
Learning Single Index Models with Diffusion Priors

Anqi Tang ^{*1} Youming Chen ^{*1} Shuchen Xue ^{2,3} Zhaoqiang Liu ¹

Abstract

Diffusion models (DMs) have demonstrated remarkable ability to generate diverse and high-quality images by efficiently modeling complex data distributions. They have also been explored as powerful generative priors for signal recovery, resulting in a substantial improvement in the quality of reconstructed signals. However, existing research on signal recovery with diffusion models either focuses on specific reconstruction problems or is unable to handle nonlinear measurement models with discontinuous or unknown link functions. In this work, we focus on using DMs to achieve accurate recovery from semi-parametric single index models, which encompass a variety of popular nonlinear models that may have *discontinuous* and *unknown* link functions. We propose an efficient reconstruction method that only requires one round of unconditional sampling and (partial) inversion of DMs. Theoretical analysis on the effectiveness of the proposed methods has been established under appropriate conditions. We perform numerical experiments on image datasets for different nonlinear measurement models. We observe that compared to competing methods, our approach can yield more accurate reconstructions while utilizing significantly fewer neural function evaluations.

1. Introduction

The objective of compressed sensing is to accurately recover the underlying high-dimensional sparse signal from a small number of linear measurements, with the measurement model as follows (Candes & Wakin, 2008; Foucart &

^{*}Equal contribution ¹University of Electronic Science and Technology of China ²University of Chinese Academy of Sciences ³Academy of Mathematics and Systems Science, CAS. Correspondence to: Zhaoqiang Liu <zqliu12@gmail.com>.

Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

Rauhut, 2013; Scarlett & Cevher, 2021):

$$\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the measurement matrix, $\mathbf{x}^* \in \mathbb{R}^n$ is the signal to be recovered, $\mathbf{e} \in \mathbb{R}^m$ is the additive noise vector, and $\mathbf{y} \in \mathbb{R}^m$ is the observed vector.

Although the linear measurement model utilized in compressed sensing can serve as a good testbed for demonstrating conceptual phenomena, in many real-world problems, it may not be justifiable or even feasible. For instance, the binary measurement model employed in 1-bit compressed sensing (Boufounos & Baraniuk, 2008) has attracted substantial interest. This is due to its cost-effective and efficient hardware. It has also been shown that 1-bit compressed sensing is robust against certain nonlinear distortions (Laska & Baraniuk, 2012). The limitations of the linear data model have prompted the study of general nonlinear measurement models. Among these, the semi-parametric single index model (SIM) is arguably the most popular (Horowitz, 2009). The SIM models the observations as follows:

$$\mathbf{y} = f(\mathbf{A}\mathbf{x}^*), \quad (2)$$

where \mathbf{A} consists of i.i.d. standard Gaussian entries, and $f : \mathbb{R} \rightarrow \mathbb{R}$ is an *unknown* and possibly random nonlinear link function that is applied element-wise. The aim is to estimate the signal \mathbf{x}^* using the knowledge of \mathbf{A} and \mathbf{y} , despite the *unknown* nonlinear link f . It is well-known that in the SIM, \mathbf{x}^* is generally not identifiable (Plan & Vershynin, 2016), since any scaling of \mathbf{x}^* can be incorporated into the unknown f . Consequently, it is common to impose the identifiability constraint $\|\mathbf{x}^*\|_2 = 1$.

To enable efficient and reliable recovery of the underlying signal \mathbf{x}^* , the predominant approach is to introduce structural modeling assumptions, such as sparse and conventional generative priors. For instance, the research works following (Bora et al., 2017) typically achieve signal reconstruction under the assumption that the underlying signal lies within the range of a variational autoencoder (VAE) or a generative adversarial network (GAN) (Van Veen et al., 2018; Hand et al., 2018; Heckel & Hand, 2019; Wu et al., 2019; Asim et al., 2020; Ongie et al., 2020; Whang et al., 2020; Jalal et al., 2021b; Nguyen et al., 2021; Liu et al., 2021b; 2022a; Genzel et al., 2022; Liu et al., 2024).

Compared to conventional generative models like VAEs and GANs, diffusion models (DMs) (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Ho et al., 2020) have recently demonstrated their superiority in terms of robust generative capabilities and learning complex data distributions (Dhariwal & Nichol, 2021). DMs have been integrated into a wide variety of domains, leading to transformative progress and reshaping traditional approaches in areas such as vision and audio generation (Saharia et al., 2022b; Kong et al., 2021), and reinforcement learning (Chi et al., 2023; Hansen-Estruch et al., 2023). In particular, when it comes to addressing signal recovery problems, diffusion models offer a more flexible and robust solution, facilitating high-quality reconstructions and improved performance (Kawar et al., 2022; Whang et al., 2022; Saharia et al., 2022c;a; Alkan et al., 2023; Chung et al., 2023a; Cardoso et al., 2023; Wang et al., 2023; Chung et al., 2023b; Feng & Bouman, 2023; Rout et al., 2023; Wu et al., 2023; Aali et al., 2024; Chung et al., 2024; Dou & Song, 2024; Song et al., 2024; Wu et al., 2024; Sun et al., 2024).

Inspired by the strong capabilities of DMs in diverse real-world applications, in this study, we investigate the application of DMs to learn SIMs.

1.1. Related Work

Relevant works on signal recovery with conventional generative models (such as GANs and VAEs) have shown that learned priors can drastically reduce sample complexity in compressed sensing and related problems. A comprehensive discussion of diffusion-based and conventional generative approaches is provided in Appendix A. In the following, we highlight diffusion-based signal recovery methods.

Signal recovery with diffusion models: Diffusion models (DMs) have advanced signal recovery by modeling complex data distributions, with two main paradigms: task-specific and pre-trained approaches. Task-specific DMs are trained for particular problems, such as super-resolution (Saharia et al., 2022c) or deblurring (Ren et al., 2023). Pre-trained DMs, used without task-specific training, guide reconstruction via score estimation (e.g., (Feng et al., 2023)) or adapt frameworks like DDPM for linear measurements (e.g., DDRM (Kawar et al., 2022)). MCG (Chung et al., 2022a) and ΠIGDM (Song et al., 2023) are mainly designed for the linear setting. Additionally, ΠIGDM can be extended to certain nonlinear settings (where the link function is known) by leveraging a combination of pseudoinverse operations and nonlinear transformations. DPS (Chung et al., 2023b) and DAPS (Zhang et al., 2024a) are applicable in nonlinear settings as well, also primarily under the assumption that the link function is known.

However, the works mentioned above either concentrate on

specific linear signal recovery problems or are incapable of handling nonlinear measurement models with *discontinuous* or *unknown* link functions. Perhaps most relevant to the present paper, the work (Meng & Kabashima, 2022) uses SGMs to learn the underlying distribution of natural signals. By estimating the conditional score function, it enables better reconstruction quality when dealing with quantized compressed sensing problems. Nevertheless, the approach in (Meng & Kabashima, 2022) is restricted to quantized measurements and is relatively slow in reconstruction. Even if utilizing the knowledge of the link function, it requires thousands of neural function evaluations (NFEs) to achieve a reasonably accurate recovery.

1.2. Contributions

The main contributions of this paper are threefold:

- We propose an efficient method for accurately recovering the underlying signal from nonlinear observations of SIMs by utilizing diffusion priors. Our approach *does not require the knowledge of the link function*, and it only requires one round of unconditional sampling and (partial) inversion of DMs, resulting in a relatively small number of NFEs for the entire process.
- Based on heuristic theoretical results, we opt to initiate the inversion of DMs from an intermediate time rather than performing the full inversion of DMs. Under suitable conditions, we provide a theoretical analysis on the effectiveness of the corresponding approach. Additionally, we empirically observe that carrying out partial inversion results in significantly better reconstructions compared to performing full inversion.
- To validate the effectiveness of our method, we conduct numerical experiments on image datasets under distinct nonlinear measurement models. Notably, in the noisy 1-bit measurement setting, our approach not only achieves substantially faster computational performance but also outperforms competing methods in reconstruction accuracy, even when these methods explicitly incorporate knowledge of the link function.

1.3. Notation

We use upper and lower case boldface letters to denote matrices and vectors respectively. For any $M \in \mathbb{N}$, we use the shorthand notation $[M] = \{1, 2, \dots, M\}$, and we use \mathbf{I}_M to denote the identity matrix in $\mathbb{R}^{M \times M}$. Given two sequences of real values $\{a_i\}$ and $\{b_i\}$, we write $a_i = O(b_i)$ if there exists an absolute constant C_1 and a positive integer i_1 such that for any $i > i_1$, $|a_i| \leq C_1 b_i$. We use \mathcal{S}^{n-1} to denote the unit sphere in \mathbb{R}^n , i.e., $\mathcal{S}^{n-1} = \{\mathbf{s} \in \mathbb{R}^n : \|\mathbf{s}\|_2 = 1\}$. For a generator $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$, we use $\mathcal{R}(G)$ to denote the range of G , i.e., $\mathcal{R}(G) = \{G(\mathbf{z}) : \mathbf{z} \in \mathbb{R}^n\}$.

2. Preliminaries for Diffusion Models

Diffusion Models (DMs) define a forward process $\{\mathbf{x}_t\}_{t \in [0, T]}$ starting from $t = 0$ and \mathbf{x}_0 . For any $t \in [0, T]$, the distribution of \mathbf{x}_t conditioned on \mathbf{x}_0 satisfies the following equation:

$$q_{0t}(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}_n). \quad (3)$$

Here, α_t and σ_t are non-negative, differentiable functions possessing bounded derivatives. Moreover, α_t is monotonically non-increasing with $\alpha_0 = 1$, while σ_t is monotonically increasing. It has been demonstrated that the following stochastic differential equation (SDE) has the same transition distribution $q_{0t}(\mathbf{x}_t | \mathbf{x}_0)$ for any $t \in [0, T]$ (Song et al., 2021b; Kingma et al., 2021):

$$d\mathbf{x}_t = f(t)\mathbf{x}_t dt + g(t)d\mathbf{w}_t, \quad \mathbf{x}_0 \sim q_0(\mathbf{x}_0), \quad (4)$$

where $\mathbf{w}_t \in \mathbb{R}^n$ is the standard Wiener process, and q_t represents the marginal distribution of \mathbf{x}_t . Moreover, $f(t)$ is a scalar function known as the drift coefficient and $g(t)$ is a scalar function called the diffusion coefficient, which are specified as follows:

$$f(t) = \frac{d \log \alpha_t}{dt}, \quad g^2(t) = \frac{d \sigma_t^2}{dt} - 2 \frac{d \log \alpha_t}{dt} \sigma_t^2. \quad (5)$$

Under certain regularity conditions, Song et al. (2021b) demonstrate that the forward process in Eq. (4) has an equivalent reverse process running from time T to 0, which starts with the marginal distribution $q_T(\mathbf{x}_T)$:

$$d\mathbf{x}_t = [f(t)\mathbf{x}_t - g^2(t)\nabla_{\mathbf{x}} \log q_t(\mathbf{x}_t)]dt + g(t)d\bar{\mathbf{w}}_t, \quad (6)$$

where $\bar{\mathbf{w}}_t \in \mathbb{R}^n$ represents the Wiener process in reverse time, and $\nabla_{\mathbf{x}} \log q_t(\mathbf{x}_t)$ stands for the time-varying score function.

For more efficient sampling, Song et al. (2021b) further show that the following probability flow ordinary differential equation (ODE) possesses the same marginal distribution at each time t as that of reverse SDE in Eq. (6):

$$\frac{d\mathbf{x}_t}{dt} = f(t)\mathbf{x}_t - \frac{1}{2}g^2(t)\nabla_{\mathbf{x}} \log q_t(\mathbf{x}_t), \quad \mathbf{x}_T \sim q_T(\mathbf{x}_T). \quad (7)$$

The only unknown term in Eq. (7) is the time-varying score function $\nabla_{\mathbf{x}} \log q_t(\mathbf{x}_t)$. In practice, DMs typically estimate the scaled score function $-\sigma_t \nabla_{\mathbf{x}} \log q_t(\mathbf{x}_t)$ by using a noise prediction network $\epsilon_{\theta}(\mathbf{x}_t, t)$, and optimizes the parameter θ through minimizing the following objective:

$$\int_0^T \mathbb{E}_{\mathbf{x}_0 \sim q_0} \mathbb{E}_{\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)} [\|\epsilon_{\theta}(\alpha_t \mathbf{x}_0 + \sigma_t \epsilon, t) - \epsilon\|_2^2] dt. \quad (8)$$

By substituting $\nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t) = -\epsilon_{\theta}(\mathbf{x}_t, t)/\sigma_t$ into Eq. (7) and taking advantage of its semi-linear structure, we obtain the following numerical integration formula (Lu et al., 2022a; Zhang & Chen, 2022):

$$\mathbf{x}_t = e^{\int_s^t f(\tau) d\tau} \mathbf{x}_s + \int_s^t \left(e^{\int_{\tau}^t f(r) dr} \cdot \frac{g^2(\tau)}{2\sigma_{\tau}} \cdot \epsilon_{\theta}(\mathbf{x}_{\tau}, \tau) \right) d\tau. \quad (9)$$

Alternatively, we can use a data prediction network $\mathbf{x}_{\theta}(\mathbf{x}_t, t)$ that satisfies the condition $\epsilon_{\theta}(\mathbf{x}_t, t) = (\mathbf{x}_t - \alpha_t \mathbf{x}_{\theta}(\mathbf{x}_t, t))/\sigma_t$ (Kingma et al., 2021; Salimans & Ho, 2022). By substituting Eq. (5) into Eq. (9) and using the transition between ϵ_{θ} and \mathbf{x}_{θ} , we obtain the following integration formula (Lu et al., 2022b):

$$\mathbf{x}_t = \frac{\sigma_t}{\sigma_s} \mathbf{x}_s + \sigma_t \int_{\lambda_s}^{\lambda_t} e^{\lambda} \hat{\mathbf{x}}_{\theta}(\hat{\mathbf{x}}_{\lambda}, \lambda) d\lambda, \quad (10)$$

where $\lambda_t := \log(\alpha_t/\sigma_t)$ is strictly decreasing with respect to t and has an inverse function $t_{\lambda}(\cdot)$. Thus, $\mathbf{x}_{\theta}(\mathbf{x}_t, t)$ can be written as $\mathbf{x}_{\theta}(\mathbf{x}_{t_{\lambda}(\lambda)}, t_{\lambda}(\lambda)) = \hat{\mathbf{x}}_{\theta}(\hat{\mathbf{x}}_{\lambda}, \lambda)$.

