

A Survey on Diffusion Models for Inverse Problems

Giannis Daras¹, Hyungjin Chung², Chieh-Hsin Lai³, Yuki Mitsufuji³,
Jong Chul Ye², Peyman Milanfar⁴, Alexandros G. Dimakis¹, Mauricio Delbracio⁴

¹UT Austin ²KAIST ³Sony AI ⁴Google

Abstract

Diffusion models have become increasingly popular for generative modeling due to their ability to generate high-quality samples. This has unlocked exciting new possibilities for solving inverse problems, especially in image restoration and reconstruction, by treating diffusion models as unsupervised priors. This survey provides a comprehensive overview of methods that utilize pre-trained diffusion models to solve inverse problems without requiring further training. We introduce taxonomies to categorize these methods based on both the problems they address and the techniques they employ. We analyze the connections between different approaches, offering insights into their practical implementation and highlighting important considerations. We further discuss specific challenges and potential solutions associated with using latent diffusion models for inverse problems. This work aims to be a valuable resource for those interested in learning about the intersection of diffusion models and inverse problems.

1 Introduction

1.1 Problem Setting.

Inverse problems are ubiquitous and the associated reconstruction problems have tremendous applications across different domains such as seismic imaging [81, 145], weather prediction [65], oceanography [152], audio signal processing [83, 123, 96, 97, 98, 62], medical imaging [133, 26, 2, 29], etc. Despite their generality, inverse problems across different domains follow a fairly unified mathematical setting. Specifically, in inverse problems, the goal is to recover an unknown sample $\mathbf{x} \in \mathbb{R}^n$ from a distribution $p_{\mathbf{X}}$, assuming access to measurements $\mathbf{y} \in \mathbb{R}^m$ and a corruption model

$$\mathbf{Y} = \mathcal{A}(\mathbf{X}) + \sigma_{\mathbf{y}}\mathbf{Z}, \quad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_m). \quad (1.1)$$

In what follows, we present some well-known examples of measurement models that fit under this general formulation.

Example 1.1 (Denoising). The simplest interesting example is the denoising inverse problem, i.e. when A is the identity matrix and $\sigma_{\mathbf{y}} > 0$. In fact, the noise model does not have to be Gaussian and it can be generalized to other distributions, including the Laplacian Distribution or the Poisson Distribution [52]. For the purposes of this survey, we focus on additive Gaussian noise.

A lot of practical applications arise from the non-invertible linear setting, i.e. for $\mathcal{A}(\mathbf{X}) = A\mathbf{X}$ and A being an $m \times n$ matrix with $m < n$.

Example 1.2 (Inpainting). A is a masking matrix, i.e. $A_{ij} = 0$ for $i \neq j$ and A_{ii} is either 0 or 1, based on whether the value at this location is observed.

Example 1.3 (Compressed Sensing). A is a matrix with entries sampled from a Gaussian random variable.

Example 1.4 (Convolutions). Here $\mathcal{A}(\mathbf{X})$ represents the convolution of \mathbf{X} with a (Gaussian or other) kernel, which is again a linear operation.

Category	Method	Non-linear	Blind	Handle noise	Pixel/Latent	Text-conditioned	Optimization Technique	Code ¹
Explicit approximations for measurement matching	Score-ALD [67]	✗	✗	✓	Pixel	✗	Grad	code
	Score-SDE [134]	✗	✗	✗	Pixel	✗	Proj	code
	ILVR [22]	✗	✗	✗	Pixel	✗	Proj	code
	DPS [24]	✓	✗	✓	Pixel	✗	Grad	code
	IIGDM [129]	✓	✗	✓	Pixel	✗	Grad	code
	Moment Matching [119]	✓	✗	✓	Pixel	✗	Grad	code
	BlindDPS [23]	✓	✓	✓	Pixel	✗	Grad	code
	SNIPS [74]	✗	✗	✓	Pixel	✗	Grad	code
	DDRM [73]	✗	✗	✓	Pixel	✗	Grad	code
	GibbsDDRM [99]	✗	✓	✓	Grad	✗	Samp	code
	DDNM [148]	✗	✗	✓	Pixel	✗	Proj	code
	DDS [25]	✗	✗	✓	Pixel	✗	Opt	code
	DiffPIR [163]	✗	✗	✓	Pixel	✗	Opt	code
	PSLD [118]	✓	✗	✓	Latent	✗	Grad	code
	STSL [116]	✓	✗	✓	Latent	✗	Grad	✗
Variational Inference	RED-Diff [94]	✓	✗	✓	Pixel	✗	Opt	code
	Blind RED-Diff [6]	✓	✓	✓	Pixel	✗	Opt	✗
	Score Prior [54]	✓	✗	✓	Pixel	✗	Opt	code
	Efficient Score Prior [53]	✓	✗	✓	Pixel	✗	Opt	code
CSGM methods	DMPUG [146]	✓	✗	✓	Pixel	✗	Opt	code
	SHRED [21]	✓	✗	✓	Pixel	✗	Opt	✗
	Consistent-CSGM [153]	✓	✗	✓	Pixel	✗	Opt	✗
	Score-ILO [32]	✓	✗	✓	Pixel	✗	Opt	code
Asymptotically Exact Methods	PnP-DM [151]	✓	✗	✓	Pixel	✗	Opt	✗
	FPS [47]	✓	✗	✓	Pixel	✗	Samp	code
	PMC [137]	✓	✗	✓	Pixel	✗	Samp	code
	SMCDiff [141]	✗	✗	✓	Pixel	✗	Samp	code
	MCGDiff [17]	✗	✗	✓	Pixel	✗	Samp	code
	TDS [150]	✗	✗	✓	Pixel	✗	Samp	code
Other methods	Implicit denoiser prior [68]	✗	✗	✗	Pixel	✗	Proj	code
	MCG [27]	✗	✗	✗	Pixel	✗	Grad/Proj	code
	Resample [127]	✓	✗	✓	Latent	✗	Grad/Opt	code
	MPGD [61]	✓	✗	✓	Pixel/Latent	✓	Grad/Opt	code
	P2L [30]	✓	✗	✓	Latent	✓	Grad/Opt	✗
	TReg [77]	✓	✗	✓	Latent	✓	Grad/Opt	✗
	DreamSampler [78]	✓	✗	✓	Latent	✓	Grad/Opt	code

Table 1: **Categorization of Diffusion-Based Inverse Problem Solvers.** This table categorizes methods by their approach to solving inverse problems with diffusion models. We identified four families of methods. *Explicit Approximations for Measurement Matching:* These methods approximate the measurement matching score, $\nabla \log p_t(\mathbf{y}|\mathbf{x}_t)$, with a closed-form expression. *Variational Inference:* These methods approximate the true posterior distribution, $p(\mathbf{x}|\mathbf{y})$, with a simpler, tractable distribution. Variational formulations are then used to optimize the parameters of this simpler distribution. *CSGM-type methods:* The works in this category use backpropagation to change the initial noise of the deterministic diffusion sampler, essentially optimizing over a latent space for the diffusion model. *Asymptotically Exact Methods:* These methods aim to sample from the true posterior distribution. This is typically achieved by constructing Markov chains (MCMC) or by propagating particles through a sequence of distributions (SMC) to obtain samples that approximate the posterior. Further categorization is based on being able to address non-linear problems, blind formulations (unknown forward model), noise handling, pixel/latent space operation, text-conditioning, and the type of optimization technique used (gradient-based, projection, etc.). Code availability is also indicated.

The same inverse problem can appear across vastly different scientific fields. To illustrate this point, we can take the inpainting case as an example. In Computer Vision, inpainting can be useful for applications such as object removal or object replacement [109, 118, 158]. In the proteins domain, inpainting can be useful for protein engineering, e.g. by mutating certain aminoacids of the protein sequence to achieve better thermodynamical properties [104, 42, 155, 154]. MRI acceleration is also an inpainting problem but in the Fourier domain [1, 161, 40, 162, 140]. Particularly, for each coil measurement y_i within the multi-coil setting, we have $A_i = PFS_i$, where P is the masking operator, F is the 2D discrete Fourier transform, and S_i denotes the element-wise sensitivity value. For single-coil, S_i is the identity matrix [92]. Similarly, CT can be considered an inpainting problem in the Radon-transformed domain $A = PR$, where R is the Radon transform [105, 57, 13]. Depending on the circumstances such as sparse-view or limited-angle, the pattern of the masking operator P differs [70]. Finally, in the audio domain, the bandwidth extension problem, i.e. the task of recovering high-frequency content from an observed signal, is another example of inpainting in the spectrogram domain [43].

