TAMING SCORE-BASED DENOISERS IN ADMM: A CONVERGENT PLUG-AND-PLAY FRAMEWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

While score-based generative models have emerged as powerful priors for solving inverse problems, directly integrating them into optimization algorithms such as ADMM remains nontrivial. Two central challenges arise: i) the mismatch between the noisy data manifolds used to train the score functions and the geometry of ADMM iterates, especially due to the influence of dual variables, and ii) the lack of convergence understanding when ADMM is equipped with score-based denoisers. To address the manifold mismatch issue, we propose ADMM plug-and-play (ADMM-PnP) with the AC-DC denoiser, a new framework that embeds a three-stage denoiser into ADMM: (1) auto-correction (AC) via additive Gaussian noise, (2) directional correction (DC) using conditional Langevin dynamics, and (3) score-based denoising. In terms of convergence, we establish two results: first, under proper denoiser parameters, each ADMM iteration is a weakly nonexpansive operator, ensuring high-probability fixed-point *ball convergence* using a constant step size; second, under more relaxed conditions, the AC-DC denoiser is a bounded denoiser, which leads to convergence under an adaptive step size schedule. Experiments on a range of inverse problems demonstrate that our method consistently improves solution quality over a variety of baselines.

1 INTRODUCTION

Inverse problems arise in many fields, including medical imaging (Song et al., 2022b; Jin et al., 2017; Arridge, 1999), remote sensing (Entekhabi et al., 1994; Combal et al., 2003), oceanography (Bennett, 1992), and computational physics (Raissi et al., 2019; Tarantola, 2005). Their solutions typically rely on incorporating prior knowledge or structural assumptions about the target signals, either through explicit regularization or data-driven models.

Classical approaches to inverse problems often rely on handcrafted regularizers, such as the ℓ_1 norm for sparsity (Yang et al., 2010; Elad & Aharon, 2006; Dabov et al., 2007) and the nuclear norm for low-rank structure (Semerci et al., 2014; Hu et al., 2017). Deep learning introduced a new paradigm of using learned generative models—VAEs, GANs, and normalizing flows—as data-driven regularizers (Ulyanov et al., 2020; Alkhouri et al., 2024; Shah & Hegde, 2018), offering more expressive priors by capturing complex distributions. More recently, pre-trained score functions from diffusion models have gained attention for inverse problems (Song et al., 2022b; Chung et al., 2023), as they effectively approximate data distributions and align solutions with the underlying data geometry (Xiao et al., 2022).

The use of pre-trained score functions in inverse problems mainly falls into two categories. The first modifies the MCMC process of diffusion sampling to incorporate observation information, as in DPS (Chung et al., 2023) and DDRM (Kawar et al., 2022), where observations guide unconditional score functions to perform posterior sampling. The second integrates score functions into deterministic or stochastic optimization algorithms; for example, Wang et al. (2024) and Song et al. (2023) use them as "projectors" to keep iterates on the desired data manifold. Furthermore, building on Tweedie's lemma, which links score functions to signal denoising, works such as (Zhu et al., 2023; Mardani et al., 2024; Renaud et al., 2024b) employ score as denoisers in proximal-gradient-like steps.

Challenges. The works in (Zhu et al., 2023; Mardani et al., 2024; Li et al., 2024) present flexible "plug-and-play (PnP)" paradigms that integrate diffusion models with optimization algorithms. However, two challenges remain in this line of work. First, score functions are trained on noisy data manifolds

constructed via Gaussian perturbations, whereas optimization iterates need not lie on such manifolds, leading to geometry mismatch and degraded denoising performance. Remedies such as stochastic regularization (Renaud et al., 2024b) or purification (Li et al., 2024) add Gaussian noise to the iterates, but this does not guarantee alignment with the score manifolds. Second, the theoretical understanding of these methods—particularly their convergence properties when combining the score denoisers with various optimization paradigms—remains limited.

Contributions. We propose to integrate score-based denoisers with the ADMM framework. Using ADMM iterates with score-based denoising is particularly challenging, as the presence of dual variables further distorts the "noise" geometry—likely explaining why score-based denoising has rarely been combined with primal–dual methods. Nevertheless, ADMM remains attractive for its flexibility in handling diverse inverse problems with multiple regularizers. Our contributions are:

▶ *Score-Based AC-DC Denoiser*: To mitigate the manifold mismatches, we propose a three-stage denoiser consisting of (1) additive Gaussian noise *auto-correction* (AC), (2) conditional Langevin dynamics-based *directional correction* (DC), and (3) score-based denoising. The AC stage pulls ADMM iterates toward neighborhoods of noise-trained manifolds, while DC refines alignment without losing signal information. This combination balances efficiency and accuracy, making score-based denoising effective within ADMM.

▶ *Convergence Analysis*: We show that, under proper AC-DC parameters, each ADMM iteration is weakly nonexpansive, ensuring convergence to a fixed-point neighborhood under constant step sizes under strongly convex losses. We further relax convexity and prove that an adaptive step-size scheme (Chan et al., 2016) guarantees convergence with high probability. These results extend prior ADMM-PnP convergence theory (Ryu et al., 2019; Chan et al., 2016) to score-based settings.

Our method is validated on diverse applications—including inpainting, phase retrieval, Gaussian and motion deblurring, super-resolution, and high dynamic range (HDR).

Notation. The detailed notation designation is listed in Appendix A.1.

2 Background

Inverse Problems. We consider the typical inverse problem setting where

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{x}) + \boldsymbol{\xi} \tag{1}$$

where $\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^n$ is the *measurement operator*, $n \leq d$, and $\boldsymbol{\xi}$ is additive noise. In some inverse problems, e.g., signal denoising and image deblurring, we have $n = d$; while for some other problems, e.g., data compression and recovery, we have $n < d$. The goal is to recover $\boldsymbol{x}$ from the $\boldsymbol{y}$, with the knowledge of $\mathcal{A}$. Structural regularization on $\boldsymbol{x}$ is often used to underpin the desired solution:

$$\min_{\boldsymbol{x}} \ \ell(\boldsymbol{y} \,||\, \mathcal{A}(\boldsymbol{x})) + h(\boldsymbol{x}), \tag{2}$$

where $\ell(\boldsymbol{y} \,||\, \mathcal{A}(\boldsymbol{x}))$ is a divergence term that measures the similarity of $\boldsymbol{y}$ and $\mathcal{A}(\boldsymbol{x})$ (e.g., $\|\boldsymbol{y} - \mathcal{A}(\boldsymbol{x})\|^2$), and $h(\boldsymbol{x})$ is a structural regularization term (e.g., $\|\boldsymbol{x}\|_1$ for sparse $\boldsymbol{x}$).

Diffusion-Based Inverse Problem Solving. Diffusion models can also be used for solving inverse problems, in ways more subtle than direct regularization. Consider training a diffusion model on $\boldsymbol{x}_0 \sim p_{\text{data}}$ via *denoising score matching* (Song et al., 2021), where the forward process is $\boldsymbol{x}_t|\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{x}_0, \sigma(t)\boldsymbol{I}), \quad t \in [0, T]$, with variance schedule $\sigma(0) = 0$ and $\sigma(t)$ increasing in t. After training, the model provides a score function $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma(t)) \approx \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t)$, where $p(\boldsymbol{x}_t)$ is the marginal density of $\boldsymbol{x}_t$. This induces noisy data manifolds

$$\mathcal{M}_{\sigma(t)} = \text{supp}(\boldsymbol{x}_t), \quad \forall t \in [T],$$

which are continuous since $\boldsymbol{x}_t$ is generated by Gaussian perturbations of $\boldsymbol{x}_0 \sim p_{\text{data}}$. These score functions can then be leveraged in different ways to assist inverse problem solving.

▶ *Posterior Sampling*: Many works formulate inverse problems as posterior sampling from $p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})$. These methods approximate $\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t)$ and combine it with the learned score $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, \sigma(t)) \approx \nabla \log p(\boldsymbol{x}_t)$ to perform stochastic sampling as t decreases (Chung et al., 2023; Song et al., 2022a; Kawar et al., 2021; 2022; Wang et al., 2023). While effective, their performance is often limited by the accuracy of the approximation to $\nabla \log p(\boldsymbol{y}|\boldsymbol{x}_t)$.

► *Plug-and-Play (PnP) Approaches*: Instead of sampling schemes, another line of work employs deterministic or stochastic optimization to solve (2), plugging score functions into the updates as structural regularizers. A representative example is DiffPIR (Zhu et al., 2023), which adopts a variable-splitting reformulation, $\min_{\boldsymbol{x},\boldsymbol{z}} \ell(\boldsymbol{y}\|\mathcal{A}(\boldsymbol{x})) + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + h(\boldsymbol{z})$, where the $\boldsymbol{z}$-subproblem at iteration k reduces to a standard denoising step:

$$\textbf{Denoising: } \boldsymbol{z}^{(k+1)} = \arg\min_{\boldsymbol{z}} \frac{\mu}{2}\left\|\boldsymbol{x}^{(k+1)} - \boldsymbol{z}\right\|_2^2 + h(\boldsymbol{z}). \tag{3}$$

This step is then tackled using a score-based denoiser:

$$\boldsymbol{z}^{(k+1)} \leftarrow D_{\sigma^{(k)}}(\widetilde{\boldsymbol{x}}^{(k+1)}) = \widetilde{\boldsymbol{x}}^{(k+1)} + (\sigma^{(k)})^2 \boldsymbol{s}_{\boldsymbol{\theta}}(\widetilde{\boldsymbol{x}}^{(k+1)}, \sigma^{(k)}) \tag{4}$$

where $\widetilde{\boldsymbol{x}}^{(k+1)} = \boldsymbol{x}^{(k+1)} + \sigma^{(k)}(\zeta^{(\widetilde{\boldsymbol{x}}^{(k)} - \boldsymbol{x}^{(k+1)})/\sigma^{(k)}} + (1-\zeta)\boldsymbol{n})$ with $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\zeta \in [0,1]$; also see (Li et al., 2024) for a similar method. These denoisers are designed following the *Tweedie's lemma* (Robbins, 1992) (see Appendix A). The construction of $\widetilde{\boldsymbol{x}}^{(k)}$ is meant to make the inputs to the score function closer to a certain $\mathcal{M}_{\sigma^{(k)}}$. Using score-based denoising, $h(\boldsymbol{z})$ is implicitly reflected in the denoising process and thus does not need to be specified analytically.

Another line of approaches explicitly construct regularizers $h(\boldsymbol{x})$ whose gradients correspond to applying the score function. Examples include RED-diff (Mardani et al., 2024) and SNORE (Renaud et al., 2024a). In SNORE, the regularizer is defined as $h(\boldsymbol{x}) = \mathbb{E}_{\widetilde{\boldsymbol{x}}|\boldsymbol{x}} \log p_\sigma(\widetilde{\boldsymbol{x}})$, $\widetilde{\boldsymbol{x}} \leftarrow \boldsymbol{x} + \sigma\boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and taking its gradient in (2) yields:

$$\textbf{SNORE update: } \boldsymbol{x}^{(k+1)} \leftarrow \boldsymbol{x}^{(k)} - \delta\nabla\ell(\boldsymbol{y}||\mathcal{A}(\boldsymbol{x}^{(k)})) - \eta(\widetilde{\boldsymbol{x}}^{(k)} - D_{\sigma_i}(\widetilde{\boldsymbol{x}}^{(k)})), \tag{5}$$

where, again, $D_\sigma(\widetilde{\boldsymbol{x}}) = \widetilde{\boldsymbol{x}} + \sigma^2 \boldsymbol{s}_\theta(\widetilde{\boldsymbol{x}}, \sigma)$ by the Tweedie's lemma.

Challenges — Manifold Mismatch and Convergence. Score-based PnP methods face two main challenges. First, the score is not trained on algorithm-induced iterates (e.g., $\boldsymbol{x}^{(k+1)}$ in (3)). While both $\boldsymbol{x}_t$ and $\boldsymbol{x}^{(k)}$ can be seen as noisy versions of $\boldsymbol{x} \sim p_{\text{data}}$, $\boldsymbol{x}_t$ follows Gaussian noise whereas the distribution of $\boldsymbol{x}^{(k)}$ is unclear. Many works attempt to bridge this gap by injecting Gaussian noise before applying the score function (cf. $\widetilde{\boldsymbol{x}}^{(k+1)}$ in (3), $\widetilde{\boldsymbol{x}}^{(k)}$ in (5)), or by purification-based schemes (Nie et al., 2022; Alkhouri et al., 2023; Meng et al., 2022). Yet noise injection alone is insufficient, and overfitting to measurement noise remains an issue (Wang et al., 2024). Second, the understanding to convergence of score-based PnP remains limited. Unlike classical denoisers with established theory, the geometry mismatch above makes it unclear whether iterates stabilize or under what conditions convergence can be ensured. Existing analyses mostly cover primal algorithms (see, e.g., (Renaud et al., 2024a)). Primal–dual methods such as ADMM offer greater flexibility for handling multiple regularizers and constraints, but their convergence with score-based denoisers remains unclear, as dual variables further complicate the manifold geometry of the iterates.

3 Proposed Approach

In this section, we propose our score-based denoiser and embed it in the ADMM framework. While we focus on ADMM due to its flexibility, the denoiser can be plugged into any other proximal operator based schemes (e.g. proximal gradient or variable-splitting as in DiffPIR).

Preliminaries of ADMM-PnP. ADMM-based inverse problem solvers start by rewriting (2) as

$$\min_{\boldsymbol{x},\boldsymbol{z}} \ell(\boldsymbol{y}||\mathcal{A}(\boldsymbol{x})) + \gamma h(\boldsymbol{z}) \text{ s.t. } \boldsymbol{x} = \boldsymbol{z} \tag{6}$$

The augmented Lagrangian of (6) is given by $L_\rho(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}; \boldsymbol{y}) = \ell(\boldsymbol{y}||\mathcal{A}(\boldsymbol{x})) + \gamma h(\boldsymbol{z}) + \boldsymbol{\lambda}^T(\boldsymbol{x}-\boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2$, where $\boldsymbol{\lambda}$ is the dual

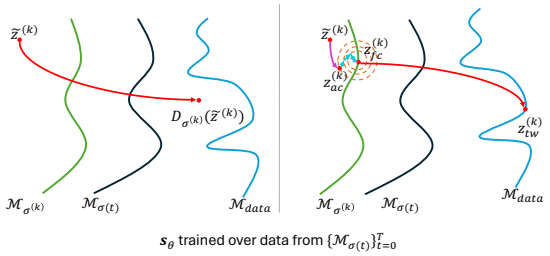

Figure 1: Left: direct denoising of $\widetilde{\boldsymbol{z}}^{(k)}$ using score functions could lead to unnatural recovered signals with artifacts. Right: AC-DC denoising brings $\widetilde{\boldsymbol{z}}^{(k)}$ closer to $\mathcal{M}_{\sigma^{(k)}}$, and then uses the score function to bring $\widetilde{\boldsymbol{z}}^{(k)}$ to the data manifold $\mathcal{M}_{\text{data}}$.

Algorithm 1 AC-DC Denoiser at iteration k of ADMM in (7)

1: **auto correction (AC):** $\boldsymbol{z}_{\text{ac}}^{(k)} \leftarrow \widetilde{\boldsymbol{z}}^{(k)} + \sigma^{(k)}\boldsymbol{n},\ \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
2: **directional correction (DC):**
3: $\boldsymbol{w}^{(0)} \leftarrow \boldsymbol{z}_{\text{ac}}^{(k)}$
4: **for** $j = 0$ to $J-1$ **do**
5: $\quad \boldsymbol{w}^{(k,j+1)} \leftarrow \boldsymbol{w}^{(k,j)} + \eta^{(k)} \left({}^{1}/_{\sigma^2_{\boldsymbol{s}^{(k)}}} (\boldsymbol{z}_{\text{ac}}^{(k)} - \boldsymbol{w}^{(k,j)}) + \boldsymbol{s}_\theta(\boldsymbol{w}^{(k,j)}, \sigma^{(k)}) \right) + \sqrt{2\eta^{(k)}}\boldsymbol{n},$
6: **end for**
7: $\boldsymbol{z}_{\text{dc}}^{(k)} \leftarrow \boldsymbol{w}^{(J)}$
8: **Denoising**: $\boldsymbol{z}_{\text{tw}}^{(k)} \leftarrow \mathbb{E}[\boldsymbol{z}_0 | \boldsymbol{z}_t = \boldsymbol{z}_{\text{dc}}^{(k)}] = \boldsymbol{z}_{\text{dc}}^{(k)} + (\sigma^{(k)})^2 \boldsymbol{s}_\theta(\boldsymbol{z}_{\text{dc}}^{(k)}, \sigma^{(k)})$ (Tweedie Denoising)
(Alternative: $\boldsymbol{z}_{\text{ode}}^{(k)} \leftarrow z_0$ by solving: $\frac{d\boldsymbol{z}_t}{dt} = \lambda(t) s_\theta(\boldsymbol{z}_t, t)$, with $\boldsymbol{z}_{\sigma^{(k)}} = \boldsymbol{z}_{\text{dc}}^{(k)}$) (ODE Denoising)

variable and $\rho > 0$ is the penalty parameter.
At iteration k, ADMM updates are as follows:

$$\boldsymbol{x}^{(k+1)} = \arg\min_{\boldsymbol{x}} \frac{1}{\rho}\ell(\boldsymbol{y}||\mathcal{A}(\boldsymbol{x})) + \frac{1}{2}\left\|\boldsymbol{x} - \boldsymbol{z}^{(k)} + \boldsymbol{u}^{(k)}\right\|_2^2 \tag{7a}$$

$$\boldsymbol{z}^{(k+1)} = \arg\min_{\boldsymbol{z}} \frac{\gamma}{\rho}h(\boldsymbol{z}) + \frac{1}{2}\left\|\boldsymbol{x}^{(k+1)} - \boldsymbol{z} + \boldsymbol{u}^{(k)}\right\|_2^2 \tag{7b}$$

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + (\boldsymbol{x}^{(k+1)} - \boldsymbol{z}^{(k+1)}) \tag{7c}$$

where $\boldsymbol{u}^{(k)} = \boldsymbol{\lambda}^{(k)}/\rho$ is the scaled dual variable. The subproblem (7b) is a denoising problem, and thus can be replaced by

$$\boldsymbol{z}^{(k+1)} = \text{Prox}_{\frac{\gamma}{\rho}h}(\boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)}) = D_{\sigma^{(k)}}(\widetilde{\boldsymbol{z}}^{(k)}). \tag{8}$$

The above is the classical ADMM-PnP method; see (Chan et al., 2016; Ryu et al., 2019). Same as before, Eq. (8) can be replaced by the score-based denoising following the Tweedie's lemma. However, as $\widetilde{\boldsymbol{z}}^{(k)} = \boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)}$ could be in any of the manifolds $\mathcal{M}_{\sigma(t)}$ on which the score function was trained, such naive replacement does not ensure effective denoising. The existence of the dual variable $\boldsymbol{u}^{(k)}$ makes the noise distribution in $\widetilde{\boldsymbol{z}}^{(k)}$ even harder to understand.

Proposed Approach: The AC-DC Denoiser. To address the manifold mismatch issues, we propose a three-stage denoiser. To be specific, in the kth iteration of the ADMM algorithm, we use the denoising process shown in Algorithm 1 . Note that the Tweedie's lemma step (line 8) can also be substituted by a score ODE based process (Karras et al., 2022) initialized at $\boldsymbol{z}_{\text{dc}}^{(k)}$. Our algorithm using these two different denoisers will be referred to as `Ours-tweedie` and `Ours-ode`, respectively.

The rationale of the AC-DC denoiser is illustrated in Fig.1. Recall that the score function is most effective on the noisy data manifolds $\{\mathcal{M}_{\sigma(t)}\}_{t=1}^{T}$, as it is trained over them. Since ADMM-induced iterates $\widetilde{\boldsymbol{z}}^{(k)}$ need not lie on these manifolds, directly applying score-based denoising may be ineffective. The AC step addresses this by adding Gaussian noise, making $\boldsymbol{z}_{\text{ac}}^{(k)}$ closer to some $\mathcal{M}_{\sigma(t)}$ (see AppendixB). This idea is related to the "purification" step in (Nie et al., 2022; Alkhouri et al., 2023; Li et al., 2024) and noise-added denoising in (Mardani et al., 2024; Renaud et al., 2024a; Zhu et al., 2023) (cf. (5) and (3)). However, AC alone does not guarantee manifold alignment. The proposed DC step, based on Langevin dynamics, further refines $\boldsymbol{z}_{\text{dc}}^{(k)}$ toward $\mathcal{M}_{\sigma^{(k)}}$.

To see the idea, let us break down the three steps. First, the AC step gives

$$\boldsymbol{z}_{\text{ac}}^{(k)} = \boldsymbol{z}_{\sigma^{(k)}} + \tilde{\boldsymbol{s}}^{(k)}, \quad \boldsymbol{s}^{(k)} = \widetilde{\boldsymbol{z}}^{(k)} - \boldsymbol{z}_{\natural}^{(k)},\ \boldsymbol{z}_{\natural}^{(k)} \sim p_{\text{data}} \tag{9}$$

where $\boldsymbol{z}_{\natural}^{(k)}$ is denoised signal of $\widetilde{\boldsymbol{z}}^{(k)}$, $\boldsymbol{z}_{\sigma^{(k)}} = \boldsymbol{z}_{\natural}^{(k)} + \sigma^{(k)}\boldsymbol{n}_1$, $\tilde{\boldsymbol{s}}^{(k)} = \sqrt{2}\sigma^{(k)}\boldsymbol{n}_2 + \boldsymbol{s}^{(k)}$, and $\boldsymbol{n}_1, \boldsymbol{n}_2 \overset{\text{i.i.d}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$. Given a sufficiently large $\sigma^{(k)}$, $\boldsymbol{z}_{\text{ac}}^{(k)}$ would have dominated by Gaussian noise—but not necessarily on any of $\mathcal{M}_{\sigma_t}$ where the score was trained. Starting from $\boldsymbol{z}_{\text{ac}}^{(k)}$, the DC step runs a few iterations of *Langevin dynamics* targeting the distribution $p(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_{\text{ac}}^{(k)})$. This is because $\text{supp}(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_{\text{ac}}^{(k)}) \subseteq \text{supp}(\boldsymbol{z}_{\sigma^{(k)}}) = \mathcal{M}_{\sigma^{(k)}}$. In addition, $p(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_{\text{ac}}^{(k)})$ at the same time retains the

information of $\boldsymbol{z}_{\text{ac}}^{(k)}$ (thereby the information from the measurements). Assume that the forward process used in training the score has sufficiently small time intervals, $\mathcal{M}_{\sigma^{(k)}}$ is approximately contained in $\{\mathcal{M}_{\sigma_t}\}_{t=1}^{T}$. This way, when applying Tweedie's lemma for denoising, the step is expected to be effective, as the score was trained over $\{\mathcal{M}_{\sigma_t}\}_{t=1}^{T}$.

Note that the conditional score for the Langevin dynamics step can be expressed as follows

$$\nabla \log p(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_{\text{ac}}^{(k)}) = \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{z}_{\sigma^{(k)}}, \sigma^{(k)}) + \nabla \log p(\boldsymbol{z}_{\text{ac}}^{(k)}|\boldsymbol{z}_{\sigma^{(k)}}). \tag{10}$$

Ideally, one would use the exact $\nabla \log p(\boldsymbol{z}_{\text{ac}}^{(k)}|\boldsymbol{z}_{\sigma^{(k)}})$ for the DC step—which is unavailable. In practice, we approximate $p(\boldsymbol{z}_{\text{ac}}^{(k)}|\boldsymbol{z}_{\sigma^{(k)}})$ using a Gaussian distribution. Note that under proper scheduling of $\sigma^{(k)}$ and mild regularity conditions on $\boldsymbol{s}^{(k)}$, e.g., when $\text{Var}(\boldsymbol{s}^{(k)})^{1/2} \ll \sigma^{(k)}$, the likelihood can be well-approximated by a locally quadratic form, leading to $\nabla \log p(\boldsymbol{z}_{\text{ac}}^{(k)}|\boldsymbol{z}_{\sigma^{(k)}}) \approx -1/\sigma^2_{\boldsymbol{s}^{(k)}}(\boldsymbol{z}_{\sigma^{(k)}} - \boldsymbol{z}_{\text{ac}}^{(k)})$ and the DC step in Algorithm 1.

4 Convergence Analysis

4.1 Convergence of Under Weakly Non-Expansive Residuals

Following the established convention in ADMM-PnP, e.g., (Buzzard et al., 2018; Sun et al., 2019; Chan, 2019; Teodoro et al., 2017), we aim at understanding the convergence properties when the AC-DC denoiser is used. We will use the following definitions:

Definition 1 (Fixed point convergence)**.** *Let $T : \mathcal{X} \to \mathcal{X}$ be the update map of an iterative algorithm, and let $\boldsymbol{x}^{(0)} \in \mathcal{X}$ be arbitrary initialization. The algorithm is said to converge to a fixed point $\boldsymbol{x}^*$ if for any $\delta > 0$ there exists $K_\delta > 0$ such that the sequence generated by the algorithm $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}^+}$ satisfies $\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\right\|_2 < \delta$ for all $k \geq K_\delta$. Equivalently, $\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}^*$ with $T(\boldsymbol{x}^*) = \boldsymbol{x}^*$.*

Definition 2 (Sequence convergence to a δ-ball)**.** *For a certain $\delta > 0$, a sequence $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}^+}$ is said to converge within a δ-ball if there exists $K > 0$ and $\boldsymbol{x}^*$ such that the following holds for all $k \geq K$.*

$$\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\right\|_2 \leq \delta \tag{11}$$

Comparing the two definitions, Definition 2 is a weaker statement; that is, even when $k \to \infty$, $\boldsymbol{x}^{(k)} \to \boldsymbol{x}^*$ does not necessarily happen. Nonetheless, convergence to a δ-ball is still meaningful. The notion of δ-ball convergence is often used in numerical analysis for stability characterization; see, e.g., Ren & Argyros (2021); Ren & Wu (2009); Liang (2007).