2.1. The Sampling Process of Diffusion Models

The sampling process in DMs gradually denoises a noise vector to an image vector that approximately follows the same distribution as the training data. More precisely, if we divide the time interval $[\epsilon, T]^1$ into N sub-intervals with $\epsilon = t_N < t_{N-1} < \dots < t_1 < t_0 = T$, and use a first-order numerical scheme for Eq. (10), for $i \in [N]$, we can derive the following iterative formula for the transition from time t_{i-1} to t_i :

$$\tilde{\mathbf{x}}_{t_i} = \frac{\sigma_{t_i}}{\sigma_{t_{i-1}}} \tilde{\mathbf{x}}_{t_{i-1}} + \sigma_{t_i} \left(\frac{\alpha_{t_i}}{\sigma_{t_i}} - \frac{\alpha_{t_{i-1}}}{\sigma_{t_{i-1}}} \right) \cdot \mathbf{x}_{\theta}(\tilde{\mathbf{x}}_{t_{i-1}}, t_{i-1}), \quad (11)$$

which is in line with the widely-used DDIM sampling method (Song et al., 2021a).

Moreover, second- or third-order numerical schemes can be employed to enhance the sampling efficiency further (Lu et al., 2022b; Zhao et al., 2024). Specifically, the second-order multi-step numerical solver for Eq. (10) yields the following iterative formula for $i \geq 2$ (Lu et al., 2022b):

$$\begin{aligned} \tilde{\mathbf{x}}_{t_i} = & \frac{\sigma_{t_i}}{\sigma_{t_{i-1}}} \tilde{\mathbf{x}}_{t_{i-1}} - \alpha_{t_i} (e^{-h_i} - 1) \\ & \times \left(\left(1 + \frac{1}{2r_i} \right) \mathbf{x}_{\theta}(\tilde{\mathbf{x}}_{t_{i-1}}, t_{i-1}) - \frac{1}{2r_i} \mathbf{x}_{\theta}(\tilde{\mathbf{x}}_{t_{i-2}}, t_{i-2}) \right), \end{aligned} \quad (12)$$

where $h_i = \lambda_{t_i} - \lambda_{t_{i-1}}$ and $r_i = h_{i-1}/h_i$. We refer to the corresponding sampling method as DM2M for brevity.

¹In practice, it is common to set the end time of sampling as a small $\epsilon > 0$ instead of 0 to avoid numerical instability, as described in (Lu et al., 2022a, Appendix D.1).

Let κ_i be the function corresponding to the i -th sampling step. For instance, for the DDIM sampling method in Eq. (11),² we have

$$\kappa_i(\mathbf{x}) := \frac{\sigma_{t_i}}{\sigma_{t_{i-1}}} \mathbf{x} + \sigma_{t_i} \left(\frac{\alpha_{t_i}}{\sigma_{t_i}} - \frac{\alpha_{t_{i-1}}}{\sigma_{t_{i-1}}} \right) \cdot \mathbf{x}_\theta(\mathbf{x}, t_{i-1}). \quad (13)$$

Then, the entire sampling process can be expressed as $\tilde{\mathbf{x}}_\epsilon = G(\tilde{\mathbf{x}}_T)$, where the generator $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the composition of $\kappa_1, \kappa_2, \dots, \kappa_N$, i.e.,

$$G(\mathbf{x}) = \kappa_N \circ \dots \circ \kappa_2 \circ \kappa_1(\mathbf{x}). \quad (14)$$

In addition, for convenience, for any $t \in (\epsilon, T]$, we use G_t to denote the sampling process from t to ϵ (and thus $G_T = G$). Specifically, if letting $i_t = \max\{i \in [N] : t_{i-1} \geq t\}$, we have that G_t is the composition of $\kappa_{i_t}, \kappa_{i_t+1}, \dots, \kappa_N$, i.e.,

$$G_t(\mathbf{x}) = \kappa_N \circ \dots \circ \kappa_{i_t+1} \circ \kappa_{i_t}(\mathbf{x}). \quad (15)$$

2.2. The Inversion of Diffusion Models

The inversion in DMs retraces the ODE sampling process with the aim of identifying the specific noise vector that generates a given image vector. This inversion is a crucial aspect in various applications, including image editing (Hertz et al., 2023; Kim et al., 2022; Wallace et al., 2023; Patashnik et al., 2023), watermark detection (Wen et al., 2023), and image-to-image translation (Kawar et al., 2022; Su et al., 2023). Recently, the inversion of DMs has attracted increasing attention, and several relevant methods have emerged (Pan et al., 2023; Zhang et al., 2024b; Wallace et al., 2023; Hong et al., 2024; Wang et al., 2024).

For example, the naive DDIM inversion method (Hertz et al., 2023; Kim et al., 2022) has the following iterative procedure:³

$$\hat{\mathbf{x}}_{t_{i-1}} = \frac{\sigma_{t_{i-1}}}{\sigma_{t_i}} \hat{\mathbf{x}}_{t_i} + \sigma_{t_{i-1}} \left(\frac{\alpha_{t_{i-1}}}{\sigma_{t_{i-1}}} - \frac{\alpha_{t_i}}{\sigma_{t_i}} \right) \mathbf{x}_\theta(\hat{\mathbf{x}}_{t_i}, t_{i-1}). \quad (16)$$

The naive DDIM inversion method circumvents the computational overhead of the implicit numerical scheme by substituting $\mathbf{x}_\theta(\hat{\mathbf{x}}_{t_{i-1}}, t_{i-1})$ with $\mathbf{x}_\theta(\hat{\mathbf{x}}_{t_i}, t_{i-1})$. It has been widely used in imaging applications like image editing. Alternatively, by taking into account the first-order numerical discretization of Eq. (10) from $t = \epsilon$ to $t = T$, it is straightforward to derive an iterative formula that is slightly

²For higher-order sampling methods, we can derive analogous representations for G in Eq. (14) and G_t in Eq. (15) using a recursive formula. An illustration of this process can be found in Lemma 5 within Appendix C.

³For brevity, we assume that the sampling and inversion of DMs utilize the same time steps. In practice, they might employ different time steps and have a varying number of steps.

different from Eq. (16):

$$\hat{\mathbf{x}}_{t_{i-1}} = \frac{\sigma_{t_{i-1}}}{\sigma_{t_i}} \hat{\mathbf{x}}_{t_i} + \sigma_{t_{i-1}} \left(\frac{\alpha_{t_{i-1}}}{\sigma_{t_{i-1}}} - \frac{\alpha_{t_i}}{\sigma_{t_i}} \right) \mathbf{x}_\theta(\hat{\mathbf{x}}_{t_i}, t_i). \quad (17)$$

Furthermore, similar to DM2M in Eq. (12), by considering the second-order numerical discretization of Eq. (10) from $t = \epsilon$ to $t = T$, we can obtain the corresponding iterative formula for the inversion of DMs.

Letting v_i be the function corresponding to the i -th inversion step. Then, the entire inversion procedure can be written as $\hat{\mathbf{x}}_T = G^\dagger(\hat{\mathbf{x}}_\epsilon)$, where the inversion operator $G^\dagger : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the composition of v_1, v_2, \dots, v_N , i.e.,

$$G^\dagger(\mathbf{x}) = v_1 \circ v_2 \circ \dots \circ v_N(\mathbf{x}). \quad (18)$$

In addition, for convenience, for any $t \in [\epsilon, T)$, we use G_t^\dagger to denote the inversion from t to T (and thus $G_\epsilon^\dagger = G^\dagger$). Specifically, if letting $j_t = \min\{j \in [M] : t_j \leq t\}$, we have that G_t^\dagger is the composition of $v_{j_t}, v_{j_t-1}, \dots, v_1$, i.e.,

$$G_t^\dagger(\mathbf{x}) = v_1 \circ v_2 \circ \dots \circ v_{j_t}(\mathbf{x}). \quad (19)$$

3. Setup and Approaches

Recall that the nonlinear observation vector $\mathbf{y} \in \mathbb{R}^m$ is assumed to be generated in accordance with the SIM in Eq. (2), where $\mathbf{x}^* \in \mathcal{S}^{n-1}$ is the signal to be estimated. We follow (Liu & Liu, 2022) to assume that the *unknown* link function f satisfies the following conditions:

$$\mu := \mathbb{E}_{\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)} [f(\mathbf{a}^T \mathbf{x}^*) \mathbf{a}^T \mathbf{x}^*] \neq 0, \quad (20)$$

$$\mathbb{E}_{\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)} [f(\mathbf{a}^T \mathbf{x}^*)^4] < \infty. \quad (21)$$

The conditions in Eqs. (20) and (21) are mild and hold true for various forms of f , such as $f(x) = \text{sign}(x)$ for 1-bit measurements and $f(x) = x^3$ for cubic measurements. However, as noted in prior works such as (Liu & Scarlett, 2020a; Liu & Liu, 2022), the condition in Eq. (20) is not satisfied for phase retrieval models with $f(x) = x^2$ or $f(x) = |x|$ (or their noisy versions).

Remark 1. The condition in Eq. (20) is a classic and crucial condition for SIMs. For example, it is (albeit implicitly) assumed in the seminal work (Plan & Vershynin, 2016) and in subsequent research that builds upon it. If this condition fails to hold, specifically when $\mu = 0$, the recovery of $\mu \mathbf{x}^*$ as (Plan & Vershynin, 2016) and in our Eq. (24) below becomes meaningless. Additionally, we follow (Liu & Liu, 2022) to assume the condition in Eq. (21), which generalizes the assumption that $f(\mathbf{a}^T \mathbf{x}^*)$ is sub-Gaussian (which is satisfied by quantized measurement models), and accommodates more general nonlinear measurement models such as cubic measurements with $f(x) = x^3$ and their noisy counterparts.

We utilize a generator $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ that corresponds to the sampling process of DMs (cf. Eq. (14)) to model the underlying signal. Specifically, we assume that $\mathbf{x}^* \in \mathcal{R}(G)$,⁴ where $\mathcal{R}(G)$ denotes the range of G . This assumption is standard for nonlinear measurement models with generative priors and has also been made in prior works such as (Wei et al., 2019; Liu & Scarlett, 2020a).

Let $G^\dagger : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be an operator corresponding to an inversion method for DMs (cf. Eq. (18)). The work (Liu & Liu, 2022) sets the estimated vector as $\mathcal{P}_G(\mathbf{A}^T \mathbf{y}/m)$, where \mathcal{P}_G denotes the projection operator onto the range of G . Since performing the projection step can be time-consuming even for conventional generative models like GANs, the authors of (Raj et al., 2019) use the composition of the pseudoinverse of the generator and the generator itself to approximate the projection operator. Building on these works, to reconstruct the direction of the signal \mathbf{x}^* from the knowledge of the measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and the observed vector $\mathbf{y} \in \mathbb{R}^m$ (despite the *unknown* link f), a natural idea is to set the estimated vector $\hat{\mathbf{x}}$ as follows:

$$\hat{\mathbf{x}} = G \circ G^\dagger \left(\frac{1}{m} \mathbf{A}^T \mathbf{y} \right). \quad (22)$$

However, such a simple idea does not fully exploit the relationship between $\frac{1}{m} \mathbf{A}^T \mathbf{y}$ and the underlying signal \mathbf{x}^* , and it leads to unsatisfactory empirical results for diffusion models, as we will see in Section 5. In particular, in the approach corresponding to Eq. (22), the inversion step starts from time ϵ (for an ϵ close to 0; see Section 2.1), and this implicitly assumes that $\frac{1}{m} \mathbf{A}^T \mathbf{y}$ approximately follows the target data distribution q_0 . However, based on the following lemmas, we observe that $\frac{1}{m} \mathbf{A}^T \mathbf{y}$ should instead be regarded as a noisy version of $\mathbf{x}^* \sim q_0$, especially when the number of samples m is relatively small compared to the ambient dimension n . First, it is easy to obtain the following lemma.

