100 steps) for ImageNet. Each DAPS++ iteration performs 4–8 MCMC refinement steps, except for nonlinear deblurring, which requires around 50 refinement steps in the refinement stage. Learning rates are tuned individually for each task. Additional details regarding model configurations, samplers, and hyperparameters appear in the Appendix, along with a discussion of sampling efficiency in Fig. 5.

Because the diffusion model already captures the underlying structure and fine-scale statistics of the data distribution, it effectively denoises the input while

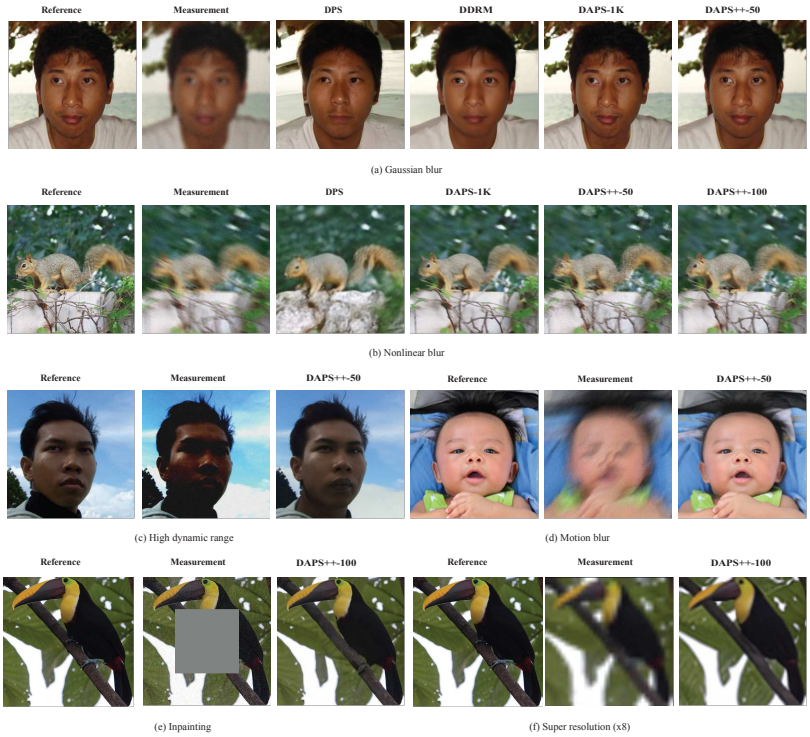

Fig. 3: (a) Gaussian blur reconstruction on the FFHQ-256 dataset compared with selected baselines. (b) Nonlinear blur reconstruction on ImageNet, comparing **DAPS++-50** and **DAPS++-100** alongside selected baselines. (c) High dynamic range (HDR) reconstruction on the FFHQ-256 dataset. (d) Motion blur reconstruction on the FFHQ-256 dataset. (e) Box inpainting results on ImageNet, where **DAPS++** accurately restores structural details. (f) 8× super-resolution using a pre-trained ImageNet diffusion model, demonstrating strong visual fidelity.

restoring plausible high-frequency details. Since the diffusion output approximates $\mathcal{A}(\mathbf{x}_0)$, the residual between this prediction and the measurement $\mathbf{y}$ is mostly dominated by noise. To ensure stable convergence and avoid overfitting to measurement noise, the injected diffusion noise at the end of each iteration must be smaller than the additive noise level. In practice, we use a noise schedule decaying from $\sigma_{\max} = 100$ to $\sigma_{\min} = 0.1$ with measurement noise $\gamma = 0.05$, which effectively suppresses overfitting and preserves structural fidelity throughout the DAPS family.

We compare our method against several state-of-the-art diffusion-based inverse problem solvers, including DAPS-1K, DAPS-100 [41], DPS [3], DDRM [19], DDNM [19], DiffPIR [43], and DCDP [20]. Note that SVD-based methods such as DDRM and DDNM apply only to linear inverse problems.

Table 1: Quantitative comparison on 100 validation images of **FFHQ** across four linear and two nonlinear inverse problems. Metrics are reported as SSIM↑ / LPIPS↓ / FID↓. Bold indicates the best result; underlined values denote the second best. All measurements include additive Gaussian noise with $\gamma = 0.05$.