Definition 3 (ADMM convergence to a δ-ball)**.** *ADMM is said to converge within a δ ball if the sequences $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}^+}$ and $\{\boldsymbol{u}^{(k)}\}_{k\in\mathbb{N}^+}$ obtained from ADMM converges within a δ-ball.*

To proceed, consider the following assumption:

Assumption 1. *For a certain $\delta > 0$ there exists $\epsilon \leq 1$ such that for all $\widetilde{\boldsymbol{z}}_1, \widetilde{\boldsymbol{z}}_2 \in \mathbb{R}^d$, the following holds:*

$$\|R_\sigma(\widetilde{\boldsymbol{z}}_1) - R_\sigma(\widetilde{\boldsymbol{z}}_2)\|_2^2 \leq \epsilon^2 \|\widetilde{\boldsymbol{z}}_1 - \widetilde{\boldsymbol{z}}_2\|_2^2 + \delta^2 \tag{12}$$

where $R_\sigma(\widetilde{\boldsymbol{z}}) = (D_{\sigma^{(k)}} - I)(\widetilde{\boldsymbol{z}})$ with I being the identity function (i.e., $I(\boldsymbol{z}) = \boldsymbol{z}$).

Here, the notation $(D_{\sigma^{(k)}} - I)(\widetilde{\boldsymbol{z}})$ denotes the residual of $D_{\sigma^{(k)}}$ i.e. $D_{\sigma^{(k)}}(\widetilde{\boldsymbol{z}}) - \widetilde{\boldsymbol{z}}$. The next theorem extends the fixed point convergence of ADMM-PnP in Ryu et al. (2019). Unlike Ryu et al. (2019) where R_σ needs to be strictly contractive, our result allows R_σ to be weakly contractive:

Theorem 1. *Under Assumption 1, assume that ℓ is μ-strongly convex. Then, there exists $\boldsymbol{x}^*$, $\boldsymbol{u}^*$ and $K > 0$ such that the sequences $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}^+}$ and $\{\boldsymbol{u}^{(k)}\}_{k\in\mathbb{N}^+}$ generated by ADMM-PnP using a fixed step size ρ satisfies $\left\|\boldsymbol{u}^{(k)} - \boldsymbol{u}^*\right\|_2 \leq r$ and $\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\right\|_2 \leq r$ with $r = (1 + \frac{\rho}{\rho+\mu})\bar{\delta}/\sqrt{1-\bar{\epsilon}^2}$ for all $k \geq K$ when $\epsilon/\mu(1+\epsilon-2\epsilon^2) < 1/\rho$ where $\bar{\delta}^2 = \frac{\delta^2\bar{\epsilon}}{\epsilon}$ and $\bar{\epsilon} = \frac{\rho+\rho\epsilon+\mu\epsilon+2\mu\epsilon^2}{\rho+\mu+2\mu\epsilon}$.*

The proof is relegated to Appendix C. Note that when $\delta = 0$, it implies the result in Ryu et al. (2019).

4.2 Convergence Under Weakly Non-Expansiveness with AC-DC

In this subsection, we will show that the AC-DC denoiser satisfies Assumption 1 under mild conditions. To this end, consider the following:

Assumption 2 (Smoothness of $\log p_{\text{data}}$)**.** *The log data density $\log p_{\text{data}}$ is M-smooth for a constant $M > 0$, i.e., $\|\nabla \log p_{\text{data}}(\boldsymbol{x}) - \nabla \log p_{\text{data}}(\boldsymbol{y})\|_2 \leq M\|\boldsymbol{x} - \boldsymbol{y}\|_2$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$.*

Assumption 3 (Coercivity for $-\log p_{\text{data}}$)**.** *There exists constants $c_1 > 0$ and $c_2 \geq 0$ such that*

$$\|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_2^2 \geq -c_1 \log p_{\text{data}} - c_2,\ \|\boldsymbol{x}\|_2 \leq -c_1 \log p_{\text{data}}(\boldsymbol{x}) + c_2,\ \forall \boldsymbol{x} \in \mathcal{X} \tag{13}$$

This coercivity assumption means the negative log-density grows sufficiently fast at infinity, which prevents the Langevin dynamics from "escaping to infinity". This assumption guarantees stability and ensure ergodicity leading to convergence to the stationary distribution (Mattingly et al., 2002).

Theorem 2. *Suppose that the assumptions in Theorem 1, Assumption 2 and Assumption 3 hold. Further, assume that the DC step reaches the stationary distribution for each k. Let $D_{\sigma^{(k)}} : \widetilde{\boldsymbol{z}}^{(k)} \mapsto \boldsymbol{z}_{\text{tw}}^{(k)}$ denote the AC-DC denoiser. Then, we have:*

(a) With probability at least $1 - 2e^{-\nu_k}$, the following holds for iteration k of ADMM-PnP:

$$\|(D_{\sigma^{(k)}} - I)(\boldsymbol{x}) - (D_{\sigma^{(k)}} - I)(\boldsymbol{y})\|_2^2 \leq \epsilon_k^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \delta_k^2 \tag{14}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ and $k \in \mathbb{N}^+$ when $\sigma_{\boldsymbol{s}^{(k)}}^2 + (\sigma^{(k)})^2 < 1/M$ with

$$\epsilon_k^2 = 3\big((\sqrt{2}M\sigma_{\boldsymbol{s}^{(k)}}^2 / 1 - \sigma_{\boldsymbol{s}^{(k)}}^2 M)^2 + (\sigma^{(k)})^4 M^2\big) \tag{15}$$

$$\delta_k^2 = 3\big(2(\sigma^{(k)})^2(d + 2\sqrt{d\nu_k} + 2\nu_k) + {}^{32d\sigma_{\boldsymbol{s}^{(k)}}^2}/_{(1 - M\sigma_{\boldsymbol{s}^{(k)}}^2)} \log {}^2/_{\nu_k}\big). \tag{16}$$

In other words, with $\nu_k = \ln {}^{2\pi}/_{6\eta} + 2\ln k$, the denoiser $D_{\sigma^{(k)}}$ satisfies part (a) for all $k \in \mathbb{N}^+$ with probability at least $1 - \eta$.

(b) Assume that $\sigma^{(k)}$ is scheduled such that $\lim_{k\to\infty}(\sigma^{(k)})^2\nu_k = 0$ for $\nu_k = \ln {}^{2\pi}/_{6\eta} + 2\ln k$, $\epsilon < 1$, and ${}^{\epsilon}/_{\mu(1+\epsilon-2\epsilon^2)} < {}^1/_{\rho}$ all hold, where $\epsilon = \lim_{k\to\infty} \sup \epsilon_k$ with ϵ_k defined in (15). Consequently, $\delta = \lim_{k\to\infty} \sup \delta_k$ is finite and ADMM-PnP with the AC-DC denoiser converges to an r-ball (see r in Theorem 1) with probability at least $1 - \eta$.

The proof is relegated to Appendix D. Theorem 2 (a) establishes that the AD-DC denoiser is weakly non-expansive with probability $1 - 2e^{-\nu_k}$ in iteration k. The (b) part states that when $\sigma^{(k)}$ is carefully scheduled to approach zero as k grows, then, with high probability, all the iterations satisfy the weakly non-expansiveness together—this leads to the convergence of the ADMM-PnP algorithm.

4.3 Convergence without Convexity of ℓ

The weakly non-expansiveness based convergence analysis holds under fixed step size (i.e., ρ) of the ADMM-PnP algorithm, which is consistent with practical implementations in many cases. However, the assumption that the ℓ term is μ-strongly convex is only met by some inverse problems, e.g., signal denoising and deblurring, but not met by others such as signal compression/recovery and data completion. In this subsection, we remove the convexity assumption and analyze the AC-DC denoiser's properties under the adaptive ρ-scheme following that in (Chan et al., 2016).

Theorem 3. *Suppose that Assumptions 2-3 hold. Let $D := \operatorname{diam}(\mathcal{X}) = \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{X}} \|\boldsymbol{x} - \boldsymbol{y}\|_2 < \infty$, $S := \inf_{\boldsymbol{x}\in\mathcal{X}} \|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_2 < \infty$ and define $L := MD + S$. Let $D_{\sigma^{(k)}} : \widetilde{\boldsymbol{z}}^{(k)} \mapsto \boldsymbol{z}_{\text{tw}}^{(k)}$ denote the AC-DC denoiser. Also, assume that the DC step reaches the stationary distribution for each k[1]. Then, the following hold:*

*(a) (**Boundedness**) With probability at least $1 - 2e^{-\nu_k}$, the denoiser $D_{\sigma^{(k)}}$ is bounded at each iteration k i.e. $\frac{1}{d}\|(D_{\sigma^{(k)}} - I)(\boldsymbol{x})\|_2^2 \leq c_k^2$ whenever $\sigma_{\boldsymbol{s}^{(k)}}^2 + (\sigma^{(k)})^2 < 1/M$, where $c_k = (\sigma^{(k)})^2(2 + 4\sqrt{\nu_k} + 4\nu_k) + {}^{16\sigma_{\boldsymbol{s}^{(k)}}^2}/_{1 - M\sigma_{\boldsymbol{s}^{(k)}}^2} \log {}^2/_{\nu_k} + 2\sigma_{\boldsymbol{s}^{(k)}}^4 L^2 + 2(\sigma^{(k)})^4 L^2$, and $\nu_k > 0$.*

[1] Note that Theorems 2 and 3 use this stationary distribution assumption for notation conciseness. For their counterparts removing this assumption, see Appendix E.2.

Let $\nu_k = \ln \frac{2\pi^2}{6\eta} + 2\ln k$ *with* $\eta \in (0, 1]$. *Consequently, the denoiser* $D_{\sigma^{(k)}}$ *is bounded for all* $k \in \mathbb{N}_+$ *with corresponding* c_k *and probability at least* $1 - \eta$.

*(b) (**Convergence**) Assume there exists* $R < \infty$ *such that* $\|\nabla \ell(\boldsymbol{x})\|_2/\sqrt{d} \leq R$. *Apply the* ρ*-increasing rule in (Chan et al., 2016) and schedule* $(\sigma^{(k)}, \sigma_{\boldsymbol{s}^{(k)}})$ *such that* $\lim_{k\to\infty}(\sigma^{(k)})^2(2 + 4\sqrt{\nu_k} + 4\nu_k) = 0$, $\lim_{k\to\infty} \frac{\sigma^2_{\boldsymbol{s}^{(k)}}}{1-M\sigma^2_{\boldsymbol{s}^{(k)}}} \log \frac{2}{\nu_k} = 0$, $\lim_{k\to\infty} \sigma^{(k)} = 0, \lim_{k\to\infty} \sigma_{\boldsymbol{s}^{(k)}} = 0$ $\sigma^2_{\boldsymbol{s}^{(k)}} + (\sigma^{(k)})^2 < 1/M$, $\forall k \in \mathbb{N}_+$ *for* $\nu_k = \ln \frac{2\pi^2}{6\eta} + 2\ln k$ *with* $\eta \in (0, 1]$. *Then, the solution sequence converges to a fixed point with probability at least* $1 - \eta$.

The proof is relegated to Appendix E. Theorem 3 (a) shows that, with high probability the denoiser is bounded uniformly across all iterations k. Part (b) further shows that, under the proper scheduling of $(\sigma^{(k)}, \sigma_{\boldsymbol{s}^{(k)}})$, the AD-DC ADMM-PnP algoirthm converges to a fixed point with high probability.

The condition of $D < \infty$ implies the data space $\mathcal{X}$ has bounded support, which is natural in practice: for images, pixel intensities typically lie within a bounded range such as $[0, 1]$. Additionally, the condition $S < \infty$ ensures that there exists at least one point in $\mathcal{X}$ where the score norm is finite. This prevents pathological cases where the score diverges everywhere (making the distribution degenerate). Together, these conditions guarantee that the score is "well-behaved".

A remark is that all theoretical results in this section focus on fixed-point convergence, which is not the strongest form of convergence guarantees. Establishing stronger convergence results, e.g., stationary-point convergence, for PnP approaches is considered challenging as the objective function is implicit (more specifically, $h(\cdot)$ in (2) is implicit). Nonetheless, in recent years, some efforts have been made towards establishing stationary-point convergence for PnP methods under certain types of denoisers (see, e.g., (Hurault et al., 2022a;b; Wei et al., 2025; Xu et al., 2025)); more discussions are in Sec. 5.

5 Related Works

ADMM-PnP has gained much popularity, due to access to data-driven effective denoisers. It has been used in various applications like image restoration (Chan et al., 2016), data compression (Yuan et al., 2022), hyperspectral imaging (Liu et al., 2022), and medical imaging (Ahmad et al., 2020). Theoretical understanding of PnP algorithms with general "black-box" denoisers remains limited. Unlike classical proximal operators, the implicit regularizer $h(\cdot)$ in (2) handled by data-driven denoisers is typically unknown. Hence, many results are therefore restricted to fixed-point convergence; see, e.g., (Ryu et al., 2019; Chan et al., 2016). Nonetheless, in certain cases where the denoisers have interesting structures, stronger convergence results can be established. For example, (Xu et al., 2025) used classical results from image denoising connecting linear denoisers with quadratic $h(\cdot)$ to show that when linear denoisers are employed, ADMM-PnP converges to KKT points. Hurault et al. (2022a) showed stationary-point convergence of PnP methods gradient-type denoisers, leveraging the fact that this type of denoisers can be written as a proximal operator of a special function (see the nonconvex counterpart in (Hurault et al., 2022b)); Wei et al. (2025) trained denoisers to satisfy a cocoercive conservativity condition, which also ensures convergence of PnP to stationary points associated with an implicit convex $h(\cdot)$. Nonetheless, these results do not cover diffusion-based denoisers. In this work, we generalize the fixed-point convergence proofs in (Chan et al., 2016; Ryu et al., 2019) to accommodate the diffusion score-based AC–DC denoiser.

Recent advances in score-based generative modeling have motivated their integration into PnP algorithms. One line of work directly replaces the proximal denoiser with a pre-trained scores (Zhu et al., 2023; Li et al., 2024). Alternatively, others embed the score function as an explicit regularizer with task-specific loss (Mardani et al., 2024; Renaud et al., 2024a). Deterministic version PnP have also been considered. For example, Wang et al. (2024); Song et al. (2023) use unrolled ODE and consistency model-distilled one-step representation to express the target signal, respectively. These methods are similar to (Bora et al., 2017), but with diffusion-driven parameterization.

Prior works have emphasized the importance of matching the residual noise to the operating range of the PnP denoiser. D-AMP (Metzler et al., 2016; Eksioglu & Tanc, 2018) achieve this via the Onsager correction, which approximately Gaussianizes the residual under compressive sensing problem structures. Wei et al. (2021) learns a reinforcement learning-based policy to automatically tune all

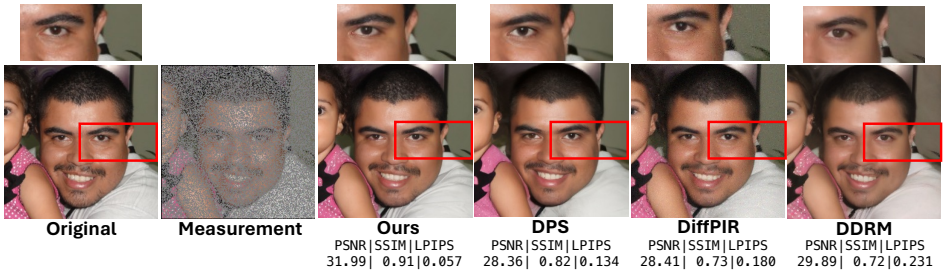

Figure 2: Inpainting under random missings.

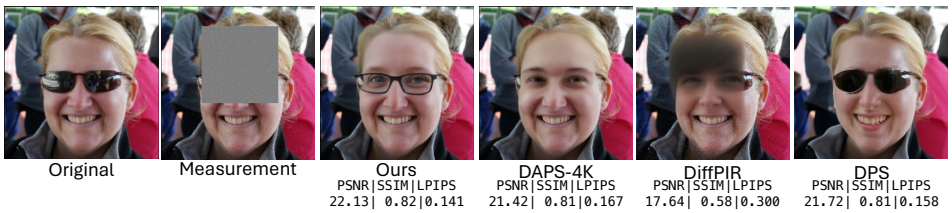

Figure 3: Inpainting under box missing.

internal parameters, including denoising strength. Unlike AC–DC that provides a generic correction mechanism for a variety inverse problems, these methods either problem specific or require training additional models. Score-based inverse problem solvers have also attempted to "bring" iterates to noisy data manifolds used during training. The work Chung et al. (2022) uses a manifold constraint based on gradient of data-fidelity, while He et al. (2024) uses an off-the-shelf pretrained neural network to impose a manifold constraint. On the other hand, Zirvi et al. (2025) uses the projection of measurement guidance to low-rank subspace, using SVD on the intermediate diffusion state, for similar purposes.

The idea of adding noise before evaluating score functions during optimization procedures (similar to our AC step) has been widely considered (Li et al., 2024; Graikos et al., 2022; Renaud et al., 2024b; Mardani et al., 2024). A variant of this called estimation-correction idea proposed in (Karras et al., 2022) is used in (Zhu et al., 2023) for this purpose.

6 Experiments

Dataset and Evaluation Metrics. For all these tasks, we use two datasets: FFHQ 256×256 (Karras et al., 2021) and ImageNet 256×256 (Deng et al., 2009). During testing, we randomly sample 100 images from the validation set of each dataset. All the methods use the pre-trained score model in Chung et al. (2023). We use *Peak Signal-to-Noise Ratio* (PSNR) as a pixel-wise similarity metric, and *Structural Similarity Index* (SSIM) and *Learned Perceptual Image Patch Similarity* (LPIPS) (Zhang et al., 2018) as perceptual similarity metrics. We report these metrics averaged over the 100 test images for each method and inverse problem.

Task Description. We consider $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \boldsymbol{I})$ with $\sigma_n = 0.05$ for all the tasks. (a) For *super-resolution*, we use cubic interpolation method with kernel size 4 for downsampling the resolution by 4 times. (b) For *recovery under Gaussian blurring* (Gaussian deblurring), a kernel of size 61 and standard deviation 3 is used. (c) As for *recovery under motion blurring* (motion deblurring), a kernel of size 61 and standard deviation of 0.5 is used. (d) In *inpainting under box mask* (box inpainting), an approximately centered mask of size 128×128 is sampled in image while maintaining the 32 pixel margin in both spatial dimensions of the input image. (e) For *inpainting under random missings* (random inpainting), 70% of the pixels are uniformly sampled to be masked, and a scaling of 2 was used in high dynamic range (HDR) before clipping the values. (f) For *phase retrieval*, similar as in prior works (Wu et al., 2024; Mardani et al., 2024), we use oversampling by factor of 2. g) For deblurring under nonlinear blurring, we use the operator in (Tran et al., 2021) with default settings.

Baselines. We use a set of baselines, namely, DPS (Chung et al., 2023), DAPS (Zhang et al., 2024), DDRM (Kawar et al., 2022), DiffPIR (Zhu et al., 2023), RED-diff (Mardani et al., 2024), DPIR (Zhang et al., 2022), DCDP (Li et al., 2025), PMC (Sun et al., 2024).

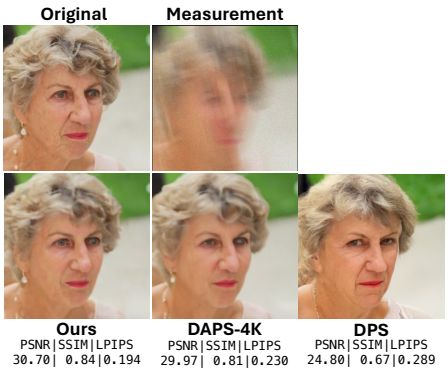

Figure 4: Recovery under motion blurring.

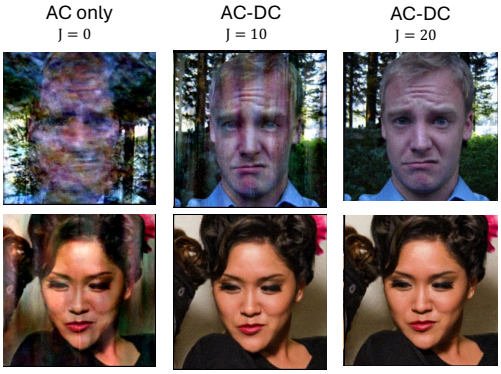

Figure 5: Influence of DC steps in the denoiser.

Table 1: Reconstruction metrics (100 images) on FFHQ / ImageNet. **Bold**: best, blue: 2nd best.

Task	Method	FFHQ PSNR↑	FFHQ SSIM↑	FFHQ LPIPS↓	ImageNet PSNR↑	ImageNet SSIM↑	ImageNet LPIPS↓
Superresolution (4x)	Ours–tweedie	**30.439**	**0.857**	0.178	**27.318**	**0.717**	0.280
	Ours–ode	29.991	0.845	**0.156**	26.919	0.700	0.276
	DAPS	29.529	0.814	0.167	26.653	0.680	**0.266**
	DPS	24.828	0.705	0.257	22.785	0.549	0.411
	DDRM	27.145	0.782	0.261	26.105	0.683	0.306
	DiffPIR	26.771	0.749	0.208	23.884	0.543	0.336
	RED-diff	16.833	0.422	0.547	18.662	0.309	0.519
	DPIR	28.849	0.826	0.254	26.524	0.699	0.334
	DCDP	27.761	0.639	0.332	24.517	0.525	0.361
	PMC	23.774	0.421	0.407	22.534	0.334	0.456
Inpainting (Random)	Ours–tweedie	**32.844**	**0.906**	0.122	**29.564**	**0.817**	0.184
	Ours–ode	32.127	0.894	**0.095**	28.733	0.795	**0.148**
	DAPS	31.652	0.847	0.124	28.137	0.751	0.162
	DPS	29.084	0.828	0.181	26.049	0.678	0.318
	DDRM	28.969	0.847	0.178	27.883	0.778	0.203
	DiffPIR	28.558	0.709	0.230	26.923	0.639	0.222
	RED-diff	20.361	0.630	0.275	20.948	0.464	0.315
	PMC	23.289	0.755	0.263	25.965	0.636	0.342
Motion Deblur	Ours–tweedie	**30.003**	**0.854**	0.179	**27.149**	**0.717**	0.280
	Ours–ode	29.648	0.841	**0.154**	26.615	0.694	**0.275**
	DAPS	29.051	0.815	0.175	26.571	0.689	0.276
	DPS	23.257	0.663	0.265	19.613	0.451	0.451
	PMC	19.480	0.590	0.426	21.608	0.480	0.510

Task	Method	FFHQ PSNR↑	FFHQ SSIM↑	FFHQ LPIPS↓	ImageNet PSNR↑	ImageNet SSIM↑	ImageNet LPIPS↓
Gaussian Blur	Ours-tweedie	**30.402**	**0.853**	0.175	**27.199**	**0.705**	0.281
	Ours-ode	30.019	0.841	0.158	26.899	0.690	0.282
	DAPS	29.790	0.813	**0.157**	26.886	0.678	**0.260**
	DPS	26.106	0.730	0.207	23.995	0.575	0.328
	DiffPIR	25.148	0.699	0.230	22.756	0.508	0.374
	DPIR	28.875	0.833	0.228	26.702	0.700	0.314
	DCDP	16.821	0.171	0.642	15.102	0.136	0.620
	PMC	20.172	0.638	0.344	24.103	0.545	0.415
Inpainting (Box)	Ours-tweedie	24.025	**0.859**	**0.131**	**21.626**	**0.789**	0.222
	Ours-ode	23.342	0.837	0.136	20.618	0.743	0.227
	DAPS	23.643	0.815	0.146	21.303	0.774	0.199
	DPS	23.488	0.817	0.164	19.933	0.677	0.309
	DiffPIR	20.934	0.561	0.294	19.565	0.562	0.342
	RED-diff	18.713	0.523	0.364	18.075	0.499	0.371
	DCDP	**25.230**	0.754	0.163	20.991	0.727	**0.195**
	PMC	14.828	0.697	0.318	15.550	0.666	0.326
Phase Retrieval	Ours-tweedie	**27.944**	**0.793**	**0.209**	**17.770**	**0.440**	**0.471**
	Ours-ode	27.095	0.757	0.237	16.013	0.339	0.539
	DAPS	26.707	0.749	0.230	16.444	0.395	0.512
	DPS	11.627	0.366	0.658	9.434	0.216	0.768
	RED-diff	15.411	0.490	0.480	12.852	0.204	0.695
	DCDP	20.026	0.540	0.424	12.257	0.212	0.665
	PMC	10.421	0.287	0.783	8.636	0.129	0.890

Qualitative performance. Figs 2, 3 and 4 show reconstructions under inpainting under random missings, inpainting under box missing, and motion deblurring. It can be seen that our method is able to recover the image that is comparatively natural looking with less noise and artifacts, while being consistent with the measurements. On the other hand, images recovered with DiffPIR appears to suffer from noise and artifacts, whereas DPS leads to measurement-inconsistent reconstructions.

Our method outperforms others while other methods appear to either be blurred or contain noisy artifacts in the recovered image. Recovery by DPS is less consistent with the original image; the pattern on the child's clothing is completely lost.

Quantitative performance. Table 1 summarizes PSNR, SSIM and LPIPS averaged over 100 images on FFHQ and Imagenet datasets. In almost all of the inverse problems, both of our variants (Ours-tweedie and Ours-ode) achieve the best or second-best performance in terms of all metrics. Our method significantly outperforms other PnP baseline methods considered, namely, DDRM, DiffPIR and RED-diff. This demonstrates the effectiveness of our AC-DC denoiser.

Effectiveness of DC. To perform albation study on the DC stage, we consider the challenging phase retrieval problem. Fig. 5 shows the output of ADMM-PnP with our AC-DC denoiser with different numbers of DC iterations J. With $J = 0$ (disabling DC step), artifacts remain severe. Increasing J progressively results in cleaner images.

