Frequency-Guided Posterior Sampling for Diffusion-Based Image Restoration

Darshan Thaker
University of Pennsylvania
dbthaker@seas.upenn.edu

Abhishek Goyal
University of Pennsylvania
abhi2358@seas.upenn.edu

René Vidal
University of Pennsylvania
vidalr@seas.upenn.edu

Abstract

Image restoration aims to recover high-quality images from degraded observations. When the degradation process is known, the recovery problem can be formulated as an inverse problem, and in a Bayesian context, the goal is to sample a clean reconstruction given the degraded observation. Recently, modern pretrained diffusion models have been used for image restoration by modifying their sampling procedure to account for the degradation process. However, these methods often rely on certain approximations that can lead to significant errors and compromised sample quality. In this paper, we propose a simple modification to existing diffusion-based restoration methods that exploits the frequency structure of the reverse diffusion process. Specifically, our approach, denoted as Frequency Guided Posterior Sampling (FGPS), introduces a time-varying low-pass filter in the frequency domain of the measurements, progressively incorporating higher frequencies during the restoration process. We provide the first rigorous analysis of the approximation error of FGPS for linear inverse problems under distributional assumptions on the space of natural images, demonstrating cases where previous works can fail dramatically. On real-world data, we develop an adaptive curriculum for our method's frequency schedule based on the underlying data distribution. FGPS significantly improves performance on challenging image restoration tasks including motion deblurring and image dehazing. The source code is available at https://github.com/darshanthaker/FGPS.

1. Introduction

Image restoration is a fundamental problem in computer vision whose goal is to recover high-quality images from degraded observations. Many image restoration problems can be cast as inverse problems, where the task is to recover an underlying signal $\mathbf{x}_0 \in \mathbb{R}^n$ from noisy measurements $\mathbf{y} \in \mathbb{R}^m$ that are obtained from $\mathbf{x}_0$ via

$$\mathbf{y} = \mathcal{A}(\mathbf{x}_0) + \mathbf{z}. \tag{1}$$

In the above, $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ denotes the forward operator, which we assume to be known, and $\mathbf{z} \in \mathbb{R}^m$ is additive Gaussian noise. In most applications, recovering $\mathbf{x}_0$ is difficult because the forward operator is a many-to-one mapping. To address this challenge, researchers incorporate prior knowledge about $\mathbf{x}_0$, including sparsity [5, 16], low-rank structure [4, 6], total variation [32], or deep generative priors [3, 29]. One way of enforcing this knowledge is by solving a regularized optimization problem where the objective is to find an $\mathbf{x}_0$ that maximizes the likelihood under the chosen prior model while satisfying the constraint $\mathbf{y} = \mathcal{A}(\mathbf{x}_0)$. Another approach is to follow the Bayesian viewpoint where instead of recovering an exact $\mathbf{x}_0$, we sample from $p(\mathbf{x} \mid \mathbf{y})$ where Equation (1) is the model for $p(\mathbf{y} \mid \mathbf{x}_0)$ [38]. The advantage of this interpretation is that we can draw multiple plausible samples. In image restoration tasks, the diversity of samples can help the user pick a visually appealing reconstruction.

Recently, diffusion models [20, 36] have shown tremendous success in modeling complex probability distributions, allowing one to generate diverse and visually coherent samples in image, video, and audio tasks [14, 21, 25]. During training, diffusion models learn to reverse a gradual noising process that transforms data $\mathbf{x}_0$ to pure noise $\mathbf{x}_T$ over T steps, where $\mathbf{x}_t$ denotes an intermediate step of the reverse diffusion process. At test time, they generate samples by iteratively denoising random noise, progressively refining it into coherent data. Naturally, these models can also be used to develop samplers for $p(\mathbf{x} \mid \mathbf{y})$ in inverse problem settings. However, while one can train a problem-specific diffusion model to sample directly from $p(\mathbf{x} \mid \mathbf{y})$ [33], this requires training new diffusion models for every inverse problem, limiting the practical use of such approaches.

A popular approach which does not require training a new diffusion model for $p(\mathbf{x} \mid \mathbf{y})$ is to modify the sampling process of a pretrained unconditional diffusion model so that it produces samples that are consistent with the given measurement $\mathbf{y}$ [9, 24, 35, 37, 41]. These methods are based on decomposing the conditional score function $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{y})$ as $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t)$ using the Bayes rule. While the first term is learned by a

pretrained unconditional diffusion model, the second term is analytically intractable in general because $\mathbf{x}_t$ is the result of a complex denoising process. Thus, in practice, we use approximations for this term, typically by first computing $\boldsymbol{\mu}_{0|t} = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]$ using the learned diffusion model, known as the Tweedie estimate [17], and following the gradient of a loss that measures the deviation between $\mathcal{A}(\boldsymbol{\mu}_{0|t})$ and $\mathbf{y}$. While intuitive, the approximation error from the Tweedie estimate is difficult to control. Previous work has attempted to address this by designing heuristics that ensure that the modifications to $\mathbf{x}_t$ respect the intermediate data manifold learned by the diffusion model. However, these approaches make strong assumptions about the data manifold's structure, such as linearity, or require that $\boldsymbol{\mu}_{0|t}$ remains on the clean data manifold throughout all iterations [10, 42].