Inpainting is just one of many useful linear inverse problems in scientific applications and there are plenty of other important examples to consider. Cryo-EM [49] is a blind inverse problem that is

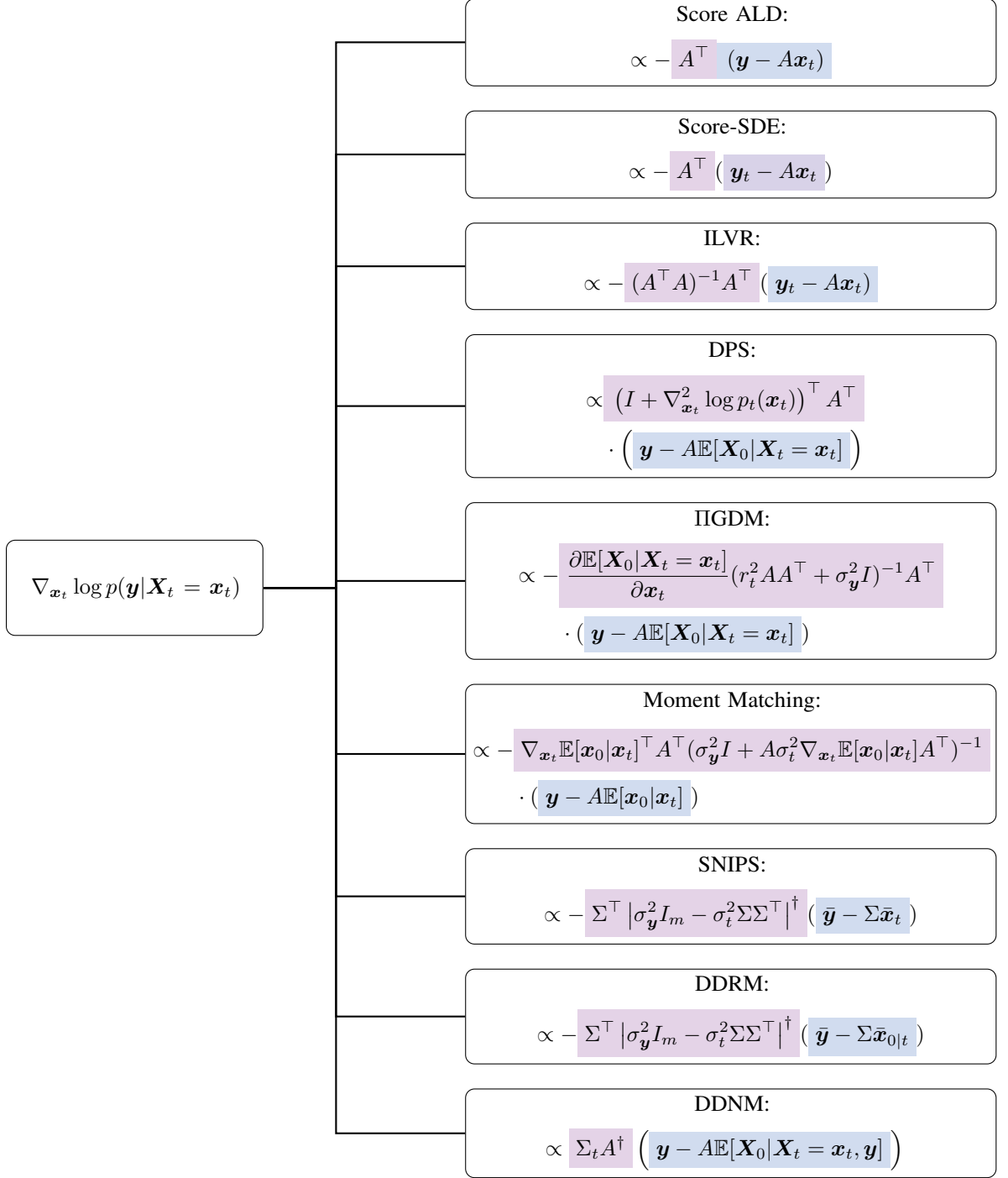


Figure 1: Approximations for the measurements score proposed by different methods.

defined by $A = CSR$, where C is a blur kernel and S is a shifting matrix, i.e. additional (unknown) shift and blur is applied to the projections. Deconvolution appears in several applications such as super-resolution [106, 122] of images and removing reverberant corruption [101] in audio signals.

There are many interesting non-linear inverse problems too, i.e. where \mathcal{A} is a nonlinear operator.

Example 1.5 (Phase Retrieval [55]). Phase retrieval considers the nonlinear operator $\mathcal{A}(\mathbf{X}) := |\mathbf{F}\mathbf{X}|$, where the measurement contains only the magnitude of the Fourier signal.

Example 1.6 (Compression Removal). Here $\mathcal{A}(\mathbf{X}; \alpha)$ represents a (non-linear) compression operator (e.g., JPEG) whose strength is controlled by the parameter α .

A famous non-linear inverse problem is the problem of imaging a black hole, where the relationship between the image to be reconstructed and the interferometric measurement can be considered as a sparse and noisy Fourier phase retrieval problem [3].

1.2 Recovery types

One common characteristic of these problems is that information is lost and perfect recovery is impossible [139], i.e. they are *ill-posed*. Hence, the type of “recovery” we are looking for should be carefully defined. For instance, one might be looking for the point that maximizes the posterior distribution $p(\mathbf{x}|\mathbf{y})$ [10, 107]. Often, the Maximum a posteriori (MAP) estimation coincides with the Minimum Mean Squared Error Estimator, i.e. the conditional expectation $\mathbb{E}[\mathbf{x}|\mathbf{y}]$ [157, 100]. MMSE estimation attempts to minimize distortion of the unknown signal \mathbf{x} , but often lead to unrealistic recoveries. A different approach is to sample from the full posterior distribution, $p(\mathbf{x}|\mathbf{y})$. Posterior sampling accounts for the uncertainty of the estimation, and typically produces samples that have higher perception quality. [14] show that, in general, it is impossible to find a sample that maximizes perception and minimizes distortion at the same time. Yet, posterior sampling is nearly optimal [67] in terms of distortion error.

1.3 Approaches for Solving Inverse Problems

Inverse problems have a rich history, with approaches evolving significantly over the decades [113, 9]. While a comprehensive review is beyond the scope of this survey, we highlight key trends to provide context. Early approaches, prevalent in the 2000s, often framed inverse problems as optimization tasks [38, 16, 46, 56, 38, 59, 126]. These methods sought to balance data fidelity with regularization terms that encouraged desired solution properties like smoothness [120, 12] or sparsity in specific representations (e.g., wavelets, dictionaries) [56, 38, 16, 46, 59].

The advent of deep learning brought a paradigm shift [103]. Researchers began leveraging large paired datasets to directly learn mappings from measurements to clean signals using neural networks [45, 85, 138, 19, 159, 20, 142, 160]. These approaches focus on minimizing some reconstruction loss during training, with various techniques employed to penalize distortions, and optimize for specific application goals (e.g., perceptual quality [66, 80]). Traditional point estimates aim to recover a single reconstruction by for example minimizing the average reconstruction error (i.e., MMSE) or by finding the most probable reconstruction through Maximum a Posteriori estimate (MAP), i.e., finding the \mathbf{x} that *maximizes* $p(\mathbf{x}|\mathbf{y})$. While powerful, this approach can suffer from “regression to the mean”, where the network predicts an average solution that may lack important details or even be outside the desired solution space [14, 39]. In fact, learning a mapping to minimize a certain distortion metric will lead, in the best case, to an average of all the plausible reconstructions (e.g., when using a L2 reconstruction loss, the best-case solution will be the posterior mean). This reconstruction might not be in the target space (e.g., a blurry image being the average of all plausible reconstructions) [14].

Recent research has revealed a striking connection between denoising algorithms and inverse problems. Powerful denoisers, often based on deep learning, implicitly encode valuable information about natural signals. By integrating these denoisers into optimization frameworks, we can harness their learned priors to achieve exceptional results in a variety of inverse problems [143, 135, 18, 114, 31, 69, 71, 95]. This approach bridges the gap between traditional regularization methods and modern denoising techniques, offering a promising new paradigm for solving these challenging tasks.

An alternative perspective views inverse problems through the lens of Bayesian inference. Given measurements \mathbf{y} , the goal becomes generating plausible reconstructions by sampling from the posterior distribution $p(\mathbf{X}|\mathbf{Y} = \mathbf{y})$ – the distribution of possible signals \mathbf{x} given the observed measurements \mathbf{y} .

In this survey we explore a specific class of methods that utilize diffusion models as priors for $p_{\mathbf{X}}$, and then try to generate plausible reconstructions (e.g., by sampling from the posterior). While other approaches exist, such as directly learning conditional diffusion models or flows for specific inverse problems [84, 122, 121, 149, 90, 91, 5, 4, 86, 87, 88, 125], these often require retraining for each

new application. In contrast, the methods covered in this survey offer a more general framework applicable to arbitrary inverse problems without retraining or fine-tuning.