Method	SR ×4	Inpainting	Gaussian Blur	Motion Blur	Nonlinear Blur	HDR
	SSIM / LPIPS / FID	SSIM / LPIPS / FID	SSIM / LPIPS / FID	SSIM / LPIPS / FID	SSIM / LPIPS / FID	SSIM / LPIPS / FID
DDNM	0.720 / 0.290 / 152.9	<u>0.810</u> / <u>0.162</u> / 94.7	**0.804** / 0.216 / 57.8	—	—	—
DDRM	**0.820** / <u>0.191</u> / 76.1	0.792 / 0.210 / 88.9	<u>0.786</u> / 0.218 / 89.4	—	—	—
DiffPIR	0.545 / 0.269 / 60.0	0.525 / 0.344 / 63.8	0.514 / 0.293 / 65.3	0.553 / 0.257 / 52.6	—	—
DCDP	0.642 / 0.333 / 92.9	0.757 / 0.167 / 42.3	0.719 / 0.282 / 82.2	0.508 / 0.384 / 115.2	**0.803** / **0.160** / **44.8**	—
DPS	0.591 / 0.357 / 81.1	0.727 / 0.259 / 75.6	0.647 / 0.285 / 73.3	0.588 / 0.327 / 76.8	0.648 / 0.281 / 74.3	0.693 / 0.284 / 77.6
DAPS-100	0.721 / 0.233 / 69.3	0.716 / 0.194 / 52.9	0.731 / 0.220 / 63.4	0.766 / 0.182 / 54.8	0.685 / 0.239 / 68.1	0.819 / 0.184 / 44.9
DAPS-1K	<u>0.782</u> / 0.192 / <u>55.5</u>	0.747 / 0.176 / <u>50.1</u>	<u>0.786</u> / <u>0.179</u> / <u>52.7</u>	**0.836** / <u>0.137</u> / <u>38.4</u>	<u>0.762</u> / <u>0.191</u> / <u>57.8</u>	**0.839** / **0.163** / **41.1**
DAPS++-50	0.781 / **0.176** / **46.0**	**0.812** / **0.141** / **42.1**	0.784 / **0.171** / **51.1**	<u>0.829</u> / **0.136** / **37.9**	0.745 / 0.194 / 54.9	<u>0.834</u> / <u>0.169</u> / <u>42.3</u>

Forward measurement operators. We evaluate our approach on a variety of linear and nonlinear inverse problems. Linear tasks include super-resolution, Gaussian and motion deblurring, and inpainting with box or random masks. For Gaussian and motion blur, we use 61×61 kernels with standard deviations of 3.0 and 0.5, respectively. Super-resolution employs bicubic downsampling by a factor of 4, while inpainting tasks use the standard 128×128 box mask. Nonlinear tasks include high dynamic range (HDR) reconstruction and nonlinear deblurring, following configurations described in [41].[1] All measurements are corrupted by additive white noise with standard deviation $\gamma = 0.05$.

Datasets and metrics. We evaluate the proposed method on two benchmark datasets: FFHQ (256×256) [18] and ImageNet (256×256) [7]. For both datasets, 100 validation images are used for quantitative evaluation under identical conditions. To comprehensively assess reconstruction quality, we report SSIM [36], LPIPS [42], and FID [14] as our primary metrics. All methods—including ours and the baselines—are evaluated on images normalized to $[0, 1]$ to ensure consistency and comparability across datasets.

4.2 Results

Main results. Tabs. 1 and 2 present the quantitative comparisons against various baselines, with qualitative examples shown in Fig. 3. Across both linear and nonlinear inverse problems, our method achieves competitive—or superior—reconstruction quality with higher SSIM and lower LPIPS and FID, while requiring substantially fewer NFEs than DAPS-1K. Under identical NFE budgets, DAPS++ consistently attains higher fidelity, demonstrating the effectiveness of fully decoupling the prior and likelihood stages and avoiding overfitting in either stage. DAPS++ also shows strong robustness under heavily degraded conditions, such as bicubic 8× super-resolution, where it can still recover meaningful structural information using only the ImageNet pre-trained model.

[1] Frequency-domain-based phase retrieval is discussed in the Supplementary Material, Section S2.2.

Table 2: Quantitative comparison on 100 validation images of **ImageNet** with four linear and two nonlinear tasks. Metrics follow SSIM↑ / LPIPS↓ / FID↓. Bold values denote the best performance; underlined entries indicate the second best. Each measurement includes additive Gaussian noise with $\gamma = 0.05$.