Our work aims to better understand and control this approximation error by analyzing images in the frequency domain. Our main contributions are:

1. We develop a new unsupervised diffusion-based image restoration algorithm, denoted as **F**requency **G**uided **P**osterior **S**ampling (FGPS). Our algorithm leverages the observation that diffusion models generate images hierarchically in the frequency domain [30]. Thus, to control the approximation error from estimating $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t)$, we apply a time-varying low-pass filter to $\mathbf{y}$ which progressively includes higher frequency ranges during the reverse diffusion process.
2. We then perform a precise theoretical analysis of the approximation errors in posterior sampling methods for linear inverse problems under distributional assumptions on the data. Next, exploiting the frequency distribution of natural images, we give precise examples of forward operators where FGPS has small approximation error compared to prior work.
3. In practice, we develop a data-dependent curriculum for the low-pass filters which allows the reverse process to adapt to spectral properties of the data distribution. While being easy to implement, our algorithm leads to a significant boost in performance on challenging linear and nonlinear image restoration problems, such as motion deblurring and image dehazing.

2. Background

2.1. Diffusion Models

Diffusion models are a class of score-based generative models that sample from a given data distribution by learning to reverse a gradual noising process [20, 36]. This noising process, known as the forward process, is represented in continuous time as an Itô Stochastic Differential Equation (SDE) for $t \in [0, 1]$

$$d\mathbf{x}_t = -\frac{1}{2}\beta(t)\mathbf{x}_t dt + \sqrt{\beta(t)} d\mathbf{w}_t. \tag{2}$$

The SDE is characterized by its drift coefficient $-\frac{1}{2}\beta(t)$ and diffusion coefficient $\sqrt{\beta(t)}$, with $\mathbf{w}_t$ representing the standard Wiener process over the interval $[0, 1]$. The initial condition of this SDE is a datapoint $\mathbf{x}_0 \sim p(\mathbf{x})$. The drift and diffusion coefficients are chosen such that $\mathbf{x}_t$ goes from the clean data distribution at $t = 0$ to pure noise $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ at $t = 1$. A special case is known as the Variance-Preserving SDE (VP-SDE), where we can write the transition density $p_t(\mathbf{x}_t \mid \mathbf{x}_0)$ as

$$p(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I}), \tag{3}$$

for $\bar{\alpha}_t = e^{\int_0^t \beta(s)\, ds}$ [20].

The reverse process aims to remove the noise from $\mathbf{x}_1$ and draw a sample from $\mathbf{x}_0$ following the reverse-time SDE

$$d\mathbf{x}_t = \left[-\frac{\beta(t)}{2}\mathbf{x}_t - \beta(t)\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\right] dt + \sqrt{\beta(t)} d\bar{\mathbf{w}}_t, \tag{4}$$

where t and the new Wiener process $\bar{\mathbf{w}}_t$ both run in reverse. A seminal result of Anderson gives that this reverse process has the same probability distribution as the forward process for every t [1]. In order to estimate the time-dependent score function $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$, diffusion models use the denoising score matching technique to train a neural network $s_\theta(\cdot)$ that predicts the noise added to a datapoint, which approximates the score function [40]. Sampling from $p(\mathbf{x})$ is performed by solving a discrete-time version of Equation (4) over T steps, where $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ [20, 37].

2.2. Posterior Sampling

In inverse problems, we are given measurements $\mathbf{y}$ as in Equation (1), and the goal is to recover the underlying clean $\mathbf{x}_0$ given $\mathbf{y}$. Throughout the paper, we will refer to $\mathcal{A}$ as the known forward operator. Because $\mathcal{A}$ is not invertible in general, the inverse problem can be highly ill-posed, necessitating data modelling assumptions.

A popular Bayesian strategy to solve the problem is to develop samplers for $p(\mathbf{x} \mid \mathbf{y})$ assuming we have a prior for $p(\mathbf{x})$ modeled by a pretrained generative model such as a diffusion model. For instance we can extend Equation (4) to sample from $p(\mathbf{x} \mid \mathbf{y})$ by using a conditional score function

$$d\mathbf{x}_t = \left[-\frac{\beta(t)}{2}\mathbf{x}_t - \beta(t)\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{y})\right] dt + \sqrt{\beta(t)} d\bar{\mathbf{w}}_t. \tag{5}$$

While one could train a conditional diffusion model to approximate the conditional score, this would require retraining models for every inverse problem we wish to solve [33].

Instead, we use Bayes rule to write the conditional score as

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{y}) = \underbrace{\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)}_{\text{unconditional score}} + \underbrace{\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t)}_{\text{noisy likelihood score}} \tag{6}$$

While a pretrained diffusion model effectively approximates the unconditional score, the noisy likelihood score remains intractable due to its factorization

$$p(\mathbf{y} \mid \mathbf{x}_t) = \int p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t) d\mathbf{x}_0, \tag{7}$$

where $p(\mathbf{x}_0 \mid \mathbf{x}_t)$ represents the complex distribution of denoised estimates of $\mathbf{x}_t$. Even though $p(\mathbf{x}_0 \mid \mathbf{x}_t)$ can be complex, the Tweedie's estimate gives us the posterior mean $\boldsymbol{\mu}_{0|t} = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]$ as a function of the unconditional score function [17] (refer to Appendix A for more background). Thus, some unsupervised methods approximate $p(\mathbf{x}_0 \mid \mathbf{x}_t)$ using $\boldsymbol{\mu}_{0|t}$, for example as a Dirac delta around $\boldsymbol{\mu}_{0|t}$ [9] or an isotropic Gaussian distribution with mean $\boldsymbol{\mu}_{0|t}$ [35]. This results in tractable approximations for the conditional score $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{y})$, but introduces a potentially large approximation error. In this work, we propose a novel approximation for the conditional score and quantify its approximation gap exactly under distributional assumptions on $\mathbf{x}_0$.