Unsupervised methods. We refer as *unsupervised methods* to those that focus on characterizing the distribution of target signals, $p_{\mathbf{X}}$, and applying this knowledge during the inversion process. Since they don't rely on paired data, they can be flexibly applied to different inverse problems using the same prior knowledge.

Unsupervised methods can be used to maximize the likelihood of $p(\mathbf{x}|\mathbf{y})$ or to sample from this distribution. Algorithmically, to solve the former problem we typically use (some variation of) Gradient Descent and to solve the latter (some variation of) Monte Carlo Simulation (e.g., Langevin Dynamics). Either way, one typically requires to compute the gradient of the conditional log-likelihood, i.e., $\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y})$.

A simple application of Bayes Rule reveals that:

$$\underbrace{\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y})}_{\text{conditional score}} = \underbrace{\nabla_{\mathbf{x}} \log p(\mathbf{x})}_{\text{unconditional score}} + \underbrace{\nabla_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x})}_{\text{measurements matching term}}. \quad (1.2)$$

The last term typically has a closed-form expression, e.g. for the linear case, we have that: $\nabla_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) = \frac{\mathbf{y} - A\mathbf{x}}{\sigma_y^2}$. However, the first term, known as the score function, might be hard to estimate when the data lie on low-dimensional manifolds. The problem arises from the fact that we do not get observations outside of the manifold and hence the vector-field estimation is inaccurate in these regions.

One way to sidestep this issue is by using a “smoothed” version of the score function, representing the score function of noisy data that will be supported everywhere. The central idea behind diffusion generative models is to learn score functions that correspond to different levels of smoothing. Specifically, in diffusion modeling, we attempt to learn the smoothed score functions, $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$, where $\mathbf{X}_t = \mathbf{X}_0 + \sigma_t \mathbf{Z}$, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I)$, for different noise levels t . During sampling, we progressively move from more smoothed vector fields to the true score function. At the very end, the score function corresponding to the data distribution is only queried at points for which the estimation is accurate because of the warm-start effect of the sampling method.

Even though estimating the unconditional score becomes easier (because of the smoothing), the measurement matching term becomes *time dependent* and loses its closed form expression. Indeed, the likelihood of the measurements is given by the intractable integral:

$$p_t(\mathbf{y}|\mathbf{x}_t) = \int p(\mathbf{y}|\mathbf{x}_0)p(\mathbf{x}_0|\mathbf{x}_t)d\mathbf{x}_0. \quad (1.3)$$

The computational challenge that emerges from the intractability of the conditional likelihood has led to the proposal of numerous approaches to use diffusion models to solve inverse problems [67, 24, 129, 134, 22, 73, 74, 148, 25, 163, 119, 89, 117, 94, 54, 53, 151, 47, 17, 150, 68, 146, 21, 28, 124]. The sheer number of the proposed methods, but also the different perspectives under which these methods have been developed, make it hard for both newcomers and experts in the field to understand the connections between them and the unifying underlying principles. This work attempts to *explain*, *taxonomize* and *relate* prominent methods in the field of using diffusion models for inverse problems. Our list of methods is by no means exhaustive. The goal of this manuscript is not to list all the methods that have been proposed but to review some representative methods of different approaches and present them under a unifying framework. We believe this survey will be useful as a reference point for people interested in this field.

2 Background

2.1 Diffusion Processes

Forward and Reverse Processes. The idea of a diffusion model is to transform a simple distribution (e.g., normal distribution) into the unknown data distribution $p_0(\mathbf{x})$, that we don't know explicitly but we have access to some of its samples. The first step is to define a *corruption process*. The popular Denoising Diffusion Probabilistic Models (DDPM) [63, 132], adopt a discrete time

Markovian process to transform the input Normal distribution into the target one by incrementally adding Gaussian noise. More generally, the corruption processes of interest can be generalized to continuous time by a stochastic differential equation (SDE) [134]:

$$d\mathbf{x}_t = \underbrace{\mathbf{f}(\mathbf{x}_t, t)}_{\text{drift coeff.}} dt + \underbrace{g(t)}_{\text{diffusion coeff.}} d\mathbf{W}_t, \quad (2.1)$$

with $\mathbf{x}_0 \sim p_0, \mathbf{x}_0 \in \mathbb{R}^n$, and \mathbf{W}_t denotes a Wiener process (i.e., Brownian motion). This SDE gradually transforms the data distribution into Gaussian noise. We denote with p_t the distribution that arises by running this dynamical system up to time t .

A remarkable result by Anderson [7] shows that we can sample from p_0 by running backwards in time the reverse SDE:

$$d\mathbf{x}_t = \left(\mathbf{f}(\mathbf{x}_t, t) - g^2(t) \underbrace{\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)}_{\text{score}} \right) dt + g(t) d\mathbf{W}_t, \quad (2.2)$$

initialized at $\mathbf{x}_T \sim p_T$. For sufficiently large T and for linear drift functions $\mathbf{f}(\cdot, \cdot)$, the latter distribution approaches a Gaussian distribution with known parameters that can be used for initializing the process. Hence, the remaining goal becomes to estimate the score function $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$.

Probability Flow ODE. Song et al. [134] and Maoutsa, Reich, and Oppor [93] observe that the (deterministic) differential equation:

$$\frac{d\mathbf{x}_t}{dt} = \left(\mathbf{f}(\mathbf{x}_t, t) - \frac{g^2(t)}{2} \underbrace{\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)}_{\text{score}} \right) \quad (2.3)$$

corresponds to the same Fokker-Planck equations as the SDE of Equation 2.2. An implication of this is we can use the *deterministic* sampling scheme of Equation 2.3. Any well-built numerical ODE solver can be used to solve Equation 2.3, such as the Euler solver:

$$\mathbf{x}_{t-\Delta t} = \mathbf{x}_t + \Delta t \left(\mathbf{f}(\mathbf{x}_t, t) - \frac{g^2(t)}{2} \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \right). \quad (2.4)$$

SDE variants: Variance Exploding and Variance Preserving Processes. The drift coefficients, $\mathbf{f}(\mathbf{x}_t, t)$, and the diffusion coefficients $g(t)$ are design choices. One popular choice, known as the Variance Exploding SDE, is setting $\mathbf{f}(\mathbf{x}_t, t) = \mathbf{0}$ and $g(t) = \sqrt{\frac{d\sigma_t^2}{dt}}$ for some variance scheduling $\{\sigma_t\}_{t=0}^T$. Under these choices, the marginal distribution at time t of the forward process of Equation 2.1 can be alternatively described as:

$$\mathbf{X}_t = \mathbf{X}_0 + \sigma_t \mathbf{Z}, \quad \mathbf{X}_0 \sim p(\mathbf{X}_0), \quad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_n). \quad (2.5)$$

The typical noise scheduling for this SDE is $\sigma_t = \sqrt{t}$ (that corresponds to $g(t) = 1$).

Another popular choice is to set the drift function to be $\mathbf{f}(\mathbf{x}_t, t) = -\mathbf{x}_t$, which is known as the Variance Preserving (VP) SDE. A famous process in the VP SDE family is the Ornstein–Uhlenbeck (OU) process:

$$d\mathbf{x}_t = -\mathbf{x}_t dt + \sqrt{2} d\mathbf{W}_t, \quad (2.6)$$

which gives:

$$\mathbf{X}_t = \exp(-t) \mathbf{X}_0 + \sqrt{1 - \exp(-2t)} \mathbf{Z}, \quad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_n). \quad (2.7)$$

The VP SDE [63] takes a more general form:

$$\mathbf{X}_t = \sqrt{\alpha_t} \mathbf{X}_0 + (1 - \alpha_t) \mathbf{Z}, \quad \mathbf{X}_0 \sim p(\mathbf{X}_0), \quad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_n). \quad (2.8)$$

With reparametrization and the Euler solver, this leads to an efficient solution to Equation 2.3, known as DDIM [128]:

$$\mathbf{x}_{t-1} = \underbrace{\sqrt{\alpha_{t-1}} \left(\frac{\mathbf{x}_t + (1 - \alpha_t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)}{\sqrt{\alpha_t}} \right)}_{=:\hat{\mathbf{x}}_0 = \text{predicted } \mathbf{x}_0} + \underbrace{\sqrt{1 - \alpha_{t-1} - \sigma_t^2} \left(-\sqrt{1 - \alpha_t} \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \right)}_{\text{direction toward } \mathbf{x}_t}. \quad (2.9)$$

For convenience, in the rest of the paper, this update will be written as: $\mathbf{x}_{t-1} \leftarrow \text{UnconditionalDDIM}(\hat{\mathbf{x}}_0, \mathbf{x}_t)$.