More Details and Additional Experiments. More details and experiments are in appendices.

7 CONCLUSION

We introduced the AC-DC denoiser, a score-based denoiser designed for integration within the ADMM-PnP framework. The denoiser adopts a three-stage structure aimed at mitigating the mismatch between ADMM iterates and the noisy manifolds on which score functions are trained. We established convergence guarantees for ADMM-PnP with the AC-DC denoiser under both fixed and adaptive step size schedules. Empirical results across a range of inverse problems demonstrate that the proposed method consistently improves solution quality over existing baselines.

Limitations. While our analysis provides initial insights, several aspects merit deeper understanding. The second convergence result relaxes convexity by allowing adaptive step sizes, though such schedules are arguably less appealing in practice. Our experiments, however, suggest that constant step sizes also perform well for nonconvex objectives; it is therefore desirable to establish convergence guarantees for constant step sizes in such settings. In addition, our result ensures the *stability* of the ADMM method, but does not directly explain the reason *why* the AC–DC denoiser attains high-quality recovery; recoverability and estimation error analyses are also desirable. On the implementation side, the noise schedules used in the AC and DC stages are currently guided by empirical heuristics. Designing problem-adaptive scheduling strategies may further improve both convergence speed and robustness. Additionally, each iteration of AC–DC denoiser needs multiple score evaluations. Reducing the required NFEs could significantly improve its efficiency.

Ethics Statement: This work focuses exclusively on the theory and methodology of solving inverse problems. It does not involve human subjects, personal data, or any sensitive procedures.

Reproducibility Statement: The source code is provided as a part of the supplementary material. All assumptions, derivations and necessary details regarding the theory and experiments are included in the appendices.

REFERENCES

Rizwan Ahmad, Charles A. Bouman, Gregery T. Buzzard, Stanley Chan, Sizhuo Liu, Edward T. Reehorst, and Philip Schniter. Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery. *IEEE Signal Processing Magazine*, 37(1):105–116, January 2020. ISSN 1558-0792. doi: 10.1109/msp.2019.2949470. URL http://dx.doi.org/10.1109/MSP.2019.2949470.

Ismail Alkhouri, Shijun Liang, Rongrong Wang, Qing Qu, and Saiprasad Ravishankar. Robust physics-based deep mri reconstruction via diffusion purification, 2023. URL https://arxiv.org/abs/2309.05794.

Ismail Alkhouri, Shijun Liang, Evan Bell, Qing Qu, Rongrong Wang, and Saiprasad Ravishankar. Image reconstruction via autoencoding sequential deep image prior. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=K1EG2ABzNE.

Simon R Arridge. Optical tomography in medical imaging. *Inverse problems*, 15(2):R41, 1999.

Heinz H Bauschke, Patrick L Combettes, Heinz H Bauschke, and Patrick L Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2017.

Andrew F Bennett. *Inverse methods in physical oceanography*. Cambridge university press, 1992.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International conference on machine learning*, pp. 537–546. PMLR, 2017.

Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Herm Jan Brascamp and Elliott H Lieb. On extensions of the brunn-minkowski and prékopa-leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, 1976. ISSN 0022-1236. doi: https://doi.org/10.1016/0022-1236(76)90004-5. URL https://www.sciencedirect.com/science/article/pii/0022123676900045.

Felix E. Browder. Nonexpansive nonlinear operators in a banach space. *Proceedings of the National Academy of Sciences of the United States of America*, 54(4):1041–1044, 1965. ISSN 00278424, 10916490. URL http://www.jstor.org/stable/73047.

Gregery T Buzzard, Stanley H Chan, Suhas Sreehari, and Charles A Bouman. Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium. *SIAM Journal on Imaging Sciences*, 11(3):2001–2020, 2018.

Stanley Chan, Xiran Wang, and Omar Elgendy. Plug-and-play admm for image restoration: Fixed point convergence and applications. *IEEE Transactions on Computational Imaging*, PP, 05 2016. doi: 10.1109/TCI.2016.2629286.

Stanley H. Chan. Performance analysis of plug-and-play admm: A graph signal processing perspective. (arXiv:1809.00020), May 2019. doi: 10.48550/arXiv.1809.00020. URL http://arxiv.org/abs/1809.00020. arXiv:1809.00020 [eess].

Xi Chen, Simon S. Du, and Xin T. Tong. On stationary-point hitting time and ergodicity of stochastic gradient langevin dynamics. *Journal of Machine Learning Research*, 21(68):1–41, 2020. URL http://jmlr.org/papers/v21/19-327.html.

Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *Advances in Neural Information Processing Systems*, 35:25683–25696, 2022.

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

B Combal, Frédéric Baret, M Weiss, Alain Trubuil, D Macé, A Pragnere, R Myneni, Y Knyazikhin, and L Wang. Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote sensing of environment*, 84(1):1–15, 2003.

Patrick L. Combettes and Isao Yamada. Compositions and convex combinations of averaged nonexpansive operators. *Journal of Mathematical Analysis and Applications*, 425(1):55–70, 2015. ISSN 0022-247X. doi: https://doi.org/10.1016/j.jmaa.2014.11.044. URL https://www.sciencedirect.com/science/article/pii/S0022247X14010865.

Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16 (8):2080–2095, 2007. doi: 10.1109/TIP.2007.901238.

Arnak S. Dalalyan and Avetik G. Karagulyan. User-friendly guarantees for the langevin monte carlo with inaccurate gradient. *Stochastic Processes and their Applications*, 129(12):5278–5311, December 2019. ISSN 03044149. doi: 10.1016/j.spa.2019.02.016. arXiv:1710.00095 [math].

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

Alex Dytso, H. Vincent Poor, and Shlomo Shamai. A general derivative identity for the conditional mean estimator in gaussian noise and some applications, 2021. URL https://arxiv.org/abs/2104.01883.

Jonathan Eckstein and Dimitri P Bertsekas. On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical programming*, 55: 293–318, 1992.

Ender M. Eksioglu and A. Korhan Tanc. Denoising amp for mri reconstruction: Bm3d-amp-mri. *SIAM Journal on Imaging Sciences*, 11(3):2090–2109, January 2018. doi: 10.1137/18M1169655.

Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. doi: 10.1109/TIP.2006.881969.

Dara Entekhabi, Hajime Nakamura, and Eni G Njoku. Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2):438–448, 1994.

Pontus Giselsson. Tight global linear convergence rate bounds for douglas–rachford splitting. *Journal of Fixed Point Theory and Applications*, 19:2241 – 2270, 2015. URL https://api.semanticscholar.org/CorpusID:56162917.

Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras. Diffusion models as plug-and-play priors. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=yhlMZ3iR7Pu.

C. Hatsell and L. Nolte. Some geometric properties of the likelihood ratio (corresp.). *IEEE Transactions on Information Theory*, 17(5):616–618, 1971. doi: 10.1109/TIT.1971.1054672.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J Zico Kolter, Ruslan Salakhutdinov, and Stefano Ermon. Manifold preserving guided diffusion. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=o3BxOLoxm1.

Wenrui Hu, Dacheng Tao, Wensheng Zhang, Yuan Xie, and Yehui Yang. The twist tensor nuclear norm for video completion. *IEEE Transactions on Neural Networks and Learning Systems*, 28(12): 2961–2973, 2017. doi: 10.1109/TNNLS.2016.2611525.

Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Gradient step denoiser for convergent plug-and-play. (arXiv:2110.03220), February 2022a. doi: 10.48550/arXiv.2110.03220. URL http://arxiv.org/abs/2110.03220. arXiv:2110.03220 [cs].

Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 9483–9505. PMLR, 2022b. URL https://proceedings.mlr.press/v162/hurault22a.html.

Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing*, 26(9): 4509–4522, 2017.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(12):4217–4228, December 2021. ISSN 0162-8828. doi: 10.1109/TPAMI.2020.2970919. URL https://doi.org/10.1109/TPAMI.2020.2970919.

Tero Karras, Miika Aittala, Samuli Laine, and Timo Aila. Elucidating the design space of diffusion-based generative models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Bahjat Kawar, Gregory Vaksman, and Michael Elad. SNIPS: Solving noisy inverse problems stochastically. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=pBKOx_dxYAN.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=kxXvopt9pWK.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, October 2000. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1015957395.

Xiang Li, Soo Min Kwon, Ismail R. Alkhouri, Saiprasad Ravishankar, and Qing Qu. Decoupled data consistency with diffusion purification for image restoration. (arXiv:2403.06054), May 2024. doi: 10.48550/arXiv.2403.06054. URL http://arxiv.org/abs/2403.06054. arXiv:2403.06054 [eess].

Xiang Li, Soo Min Kwon, Shijun Liang, Ismail R. Alkhouri, Saiprasad Ravishankar, and Qing Qu. Decoupled data consistency with diffusion purification for image restoration, 2025. URL https://arxiv.org/abs/2403.06054.

Kewei Liang. Homocentric convergence ball of the secant method. *Applied Mathematics-A Journal of Chinese Universities*, 22:353–365, 2007.

Yun-Yang Liu, Xi-Le Zhao, Yu-Bang Zheng, Tian-Hui Ma, and Hongyan Zhang. Hyperspectral image restoration by tensor fibered rank constrained optimization and plug-and-play regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022. doi: 10.1109/TGRS.2020.3045169.

Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A variational perspective on solving inverse problems with diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=1YO4EE3SPB.

J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 101(2): 185–232, October 2002. ISSN 0304-4149. doi: 10.1016/S0304-4149(02)00150-3.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=aBsCjcPu_tE.

Christopher A. Metzler, Arian Maleki, and Richard G. Baraniuk. From denoising to compressed sensing. (arXiv:1406.4175), April 2016. doi: 10.48550/arXiv.1406.4175. URL http://arxiv.org/abs/1406.4175. arXiv:1406.4175 [cs].

Weili Nie, Brandon Guo, Yujia Huang, Chaowei Xiao, Arash Vahdat, and Animashree Anandkumar. Diffusion models for adversarial purification. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 16805–16827. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/nie22a.html.

D.P. Palomar and S. Verdu. Gradient of mutual information in linear vector gaussian channels. *IEEE Transactions on Information Theory*, 52(1):141–154, 2006. doi: 10.1109/TIT.2005.860424.

Leandro Pardo. *Statistical inference based on divergence measures*. Chapman and Hall/CRC, 2018.

Xinyu Peng, Ziyang Zheng, Wenrui Dai, Nuoqian Xiao, Chenglin Li, Junni Zou, and Hongkai Xiong. Improving diffusion models for inverse problems using optimal posterior covariance, 2024. URL https://arxiv.org/abs/2402.02149.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Hongmin Ren and Ioannis K. Argyros. On the complexity of extending the convergence ball of wang's method for finding a zero of a derivative. *Journal of Complexity*, 64:101526, 2021. ISSN 0885-064X. doi: https://doi.org/10.1016/j.jco.2020.101526. URL https://www.sciencedirect.com/science/article/pii/S0885064X20300704.

Hongmin Ren and Qingbiao Wu. Convergence ball and error analysis of a family of iterative methods with cubic convergence. *Applied Mathematics and Computation*, 209(2):369–378, 2009. ISSN 0096-3003. doi: https://doi.org/10.1016/j.amc.2008.12.057. URL https://www.sciencedirect.com/science/article/pii/S0096300308009739.

Marien Renaud, Jean Prost, Arthur Leclaire, and Nicolas Papadakis. Plug-and-play image restoration with stochastic denoising regularization. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024a.

Marien Renaud, Jean Prost, Arthur Leclaire, and Nicolas Papadakis. Plug-and-play image restoration with stochastic denoising regularization. *CoRR*, abs/2402.01779, 2024b. URL https://doi.org/10.48550/arXiv.2402.01779.

Herbert Robbins. An empirical bayes approach to statistics. *Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability*, 1, 01 1992. doi: 10.1007/978-1-4612-0919-5_26.

Ernest Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Plug-and-play methods provably converge with properly trained denoisers. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5546–5557. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/ryu19a.html.

Oguz Semerci, Ning Hao, Misha E. Kilmer, and Eric L. Miller. Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Transactions on Image Processing*, 23(4):1678–1693, 2014. doi: 10.1109/TIP.2014.2305840.

Viraj Shah and Chinmay Hegde. 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.

Bowen Song, Soo Min Kwon, Zecheng Zhang, Xinyu Hu, Qing Qu, and Liyue Shen. Solving inverse problems with latent diffusion models via hard data consistency. October 2023. URL https://openreview.net/forum?id=j8hdRqOUhN.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. September 2022a. URL https://openreview.net/forum?id=9_gsMA8MRKQ.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations*, 2022b. URL https://openreview.net/forum?id=vaRCHVj0uGI.

Yu Sun, Brendt Wohlberg, and Ulugbek S. Kamilov. An online plug-and-play algorithm for regularized image reconstruction. *IEEE Transactions on Computational Imaging*, 5(3):395–408, September 2019. ISSN 2333-9403, 2334-0118, 2573-0436. doi: 10.1109/TCI.2019.2893568. arXiv:1809.04693 [cs].

Yu Sun, Zihui Wu, Yifan Chen, Berthy T. Feng, and Katherine L. Bouman. Provable probabilistic imaging using score-based generative priors, 2024. URL https://arxiv.org/abs/2310.10835.

Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

Afonso M. Teodoro, José M. Bioucas-Dias, and Mário A. T. Figueiredo. Scene-adapted plug-and-play algorithm with convergence guarantees. (arXiv:1702.02445), November 2017. doi: 10.48550/arXiv.1702.02445. URL http://arxiv.org/abs/1702.02445. arXiv:1702.02445 [cs].

Edward Charles Titchmarsh and David Rodney Heath-Brown. *The theory of the Riemann zeta-function*. Oxford university press, 1986.

Phong Tran, A. Tran, Quynh Phung, and Minh Hoai. Explore image deblurring via encoded blur kernel space. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11951–11960, 2021. URL https://api.semanticscholar.org/CorpusID:235328539.

G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Phys. Rev.*, 36:823–841, Sep 1930. doi: 10.1103/PhysRev.36.823. URL https://link.aps.org/doi/10.1103/PhysRev.36.823.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. *International Journal of Computer Vision*, 128(7):1867–1888, March 2020. ISSN 1573-1405. doi: 10.1007/s11263-020-01303-4. URL http://dx.doi.org/10.1007/s11263-020-01303-4.

Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2008.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

Hengkang Wang, Xu Zhang, Taihui Li, Yuxiang Wan, Tiancong Chen, and Ju Sun. DMPlug: A plug-in method for solving inverse problems with diffusion models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=81IFFsfQUj.

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

Deliang Wei, Peng Chen, Haobo Xu, Jiale Yao, Fang Li, and Tieyong Zeng. Learning cocoercive conservative denoisers via helmholtz decomposition for poisson inverse problems. (arXiv:2505.08909), October 2025. doi: 10.48550/arXiv.2505.08909. URL http://arxiv.org/abs/2505.08909. arXiv:2505.08909 [cs].

Kaixuan Wei, Angelica Aviles-Rivero, Jingwei Liang, Ying Fu, Hua Huang, and Carola-Bibiane Schönlieb. Tfpnp: Tuning-free plug-and-play proximal algorithm with applications to inverse imaging problems. (arXiv:2012.05703), 2021. doi: 10.48550/arXiv.2012.05703. URL http://arxiv.org/abs/2012.05703. arXiv:2012.05703 [cs].

Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=Xq9HQf7VNV.

Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion GANs. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=JprM0p-q0Co.

Le Xu, Lei Cheng, Junting Chen, Wenqiang Pu, and Xiao Fu. Radio map estimation via latent domain plug-and-play denoising, 2025. URL https://arxiv.org/abs/2501.13472.

Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010. doi: 10.1109/TIP.2010.2050625.

Yonghong Yao, Haiyun Zhou, and Yeong-Cheng Liou. Iterative algorithms with variable anchors for non-expansive mappings. *Journal of Applied Mathematics and Computing*, 28:39–49, 2008.

Xin Yuan, Yang Liu, Jinli Suo, Frédo Durand, and Qionghai Dai. Plug-and-play algorithms for video snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7093–7111, 2022. doi: 10.1109/TPAMI.2021.3099035.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. *CoRR*, abs/2407.01521, 2024. URL https://doi.org/10.48550/arXiv.2407.01521.

Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2022. doi: 10.1109/TPAMI.2021.3088914.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric . In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595, Los Alamitos, CA, USA, June 2018. IEEE Computer Society. doi: 10.1109/CVPR.2018.00068. URL https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00068.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023 - Workshops, Vancouver, BC, Canada, June 17-24, 2023*, pp. 1219–1229. IEEE, 2023. doi: 10.1109/CVPRW59228.2023.00129. URL https://doi.org/10.1109/CVPRW59228.2023.00129.

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

A PRELIMINARIES

A.1 NOTATION

Table 2: Summary of notation

Symbol	Description
$\boldsymbol{x} \in \mathbb{R}^d$	The unknown signal or image to be recovered
$\boldsymbol{y} \in \mathbb{R}^n$	Measurements, with $n \leq d$
y_i	ith element of measurement $\boldsymbol{y}$
$\mathcal{A} : \mathbb{R}^d \to \mathbb{R}^n$	Measurement operator
$\mathcal{X}$	Support of $\boldsymbol{x}$
$\mathcal{X}_t$	Support of $\boldsymbol{x}_t$
$\boldsymbol{\xi}$	Additive measurement noise
$\boldsymbol{x}_t$	noisy data by using forward diffusion process with noise $\sigma(t)$
$\boldsymbol{x}_{\sigma^{(k)}}$	noisy data by using forward diffusion process with noise $\sigma^{(k)}$
$\ell(\boldsymbol{y}\|\|\mathcal{A}(\boldsymbol{x}))$	data-fidelity loss (e.g. $\|\boldsymbol{y} - \mathcal{A}(\boldsymbol{x})\|_2^2$)
$h(\boldsymbol{x})$	structural regularization prior (enforced via denoiser)
$\rho > 0$	ADMM penalty parameter
Prox$(\cdot)$	Proximal operator
$\boldsymbol{u}^{(k)}$	Scaled dual variable in iteration k of ADMM
$\boldsymbol{z}^{(k)}$	Auxiliary variable in iteration k of ADMM
$\widetilde{\boldsymbol{z}}^{(k)} = \boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)}$	Pre-denoising input to the PnP denoiser
$\sigma^{(k)}$	Noise level schedule for the AC-DC denoiser
$\sigma_{\boldsymbol{s}^{(k)}}$	Variance parameter in the AC-DC prior for directional correction.
$\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$	Multivariate standard gaussian random variable
$\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}, \sigma) \approx \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} + \sigma \boldsymbol{n})$	Pretrained score function
K	Maximum iteration of ADMM
J	Total iteration of directional correction at for each denoising
M	Smoothness constant of $\nabla \log p_{\text{data}}$ (Assumption 2)
M_t	Smoothness constant of $\nabla \log p_t$
$T > 0$	Maximum time steps used for diffusion
$\mathcal{M}_{\sigma(t)}$	Manifold of $\boldsymbol{x}_t$
$\mathcal{M}_{\sigma^{(k)}}$	Manifold of $\boldsymbol{x}_t$ where $t \in [0, T]$ such that $\sigma(t) = \sigma^{(k)}$
$D_{\sigma^{(k)}}(\boldsymbol{z})$	AC-DC denoiser at kth iteration
$R_{\sigma^{(k)}}(\boldsymbol{z}) = D_{\sigma^{(k)}}(\boldsymbol{z}) - \boldsymbol{z}$	Residual of AC-DC denoiser at kth iteration
$I(\boldsymbol{x}) = \boldsymbol{x}$	Identity mapping function
$\boldsymbol{I}$	Identity matrix
$\boldsymbol{0}$	vector of values 0
$T_1 \circ T_2(\boldsymbol{z}) = T_1(T_2(\boldsymbol{z}))$	Concatentation of two functions T_1 and T_2
$\|\boldsymbol{x}\|_2$	2-norm of a vector $\boldsymbol{x}$
Cov$(\cdot)$	Covariance matrix

A.2 DEFINITIONS

μ-strongly convex function (Boyd & Vandenberghe, 2004). A differentiable function $f : \mathbb{R}^m \to \mathbb{R}$ is *μ-strongly convex* for a certain $\mu > 0$ if

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) + \frac{\mu}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \tag{17}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$.

The notion of of *nonexpansive* and *averaged nonexpansive* have been widely used in the convergence analysis of various nonlinear problems (Combettes & Yamada, 2015; Yao et al., 2008; Eckstein &

Bertsekas, 1992). We use the generalized form of both *nonexpansive* and *averaged nonexpansive* operator for establishing the ball convergence in our method.

Nonexpansive function (Browder, 1965). A function $T : \mathbb{R}^d \to \mathbb{R}^d$ is *nonexpansive* if T is *nonexpansive* function if there exists $\epsilon \in [0, 1]$ such that

$$\|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 \leq \epsilon^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \tag{18}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$.

θ-averaged function (Combettes & Yamada, 2015). A mapping $T : \mathbb{R}^d \to \mathbb{R}^d$ is defined to be *θ-averaged* for a constant $\theta \in (0, 1)$ if there exists a nonexpansive operator $R : \mathbb{R}^d \to \mathbb{R}^d$ such that $T = (1 - \alpha)I + \alpha R$.

The notion of relaxed bound $\|T_k(\boldsymbol{x}) - T_k(\boldsymbol{y})\| \leq \epsilon^{(k)} \|\boldsymbol{x} - \boldsymbol{y}\| + \delta^{(k)},\ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ was used to study and show the convergence in Yao et al. (2008) when $\sum_{k=1}^{\infty} |\delta^{(k)}| < \infty$. We define a similar weaker form of nonexpansive function and θ-averaged functions below.

δ-weakly nonexpansive function. A mapping $T : \mathbb{R}^d \to \mathbb{R}^d$ is said to be *δ-weakly nonexpansive* for $\delta \geq 0$ if there exists $\epsilon \in [0, 1]$ such that

$$\|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 \leq \epsilon^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \delta^2 \tag{19}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$.

δ-weakly θ-averaged function. A mapping $T : \mathbb{R}^d \to \mathbb{R}^d$ is defined to be *δ-weakly θ-averaged* function for a certain $\delta \geq 0$ and $\theta \in (0, 1)$, if there exists a δ-weakly nonexpansive function $R : \mathbb{R}^d \to \mathbb{R}^d$ such that $T = \theta R + (1 - \theta)I$.

Sub-Gaussian random vector. A random vector $\boldsymbol{x} \in \mathbb{R}^d$ (with mean $\mathbb{E}[\boldsymbol{x}]$) is called *sub-Gaussian with parameter σ^2* if its Euclidean norm satisfies a sub-Gaussian tail bound:

$$\Pr(\|\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}]\|_2 > \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right), \quad \forall\, \varepsilon > 0. \tag{20}$$

2-Wasserstein Distance. Let μ and ν be probability measures on $\mathbb{R}^d$ with finite second moments. The 2-Wasserstein distance between μ and ν is defined as

$$W_2(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|_2^2 \, d\gamma(x, y)\right)^{1/2}, \tag{21}$$

where $\Gamma(\mu, \nu)$ denotes the set of all couplings of μ and ν, i.e.,

$$\Gamma(\mu, \nu) = \left\{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) :\ \gamma(A \times \mathbb{R}^d) = \mu(A),\ \gamma(\mathbb{R}^d \times B) = \nu(B),\ \forall A, B \subseteq \mathbb{R}^d \text{ measurable}\right\}. \tag{22}$$

A.3 Supporting Lemmas

Tweedie's lemma establishes an important connection between the score of the marginal distribution and expectation of posterior when the likelihood function is gaussian. This allows the score function of the diffusion model to be used as a *minimum-mean-square-error* (MMSE) denoiser.

Lemma 1 (Tweedie's lemma (Robbins, 1992)). *Let $p_0(\boldsymbol{x}_0)$ be the prior distribution and then $\boldsymbol{x}_t \sim \mathcal{N}(\boldsymbol{x}_0, \boldsymbol{\Sigma})$ be observed with $\boldsymbol{\Sigma}$ known. Suppose $p_t(\boldsymbol{x}_t)$ be the marginal distribution of $\boldsymbol{x}_t$. Then, Tweedie's lemma computes the posterior expectation of $\boldsymbol{x}_0$ given $\boldsymbol{x}_t$ as*

$$\mathbb{E}[\boldsymbol{x}_0 | \boldsymbol{x}_t] = \boldsymbol{x}_t + \boldsymbol{\Sigma} \nabla \log p_t(\boldsymbol{x}_t) \tag{23}$$

The lemmas related to θ-averaged from Combettes & Yamada (2015) are used to show fixed point convergence in Ryu et al. (2019). In the following, we extend all these lemmas to a more general δ-weakly θ-averaged cases that will be used later to show our ball convergence.

Lemma 2. *$T : \mathbb{R}^d \to \mathbb{R}^d$ be a function. Then, the following statements are equivalent:*

(a) T is δ-weakly θ-averaged for $\delta \geq 0$ and $\theta \in (0, 1)$.

(b) $\|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 + (1-2\theta)\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - 2(1-\theta)\langle T(\boldsymbol{x}) - T(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle \leq \delta^2\theta^2$, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$.

(c) $(1 - 1/\theta)I + (1/\theta)T$ is δ-weakly nonexpansive.