2.3. Signal Processing Basics

Our method relies on the frequency decomposition of natural images, so in this section, we review the terminology used for the remainder of the paper. The Discrete Fourier Transform (DFT) of a finite signal $\mathbf{x} \in \mathbb{R}^n$, denoted as $\mathcal{F}(\mathbf{x})$, maps the signal to the frequency domain. At each DFT sample frequency $f_k = \frac{k}{n}$ for $k = 1, \ldots, n-1$, the DFT value $\mathcal{F}(\mathbf{x})[f_k]$ represents the complex-valued frequency component in the signal. We define the signal's *power spectral density* as the expected squared magnitude (the power) of the signal at different sample frequencies

$$S(f_k) = \mathbb{E}_{\mathbf{x}} \left[\frac{|\mathcal{F}(\mathbf{x})[f_k]|^2}{n} \right]. \tag{8}$$

When $S(f_k) = c|f_k|^{-\beta}$ for constants $c, \beta > 0$, we say that the signal follows a power law in the frequency domain with parameters c, β.

3. Related Work

Inverse problems of the form in Equation (1) have been solved for many years with different priors, such as sparse/low-rank priors [4–6] or generative priors [3, 29]. In this work, we focus on posterior sampling methods that use pretrained diffusion models and approximate the conditional score to perform guidance [8, 9, 19, 31, 35, 37, 41]. These have been applied to inverse problems such as image deblurring, super-resolution [11], and image inpainting [9, 27]. However, they can incur significant approximation errors as we will illustrate in Section 5. Works such as [10, 42] propose methods to mitigate these approximation errors, but rely on strong assumptions on the data manifold. In this work, we use a more practical assumption that the data obeys a power law in the frequency domain [34] in order to develop a new approximation.

A line of work has studied diffusion models in a transformed space to perform guidance for inverse problems [23, 24]. This is in fact closely related to our work as our work is akin to performing diffusion in the Fourier domain instead, similar to [28]. Building on recent connections between diffusion models and spectral autoregression [15], our work differs from previous approaches [8, 37] by using a frequency-guided measurement schedule rather than one based on the unconditional diffusion model's variance schedule. We refer the reader to [12] for a comprehensive survey on using diffusion models for inverse problems.

4. Frequency Guided Posterior Sampling

As illustrated in Section 2.2, the main challenge in developing unsupervised inverse problem solvers is estimating the noisy likelihood score $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t)$, which depends on the intractable denoising distribution $p(\mathbf{x}_0 \mid \mathbf{x}_t)$. It is common to substitute this denoising distribution with an approximation that depends only on the posterior mean $\boldsymbol{\mu}_{0|t} = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]$, which is tractable using Tweedie's formula. For example, Diffusion Posterior Sampling (DPS) approximates the noisy likelihood score with $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \boldsymbol{\mu}_{0|t})$ [9]. However, these methods can incur a significant approximation error by using simple distributions centered around the posterior mean instead of the true distribution.

The key challenge is ensuring that $\mathbf{x}_t$ stays on the data manifold learned by the diffusion model even after it is updated using an approximation of the gradient $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t)$. Prior approaches address this by assuming that the data manifold is linear or that $\boldsymbol{\mu}_{0|t}$ remains on the clean data manifold [10, 42]. However, these assumptions are often violated in practice. Our main insight is to use the observation that diffusion models generate images hierarchically. Specifically, recent work has connected diffusion models and inverse heat dissipation, concluding that diffusion models naturally generate images from low frequencies (corresponding to coarse features) to high frequencies (corresponding to fine-grained features) [30] [1]. If there is a mismatch in frequency content between $\boldsymbol{\mu}_{0|t}$ and the measurement $\mathbf{y}$, updating $\mathbf{x}_t$ by following an approximation of $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t)$ can move $\mathbf{x}_t$ off the learned data manifold. This problem is exacerbated during the early stages of

[1] This effect is because the forward process noise schedule removes high frequency components proportional to the variance of the noise.

the reverse process since the early stages play a crucial role in ensuring the generated output matches the fundamental low-frequency structure of the measurements.

We propose a simple fix to this issue by retaining only low frequency components of the measurements at the initial steps of the reverse process and progressively adding high frequency components in later steps. To implement this, we convolve the original measurement $\mathbf{y}$ with a time-varying low-pass filter $\mathbf{k}_t$. Denoting $\phi_t(\mathbf{y})$ as this convolution, the approximation we propose for the noisy likelihood score is

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t) \approx \nabla_{\mathbf{x}_t} \log p(\phi_t(\mathbf{y}) \mid \boldsymbol{\mu}_{0|t}), \tag{9}$$

where $\boldsymbol{\mu}_{0|t}$ denotes the Tweedie estimate $\mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t]$ predicted by the diffusion model. In order to compute the above approximation, we need to compute the model likelihood for the new measurements. Since convolution of a signal $\mathbf{y}$ with a low pass filter can be represented as a circulant matrix multiplied by $\mathbf{y}$, we have that $\phi_t(\mathbf{y}) = \mathbf{C}_t\mathbf{y}$ where $\mathbf{C}_t$ is a time-dependent circulant matrix representing $\mathbf{k}_t$. This circulant matrix has its first row as the flattened $\mathbf{k}_t$ and all other rows as cyclic shifts. Thus, using properties of the multivariate normal distribution, we have that the updated model likelihood is

$$p(\phi_t(\mathbf{y}) \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{C}_t\mathcal{A}(\mathbf{x}_0), \sigma_y^2\mathbf{C}_t\mathbf{C}_t^T). \tag{10}$$