2.2 Tweedie’s Formula and Denoising Score Matching

In what follows, we will discuss how one can learn the score function $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ that appears in Equation 2.17. We will focus on the VE SDE, since the mathematical calculations are simpler.

Tweedie’s formula [50] is a famous result in statistics that shows that for an additive Gaussian corruption, $\mathbf{X}_t = \mathbf{X}_0 + \sigma_t \mathbf{Z}$, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_n)$, it holds that:

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) = \frac{\mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t] - \mathbf{x}_t}{\sigma_t^2}. \quad (2.10)$$

The formal statement and a self-contained proof can be found in the Appendix, Lemma A.2.

Tweedie’s formula gives us a way to derive the unconditional score function needed in Equation 2.17, by optimizing for the conditional expectation, $\mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t]$. The conditional expectation $\mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t]$, is nothing more than the minimum mean square error estimator (MMSE) of the clean image given the noisy observation \mathbf{x}_t , that is a *denoiser*.

In practice, we don’t know analytically this denoiser but we can parametrize it using a neural network $\mathbf{h}_\theta(\mathbf{x}_t)$ and learn it in a supervised way by minimizing the following objective:

$$J_{\text{DSM}}(\theta) = \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_t} [\|\mathbf{h}_\theta(\mathbf{x}_t) - \mathbf{x}_0\|^2]. \quad (2.11)$$

Assuming a rich enough family Θ , the minimizer of Equation 2.11 is $\mathbf{h}_\theta(\mathbf{x}_t) = \mathbb{E}[\mathbf{x}_0 | \mathbf{X}_t = \mathbf{x}_t]$ (see Lemma A.1) and the score in Equation 2.10 is approximated as $(\mathbf{h}_\theta(\mathbf{x}_t) - \mathbf{x}_t)/\sigma_t^2$. Note that for each σ_t we would need to learn a different denoiser (since the noise strength is different), or alternative the neural network \mathbf{h}_θ should also take as input the value of t or σ_t . Diffusion models are trained following the later paradigm, i.e. the same neural network approximates the optimal denoisers at all noise levels by conditioning it on the noise level through t .

Interestingly, Vincent [144] independently discovered that the score function can be learned by minimizing an l_2 objective, similar to Equation 2.11. The formal statement and a self-contained proof of this alternative derivation is included in the Appendix, Theorem A.3.

2.3 Latent Diffusion Processes

For high-dimensional distributions, diffusion models training (see Equation 2.11) and sampling (see Equation 2.3) require massive computational resources. To make the training and sampling more efficient, the authors of Stable Diffusion [115] propose performing the diffusion in the latent space of a pre-trained powerful autoencoder. Specifically, given an encoder $\text{Enc} : \mathbb{R}^n \rightarrow \mathbb{R}^k$ and a decoder $\text{Dec} : \mathbb{R}^k \rightarrow \mathbb{R}^n$, one can create noisy samples:

$$\mathbf{X}_t^{\text{E}} = \underbrace{\mathbf{X}_0^{\text{E}}}_{\text{Enc}(\mathbf{X}_0)} + \sigma_t \mathbf{Z}, \quad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_k), \quad (2.12)$$

and train a denoiser network in the latent space. At inference time, one starts with pure noise, samples a clean latent $\hat{\mathbf{x}}_0^{\text{E}}$ by running the reverse process, and outputs $\mathbf{x}_0 = \text{Dec}(\hat{\mathbf{x}}_0^{\text{E}})$ ¹. Solving inverse problems with Latent Diffusion models requires special treatment. We discuss the reasons and approaches in this space in Section 3.5.

2.4 Conditional Sampling

2.4.1 Stochastic Samplers for Inverse Problems

The goal in inverse problems is to sample from $p_0(\cdot | \mathbf{y})$ assuming a corruption model $\mathbf{Y} = \mathcal{A}(\mathbf{X}_0) + \sigma_{\mathbf{y}} \mathbf{Z}$, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, I_m)$. We can easily adapt the original unconditional formulation given by Equation 2.2 into a conditional one to generate samples from $p_0(\cdot | \mathbf{y})$. Specifically, the associated reverse process is given by the stochastic dynamical system [102]:

$$d\mathbf{x}_t = \left(\mathbf{f}(\mathbf{x}_t, t) - g^2(t) \underbrace{\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t | \mathbf{y})}_{\text{conditional score}} \right) dt + g(t) d\mathbf{W}_t, \quad (2.13)$$

¹To give an idea, the original work introducing latent diffusion models [115], propose to use a latent space of dimension $64 \times 64 \times 4$ to represent images of size $512 \times 512 \times 3$.

initialized at $\mathbf{x}_T \sim p_T(\cdot|\mathbf{y})$. For sufficiently large T and for linear drift functions $\mathbf{f}(\cdot, \cdot)$, the distribution $p_T(\cdot|\mathbf{y})$ is a Gaussian distribution with parameters independent of \mathbf{y} . In the conditional case, the goal becomes to estimate the score function $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y})$.

2.4.2 Deterministic Samplers for Inverse Problems

It is worth noting that (as in the unconditional setting) it is possible to derive deterministic sampling algorithms as well. Particularly, one can use the following dynamical system [134, 102]:

$$\frac{d\mathbf{x}_t}{dt} = -\frac{g^2(t)}{2} \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y}). \quad (2.14)$$

initialized at $p_T(\cdot|\mathbf{y})$ to get sample from the conditional distribution $p_0(\cdot|\mathbf{y})$. Once again, to run this discrete dynamical system, one needs to know the conditional score, $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y})$.

2.4.3 Conditional Diffusion Models

Similarly to the unconditional setting, one can directly train a network to approximate the conditional score, $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y})$. A generalization of Tweedie’s formula gives that:

$$\nabla \log p_t(\mathbf{x}_t|\mathbf{y}) = \frac{\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t, \mathbf{Y} = \mathbf{y}] - \mathbf{x}_t}{\sigma_t^2}. \quad (2.15)$$

Hence, one can train a network using a generalized version of the Denoising Score Matching,

$$J_{\text{cond, DSM}}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_t, \mathbf{y}} \left[\|\mathbf{h}_{\boldsymbol{\theta}}(\mathbf{x}_t, \mathbf{y}) - \mathbf{x}_0\|^2 \right], \quad (2.16)$$

and then use it in Equation 2.15 in place of the conditional expectation. The main issue with this approach is that the forward model (degradation operator) needs to be known at training time. If the corruption model $\mathcal{A}(\mathbf{X})$ changes, then the model needs to be retrained. Further, with this approach we need to train new models and we cannot directly leverage powerful unconditional models that are already available. The focus of this work is on methods that use pre-trained unconditional diffusion models to solve inverse problems, without further training.

2.4.4 Using pre-trained diffusion models to solve inverse problems

As we showed earlier, the conditional score can be decomposed using Bayes Rule into:

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y}) = \underbrace{\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)}_{\text{score}} + \underbrace{\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)}_{\text{measurements matching term}}. \quad (2.17)$$

that is, the (smoothed) score function, and the measurements matching term that is given by the inverse problem we are interested in solving. Applying this to equation 2.13, we get that:

$$d\mathbf{x}_t = (\mathbf{f}(\mathbf{x}_t, t) - g^2(t) (\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla \log p(\mathbf{y}|\mathbf{x}_t))) dt + g(t) d\mathbf{W}_t. \quad (2.18)$$

Similarly, one can use the deterministic process:

$$d\mathbf{x}_t = \left(\mathbf{f}(\mathbf{x}_t, t) - \frac{1}{2} g^2(t) (\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla \log p(\mathbf{y}|\mathbf{x}_t)) \right) dt. \quad (2.19)$$

We have already discussed how to train a neural network to approximate $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ (using Tweedie’s Formula / Denoising Score Matching). However, here we further need access to the term $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$. The likelihood of the measurements is given by the intractable integral:

$$p_t(\mathbf{y}|\mathbf{x}_t) = \int p(\mathbf{y}|\mathbf{x}_0) p(\mathbf{x}_0|\mathbf{x}_t) d\mathbf{x}_0. \quad (2.20)$$

Gupta et. al [58] prove that there are instances of the posterior sampling problem for which *every* algorithm takes superpolynomial time, even though *unconditional* sampling is provably fast. Hence, diffusion models excel at performing unconditional sampling but are hard to use as priors for solving inverse problems because of the time dependence in the measurements matching term. Since the very introduction of diffusion models, there has been a plethora of methods proposed to use them to solve inverse problems without retraining. This survey serves as a reference point for different techniques that have been developed in this space.