Lemma 1. *If $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, then we have that the following holds with high probability⁵*

$$\|\epsilon\|_\infty \leq C \sqrt{\log(2n)}, \quad (23)$$

where C is a sufficiently large positive constant.

Furthermore, we have the following lemma, which demonstrates that $\frac{1}{m} \mathbf{A}^T \mathbf{y}$ is a noisy version of $\mu \mathbf{x}^*$ (see Eq. (20) for the definition of μ) and approximately characterizes the noise level. The proofs of Lemmas 1 and 2 are placed in Appendix B.

⁴This assumption can also be approximately expressed as $\mathbf{x}^* \sim q_0$ when G corresponds to an excellent diffusion modeling of the target data distribution q_0 . The assumption that $\mathbf{x}^* \sim q_0$ is typically implicitly assumed by works that utilize diffusion models to solve compressed sensing problems.

⁵Here and in the rest of the paper, a statement is said to hold with high probability (w.h.p.) if it holds with probability at least 0.99.

Lemma 2. *Under conditions for the link function f in Eqs. (20) and (21), we have the following holds w.h.p.:*

$$\left\| \frac{1}{m} \mathbf{A}^T \mathbf{y} - \mu \mathbf{x}^* \right\|_\infty \leq \frac{C' \sqrt{\log(2n)}}{\sqrt{m}}, \quad (24)$$

where C' is a sufficiently large positive constant.

The comparison between Lemmas 1 and 2 inspires us to express $\frac{1}{m\mu} \mathbf{A}^T \mathbf{y}$ as

$$\frac{1}{m\mu} \mathbf{A}^T \mathbf{y} \approx \mathbf{x}^* + \frac{C'}{C\mu\sqrt{m}} \epsilon \quad (25)$$

for $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. Additionally, from Eq. (3), we know that \mathbf{x}_t can be written as $\mathbf{x}_t = \alpha_t \mathbf{x}_0 + \sigma_t \epsilon_t = \alpha_t (\mathbf{x}_0 + (\sigma_t/\alpha_t) \epsilon_t)$ with $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. Since the nonlinear link function f is *unknown* (and consequently, μ is also unknown), and C, C' are undetermined positive constants, we employ a tuning parameter C_s and find a time t^* such that⁶

$$\frac{\sigma_{t^*}}{\alpha_{t^*}} = \frac{C_s}{\sqrt{m}}, \quad (26)$$

and start the inversion from time t^* with the input vector being $\alpha_{t^*} C'_s \mathbf{A}^T \mathbf{y}/m$, where $C'_s > 0$ is a tuning parameter.

Remark 2. Unlike prior works such as (Chung et al., 2022b; Fabian et al., 2024; Wu et al., 2024) that identify an intermediate time for sampling, we seek an intermediate time mainly for the inversion of DMs. Moreover, this intermediate time is computed using theoretical findings in Lemma 2.

Then, the corresponding estimated vector is

$$\hat{\mathbf{x}} = G \circ G_{t^*}^\dagger \left(\frac{\alpha_{t^*} C'_s}{m} \mathbf{A}^T \mathbf{y} \right), \quad (27)$$

where $G_{t^*}^\dagger$ denotes the partial inversion operator (from t^* to T ; see Eq. (19)). One may also notice that since $\frac{1}{m} \mathbf{A}^T \mathbf{y}$ can be regarded as a noisy version of \mathbf{x}^* and the sampling process has a denoising effect, we can simply sample from t^* to ϵ without taking the inversion step. In this case, the corresponding estimated vector is

$$\hat{\mathbf{x}} = G_{t^*} \left(\frac{\alpha_{t^*} C'_s}{m} \mathbf{A}^T \mathbf{y} \right), \quad (28)$$

where G_{t^*} denotes the partial sampling operator (from t^* to ϵ ; see Eq. (15)). Indeed, such a simple and efficient approach can also lead to reasonably good reconstructions, although its performance is inferior to that of the approach corresponding to Eq. (27).

We refer to the approach for Eq. (22) as SIM-DMFIS (SIMs using diffusion models with full inversion and sampling

⁶If $\frac{C_s}{\sqrt{m}} > \frac{\sigma_T}{\alpha_T}$, set t^* to T , and if $\frac{C_s}{\sqrt{m}} < \frac{\sigma_\epsilon}{\alpha_\epsilon}$, set t^* to ϵ .

procedures), the approach for Eq. (28) as SIM-DMS since it only performs a (partial) sampling procedure, and the approach for Eq. (27) as SIM-DMIS. We illustrate the differences among these three approaches in Figure 1. For convenience, we present the SIM-DMIS approach in Algorithm 1.

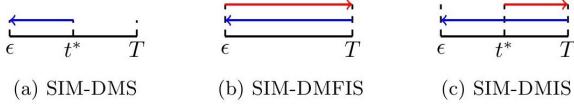


Figure 1. An illustration of our three approaches. For SIM-DMS, we only perform the sampling from t^* to ϵ . For SIM-DMFIS, we perform the full inversion and sampling procedures. For SIM-DMIS, we first perform the inversion from t^* to T , and then perform the sampling from T to ϵ .

Algorithm 1 The SIM-DMIS approach

Input: $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{y} \in \mathbb{R}^m$, \mathbf{x}_θ , time steps $\epsilon = t_N < t_{N-1} < \dots < t_1 < t_0 = T$, generator G corresponding to the sampling process as in Eq. (14), partial inversion operator G_t^\dagger as in Eq. (19), tuning parameters C_s and C'_s

1: Calculate $t^* \in [\epsilon, T]$ as in Eq. (26).

2: Calculate the estimated vector $\hat{\mathbf{x}}$ as in Eq. (27).

Output: $\hat{\mathbf{x}}$

4. Theoretical Analysis

In this section, we provide a theoretical analysis of the effectiveness of our proposed SIM-DMIS approach. Before presenting the analysis, we first present some preliminary results. Firstly, the following assumption states that the neural function $\mathbf{x}_\theta(\mathbf{x}, t)$ is Lipschitz continuous with respect to its first argument. This assumption has been widely adopted in relevant works such as (Lu et al., 2022a; Chen et al., 2023a;c; 2024; 2023d; Bortoli, 2022; Lee et al., 2022; Li et al., 2023).

Assumption 1 (Lipschitz Continuity of the Neural Function). *For any $t \in [0, T]$, there exists a positive constant $L_t > 0$ such that the neural function $\mathbf{x}_\theta(\mathbf{x}, t)$ is L_t -Lipschitz continuous with respect to its first parameter \mathbf{x} . That is, it holds for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ that $\|\mathbf{x}_\theta(\mathbf{x}_1, t) - \mathbf{x}_\theta(\mathbf{x}_2, t)\|_2 \leq L_t \|\mathbf{x}_1 - \mathbf{x}_2\|_2$.*

Based on Assumption 1, it can be easily demonstrated that the generator corresponding to popular sampling methods such as DDIM and DM2M (cf. Sec. 2.1) is Lipschitz continuous. Specifically, we have the following lemma.

Lemma 3. *Suppose that the data prediction network \mathbf{x}_θ satisfies Assumption 1. Then, if $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the generator as in Eq. (14), we have that G is L -Lipschitz continuous with $L > 0$ being dependent on $\{\alpha_{t_i}\}_{i=0}^{N-1}$, $\{\sigma_{t_i}\}_{i=0}^{N-1}$, and $\{L_{t_i}\}_{i=0}^{N-1}$.*

Lemma 3 shows that it is reasonable to assume that G is Lipschitz continuous. The proof of Lemma 3 and the precise characterization of the Lipschitz constant L for DDIM and DM2M are deferred to Appendix C.

Then, we have the following theorem regarding the effectiveness of the SIM-DMIS approach as presented in Eq. (27). Specifically, for $t \in [\epsilon, T]$, it offers an upper bound for the distance between $G \circ G_t^\dagger(\bar{\mathbf{x}}_t)$ and $\bar{\mathbf{x}}_\epsilon$, where $\bar{\mathbf{x}}_t$ is a sample of q_t and $\bar{\mathbf{x}}_\epsilon$ is on the same ODE trajectory as $\bar{\mathbf{x}}_t$. When ϵ is sufficiently small and the (scaled) data prediction network \mathbf{x}_θ provides a good approximation to the ground-truth score $\nabla \log q_t$ (Chen et al., 2023d; Li et al., 2023), $\bar{\mathbf{x}}_\epsilon$ will be close to the ground-truth data following q_0 that corresponds to $\bar{\mathbf{x}}_t$. Additionally, it has been demonstrated in (Lu et al., 2022a;b) that under certain regularity conditions, DDIM and DM2M are first- and second-order numerical sampling methods for Eq. (10), respectively. Moreover, by considering the first- and second-order numerical discretization of Eq. (10) from $t = \epsilon$ to $t = T$, we similarly obtain the first- and second-order numerical inversion methods for Eq. (10) respectively (for instance, the inversion method in Eq. (17) is a first-order numerical inversion method).

Therefore, based on Theorem 3, if the term $\alpha_{t^*} C'_s \mathbf{A}^T \mathbf{y} / m$ in Eq. (27) can be approximately expressed as $\alpha_{t^*} \mathbf{x}^* + \sigma_{t^*} \epsilon$ with ϵ being a standard normal vector, $\hat{\mathbf{x}} := G \circ G_{t^*}^\dagger(\alpha_{t^*} C'_s \mathbf{A}^T \mathbf{y} / m)$ is close to \mathbf{x}^* under appropriate conditions. The proof of Theorem 3 can be found in Appendix C.

Theorem 3. *For $t \in [\epsilon, T]$, let $\bar{\mathbf{x}}_t \in \mathbb{R}^n$ be a sample of q_t , and let $\bar{\mathbf{x}}_\epsilon = \frac{\sigma_\epsilon}{\sigma_t} \bar{\mathbf{x}}_t + \sigma_\epsilon \int_{\lambda_t}^{\lambda_\epsilon} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda$, which is the analytic solution of Eq. (10) with respect to the initial vector $\bar{\mathbf{x}}_t$. For $k_1, k_2 \in \{1, 2, 3\}$, suppose that G^\dagger corresponds to a k_1 -th order numerical inversion method for Eq. (10), and G_t^\dagger denotes the partial inversion operator from t to T (see Eq. (19)). Suppose that the generator G corresponds to a k_2 -th order numerical sampling method for Eq. (10) and G is L -Lipschitz continuous. Then, under certain regularity conditions,⁷ we have the following:*

$$\|\bar{\mathbf{x}}_\epsilon - G \circ G_t^\dagger(\bar{\mathbf{x}}_t)\|_2 = O(\sqrt{n}(h_{\max}^{k_2} + Lh_{\max}^{k_1})), \quad (29)$$

where $h_{\max} = \max_{i \in [N]} (\lambda_{t_i} - \lambda_{t_{i-1}})$.

5. Experiments

In this section, we conduct a series of experiments to validate the effectiveness of the proposed SIM-DMIS approach (see Algorithm 1). Specifically, we evaluate our method on two datasets: FFHQ 256×256 (Karras et al., 2019), ImageNet 256×256 (Deng et al., 2009).⁸ For FFHQ and Im-

⁷These conditions are similar to those in (Lu et al., 2022a, Appendix B.1) and are listed in Appendix C.

⁸Due to the page limit, the results of CIFAR-10 (Krizhevsky & Hinton, 2009) are included in Appendix D.

ageNet, the ambient dimension is $n = 3 \times 256 \times 256 = 196608$.

We perform experiments on nonlinear measurement models with 1-bit and cubic measurements, for which we have $\mathbf{y} = \text{sign}(\mathbf{A}\mathbf{x}^* + \mathbf{e})$ and $\mathbf{y} = (\mathbf{A}\mathbf{x}^*)^3 + \mathbf{e}$ respectively, where $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$. We compare SIM-DMIS with QCS-SGM proposed in (Meng & Kabashima, 2022), which uses DMs to handle quantized compressed sensing problems. We also compare our approach with the posterior sampling methods DPS (Chung et al., 2023b) and DAPS (Zhang et al., 2024a). DPS is a popular baseline for general signal recovery problems, and DAPS has shown excellent performance in nonlinear signal recovery problems. When DPS and DAPS utilize the knowledge of the link function f ,⁹ the corresponding methods are denoted as DPS-N and DAPS-N respectively. In addition, we compare with DPS and DAPS for the linear operator without using the knowledge of f , and the corresponding methods are denoted as DPS-L and DAPS-L respectively.

For the QCS-SGM method, we adhere to the experimental settings specified in (Meng & Kabashima, 2022) to ensure a fair comparison. Additionally, we compare with SIM-DMFIS and SIM-DMS (see Eqs. (22) and (28)), which are two variants of SIM-DMIS. We employ three metrics to assess the performance of these methods: Peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS). The reported quantitative results are averaged over 100 testing images. In the experimental process, we use DDIM for sampling and the DM2M inversion method (refer to Sec. 2.1 for details). The reported NFEs represent the total number of NFEs utilized by each approach. For instance, in the case of SIM-DMIS and SIM-DMFIS, the reported NFE is the sum of the number of steps in the inversion process and the number of steps in the sampling process. We adopt the variance preserving (VP) schedule with $\beta_{\min} = 0.1$ and $\beta_{\max} = 20$. For SIM-DMS, we use 50 NFEs, while for SIM-DMIS and SIM-DMFIS, we use 150 NFEs.¹⁰ The remaining parameters are set as $\epsilon = 0.001$, $T = 1$, and $\sigma = 0.05$. For the detailed parameter settings of C_s and C'_s in SIM-DMS and SIM-DMIS, please refer to Appendices F and G.