Given this likelihood, our approximation of the noisy likelihood score becomes

$$p(\phi_t(\mathbf{y}) \mid \boldsymbol{\mu}_{0|t}) = \mathcal{N}(\mathbf{C}_t\mathcal{A}(\boldsymbol{\mu}_{0|t}), \sigma_y^2\mathbf{C}_t\mathbf{C}_t^T), \tag{11}$$

and its corresponding gradient is

$$\nabla_{\mathbf{x}_t} \log p(\phi_t(\mathbf{y}) \mid \boldsymbol{\mu}_{0|t}) = \mathbf{S}_t\nabla_{\mathbf{x}_t} \left\|\mathbf{C}_t\mathbf{y} - \mathbf{C}_t\mathcal{A}(\boldsymbol{\mu}_{0|t})\right\|_2^2, \tag{12}$$

where $\mathbf{S}_t = (\sigma_y^2\mathbf{C}_t\mathbf{C}_t^T)^{-1}$. Since $\boldsymbol{\mu}_{0|t}$ can be computed using one network evaluation of the diffusion model, its gradient with respect to $\mathbf{x}_t$ can be computed via backpropagation, as in DPS. Alternating unconditional sampling and conditional sampling using their corresponding score functions gives us our algorithm, which we call **F**requency **G**uided **P**osterior **S**ampling (FGPS) outlined in Algorithm 1.

Dataset-informed choice of $\mathbf{k}_t$. An important choice for our method is choosing the sequence $\{\mathbf{k}_t\}_{t=1}^T$. Each $\mathbf{k}_t$ is a low-pass filter that should allow frequencies up till a threshold τ_t and then strongly attenuate any frequencies above τ_t. In order to encode that the initial steps of the reverse process should only use low frequency components of $\mathbf{y}$, we require that τ_T should start off small and increase as t decreases. We denote the sequence of τ_t as our frequency curriculum. As datasets can have varying frequency characteristics, our method gives flexibility to set the frequency curriculum in a data-dependent way, as explained in Section 6.

Efficient Implementation. We efficiently compute $\phi_t(\cdot)$ in the Fourier domain. Specifically, $\mathbf{C}_t$ can be diagonalized as $\mathbf{C}_t = \mathbf{F}^{-1}\mathbf{M}\mathbf{F}$, where $\mathbf{F}$ denotes the Discrete Fourier Transform matrix [39]. Using this property along with the fact that $\mathbf{F}$ is an orthogonal matrix with determinant 1, we have that $p(\phi_t(\mathbf{y}) \mid \boldsymbol{\mu}_{0|t}) = p(\mathbf{F}\phi_t(\mathbf{y}) \mid \boldsymbol{\mu}_{0|t}) = p(\mathbf{M}\mathbf{F}\mathbf{y} \mid \boldsymbol{\mu}_{0|t})$. Thus, we bypass storing the large circulant matrix and compute $\phi_t(\mathbf{y})$ and $\phi_t(\mathcal{A}(\boldsymbol{\mu}_{0|t}))$ by first taking the Fast Fourier Transform of the input and applying a binary mask in the frequency domain, which adds minimal computational overhead (see Appendix E).

Algorithm 1 Frequency Guided Posterior Sampling

Require: $\mathbf{y}$, $\mathcal{A}$, $\{\phi_t(\cdot), \mathbf{S}_t, \alpha_t, \tilde{\sigma}_t\}_{t=1}^T$
1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t \leftarrow T$ to 1 **do**
3: $\quad\hat{\boldsymbol{s}} \leftarrow s_\theta(\mathbf{x}_t, t)$ // Predict $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$
4: $\quad\bar{\alpha}_t \leftarrow \prod_{j=1}^i \alpha_j$
5: $\quad\boldsymbol{\mu}_{0|t} \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t + (1-\bar{\alpha}_t)\hat{\boldsymbol{s}})$ // Tweedie estimate
6: $\quad\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
7: $\quad\mathbf{x}'_{t-1} \leftarrow \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\boldsymbol{\mu}_{0|t} + \tilde{\sigma}_t\mathbf{z}$ // DDIM Sampling
8: $\quad\mathbf{x}_{t-1} \leftarrow \mathbf{x}'_{t-1} - \mathbf{S}_t\nabla_{\mathbf{x}_t} \left\|\phi_t(\mathbf{y}) - \phi_t(\mathcal{A}(\boldsymbol{\mu}_{0|t}))\right\|_2^2$ // FGPS Approximation of $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t)$
9: **end for**

5. Theoretical Analysis

In this section, we utilize distributional assumptions to provide a precise theoretical analysis of the approximation errors induced by traditional posterior sampling methods as well as our proposed method.

5.1. Mind the (Approximation) Gap

Since we rely on a spectral characterization of the reverse diffusion process, our analysis requires two key properties.

1. We need to analyze the frequency characteristics of the data as well as the interplay between the frequency characteristics of the data and the forward operator.
2. To quantify the approximation gap, we need to compute the true conditional score.

For the first property, we look at the power spectral density of the data, as defined in Equation (8), which tells us how much the signal's power is distributed across different frequencies. For the second property, it is natural to suppose the clean data $\mathbf{x}_0$ comes from a multivariate Gaussian distribution. Although these two properties seem conflicting, the following lemma establishes an equivalence between the data power spectral density and the data covariance.

Lemma 1. *Let $f_k = \frac{k}{n}$ for $k = 0, \ldots, n-1$ denote the DFT sample frequencies for a signal of length n. Then, there exists a covariance matrix $\mathbf{\Sigma}_f$ such that for $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_f)$, the following two properties hold. First, the signal $\mathbf{x}$ follows a power law in the frequency domain with parameters $c, \beta > 0$, i.e., it has a power spectral density $S(f_k) = c|f_k|^{-\beta}$ for the non-zero DFT sample frequencies f_k. Second, the eigenvalues of $\mathbf{\Sigma}_f$ are precisely $c|f_k|^{-\beta}$.*

This is a special case of a classical result known as the Wiener-Khinchin Theorem [7]. Note that this does not imply that all random vectors that follow a power law in the frequency domain are Gaussian. For example, natural image data tends to follow a radially averaged power law in the frequency domain, but is far from Gaussian [34]. Nevertheless, this lemma gives a distributional assumption that mimics the spectral properties of natural image data, which we will show already elucidates failure cases of existing methods. Next, we use these assumptions to exactly quantify the approximation gap of our method.