2.4.5 Ambient Diffusion: Learning to solve inverse problems using only measurements

The goal of the unsupervised learning approach for solving inverse problems (Section 2.4.4) is to use a prior $p(\mathbf{x})$ to approximate the measurements matching term, $\nabla \log p_t(\mathbf{y}|\mathbf{x}_t)$. However, in certain applications, it is expensive or even impossible to get data from (and hence learn) $p(\mathbf{x})$ in the first place. For instance, in MRI the quality of the data is proportionate to the time spent under the scanner [161] and it is infeasible to acquire full measurements from black holes [3]. This creates a chicken-egg problem: we need access to $p(\mathbf{x})$ to solve inverse problems and we do not have access to samples from $p(\mathbf{x})$ unless we can solve inverse problems. In certain scenarios, it is possible to break this seemingly impossible cycle.

Ambient Diffusion [37] was one of the first frameworks to train diffusion models with linearly corrupted data. The key concept behind the Ambient Diffusion framework is the idea of further corruption. Specifically, the given measurements get further corrupted and the model is trained to predict a clean image by using the measurements before further corruption for validation. Ambient DPS [2] shows that priors learned from corrupted data can even outperform (in terms of usefulness for inverse problems), at the high-corruption regime, priors learned from clean data. Ambient Diffusion was extended to handle additive Gaussian Noise in the measurements. The paper Consistent Diffusion Meets Tweedie [36] was the first diffusion-based framework to provide guarantees for sampling from the distribution of interest, given only access to noisy data. This paper extends the idea of further corruption to the noisy case and proposes a novel consistency loss [33] to learn the score function for diffusion times that correspond to noise levels below the level of the noise in the dataset.

Both Ambient Diffusion and Consistent Diffusion Meets Tweedie have connections to deep ideas from the literature in learning restoration models from corrupted data, such as Stein’s Unbiased Risk Estimate (SURE) [51, 136] and Noise2X [82, 79, 11]. These connections are also leveraged by alternative frameworks to Ambient Diffusion, as in [74, 1]. A different approach for learning diffusion models from measurements is based on the Expectation-Maximization (EM) algorithm [8, 119, 147]. The convergence of these methods to the true distribution depends on the convergence of the EM algorithm, which might get stuck in a local minimum.

In this survey, we focus on the setting where a pre-trained prior $p(\mathbf{x})$ is available, regardless of whether it was learned from clean or corrupted data.

3 Reconstruction Algorithms

We summarize all the methods analyzed in this work in Table 1. The methods have been taxonomized based on the approach they use to solve the inverse problem (explicit score approximations, variational methods, CSGM-type methods and asymptotically exact methods), the type of inverse problems they can solve and the optimization techniques used to solve the problem at hand (gradient descent, sampling, projections, parameter optimization). Additionally, we provide links to the official code repositories associated with the papers included in this survey. Please note that we have not conducted a review or evaluation of these codebases to verify their consistency with the corresponding papers. These links are included for informational purposes only.

Taxonomy based on the type of the reconstruction algorithm. We identified four families of methods. *Explicit Approximations for Measurement Matching:* These methods approximate the measurement matching score, $\nabla \log p_t(\mathbf{y}|\mathbf{x}_t)$, with a closed-form expression. *Variational Inference:* These methods approximate the true posterior distribution, $p(\mathbf{x}|\mathbf{y})$, with a simpler, tractable distribution. Variational formulations are then used to optimize the parameters of this simpler distribution. *CSGM-type methods:* The works in this category use backpropagation to change the initial noise of the deterministic diffusion sampler, essentially optimizing over a latent space for the diffusion model. *Asymptotically Exact Methods:* These methods aim to sample from the true posterior distribution. This is typically achieved by constructing Markov chains (MCMC) or by propagating particles through a sequence of distributions (SMC) to obtain samples that approximate the posterior. Methods that do not fall into any of these categories are classified as *Others*.

Taxonomy based on the type of optimization techniques used. The objective of all methods is to explain the measurements. The measurement consistency can be enforced with different optimization techniques, e.g. through gradients (Grad), projections (Proj), sampling (Samp), or other optimization

techniques (Opt). Methods that belong to the Grad-type take a single gradient step (either it be deterministic or stochastic) to \mathbf{x}_t to enforce measurement consistency. Proj-type projects \mathbf{x}_t or $\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]$ to the measurement subspace. Samp-type samples the next particles by defining a proposal distribution, and propagates multiple chains of particles to solve the problem. Opt-type either defines and solves an optimization problem for every timestep, or defines a global optimization problem that encompasses all timesteps. When the method belongs to more than one type, we separate them with /. Note that the categorization of different “types” is subjective, and more often than not, the category that the method belongs to may be interpreted in multiple ways. For instance, a projection step is also a gradient descent step with a specific step size.

Taxonomy based on the type of the inverse problem. Based on the linearity of the corruption operator \mathcal{A} , the inverse problems can be classified as linear or nonlinear. The inverse problems can be further categorized based on whether there is noise in the measurements. Additionally, they are classified as non-blind or blind depending on whether full information about \mathcal{A} is available. In blind problems, the degradation operator (e.g., convolution kernel, inpainting kernel) is known, while its coefficients are unknown but parametrized. For example, we might know that we have measurements with additive Gaussian noise, but the variance of the noise might be unknown. Finally, in certain inverse problems, there is additional text-conditioning. Such inverse problems are typically solved with text-to-image latent diffusion models [115].

3.1 Explicit Approximations for the Measurements Matching Term

The first family of reconstruction algorithms we identify is the one where explicit approximations for the measurements matching term, $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t)$, are made. It is important to underline that these approximations are not always clearly stated in the works that propose them, which makes it hard to understand the differences and commonalities between different methods. In what follows, we attempt to elucidate the different approximations that are being made and present different works under a common framework. To provide some insights, we often provide the explicit approximation formulas for the measurements matching term in the setting of linear inverse problems. In general, it follows the template form:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx - \frac{\mathcal{L}_t \mathcal{M}_t}{\mathcal{G}_t}. \quad (3.1)$$

Here,

- \mathcal{M}_t represents the error vector measuring the discrepancy between the observation \mathbf{y} and the estimated restored vector; for example, in Score ALD [67], $\mathcal{M}_t = \mathbf{y} - \mathbf{A}\mathbf{x}_t$.
- \mathcal{L}_t denotes a matrix that projects the error vector \mathcal{M}_t from \mathbb{R}^m back into an appropriate space in \mathbb{R}^n ; for instance, in Score ALD, $\mathcal{L}_t = \mathbf{A}^\top$.
- \mathcal{G}_t is the re-scaling scalar for the guidance vector $\mathcal{L}_t \mathcal{M}_t$; for example, in Score ALD, $\mathcal{G}_t = \sigma_{\mathbf{y}}^2 + \gamma_t^2$ with a hyperparameter γ_t .

In Figure 1, we summarize the approximation-based methods in this section using the template above. We use \propto to omit the guidance strength terms \mathcal{G}_t .

3.1.0 Sampling from a Denoiser [68]

Kadkhodaie and Simoncelli [68] introduce a method for solving linear inverse problems by using the implicit prior knowledge captured by a pre-trained denoiser on multiple noise levels. The method is anchored on Tweedie’s formula that connects the least-squares solution for Gaussian denoising to the gradient of the log-density of noisy images given in Equation 2.10

$$\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{y} + \sigma^2 \nabla_{\mathbf{y}} \log p(\mathbf{y}), \quad (3.2)$$

where $\mathbf{y} = \mathbf{x} + \mathbf{n}$, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$.

By interpreting the denoiser’s output as an approximation of this gradient, the authors develop a stochastic gradient ascent algorithm to generate high-probability samples from the implicit prior

$$\mathbf{y}_t = \mathbf{y}_{t-1} + h_t \mathbf{r}(\mathbf{y}_{t-1}) + \epsilon_t \mathbf{z}_t, \quad (3.3)$$

where $\mathbf{r}(\mathbf{y}) = \hat{\mathbf{x}}(\mathbf{y}) - \mathbf{y}$ is the denoiser residual, h_t is a step size (parameter), and ϵ_t controls the amount of newly introduced Gaussian noise \mathbf{z}_t .

To solve linear inverse problems such as deblurring, super-resolution, and compressive sensing, the generative method is extended to handle constrained sampling. Given a set of linear measurements $\mathbf{x}_c = M^\top \mathbf{x}$ of an image \mathbf{x} , where M is a low-rank measurement matrix, the goal is to reconstruct the original image by utilizing the following gradient:

$$\nabla_{\mathbf{y}} \log p(\mathbf{y}|\mathbf{x}_c) = (I - MM^\top) \mathbf{r}(\mathbf{y}) + M(\mathbf{x}_c - M^\top \mathbf{y}), \quad (3.4)$$

This approach is particularly interesting because its mathematical foundation relies solely on Tweedie’s formula, providing a simple yet powerful framework for tackling inverse problems using denoisers.