(d) $\|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 \leq \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - \frac{1-\theta}{\theta}\|(I - T)(\boldsymbol{x}) - (I - T)(\boldsymbol{y})\|_2^2 + \delta^2\theta$, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$

Proof. Equivalence between (a) and (b): Provided T is *δ-weakly θ-averaged*, let's find the LHS - RHS in (b)

$$
\begin{aligned}
&(1-2\theta)\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 - 2(1-\theta)\langle \boldsymbol{x} - \boldsymbol{y}, T(\boldsymbol{x}) - T(\boldsymbol{y})\rangle \\
\leq &(1-2\theta)\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \|\theta R(\boldsymbol{x}) - \theta R(\boldsymbol{y}) + (1-\theta)(\boldsymbol{x} - \boldsymbol{y})\|_2^2 \\
&\qquad - 2(1-\theta)\langle \boldsymbol{x} - \boldsymbol{y}, \theta(R(\boldsymbol{x}) - R(\boldsymbol{y})) + (1-\theta)(\boldsymbol{x} - \boldsymbol{y})\rangle \\
= &(1-2\theta)\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \theta^2\|R(\boldsymbol{x}) - R(\boldsymbol{y})\|_2^2 + (1-\theta)^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - 2(1-\theta)^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \\
&\qquad + 2\theta(1-\theta)\langle \boldsymbol{x} - \boldsymbol{y}, R(\boldsymbol{x}) - R(\boldsymbol{y})\rangle - 2(1-\theta)\theta\langle \boldsymbol{x} - \boldsymbol{y}, R(\boldsymbol{x}) - R(\boldsymbol{y})\rangle \\
\leq &(1-2\theta)\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \theta^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \theta^2\delta^2 + (1-\theta)^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - 2(1-\theta)^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \\
&\qquad + 2\theta(1-\theta)\langle \boldsymbol{x} - \boldsymbol{y}, R(\boldsymbol{x}) - R(\boldsymbol{y})\rangle - 2(1-\theta)\theta\langle \boldsymbol{x} - \boldsymbol{y}, R(\boldsymbol{x}) - R(\boldsymbol{y})\rangle \\
= &(1 - 2\theta + \theta^2 + (1-\theta)^2 - 2(1-\theta)^2)\|\boldsymbol{x} - \boldsymbol{y}^2\| + \theta^2\delta^2 \\
= &\theta^2\delta^2
\end{aligned}
$$

For another direction, let us suppose T satisfies (a). Let $R = \frac{1}{\theta}(T - (1-\theta)I)$ so that we have $T = \theta R + (1-\theta)I$. Now, we need to show that $\|R(\boldsymbol{x}) - R(\boldsymbol{y})\|_2^2 \leq \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \delta^2$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ i.e. *δ-weakly nonexpansive*.

$$
\begin{aligned}
&\|R(\boldsymbol{x}) - R(\boldsymbol{y})\|_2^2 \\
= &\frac{1}{\theta^2}\|T(\boldsymbol{x}) - T(\boldsymbol{y}) - (1-\theta)(\boldsymbol{x} - \boldsymbol{y})\|_2^2 \\
= &\frac{1}{\theta^2}\left(\|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 + (1-\theta)^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - 2(1-\theta)\langle T(\boldsymbol{x}) - T(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle\right) \\
= &\frac{1}{\theta^2}\left(\|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 + (1-2\theta)\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - 2(1-\theta)\langle T(\boldsymbol{x}) - T(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle + \theta^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right) \\
\leq &\frac{1}{\theta^2}\left(\theta^2\delta^2 + \theta^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right) \\
= &\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \delta^2
\end{aligned}
$$

where, the inequality is due to T satisfying (b).

Equivalence between (a) and (c): Note that T is δ-weakly θ-averaged $\iff$ $T = \theta R + (1-\theta)I$ with R being δ-weakly nonexpansive function. Now, we have

$$
\begin{aligned}
(1 - 1/\theta)I + (1/\theta)T &= (1 - 1/\theta)I + 1/\theta \cdot (\theta R + (1-\theta)I) \\
&= (1 - 1/\theta)I + R - (1 - 1/\theta)I \\
&= R
\end{aligned}
$$

Hence, T being δ-weakly θ-average is equivalent to $(1 - 1/\theta)I + (1/\theta)T$ being δ-weakly nonexpansive.

Equivalence between (a) and (d): From equivalence between (a) and (b), we have

$$\begin{aligned}
&\|T(\boldsymbol{x})-T(\boldsymbol{y})\|_2^2+(1-2\theta)\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-2(1-\theta)\langle T(\boldsymbol{x})-T(\boldsymbol{y}),\boldsymbol{x}-\boldsymbol{y}\rangle\leq\delta^2\theta^2\\
\iff&\|T(\boldsymbol{x})-T(\boldsymbol{y})\|_2^2+(1-2\theta)\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-\\
&\qquad(1-\theta)\left(\|T(\boldsymbol{x})-T(\boldsymbol{y})\|_2^2+\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-\|(T-I)(\boldsymbol{x})-(T-I)(\boldsymbol{y})\|_2^2\right)\leq\delta^2\theta^2\\
\iff&\theta\|T(\boldsymbol{x})-T(\boldsymbol{y})\|_2^2-\theta\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\leq\delta^2\theta^2-(1-\theta)\|(T-I)(\boldsymbol{x})-(T-I)(\boldsymbol{y})\|_2^2\\
\iff&\|T(\boldsymbol{x})-T(\boldsymbol{y})\|_2^2\leq\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-\frac{1-\theta}{\theta}\|(I-T)(\boldsymbol{x})-(I-T)(\boldsymbol{y})\|_2^2+\delta^2\theta
\end{aligned}$$

□

Lemma 3 (Concatenation of δ-weakly θ-averaged functions)**.** *Assume $T_1:\mathbb{R}^d\to\mathbb{R}^d$ and $T_2:\mathbb{R}^d\to\mathbb{R}^d$ are δ_1-weakly θ_1-averaged and δ_2-weakly θ_2-averaged respectively. Then, $T_1\circ T_2$ is δ-weakly θ-averaged, with $\theta=\frac{\theta_1+\theta_2-2\theta_1\theta_2}{1-\theta_1\theta_2}$, and $\delta^2=\frac{1}{\theta}(\delta_1^2\theta_1+\delta_2^2\theta_2)$.*

Proof. Here, we follow the proof structure of Combettes & Yamada (2015). Since $\theta_1(1-\theta_2)\leq(1-\theta_2)$, we have $\theta_1+\theta_2\leq 1+\theta_1\theta_2$, and therefore, $\theta=\frac{\theta_1+\theta_2-2\theta_1\theta_2}{1-\theta_1\theta_2}\in(0,1)$, and let $\delta^2=\frac{\delta_1^2\theta_1+\delta_2^2\theta_2}{\theta}$

Now, from Lemma 2, for $i\in\{1,2\}$, we have,

$$\|T_i(\boldsymbol{x})-T_i(\boldsymbol{y})\|_2^2\leq\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-\frac{1-\theta_i}{\theta_i}\|(I-T_i)(\boldsymbol{x})-(I-T_i)(\boldsymbol{y})\|_2^2+\delta_i^2\theta_i \tag{24}$$

Then, let us evaluate the composition function using this property.

$$\begin{aligned}
&\|T_1\circ T_2(\boldsymbol{x})-T_1\circ T_2(\boldsymbol{y})\|_2^2\\
\leq&\|T_2(\boldsymbol{x})-T_2(\boldsymbol{y})\|_2^2-\frac{1-\theta_1}{\theta_1}\|(I-T_1)(T_2(\boldsymbol{x}))-(I-T_1)(T_2(\boldsymbol{y})\|_2^2\\
\leq&\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-\frac{1-\theta_2}{\theta_2}\|(I-T_2)(\boldsymbol{x})-(I-T_2)(\boldsymbol{y})\|_2^2+\delta_2^2\theta_2\\
&\qquad-\frac{1-\theta_1}{\theta_1}\|(I-T_1)(T_2(\boldsymbol{x}))-(I-T_1)(T_2(\boldsymbol{y})\|_2^2+\delta_1^2\theta_1
\end{aligned}$$

From Bauschke et al. (2017)[Corollary 2.15], we have, for $\alpha\in\mathbb{R}$,

$$\begin{aligned}
&\|\alpha\boldsymbol{u}+(1-\alpha)\boldsymbol{v}\|_2^2+\alpha(1-\alpha)\|\boldsymbol{u}-\boldsymbol{v}\|_2^2=\alpha\|\boldsymbol{u}\|_2^2+(1-\alpha)\|\boldsymbol{v}\|_2^2\\
\implies&\alpha(1-\alpha)\|\boldsymbol{u}+\boldsymbol{v}\|_2^2\leq\alpha\|\boldsymbol{u}\|_2^2+(1-\alpha)\|\boldsymbol{v}\|_2^2
\end{aligned}$$

Now, let $\boldsymbol{u}=(I-T_2)(\boldsymbol{x})-(I-T_2)(\boldsymbol{y})$, $\boldsymbol{v}=(I-T_1)(T_2(\boldsymbol{x}))-(I-T_1)(T_2(\boldsymbol{y}))$, $a=\frac{1-\theta_2}{\theta_2}$, and $b=\frac{1-\theta_1}{\theta_1}$.

$$\begin{aligned}
&\|T_1\circ T_2(\boldsymbol{x})-T_1\circ T_2(\boldsymbol{y})\|_2^2\\
\leq&\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-a\|\boldsymbol{u}\|_2^2-b\|\boldsymbol{v}\|_2^2+\delta_1^2\theta_1+\delta_2^2\theta_2\\
=&\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-(a+b)\left(\frac{a}{a+b}\|\boldsymbol{u}\|_2^2+\frac{b}{a+b}\|\boldsymbol{v}\|_2^2\right)+\delta_1^2\theta_1+\delta_2^2\theta_2\\
=&\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-(a+b)\left(\frac{a}{a+b}\|\boldsymbol{u}\|_2^2+1-\frac{a}{a+b}\|\boldsymbol{v}\|_2^2\right)+\delta_1^2\theta_1+\delta_2^2\theta_2
\end{aligned}$$

Using the above results, we get,

$$\begin{aligned}
&\|T_1\circ T_2(\boldsymbol{x})-T_1\circ T_2(\boldsymbol{y})\|_2^2\\
\leq&\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-(a+b)\frac{ab}{(a+b)^2}\|\boldsymbol{u}+\boldsymbol{v}\|_2^2+\delta_1^2\theta_1+\delta_2^2\theta_2\\
=&\|\boldsymbol{x}-\boldsymbol{y}\|_2^2-\frac{ab}{(a+b)}\|(I-T_1\circ T_2)(\boldsymbol{x})-(I-T_1\circ T_2)(\boldsymbol{y})\|_2^2+\delta_1^2\theta_1+\delta_2^2\theta_2
\end{aligned}$$

Let $\theta = \frac{\theta_1+\theta_2-2\theta_1\theta_2}{1-\theta_1\theta_2}$. Then, we can see that $\frac{ab}{a+b} = \frac{1-\theta}{\theta}$.

$$\|T_1 \circ T_2(\boldsymbol{x}) - T_1 \circ T_2(\boldsymbol{y})\|_2^2 \leq \|\boldsymbol{x}-\boldsymbol{y}\|_2^2 - \frac{1-\theta}{\theta}\|(I - T_1 \circ T_2)(\boldsymbol{x}) - (I - T_1 \circ T_2)(\boldsymbol{y})\|_2^2 + \delta^2\theta \tag{25}$$

where, $\delta^2\theta = \delta_1^2\theta_1 + \delta_2^2\theta_2$. This implies that $T_1 \circ T_2$ is δ-weakly θ-averaged with

$$\theta = \frac{\theta_1+\theta_2-2\theta_1\theta_2}{1-\theta_1\theta_2},\ \delta^2 = \frac{1}{\theta}(\delta_1^2\theta_1 + \delta_2^2\theta_2) \tag{26}$$

□

Lemma 4 (Proposition 5.4 of Giselsson (2015))**.** *Assume ℓ is μ-strongly convex, closed, and proper. Then, $-(2\text{Prox}_{\frac{1}{\rho}\ell} - I)$ is $\frac{\rho}{\rho+\mu}$-averaged.*

Lemma 5 (Pardo (2018))**.** *The* KL divergence *between two gaussian distributions $q_1 = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $q_2 = \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ in $\mathbb{R}^d$ space is given by*

$$KL(q_1||q_2) = \frac{1}{2}\left[\log\frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|} - d + tr(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)\right] \tag{27}$$

where $|\cdot|$ denotes the determinant, and tr denotes the trace of the matrix.

B Influence of AC-step on bringing close to $\{\mathcal{M}_{\sigma(t)}\}_{t=0}^T$

Given a noisy image $\widetilde{\boldsymbol{z}}^{(k)}$ at each iteration k, the denoising aims to recover the underlying clean image $\boldsymbol{z}_\natural^{(k)} \sim p_0(\boldsymbol{z})$ such that $\widetilde{\boldsymbol{z}}^{(k)} = \boldsymbol{z}_\natural^{(k)} + \boldsymbol{s}^{(k)}$, where $\boldsymbol{s}^{(k)}$ is the noise contained in $\widetilde{\boldsymbol{z}}^{(k)}$. The AC-step aims to bring $\boldsymbol{z}_{\text{ac}}^{(k)}$ closer to the noisy distribution $\mathcal{M}_{\sigma^{(k)}}$ on which the $\boldsymbol{s}_{\boldsymbol{\theta}}(\cdot, \sigma^{(k)})$ was trained on. Lemma 6 shows that the AC-step tries to match with the distribution induced by the forward diffusion process.

Lemma 6. *The KL divergence between the target distribution $p(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_\natural^{(k)})$ for correction steps and the distribution $p(\boldsymbol{z}_{\text{ac}}^{(k)}|\widetilde{\boldsymbol{z}}^{(k)})$ induced by the approximate correction step in Algorithm 1 is given by*

$$KL(p(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_\natural^{(k)})||p(\boldsymbol{z}_{\text{ac}}^{(k)}|\widetilde{\boldsymbol{z}}^{(k)})) = \frac{1}{2\left(\sigma^{(k)}\right)^2}\left\|\boldsymbol{s}^{(k)}\right\|_2^2 \tag{28}$$

Proof.

$$\textbf{Target distribution: } p\left(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_\natural^{(k)}\right) = \mathcal{N}(\boldsymbol{z}_\natural^{(k)}, \sigma^{(k)}\boldsymbol{I}) \tag{29}$$

$$\textbf{AC induced distribution: } p\left(\boldsymbol{z}_{\text{ac}}^{(k)}|\widetilde{\boldsymbol{z}}^{(k)}\right) = \mathcal{N}(\widetilde{\boldsymbol{z}}^{(k)}, \sigma^{(k)}\boldsymbol{I}) \tag{30}$$

The KL divergence between these two distribution can be computed in closed form using Lemma 5.

$$\begin{aligned}
\text{KL}(q_1||q_2) &= \frac{1}{2}\left(\log 1 - d + \text{tr}(\boldsymbol{I}) + (\widetilde{\boldsymbol{z}}^{(k)} - \boldsymbol{z}_\natural^{(k)})^T\left(\sigma^{(k)}\right)^{-2}\boldsymbol{I}(\widetilde{\boldsymbol{z}}^{(k)} - \boldsymbol{z}_\natural^{(k)})^T\right) \\
&= \frac{1}{2}\left(0 - d + d + \left(\sigma^{(k)}\right)^{-2}\left\|\widetilde{\boldsymbol{z}}^{(k)} - \boldsymbol{z}_\natural^{(k)}\right\|_2^2\right) \\
&= \frac{1}{2\left(\sigma^{(k)}\right)^2}\left\|\boldsymbol{s}^{(k)}\right\|_2^2
\end{aligned}$$

where, $\boldsymbol{s}^{(k)} = \widetilde{\boldsymbol{z}}^{(k)} - \boldsymbol{z}_\natural^{(k)}$. □

Lemma 6 shows that KL-gap of our approximate AC update; as long as $\sigma^{(k)}$ is sufficiently large, the two distributions remain close. Alkhouri et al. (2023)[Theorem 1] showed a result with a similar flavor. Larger noise $\sigma^{(k)}$ makes the posterior nearly indistinguishable, but it also washes out fine

structural details originally present (low *Signal-to-Noise Ratio* with larger $\sigma^{(k)}$). Existing works often use annealed scheduling $\sigma^{(k)} \downarrow 0$ (Zhu et al., 2023; Renaud et al., 2024a; Wang et al., 2024) to preserve image details, implicitly assuming $\left\|\boldsymbol{s}^{(k)}\right\|_2^2$ decays at least as fast as $(\sigma^{(k)})^2$. With just the use of annealing $\sigma^{(k)}$ schedule, it is not sufficient to ensure that $\boldsymbol{z}_{\text{ac}}^{(k)}$ lands in a desired manifold in each ADMM iteration. To bridge this gap, we propose to use DC-step in addition to the widely used annealing $\sigma^{(k)}$ schedule that explicitly corrects this gap.

C Proof of Theorem 1

The proof involves showing that the each iteration of ADMM-PnP is weakly non-expansive when the denoiser satisfies Assumption 1. This weakly nonexpansiveness of each step leads to ball convergence of the algorithm.

Recall that the subproblems at kth iteration of ADMM-PnP is given by:

$$\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \arg\min_{\boldsymbol{x}} \frac{1}{\rho}\ell(\boldsymbol{y}||\mathcal{A}(\boldsymbol{x})) + \frac{1}{2}\left\|\boldsymbol{x} - \boldsymbol{z}^{(k)} + \boldsymbol{u}^{(k)}\right\|_2^2 \\
&= \text{Prox}_{\frac{1}{\rho}\ell}(\boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)}) \qquad (31a) \\
\boldsymbol{z}^{(k+1)} &= \arg\min_{\boldsymbol{z}} \frac{\gamma}{\rho}h(\boldsymbol{z}) + \frac{1}{2}\left\|\boldsymbol{x}^{(k+1)} - \boldsymbol{z} + \boldsymbol{u}^{(k)}\right\|_2^2 \\
&= \text{Prox}_{\frac{\gamma}{\rho}h}(\boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)}) \\
&= D_{\sigma^{(k)}}(\boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)}) \qquad (31b) \\
\boldsymbol{u}^{(k+1)} &= \boldsymbol{u}^{(k)} + (\boldsymbol{x}^{(k+1)} - \boldsymbol{z}^{(k+1)}) \qquad (31c)
\end{aligned}$$

Lemma 7 (Ryu et al. (2019))**.** *The steps of ADMM-PnP in (31) can be expressed as* $\boldsymbol{v}^{(k+1)} = T(\boldsymbol{v}^{(k)})$ *with* $\boldsymbol{v}^{(k)} = \boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)}$ *and*

$$T = \frac{1}{2}I + \frac{1}{2}(2D_{\sigma^{(k)}} - I)(2\text{Prox}_{\frac{1}{\rho}\ell} - I) \qquad (32)$$

Lemma 8. *$D_\sigma : \mathbb{R}^d \to \mathbb{R}^d$ satisfies Assumption 1 if and only if $\frac{1}{1+2\epsilon}(2D_{\sigma^{(k)}} - I)$ is Δ-weakly θ-averaged with $\theta = \frac{2\epsilon}{1+2\epsilon}$ and $\Delta^2 = 4\delta^2\frac{(1-\theta)^2}{\theta^2}$.*

Proof. We follow the similar proof structure as in Ryu et al. (2019). Let $\theta = \frac{2\epsilon}{1+2\epsilon}$ which implies $\epsilon = \frac{\theta}{2(1-\theta)}$. Here, we can clearly see that $\theta \in [0, 1)$. Let us define $G = \frac{1}{1+2\epsilon}(2D_{\sigma^{(k)}} - I)$ which implies $D_{\sigma^{(k)}} = \frac{1}{2(1-\theta)}G + \frac{1}{2}I$. Then,

$$\begin{aligned}
&\|(D_{\sigma^{(k)}} - I)(\boldsymbol{x}) - (D_{\sigma^{(k)}} - I)(\boldsymbol{y})\|^2 - \epsilon^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \\
=&\|(D_{\sigma^{(k)}}(\boldsymbol{x}) - D_{\sigma^{(k)}}(\boldsymbol{y})) - (\boldsymbol{x} - \boldsymbol{y})\|_2^2 - \epsilon^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \\
=&\left\|\left(D_{\sigma^{(k)}} - \frac{1}{2}I\right)(\boldsymbol{x}) - \left(D_{\sigma^{(k)}} - \frac{1}{2}I\right)(\boldsymbol{y})\right\|_2^2 + \frac{1}{4}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 - \frac{\theta^2}{4(1-\theta)^2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \\
&\qquad - 2\langle\left(D_{\sigma^{(k)}} - \frac{1}{2}I\right)(\boldsymbol{x}) - \left(D_{\sigma^{(k)}} - \frac{1}{2}\right)(\boldsymbol{y}), \frac{1}{2}(\boldsymbol{x} - \boldsymbol{y})\rangle \\
=&\frac{1}{4(1-\theta)^2}\|G(\boldsymbol{x}) - G(\boldsymbol{y})\|_2^2 + \frac{1}{4}\left(1 - \frac{\theta^2}{(1-\theta)^2}\right)\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \\
&\qquad - \frac{1}{2(1-\theta)}\langle G(\boldsymbol{x}) - G(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle \\
=&\frac{1}{4(1-\theta)^2}\left(\|G(\boldsymbol{x}) - G(\boldsymbol{y})\|_2^2 - 2(1-\theta)\langle G(\boldsymbol{x}) - G(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle + (1-2\theta)\|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right)
\end{aligned}$$

Now,

$$\frac{1}{4(1-\theta)^2}\left(\|G(\boldsymbol{x})-G(\boldsymbol{y})\|_2^2-2(1-\theta)\langle G(\boldsymbol{x})-G(\boldsymbol{y}),\boldsymbol{x}-\boldsymbol{y}\rangle+(1-2\theta)\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\right)\leq\delta^2$$
$$\Leftrightarrow\|G(\boldsymbol{x})-G(\boldsymbol{y})\|_2^2-2(1-\theta)\langle G(\boldsymbol{x})-G(\boldsymbol{y}),\boldsymbol{x}-\boldsymbol{y}\rangle+(1-2\theta)\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\leq 4\delta^2(1-\theta)^2$$
$$\Leftrightarrow\|G(\boldsymbol{x})-G(\boldsymbol{y})\|_2^2-2(1-\theta)\langle G(\boldsymbol{x})-G(\boldsymbol{y}),\boldsymbol{x}-\boldsymbol{y}\rangle+(1-2\theta)\|\boldsymbol{x}-\boldsymbol{y}\|_2^2\leq \Delta^2\theta^2$$

where, $\Delta^2=4\delta^2\frac{(1-\theta)^2}{\theta^2}$. From Lemma 2, this is equivalent to G being Δ-weakly θ-averaged. □

C.1 Proof of the theorem

We follow the procedures in Ryu et al. (2019) and expand the results in the δ-weakly expansive denoisers. We show that each iteration of PnP ADMM is also wealkly nonexpansive when the denoiser satisfies Assumption 1.

Proof. From Assumption 1.

$$\|(D_{\sigma^{(k)}}-I)(\boldsymbol{x})-(D_{\sigma^{(k)}}-I)(\boldsymbol{y})\|_2^2\leq\epsilon^2\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+\delta^2 \tag{33}$$

From Lemma 4, we have $-(2\text{Prox}_{\frac{1}{\rho}\ell}-I)$ is $\frac{\rho}{\rho+\mu}$-averaged.

Then, from Lemma 8, we have

$$\frac{1}{1+2\epsilon}\left(2D_{\sigma^{(k)}}-I\right) \tag{34}$$

is δ_1-weakly θ-averaged with $\theta=\frac{2\epsilon}{1+2\epsilon}$ and $\delta_1^2=4\delta^2\frac{(1-\theta)^2}{\theta^2}$.

By Lemma 3, it implies

$$-\frac{1}{1+2\epsilon}(2D_{\sigma^{(k)}}-I)(2\text{Prox}_{\frac{1}{\rho}\ell}-I) \tag{35}$$

is $\delta_\circ$-weakly $\theta_\circ$-averaged with $\theta_\circ=\frac{\rho+2\mu\epsilon}{\rho+\mu+2\mu\epsilon}$ and $\delta_\circ=\frac{1}{\theta_\circ}\cdot\frac{4\delta^2}{2\epsilon(1+2\epsilon)}$.

Now, using the definition of $\delta_\circ$-weakly $\theta_\circ$-averagedness, we have

$$\begin{aligned}(2D_{\sigma^{(k)}}-I)(2\text{Prox}_{\frac{1}{\rho}}-I)&=-(1+2\epsilon)\left((1-\theta_\circ)I+\theta_\circ R\right)\\&=-(1+2\epsilon)\left(\frac{\mu}{\rho+\mu+2\mu\epsilon}I+\frac{\rho+2\mu\epsilon}{\rho+\mu+2\mu\epsilon}R\right)\end{aligned}$$

where, R is a certain $\delta_\circ$-weakly nonexpansive function.

Plugging this result into ADMM-PnP operator (Lemma 7), we get

$$\begin{aligned}T&=\frac{1}{2}I+\frac{1}{2}(2D_{\sigma^{(k)}}-I)(2\text{Prox}_{\frac{1}{\rho}\ell}-I)\\&=\frac{1}{2}I-\frac{1}{2}(1+2\epsilon)\left(\frac{\mu}{\rho+\mu+2\mu\epsilon}I+\frac{\rho+2\mu\epsilon}{\rho+\mu+2\mu\epsilon}R\right)\\&=\underbrace{\frac{\rho}{2(\rho+\mu+2\mu\epsilon)}}_{a}I-\underbrace{\frac{(1+2\epsilon)(\rho+2\mu\epsilon)}{2(\rho+\mu+2\mu\epsilon)}}_{b}R\end{aligned}$$

where, clearly $a>0$ and $b>0$.

Now,

$$\|T(\boldsymbol{x})-T(\boldsymbol{y})\|_2^2=a^2\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+b^2\|R(\boldsymbol{x})-R(\boldsymbol{y})\|_2^2-2\langle a(\boldsymbol{x}-\boldsymbol{y}),b(R(\boldsymbol{x})-R(\boldsymbol{y}))\rangle \tag{36}$$

From Young's inequality, for any $\gamma>0$, we have

$$\langle a(\boldsymbol{x}-\boldsymbol{y}),b(R(\boldsymbol{x})-R(\boldsymbol{y}))\rangle\leq\frac{1}{2\gamma}a^2\|\boldsymbol{x}-\boldsymbol{y}\|_2^2+\frac{\gamma b^2}{2}\|R(\boldsymbol{x})-R(\boldsymbol{y})\|_2^2 \tag{37}$$

Plugging this, we get,

$$\|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 \leq a^2 \left(1 + \frac{1}{\gamma}\right) \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + b^2(1+\gamma)\|R(\boldsymbol{x}) - R(\boldsymbol{y})\|_2^2 \tag{38}$$

$$\leq \left(a^2 \left(1 + \frac{1}{\gamma}\right) + b^2(1+\gamma)\right) \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + b^2(1+\gamma)\delta_\circ^2 \tag{39}$$

where, the second inequality is due to $\delta_\circ$-weak nonexpansiveness of R.