Theorem 2. *Suppose $\mathbf{x}_0$ is drawn from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and we are given linear measurements $\mathbf{y} = \mathcal{A}(\mathbf{x}_0) + \mathbf{z}$, where $\mathcal{A}(\mathbf{x}_0) = \mathbf{A}\mathbf{x}_0$ and $\mathbf{z} \sim \mathcal{N}(0, \sigma_y^2 I)$. Suppose that the intermediate value $\mathbf{x}_t$ of the continuous-time reverse diffusion from Equation (5) can be written as $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$ where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then, we have that the true noisy likelihood score $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t)$ is*

$$(\mathbf{A}\mathbf{\Gamma}_t)^T(\mathbf{A}\mathbf{\Sigma}_{0|t}\mathbf{A}^T + \sigma_y^2 \boldsymbol{I})^{-1}(\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_{0|t}), \tag{13}$$

where $\mathbf{\Gamma}_t = \sqrt{\bar{\alpha}_t}\mathbf{\Sigma}(\bar{\alpha}_t\mathbf{\Sigma} + (1-\bar{\alpha}_t)\boldsymbol{I})^{-1}$, $\boldsymbol{\mu}_{0|t} = \mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t] = \mathbf{\Gamma}_t\mathbf{x}_t$ and $\mathbf{\Sigma}_{0|t} = \mathrm{Cov}[\mathbf{x}_0 \mid \mathbf{x}_t] = \mathbf{\Sigma} - \sqrt{\bar{\alpha}_t}\mathbf{\Gamma}_t\mathbf{\Sigma}$. Moreover, the FGPS approximation $\nabla_{\mathbf{x}_t} \log p(\mathbf{C}_t\mathbf{y} \mid \boldsymbol{\mu}_{0|t})$ can be analytically calculated as

$$(\mathbf{C}_t\mathbf{A}\mathbf{\Gamma}_t)^T\,(\sigma_y^2\mathbf{C}_t\mathbf{C}_t^T)^{-1}(\mathbf{C}_t\mathbf{y} - \mathbf{C}_t\mathbf{A}\boldsymbol{\mu}_{0|t}). \tag{14}$$

The proof is in Appendix B. The difference of Equation (13) and Equation (14) gives a precise expression for the approximation gap of our method. The $\bar{\alpha}_t$ corresponds to the variance schedule of standard diffusion models such as DDPM [20]. Our assumption that $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$ isolates the effects of the approximation gap as the generated sample is exactly $\mathbf{x}_0$. Our framework also allows us to analytically study the approximation gap for other inverse problem solvers. Below, we give one example for the popular Diffusion Posterior Sampling (DPS) solver.

Corollary 3. *Under the same assumptions and notation as in Theorem 2, we have that the DPS approximation $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \boldsymbol{\mu}_{0|t})$ can be analytically calculated as*

$$\sigma_y^{-2}(\mathbf{A}\mathbf{\Gamma}_t)^T(\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_{0|t}). \tag{15}$$

To the best of our knowledge, this is the first time any distributional assumptions have been used to characterize the exact approximation gap of any inverse problem solver. While [9] and [42] provide upper and lower bounds on this gap, these bounds do not exploit the structure of the data distribution and as such, can be loose. We formally compare the approximation gap of FGPS and DPS in Appendix C.

5.2. Approximation Gap on Synthetic Data

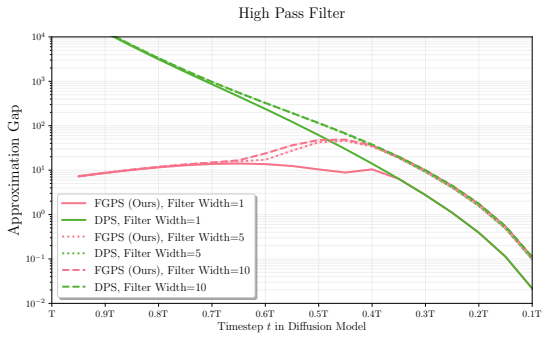

Figure 1. We show the exact approximation gap for FGPS and DPS across diffusion timesteps when the data follows a power law in the frequency domain and the forward operator is a a high-pass filter (Dirac kernel minus a Gaussian kernel) of varying width. The width denotes the σ of the Gaussian kernel. Note the y-axis is in log-scale.

In this subsection, we probe the approximation gap of our method and DPS on synthetic data, and we give examples of forward operators where FGPS results in a significantly less approximation gap than DPS. In order to construct these examples, we note that Theorem 2 depends crucially on how the forward operator $\mathbf{A}$ interacts with the data covariance $\mathbf{\Sigma}$, particularly through the term $\mathbf{A}\mathbf{\Sigma}_{0|t}\mathbf{A}^T$. Intuitively, if the data covariance is chosen to be $\mathbf{\Sigma}_f$ from Lemma 1, the frequency characteristics of the data and the forward operator would be misaligned when the forward operator is a high-pass filter, so the measurement contains primarily high-frequency components. Some natural imaging systems fall into this regime, e.g., phase contrast microscopy where intensity measurements are high-pass filtered versions of refractive index.