Method	SR ×4	Inpainting	Gaussian Blur	Motion Blur	Nonlinear Blur	HDR
	SSIM / LPIPS / FID	SSIM / LPIPS / FID	SSIM / LPIPS / FID	SSIM / LPIPS / FID	SSIM / LPIPS / FID	SSIM / LPIPS / FID
DDNM	0.487 / 0.458 / 215.3	0.656 / 0.289 / 157.9	0.650 / **0.231** / **90.8**	—	—	—
DDRM	0.652 / 0.296 / 111.2	0.602 / 0.336 / 163.7	0.584 / 0.361 / 161.0	—	—	—
DiffPIR	0.443 / 0.370 / **106.8**	0.500 / 0.390 / 149.2	0.395 / 0.428 / 154.6	0.152 / 0.683 / 264.8	—	—
DPS	0.454 / 0.474 / 208.4	0.623 / 0.343 / 157.5	0.526 / 0.355 / 155.6	0.542 / 0.365 / 161.4	0.520 / 0.365 / 151.5	0.349 / 0.552 / 213.8
DAPS-100	0.610 / 0.309 / 114.4	0.693 / 0.248 / 123.6	0.619 / 0.296 / 117.0	0.710 / 0.209 / 62.0	0.650 / 0.263 / 105.7	0.810 / 0.189 / **43.2**
DAPS-1K	0.638 / 0.295 / 109.1	0.715 / 0.229 / 114.0	0.658 / 0.268 / 107.6	**0.769** / **0.175** / **47.1**	**0.720** / 0.212 / 75.9	**0.824** / **0.171** / 44.7
DAPS++-50	0.653 / 0.283 / 111.7	0.760 / 0.208 / 113.3	**0.666** / 0.269 / 118.0	0.754 / 0.185 / 57.4	0.686 / 0.234 / 85.5	0.807 / 0.189 / 46.9
DAPS++-100	**0.661** / **0.276** / 114.2	**0.771** / **0.195** / **105.3**	0.663 / 0.273 / 122.9	0.763 / 0.179 / 56.1	0.718 / **0.211** / **75.7**	0.807 / 0.191 / 48.5

During the alternating initialization-refinement cycles, the optimization stages are deliberately non-overlapping. Conventional MAP-style updates in the refinement stage typically overfit measurement noise unless carefully regularized. In contrast, DAPS++ employs larger step sizes and fewer refinement iterations, leveraging the structure provided by the diffusion-based initialization while avoiding noise amplification. This design yields stable convergence and improved reconstruction quality at lower computational cost.

Small-step ODE solvers introduce notable discretization error when only a limited number of updates are performed. For example, DAPS-1K uses five Euler steps, resulting in an $\mathcal{O}(\Delta t)$ error that produces a weaker initialization compared with Tweedie-based estimation. As a consequence, more MCMC iterations are needed to return to the posterior manifold. DAPS++, by directly predicting the denoised mean in the initialization stage, avoids this discretization bottleneck and converges more rapidly with significantly fewer total steps. All baseline methods are evaluated on images normalized to $[0, 1]$ to ensure consistency and comparability across datasets.

Noise Tolerance. DAPS++ exhibits strong robustness under increasing noise levels. As shown in Tab. 3 and Fig. 4, when $\sigma_{\min} = 0.1$ and $\gamma = 0.1$, DAPS tends to overfit the noise, producing visible artifacts. In contrast, DAPS++ preserves structural fidelity and perceptual quality even as the noise level increases. As γ rises from 0.05 to 0.4, DAPS++ consistently exhibits stronger robustness, especially when $\sigma_{\min}$ is matched to γ, demonstrating its ability to prevent overfitting. This robustness arises from the structured re-noising mechanism and the reduced number of refinement steps, which together prevent overfitting and maintain diffusion-driven variability.