Note, that this holds for any $\gamma > 0$. When $\gamma = \frac{a}{b}$, we have

$$\left(a^2 \left(1 + \frac{1}{\gamma}\right) + b^2(1+\gamma)\right) = (a+b)^2 \tag{40}$$

$$\begin{aligned}
&\|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 \\
\leq &(a+b)^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + b^2 \left(1 + \frac{a}{b}\right) \delta_\circ^2 \\
= &\underbrace{\left(\frac{\rho + \rho\epsilon + \mu\epsilon + 2\mu\epsilon^2}{\rho + \mu + 2\mu\epsilon}\right)^2}_{\epsilon_T^2} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \underbrace{\frac{(\rho + \rho\epsilon + \mu\epsilon + 2\mu\epsilon^2)\delta^2}{\epsilon(\rho + \mu + 2\mu\epsilon)}}_{\delta_T^2}
\end{aligned} \tag{41}$$

Hence, we have

$$\|T(\boldsymbol{x}) - T(\boldsymbol{y})\|_2^2 \leq \epsilon_T^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \delta_T^2 \tag{42}$$

This shows that when $\epsilon_T \leq 1$, then, $\exists_{\boldsymbol{v}^*} \in \mathbb{R}^d$, and $K > 0$ such that $\forall k \geq K$ the following holds:

$$\begin{aligned}
&\left\|\boldsymbol{v}^{(k)} - \boldsymbol{v}^*\right\|_2^2 \leq \epsilon_T^{2k} \left\|\boldsymbol{v}^{(0)} - \boldsymbol{v}^*\right\|_2^2 + \frac{\delta_T^2}{1 - \epsilon_T^2} \\
\implies &\lim_{k\to\infty} \left\|\boldsymbol{v}^{(k)} - \boldsymbol{v}^*\right\| \leq \frac{\delta_T}{\sqrt{1 - \epsilon_T^2}}
\end{aligned} \tag{43}$$

Hence with this we have the sequence $\{\boldsymbol{v}^{(k)} = \boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)}\}_{k\in\mathbb{N}^+}$ converges within a ball of radius $\frac{\delta_T}{\sqrt{1-\epsilon_T^2}}$. Since, $-(2\text{Prox}_{\frac{1}{\rho}\ell} - I)$ is $\frac{\rho}{\rho+\mu}$-averaged, this implies

$$\text{Prox}_{\frac{1}{\rho}\ell} = \frac{1}{2}\frac{\rho}{\rho + \mu}(R - I) \tag{44}$$

for some nonexpansive function R. With this, we have

$$\begin{aligned}
&\lim_{k\to\infty} \left\|\text{Prox}_{\frac{1}{\rho}\ell}(\boldsymbol{v}^{(k)}) - \text{Prox}_{\frac{1}{\rho}\ell}(\boldsymbol{v}^*)\right\|_2^2 \leq \left(\frac{\rho}{\rho + \mu}\right)^2 \left\|\boldsymbol{v}^{(k)} - \boldsymbol{v}^*\right\|_2^2 \\
\implies &\lim_{k\to\infty} \left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\right\|_2 \leq \frac{\rho}{\rho + \mu} \frac{\delta_T}{\sqrt{1 - \epsilon_T^2}}
\end{aligned} \tag{45}$$

where $\boldsymbol{x}^* = \text{Prox}_{\frac{1}{\rho}\ell}(\boldsymbol{v}^*)$.

With these results, we know there exists $\boldsymbol{u}^*$ such that

$$\lim_{k\to\infty} \left\|\boldsymbol{u}^k - \boldsymbol{u}^*\right\|_2 \leq \left(1 + \frac{\rho}{\rho + \mu}\right) \frac{\delta_T}{\sqrt{1 - \epsilon_T^2}} \tag{46}$$

□

D Proof of Theorem 2

Here, we show that our 3-step AC-DC denoiser satisfies Assumption 1 for constants ϵ and δ. In the following, we first show that each step satisfies the weakly nonexpansive assumption. Therefore, the concatenation of these 3 steps meets Assumption 1.

Lemma 9. *Assume the* Variance Exploding (VE) scheduling *(Karras et al., 2022) is used in the diffusion model. Given that the log-density* $\log p_0$ *(i.e.* $\log p_{\text{data}}$*) is* M*-smooth (Assumption 2), the intermediate noisy log-densities* $\{\log p_t\}$ *are* M_t*-smooth for* $t \in [0, T]$ *i.e.* $\|\nabla \log p_t(\boldsymbol{x}) - \nabla \log p_t(\boldsymbol{y})\|_2 \leq M_t \|\boldsymbol{x} - \boldsymbol{y}\|_2$ *for all* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$*. For* t *such that* $\sigma^2(t) < 1/M$*, the smoothness constant* M_t *can be upperbounded as*

$$M_t \leq \frac{M}{1 + M\sigma^2(t)} \leq M \tag{47}$$

Proof. From Tweedie's lemma, we have,

$$\begin{aligned} &\implies \nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t) = -\frac{1}{\sigma^2(t)} \left(\boldsymbol{x}_t - \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]\right) \\ &\implies \nabla^2_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t) = -\frac{1}{\sigma^2(t)} (\boldsymbol{I} - \nabla_{\boldsymbol{x}_t} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]) \end{aligned} \tag{48}$$

Now, let us evaluate the Jacobian $\nabla_{\boldsymbol{x}_t} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]$,

$$\begin{aligned} \nabla_{\boldsymbol{x}_t} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] &= \nabla_{\boldsymbol{x}_t} \int_{\boldsymbol{x}_0 \in \mathcal{X}} \boldsymbol{x}_0 p(\boldsymbol{x}_0|\boldsymbol{x}_t) d\boldsymbol{x}_0 \\ &= \int_{\boldsymbol{x}_0 \in \mathcal{X}} \boldsymbol{x}_0 \left(\nabla_{\boldsymbol{x}_t} \frac{p(\boldsymbol{x}_t|\boldsymbol{x}_0)}{p_t(\boldsymbol{x}_t)} \right) p_0(\boldsymbol{x}_0) d\boldsymbol{x}_0 \\ &= \int_{\boldsymbol{x}_0 \in \mathcal{X}} \boldsymbol{x}_0 \left(\frac{1}{p_t(\boldsymbol{x}_t)} \nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t|\boldsymbol{x}_0) - p(\boldsymbol{x}_t|\boldsymbol{x}_0) \frac{1}{p_t^2(\boldsymbol{x}_t)} \nabla_{\boldsymbol{x}_t} p_t(\boldsymbol{x}_t) \right) p_0(\boldsymbol{x}_0) d\boldsymbol{x}_0 \end{aligned} \tag{49}$$

Given $\boldsymbol{x}_t|\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{x}_0, \sigma(t)^2 \boldsymbol{I})$ and $p(\boldsymbol{x}_t) = \int_{\boldsymbol{x}_0 \in \mathcal{X}} p(\boldsymbol{x}_0) p(\boldsymbol{x}_t|\boldsymbol{x}_0) d\boldsymbol{x}_0$, we can compute their gradient (similar to Peng et al. (2024)) as:

$$\begin{aligned} \nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t|\boldsymbol{x}_0) &= -\frac{1}{\sigma^2(t)} p(\boldsymbol{x}_t|\boldsymbol{x}_0)(\boldsymbol{x}_t - \boldsymbol{x}_0) \\ \nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t) &= \int_{\boldsymbol{x}_0 \in \mathcal{X}} p_0(\boldsymbol{x}_0) \nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t|\boldsymbol{x}_0) d\boldsymbol{x}_0 \\ &= \int_{\boldsymbol{x}_0 \in \mathcal{X}} p_0(\boldsymbol{x}_0) \left(-\frac{1}{\sigma^2(t)} p(\boldsymbol{x}_t|\boldsymbol{x}_0)(\boldsymbol{x}_t - \boldsymbol{x}_0) \right) d\boldsymbol{x}_0 \\ &= -\frac{1}{\sigma^2(t)} p_t(\boldsymbol{x}_t) \mathbb{E}[\boldsymbol{x}_t - \boldsymbol{x}_0|\boldsymbol{x}_t] \end{aligned} \tag{50}$$

Plugging these results in (49) and using integration by parts,

$$\nabla_{\boldsymbol{x}_t} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] = \frac{1}{p(\boldsymbol{x}_t)} \int_{\boldsymbol{x}_0 \in \mathcal{X}} \boldsymbol{x}_0 \nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t|\boldsymbol{x}_0) p(\boldsymbol{x}_0) d\boldsymbol{x}_0 - \frac{\nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t)}{p(\boldsymbol{x}_t)} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] \tag{51}$$

Substituting $\nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t|\boldsymbol{x}_0) = -1/(\sigma^2(t))(\boldsymbol{x}_t - \boldsymbol{x}_0) p(\boldsymbol{x}_t|\boldsymbol{x}_0)$, we have

$$\begin{aligned} \nabla_{\boldsymbol{x}_t} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] &= -\frac{1}{\sigma^2(t) p(\boldsymbol{x}_t)} \int_{\boldsymbol{x}_0 \in \mathcal{X}} \boldsymbol{x}_0 (\boldsymbol{x}_t - \boldsymbol{x}_0) p(\boldsymbol{x}_t|\boldsymbol{x}_0) p(\boldsymbol{x}_0) d\boldsymbol{x}_0 - \frac{\nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t)}{p(\boldsymbol{x}_t)} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] \\ &= \frac{-1}{\sigma^2(t)} \mathbb{E}[\boldsymbol{x}_0 (\boldsymbol{x}_t - \boldsymbol{x}_0)|\boldsymbol{x}_t] - \frac{\nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t)}{p(\boldsymbol{x}_t)} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] \end{aligned} \tag{52}$$

Substituting $\nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t) = \int_{\boldsymbol{x}_0 \in \mathcal{X}} p(\boldsymbol{x}_0) \nabla_{\boldsymbol{x}_t} p(\boldsymbol{x}_t|\boldsymbol{x}_0) d\boldsymbol{x}_0 = -p_t(\boldsymbol{x}_t)/\sigma^2(t) \mathbb{E}[\boldsymbol{x}_t - \boldsymbol{x}_0|\boldsymbol{x}_t]$, we have

$$\begin{aligned} \nabla_{\boldsymbol{x}_t} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] &= \frac{-1}{\sigma^2(t)} \mathbb{E}[\boldsymbol{x}_0 (\boldsymbol{x}_t - \boldsymbol{x}_0)|\boldsymbol{x}_t] + \frac{1}{\sigma^2(t)} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] \mathbb{E}[\boldsymbol{x}_t - \boldsymbol{x}_0 \boldsymbol{x}_t] \\ &= -\frac{1}{\sigma^2(t)} \text{Cov}(\boldsymbol{x}_0, \boldsymbol{x}_t - \boldsymbol{x}_0) \\ &= \frac{1}{\sigma^2(t)} \text{Cov}(\boldsymbol{x}_0|\boldsymbol{x}_t) \end{aligned} \tag{53}$$

Note that this result in (53) is similar to Dytso et al. (2021)[Proposition 1], and has also been derived in Hatsell & Nolte (1971); Palomar & Verdu (2006).

From Assumption 2, we have

$$-M\boldsymbol{I} \preceq \nabla^2_{\boldsymbol{x}_0} \log p_0(\boldsymbol{x}_0) \preceq M\boldsymbol{I},\ \forall \boldsymbol{x}_0 \in \mathcal{X} \tag{54}$$

where, $M > 0$ is a constant.

Then, let's analyze the hessian of the log of posterior distribution $p(\boldsymbol{x}_0|\boldsymbol{x}_t)$,

$$\nabla^2_{\boldsymbol{x}_0} \log p(\boldsymbol{x}_0|\boldsymbol{x}_t) = \nabla^2_{\boldsymbol{x}_0} \log p_0(\boldsymbol{x}_0) + \nabla^2_{\boldsymbol{x}_0} \log p(\boldsymbol{x}_t|\boldsymbol{x}_0) \tag{55}$$

$$\implies -M\boldsymbol{I} + \frac{1}{\sigma^2(t)}\boldsymbol{I} \preceq \nabla^2_{\boldsymbol{x}_0} \log p(\boldsymbol{x}_0|\boldsymbol{x}_t) \preceq M\boldsymbol{I} + \frac{1}{\sigma^2(t)}\boldsymbol{I} \tag{56}$$

When $M < \dfrac{1}{\sigma^2(t)}$, then the distribution $\log p(\boldsymbol{x}_0|\boldsymbol{x}_t)$ is strongly log-concave. In this case, the covariance of distribution $p(\boldsymbol{x}_0|\boldsymbol{x}_t)$ can be bounded (Brascamp & Lieb, 1976) as

$$\left(M\boldsymbol{I} + \frac{1}{\sigma^2(t)}\boldsymbol{I}\right)^{-1} \preceq \text{Cov}(\boldsymbol{x}_0|\boldsymbol{x}_t) \preceq \left(-M\boldsymbol{I} + \frac{1}{\sigma^2(t)}\boldsymbol{I}\right)^{-1} \tag{57}$$

Combining result from equations (48), (53) and (57), we get

$$\begin{aligned}
\nabla^2_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t) &= -\frac{1}{\sigma^2(t)} \left(\boldsymbol{I} - \nabla_{\boldsymbol{x}_t} \mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t]\right) \\
&= -\frac{1}{\sigma^2(t)} \left(\boldsymbol{I} - \frac{1}{\sigma^2(t)} \text{Cov}(\boldsymbol{x}_0|\boldsymbol{x}_t)\boldsymbol{I}\right) \\
\left\|\nabla^2_{\boldsymbol{x}_t} \log q_t(\boldsymbol{x}_t)\right\|_2 &\leq \frac{1}{\sigma^2(t)} \left|\left(1 - \frac{1}{\sigma^2(t)} \cdot \frac{\sigma^2(t)}{M\sigma^2(t)+1}\right)\right| \\
&= \frac{1}{\sigma^2(t)} \cdot \frac{M\sigma^2(t)}{1 + M(\sigma^2(t))} \\
&= \frac{M}{1 + M\sigma^2(t)}
\end{aligned} \tag{58}$$

Hence, the smoothness constant of $\log q(\boldsymbol{x}_t)$ is upper bounded as $M_t \leq \dfrac{M}{1 + M\sigma^2(t)}$ i.e. $M_t \leq M$.

□

Lemma 10. *Let $H^{(k)}_{\text{ac}} : \widetilde{\boldsymbol{z}}^{(k)} \mapsto \boldsymbol{z}^{(k)}_{\text{ac}}$ denote the function corresponding to approximate correction in Algorithm 1. Then, with probability at least $1 - e^{-\nu_k}$, the following holds for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$*

$$\left\|(H^{(k)}_{\text{ac}} - I)(\boldsymbol{x}) - (H^{(k)}_{\text{ac}} - I)(\boldsymbol{y})\right\|^2_2 \leq (\delta^{(k)}_{\text{ac}})^2 + (\epsilon^{(k)}_{\text{ac}})^2 \|\boldsymbol{x} - \boldsymbol{y}\|^2_2 \tag{59}$$

where, ${\delta^{(k)}_{\text{ac}}}^2 = 2(\sigma^{(k)})^2(d + 2\sqrt{d\nu_k} + 2\nu_k)$, and $(\epsilon^{(k)}_{\text{ac}})^2 = 0$.

Proof. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, we have the residuals $R^{(k)}_{\text{ac}}(\boldsymbol{x}) = (H^{(k)}_{\text{ac}} - I)(\boldsymbol{x}) = \sigma^{(k)}\boldsymbol{n}_1$ and $R^{(k)}_{\text{ac}}(\boldsymbol{y}) = (H^{(k)}_{\text{ac}} - I)(\boldsymbol{y}) = \sigma^{(k)}\boldsymbol{n}_2$ where, $\boldsymbol{n}_1, \boldsymbol{n}_2 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

Then, we can bound the norm of difference of these two residuals as

$$\begin{aligned}
\left\|R^{(k)}_{\text{ac}}(\boldsymbol{x}) - R^{(k)}_{\text{ac}}(\boldsymbol{y})\right\|^2_2 &= \left\|\sigma^{(k)}\boldsymbol{n}_1 - \sigma^{(k)}\boldsymbol{n}_2\right\|^2_2 \\
&= (\sigma^{(k)})^2 \|\boldsymbol{n}_{12}\|^2_2 \\
&= 2(\sigma^{(k)})^2 \chi^2_d
\end{aligned} \tag{60}$$

where, $\boldsymbol{n}_{12} = \boldsymbol{n}_1 - \boldsymbol{n}_2 \sim \mathcal{N}(\mathbf{0}, 2\boldsymbol{I})$ and χ^2_d is standard chi-square distribution with d degree of freedom.

From Laurent & Massart (2000)[Lemma 1], the following holds with probability at least $1 - e^{-\nu_k}$

$$\chi^2_d \leq d + 2\sqrt{d\nu_k} + 2\nu_k \tag{61}$$

Plugging this in proves the lemma.

□

Lemma 11. *Let $H^{(k)}_{\mathrm{dc}} : \boldsymbol{z}^{(k)}_{\mathrm{ac}} \mapsto \boldsymbol{z}^{(k)}_{\mathrm{dc}}$ denote the function corresponding to fine correction as defined in Algorithm 1. Then, with probability at least $1 - e^{-\nu_k}$, the following holds for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ if $\sigma^2_{\boldsymbol{s}^{(k)}} < \frac{1}{M_t}$:*

$$\left\|(H^{(k)}_{\mathrm{dc}} - I)(\boldsymbol{x}) - (H^{(k)}_{\mathrm{dc}} - I)(\boldsymbol{y})\right\|^2_2 \leq (\delta^{(k)}_{\mathrm{dc}})^2 + (\epsilon^{(k)}_{\mathrm{dc}})^2 \|\boldsymbol{x} - \boldsymbol{y}\|^2_2 \tag{62}$$

where, $(\delta^{(k)}_{\mathrm{dc}})^2 = \frac{32d\sigma^2_{\boldsymbol{s}^{(k)}}}{(1 - M_t\sigma^2_{\boldsymbol{s}^{(k)}})} \log \frac{2}{\nu_k}$, and $(\epsilon^{(k)}_{\mathrm{dc}})^2 = \left(\frac{\sqrt{2}M_t\sigma^2_{\boldsymbol{s}^{(k)}}}{1 - \sigma^2_{\boldsymbol{s}^{(k)}} M_t}\right)^2$.

Proof. Recall that the target distribution for this step is given by

$$\log p(\boldsymbol{z}_{\sigma^{(k)}} | \boldsymbol{z}^{(k)}_{\mathrm{ac}}) \propto \log p(\boldsymbol{z}^{(k)}_{\mathrm{ac}} | \boldsymbol{z}_{\sigma^{(k)}}) + \log p(\boldsymbol{z}_{\sigma^{(k)}}) \tag{63}$$

where, $p(\boldsymbol{z}^{(k)}_{\mathrm{ac}} | \boldsymbol{z}_{\sigma^{(k)}}) = \mathcal{N}(\boldsymbol{z}_{\sigma^{(k)}}, \sigma^2_{\boldsymbol{s}^{(k)}} \boldsymbol{I})$.

Under Assumptions 2 and 3, the target distribution $p(\boldsymbol{z}_{\sigma^{(k)}} | \boldsymbol{z}^{(k)}_{\mathrm{ac}})$ also inherits smoothness and coercivity properties. These conditions imply the ergodicity of corresponding Langevin diffusion (Mattingly et al., 2002; Chen et al., 2020). In particular, Fokker-Planck equation (Uhlenbeck & Ornstein, 1930) characterizes $p(\boldsymbol{z}_{\sigma^{(k)}} | \boldsymbol{z}^{(k)}_{\mathrm{ac}})$ as its unique stationary distribution. Consequently, the iterates $\boldsymbol{z}^{(k)}_{\mathrm{dc}}$ obtained through Langevin dynamics converge to this distribution as the step size $\eta^{(k)} \to 0$ and the number of iterations $J \to \infty$. The gradient and hessian of the log of this desired distribution are given by

$$\nabla_{\boldsymbol{z}^{(k)}_{\mathrm{dc}}} \log p(\boldsymbol{z}^{(k)}_{\mathrm{dc}} | \boldsymbol{z}^{(k)}_{\mathrm{ac}}) = \frac{1}{\sigma^2_{\boldsymbol{s}^{(k)}}} (\boldsymbol{z}^{(k)}_{\mathrm{ac}} - \boldsymbol{z}^{(k)}_{\mathrm{dc}}) + \nabla_{\boldsymbol{z}^{(k)}_{\mathrm{dc}}} \log p_t(\boldsymbol{z}^{(k)}_{\mathrm{dc}}) \tag{64}$$

$$\nabla^2_{\boldsymbol{z}^{(k)}_{\mathrm{dc}}} \log p(\boldsymbol{z}^{(k)}_{\mathrm{dc}} | \boldsymbol{z}^{(k)}_{\mathrm{ac}}) = -\frac{1}{\sigma^2_{\boldsymbol{s}^{(k)}}} + \nabla^2_{\boldsymbol{z}^{(k)}_{\mathrm{dc}}} \log p_t(\boldsymbol{z}^{(k)}_{\mathrm{dc}}) \tag{65}$$

Here, t refers to the noise level such that $\sigma(t) = \sigma^{(k)}$. By the M_t-smoothness of $\log p_t$ distribution (Lemma 9), we have

$$-\left(M_t + \frac{1}{\sigma^2_{\boldsymbol{s}^{(k)}}}\right) \boldsymbol{I} \preceq \nabla^2_{\boldsymbol{z}^{(k)}_{\mathrm{dc}}} \log p(\boldsymbol{z}^{(k)}_{\mathrm{dc}} | \boldsymbol{z}^{(k)}_{\mathrm{ac}}) \preceq \left(M_t - \frac{1}{\sigma^2_{\boldsymbol{s}^{(k)}}}\right) \boldsymbol{I} \tag{66}$$

When $M_t < \frac{1}{\sigma^2_{\boldsymbol{s}^{(k)}}}$, the hessian is negative semi-definite that implies the distribution being log concave. This also implies that when $M_t \ll \frac{1}{\sigma^2_{\boldsymbol{s}^{(k)}}}$, the likelihood term dominates in the posterior (63).

Using (Wainwright, 2019)[Theorem 3.16], the following holds with probability at least $1 - \nu_k$

$$\left\|\boldsymbol{z}^{(k)}_{\mathrm{dc}} - \mathbb{E}[\boldsymbol{z}^{(k)}_{\mathrm{dc}} | \boldsymbol{z}^{(k)}_{\mathrm{ac}}]\right\|_2 \leq \sqrt{\frac{4}{\lambda_t} \log \frac{2}{\nu_k}} \tag{67}$$

where $\lambda_t = -M_t + \frac{1}{\sigma^2_{\boldsymbol{s}^{(k)}}}$.

Now, using Tweedie's lemma, we have

$$\begin{aligned} \mathbb{E}[\boldsymbol{z}^{(k)}_{\mathrm{dc}} | \boldsymbol{z}^{(k)}_{\mathrm{ac}}] &= \boldsymbol{z}^{(k)}_{\mathrm{ac}} + \sigma^2_{\boldsymbol{s}^{(k)}} \nabla_{\boldsymbol{z}^{(k)}_{\mathrm{ac}}} \log p(\boldsymbol{z}^{(k)}_{\mathrm{ac}}) \\ \implies \mathbb{E}[\boldsymbol{z}^{(k)}_{\mathrm{dc}} | \boldsymbol{z}^{(k)}_{\mathrm{ac}}] - \boldsymbol{z}^{(k)}_{\mathrm{ac}} &= \sigma^2_{\boldsymbol{s}^{(k)}} \nabla_{\boldsymbol{z}^{(k)}_{\mathrm{ac}}} \log p(\boldsymbol{z}^{(k)}_{\mathrm{ac}}) \end{aligned} \tag{68}$$

Combining the above two results, with probability at least $1 - 2\nu_k$, the difference of residual for $\boldsymbol{x}$ and $\boldsymbol{y}$ can be bounded as

$$\begin{aligned}\left\|R_{\mathrm{dc}}^{(k)}(\boldsymbol{x}) - R_{\mathrm{dc}}^{(k)}(\boldsymbol{y})\right\|_2^2 &\leq 2\left(2\sqrt{\frac{4}{\lambda_t}\log\frac{2}{\nu_k}}\right)_2^2 + \left\|\sigma_{\boldsymbol{s}^{(k)}}^2\left(\nabla_{\boldsymbol{x}}\log p_{\boldsymbol{z}_{\mathrm{ac}}^{(k)}}(\boldsymbol{x}) - \nabla_{\boldsymbol{y}}\log p_{\boldsymbol{z}_{\mathrm{ac}}^{(k)}}(\boldsymbol{y})\right)\right\|_2^2 \\ &\leq \frac{32d}{\lambda_t}\log\frac{2}{\nu_k} + 2\left\|\sigma_{\boldsymbol{s}^{(k)}}^2\left(\nabla_{\boldsymbol{x}}\log p_{\boldsymbol{z}_{\mathrm{ac}}^{(k)}}(\boldsymbol{x}) - \nabla_{\boldsymbol{y}}\log p_{\boldsymbol{z}_{\mathrm{ac}}^{(k)}}(\boldsymbol{y})\right)\right\|^2 \\ &\leq \frac{32d}{\lambda_t}\log\frac{2}{\nu_k} + 2\sigma_{\boldsymbol{s}^{(k)}}^4 M_{\boldsymbol{z}_{\mathrm{ac}}^{(k)}}^2\|\boldsymbol{x}-\boldsymbol{y}\|^2 \end{aligned} \tag{69}$$

where, $M_{\boldsymbol{z}_{\mathrm{ac}}^{(k)}}$ is the smoothness constant of $\log p_{\boldsymbol{z}_{\mathrm{ac}}^{(k)}}$ that can be derived using a similar proof procedure as Lemma 9, and $\lambda_t = -M_t + \frac{1}{\sigma_{\boldsymbol{s}^{(k)}}^2}$.