To verify the above intuition, we begin by simulating 1-D signals $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_f)$ that follow a power law in the frequency domain, which uses the construction of $\mathbf{\Sigma}_f$ given in Lemma 1. Then, the corresponding measurements for these signals are $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$ with $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where $\mathbf{A}$ is the convolution matrix corresponding to convolution with a Dirac kernel minus a Gaussian kernel of standard deviation σ. This kernel approximates a high-pass filter.

Figure 1 shows the approximation gap of DPS and FGPS as a function of the diffusion timesteps (where higher

timesteps correspond to a lower value of $\bar{\alpha}_t$) for the variance schedule used in the DDPM model for $T = 1000$ [20]. In accordance with our intuition, we see that the DPS approximation worsens significantly as we increase t, likely because higher frequency noise components in $\mathbf{x}_t$ are amplified by the high-pass filter. In contrast, we find our method has an approximation gap that is several orders of magnitude smaller than DPS in the early stages of the reverse process. As the early stages of the reverse process are when the generated image develops coarse features, we argue it is crucial that the approximation gap is small in these phases. Refer to Appendix E for the full experimental details.

6. Experiments

Although our theoretical results were restricted to cases where FGPS provably improved over existing methods, we now turn to empirical validation of FGPS on several real-world datasets and image restoration tasks.

6.1. Experimental Setup

Datasets. We evaluate our method on two datasets: the Flickr-Faces-HQ (FFHQ 256x256) dataset of high-resolution faces [22], and the ImageNet-1000 dataset of diverse objects [13]. We utilize pretrained diffusion models from [9] and [14] for the FFHQ and ImageNet datasets respectively, using a class-conditional model for ImageNet.

Tasks. We consider four challenging image restoration tasks for our experimental evaluation. The first three are linear inverse problems: Gaussian deblurring, motion deblurring, and deconvolution with a high-pass filter. The last is a nonlinear inverse problem: image dehazing. The details of the four forward operators we consider can be found in Appendix E. All measurements have Gaussian measurement noise added with standard deviation 0.05.

Baselines and Metrics. We compare against several competitive unsupervised methods. The first two baselines are popular unsupervised diffusion-based solvers: Diffusion Posterior Sampling (DPS) [9], a method that approximates the conditional score with a Dirac delta around the posterior mean, and Score-SDE/ILVR [8, 37], a method that uses a sequence of noisy measurements computed using the diffusion model variance schedule. Then, we compare to two state-of-the-art methods focused on mitigating the approximation gap problem: Manifold Constrained Gradient (MCG) [10] and Diffusion with Spherical Guidance (DSG) [42]. For image dehazing, we further compare to two problem-specific baselines: DoubleDIP [18] and AOD-Net [26], a supervised method trained to dehaze images from pairs of clean-hazy images. We use the same pretrained score function for all diffusion-based methods. We evaluate our method on two perceptual metrics, FID and LPIPS distance, and two distortion-based metrics, PSNR and SSIM.

Frequency Curriculum. For both FFHQ and ImageNet across all tasks, in our method we set τ_1 and τ_T to be $\frac{75*100}{256}\%$ and $\frac{10*100}{256}\%$ of the overall frequency range respectively. We set different frequency curricula for the two datasets as described in Section 4. We find the exponential schedule works well for face datasets with many high-frequency details, while the uniform schedule is better for diverse, multiclass datasets like ImageNet. We emphasize that $\mathbf{k}_t$ does not require extensive tuning, as a simple schedule already beats baseline methods (see Appendix G).

Step Size Schedule. In practice, we take $\mathbf{S}_t$ to be a hyperparameter as opposed to being fixed as $(\sigma_y^2 \mathbf{C}_t \mathbf{C}_t^T)^{-1}$ in order to further control the effect of the approximation error. We adopt the choice that DPS makes where the step size is $\frac{\kappa}{\left\|\phi_t(\mathbf{y}) - \phi(\mathcal{A}(\boldsymbol{\mu}_{0|t}))\right\|_2}\mathbf{I}$ for a hyperparameter κ. Using the intuition from Figure 1 that the approximation gap is small in the initial steps of the reverse process, we set κ large initially and decay it following a cosine curve.

6.2. Results

Empirical Validation of Theoretical Results. In Figure 2, we consider two forward operators on an example from the FFHQ dataset: a) a standard Gaussian blurring low-pass filter, and b) the high-pass filter we studied in our theoretical analysis. As expected, prior approaches such as Diffusion Posterior Sampling [9] and variants [10, 42] perform well at Gaussian deblurring. However, when the forward operator is a high-pass filter, remarkably these approaches all fail to produce coherent outputs, which aligns with our theoretical results about the large approximation gap of existing methods. FGPS on the other hand is still able to produce high-quality reconstructions.

Results on Linear Inverse Problems. Table 1 shows our main results for linear inverse problems. On the FFHQ dataset, for simple tasks such as Gaussian deblurring, FGPS gives comparable performance to baselines, which already have strong reconstruction performance. However, on more complex image restoration tasks such as motion deblurring, our method outperforms the baselines in almost all metrics. Notably, even though Score-SDE/ILVR uses a sequence of noisy measurements, this sequence is not adapted to the frequency characteristics of the data, which results in lower quality reconstructions as compared to our method (see Appendix F for further discussion). In Figure 3, we see the qualitative differences between the methods. For high-pass filter operators, existing methods generate visually implausible images, indicating that the approximation errors push the denoised outputs away from the natural image manifold. In contrast, our method is able to accurately reconstruct the underlying ground truth image. For blurry images, we see that DPS tends to produce smoothed images, losing high-frequency details. While DSG is able to improve upon this