4.3 Ablation Studies

Efficiency comparison. We assess the efficiency of DAPS++ under varying computational budgets by evaluating configurations with diffusion NFEs ranging from 20 to 200, and comparing them against DAPS variants using 50 to 1K NFEs as well as several representative baselines. All methods are evaluated in

(a) DAPS++

(b) DAPS

Fig. 4: Example under γ=0.1, $\sigma_{\min}$=0.1. Comparison between (a) **DAPS++** and (b) **DAPS** for Gaussian blur shows that DAPS overfits noise, producing more artifacts in the final outputs.

terms of single-image sampling time on an NVIDIA A100 (80GB PCIe) GPU and reconstruction quality using the LPIPS metric. As shown in Fig. 5, the trade-off between sampling cost and reconstruction quality consistently favors DAPS++, which achieves higher fidelity at substantially lower computational cost across both linear and nonlinear inverse problems. Even at extremely low computational settings, such as DAPS++–20, the method converges rapidly and produces competitive reconstructions for linear degradation models, suggesting that the fully decoupled formulation provides an effective path toward stable likelihood-driven refinement without relying on heavy diffusion sampling. As computational budgets increase, DAPS++ exhibits progressively improved performance while maintaining a significantly lower cost profile than diffusion-based alternatives that interleave prior and likelihood updates at every timestep. Overall, these results demonstrate that DAPS++ achieves a favorable balance between accuracy and efficiency, making it a scalable and practical solution for real-time or resource-limited inverse imaging applications.

Comparison with DAPS on prior and data-consistency design. We examine the key design difference between DAPS++ and DAPS by evaluating how the diffusion prior is used during reconstruction. DAPS applies diffusion score together with likelihood updates and requires a large number of MCMC steps to approximately preserve the assumed time-marginal distribution. In contrast, DAPS++ uses prior guidance only in the initialization stage, after which

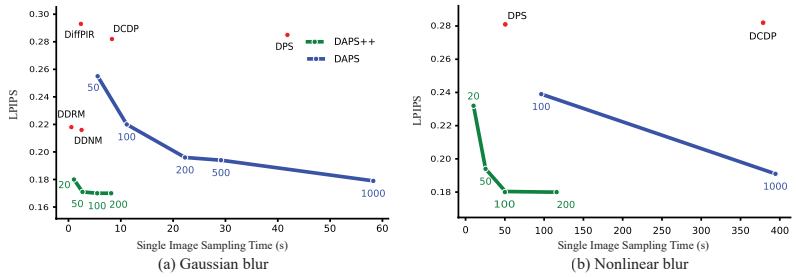

Fig. 5: Sampling time vs. reconstruction quality on an NVIDIA A100 (80GB PCIe) GPU. The y-axis shows LPIPS; results are averaged over 100 FFHQ images for Gaussian blur and nonlinear blur.

Table 3: Comparison of **DAPS-1K** and **DAPS++-50** under varying noise levels on the FFHQ Gaussian blur task, where $\sigma_{\min}$ is matched to γ.

Noise Level	γ=0.05, $\sigma_{\min}$=0.1	γ=0.1, $\sigma_{\min}$=0.1	γ=0.2, $\sigma_{\min}$=0.2	γ=0.4, $\sigma_{\min}$=0.4
	SSIM↑ / LPIPS↓	SSIM↑ / LPIPS↓	SSIM↑ / LPIPS↓	SSIM↑ / LPIPS↓
DAPS-1K	0.786 / 0.179	0.615 / 0.330	0.410 / 0.495	0.192 / 0.633
DAPS++-50	**0.784** / **0.171**	**0.728** / **0.191**	**0.674** / **0.229**	**0.579** / **0.303**

a small number of likelihood-only refinement steps is sufficient to enforce data consistency.

Tab. 4 shows the results of a focused FFHQ inpainting ablation under the same 100 diffusion-NFE budget. Comparing DAPS-100 with and without prior-gradient refinement shows nearly unchanged image quality and perceptual metrics, suggesting that the score gradient from diffusion guidance contributes little during data refinement in this setting. In contrast, the initialization strategy has a clearer effect. Direct Tweedie initialization improves the metrics over the DAPS baseline but tends to produce overly smooth reconstructions with insufficient details, motivating our thresholded transition from Tweedie estimation to a higher-order solver for initialization. Under the same computation budget, DAPS++ achieves better reconstruction quality and lower runtime than the DAPS baseline, indicating that the roles and design of the diffusion prior and data-consistency term are critical.

Hyperparameter choice. Two hyperparameters play a central role in our method: the threshold noise level $\bar{\sigma}$ and the polynomial scheduler exponent. Both parameters directly influence reconstruction quality, particularly when the number of sampling steps is limited and the interaction between diffusion initialization and likelihood refinement becomes more sensitive. The threshold $\bar{\sigma}$ determines when Tweedie-based estimation transitions to a higher-order ODE solver within the diffusion process, allowing efficient denoising at large noise levels while

Table 4: FFHQ inpainting ablation under the same 100 diffusion-NFE budget, isolating the effects of initialization and refinement strategy.