3.1.1 Score ALD [67]

One of the first proposed methods for solving *linear* inverse problems with diffusion models is the Score-Based Annealed Langevin Dynamics (Score ALD) [67] method. The approximation of this work is that:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx - \frac{\overbrace{A^\top}^{\text{"lifting" matrix}} (\underbrace{\mathbf{y} - A\mathbf{x}_t}_{\text{measurements error}})}{\underbrace{\sigma_{\mathbf{y}}^2 + \gamma_t^2}_{\text{guidance strength}}}, \quad (3.5)$$

where γ_t is a parameter to be tuned.

It is pretty straightforward to understand what this term is doing. The diffusion process is guided towards the opposite direction of the “lifting” (application of the A^\top operator) of the measurements error, i.e. $(\mathbf{y} - A\mathbf{x}_t)$, where the denominator controls the guidance strength.

3.1.2 Score-SDE [134]

Score-SDE [134] is another one of the first works that discussed solving inverse problems with pre-trained diffusion models. For linear inverse problems, the difference between Score-ALD and Score-SDE is that the latter noises the measurements before computing the measurements error. Specifically, for $t : \sigma_t > \sigma_{\mathbf{y}}$, the approximation becomes:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx - \overbrace{A^\top}^{\text{"lifting" matrix}} (\underbrace{\mathbf{y} + \sigma_t \epsilon}_{\mathbf{y}_t \text{ (noised measurements)}} - A\mathbf{x}_t) \quad (3.6)$$

where ϵ is sampled from $\mathcal{N}(\mathbf{0}, I_m)$. Here, A is an orthogonal matrix, and taking a gradient step with Equation 3.6 yields a noisy projection to $\mathbf{y}_t = A\mathbf{x}_t$ where $\mathbf{y}_t = \mathbf{y} + \sigma_t \epsilon$. Hence, we categorize Score-SDE as “projection”.

Disregarding the guidance strength of Equation 3.5, Equation 3.5 and Equation 3.6 look very similar. Indeed, the only difference is that the latter has stochasticity that arises from the noising of the measurements.

Special case: Inpainting (Repaint [89]) Observe that for the simplest case of inpainting, Equation 3.6 would be replacing the pixel values in the current estimate \mathbf{x}_t with the known pixel values from the *noised* \mathbf{y}_t . Coincidentally, this is exactly the Repaint [89] algorithm that was proposed for solving the inpainting inverse problem with pre-trained diffusion models. RePaint++ [117] improves upon this approximation to run the forward-reverse diffusion processes multiple times, so that the errors arising (e.g. boundaries) can be mitigated. This can be thought of as analogous to running MCMC corrector steps as in predictor-corrector sampling [134].

3.1.3 ILVR [22]

ILVR is a similar approach that was initially proposed for the task of super-resolution. The approximation made here is the following:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx -A^\dagger(\mathbf{y}_t - A\mathbf{x}_t) = - \overbrace{(A^\top A)^{-1} A^\top}^{\text{"lifting" matrix}} \underbrace{(\mathbf{y}_t - A\mathbf{x}_t)}_{\text{measurements error}}, \quad (3.7)$$

where A^\dagger is the Moore-Penrose pseudo-inverse of A , and similar to Score-SDE, $\mathbf{y}_t = \mathbf{y} + \sigma_t \epsilon$.

ILVR can be regarded as a pre-conditioned version of score-SDE. In ILVR, the projection to the space of images happens using the Moore-Penrose pseudo-inverse of A , instead of the simple A^\top .

3.1.4 DPS

All of the previous algorithms were proposed for linear inverse problems. Diffusion Posterior Sampling (DPS) is one of the most well known reconstruction algorithms for solving non-linear inverse problems. The underlying approximation behind DPS is that:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_0 = \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]). \quad (3.8)$$

It is easy to see that:

$$p(\mathbf{y}|\mathbf{X}_0 = \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]) = \mathcal{N}(\mathbf{y}; \boldsymbol{\mu} = \mathcal{A}(\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]), \Sigma = \sigma_{\mathbf{y}}^2 I). \quad (3.9)$$

Hence, the DPS approximation can be stated as:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx \nabla_{\mathbf{x}_t} \log \mathcal{N}(\mathbf{y}; \boldsymbol{\mu} = \mathcal{A}(\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]), \Sigma = \sigma_{\mathbf{y}}^2 I) \quad (3.10)$$

$$= \nabla_{\mathbf{x}_t} \left(\frac{1}{2\sigma_{\mathbf{y}}^2} \|\mathbf{y} - \mathcal{A}(\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t])\|^2 \right) \quad (3.11)$$

$$= \frac{1}{2\sigma_{\mathbf{y}}^2} \nabla_{\mathbf{x}_t}^\top \mathcal{A}(\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]) (\mathcal{A}(\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]) - \mathbf{y}). \quad (3.12)$$

For linear inverse problems, this simplifies to:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx - \underbrace{\frac{1}{2\sigma_{\mathbf{y}}^2}}_{\text{guidance strength}} \underbrace{\nabla_{\mathbf{x}_t}^\top \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t] A^\top}_{\text{"lifting" matrix}} \underbrace{(\mathbf{y} - A\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t])}_{\text{measurements error}}. \quad (3.13)$$

We can further use Tweedie's formula to further write it as:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx - \frac{1}{2\sigma_{\mathbf{y}}^2} (I + \nabla_{\mathbf{x}_t}^2 \log p_t(\mathbf{x}_t))^\top A^\top (\mathbf{y} - A\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]). \quad (3.14)$$

In practice, DPS does not use the theoretical guidance strength but instead proposes to use a reweighting with a step size inversely proportional to the norm of the measurement error.

MCG [27] provides a geometric interpretation of DPS by showing that the approximation used in DPS can guarantee the noisy samples stay on the manifold. DSG [156] showed that one can choose a theoretically "correct" step size under the geometric view of MCG, and combined with projected gradient descent, one can achieve superior sample quality. MPGD [61] showed that by constraining the gradient update step to stay on the low dimensional subspace by autoencoding, one can acquire better results.

3.1.5 Π GDM [129]

Recall the intractable integral in Equation 1.3. According to this relation, the DPS approximation is achieved by setting

$$p(\mathbf{X}_0|\mathbf{X}_t) \approx \delta(\mathbf{X}_0 - \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]). \quad (3.15)$$

In Π GDM, the authors propose to use a Gaussian distribution for approximation

$$p(\mathbf{X}_0|\mathbf{X}_t) \approx \mathcal{N}(\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t], r_t^2 I_n), \quad (3.16)$$

where r_t is a hyperparameter. For linear inverse problems, this leads to

$$p(\mathbf{y}|\mathbf{X}_t) \approx \mathcal{N}(A\hat{\mathbb{E}}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t], r_t^2 AA^\top + \sigma_y^2 I_n). \quad (3.17)$$

Subsequently, we have

$$\begin{aligned} & \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \\ & \approx - \underbrace{\frac{\partial \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]}{\partial \mathbf{x}_t}}_{\text{"lifting" matrix}} (r_t^2 AA^\top + \sigma_y^2 I)^{-1} A^\top (\underbrace{\mathbf{y} - A\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]}_{\text{measurements error}}). \end{aligned} \quad (3.18)$$

3.1.6 Moment Matching [119]

In Π GDM, the distribution $p(\mathbf{x}_0|\mathbf{x}_t)$ was assumed to be isotropic Gaussian. However, one can calculate explicitly the variance matrix, $V[\mathbf{x}_0|\mathbf{x}_t]$. As shown in Lemma A.4, it holds that:

$$V[\mathbf{x}_0|\mathbf{x}_t] = \sigma_t^4 H(\log p_t(\mathbf{x}_t)) + \sigma_t^2 I_n \quad (3.19)$$

$$= \sigma_t^2 \nabla_{\mathbf{x}_t} \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]. \quad (3.20)$$

The Moment Matching [119] method approximates the distribution $p(\mathbf{x}_0|\mathbf{x}_t)$ with an anisotropic Gaussian:

$$p(\mathbf{x}_0|\mathbf{x}_t) \approx \mathcal{N}(\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t], V[\mathbf{x}_0|\mathbf{x}_t]). \quad (3.21)$$

For linear inverse problems, this leads to the following approximation for the measurements' score:

$$\begin{aligned} & \nabla \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \\ & \approx - \underbrace{\nabla_{\mathbf{x}_t} \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]^\top A^\top (\sigma_y^2 I + A \sigma_t^2 \nabla_{\mathbf{x}_t} \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t] A^\top)^{-1}}_{\text{"lifting" matrix}} (\underbrace{\mathbf{y} - A\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]}_{\text{measurements error}}). \end{aligned} \quad (3.22)$$

In high-dimensions, even materializing the matrix $\nabla_{\mathbf{x}_t} \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]$ is computationally intensive. Instead, the authors of [119] use automatic differentiation to compute the Jacobian-vector products.