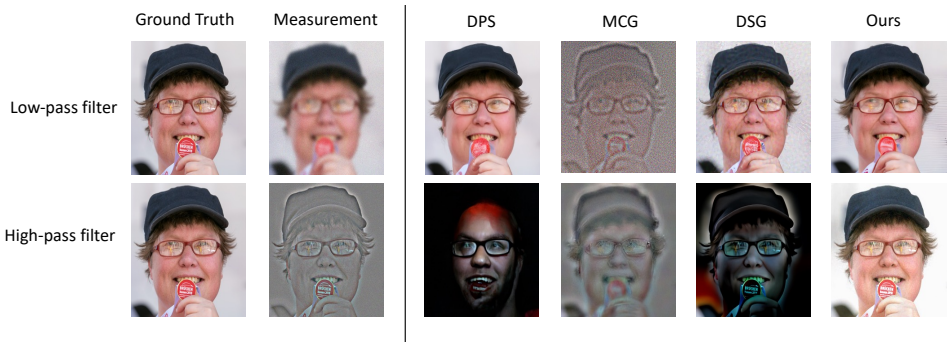

Figure 2. Even for linear inverse problems, SOTA diffusion-based methods (DPS, MCG, DSG) can struggle to solve inverse problems when the forward operator is convolution with a high-pass filter. In contrast, our method is still able to achieve a high quality reconstruction.

Method	Gaussian Blur				Motion Blur				High-Pass Filter			
	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑	FID↓	LPIPS↓	PSNR↑	SSIM↑
FFHQ Dataset												
Score-SDE/ILVR [8, 37]	58.79	0.266	23.34	0.648	69.10	0.387	18.59	0.457	85.35	0.356	11.86	0.600
MCG [10]	310.33	1.035	11.64	0.143	296.93	0.771	11.54	0.127	151.77	0.581	12.48	0.466
DPS [9]	**21.69**	**0.116**	**26.01**	**0.743**	27.15	0.153	24.17	0.688	121.57	0.694	6.17	0.222
DSG [42]	44.63	0.294	22.75	0.633	32.74	0.257	**25.66**	0.687	17.81	0.246	12.12	0.695
FGPS (Ours)	21.89	0.139	25.44	0.727	**20.48**	**0.127**	25.48	**0.723**	**10.02**	**0.124**	**18.79**	**0.739**
ImageNet Dataset												
Score-SDE/ILVR [8, 37]	104.81	0.676	16.27	0.318	100.54	0.665	16.62	0.290	226.51	0.784	10.69	0.345
MCG [10]	320.63	1.005	12.01	0.121	306.43	0.963	11.80	0.076	295.10	1.000	11.62	0.278
DPS [9]	98.02	0.517	17.35	0.364	77.74	0.407	20.90	0.551	111.73	0.569	10.66	0.396
DSG [42]	87.51	0.375	20.75	0.524	107.25	0.453	19.58	0.488	110.85	0.546	8.425	0.382
FGPS (Ours)	**56.46**	**0.294**	**21.70**	**0.574**	**49.25**	**0.267**	**22.01**	**0.601**	**24.44**	**0.192**	**15.96**	**0.686**

Table 1. Quantitative comparison of different methods on FFHQ and ImageNet validation datasets. Bolded entries indicate best performance and underlined entries indicate second best performance.

and generate images with fine-grained characteristics of the underlying image, it tends to also generate artifacts. In contrast, due to our frequency curriculum, our method is able to generate images with high perceptual quality. We give further qualitative comparisons to baselines in Appendix G.

Role of Dataset. For the ImageNet dataset, we observe that FGPS significantly outperforms all baselines, especially in perceptual quality. We conjecture that this is because on more complex datasets with varying frequency characteristics, the explicit frequency schedule for the measurement is essential to ensure the reverse process iterates do not deviate far from the intermediate noisy data manifolds learned by the diffusion models.

Results on Image Dehazing. Table 2 shows our main results for the nonlinear image dehazing task on the FFHQ dataset. A key strength of our method is the ability to handle nonlinear inverse problems, unlike methods such as [8, 24, 37]. In Figure 3, we observe that our method successfully reconstructs finer details such as background details and facial characteristics similar to what we observed for linear inverse problems. DSG produces slightly grainy reconstructions, but performs much better on the dehazing task than the other tasks. Due to the powerful generative capacity of diffusion models, our method also performs better than problem-specific supervised methods such as AOD-Net [26], a phenomenon also observed in [35]. We provide more qualitative examples in Appendix G.

Method	FID↓	LPIPS↓	PSNR↑	SSIM↑
DoubleDIP [18]	50.51	0.500	13.80	0.483
AOD-Net [26]	33.35	0.325	17.42	0.532
DPS [9]	24.24	0.137	25.88	0.806
DSG [42]	10.55	0.134	26.07	0.795
FGPS (Ours)	**9.37**	**0.066**	**26.88**	**0.840**

Table 2. Comparison of different haze removal methods. Bolded entries indicate best performance and underlined entries indicate second best performance.

Visualizing Measurement Sequence. In Figure 4, we show the transformed measurements at three timesteps in the reverse diffusion process when running FGPS. In the early stages of the reverse process, the measurement retains

Figure 3. Qualitative results on FFHQ and ImageNet datasets. The dotted red boxes highlight areas of the image where our method results in higher-quality reconstructions than DPS and DSG. Zoomed in versions can be found in the Appendix.

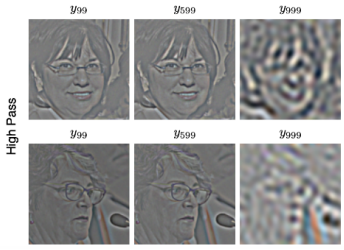

Figure 4. Visualization of Transformed Measurements y_t at different timesteps t, demonstrating our coarse-to-fine strategy.

only a coarse outline of the measurement. This coarse-to-fine strategy is essential to ensuring $\mathbf{x}_t$ remains on the data manifold learned by the diffusion model.