Results for FFHQ (256×256) Since the experiments for FFHQ are time-consuming, we only report the results for $m = n/8 = 24576$. For the QCS-SGM method, we leverage a pre-trained unconditional Score-SDE model with the

⁹For 1-bit measurements where f is not always differentiable, PyTorch can still enforce automatic differentiation.

¹⁰The same number of NFEs is used for the inversion process of SIM-DMIS and SIM-DMFIS. As demonstrated in Appendices F and G, this choice has a negligible impact on the experimental results.

variance exploding (VE) schedule¹¹ for the FFHQ dataset. As for the other methods, we employ the model from DPS, which has been trained on 49,000 FFHQ 256×256 images and validated on the first 1,000 images.

It is evident from Table 1 and Figure 2 that SIM-DMIS exhibits superior reconstruction performance with 1-bit measurements. Notably, SIM-DMIS requires only 150 NFEs, in contrast to QCS-SGM, which demands a significantly higher 11,555 NFEs. Additionally, both DPS and DAPS utilize 1,000 NFEs. The results for cubic measurements presented in Table 2 and Figure 3 further demonstrate the effectiveness of our proposed methods. In particular, SIM-DMIS consistently achieves the highest reconstruction quality across all evaluated metrics while requiring only 150 NFEs, demonstrating both the efficiency and robustness of our approach under the cubic measurement setting on the FFHQ dataset.

Note that in 1-bit measurements, even in the noiseless scenario, the link function $f(x) = \text{sign}(x)$ is non-differentiable at $x = 0$ (though as mentioned in Footnote 9, PyTorch can still enforce automatic differentiation). The non-differentiability also poses challenges for DPS-N and DAPS-N as they rely on f in gradient based updates. This can lead to inaccurate gradients and ultimately resulting in subpar performance.¹² For cubic measurements, since we have observed from the results for 1-bit measurements that SIM-DMFIS, DPS-L, and DAPS-L do not perform well, we do not compare against them. Furthermore, we do not include comparisons with DAPS-N since although the link function for cubic measurements is differentiable, its pronounced non-linearity produces unstable gradient directions that make optimization with DAPS-N unreliable and significantly degrade reconstruction performance.

Table 1. Quantitative results for FFHQ with 1-bit measurements ($m = n/8 = 24576$). We mark **bold** the best scores, and underline the second best scores.

Method	NFE	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
QCS-SGM	11555	12.91 ± 2.36	<u>0.51 ± 0.07</u>	0.50 ± 0.08
DPS-N	1000	11.14 ± 1.46	0.37 ± 0.09	0.69 ± 0.05
DPS-L	1000	8.57 ± 2.05	0.22 ± 0.08	0.69 ± 0.09
DAPS-N	1000	16.59 ± 0.54	0.33 ± 0.05	<u>0.48 ± 0.05</u>
DAPS-L	1000	5.63 ± 0.71	0.04 ± 0.03	0.61 ± 0.03
SIM-DMS	50	<u>17.14 ± 2.41</u>	0.44 ± 0.07	<u>0.48 ± 0.05</u>
SIM-DMFIS	150	8.48 ± 0.13	0.03 ± 0.00	0.90 ± 0.02
SIM-DMIS	150	19.87 ± 2.77	0.60 ± 0.09	0.37 ± 0.05

¹¹<https://huggingface.co/google/ncsnpp-ffhq-256>

¹²The supplementary material of DAPS suggests using Metropolis Hasting for non-differentiable forward operators, but it is also mentioned to have inferior performance and low efficiency, with results only in the supplementary material.

Table 2. Quantitative results on the FFHQ with cubic measurements ($m = n/8 = 24576$). We mark **bold** the best scores, and underline the second best scores.

Method	NFE	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
DPS-N	1000	<u>15.41</u> \pm 1.38	0.37 \pm 0.08	0.50 \pm 0.07
SIM-DMS	50	14.44 \pm 2.13	<u>0.51</u> \pm 0.10	0.47 \pm 0.08
SIM-DMIS	150	17.56 \pm 2.56	0.57 \pm 0.10	0.40 \pm 0.07

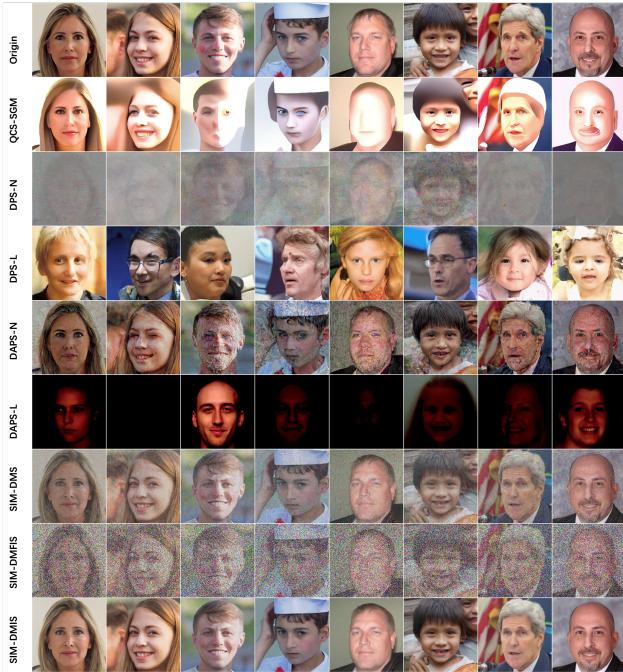


Figure 2. Examples of 1-bit reconstructed images for FFHQ.



Figure 3. Examples of cubic reconstructed images for FFHQ.

Results for ImageNet (256 \times 256) For ImageNet, we report the results for $m = n/16 = 12288$. We utilize a pre-trained conditional ImageNet 256 \times 256 model along with its corresponding classifier sourced from ADM (Dhariwal & Nichol, 2021). We adhere to the default recommended configuration settings of these models, substituting the unconditional ImageNet 256 \times 256 model initially provided by DPS.¹³ The tests are conducted on 100 images selected from the Validation1K Set.¹⁴ Given that QCS-SGM lacks pre-trained models for ImageNet, we do not compare with it.

Table 3 and Figure 4 demonstrate that SIM-DMIS achieves superior reconstruction with 1-bit measurements compared to DPS and DAPS variants, while requiring only 150 NFEs. The results with cubic measurements, summarized in Table 4 and Figure 5, further validate the advantages of our methods. Notably, SIM-DMIS achieves the best performance across all metrics while maintaining a low computational cost of only 150 NFEs. These results underscore the effectiveness and efficiency of our approach in handling cubic measurement settings on the ImageNet dataset.

Table 3. Quantitative results for ImageNet with 1-bit measurements ($m = n/16 = 12288$). We mark **bold** the best scores, and underline the second best scores.

Method	NFE	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
DPS-N	1000	11.57 \pm 2.63	0.24 \pm 0.11	0.64 \pm 0.04
DPS-L	1000	12.67 \pm 3.43	0.11 \pm 0.16	0.81 \pm 0.12
DAPS-N	1000	14.62 \pm 1.20	0.12 \pm 0.01	0.61 \pm 0.04
DAPS-L	1000	6.20 \pm 1.66	0.08 \pm 0.14	0.64 \pm 0.02
SIM-DMS	50	<u>16.93</u> \pm 3.16	<u>0.32</u> \pm 0.09	<u>0.54</u> \pm 0.09
SIM-DMFIS	150	11.89 \pm 2.24	0.08 \pm 0.02	0.76 \pm 0.06
SIM-DMIS	150	18.06 \pm 3.16	0.45 \pm 0.13	0.46 \pm 0.08

Table 4. Quantitative results on the ImageNet with cubic measurements ($m = n/16 = 12288$). We mark **bold** the best scores, and underline the second best scores.

Method	NFE	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
DPS-N	1000	11.47 \pm 2.97	0.22 \pm 0.08	0.62 \pm 0.06
SIM-DMS	50	<u>14.99</u> \pm 2.82	<u>0.33</u> \pm 0.06	<u>0.59</u> \pm 0.08
SIM-DMIS	150	16.28 \pm 3.38	0.40 \pm 0.15	0.55 \pm 0.10

¹³For DAPS, we follow the settings described in the original paper and employ a pre-trained unconditional model from ADM (Dhariwal & Nichol, 2021).

¹⁴<https://github.com/XingangPan/deep-generative-prior/>

6. Conclusion

In this paper, we present approaches for learning SIMs by making use of the sampling and inversion methods for pre-trained unconditional DMs. Theoretical analysis and numerical results are provided to illustrate the effectiveness of the proposed approaches.

Impact Statement

This paper presents work aimed at advancing the field of Machine Learning, particularly in the area of signal recovery using diffusion models (DMs). The proposed method applies DMs to enhance the recovery of signals from Single Index Models (SIMs), offering more flexible and robust solutions for complex signal reconstruction tasks. It could have broad applications in fields such as imaging, signal processing, and data reconstruction, with potential to improve the efficiency and accuracy of various real-world tasks. While no direct ethical concerns are associated with the methodology, the potential societal impacts of this work include improvements in computational efficiency and the enhancement of technologies reliant on image reconstruction and signal recovery.

Figure 4. Examples of 1-bit reconstructed images for ImageNet.



Figure 5. Examples of cubic reconstructed images for ImageNet.

References

- Aali, A., Daras, G., Levac, B., Kumar, S., Dimakis, A. G., and Tamir, J. I. Ambient diffusion posterior sampling: Solving inverse problems with diffusion models trained on corrupted data. <https://arxiv.org/2403.08728>, 2024.
- Alkan, C., Oscanoa, J., Abraham, D., Gao, M., Nurdinova, A., Setsompop, K., Pauly, J. M., Mardani, M., and Vasanawala, S. Variational diffusion models for blind MRI inverse problems. In *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems*, 2023.
- Asim, M., Daniels, M., Leong, O., Ahmed, A., and Hand, P. Invertible generative models for inverse problems: Mitigating representation error and dataset bias. In *ICML*, 2020.
- Aubin, B., Loureiro, B., Baker, A., et al. Exact asymptotics for phase retrieval and compressed sensing with random generative priors. In *MSML*, 2020.
- Bora, A., Jalal, A., Price, E., and Dimakis, A. G. Compressed sensing using generative models. In *ICML*, 2017.
- Bortoli, V. D. Convergence of denoising diffusion models under the manifold hypothesis. *Transactions on Machine Learning Research*, 2022. Expert Certification.
- Boufounos, P. T. and Baraniuk, R. G. 1-bit compressive sensing. In *CIS*, 2008.

- Candes, E. J. and Wakin, M. B. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2): 21–30, 2008. doi: 10.1109/MSP.2007.914731.
- Cardoso, G., Le Corff, S., Moulines, E., et al. Monte Carlo guided denoising diffusion models for Bayesian linear inverse problems. In *ICLR*, 2023.
- Chen, H., Lee, H., and Lu, J. Improved analysis of score-based generative modeling: User-friendly bounds under minimal smoothness assumptions. In *ICML*, 2023a.
- Chen, J., Scarlett, J., Ng, M., and Liu, Z. A unified framework for uniform signal recovery in nonlinear generative compressed sensing. In *NeurIPS*, 2023b.
- Chen, M., Huang, K., Zhao, T., and Wang, M. Score approximation, estimation and distribution recovery of diffusion models on low-dimensional data. In *ICML*, 2023c.
- Chen, S., Chewi, S., Li, J., Li, Y., Salim, A., and Zhang, A. R. Sampling is as easy as learning the score: Theory for diffusion models with minimal data assumptions. In *ICLR*, 2023d.
- Chen, S., Chewi, S., Lee, H., Li, Y., Lu, J., and Salim, A. The probability flow ODE is provably fast. In *NeurIPS*, 2024.
- Chi, C., Feng, S., Du, Y., Xu, Z., Cousineau, E., Burchfiel, B., and Song, S. Diffusion policy: Visuomotor policy learning via action diffusion. In *Robotics: Science and Systems*, 2023.
- Chung, H., Sim, B., Ryu, D., and Ye, J. C. Improving diffusion models for inverse problems using manifold constraints. In *NeurIPS*, 2022a.
- Chung, H., Sim, B., and Ye, J. C. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *CVPR*, 2022b.
- Chung, H., Kim, J., Kim, S., and Ye, J. C. Parallel diffusion models of operator and image for blind inverse problems. In *CVPR*, 2023a.
- Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C. Diffusion posterior sampling for general noisy inverse problems. In *ICLR*, 2023b.
- Chung, H., Lee, S., and Ye, J. C. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *ICLR*, 2024.
- Daras, G., Dean, J., Jalal, A., and Dimakis, A. Intermediate layer optimization for inverse problems using deep generative models. In *ICML*, 2021.
- Daras, G., Dagan, Y., Dimakis, A., and Daskalakis, C. Score-guided intermediate level optimization: Fast Langevin mixing for inverse problems. In *ICML*, 2022.
- Daras, G., Chung, H., Lai, C.-H., Mitsuji, Y., Ye, J. C., Milanfar, P., Dimakis, A. G., and Delbracio, M. A survey on diffusion models for inverse problems. <https://arxiv.org/2410.00083>, 2024.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In *CVPR*, 2009.
- Dhar, M., Grover, A., and Ermon, S. Modeling sparse deviations for compressed sensing using generative models. In *ICML*, 2018.
- Dhariwal, P. and Nichol, A. Diffusion models beat GANs on image synthesis. In *NeurIPS*, 2021.
- Dou, Z. and Song, Y. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *ICLR*, 2024.
- Fabian, Z., Tinaz, B., and Soltanolkotabi, M. DiracDiffusion: Denoising and incremental reconstruction with assured data-consistency. In *ICML*, 2023.
- Fabian, Z., Tinaz, B., and Soltanolkotabi, M. Adapt and diffuse: Sample-adaptive reconstruction via latent diffusion models. *Proceedings of machine learning research*, 235: 12723, 2024.
- Fei, B., Lyu, Z., Pan, L., Zhang, J., Yang, W., Luo, T., Zhang, B., and Dai, B. Generative diffusion prior for unified image restoration and enhancement. In *CVPR*, 2023.
- Feng, B. T. and Bouman, K. L. Efficient Bayesian computational imaging with a surrogate score-based prior. In *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems*, 2023.
- Feng, B. T., Smith, J., Rubinstein, M., Chang, H., Bouman, K. L., and Freeman, W. T. Score-based diffusion models as principled priors for inverse imaging. In *ICCV*, 2023.
- Foucart, S. and Rauhut, H. *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013.
- Genzel, M., Macdonald, J., and März, M. Solving inverse problems with deep neural networks—robustness included? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1119–1134, 2022.
- Hand, P., Leong, O., and Voroninski, V. Phase retrieval under a generative prior. In *NeurIPS*, 2018.