3.1.7 BlindDPS [23]

Methods that were considered so far were designed for non-blind inverse problems, where A is fully known. BlindDPS targets the case where we have a parametrized *unknown* forward model A_ϕ (e.g. blurring with an unknown kernel ϕ). In BlindDPS, on top of the posterior mean approximation of \mathbf{x}_t , one approximates the parameter of the forward model, again, with the posterior mean. Specifically, we design two parallel generative SDEs

$$d\mathbf{x}_t = (\mathbf{f}(\mathbf{x}_t, t) - g^2(t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t, \phi_t|\mathbf{y})) dt + g(t) d\mathbf{W}_t \quad (3.23)$$

$$d\phi_t = (\mathbf{f}(\phi_t, t) - g^2(t) \nabla_{\phi_t} \log p_t(\mathbf{x}_t, \phi_t|\mathbf{y})) dt + g(t) d\mathbf{W}_t, \quad (3.24)$$

where the two SDEs are coupled through $\log p_t(\mathbf{x}_t, \phi_t|\mathbf{y})$, where under the independence between \mathbf{X}_t and Φ_t , the Bayes rule reads

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t, \phi_t|\mathbf{y}) = \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t, \Phi_t = \phi_t) \quad (3.25)$$

$$\nabla_{\phi_t} \log p_t(\mathbf{x}_t, \phi_t|\mathbf{y}) = \nabla_{\phi_t} \log p_t(\phi_t) + \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t, \Phi_t = \phi_t), \quad (3.26)$$

where we see that \mathbf{X}_t and Φ_t are coupled through the likelihood $p(\mathbf{y}|\mathbf{X}_t, \Phi_t)$. In BlindDPS, the approximation used in DPS is applied to both the image and the operator, leading to

$$p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t, \Phi_t = \phi_t) \approx p(\mathbf{y}|\mathbf{X}_0 = \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t], \Phi_0 = \mathbb{E}[\Phi_0|\Phi_t = \phi_t]). \quad (3.27)$$

The gradient of the coupled likelihood with respect to \mathbf{x}_t leads to

$$\begin{aligned} & \text{guidance strength} \rightarrow \frac{1}{2\sigma_y^2} \xrightarrow{\text{"lifting" matrix}} \nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t, \Phi_t = \phi_t) \xrightarrow{\text{measurements error}} \\ & \approx - \frac{1}{2\sigma_y^2} \nabla_{\mathbf{x}_t}^\top \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t] A_{\mathbb{E}[\Phi_0|\Phi_t = \phi_t]}^\top \left(\mathbf{y} - A_{\mathbb{E}[\Phi_0|\Phi_t = \phi_t]} \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t] \right). \end{aligned} \quad (3.28)$$

Similarly, for ϕ_t , we have

$$\begin{aligned} & \text{guidance strength} \rightarrow \frac{1}{2\sigma_y^2} \xrightarrow{\text{"lifting" matrix}} \nabla_{\phi_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t, \Phi_t = \phi_t) \xrightarrow{\text{measurements error}} \\ & \approx - \frac{1}{2\sigma_y^2} \nabla_{\phi_t}^\top \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t] A_{\mathbb{E}[\Phi_0|\Phi_t = \phi_t]}^\top \left(\mathbf{y} - A_{\mathbb{E}[\Phi_0|\Phi_t = \phi_t]} \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t] \right). \end{aligned} \quad (3.29)$$

3.1.8 DDRM Family

The methods under the DDRM family poses all linear inverse problems to a noisy inpainting problem, by decomposing the measurement matrix with singular value decomposition (SVD), i.e. $A = U\Sigma V^\top$, where $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix with singular values $\{s_j\}_{j=1}^m$ as the elements. One can then rewrite $\mathbf{y} = A\mathbf{x} + \sigma_y \mathbf{z}$, $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I_m)$ as

$$\bar{\mathbf{y}} = \Sigma \bar{\mathbf{x}} + \sigma_y \bar{\mathbf{z}}, \quad \text{where} \quad \bar{\mathbf{y}} := U^\top \mathbf{y}, \quad \bar{\mathbf{x}} := V^\top \mathbf{x}, \quad \bar{\mathbf{z}} := U^\top \mathbf{z}. \quad (3.30)$$

Subsequently, Equation 3.30 becomes an inpainting problem in the spectral space.

SNIPS [74]. SNIPS proceeds by first solving the inverse problem posed as Equation 3.30 in the spectral space to achieve a sample $\bar{\mathbf{x}} \sim p(\bar{\mathbf{x}}|\bar{\mathbf{y}})$, then retrieving the posterior sample with $\hat{\mathbf{x}} = V\bar{\mathbf{x}}$. The key approximation can be concisely represented as

$$\nabla_{\bar{\mathbf{x}}_t} \log p(\bar{\mathbf{y}}|\bar{\mathbf{X}}_t = \mathbf{x}_t) \approx - \frac{\Sigma^\top |\sigma_y^2 I_m - \sigma_t^2 \Sigma \Sigma^\top|^\dagger}{\text{"lifting" matrix}} (\bar{\mathbf{y}} - \Sigma \bar{\mathbf{x}}_t), \quad (3.31)$$

↑ measurements error

For the simplest case of denoising where $m = n$ and $\Sigma = A = I$, the method becomes [75]

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx \frac{\mathbf{y} - \mathbf{x}_t}{|\sigma_y^2 - \sigma_t^2|}. \quad (3.32)$$

which produces a vector direction that is weighted by the absolute difference between the diffusion noise level σ_t^2 , and the measurement noise level σ_y^2 . For the fully general case in Equation 3.31, elements in different indices are weighted according to the singular values contained in Σ . In practice, SNIPS uses pre-conditioning with the approximate negative inverse Hessian of $\log p(\bar{\mathbf{x}}_t|\bar{\mathbf{y}})$ when running annealed Langevin dynamics.

DDRM [73]. DDRM extends SNIPS by leveraging the posterior mean $\bar{\mathbf{x}}_{0|t} := V\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]$ in the place of $\bar{\mathbf{x}}_t$ used in SNIPS. i.e.,

$$\nabla_{\bar{\mathbf{x}}_t} \log p(\bar{\mathbf{y}}|\bar{\mathbf{X}}_t = \mathbf{x}_t) \approx - \frac{\Sigma^\top |\sigma_y^2 I_m - \sigma_t^2 \Sigma \Sigma^\top|^\dagger}{\text{"lifting" matrix}} (\bar{\mathbf{y}} - \Sigma \bar{\mathbf{x}}_{0|t}). \quad (3.33)$$

↑ measurements error

Expressing Equation 3.33 element-wise, we get

$$p(\bar{\mathbf{x}}_t^{(i)}|\mathbf{X}_{t+1} = \mathbf{x}_{t+1}, \mathbf{y}) = \begin{cases} \mathcal{N}(\bar{\mathbf{x}}_t^{(i)}; \bar{\mathbf{x}}_{0|t+1}^{(i)}, \sigma_t^2) & \text{if } s_i = 0 \\ \mathcal{N}(\bar{\mathbf{x}}_t^{(i)}; \bar{\mathbf{x}}_{0|t+1}^{(i)}, \sigma_t^2) & \text{if } \sigma_t < \frac{\sigma_y}{s_i} \\ \mathcal{N}(\bar{\mathbf{x}}_t^{(i)}; \bar{\mathbf{y}}^{(i)}, \sigma_t^2 - \frac{\sigma_y^2}{s_i^2}) & \text{if } \sigma_t \geq \frac{\sigma_y}{s_i} \end{cases}, \quad (3.34)$$

where $\mathbf{x}^{(i)}$ denotes the i -th element of the vector, and s_i its corresponding singular value. Here, DDRM introduces another hyper-parameter η to control the stochasticity of the sampling process

$$p(\bar{\mathbf{x}}_t^{(i)} | \mathbf{X}_{t+1} = \mathbf{x}_{t+1}, \mathbf{y}) = \begin{cases} \mathcal{N}(\bar{\mathbf{x}}_t^{(i)}; \bar{\mathbf{x}}_{0|t+1}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{x}}_{t+1}^{(i)} - \bar{\mathbf{x}}_{0|t+1}^{(i)}}{\sigma_{t+1}}, \eta^2 \sigma_t^2) & \text{if } s_i = 0 \\ \mathcal{N}(\bar{\mathbf{x}}_t^{(i)}; \bar{\mathbf{x}}_{0|t+1}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{y}}^{(i)} - \bar{\mathbf{x}}_{0|t+1}^{(i)}}{\sigma_{\mathbf{y}}/s_i}, \eta^2 \sigma_t^2) & \text{if } \sigma_t < \frac{\sigma_{\mathbf{y}}}{s_i}, \\ \mathcal{N}(\bar{\mathbf{x}}_t^{(i)}; \bar{\mathbf{y}}^{(i)}, \sigma_t^2 - \frac{\sigma_{\mathbf{y}}^2}{s_i^2}) & \text{if } \sigma_t \geq \frac{\sigma_{\mathbf{y}}}{s_i} \end{cases} \quad (3.35)$$

with $\eta \in (0, 1]$ such that $\eta = 1.0$ recovers Equation 3.34.