Effect of Frequency Curriculum. A key ingredient of our method is the frequency curriculum with a time-varying low pass filter $\mathbf{k}_t$. To understand the effect of this curriculum, we compare taking a time-dependent $\mathbf{k}_t$ vs. a fixed $\mathbf{k}$ that zeroes out frequencies above τ_1 for all t. In Figure 5, we see that for the high-pass filter operator, the generated reconstructions are still visually implausible images for a fixed $\mathbf{k}$, indicating the frequency curriculum is crucial to our method. Our method only requires $\mathbf{k}_1$ to be a low frequency cut-off, otherwise the reconstructions look like Figure 5. This corroborates our theoretical results, which show that the approximation gap early on in the reverse process greatly influences reconstruction quality. Finally, our method does not qualitatively change with higher cutoffs for $\mathbf{k}_T$. We examine this in detail in Appendix G as well as conduct further ablation studies.

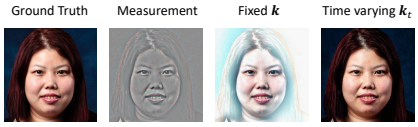

Figure 5. Qualitative Ablation Study: fixed low-pass filter vs. time-dependent low pass filter.

7. Conclusion

In this paper, we introduce Frequency Guided Posterior Sampling (FGPS), a method that progressively incorporates frequency information in a data-dependent manner. We provide a theoretical analysis characterizing the approximation error of FGPS and existing diffusion-based restoration methods, demonstrating how these errors can become severe when forward operators interact with high-frequency components of images. Our experiments on challenging image restoration tasks, such as motion deblurring and image dehazing, demonstrate that FGPS significantly outperforms existing methods. This work opens up new directions for analyzing diffusion-based inverse problem solvers by carefully considering the frequency characteristics of both the data and the degradation process.

Acknowledgements

D.T. was supported by a gift from AWS to Penn Engineering's ASSET Center for Trustworthy AI. R.V. thanks the support of Penn startup funds and the Research Collaboration on the Mathematical and Scientific Foundations of Deep Learning under grants NSF 2031985 and Simons 814201.

References

[1] Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. 2

[2] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*. Springer, 2006. 2

[3] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International conference on machine learning*, pages 537–546. PMLR, 2017. 1, 3

[4] Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012. 1, 3

[5] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006. 1

[6] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011. 1, 3

[7] David C Champeney. *A handbook of Fourier theorems*, chapter 11. Cambridge University Press, 1987. 5

[8] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021. 3, 6, 7, 4, 5

[9] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022. 1, 3, 5, 6, 7, 4, 8

[10] 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. 2, 3, 6, 7, 4

[11] Max Daniels, Tyler Maunu, and Paul Hand. Score-based generative neural networks for large-scale optimal transport. *Advances in neural information processing systems*, 34:12955–12965, 2021. 3

[12] Giannis Daras, Hyungjin Chung, Chieh-Hsin Lai, Yuki Mitsufuji, Jong Chul Ye, Peyman Milanfar, Alexandros G Dimakis, and Mauricio Delbracio. A survey on diffusion models for inverse problems. *arXiv preprint arXiv:2410.00083*, 2024. 3, 4

[13] 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*, pages 248–255. Ieee, 2009. 6

[14] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 1, 6

[15] Sander Dieleman. Diffusion is spectral autoregression, 2024. 3

[16] David L Donoho, Michael Elad, and Vladimir N Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on information theory*, 52(1):6–18, 2005. 1

[17] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. 2, 3, 1

[18] Yosef Gandelsman, Assaf Shocher, and Michal Irani. " double-dip": unsupervised image decomposition via coupled deep-image-priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11026–11035, 2019. 6, 7

[19] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 3

[20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 1, 2, 5, 6

[21] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *Advances in Neural Information Processing Systems*, 35:8633–8646, 2022. 1

[22] Tero Karras, S. Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4396–4405, 2019. 6

[23] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021. 3

[24] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022. 1, 3, 7, 5

[25] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. *arXiv preprint arXiv:2009.09761*, 2020. 1

[26] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pages 4770–4778, 2017. 6, 7

[27] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11461–11471, 2022. 3

[28] Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, and Liqiang Nie. Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25378–25388, 2024. 3

[29] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7474–7489, 2021. 1, 3

[30] Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. *arXiv preprint arXiv:2206.13397*, 2022. 2, 3

[31] François Rozet, Gérôme Andry, François Lanusse, and Gilles Louppe. Learning diffusion priors from observations by expectation maximization. *arXiv preprint arXiv:2405.13712*, 2024. 3

[32] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992. 1

[33] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022. 1, 2

[34] A. van der Schaaf and J. H. van Hateren. Modelling the Power Spectra of Natural Images: Statistics and Information. *Vision Research*, 36(17):2759–2770, 1996. 3, 5

[35] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023. 1, 3, 7

[36] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019. 1, 2

[37] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 1, 2, 3, 6, 7, 4, 5

[38] Andrew M Stuart. Inverse problems: a bayesian perspective. *Acta numerica*, 19:451–559, 2010. 1

[39] Marco Alexander Treiber. *Optimization for computer vision*. Springer, 2013. 4

[40] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011. 2

[41] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022. 1, 3

[42] Lingxiao Yang, Shutong Ding, Yifan Cai, Jingyi Yu, Jingya Wang, and Ye Shi. Guidance with spherical gaussian constraint for conditional diffusion. *arXiv preprint arXiv:2402.03201*, 2024. 2, 3, 5, 6, 7, 4