- Hansen-Estruch, P., Kostrikov, I., Janner, M., Kuba, J. G., and Levine, S. IDQL: Implicit Q-learning as an actor-critic method with diffusion policies. <https://arxiv.org/2304.10573>, 2023.
- Heckel, R. and Hand, P. Deep decoder: Concise image representations from untrained non-convolutional networks. In *ICLR*, 2019.
- Hertz, A., Mokady, R., Tenenbaum, J., Aberman, K., Pritch, Y., and Cohen-Or, D. Prompt-to-Prompt image editing with cross-attention control. In *ICLR*, 2023.
- Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.
- Hong, S., Lee, K., Jeon, S. Y., Bae, H., and Chun, S. Y. On exact inversion of DPM-Solvers. In *CVPR*, 2024.
- Horowitz, J. L. *Semiparametric and nonparametric methods in econometrics*, volume 12. 2009.
- Hyder, R., Shah, V., Hegde, C., and Asif, M. S. Alternating phase projected gradient descent with generative priors for solving compressive phase retrieval. In *ICASSP*, 2019.
- Jagatap, G. and Hegde, C. Algorithmic guarantees for inverse imaging with untrained network priors. In *NeurIPS*, volume 32, 2019.
- Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A. G., and Tamir, J. Robust compressed sensing MRI with deep generative priors. In *NeurIPS*, 2021a.
- Jalal, A., Karmalkar, S., Dimakis, A., and Price, E. Instance-optimal compressed sensing via posterior sampling. In *ICML*, 2021b.
- Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.
- Kawar, B., Elad, M., Ermon, S., and Song, J. Denoising diffusion restoration models. In *NeurIPS*, 2022.
- Killedar, V. and Seelamantula, C. S. Compressive phase retrieval based on sparse latent generative priors. In *ICASSP*, 2022.
- Kim, G., Kwon, T., and Ye, J. C. Diffusionclip: Text-guided diffusion models for robust image manipulation. In *CVPR*, 2022.
- Kingma, D., Salimans, T., Poole, B., and Ho, J. Variational diffusion models. In *NeurIPS*, 2021.
- Kong, Z., Ping, W., Huang, J., Zhao, K., and Catanzaro, B. DiffWave: A versatile diffusion model for audio synthesis. In *ICLR*, 2021.
- Krizhevsky, A. and Hinton, G. Learning multiple layers of features from tiny images. 2009.
- Laska, J. N. and Baraniuk, R. G. Regime change: Bit-depth versus measurement-rate in compressive sensing. *IEEE Trans. Sig. Proc.*, 60(7):3496–3505, 2012.
- Lee, H., Lu, J., and Tan, Y. Convergence for score-based generative modeling with polynomial complexity. In *NeurIPS*, 2022.
- Li, G., Wei, Y., Chen, Y., and Chi, Y. Towards faster non-asymptotic convergence for diffusion-based generative models. <https://arxiv.org/2306.09251>, 2023.
- Liu, J. and Liu, Z. Non-iterative recovery from nonlinear observations using generative models. In *CVPR*, 2022.
- Liu, Z. and Scarlett, J. The generalized Lasso with nonlinear observations and generative priors. In *NeurIPS*, 2020a.
- Liu, Z. and Scarlett, J. Information-theoretic lower bounds for compressive sensing with generative models. *IEEE J. Sel. Areas Inf. Theory*, 1(1):292–303, 2020b.
- Liu, Z., Gomes, S., Tiwari, A., and Scarlett, J. Sample complexity bounds for 1-bit compressive sensing and binary stable embeddings with generative priors. In *ICML*, 2020.
- Liu, Z., Ghosh, S., and Scarlett, J. Robust 1-bit compressive sensing with partial Gaussian circulant matrices and generative priors. In *ITW*, 2021a.
- Liu, Z., Ghosh, S., and Scarlett, J. Towards sample-optimal compressive phase retrieval with sparse and generative priors. In *NeurIPS*, 2021b.
- Liu, Z., Liu, J., Ghosh, S., Han, J., and Scarlett, J. Generative principal component analysis. In *ICLR*, 2022a.
- Liu, Z., Wang, X., and Liu, J. Misspecified phase retrieval with generative priors. In *NeurIPS*, 2022b.
- Liu, Z., Li, W., and Chen, J. Generalized eigenvalue problems with generative priors. In *NeurIPS*, 2024.
- Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., and Zhu, J. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *NeurIPS*, 2022a.
- Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., and Zhu, J. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models. <https://arxiv.org/2211.01095>, 2022b.
- Meng, X. and Kabashima, Y. Quantized compressed sensing with score-based generative models. <https://arxiv.org/2211.13006>, 2022.

- Nguyen, T. V., Jagatap, G., and Hegde, C. Provable compressed sensing with generative priors via Langevin dynamics. <https://arxiv.org/abs/2102.12643>, 2021.
- Ongie, G., Jalal, A., Metzler, C. A., et al. Deep learning techniques for inverse problems in imaging. *IEEE J. Sel. Areas Inf. Theory*, 1(1):39–56, 2020.
- Pan, Z., Gherardi, R., Xie, X., and Huang, S. Effective real image editing with accelerated iterative diffusion inversion. In *ICCV*, 2023.
- Patashnik, O., Garibi, D., Azuri, I., Averbuch-Elor, H., and Cohen-Or, D. Localizing object-level shape variations with text-to-image diffusion models. In *ICCV*, 2023.
- Plan, Y. and Vershynin, R. The generalized Lasso with non-linear observations. *IEEE Trans. Inf. Theory*, 62(3):1528–1537, 2016.
- Qiu, S., Wei, X., and Yang, Z. Robust one-bit recovery via ReLU generative networks: Improved statistical rates and global landscape analysis. In *ICML*, 2020.
- Raj, A., Li, Y., and Bresler, Y. GAN-based projector for faster recovery with convergence guarantees in linear inverse problems. In *ICCV*, 2019.
- Ren, M., Delbracio, M., Talebi, H., Gerig, G., and Milanfar, P. Multiscale structure guided diffusion for image deblurring. In *ICCV*, 2023.
- Rout, L., Raoof, N., Daras, G., Caramanis, C., Dimakis, A., and Shakkottai, S. Solving linear inverse problems provably via posterior sampling with latent diffusion models. In *NeurIPS*, 2023.
- Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., and Norouzi, M. Palette: Image-to-image diffusion models. In *SIGGRAPH*, 2022a.
- Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., Salimans, T., et al. Photorealistic text-to-image diffusion models with deep language understanding. In *NeurIPS*, 2022b.
- Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., and Norouzi, M. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022c.
- Salimans, T. and Ho, J. Progressive distillation for fast sampling of diffusion models. In *ICLR*, 2022.
- Scarlett, J. and Cevher, V. An introductory guide to Fano’s inequality with applications in statistical estimation. *Information-Theoretic Methods in Data Science*, 2021.
- Scarlett, J., Heckel, R., Rodrigues, M. R., Hand, P., and Eldar, Y. C. Theoretical perspectives on deep learning methods in inverse problems. *IEEE Journal on Selected Areas in Information Theory*, 3(3):433–453, 2022.
- Shah, V. and Hegde, C. Solving linear inverse problems using GAN priors: An algorithm with provable guarantees. In *ICASSP*, 2018.
- Shamshad, F. and Ahmed, A. Compressed sensing-based robust phase retrieval via deep generative priors. *IEEE Sens. J.*, 21(2):2286–2298, 2020.
- Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*, 2015.
- Song, B., Kwon, S. M., Zhang, Z., Hu, X., Qu, Q., and Shen, L. Solving inverse problems with latent diffusion models via hard data consistency. In *ICLR*, 2024.
- Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. In *ICLR*, 2021a.
- Song, J., Vahdat, A., Mardani, M., and Kautz, J. Pseudoinverse-guided diffusion models for inverse problems. In *ICLR*, 2023.
- Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. In *NeurIPS*, 2019.
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021b.
- Su, X., Song, J., Meng, C., and Ermon, S. Dual diffusion implicit bridges for image-to-image translation. *ICLR*, 2023.
- Sun, Y., Wu, Z., Chen, Y., Feng, B. T., and Bouman, K. L. Provable probabilistic imaging using score-based generative priors. *IEEE Transactions on Computational Imaging*, 2024.
- Van Veen, D., Jalal, A., Soltanolkotabi, M., et al. Compressed sensing with deep image prior and learned regularization. <https://arxiv.org/abs/1806.06438>, 2018.
- Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.
- Wallace, B., Gokul, A., and Naik, N. Edict: Exact diffusion inversion via coupled transformations. In *CVPR*, 2023.
- Wang, F., Yin, H., Dong, Y.-J., Zhu, H., Zhang, C., Zhao, H., Qian, H., and Li, C. BELM: Bidirectional explicit linear multi-step sampler for exact inversion in diffusion models. In *NeurIPS*, 2024.

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

Wei, X., Yang, Z., and Wang, Z. On the statistical rate of nonlinear recovery in generative models with heavy-tailed data. In *ICML*, 2019.

Wen, Y., Kirchenbauer, J., Geiping, J., and Goldstein, T. Tree-ring watermarks: Fingerprints for diffusion images that are invisible and robust. <https://arxiv.org/2305.20030>, 2023.

Whang, J., Lei, Q., and Dimakis, A. G. Compressed sensing with invertible generative models and dependent noise. In *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*, 2020.

Whang, J., Delbracio, M., Talebi, H., Saharia, C., Dimakis, A. G., and Milanfar, P. Deblurring via stochastic refinement. In *CVPR*, 2022.

Wu, L., Trippe, B., Naesseth, C., Blei, D., and Cunningham, J. P. Practical and asymptotically exact conditional sampling in diffusion models. In *NeurIPS*, 2023.

Wu, Y., Rosca, M., and Lillicrap, T. Deep compressed sensing. In *ICML*, 2019.

Wu, Z., Sun, Y., Chen, Y., Zhang, B., Yue, Y., and Bouman, K. L. Principled probabilistic imaging using diffusion models as plug-and-play priors. In *NeurIPS*, 2024.

Zhang, B., Chu, W., Berner, J., Meng, C., Anandkumar, A., and Song, Y. Improving diffusion inverse problem solving with decoupled noise annealing. *arXiv preprint arXiv:2407.01521*, 2024a.

Zhang, G., Ji, J., Zhang, Y., Yu, M., Jaakkola, T. S., and Chang, S. Towards coherent image inpainting using denoising diffusion implicit models. In *ICLR*, 2023.

Zhang, G., Lewis, J. P., and Kleijn, W. B. Exact diffusion inversion via bi-directional integration approximation. In *ECCV*, 2024b.

Zhang, Q. and Chen, Y. Fast sampling of diffusion models with exponential integrator. In *NeurIPS 2022 Workshop on Score-Based Methods*, 2022.

Zhao, W., Bai, L., Rao, Y., Zhou, J., and Lu, J. UniPC: A unified predictor-corrector framework for fast sampling of diffusion models. In *NeurIPS*, 2024.

A. Extended Related Work

In this subsection, we summarize some relevant works, which can roughly be divided into (i) nonlinear signal recovery with conventional generative models, and (ii) signal recovery with diffusion models.