Following the similar procedure as in proof of Lemma 9, we get,

$$M_{\boldsymbol{z}_{\mathrm{ac}}^{(k)}} \leq \frac{M_t}{1 - \sigma_{\boldsymbol{s}^{(k)}}^2 M_t} \tag{70}$$

Plugging this leads to the lemma.

□

Lemma 12. *Let $H_{\mathrm{tw}}^{(k)} : \boldsymbol{z}_{\mathrm{dc}}^{(k)} \mapsto \boldsymbol{z}_{\mathrm{tw}}^{(k)}$ denote the projection function using Tweedie's lemma defined in Algorithm 1. Then, we have the following*

$$\left\|(H_{\mathrm{tw}}^{(k)} - I)(\boldsymbol{x}) - (H_{\mathrm{tw}}^{(k)} - I)(\boldsymbol{y})\right\|_2^2 \leq (\epsilon_{\mathrm{tw}}^{(k)})^2\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + {\delta_{\mathrm{tw}}^{(k)}}^2 \tag{71}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, where $(\epsilon_{\mathrm{tw}}^{(k)})^2 = (\sigma^{(k)})^4 M_t^2$, and ${\delta_{\mathrm{tw}}^{(k)}}^2 = 0$.

Proof. From Tweedie's lemma, we have

$$\begin{aligned}\boldsymbol{z}_{\mathrm{tw}}^{(k)} &= \mathbb{E}[\boldsymbol{z}_0|\boldsymbol{z}_t = \boldsymbol{z}_{\mathrm{dc}}^{(k)}] \\ &= \boldsymbol{z}_{\mathrm{dc}}^{(k)} + (\sigma^{(k)})^2\nabla\log p_t(\boldsymbol{z}_{\mathrm{dc}}^{(k)})\end{aligned}$$

where, $t \in [0, T]$ such that $\sigma^{(k)} = \sigma(t)$.

Then, the residuals are given by

$$R_{\mathrm{tw}}^{(k)}(\boldsymbol{x}) = (\sigma^{(k)})^2\nabla\log p_t(\boldsymbol{x}) \tag{72}$$

$$R_{\mathrm{tw}}^{(k)}(\boldsymbol{y}) = (\sigma^{(k)})^2\nabla\log p_t(\boldsymbol{y}) \tag{73}$$

Now, the norm of the difference of the residuals can be written as

$$\left\|R_{\mathrm{tw}}^{(k)}(\boldsymbol{x}) - R_{\mathrm{tw}}^{(k)}(\boldsymbol{y})\right\|_2^2 = (\sigma^{(k)})^4\|\nabla\log p_t(\boldsymbol{x}) - \nabla\log p_t(\boldsymbol{y})\|_2^2 \tag{74}$$

$$\leq (\sigma^{(k)})^4 M_t^2\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \tag{75}$$

where, M_t is the smoothness constant of $\log p_t(\boldsymbol{x})$. □

D.1 Main proof:

Proof. **Part (a)** Using Lemma 10, 11, and 12, with probability at least $1 - 2e^{-\nu_k}$, we have

$$\begin{aligned}&\|R_{\sigma^{(k)}}(\boldsymbol{x}) - R_{\sigma^{(k)}}(\boldsymbol{y})\|_2^2 \\ \leq& 3\left\|R_{\mathrm{ac}}^{(k)}(\boldsymbol{x}) - R_{\mathrm{ac}}^{(k)}(\boldsymbol{y})\right\|_2^2 + 3\left\|R_{\mathrm{dc}}^{(k)}(\boldsymbol{x}) - R_{\mathrm{dc}}^{(k)}(\boldsymbol{y})\right\|_2^2 + 3\left\|R_{\mathrm{tw}}^{(k)}(\boldsymbol{x}) - R_{\mathrm{tw}}^{(k)}(\boldsymbol{y})\right\|_2^2 \\ \leq& 3((\epsilon_{\mathrm{ac}}^{(k)})^2 + (\epsilon_{\mathrm{dc}}^{(k)})^2 + (\epsilon_{\mathrm{tw}}^{(k)})^2)\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + 3({\delta_{\mathrm{ac}}^{(k)}}^2 + {\delta_{\mathrm{dc}}^{(k)}}^2 + {\delta_{\mathrm{tw}}^{(k)}}^2) \\ \leq& \epsilon_k^2\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \delta_k^2\end{aligned} \tag{76}$$

where,

$$\epsilon_k^2 = 3\left(\left(\frac{\sqrt{2}M_t\sigma^2_{\boldsymbol{s}^{(k)}}}{1-\sigma^2_{\boldsymbol{s}^{(k)}}M_t}\right)^2 + (\sigma^{(k)})^4 M_t^2\right)$$

$$\delta_k^2 = 3\left(2(\sigma^{(k)})^2(d+2\sqrt{d\nu_k}+2\nu_k) + \frac{32d\sigma^2_{\boldsymbol{s}^{(k)}}}{(1-M_t\sigma^2_{\boldsymbol{s}^{(k)}})}\log\frac{2}{\nu_k}\right)$$

Using Lemma 9 leads to the final theorem.

Part (b). Let us set $\nu_k = \ln\left(\frac{2\pi^2}{6\eta}\right) + 2\ln k$. With this, the above weakly nonexpansiveness holds for all $k \in \mathbb{N}^+$ with probability at least

$$\begin{aligned} &1 - \sum_{k=1}^{\infty} 2e^{-\ln\left(\frac{2\pi^2}{6\eta}\right)-2\ln k} \\ =&1 - \frac{6\eta}{\pi^2} \times \sum_{k=1}^{\infty}\frac{1}{k^2} \end{aligned} \tag{77}$$

Using Riemann zeta function (Titchmarsh & Heath-Brown, 1986) at value 2, we have

$$\zeta(2) = \sum_{k=1}^{\infty}\frac{1}{k^2} = \frac{\pi^2}{6} \tag{78}$$

Plugging this in we get the probability to be at least $1-\eta$. Now, combining the results with Theorem 1 leads to the final proof of this part. □

E Proof of Theorem 3

Here, show that our 3-step AC-DC denoiser is bounded with high probability. We first show that each of 3 steps are bounded, and then combined them to establish the boundedness of our AC-DC denoiser as a whole. And following the boundedness, we show that AC-DC ADMM-PnP converges to a fixed point with proper scheduling of $\sigma^{(k)}$ and $\sigma_{\boldsymbol{s}^{(k)}}$.

Lemma 13 (Uniform score bound)**.** *Suppose Assumption 2 holds. Let*

$$D := \operatorname{diam}(\mathcal{X}) = \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{X}} \|\boldsymbol{x}-\boldsymbol{y}\|_2 < \infty \text{ and } S := \inf_{\boldsymbol{x}\in\mathcal{X}} \|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_2 < \infty.$$

Then, with $L = MD + S$, we have

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_\infty \le L.$$

Proof. From Assumption 2, we have

$$\|\nabla \log p_{\text{data}}(\boldsymbol{x}) - \nabla \log p_{\text{data}}(\boldsymbol{y})\|_2 \le M\|\boldsymbol{x}-\boldsymbol{y}\|_2,\ \forall \boldsymbol{x},\boldsymbol{y}\in\mathcal{X} \tag{79}$$

Fix any $\boldsymbol{x}_0 \in \mathcal{X}$. By the triangle inequality, for all $\boldsymbol{x}\in\mathcal{X}$,

$$\begin{aligned} \|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_2 &\le \|\nabla \log p_{\text{data}}(\boldsymbol{x}) - \nabla \log p_{\text{data}}(\boldsymbol{x}_0)\|_2 + \|\nabla \log p_{\text{data}}(\boldsymbol{x}_0)\|_2 \\ &\le M\|\boldsymbol{x}-\boldsymbol{x}_0\|_2 + \|\nabla \log p_{\text{data}}(\boldsymbol{x}_0)\|_2 \end{aligned} \tag{80}$$

Taking the supremum over $\boldsymbol{x}\in\mathcal{X}$ and then the infimum over $\boldsymbol{x}_0\in\mathcal{X}$ yields

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_2 \le \sup_{\boldsymbol{x}\in\mathcal{X}} M\|\boldsymbol{x}-\boldsymbol{x}_0\|_2 + \inf_{\boldsymbol{x}_0\in\mathcal{X}} \|\nabla \log p_{\text{data}}(\boldsymbol{x}_0)\|_2 \tag{81}$$

$$\le MD + S \tag{82}$$

Because $\|\boldsymbol{u}\|_\infty \le \|\boldsymbol{u}\|_2$ for any vector,

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_\infty \le MD + S \tag{83}$$

which proves the above lemma. □

Lemma 14. *Assuming* $\|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_\infty \leq, \forall \boldsymbol{x} \in \mathcal{X}$ *, the score of intermediate noisy distributions* $\{p_t\}_{t\in[0,T]}$ *are bounded as.*

$$\|\nabla \log p_t(\boldsymbol{x})\|_2 \leq \sqrt{d}L \tag{84}$$

for all $\boldsymbol{x} \in \mathcal{X}$.

Proof. We have $\boldsymbol{x}_t = \boldsymbol{x}_0 + \sigma(t)\boldsymbol{n}$ with $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Let us denote $\boldsymbol{n}_1 = \sigma(t)\boldsymbol{n}$. Then, the marginal distribution is given by the convolution of two distributions.

$$p_t(\boldsymbol{x}) = \int_{\boldsymbol{x}_1 \in \mathcal{X}} p_{\boldsymbol{n}_1}(\boldsymbol{x}_1) p_0(\boldsymbol{x} - \boldsymbol{x}_1) d\boldsymbol{x}_1 \tag{85}$$

Then, the score is given by

$$\begin{aligned}
\nabla \log p_t(\boldsymbol{x}) &= \frac{\nabla p_t(\boldsymbol{x})}{p_t(\boldsymbol{x})} \\
&= \frac{1}{p_t(\boldsymbol{x})} \int_{\boldsymbol{x}_1 \in \mathcal{X}} p_{\boldsymbol{n}_1}(\boldsymbol{x}_1) \nabla_{\boldsymbol{x}} p_0(\boldsymbol{x} - \boldsymbol{x}_1) d\boldsymbol{x}_1 \\
&= \frac{1}{p_t(\boldsymbol{x})} \int_{\boldsymbol{x}_1 \in \mathcal{X}} p_{\boldsymbol{n}_1}(\boldsymbol{x}_1) p_0(\boldsymbol{x} - \boldsymbol{x}_1) \frac{\nabla_{\boldsymbol{x}} p_0(\boldsymbol{x} - \boldsymbol{x}_1)}{p_0(\boldsymbol{x} - \boldsymbol{x}_1)} d\boldsymbol{x}_1 \\
&= \frac{1}{p_t(\boldsymbol{x})} \int_{\boldsymbol{x}_1 \in \mathcal{X}} p_{\boldsymbol{n}_1}(\boldsymbol{x}_1) p_0(\boldsymbol{x} - \boldsymbol{x}_1) \nabla_{\boldsymbol{x}} \log p_0(\boldsymbol{x} - \boldsymbol{x}_1) d\boldsymbol{x}_1
\end{aligned} \tag{86}$$

Now the norm can be bounded as

$$\begin{aligned}
\|\nabla \log p_t(\boldsymbol{x})\|_2 &\leq \frac{1}{p_t(\boldsymbol{x})} \int_{\boldsymbol{x}_1 \in \mathcal{X}} p_{\boldsymbol{n}_1}(\boldsymbol{x}_1) p_0(\boldsymbol{x} - \boldsymbol{x}_1) \|\nabla_{\boldsymbol{x}} \log p_0(\boldsymbol{x} - \boldsymbol{x}_1)\|_2 d\boldsymbol{x}_1 \\
&\leq \sup_{\boldsymbol{x}_2 \in \mathcal{X}} \|\nabla_{\boldsymbol{x}} \log p_0(\boldsymbol{x}_2)\|_2 \frac{1}{p_t(\boldsymbol{x})} \int_{\boldsymbol{x}_1 \in \mathcal{X}} p_{\boldsymbol{n}_1}(\boldsymbol{x}_1) p_0(\boldsymbol{x} - \boldsymbol{x}_1) d\boldsymbol{x}_1 \\
&= \sqrt{d}L
\end{aligned} \tag{87}$$

The final equality is due to the fact $\|\boldsymbol{x}\|_2 \leq \sqrt{d}\|\boldsymbol{x}\|_\infty$.

□

Lemma 15. *Let* $H_{\text{ac}}^{(k)} : \widetilde{\boldsymbol{z}}^{(k)} \mapsto \boldsymbol{z}_{\text{ac}}^{(k)}$ *denote the function corresponding to approximate correction to noise level* $\sigma^{(k)}$ *defined in Algorithm 1. Then, with probability at least* $1 - e^{-\nu}$*, the following holds for any* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$

$$\frac{1}{d}\left\|(H_{\text{ac}}^{(k)} - I)(\boldsymbol{x})\right\|_2^2 \leq (\sigma^{(k)})^2(1 + 2\sqrt{\nu} + 2\nu) \tag{88}$$

Proof. For any $\boldsymbol{x} \in \mathcal{X}$, we have the residual $R_{\text{ac}}^{(k)}(\boldsymbol{x}) = (H_{\text{ac}}^{(k)} - I)(\boldsymbol{x}) = \sigma^{(k)}\boldsymbol{n},\ \boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Then,

$$\left\|R_{\text{ac}}^{(k)}(\boldsymbol{x})\right\|_2^2 = (\sigma^{(k)})^2\|\boldsymbol{n}\|_2^2 = (\sigma^{(k)})^2\chi_d^2 \tag{89}$$

where χ_d^2 is standard chi-square distribution with d degrees of freedom. From Laurent & Massart (2000)[Lemma 1], the following holds with probability at least $1 - e^{-\nu}$

$$\chi_d^2 \leq d + 2\sqrt{d\nu} + 2\nu \tag{90}$$

This implies $\frac{1}{d}\left\|H_{\text{ac}}^{(k)} - I)(\boldsymbol{x})\right\|_2^2 \leq (\sigma^{(k)})^2(1 + 2\sqrt{1\nu} + 2\nu)$ with probability at least $1 - e^{-\nu}$ due to $d \geq 1$.

□

Lemma 16. *Let* $H_{\text{dc}}^{(k)} : \boldsymbol{z}_{\text{ac}}^{(k)} \mapsto \boldsymbol{z}_{\text{dc}}^{(k)}$ *denote the function corresponding to fine correction defined in Algorithm 1. Assume* $\|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_\infty \leq L$*. Then, with probability at least* $1 - e^{-\nu}$*, the following holds for any* $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$*:*

$$\frac{1}{d}\left\|(H_{\text{dc}}^{(k)} - I)(\boldsymbol{x})\right\|_2^2 \leq \frac{8}{\lambda_t}\log\frac{2}{\nu} + \sigma_{\boldsymbol{s}^{(k)}}^4 L^2 \tag{91}$$

where, $\lambda_t = -M_t + \frac{1}{\sigma_{\boldsymbol{s}^{(k)}}^2}$.

Proof. From (67), with probability at least $1 - 2\nu$, the norm can be bounded as

$$\left\| R_{\mathrm{dc}}^{(k)}(\boldsymbol{x}) \right\|_2^2 \leq \frac{8d}{\lambda_t} \log \frac{2}{\nu} + \sigma_{\boldsymbol{s}^{(k)}}^4 \left\| \nabla_{\boldsymbol{x}} \log p_{\boldsymbol{z}_{\mathrm{ac}}^{(k)}}(\boldsymbol{x}) \right\|_2^2 \tag{92}$$

Then, using Lemma 14, we have

$$\frac{1}{d} \left\| R_{\mathrm{dc}}^{(k)}(\boldsymbol{x}) \right\|_2^2 \leq \frac{8}{\lambda_t} \log \frac{2}{\nu} + \sigma_{\boldsymbol{s}^{(k)}}^4 L^2 \tag{93}$$

□

Lemma 17. *Let $H_{\mathrm{tw}}^{(k)} : \boldsymbol{z}_{\mathrm{dc}}^{(k)} \mapsto \boldsymbol{z}_{\mathrm{tw}}^{(k)}$ denote the projection function using Tweedie's lemma defined in Algorithm 1. Assume $\|\nabla \log p_{\mathrm{data}}(\boldsymbol{x})\|_\infty \leq L$. Then, we have the following*

$$\frac{1}{d} \left\| (H_{\mathrm{tw}}^{(k)} - I)(\boldsymbol{x}) \right\|_2^2 \leq (\sigma^{(k)})^4 L^2 \tag{94}$$

for any $\boldsymbol{x} \in \mathcal{X}$.

Proof. From Tweedie's lemma, we have

$$\begin{aligned} \boldsymbol{z}_{\mathrm{tw}}^{(k)} &= \mathbb{E}[\boldsymbol{z}_0 | \boldsymbol{z}_t = \boldsymbol{z}_{\mathrm{dc}}^{(k)}] \\ &= \boldsymbol{z}_{\mathrm{dc}}^{(k)} + (\sigma^{(k)})^2 \nabla \log p_t(\boldsymbol{z}_{\mathrm{dc}}^{(k)}) \end{aligned} \tag{95}$$

where, $t \in [0, T]$ such that $\sigma^{(k)} = \sigma(t)$.

Now, the norm of residual can be written as

$$\begin{aligned} \left\| R_{\mathrm{tw}}^{(k)}(\boldsymbol{x}) \right\|_2^2 &= \left\| (\sigma^{(k)})^2 \nabla \log p_t(\boldsymbol{x}) \right\|_2^2 \\ &\leq (\sigma^{(k)})^4 L^2 \cdot d \end{aligned} \tag{96}$$

where, L bound of gradient from Lemma 14. □

E.1 Main proof

Combining Lemmas 15, 16 and 17 leads to the proof of part (a) of Theorem 3.

With probability at least $1 - 2e^{-\nu_k}$, the denoiser satisfies the bounded residual condition.

$$1/d \|(D_{\sigma^{(k)}} - I)(\boldsymbol{x})\|_2^2 \leq c_k^2. \tag{97}$$

Let's define the relative residue as:

$$\beta_k := \frac{1}{\sqrt{d}} \left(\left\| \boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k-1)} \right\|_2 + \left\| \boldsymbol{z}^{(k)} - \boldsymbol{z}^{(k-1)} \right\|_2 + \left\| \boldsymbol{u}^{(k)} - \boldsymbol{u}^{(k-1)} \right\|_2 \right) \tag{98}$$

For any $\eta \in [0, 1)$ and a constant $\gamma > 1$, the penalty parameter ρ_k is adjusted at each iteration k according to following rule (Chan et al., 2016):

$$\rho_{k+1} = \begin{cases} \gamma \rho_k & \text{if } \beta_{k+1} \geq \eta \beta_k \quad \text{(Case 1)} \\ \rho_k & \text{else (Case 2)} \end{cases} \tag{99}$$

The PnP-ADMM with adaptive penalty involves two cases as shown above.

At iteration k, if Case 1 holds, then by Lemma 18 we have

$$\beta_{k+1} \leq 6c_k + 2c_{k-1} + \frac{2R}{\rho_k} \tag{100}$$

On the other hand if Case 2 holds, then,

$$\beta_{k+1} \leq \eta \beta_k \tag{101}$$

Define $a_k = 6c_k + 2c_{k-1} + \frac{2R}{\rho_k}$. Combining two cases, we get

$$\beta_{k+1} \leq \delta\beta_k + a_k, \quad \delta = \begin{cases} \eta, & \text{if Case 2 holds at iteration } k \\ 0, & \text{if Case 1 holds at iteration } k \end{cases} \tag{102}$$

Note that for Case 1 $\rho_{k+1} = \gamma\rho_k$ and with $\gamma > 1$, we get $\lim_{k\to\infty} \frac{c}{\rho_k} = 0$. In addition, with $\nu_k = \ln \frac{2\pi^2}{6\eta} + 2\ln k$, and the scheduling of $\sigma^{(k)}$, $\sigma_{\boldsymbol{s}^{(k)}}$ that satisfies

$$\lim_{k\to\infty} (\sigma^{(k)})^2(2 + 4\sqrt{\nu_k} + 4\nu_k) = 0,\ \lim_{k\to\infty} \frac{\sigma^2_{\boldsymbol{s}^{(k)}}}{1 - M\sigma^2_{\boldsymbol{s}^{(k)}}} \log \frac{2}{\nu_k} = 0,\ \lim_{k\to\infty} \sigma^{(k)} = 0, \lim_{k\to\infty} \sigma_{\boldsymbol{s}^{(k)}} = 0$$

results in $\lim_{k\to\infty} a_k = 0$.

As $k \to \infty$, 3 different scenarios could occur. Let's analyze each of the scenarios one by one.

- Scene 1: Case 1 occurs infinitely many times and Case 2 occurs finitely many times. When Scene 1 occurs, then, there exists a constant $K_1 > 0$ such that $\beta_{k+1} \leq a_k$ for all $k \geq K_1$. Since $\lim_{k\to\infty} a_k = 0$, this leads to $\lim_{k\to\infty} \beta_k = 0$.
- Scene 2: Case 2 occurs infinitely many times and Case 1 occurs finitely many times. Similarly, there exists a constant $K_2 > 0$ such that $\beta_{k+1} \leq \eta\beta_k$ for all $k \geq K_2$. Then, we have
$$\beta_k \leq \eta^{k-K_2}\beta_{K_2} \tag{103}$$
And with $\eta \in [0, 1)$, we have $\lim_{k\to\infty} \beta_k = 0$.
- Scene 3: Both Case 1 and Case 2 occurs infinitely many times. With the Scene 1 and Scene 2 converging, the sequence $\lim_{k\to\infty} \beta_k = 0$ under this Scene as well.

This proves the part (b) of Theorem 3.