Method	Initialization	Refinement	SSIM↑ / LPIPS↓	Time (s)
DAPS-100	2-step Euler	Prior + likelihood	0.716 / 0.194	7.4
DAPS-100 w/o prior-gradient	2-step Euler	Likelihood only	0.716 / 0.194	6.8
DAPS + Tweedie init.	Tweedie	Prior + likelihood	0.741 / 0.186	10.1
DAPS + Tweedie/RK4	Tweedie/RK4	Prior + likelihood	0.712 / 0.192	5.9
DAPS++	Tweedie/RK4	Likelihood only	**0.812** / **0.141**	**2.4**

producing a reliable initialization with only a few high-order ODE evaluations. However, even under small noise settings, purely diffusion-generated estimates without data-driven refinement will drift away from the measurement; therefore, after the higher-order solver step, we still apply a refinement update at the end of the annealing schedule to correct this deviation.

The polynomial scheduler exponent governs the decay rate of the noise levels during annealing and becomes especially important in low-step regimes (*e.g.*, 20–50 steps), as it controls the trade-off between convergence stability and refinement strength. A slower decay stabilizes the refinement updates but introduces more high-noise transitions, which may reduce the effectiveness of likelihood refinement. A faster decay generally accelerates convergence and is often preferred in practice, provided that the refinement stage remains numerically stable. A comprehensive sensitivity analysis for these hyperparameters, along with practical tuning guidelines, is provided in the Appendix.

5 Conclusion

In this work, we revisit the underlying mechanism of diffusion-based inverse problem solvers and propose a fully decoupled initialization-refinement formulation that makes explicit the separation between the diffusion and data-consistency stages. Based on this perspective, we develop **DAPS++**, a fully decoupled framework that substantially improves sampling efficiency while preserving high reconstruction fidelity. The diffusion stage functions as a prior-driven initialization on a flat manifold, whereas the data-consistency stage performs likelihood-guided refinement independent of the diffusion process. This separation offers a clearer functional interpretation of diffusion-based inference and provides an explanation for the empirical behavior of existing approaches. Across both linear and nonlinear imaging tasks, DAPS++ achieves strong performance using only a fraction of the function evaluations required by conventional methods, establishing an efficient and scalable foundation for future research in diffusion-based inverse problems.

References

1. Cardoso, G., Idrissi, Y.J.E., Corff, S.L., Moulines, E.: Monte carlo guided diffusion for Bayesian linear inverse problems. In: International Conference on Learning Representations (2024)
2. Choi, J., Kim, S., Jeong, Y., Gwon, Y., Yoon, S.: Ilvr: Conditioning method for denoising diffusion probabilistic models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 14367–14376 (October 2021)
3. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: International Conference on Learning Representations (2023)
4. Chung, H., Sim, B., Ryu, D., Ye, J.C.: Improving diffusion models for inverse problems using manifold constraints. Advances in Neural Information Processing Systems **35**, 25683–25696 (2022)
5. Dalalyan, A.S.: Theoretical guarantees for approximate sampling from smooth and log-concave densities. Journal of the Royal Statistical Society Series B: Statistical Methodology **79**(3), 651–676 (04 2016)
6. Dashti, M., Stuart, A.M.: The bayesian approach to inverse problems. In: Handbook of uncertainty quantification, pp. 1–118. Springer (2015)
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
8. Dey, N., Blanc-Feraud, L., Zimmer, C., Roux, P., Kam, Z., Olivo-Marin, J.C., Zerubia, J.: Richardson–lucy algorithm with total variation regularization for 3d confocal microscope deconvolution. Microscopy research and technique **69**(4), 260–266 (2006)
9. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
10. Durmus, A., Moulines, É.: Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. The Annals of Applied Probability **27**(3), 1551 – 1587 (2017)
11. Feng, B.T., Smith, J., Rubinstein, M., Chang, H., Bouman, K.L., Freeman, W.T.: Score-based diffusion models as principled priors for inverse imaging. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10520–10531 (2023)
12. Fish, D., Brinicombe, A., Pike, E., Walker, J.: Blind deconvolution by means of the richardson–lucy algorithm. Journal of the Optical Society of America A **12**(1), 58–65 (1995)
13. Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. International statistical review **70**(3), 419–435 (2002)
14. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)
15. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
16. Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A.G., Tamir, J.: Robust compressed sensing mri with deep generative priors. Advances in neural information processing systems **34**, 14938–14954 (2021)
17. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. Advances in neural information processing systems **35**, 26565–26577 (2022)

18. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4401–4410 (2019)
19. Kawar, B., Elad, M., Ermon, S., Song, J.: Denoising diffusion restoration models. Advances in neural information processing systems **35**, 23593–23606 (2022)
20. Li, X., Kwon, S.M., Liang, S., Alkhouri, I.R., Ravishankar, S., Qu, Q.: Decoupled data consistency with diffusion purification for image restoration. arXiv preprint arXiv:2403.06054 (2024)
21. Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., Zhu, J.: Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. Advances in neural information processing systems **35**, 5775–5787 (2022)
22. Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., Zhu, J.: Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. Machine Intelligence Research pp. 1–22 (2025)
23. Lucy, L.: Optimum strategies for inverse problems in statistical astronomy. Astronomy and Astrophysics (ISSN 0004-6361), vol. 289, no. 3, p. 983-994 **289**, 983–994 (1994)
24. Lustig, M., Donoho, D., Pauly, J.M.: Sparse mri: The application of compressed sensing for rapid mr imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **58**(6), 1182–1195 (2007)
25. Porth, O., Chatterjee, K., Narayan, R., Gammie, C.F., Mizuno, Y., Anninos, P., Baker, J.G., Bugli, M., Chan, C.k., Davelaar, J., et al.: The event horizon general relativistic magnetohydrodynamic code comparison project. The Astrophysical Journal Supplement Series **243**(2), 26 (2019)
26. Shah, V., Hegde, C.: Solving linear inverse problems using gan priors: An algorithm with provable guarantees. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 4609–4613. IEEE (2018)
27. Song, B., Kwon, S.M., Zhang, Z., Hu, X., Qu, Q., Shen, L.: Solving inverse problems with latent diffusion models via hard data consistency. In: The Twelfth International Conference on Learning Representations (2024)
28. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021)
29. Song, Y., Ermon, S.: Generative modeling by estimating gradients of the data distribution. Advances in neural information processing systems **32** (2019)
30. Song, Y., Ermon, S.: Improved techniques for training score-based generative models. Advances in neural information processing systems **33**, 12438–12448 (2020)
31. Song, Y., Shen, L., Xing, L., Ermon, S.: Solving inverse problems in medical imaging with score-based generative models. In: International Conference on Learning Representations (2022)
32. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: ICLR (2021)
33. Stuart, A.M.: Inverse problems: a bayesian perspective. Acta numerica **19**, 451–559 (2010)
34. Tarantola, A.: Inverse problem theory and methods for model parameter estimation. SIAM (2005)
35. Wang, Y., Yu, J., Zhang, J.: Zero-shot image restoration using denoising diffusion null-space model. In: The Eleventh International Conference on Learning Representations (2023)

36. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
37. Wu, Z., Sun, Y., Chen, Y., Zhang, B., Yue, Y., Bouman, K.L.: Principled probabilistic imaging using diffusion models as plug-and-play priors. Advances in Neural Information Processing Systems **37**, 118389–118427 (2024)
38. Xu, T., Cai, X., Zhang, X., Ge, X., He, D., Sun, M., Liu, J., Zhang, Y.Q., Li, J., Wang, Y.: Rethinking diffusion posterior sampling: From conditional score estimator to maximizing a posterior. In: ICLR (2025)
39. Xu, X., Chi, Y.: Provably robust score-based diffusion posterior sampling for plug-and-play image reconstruction. Advances in Neural Information Processing Systems **37**, 36148–36184 (2024)
40. Yang, L., Ding, S., Cai, Y., Yu, J., Wang, J., Shi, Y.: Guidance with spherical gaussian constraint for conditional diffusion. In: Proceedings of the 41st International Conference on Machine Learning. ICML'24, JMLR.org (2024)
41. Zhang, B., Chu, W., Berner, J., Meng, C., Anandkumar, A., Song, Y.: Improving diffusion inverse problem solving with decoupled noise annealing. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 20895–20905 (2025)
42. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
43. Zhu, Y., Zhang, K., Liang, J., Cao, J., Wen, B., Timofte, R., Van Gool, L.: Denoising diffusion models for plug-and-play image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1219–1229 (2023)