Nonlinear signal recovery with conventional generative models: The seminal work (Bora et al., 2017) studies linear compressed sensing with conventional generative priors (such as GANs and VAEs) and demonstrates via numerical experiments on image datasets that using a pre-trained generative prior can significantly reduce the number of required measurements (compared to that of using the sparse prior) for accurate signal recovery. This has led to a significant volume of follow-up works in (Dhar et al., 2018; Hand et al., 2018; Shah & Hegde, 2018; Jagatap & Hegde, 2019; Asim et al., 2020; Ongie et al., 2020; Daras et al., 2021; Liu & Scarlett, 2020b; Jalal et al., 2021b;a), which explore various aspects of high-dimensional signal recovery with generative priors. A literature review in this area can be found in (Scarlett et al., 2022).

In particular, under conventional generative priors, there have been several studies on 1-bit compressed sensing (Qiu et al., 2020; Liu et al., 2020; 2021a) and phase retrieval (Hand et al., 2018; Hyder et al., 2019; Jagatap & Hegde, 2019; Aubin et al., 2020; Shamshad & Ahmed, 2020; Liu et al., 2021b; Killedar & Seelamantula, 2022; Liu et al., 2022b), among others. Moreover, SIMs that account for unknown nonlinearity and conventional generative priors have been explored in several studies, including (Wei et al., 2019; Liu & Scarlett, 2020a; Liu & Liu, 2022; Chen et al., 2023b). More specifically, the works (Liu & Scarlett, 2020a; Liu & Liu, 2022; Chen et al., 2023b) provide recovery guarantees for the generalized Lasso approach and practical algorithms under the assumption of Gaussian sensing vectors. In (Wei et al., 2019), under the extra assumption that the link function is differentiable, SIMs with link functions of first- and second-orders are studied through Stein's identity, allowing for general non-Gaussian sensing vectors.

Signal recovery with diffusion models: Due to the ability to model complex distributions, DMs have demonstrated remarkable performance in a variety of applications. The seminal work on DMs (Sohl-Dickstein et al., 2015) introduces a framework for the diffusion process, which has been further developed in multiple forms, such as denoising diffusion probabilistic models (DDPMs) (Ho et al., 2020) and score-based generative models (SGMs) (Song et al., 2021b). These models have been widely applied to signal recovery problems (Daras et al., 2024), which can be broadly classified into two main paradigms.

One paradigm is task-specific, in which DMs are trained for particular tasks. For example, the model in (Saharia et al., 2022c) is customized for super-resolution, while the model in (Ren et al., 2023) is developed for deblurring. The other paradigm makes use of pre-trained DMs that are not limited to any specific signal recovery problem. Works within this paradigm can be further classified according to how they handle the measurements. Some approaches use DMs as priors to direct the reconstruction process by estimating the score function, as shown in (Feng et al., 2023; Daras et al., 2022). Other approaches, such as DDRM (Kawar et al., 2022), adapt the DDPM framework specifically for linear signal recovery problems. By incorporating singular value decomposition to limit the solution space, DDRM can effectively handle linear measurements. However, its performance degrades when dealing with sparse or limited measurements, rendering it less effective in challenging situations.

To solve general noisy linear signal recovery problems, some recent works (Chung et al., 2022a; 2023b; Song et al., 2023; Fei et al., 2023; Fabian et al., 2023; Zhang et al., 2023) aim to estimate the posterior distribution using the unconditional diffusion model based on Bayes' rule. In particular, MCG (Chung et al., 2022a) approaches data consistency from the perspective of the data manifold. It proposes a manifold-constrained gradient to ensure that corrections stay on the data manifold. DPS (Chung et al., 2023b) abandons the projection step in the reverse process because it may cause the sampling path to deviate from the data manifold. Instead, it approximates the gradient of the posterior through a specially hand-designed strategy, where the measurements can be regarded as a signal conditioning the sampling process. ΠIGDM (Song et al., 2023) further develops this type of strategy into a unified expression that encompasses various signal recovery problems by using the Moore-Penrose pseudoinverse. ΠIGDM guides the diffusion process by matching the one-step denoising solution and the ground-truth measurements, after transforming both via a generalized pseudoinverse of the measurement model. DAPS (Zhang et al., 2024a) modifies the traditional diffusion process by decoupling consecutive steps in the sampling trajectory. This allows for greater variation between steps, improving the exploration of the solution space and leading to more accurate and stable reconstructions.

B. Proofs of Lemmas 1 and 2

Before presenting the proofs, we present the following simple tail bound for a Gaussian random variable.

Lemma 4. (Gaussian tail bound (Wainwright, 2019, Example 2.1)) *Suppose that $X \sim \mathcal{N}(\alpha, \sigma^2)$ is a Gaussian random variable with mean α and variance σ^2 . Then, for any $u > 0$, $\mathbb{P}(|X - \alpha| \geq u) \leq 2e^{-\frac{u^2}{2\sigma^2}}$.*

B.1. Proof of Lemma 1

For $u > 0$ and any $j \in [n]$, from Lemma 4, from Lemma 4, we have with probability at least $1 - 2e^{-\frac{u^2}{2}}$ that $|\epsilon_j| \leq u$. Taking a union bound over all $j \in [n]$, we obtain with probability at least $1 - 2ne^{-\frac{u^2}{2}}$ that $\|\epsilon\|_\infty \leq u$. Then, by setting $u = C\sqrt{\log(2n)}$, we obtain the desired result.

B.2. Proof of Lemma 2

Let $g_i = \mathbf{a}_i^T \mathbf{x}^*$. Since $\mathbf{x}^* \in \mathcal{S}^{n-1}$, we know that g_i are independent standard normal random variables. Additionally, note that

$$\frac{1}{m} \mathbf{A}^T \mathbf{y} - \mu \mathbf{x}^* = \frac{1}{m} \sum_{i=1}^m y_i \mathbf{a}_i - \mu \mathbf{x}^* \quad (30)$$

$$= \frac{1}{m} \sum_{i=1}^m y_i (\mathbf{a}_i - (\mathbf{a}_i^T \mathbf{x}^*) \mathbf{x}^*) + \frac{1}{m} \sum_{i=1}^m y_i (\mathbf{a}_i^T \mathbf{x}^*) \mathbf{x}^* - \mu \mathbf{x}^*. \quad (31)$$

For any $j \in [n]$, since $\mathbb{E}[a_{ij} g_i] = x_j^*$, a_{ij} can be written as

$$a_{ij} = x_j^* g_i + \sqrt{1 - (x_j^*)^2} r_{ij}, \quad (32)$$

where r_{ij} is a standard normal random variable that is independent of g_i . Then, the j -th entry of the first term in the right-hand side of Eq. (31) can be written as

$$\left[\frac{1}{m} \sum_{i=1}^m y_i (\mathbf{a}_i - (\mathbf{a}_i^T \mathbf{x}^*) \mathbf{x}^*) \right]_j = \frac{1}{m} \sum_{i=1}^m y_i (a_{ij} - g_i x_j^*) \quad (33)$$

$$= \sqrt{1 - (x_j^*)^2} \cdot \frac{1}{m} \sum_{i=1}^m y_i r_{ij}. \quad (34)$$

Note that $y_i = f(\mathbf{a}_i^T \mathbf{x}^*) = f(g_i)$. Thus r_{ij} is also independent of y_i . Let $M_2 := \mathbb{E}_{g \sim \mathcal{N}(0,1)}[f(g)^2]$ and \mathcal{E}_1 be the event that $\frac{1}{m} \sum_{i=1}^m y_i^2 \leq 2M_2$. From (Liu & Liu, 2022, Lemma 3), we have

$$\mathbb{P}(\mathcal{E}_1^c) \leq \frac{C_f}{m}, \quad (35)$$

where C_f is a positive constant depending on f . Moreover, conditioned on the event \mathcal{E}_1 , the term $\sum_{i=1}^m y_i r_{ij}$ in Eq. (34) is zero-mean Gaussian with the variance being $\sum_{i=1}^m y_i^2$. Then, from Lemma 4, we obtain

$$\mathbb{P}\left(\left|\sum_{i=1}^m y_i r_{ij}\right| > u\right) \leq 2 \exp\left(-\frac{u^2}{2 \sum_{i=1}^m y_i^2}\right) \leq 2 \exp\left(-\frac{u^2}{4mM_2}\right). \quad (36)$$

Taking a union bound over all $j \in [n]$, we obtain with probability at least $1 - 2n \exp(-\frac{u^2}{4mM_2})$ that

$$\left\| \frac{1}{m} \sum_{i=1}^m y_i (\mathbf{a}_i - (\mathbf{a}_i^T \mathbf{x}^*) \mathbf{x}^*) \right\|_\infty = \left\| \sqrt{1 - (x_j^*)^2} \cdot \frac{1}{m} \sum_{i=1}^m y_i r_{ij} \right\|_\infty \leq \frac{u}{m}. \quad (37)$$

Furthermore, the j -th entry of the second term in the right-hand side of Eq. (31) can be written as

$$\frac{1}{m} \sum_{i=1}^m y_i (\mathbf{a}_i^T \mathbf{x}^*) \mathbf{x}^* - \mu \mathbf{x}^* = \left(\frac{1}{m} \sum_{i=1}^m y_i (\mathbf{a}_i^T \mathbf{x}^*) - \mu \right) \mathbf{x}^*. \quad (38)$$

For any $\tau > 0$, let \mathcal{E}_2 be the event that $\left| \frac{1}{m} \sum_{i=1}^m y_i(\mathbf{a}_i^T \mathbf{x}^*) - \mu \right| < \tau$. From (Liu & Liu, 2022, Lemma 3), we have

$$\mathbb{P}(\mathcal{E}_2^c) \leq \frac{C'_f}{m\tau^2}, \quad (39)$$

where C'_f is a positive constant depending on f . Additionally, conditioned on \mathcal{E}_2 , we have

$$\left\| \left(\frac{1}{m} \sum_{i=1}^m y_i(\mathbf{a}_i^T \mathbf{x}^*) - \mu \right) \mathbf{x}^* \right\|_\infty = \left| \frac{1}{m} \sum_{i=1}^m y_i(\mathbf{a}_i^T \mathbf{x}^*) - \mu \right| \cdot \|\mathbf{x}^*\|_\infty \leq \tau. \quad (40)$$

Combining Eqs. (31), (35), (37), (39), and (40), we observe that by setting $u = C_1 \sqrt{\log(2n)} \cdot \sqrt{m}$ and $\tau = C_2 \sqrt{\log(2n)}/\sqrt{m}$ for some sufficiently large positive constants C_1 and C_2 and letting $C' = C_1 + C_2$, we obtain the desired result.

C. Proof of Theorem 3

First, we provide a proof for the auxiliary result in Lemma 3 for popular sampling methods such as DDIM and DM2M (cf. Sec. 2.1). Specifically, we have the following detailed version of Lemma 3.

Lemma 5. Suppose that the data prediction network \mathbf{x}_θ satisfies Assumption 1. Then, if $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the generator corresponding to the entire sampling process of DDIM (see Eq. (11)), we have that G is L -Lipschitz continuous with

$$L = \prod_{i=1}^N \left(\frac{\sigma_{t_i}}{\sigma_{t_{i-1}}} + \sigma_{t_i} \left(\frac{\alpha_{t_i}}{\sigma_{t_i}} - \frac{\alpha_{t_{i-1}}}{\sigma_{t_{i-1}}} \right) \cdot L_{t_{i-1}} \right). \quad (41)$$

Additionally, suppose that $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the generator corresponding to the entire sampling process of DM2M (see Eq. (12)), and suppose that for $i = 0, 1, 2, \dots, N$, $\tilde{\mathbf{x}}_{t_i}$ is \tilde{L}_i -Lipschitz continuous with respect to $\tilde{\mathbf{x}}_{t_0}$. Then, we have that \tilde{L}_0 can be set to 1 and \tilde{L}_1 can be set to be $\frac{\sigma_{t_1}}{\sigma_{t_0}} + \sigma_{t_1} \left(\frac{\alpha_{t_1}}{\sigma_{t_1}} - \frac{\alpha_{t_0}}{\sigma_{t_0}} \right) \cdot L_{t_0}$, and for $i \geq 2$, \tilde{L}_i can be calculated via the following recursive formula:

$$\tilde{L}_i = \left(\frac{\sigma_{t_i}}{\sigma_{t_{i-1}}} + \alpha_{t_i} (1 - e^{-h_i}) \cdot \left(1 + \frac{1}{2r_i} \right) L_{t_{i-1}} \right) \tilde{L}_{i-1} + \alpha_{t_i} (1 - e^{-h_i}) \cdot \frac{1}{2r_i} L_{t_{i-2}} \tilde{L}_{i-2}. \quad (42)$$

In particular, we have that G is L -Lipschitz continuous with $L = \tilde{L}_N$.