Lemma 18. *For any iteration k that falls into Case 1, the following holds*

$$\beta_{k+1} \leq 6c_k + 2c_{k-1} + \frac{2R}{\rho_k} \tag{104}$$

Proof. Consider the subproblem (7a) with adaptive penalty parameter ρ_k (as defined in (99)),

$$\boldsymbol{x}^{(k+1)} = \arg\min_{\boldsymbol{x}} \frac{1}{\rho_k}\ell(\boldsymbol{y}||\mathcal{A}(\boldsymbol{x})) + \frac{1}{2}\left\|\boldsymbol{x} - \boldsymbol{z}^{(k)} + \boldsymbol{u}^{(k)}\right\|_2^2 \tag{105}$$

With the first order optimality condition, the solution $\boldsymbol{x}^{(k+1)}$ satisfies

$$\frac{1}{\rho_k}\nabla_{\boldsymbol{x}}\ell(\boldsymbol{y}||\mathcal{A}(\boldsymbol{x}))\big|_{\boldsymbol{x}=\boldsymbol{x}^{(k+1)}} + (\boldsymbol{x}^{(k+1)} - \boldsymbol{z}^{(k)} + \boldsymbol{u}^{(k)}) = \boldsymbol{0} \tag{106}$$

$$\implies \frac{1}{\sqrt{d}}\left\|\boldsymbol{x}^{(k+1)} - \boldsymbol{z}^{(k)} + \boldsymbol{u}^{(k)}\right\|_2 = \frac{1}{\rho_k\sqrt{d}}\left\|\nabla_{\boldsymbol{x}}\ell(\boldsymbol{y}||\mathcal{A}(\boldsymbol{x}))\big|_{\boldsymbol{x}=\boldsymbol{x}^{(k+1)}}\right\|_2 \tag{107}$$

Using the assumption of existence of $R < \infty$ such that $\|\nabla_{\boldsymbol{x}}\ell(\boldsymbol{y}||\mathcal{A}(\boldsymbol{x}))\|_2/\sqrt{d} \leq R,\ \forall \boldsymbol{x} \in \mathcal{X}$, we get

$$\frac{1}{\sqrt{d}}\left\|\boldsymbol{x}^{(k+1)} - \boldsymbol{z}^{(k)} + \boldsymbol{u}^{(k)}\right\|_2 \leq \frac{R}{\rho_k} \tag{108}$$

Since the denoiser $D_{\sigma^{(k)}}$ is bounded with probability at least $1 - 2e^{-\nu_k}$, we have

$$\frac{1}{\sqrt{d}}\|(D_{\sigma^{(k)}} - I)(\boldsymbol{x})\|_2 \leq c_k \tag{109}$$

Now,

$$\frac{1}{\sqrt{d}}\left\|\boldsymbol{x}^{(k+1)} - \boldsymbol{z}^{(k+1)} + \boldsymbol{u}^{(k)}\right\|_2 = \frac{1}{\sqrt{d}}\left\|\boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)} - D_{\sigma^{(k)}}(\boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)})\right\|_2 \tag{110}$$

$$= \frac{1}{\sqrt{d}}\left\|(D_{\sigma^{(k)}} - I)(\boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)})\right\|_2 \tag{111}$$

$$\leq c_k \tag{112}$$

Now, using triangle inequality, we can bound $\left\|\boldsymbol{z}^{(k+1)}-\boldsymbol{z}^{(k)}\right\|_2$ as

$$\frac{1}{\sqrt{d}}\left\|\boldsymbol{z}^{(k+1)}-\boldsymbol{z}^{(k)}\right\|_2=\frac{1}{\sqrt{d}}\left\|\boldsymbol{z}^{(k+1)}-\boldsymbol{x}^{(k+1)}-\boldsymbol{u}^{(k)}+\boldsymbol{x}^{(k+1)}+\boldsymbol{u}^{(k)}-\boldsymbol{z}^{(k)}\right\|_2 \tag{113}$$

$$\leq \frac{R}{\rho_k}+c_k \tag{114}$$

Similarly, it can be shown that

$$\frac{1}{\sqrt{d}}\left\|\boldsymbol{u}^{(k+1)}\right\|_2=\frac{1}{\sqrt{d}}\left\|\boldsymbol{u}^{(k)}+\boldsymbol{x}^{(k+1)}-\boldsymbol{z}^{(k+1)}\right\|_2 \tag{115}$$

$$=\frac{1}{\sqrt{d}}\left\|\boldsymbol{u}^{(k)}+\boldsymbol{x}^{(k+1)}-D_{\sigma^{(k)}}(\boldsymbol{x}^{(k+1)}+\boldsymbol{u}^{(k)})\right\|_2 \tag{116}$$

$$=\frac{1}{\sqrt{d}}\left\|(D_{\sigma^{(k)}}-I)(\boldsymbol{x}^{(k+1)}+\boldsymbol{u}^{(k)})\right\|_2 \tag{117}$$

$$=c_k \tag{118}$$

This implies $\frac{1}{\sqrt{d}}\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{u}^{(k)}\right\|_2\leq 2c_k$. Finally, we use $\boldsymbol{x}^{(k+1)}=\boldsymbol{u}^{(k+1)}-\boldsymbol{u}^{(k)}+\boldsymbol{z}^{(k+1)}$ to obtain

$$\frac{1}{\sqrt{d}}\left\|\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}\right\|_2 \tag{119}$$

$$=\frac{1}{\sqrt{d}}\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{u}^{(k)}+\boldsymbol{z}^{(k+1)}-\boldsymbol{u}^{(k)}+\boldsymbol{u}^{(k-1)}-\boldsymbol{z}^{(k)}\right\|_2 \tag{120}$$

$$\leq\frac{1}{\sqrt{d}}\left\|\boldsymbol{u}^{(k+1)}-\boldsymbol{u}^{(k)}\right\|_2+\frac{1}{\sqrt{d}}\left\|\boldsymbol{u}^{(k)}-\boldsymbol{u}^{(k-1)}\right\|_2+\frac{1}{\sqrt{d}}\left\|\boldsymbol{z}^{(k+1)}-\boldsymbol{z}^{(k)}\right\|_2 \tag{121}$$

$$\leq 2c_k+2c_{k-1}+\frac{R}{\rho_k}+c_k \tag{122}$$

$$=3c_k+2c_{k-1}+\frac{R}{\rho_k} \tag{123}$$

Combining all the bounds results using triangle inequality results in

$$\beta_{k+1}\leq 6c_k+2c_{k-1}+\frac{2R}{\rho_k} \tag{124}$$

where, $c_k=(\sigma^{(k)})^2(2+4\sqrt{\nu_k}+4\nu_k)+{}^{16\sigma^2_{\boldsymbol{s}^{(k)}}}/_{1-M\sigma^2_{\boldsymbol{s}^{(k)}}}\log{}^{2}/_{\nu_k}+2\sigma^4_{\boldsymbol{s}^{(k)}}L^2+2(\sigma^{(k)})^4L^2$ □

Remark 1. *While the proposed method and its theoretical results are based on* Variance Exploding *(VE) scheduling, they can be easily extended to* Variance Preserving *(VP) scheduling case (Karras et al., 2022).*

E.2 Theoretical Results with finite DC steps J

Lemma 19. *Let $H^{(k)}_{\text{dc}}:\boldsymbol{z}^{(k)}_{\text{ac}}\mapsto\boldsymbol{z}^{(k)}_{\text{dc}}$ denote the function corresponding to fine correction as defined in Algorithm 1 with finite J and $\eta^{(k)}\leq 2\sigma^2_{\boldsymbol{s}^{(k)}}$. Also, let $\pi^{(k)}=p(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}^{(k)}_{\text{ac}})$ be the stationary target distribution and $\tilde{\pi}^{(k)}_0$ be initial distribution used for the DC at iteration k. Then, with probability at least $1-e^{-\nu_k}$, the following holds for any $\boldsymbol{x},\boldsymbol{y}\in\mathcal{X}$ if ${}^{1}/_{\sigma^2_{\boldsymbol{s}^{(k)}}}<M_t$:*

$$\left\|(H^{(k)}_{\text{dc}}-I)(\boldsymbol{x})-(H^{(k)}_{\text{dc}}-I)(\boldsymbol{y})\right\|_2^2\leq(\delta^{(k)}_{\text{dc}})^2+(\epsilon^{(k)}_{\text{dc}})^2\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 \tag{125}$$

where,$0<\kappa<1$, $C>0$, $(\delta^{(k)}_{\text{dc}})^2=\frac{64d\sigma^2_{\boldsymbol{s}^{(k)}}}{(1-M_t\sigma^2_{\boldsymbol{s}^{(k)}})}\log\frac{2}{\nu_k}+C(1-\kappa)^{2J}\mathcal{W}_2^2(\tilde{\pi}^{(k)}_0,\pi^{(k)})+O\left((\eta^{(k)})^2\right)$, and $(\epsilon^{(k)}_{\text{dc}})^2=\left(\frac{2\sqrt{2}M_t\sigma^2_{\boldsymbol{s}^{(k)}}}{1-M_t\sigma^2_{\boldsymbol{s}^{(k)}}}\right)^2$.

Proof. With finite J, the langevin dynamics doesn't necessarily converge to the stationary distribution $\pi^{(k)} = p(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_{\text{ac}}^{(k)})$. Let $\tilde{\pi}_0^{(k)}$ be the initial distribution used to initialize the finite step langevin dynamics and $\tilde{\pi}^{(k)}$ be the distribution of the iterate after running finite J steps of langevin dynamics. Using (Dalalyan & Karagulyan, 2019)[Theorem 1], the following holds for $\eta^{(k)} \leq 2\sigma^2_{\boldsymbol{s}^{(k)}}$

$$\mathcal{W}_2(\tilde{\pi}^{(k)}, \pi^{(k)}) \leq (1-\kappa)^J \mathcal{W}_2(\tilde{\pi}_0^{(k)}, \pi^{(k)}) + O\left(\eta^{(k)}\right) \tag{126}$$

where, $0 < \kappa < 1$ and $\mathcal{W}_2$ is the 2-Wassertein distance.

Using Kantorovich and Rubinstein dual representation (Villani et al., 2008),

$$\|\mathbb{E}_{\tilde{\pi}^{(k)}}[\boldsymbol{z}] - \mathbb{E}_{\pi^{(k)}}[\boldsymbol{z}]\|_2 = \left\| \int \boldsymbol{z}_{\text{dc}}^{(k)} d\tilde{\pi}^{(k)}(\boldsymbol{z}) - \int \boldsymbol{z} d\pi^{(k)}(\boldsymbol{z}) \right\|_2 \tag{127}$$

$$\leq \mathcal{W}_1(\tilde{\pi}^{(k)}, \pi^{(k)}) \tag{128}$$

$$\leq \mathcal{W}_2(\tilde{\pi}^{(k)}, \pi^{(k)}) \tag{129}$$

The last inequality is due to the Holder's inequality Villani et al. (2008). Using triangle inequality leads to following:

$$\left\| \boldsymbol{z}_{\text{dc}}^{(k)} - \mathbb{E}_{\pi^{(k)}}[\boldsymbol{z}] \right\|_2 \leq \left\| \boldsymbol{z}_{\text{dc}}^{(k)} - \mathbb{E}_{\tilde{\pi}^{(k)}}[\boldsymbol{z}] \right\|_2 + \|\mathbb{E}_{\tilde{\pi}^{(k)}}[\boldsymbol{z}] - \mathbb{E}_{\pi^{(k)}}[\boldsymbol{z}]\|_2 \tag{130}$$

$$\leq \left\| \boldsymbol{z}_{\text{dc}}^{(k)} - \mathbb{E}_{\tilde{\pi}^{(k)}}[\boldsymbol{z}] \right\|_2 + (1-\kappa)^J \mathcal{W}_2(\tilde{\pi}_0^{(k)}, \pi^{(k)}) + O\left(\eta^{(k)}\right) \tag{131}$$

Using this result and following the same procedure as in Lemma 11, we get the theorem.

□

Theorem 4. *Suppose that the assumptions in Theorem 1, Assumption 2, and Assumption 3 hold. Further, assume that the DC steps finite steps J and $\eta^{(k)} \leq 2\sigma^2_{\boldsymbol{s}^{(k)}}$. Also, let $\pi^{(k)} = p(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_{\text{ac}}^{(k)})$ be the stationary target distribution and $\tilde{\pi}_0^{(k)}$ be initial distribution used for the DC at iteration k. Let $D_{\sigma^{(k)}} : \widetilde{\boldsymbol{z}}^{(k)} \mapsto \boldsymbol{z}_{\text{tw}}^{(k)}$ denote the AC-DC denoiser. Then, we have:*

With probability at least $1 - 2e^{-\nu_k}$, the following holds for iteration k of ADMM-PnP:

$$\|(D_{\sigma^{(k)}} - I)(\boldsymbol{x}) - (D_{\sigma^{(k)}} - I)(\boldsymbol{y})\|_2^2 \leq \epsilon_k^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \delta_k^2 \tag{132}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, $k \in \mathbb{N}^+$, a constant $0 < \kappa < 1$ and $C > 0$, when $\sigma^2_{\boldsymbol{s}^{(k)}} + (\sigma^{(k)})^2 < 1/M$ with

$$\epsilon_k^2 = 3((2\sqrt{2}M\sigma^2_{\boldsymbol{s}^{(k)}}/(1-M\sigma^2_{\boldsymbol{s}^{(k)}}))^2 + (\sigma^{(k)})^4 M^2) \tag{133}$$

$$\delta_k^2 = 3(2(\sigma^{(k)})^2(d + 2\sqrt{d\nu_k} + 2\nu_k) + {}^{64d\sigma^2_{\boldsymbol{s}^{(k)}}}/_{(1-M\sigma^2_{\boldsymbol{s}^{(k)}})} \log {}^2/_{\nu_k}) + C(1-\kappa)^{2J}\mathcal{W}_2^2(\tilde{\pi}_0^{(k)}, \pi^{(k)}) + O\left((\eta^{(k)})^2\right). \tag{134}$$

In other words, with $\nu_k = \ln {}^{2\pi}/_{6\eta} + 2\ln k$, the denoiser $D_{\sigma^{(k)}}$ satisfies part (a) for all $k \in \mathbb{N}^+$ with probability at least $1 - \eta$.

Proof. By substituting Lemma 11 with Lemma 19 leads to the theorem. □

Theorem 5. *Suppose that Assumptions 2-3 hold. Let $D := \text{diam}(\mathcal{X}) = \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{X}} \|\boldsymbol{x} - \boldsymbol{y}\|_2 < \infty$, $S := \inf_{\boldsymbol{x}\in\mathcal{X}} \|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_2 < \infty$ and define $L := MD + S$. Let $D_{\sigma^{(k)}} : \widetilde{\boldsymbol{z}}^{(k)} \mapsto \boldsymbol{z}_{\text{tw}}^{(k)}$ denote the AC-DC denoiser. Further, assume that the DC steps finite steps J and $\eta^{(k)} \leq 2\sigma^2_{\boldsymbol{s}^{(k)}}$. Also, let $\pi^{(k)} = p(\boldsymbol{z}_{\sigma^{(k)}}|\boldsymbol{z}_{\text{ac}}^{(k)})$ be the stationary target distribution and $\tilde{\pi}_0^{(k)}$ be initial distribution used for the DC at iteration k. Then, the following hold:*

***(Boundedness)** With probability at least $1 - 2e^{-\nu_k}$, the denoiser $D_{\sigma^{(k)}}$ is bounded at each iteration k i.e. $\frac{1}{d}\|(D_{\sigma^{(k)}} - I)(\boldsymbol{x})\|_2^2 \leq c_k^2$ whenever $\sigma^2_{\boldsymbol{s}^{(k)}} + (\sigma^{(k)})^2 < 1/M$, where $c_k = (\sigma^{(k)})^2(2 + 4\sqrt{\nu_k} + 4\nu_k) + {}^{32\sigma^2_{\boldsymbol{s}^{(k)}}}/_{(1-M\sigma^2_{\boldsymbol{s}^{(k)}})} \log {}^2/_{\nu_k} + C(1-\kappa)^{2J}\mathcal{W}_2^2(\tilde{\pi}_0^{(k)}, \pi^{(k)}) + O\left((\eta^{(k)})^2\right) + 4L^2\sigma^4_{\boldsymbol{s}^{(k)}} + 2(\sigma^{(k)})^4 L^2$, $0 < \kappa < 1$, $C > 0$ and $\nu_k > 0$.*

Let $\nu_k = \ln \frac{2\pi^2}{6\eta} + 2\ln k$ with $\eta \in (0, 1]$. Consequently, the denoiser $D_{\sigma^{(k)}}$ is bounded for all $k \in \mathbb{N}_+$ with corresponding c_k and probability at least $1 - \eta$.

Proof. The proof follows similar as in Theorem 3 by incorporating the effect of finite J in Lemma 16 as done in Lemma 19. □

F THEORETICAL RESULTS FOR ODE BASED DENOISER

Refer to the Zhang et al. (2024) for details on ODE based denoiser.

F.1 THEORETICAL RESULTS EQUIVALENT TO THEOREM 2

Lemma 20. *Let $H_{\text{ode}}^{(k)} : \boldsymbol{z}_{\text{dc}}^{(k)} \mapsto \boldsymbol{z}_{\text{ode}}^{(k)}$ denote the projection function using ode based denoiser (Karras et al., 2022) in Algorithm 1. Then, we have the following*

$$\left\|(H_{\text{ode}}^{(k)} - I)(\boldsymbol{x}) - (H_{\text{ode}}^{(k)} - I)(\boldsymbol{y})\right\|_2^2 \leq (\epsilon_{\text{ode}}^{(k)})^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + {\delta_{\text{ode}}^{(k)}}^2 \tag{135}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ with $(\epsilon_{\text{ode}}^{(k)})^2 = 2\left(\int_{t=t_{\sigma^{(k)}}}^{0} \left(\sigma(t)\sigma'(t)M_t\right)^2 dt\right)$, and ${\delta_{\text{ode}}^{(k)}}^2 = 0$.

Proof. Then, the difference of residual of ode projection i.e. $R_{\text{ode}}^{(k)} = H_{\text{ode}}^{(k)} - I$ can be bounded as

$$\begin{aligned}
\left\|R_{\text{ode}}^{(k)}(\boldsymbol{x}) - R_{\text{ode}}^{(k)}(\boldsymbol{y})\right\|_2^2 =& \left\|\int_{t=t_{\sigma^{(k)}}}^{0} -\sigma(t)\sigma'(t)(\nabla \log p_t(\boldsymbol{x}) - \nabla \log p_t(\boldsymbol{y}))dt\right\|_2^2 \\
\leq& 2\left\|\int_{t=t_{\sigma^{(k)}}}^{0} -\sigma(t)\sigma'(t)(\nabla \log p_t(\boldsymbol{x}) - \nabla \log p_t(\boldsymbol{y}))dt\right\|_2^2 \\
\leq& 2\int_{t=t_{\sigma^{(k)}}}^{0} \left(\sigma(t)\sigma'(t)\right)^2 \|(\nabla \log p_t(\boldsymbol{x}) - \nabla \log p_t(\boldsymbol{y}))\|_2^2 dt \\
\leq& 2\int_{t=t_{\sigma^{(k)}}}^{0} \left(\sigma(t)\sigma'(t)\right)^2 M_t^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 dt \\
\leq& 2\left(\int_{t=t_{\sigma^{(k)}}}^{0} \left(\sigma(t)\sigma'(t)\right)^2 M_t^2 dt\right) \|\boldsymbol{x} - \boldsymbol{y}\|_2^2
\end{aligned} \tag{136}$$

□

Theorem 6. *Suppose that the assumptions in Theorem 1, Assumption 2 and Assumption 3 hold. Further, assume that the step size satisfies $\eta^{(k)} \to 0$ and the number of iterations $J \to \infty$. Let $D_{\sigma^{(k)}} : \widetilde{\boldsymbol{z}}^{(k)} \mapsto \boldsymbol{z}_{\text{tw}}^{(k)}$ denote the AC-DC denoiser. Then, we have:*

(a) With probability at least $1 - 2e^{-\nu_k}$, the following holds for iteration k of ADMM-PnP:

$$\|(D_{\sigma^{(k)}} - I)(\boldsymbol{x}) - (D_{\sigma^{(k)}} - I)(\boldsymbol{y})\|_2^2 \leq \epsilon_k^2 \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \delta_k^2 \tag{137}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ and $k \in \mathbb{N}^+$ when $\sigma_{\boldsymbol{s}^{(k)}}^2 + (\sigma^{(k)})^2 < 1/M$ with

$$\epsilon_k^2 = 3(\sqrt{2}M\sigma_{\boldsymbol{s}^{(k)}}^2 / 1 - \sigma_{\boldsymbol{s}^{(k)}}^2 M)^2 + 6\int_{t=t_{\sigma^{(k)}}}^{0} \left(\sigma(t)\sigma'(t)\right)^2 M_t^2 dt)) \tag{138}$$

$$\delta_k^2 = 3(2(\sigma^{(k)})^2(d + 2\sqrt{d\nu_k} + 2\nu_k) + 32d\sigma_{\boldsymbol{s}^{(k)}}^2 / (1 - M\sigma_{\boldsymbol{s}^{(k)}}^2) \log 2/\nu_k). \tag{139}$$

In other words, if $\nu_k = \ln 2\pi/6\eta + 2nk$, the denoiser $D_{\sigma^{(k)}}$ satisfies part (a) for all $k \in \mathbb{N}^+$ with probability at least $1 - \eta$.

(b) Assume that $\sigma^{(k)}$ is scheduled such that $\lim_{k\to\infty}(\sigma^{(k)})^2\nu_k = 0$ for $\nu_k = \ln 2\pi/6\eta + 2nk$, $\epsilon < 1$, and $\epsilon/\mu(1+\epsilon-2\epsilon^2) < 1/\rho$ all hold, where $\epsilon = \lim_{k\to\infty} \sup \epsilon_k$ with ϵ_k defined in (138). Consequently, $\delta = \lim_{k\to\infty} \sup \delta_k$ is finite and ADMM-PnP with the AC-DC denoiser with ode based denoiser converges to an r-ball (see r in Theorem 1) with probability at least $1 - \eta$.

Proof. The proof follow similar to the proof of Theorem 2 in Appendix D.1 with the residual bound of Tweedie's lemma replaced by Lemma 20. □

F.2 Theoretical Results equivalent to Theorem 3

Lemma 21. *Let $H_{\text{ode}}^{(k)} : \boldsymbol{z}_{\text{dc}}^{(k)} \mapsto \boldsymbol{z}_{\text{ode}}^{(k)}$ denote the projection function using ode based denoiser (Karras et al., 2022) in Algorithm 1. Assume $\|\nabla \log p_{\text{data}}(\boldsymbol{x})\|_\infty \leq L,\ \forall \boldsymbol{x} \in \mathcal{X}$. Then, we have the following*

$$\frac{1}{d}\left\|(H_{\text{ode}}^{(k)} - I)(\boldsymbol{x})\right\|_2^2 \leq L^2 \int_{t=t_{\sigma(k)}}^{0} (\sigma(t)\sigma'(t)^2 dt \tag{140}$$

for any $\boldsymbol{x} \in \mathcal{X}$.

Proof. Then, the residual of ode projection i.e. $R_{\text{ode}}^{(k)} = H_{\text{ode}}^{(k)} - I$ can be bounded as

$$\begin{aligned}
\left\|R_{\text{ode}}^{(k)}(\boldsymbol{x})\right\|_2^2 =& \left\|\int_{t=t_{\sigma(k)}}^{0} -\sigma(t)\sigma'(t)\nabla \log p_t(\boldsymbol{x}) dt\right\|_2^2 \\
\leq& d \cdot L^2 \left\|\int_{t=t_{\sigma(k)}}^{0} -\sigma(t)\sigma'(t) dt\right\|_2^2 \\
\leq& d \cdot L^2 \left\|\int_{t=t_{\sigma(k)}}^{0} -\sigma(t)\sigma'(t) dt\right\|_2^2 \\
\leq& d \cdot L^2 \int_{t=t_{\sigma(k)}}^{0} (\sigma(t)\sigma'(t)^2 dt
\end{aligned} \tag{141}$$

□

A theorem analogous to Theorem 3 can also be obtained for ODE-based denoiser. The only difference lies in the expression of the constant c_k, which in this case becomes

$$c_k = (\sigma^{(k)})^2(2 + 4\sqrt{\nu_k} + 4\nu_k) + {}^{16\sigma_{\boldsymbol{s}^{(k)}}^2}/_{1 - M\sigma_{\boldsymbol{s}^{(k)}}^2} \log {}^2/_{\nu_k} + 2\sigma_{\boldsymbol{s}^{(k)}}^4 L^2 + 2L^2 \int_{t=t_{\sigma(k)}}^{0} (\sigma(t)\sigma'(t)^2 dt.$$

G Usage of Large Language Models (LLM)

An LLM was used solely to assist with polishing the writing. LLM played no part in the experiments, results and conclusion.

H Experimental Details

H.1 Hyperparameter Settings

We adopt a linear schedule for $\sigma^{(k)}$ with range $[0.1, 10]$ over W decay window i.e. $\sigma^{(k)} = \max(0.1, 10 - (10 - 0.1) \cdot k/W)$. The maximal number of iterations for our proposed method is set to $K = W + 10$. At iteration k, we use $J = 10$ DC steps, and the schedules $\eta^{(k)} = 5 \times 10^{-4}\sigma^{(k)}$ and $\sigma_{\boldsymbol{s}^{(k)}} = {}^{0.1}/_{\sqrt{\sigma^{(k)}}}$. We use gradient descent with Adam optimizer (Kingma & Ba, 2015) for solving each regularized maximum likelihood subproblem (7a). This subproblem is optimized for maximum of 1000 iterations with convergence detected when the loss value increases more than $\Delta_{\text{tol}} = 1 \times 10^{-1}$ consecutively for 3 iterations window. We conduct our experiment with two variants based on the third stage: using Tweedie's lemma (denoted as "Ours-tweedie") and a 10-step ODE based denoiser (Zhang et al., 2024; Karras et al., 2022) (denoted as "Ours-ode"). We use the preconditioning in Karras et al. (2022) while using the pretrained diffusion models.

H.2 Details on task specific data-fidelity loss ℓ

We use *mean square error* (MSE) as the data-fidelity loss for every task i.e.

$$\ell(\boldsymbol{y}||\boldsymbol{x}) = -\log p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{2\sigma_n^2}\|\boldsymbol{y} - \mathcal{A}(\boldsymbol{x})\|_2^2 \tag{142}$$

Table 3: Hyperparameter settings for each task

Task	ρ	W	lr of Adam in (7a)
Superresolution ($4\times$)	100	100	3×10^{-2}
Gaussian Deblur	100	100	5×10^{-2}
HDR	500	100	3×10^{-2}
Inpainting (Random)	500	100	1×10^{-1}
Inpainting (Box)	500	100	1×10^{-1}
Motion Deblur	100	100	1×10^{-1}
Nonlinear Deblur	300	400	3×10^{-1}
Phase Retrieval	100	400	1×10^{-1}

H.3 Details on pretrained diffusion models

The pretrained models provided in Chung et al. (2023) are used in our experiment. Refer to Chung et al. (2023) for more details on these pretrained models.

H.4 Baseline Details

Unless mentioned otherwise, we conduct the experiments in the default settings of their original implementation except for maintaining consistency within the measurement operators.

- **DDRM (Kawar et al., 2021):** We use 20 steps DDIM with $\eta = 0.85$ and $\eta_b = 1$ as specified in Kawar et al. (2022).
- **DPS (Chung et al., 2023) :** The original implementation is ran in their default settings.
- **DiffPIR (Zhu et al., 2023):** The default settings are adopted in the experiments.
- **RED-diff (Mardani et al., 2024):** We use $\lambda = 0.25$ and $lr = 0.5$ as specified in the paper.
- **DAPS (Zhang et al., 2024):** We use the best performing DAPS-4K version as proposed in the paper.
- **DPIR (Zhang et al., 2022)**: We employ "drunet_color" as PnP denoiser, while keeping all the other settings at their default values.
- **DCDP (Li et al., 2025)**: All the setttings are set to their default values.
- **PMC (Sun et al., 2024)**: PMC was proposed using different score models for two different tasks with relatively high measurement SNR. For a fair comparison, we used our own implementation with the same score model checkpoints as our methods, and further tuned this method accordingly.

H.5 Evaluation Metrics

For all the methods, we use the implementation of PSNR, SSIM, and LPIPS provided in *piq* python package. The default settings for these metrics are used except the average pooling enabled for LPIPS.

H.6 Computation Resource Details

All the experiments were run on a instance equipped with one Nvidia H100 GPU, 20 cores of 2.0 Ghz Intel Xeon Platinum 8480CL CPU, and 64 GB of RAM.

I Additional Experimental Results

I.1 Illustration of proposed Denoiser

Figure 6 illustrates the effect of the proposed correction-denoising procedure. The noisy input image $\widetilde{\boldsymbol{z}}^{(k)}$ typically lies far away from the Gaussian noise manifold, leading to poor denoising

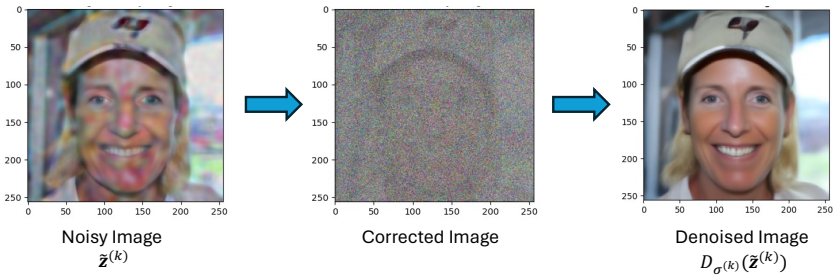

Figure 6: Illustration of correction and denoising step in proposed method.