Proof. For DDIM, by the triangle inequality, we have the following inequality for any $i \in [N]$ and any $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$:

$$\|\kappa_i(\mathbf{x}_1) - \kappa_i(\mathbf{x}_2)\|_2 \leq \left(\frac{\sigma_{t_i}}{\sigma_{t_{i-1}}} + \sigma_{t_i} \left(\frac{\alpha_{t_i}}{\sigma_{t_i}} - \frac{\alpha_{t_{i-1}}}{\sigma_{t_{i-1}}} \right) \cdot L_{t_{i-1}} \right) \cdot \|\mathbf{x}_1 - \mathbf{x}_2\|_2, \quad (43)$$

where κ_i is defined in Eq. (13). Additionally, it is easy to show that if function f is L_f -Lipschitz and g is L_g -Lipschitz, then their composition $f \circ g$ is $(L_f L_g)$ -Lipschitz (Bora et al., 2017). Then, we obtain that the generator $G = \kappa_N \circ \dots \circ \kappa_2 \circ \kappa_1$ (see Eq. (14)) is L -Lipschitz continuous, with L given by Eq. (41). Note that for the first sampling step of DM2M, we need to use the first-order numerical scheme—DDIM. Then, similarly to the case of DDIM, based on Eq. (12), we can obtain the desired recursive formula for the Lipschitz constants. \square

Then, we list the regularity conditions on \mathbf{x}_θ required for Theorem 3 as follows. These conditions are similar to those in (Lu et al., 2022a, Appendix B.1).

- The total derivatives $\frac{d^j \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda)}{d\lambda^j}$ (as a function of λ) exist and are continuous for $j = 0, 1, 2, 3$.
- The neural function $\mathbf{x}_\theta(\mathbf{x}, t)$ is Lipschitz continuous with w.r.t. its first parameter \mathbf{x} (i.e., Assumption 1).
- $h_{\max} = \max_{i \in [N]} (\lambda_{t_i} - \lambda_{t_{i-1}}) = O(1/N)$.

Based on the above auxiliary results, we are now ready to present the proof of Theorem 3.

C.1. Proof of Theorem 3

Let \bar{G}_t^\dagger be the analytic inversion operator (corresponding to Eq. (10)) from time t to T . That is, for any $\mathbf{x} \in \mathbb{R}^n$, we have

$$\bar{G}_t^\dagger(\mathbf{x}) = \frac{\sigma_T}{\sigma_t} \mathbf{x} + \sigma_T \int_{\lambda_t}^{\lambda_T} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda. \quad (44)$$

Additionally, let \bar{G} be the analytic sampling operator (corresponding to Eq. (10)) from time T to ϵ . That is, for any $\mathbf{x} \in \mathbb{R}^n$, we have

$$\bar{G}(\mathbf{x}) = \frac{\sigma_\epsilon}{\sigma_T} \mathbf{x} + \sigma_\epsilon \int_{\lambda_T}^{\lambda_\epsilon} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda. \quad (45)$$

Let $\bar{\mathbf{x}}_T = \bar{G}_t^\dagger(\bar{\mathbf{x}}_t) = \frac{\sigma_T}{\sigma_t} \bar{\mathbf{x}}_t + \sigma_T \int_{\lambda_t}^{\lambda_T} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda$. We have

$$\bar{G} \circ \bar{G}_t^\dagger(\bar{\mathbf{x}}_t) = \bar{G}(\bar{\mathbf{x}}_T) \quad (46)$$

$$= \frac{\sigma_\epsilon}{\sigma_t} \bar{\mathbf{x}}_T + \sigma_\epsilon \int_{\lambda_T}^{\lambda_\epsilon} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda \quad (47)$$

$$= \frac{\sigma_\epsilon}{\sigma_T} \left(\frac{\sigma_T}{\sigma_t} \bar{\mathbf{x}}_t + \sigma_T \int_{\lambda_t}^{\lambda_T} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda \right) + \sigma_\epsilon \int_{\lambda_T}^{\lambda_\epsilon} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda \quad (48)$$

$$= \frac{\sigma_\epsilon}{\sigma_t} \bar{\mathbf{x}}_t + \sigma_\epsilon \int_{\lambda_t}^{\lambda_T} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda + \sigma_\epsilon \int_{\lambda_T}^{\lambda_\epsilon} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda \quad (49)$$

$$= \frac{\sigma_\epsilon}{\sigma_t} \bar{\mathbf{x}}_t + \sigma_\epsilon \int_{\lambda_t}^{\lambda_\epsilon} e^\lambda \hat{\mathbf{x}}_\theta(\hat{\mathbf{x}}_\lambda, \lambda) d\lambda \quad (50)$$

$$= \bar{\mathbf{x}}_\epsilon. \quad (51)$$

Therefore, we obtain

$$\|G \circ G_t^\dagger(\bar{\mathbf{x}}_t) - \bar{\mathbf{x}}_\epsilon\|_2 = \|G \circ G_t^\dagger(\bar{\mathbf{x}}_t) - \bar{G} \circ \bar{G}_t^\dagger(\bar{\mathbf{x}}_t)\|_2 \quad (52)$$

$$\leq \|G \circ G_t^\dagger(\bar{\mathbf{x}}_t) - G \circ \bar{G}_t^\dagger(\bar{\mathbf{x}}_t)\|_2 + \|G \circ \bar{G}_t^\dagger(\bar{\mathbf{x}}_t) - \bar{G} \circ \bar{G}_t^\dagger(\bar{\mathbf{x}}_t)\|_2 \quad (53)$$

$$\leq L \|G_t^\dagger(\bar{\mathbf{x}}_t) - \bar{G}_t^\dagger(\bar{\mathbf{x}}_t)\|_2 + \|G(\bar{\mathbf{x}}_T) - \bar{G}(\bar{\mathbf{x}}_T)\|_2, \quad (54)$$

where the last inequality follows from the L -Lipschitz continuity of G and $\bar{\mathbf{x}}_T = \bar{G}_t^\dagger(\bar{\mathbf{x}}_t)$. Under the regularity conditions listed above, similarly to (Lu et al., 2022a, Theorem 3.2),¹⁵ we have

$$\|G_t^\dagger(\bar{\mathbf{x}}_t) - \bar{G}_t^\dagger(\bar{\mathbf{x}}_t)\|_2 = O(\sqrt{n} h_{\max}^{k_1}) \quad (55)$$

and

$$\|G(\bar{\mathbf{x}}_T) - \bar{G}(\bar{\mathbf{x}}_T)\|_2 = O(\sqrt{n} h_{\max}^{k_2}). \quad (56)$$

Combining Eqs. (54), (55), and (56), we obtain the desired result.

D. Experimental Results for CIFAR-10 with 1-bit and cubic Measurements

In this section, we present the experimental results obtained from the CIFAR-10 dataset (Krizhevsky & Hinton, 2009) using 1-bit and cubic measurements. For the CIFAR-10 dataset, the ambient dimension is $n = 32 \times 32 \times 3 = 3072$, and the pixel values are normalized to the range $[0, 1]$. Prior experimental findings have shown that the SIM-DMIS method outperforms SIM-DMS and SIM-DMFIS in terms of reconstruction performance. Therefore, in this study, we limit our comparison to our SIM-DMIS approach against QCS-SGM, DPS-N, and DPS-L. For 1-bit measurements, since QCS-SGM cannot handle cubic measurements, we compare against OneShot proposed in (Liu & Liu, 2022), which uses GANs to solve nonlinear signal recovery problems. Additionally, since we have observed from the results for 1-bit measurements that SIM-DMFIS does not perform well, we do not compare against it. The experimental results demonstrate the superior performance of our proposed method SIM-DMIS across different datasets and measurement settings.

¹⁵Note that the term $O(h_{\max}^k)$ in the statement of (Lu et al., 2022a, Theorem 3.2) is added element-wise. Thus when considering the ℓ_2 norm, we have an extra \sqrt{n} factor, where n is the data dimension.

Results on CIFAR-10 (32×32) with 1-Bit Measurements For 1-bit measurements, we report results under two measurement settings: $m = 500$ and $m = 1000$, corresponding to approximately 16% and 32% of the ambient dimension n . We utilize a pre-trained unconditional DDPM model with the variance preserving (VP) schedule¹⁶ for CIFAR-10. Prior experimental findings have shown that the SIM-DMIS method outperforms SIM-DMS and SIM-DMFIS in terms of reconstruction performance. Therefore, in this study, we limit our comparison to our SIM-DMIS approach against QCS-SGM, DPS-N, and DPS-L. As shown in Table 5, SIM-DMIS consistently outperforms existing approaches across both $m = 500$ and $m = 1000$. Qualitative results in Figure 6 further confirm its strong reconstruction capability.

Table 5. Quantitative results for 1-bit with $m = 500$ and $m = 1000$ measurements. We mark **bold** the best scores, and underline the second best scores.

Method	NFE	$m = 500$			$m = 1000$		
		PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
QCS-SGM	1160	8.40 ± 3.75	0.25 ± 0.19	0.55 ± 0.20	11.76 ± 4.59	0.41 ± 0.22	0.41 ± 0.18
DPS-N	1000	9.43 ± 1.89	0.04 ± 0.06	0.57 ± 0.05	9.70 ± 2.19	0.04 ± 0.06	0.65 ± 0.08
DPS-L	1000	<u>11.53 ± 1.59</u>	0.08 ± 0.02	0.61 ± 0.03	<u>12.24 ± 2.71</u>	0.10 ± 0.05	0.69 ± 0.07
SIM-DMIS	150	16.24 ± 1.65	0.55 ± 0.10	0.40 ± 0.10	18.61 ± 1.65	0.65 ± 0.08	0.31 ± 0.07

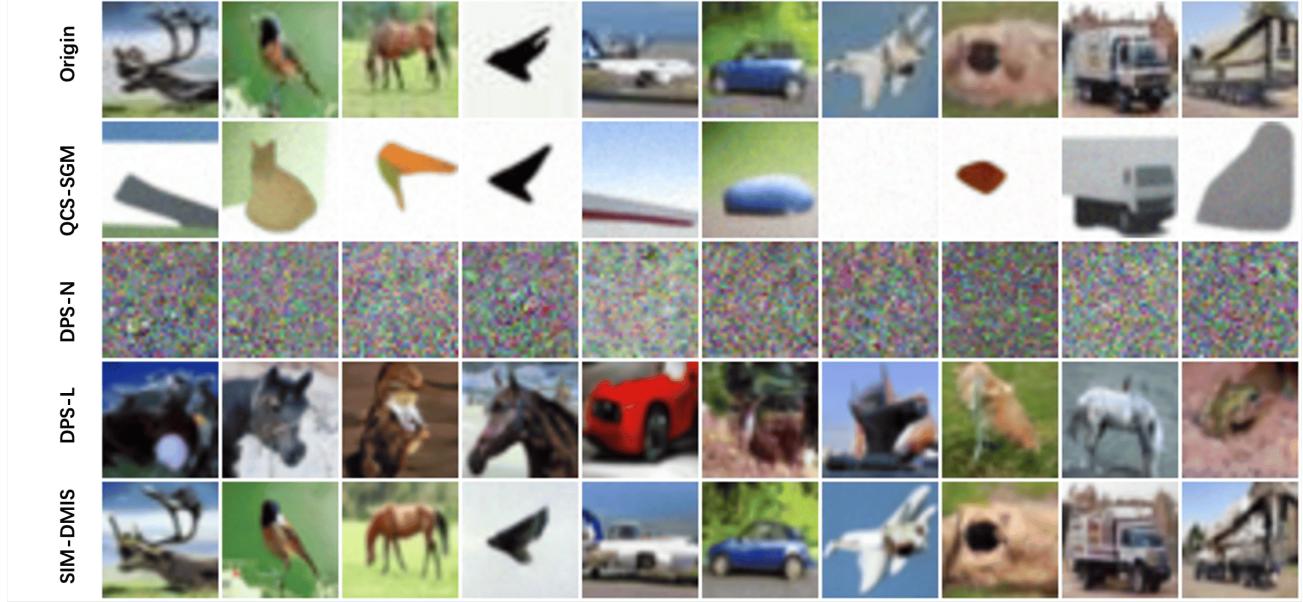


Figure 6. Examples of reconstructed images for CIFAR-10 with $m = 1000$ 1-bit measurements.

Results on CIFAR-10 (32×32) with Cubic Measurements The quantitative results are shown in Table 6, and examples of reconstructed images are presented in Figure 7. SIM-DMIS consistently outperforms OneShot across all measurement

¹⁶<https://huggingface.co/google/ddpm-cifar10-32>

settings. Specifically, with $m = 500$ measurements, SIM-DMIS achieves approximately 15% higher PSNR and 2% better SSIM compared to OneShot. With $m = 1000$ measurements, SIM-DMIS achieves approximately 25% higher PSNR and 12% better SSIM compared to OneShot. Furthermore, SIM-DMIS consistently yields a lower LPIPS value, indicating better perceptual similarity to the original images, compared to OneShot.

Table 6. Quantitative results for CIFAR-10 with cubic measurements. We mark **bold** the best scores, and underline the second best scores.

Method	$m = 500$			$m = 1000$		
	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
OneShot	13.68 \pm 2.09	0.33 \pm 0.09	0.50 \pm 0.04	14.61 \pm 2.09	0.39 \pm 0.09	0.48 \pm 0.04
SIM-DMIS	15.80\pm2.02	0.45\pm0.10	0.43\pm0.08	16.97\pm2.21	0.55\pm0.12	0.38\pm0.09

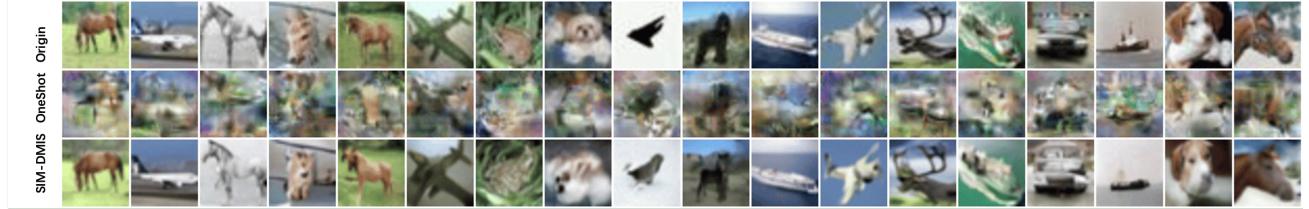


Figure 7. Examples of reconstructed images for CIFAR-10 with $m = 1000$ cubic measurements.

E. Additional Examples of Reconstructed Images for FFHQ and ImageNet with 1-bit Measurements

In this section, we present some additional examples of reconstructed images for the FFHQ 256×256 dataset in Figure 8, and also present some additional examples of reconstructed images for the ImageNet 256×256 dataset in Figure 9.



Figure 8. Examples of reconstructed images for FFHQ with $m = n/8 = 24576$ 1-bit measurements. Top: Origin, Middle: SIM-DMS, Bottom: SIM-DMIS.

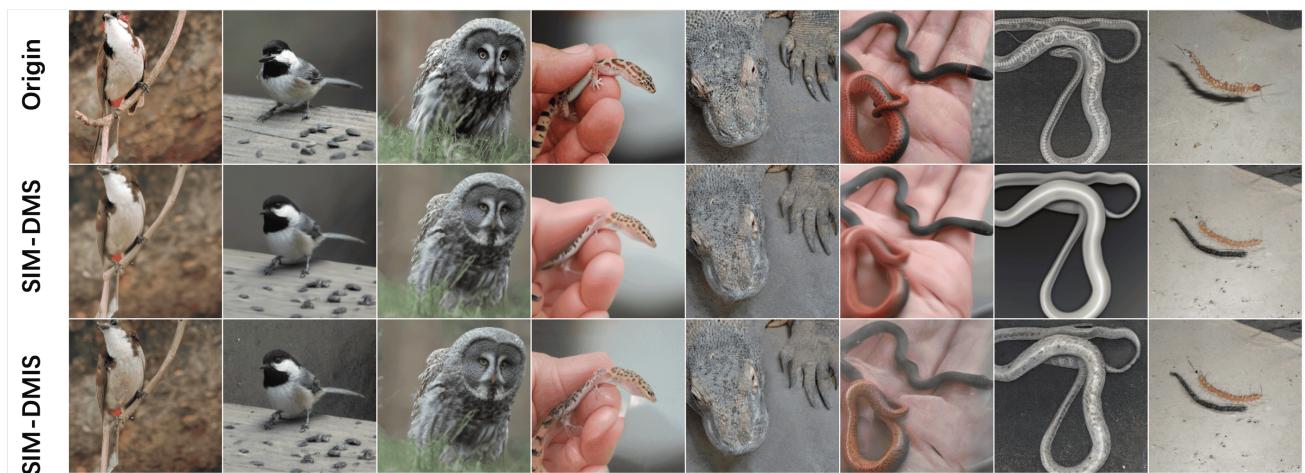


Figure 9. Examples of reconstructed images for ImageNet with $m = n/16 = 12288$ 1-bit measurements.

F. Ablation Study for SIM-DMS

In this section, we present a detailed analysis of the ablation experiments conducted on SIM-DMS using the CIFAR-10 dataset with $m = 500$ 1-bit measurements. The experimental setup follows the same framework as in the main body of the research, and the sampling technique is applied throughout.

The results shown in Table 7 provide valuable insights into the performance of SIM-DMS under different parameter configurations. Specifically, we analyze the impact of hyperparameters C_s , C'_s , and the number of function evaluations on model performance.

When varying C_s while fixing $C'_s = 55$ and NFE = 50, we observe a clear upward trend in both PSNR and SSIM, indicating improved signal quality and better structural similarity with increasing C_s . However, LPIPS reaches its minimum value at a moderate C_s , suggesting that the best perceptual quality is achieved with balanced parameter settings. On the other hand, when C_s is fixed at 55 and C'_s is varied, we find that lower values of C'_s yield better performance across all metrics, with PSNR and SSIM reaching their peak at the lowest tested value. LPIPS remains relatively stable but shows a slight decrease at lower C'_s values. Finally, when we adjust the NFE with $C_s = 55$ and $C'_s = 55$, all performance metrics show minimal fluctuation, suggesting that once NFE reaches a certain threshold, increasing it further does not yield significant improvements.

In conclusion, these results underscore the flexibility of SIM-DMS with respect to parameter tuning. The parameter C_s demonstrates a clear impact on reconstruction quality, while lower values of C'_s tend to yield better results. Additionally, the stability across different NFE values indicates the model's efficiency even with fewer iterations. These findings provide valuable guidance for parameter selection in practical applications of SIM-DMS.

Table 7. Ablation experiments for SIM-DMS on CIFAR-10 with $m = 500$ 1-bit measurements, with different values of C_s , C'_s and NFEs. We mark **bold** the best scores, and underline the second best scores.

Value	$C_s (C'_s = 55)$					$C'_s (C_s = 55)$					NFE ($C_s = 55, C'_s = 55$)			
	50	52	55	58	60	50	<u>52</u>	55	58	60	20	50	80	100
PSNR (\uparrow)	14.441 ± 1.037	14.625 ± 1.119	14.860 ± 1.201	14.956 ± 1.275	14.971 ± 1.334	15.337 <u>15.214</u>	<u>15.154</u> <u>15.125</u>	14.860 ± 1.201	14.451 ± 1.185	14.205 ± 1.146	14.860 ± 1.201	14.860 ± 1.201	14.860 ± 1.201	14.860 ± 1.201
SSIM (\uparrow)	0.427 ± 0.096	0.442 ± 0.099	0.460 ± 0.103	0.467 ± 0.106	0.470 ± 0.108	0.481 <u>0.475</u>	0.460 ± 0.104	0.440 ± 0.103	0.426 ± 0.101	0.460 ± 0.099	0.460 ± 0.103	0.460 ± 0.103	0.460 ± 0.103	0.460 ± 0.103
LPIPS (\downarrow)	0.444 ± 0.090	0.436 ± 0.090	0.433 0.094	0.441 ± 0.097	0.447 ± 0.099	0.440 ± 0.100	0.436 <u>0.433</u>	0.440 ± 0.098	0.447 ± 0.094	0.433 ± 0.091	0.433 ± 0.090	0.433 ± 0.094	0.433 ± 0.094	0.433 ± 0.094

G. Ablation Study for SIM-DMIS

In this section, we present a detailed analysis of the ablation experiments conducted on SIM-DMIS using the CIFAR-10 dataset with a measurement of $m = 500$. The experimental setup follows the same framework as in the main body of the research, and the DDIM sampling and DM2M inversion technique is applied throughout. The results are shown in Table 8.

When varying C_s while fixing $C'_s = 55$ and NFE = 50, the trend of PSNR and SSIM are consistent with the SIM-DMS where higher values of C_s also led to better signal quality and structural similarity. On the other hand, when C_s is fixed at 55 and C'_s is varied, the trends for PSNR and SSIM are largely similar to the SIM-DMS results, where lower values of C'_s lead to better performance in these metrics. Interestingly, while the improvement in PSNR and SSIM is clear, LPIPS remains relatively stable across a wide range of C'_s values, with a slight decrease at lower values of C'_s , particularly around 50 and 52. Finally, when adjusting the NFE with $C_s = 55$ and $C'_s = 55$, the results show minimal fluctuation in all performance metrics. This observation is consistent with the findings from SIM-DMS, where a similar behavior was noted.

In summary, the findings from the ablation study of SIM-DMIS underline several important insights. The parameter C_s has a notable impact on both the signal quality and structural similarity of the reconstructed images, with higher values leading to better PSNR and SSIM. The parameter C'_s plays a crucial role in enhancing PSNR and SSIM, with lower values yielding better results, while LPIPS remains stable. Additionally, the minimal effect of increasing NFE suggests that SIM-DMIS is highly efficient and provides stable performance even with fewer evaluations.

Table 8. Ablation experiments for **SIM-DMIS** on CIFAR-10 with $m = 500$ 1-bit measurements. We mark **bold** the best scores, and underline the second best scores.

Value	$C_s (C'_s = 55, \text{NFE}=150)$					$C'_s (C_s = 55, \text{NFE}=150)$					NFE ($C_s = 55, C'_s = 55$)			
	50	52	55	58	60	50	52	55	58	60	100	150	200	250
PSNR (\uparrow)	15.388 ± 1.218	15.562 ± 1.331	15.733 ± 1.420	15.809 ± 1.513	15.820 ± 1.592	16.247 <u>16.091</u>	15.733 ± 1.655	15.322 ± 1.500	15.057 ± 1.420	15.733 ± 1.355	15.733 ± 1.307	15.733 ± 1.420	15.733 ± 1.420	15.733 ± 1.420
SSIM (\uparrow)	0.489 ± 0.097	0.506 ± 0.098	0.522 ± 0.099	0.531 ± 0.101	0.534 ± 0.104	0.548 <u>0.540</u>	0.522 ± 0.101	0.500 ± 0.099	0.484 ± 0.099	0.522 ± 0.099	0.522 ± 0.099	0.522 ± 0.099	0.522 ± 0.099	0.522 ± 0.099
LPIPS (\downarrow)	0.406 ± 0.095	0.399 ± 0.093	0.402 ± 0.095	0.411 ± 0.100	0.417 ± 0.102	0.409 ± 0.103	0.405 ± 0.098	0.402 ± 0.095	0.404 ± 0.095	0.411 ± 0.093	0.402 ± 0.094	0.402 ± 0.095	0.402 ± 0.095	0.402 ± 0.095

H. Comparative Evaluation of SIM-DMS and SIM-DMIS on Performance and Efficiency

In this section, we evaluate the performance and efficiency of **SIM-DMS** and **SIM-DMIS** on the FFHQ dataset. The experimental setup follows the same framework as in the main body of the research. The results presented in the main paper (Table 1) show that **SIM-DMIS** consistently outperforms **QCS-SGM** and **SIM-DMFIS** in both reconstruction quality and NFE. Notably, **QCS-SGM** requires over 11,000 NFEs to achieve competitive results, incurring a high computational cost. Due to the performance of **SIM-DMIS** over both methods, and the particularly high computational cost of **QCS-SGM**, we exclude **QCS-SGM** and **SIM-DMFIS** from the comparisons in this section.

Performance Evaluation As discussed in Footnote 10 and Appendices F and G, ablation studies on CIFAR-10 show that further increasing the NFE for **SIM-DMS** does not lead to substantial performance enhancements.

To further verify this behavior and to compare the relative performance of **SIM-DMS** and **SIM-DMIS**, we conduct additional experiments on the FFHQ dataset. The results, summarized in Table 9, indicate that under the same NFEs, **SIM-DMIS** consistently outperforms **SIM-DMS** across all standard metrics.

Table 9. Quantitative results on FFHQ ($m = n/8 = 24576$). We mark **bold** the best scores, and underline the second best scores.

Method	NFE	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)	FID (\downarrow)
DPS-N	1000	11.14 ± 1.46	0.37 ± 0.09	0.69 ± 0.05	349.24
DPS-L	1000	8.57 ± 2.05	0.22 ± 0.08	0.69 ± 0.09	109.01
DAPS-N	1000	16.59 ± 0.54	0.33 ± 0.05	<u>0.48</u> ± 0.05	138.70
DAPS-L	1000	5.63 ± 0.71	0.04 ± 0.03	0.61 ± 0.03	322.75
SIM-DMS	50	17.14 ± 2.41	0.44 ± 0.07	<u>0.48</u> ± 0.05	105.04
SIM-DMS	150	17.72 ± 2.63	0.46 ± 0.08	<u>0.48</u> ± 0.06	95.52
SIM-DMIS	50	<u>18.78 ± 3.09</u>	0.58 ± 0.10	<u>0.41</u> ± 0.08	89.47
SIM-DMIS	150	19.87 ± 2.77	0.60 ± 0.09	0.37 ± 0.05	76.21

Reconstruction Speed We conduct additional experiments to measure the reconstruction speed of **SIM-DMS** and **SIM-DMIS**. The inference time is averaged over 10 validation images from the FFHQ dataset. All experiments are executed on a single NVIDIA GeForce RTX 4090 GPU. As shown in Table 10, the results indicate that **SIM-DMS** and **SIM-DMIS** exhibit significantly faster reconstructions when compared to the competing methods.

Table 10. Comparison of inference time (in seconds) for reconstructing 10 images on FFHQ. We mark **bold** the best scores, and underline the second best scores.

Method	NFE	Inference Time (s) (\downarrow)
DPS-N	1000	142
DPS-L	1000	142
DAPS-N	1000	160
DAPS-L	1000	160
SIM-DMS	50	1.96
SIM-DMIS	150	<u>5.66</u>