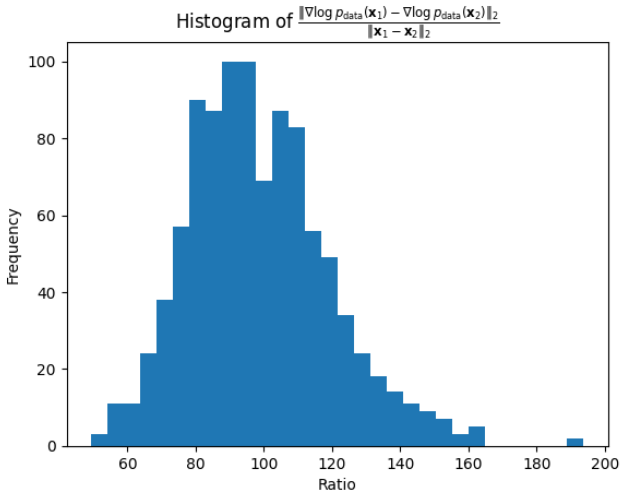

Figure 7: Histogram of score difference norm ratio: $\|\nabla \log p_{\text{data}}(\boldsymbol{x}_2) - \nabla \log p_{\text{data}}(\boldsymbol{x}_1)\|_2 / \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2$, illustrating empirical smoothness (Assumption 2).

performance if directly used. To address this mismatch, our method first performs correction to effectively gaussianize the noise which is then denoised using Tweedie's lemma or ode-style score integration, producing a high-quality clean reconstruction.

I.2 Empricial validation of Assumption 2 and Assumption 3

To assess the practicality of the smoothness and coercivity assumptions used in Theorem 2, we conduct two diagnostic experiments using a pretrained score model on the validation split of the FFHQ dataset. These experiments are designed to evaluate (i) the empirical Lipschitz behavior of the score function $\nabla \log p_{\text{data}}(\boldsymbol{x})$ (Assumption 2), and (ii) the coercivity of the energy landscape $-\log p_{\text{data}}(\boldsymbol{x})$ (Assumption 3).

Empirical smoothness of the score. We randomly 1000 samples of $\boldsymbol{x}_1, \boldsymbol{x}_2$ and compute score differences $\|\nabla \log p_{\text{data}}(\boldsymbol{x}_1) - \nabla \log p_{\text{data}}(\boldsymbol{x}_2)\|_2$ and image differences $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2$. Figure 7 plots the histogram of their ratio. The distribution concentrates around a finite value (mostly between 50 and 160), indicating that the score behaves approximately M-Lipschitz with a moderate empirical constant. This supports the smoothness requirement in Assumption 2.

Empirical coercivity. To evaluate coercivity, we scale images by factors $c \in \{1, 1.5, 2, 3\}$ and measure the quantity $\langle \boldsymbol{x}, -\nabla \log p_{\text{data}}(\boldsymbol{x}) \rangle$ as a function of the squared image norm $\|\boldsymbol{x}\|_2^2$. As shown in Figure 8, the inner product grows approximately linearly with $\|\boldsymbol{x}\|_2^2$, indicating that the learned score consistently pulls large-norm images back toward the data manifold. This behavior is consistent with the coercivity structure assumed in Assumption 3.

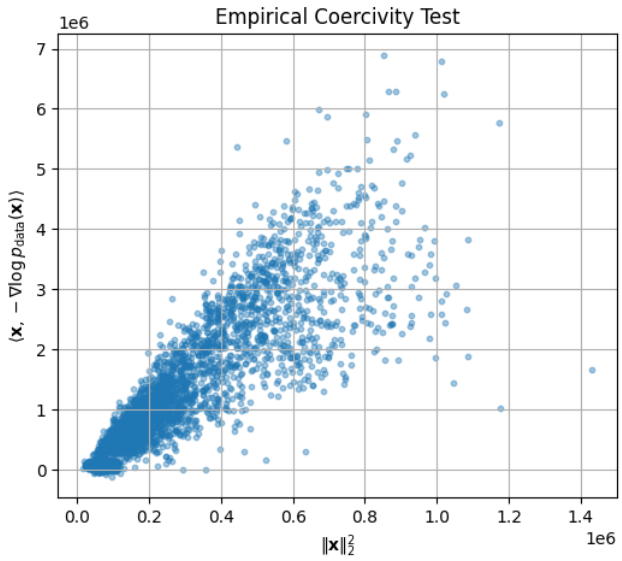

Figure 8: Empirical coercivity test: relationship between $\langle \boldsymbol{x}, -\nabla \log p_{\text{data}}(\boldsymbol{x}) \rangle$ and $\|\boldsymbol{x}\|_2^2$. The strong positive correlation indicates coercive energy behavior (Assumption 3).

Together, these empirical diagnostics demonstrate that the theoretical assumptions employed in our analysis hold approximately in practice and therefore justify the use of DC correction in our AC–DC algorithm.

I.3 Illustration of Usage of Additional Regularization

To further demonstrate the flexibility of integrating diffusion-based PnP denoisers within the ADMM framework, we present an example where we employ an additional perceptual regularization term which will be handled in the maximum-likelihood (ML) step. In particular, the $\boldsymbol{x}$-update step of ADMM with an LPIPS perceptual regularization (Zhang et al., 2018) becomes:

$$\boldsymbol{x}^{(k+1)} = \arg\min_{\boldsymbol{x}} \frac{1}{\rho}\ell(y\|\mathcal{A}(\boldsymbol{x})) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}^{(k)} + \boldsymbol{u}^{(k)}\|_2^2 + \lambda_{\text{lpips}}\,\text{LPIPS}_{\text{VGG}}(\boldsymbol{x}, \boldsymbol{x}_{\text{ref}}), \tag{143}$$

where $\boldsymbol{x}_{\text{ref}}$ is the reference image and λ_{lpips} controls the perceptual strength.

This example highlights the flexibility of the proposed method: unlike traditional diffusion-based PnP approaches that struggle in the presence of dual variables, our design enables seamless incorporation of additional regularization terms. In Fig. 9 we illustrate box inpainting reconstruction task with the perceptual LPIPS-VGG regularization which enhances semantic content consistency while allowing visual style transfer from the reference images.

I.4 Results on additional tasks

The results on additional two tasks: hdr and nonlinear deblurring are presented in the Table 4.

I.5 Ablation study

We perform the ablation study on the significance of our proposed correction steps. The results are presented in the Table 5.

I.6 Influence of decay schedule and NFE efficiency

In our ADMM-PnP scheme, the size of the decay window for $\sigma^{(k)}$ determines the total number of iterations – and thus the speed of convergence. A shorter window (small W) drives $\sigma^{(k)}$ down more

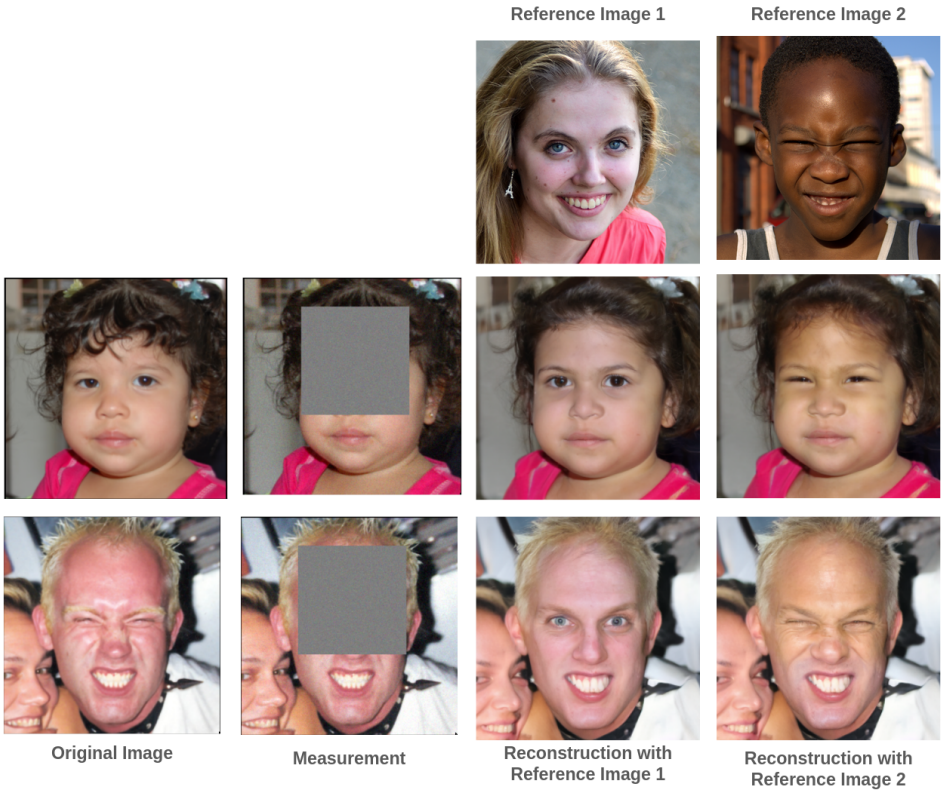

Figure 9: Demonstration of incorporating additional perceptual regularization.

Table 4: Reconstruction metrics (100 images) on FFHQ / ImageNet for additional tasks. **Bold**: best, blue: 2nd best.

		FFHQ			ImageNet		
Task	Method	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
HDR	Ours–tweedie	**27.425**	**0.853**	0.164	26.515	**0.817**	0.182
	DAPS	26.94	0.852	**0.154**	**26.848**	0.816	**0.172**
	RED-diff	26.815	0.836	0.241	20.794	0.771	0.232
	PMC	21.582	0.707	0.291	22.745	0.707	0.290
Nonlinear Deblur	Ours–tweedie	**29.326**	**0.823**	0.185	**27.837**	0.725	0.212
	DAPS	28.598	0.782	**0.172**	27.745	**0.739**	**0.201**
	DPS	23.746	0.668	0.276	22.724	0.543	0.394
	RED-diff	26.9	0.72	0.234	25.488	0.72	0.207
	PMC	21.102	0.0623	0.354	22.347	0.533	0.430

Table 5: Comparison of our method with and without correction on FFHQ. Best results are highlighted in bold.

	Ours-tweedie without correction			Ours-tweedie with correction		
Tasks	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Superresolution (4x)	26.915	0.730	0.314	**30.439**	**0.857**	**0.178**
Gaussian Blur	28.896	0.788	0.275	**30.402**	**0.853**	**0.175**
Inpainting (Box)	15.604	0.617	0.361	**24.025**	**0.859**	**0.131**
Motion Deblur	25.123	0.538	0.370	**30.003**	**0.854**	**0.179**
Nonlinear Deblur	21.731	0.561	0.375	**29.326**	**0.823**	**0.185**
Phase Retrieval	11.978	0.181	0.726	**27.944**	**0.793**	**0.209**

quickly, often reaching convergence in fewer steps but at the risk of settling in a suboptimal local minimum.

To study this trade-off, we sweep

$$W \in \{5, 10, 50, 100, 200, 300, 400, 500\}$$

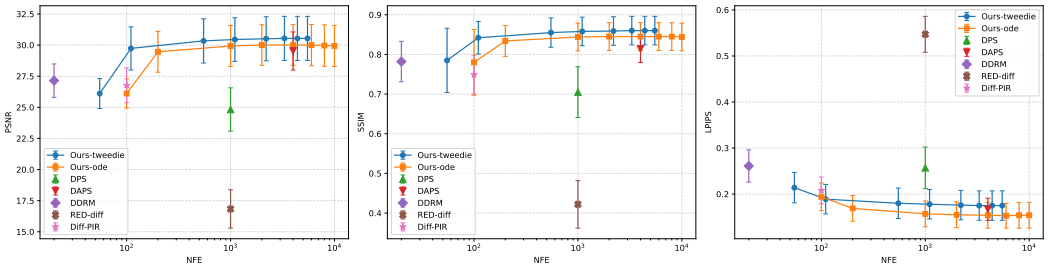

Figure 10: Performance with respect to NFE for Superresolution task (FFHQ)

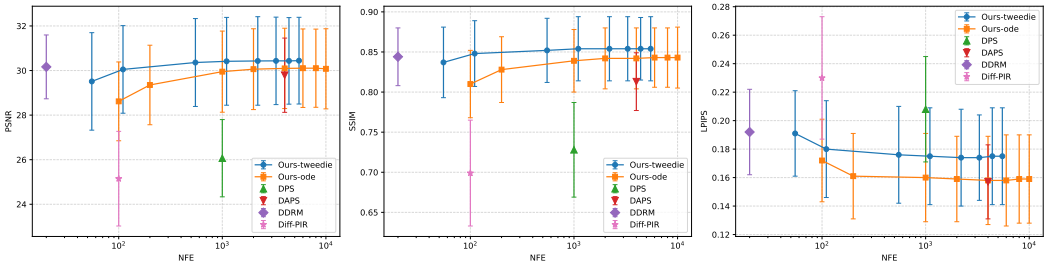

Figure 11: Performance with respect to NFE for Gaussian deblurring task (FFHQ)

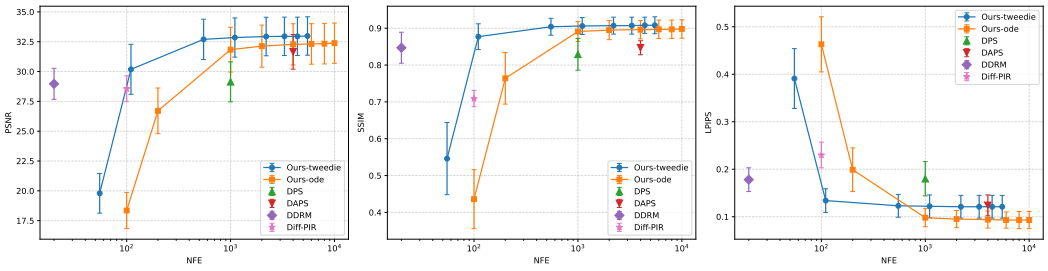

Figure 12: Performance with respect to NFE for Inpainting with random missings (FFHQ)

for each task. Since each iteration of Ours-tweedie uses 11 score evaluations (10 for the DC update and 1 for the Tweedie's lemma based denoiser), these W value translate to

$$\text{Number of Function Evaluations (NFE)} = \{55, 110, 550, 1100, 2200, 3300, 4400, 5500\}$$

By contrast, each Ours-ode iteration costs 20 NFEs, giving

$$\text{Number of Function Evaluations (NFE)} = \{100, 200, 1000, 2000, 4000, 6000, 8000, 10000\}$$

Figures 10-15 plot mean$\pm$std. (standard deviation) performance of our methods and all baselines against NFE over 100 images of FFHQ dataset. For most tasks, quality saturates after just 10 iterations (110 NFE for Ours-tweedie, 200 NFE for Ours-ode), showing a rapid decay schedule suffices to achieve near-peak results. However, on the hardest inverse problems (phase retrieval and nonlinear blur), gradually decaying noise (larger W) and more NFEs yield significantly better reconstructions–far outpacing every baseline. Thus, while aggressive schedules excel on simple tasks, challenging problems benefit from extended iteration and gentler annealing; given enough NFEs, our approach establishes state-of-the-art performance across most of the tasks.

I.7 More qualitative results

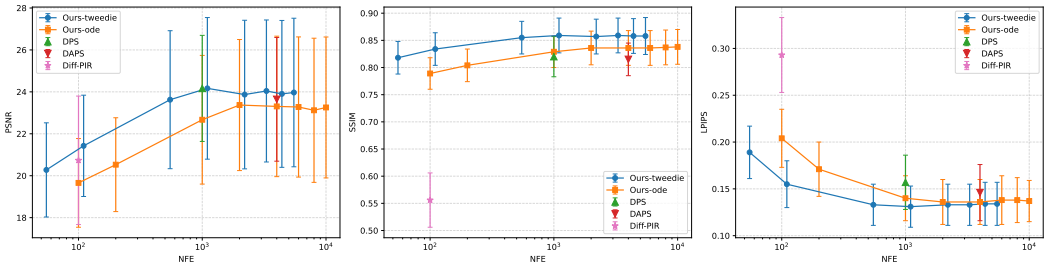

Figure 13: Performance with respect to NFE for Inpainting with box missing (FFHQ)

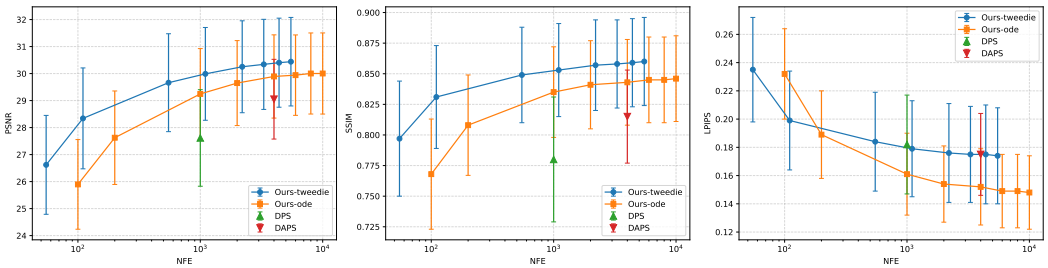

Figure 14: Performance with respect to NFE for Motion blur (FFHQ)

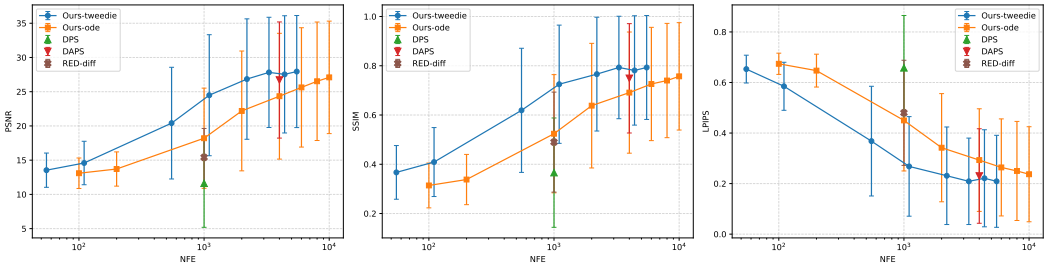

Figure 15: Performance with respect to NFE for Phase retrieval (FFHQ)

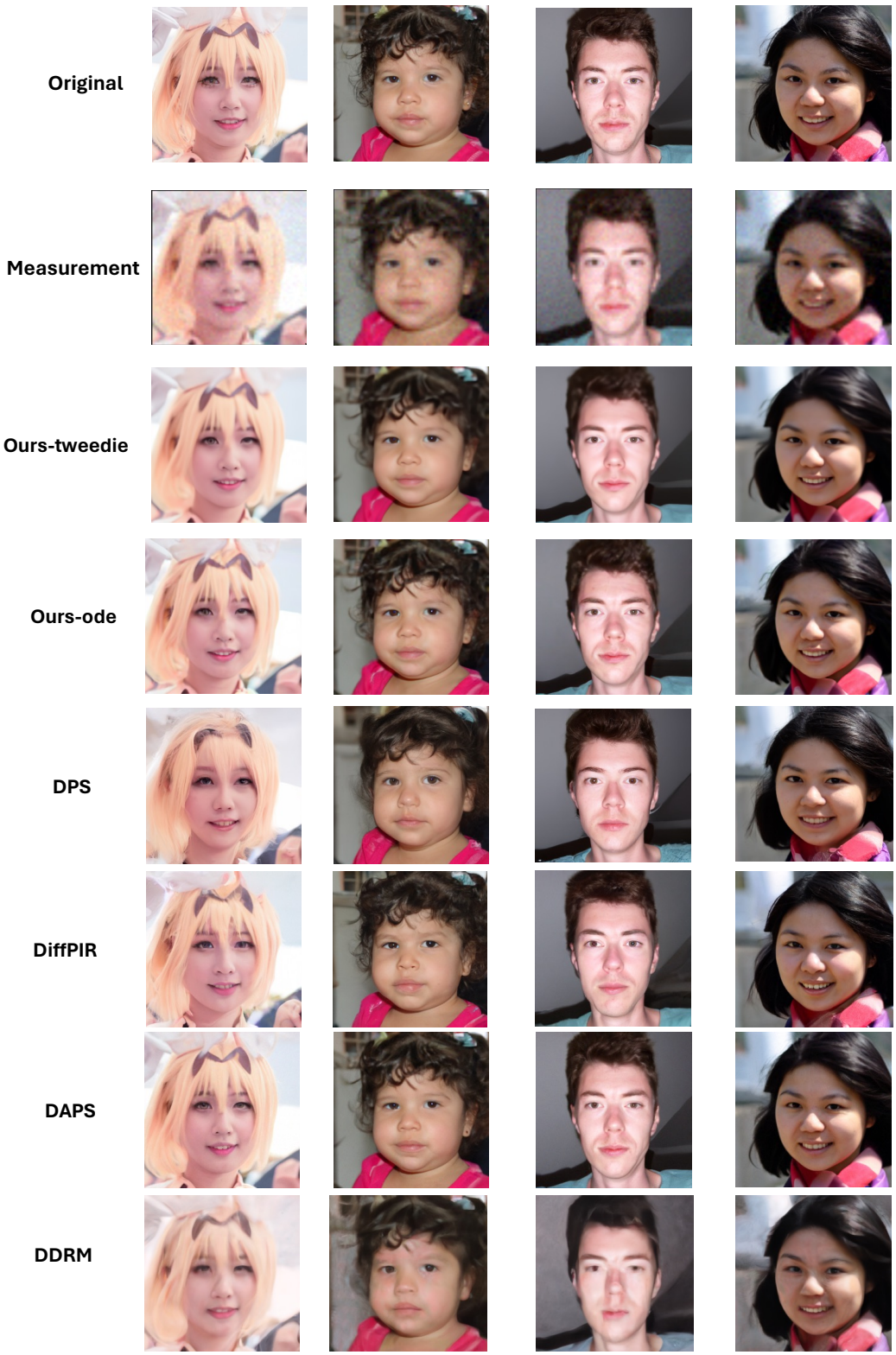

Figure 16: Recovery under $4\times$ superresolution task on FFHQ

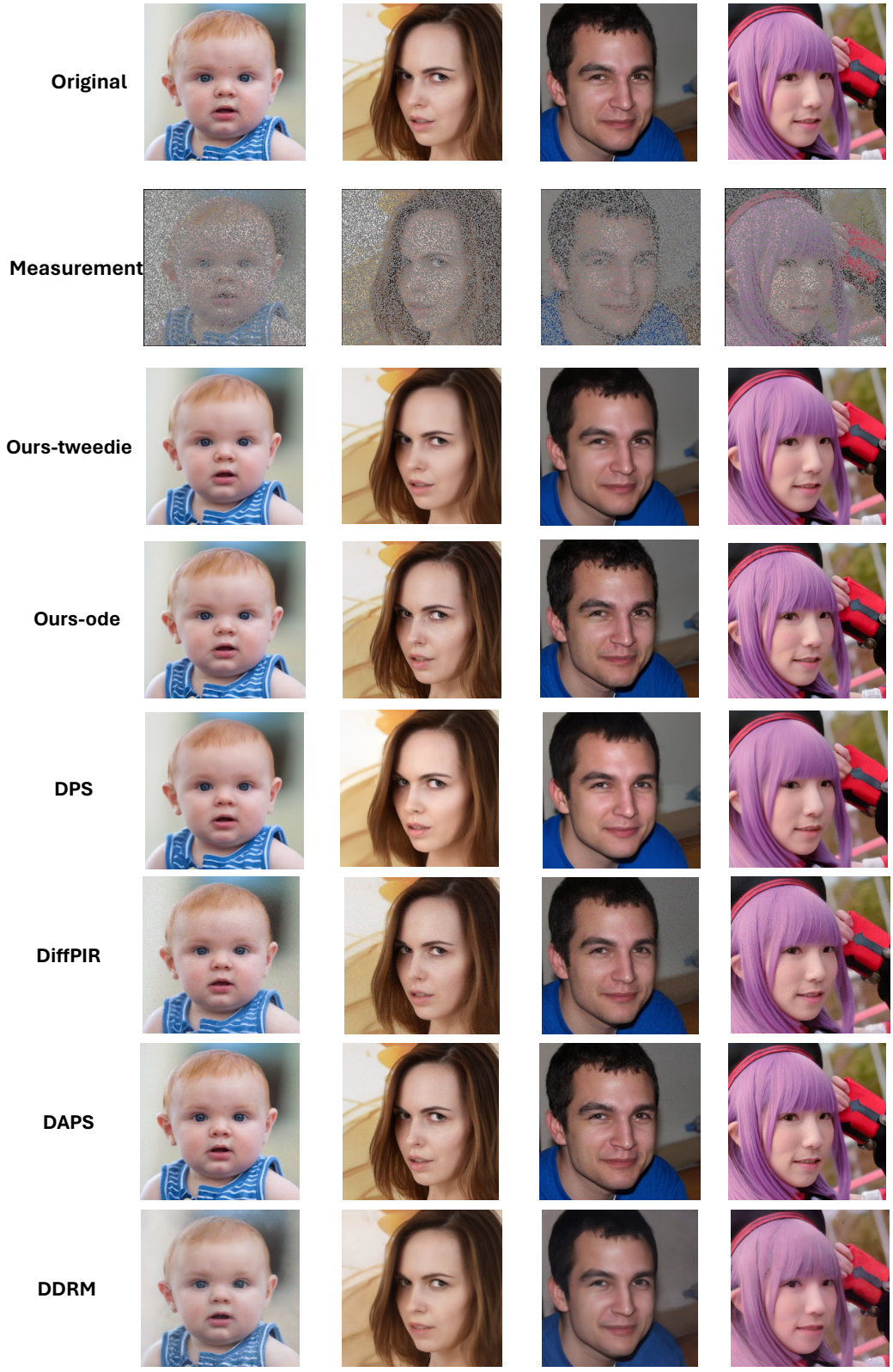

Figure 17: Recovery under inpainting with random missings on FFHQ

Figure 18: Recovery under motion blur task on FFHQ