GibbsDDRM. GibbsDDRM [99] extends DDRM to the following blind linear problem $\mathbf{y} = A_\varphi \mathbf{x} + \sigma_{\mathbf{y}} \mathbf{z}$, where A_φ is a linear operator parameterized by φ . Here, $A_\varphi = U_\varphi \Sigma_\varphi V_\varphi^\top$ has a φ dependence SVD decomposition with singular values $\{s_{j,\varphi}\}_{j=1}^m$ as the elements of the diagonal matrix Σ_φ . In the spectral space, $\bar{\mathbf{y}}_\varphi := U_\varphi^\top \mathbf{y}_\varphi$, $\bar{\mathbf{x}}_\varphi := V_\varphi^\top \mathbf{x}_\varphi$, $\bar{\mathbf{z}}_\varphi := U_\varphi^\top \mathbf{z}_\varphi$. Subsequently, the posterior mean in DDRM is replaced with $\bar{\mathbf{x}}_{0|t,\varphi} := V_\varphi \mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t]$, also depending on φ . Thus, it leads to the sampling process

$$p(\bar{\mathbf{x}}_{t,\varphi}^{(i)} | \mathbf{X}_{t+1} = \mathbf{x}_{t+1}, \mathbf{y}, \varphi) = \begin{cases} \mathcal{N}(\bar{\mathbf{x}}_{t,\varphi}^{(i)}; \bar{\mathbf{x}}_{0|t+1,\varphi}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{x}}_{t+1,\varphi}^{(i)} - \bar{\mathbf{x}}_{0|t+1,\varphi}^{(i)}}{\sigma_{t+1}}, \eta^2 \sigma_t^2) & \text{if } s_{i,\varphi} = 0 \\ \mathcal{N}(\bar{\mathbf{x}}_{t,\varphi}^{(i)}; \bar{\mathbf{x}}_{0|t+1,\varphi}^{(i)} + \sqrt{1 - \eta^2} \sigma_t \frac{\bar{\mathbf{y}}_\varphi^{(i)} - \bar{\mathbf{x}}_{0|t+1,\varphi}^{(i)}}{\sigma_{\mathbf{y}}/s_{i,\varphi}}, \eta^2 \sigma_t^2) & \text{if } \sigma_t < \frac{\sigma_{\mathbf{y}}}{s_{i,\varphi}} \\ \mathcal{N}(\bar{\mathbf{x}}_{t,\varphi}^{(i)}; \bar{\mathbf{y}}_\varphi^{(i)}, \sigma_t^2 - \frac{\sigma_{\mathbf{y}}^2}{s_{i,\varphi}^2}) & \text{if } \sigma_t \geq \frac{\sigma_{\mathbf{y}}}{s_{i,\varphi}} \end{cases} \quad (3.36)$$

At time step t , φ is sampled by using the conditional distribution $p(\varphi | \mathbf{x}_{t:T}, \mathbf{y})$ and updated for several iterations in a Langevin manner:

$$\varphi \leftarrow \varphi + \frac{\xi}{2} \nabla_\varphi \log p(\varphi | \mathbf{x}_{t:T}, \mathbf{y}) + \sqrt{\xi} \epsilon,$$

where ξ is a stepsize and $\epsilon \sim \mathcal{N}(\mathbf{0}, I_n)$. Here, $\nabla_\varphi \log p(\varphi | \mathbf{x}_{t:T}, \mathbf{y}) \approx \nabla_\varphi \log p(\varphi | \bar{\mathbf{x}}_{0|t,\varphi}, \mathbf{y})$, and the gradient can be computed as:

$$\nabla_\varphi \log p(\varphi | \bar{\mathbf{x}}_{0|t,\varphi}, \mathbf{y}) = -\frac{1}{2\sigma_{\mathbf{y}}^2} \nabla_\varphi \|\mathbf{y} - \mathbf{A}_\varphi \bar{\mathbf{x}}_{0|t,\varphi}\|^2. \quad (3.37)$$

3.1.9 DDNM [148] family

A different way to find meaningful approximations for the conditional score is to look at the conditional version of Tweedie's formula, see Equation 2.15. Using Bayes rule and rearranging [111], we have

$$\mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t, \mathbf{y}] = \mathbf{x}_t + \sigma_t^2 \nabla_{\mathbf{x}_t} \log p_t(\mathbf{X}_t | \mathbf{y}) \quad (3.38)$$

$$= \mathbf{x}_t + \sigma_t^2 \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \sigma_t^2 \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} | \mathbf{X}_t = \mathbf{x}_t) \quad (3.39)$$

$$= \mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t] + \sigma_t^2 \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} | \mathbf{X}_t = \mathbf{x}_t). \quad (3.40)$$

The methods that belong to the DDNM family make approximations to $\mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t, \mathbf{y}]$ by making certain data consistency updates to $\mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t]$.

DDNM [148]. The simplest form of update when considering no noise can be obtained through range-null space decomposition, assuming that one can compute the pseudo-inverse. In DDNM, this condition is trivially met by considering operations that are SVD-decomposable. DDNM proposes to use the following projection step to the posterior mean to obtain an approximation of the conditional posterior mean

$$\mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t, \mathbf{y}] \approx (I - A^\dagger A) \mathbb{E}[\mathbf{X}_0 | \mathbf{X}_t = \mathbf{x}_t] + A^\dagger \mathbf{y}, \quad (3.41)$$

where A^\dagger is the Moore-Penrose pseudo-inverse of A . One can also express Equation 3.41 as an approximation of the likelihood, consistent to other methods in the chapter. Specifically, notice that by using the relation in Equation 3.40,

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) = \frac{1}{\sigma_t^2} (\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t, \mathbf{y}] - \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]). \quad (3.42)$$

Plugging in Equation 3.41 to Equation 3.42,

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx \underbrace{-\frac{1}{\sigma_t^2}}_{\text{guidance strength}} \underbrace{A^\dagger}_{\text{"lifting" matrix}} \left(\underbrace{\mathbf{y} - A\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t, \mathbf{y}]}_{\text{measurements error}} \right) \quad (3.43)$$

When there is noise in the measurement, one can make soft updates

$$\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t, \mathbf{y}] \approx (I - \Sigma_t A^\dagger A) \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t] + \Sigma_t A^\dagger \mathbf{y}, \quad \Sigma \in \mathbb{R}^{n \times n}. \quad (3.44)$$

Also, similar to Equation 3.43,

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{X}_t = \mathbf{x}_t) \approx \underbrace{-\frac{1}{\sigma_t^2}}_{\text{guidance strength}} \underbrace{\Sigma_t A^\dagger}_{\text{"lifting" matrix}} \left(\underbrace{\mathbf{y} - A\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t, \mathbf{y}]}_{\text{measurements error}} \right) \quad (3.45)$$

Here, one can choose a simple $\Sigma_t = \lambda_t I$ with λ_t set as a hyper-parameter, or use different scaling for each spectral component. Observe that due to the relationship between the (conditional) score function and the posterior mean established in Equation 3.40, we can also easily rewrite the approximation in terms of the score of the posterior.

DDS [25], DiffPIR [163]. Both DDS and DiffPIR propose a proximal update to approximate the conditional posterior mean, albeit from different motivations. The resulting approximation reads

$$\mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t, \mathbf{y}] \approx \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \frac{\lambda_t}{2} \|\mathbf{x} - \mathbb{E}[\mathbf{X}_0|\mathbf{X}_t = \mathbf{x}_t]\|^2. \quad (3.46)$$

The difference between the two algorithms comes from how one solves the optimization problem in Equation 3.46, and how one chooses the hyperparameter λ_t . In DDS, the optimization is solved with a few-step conjugate gradient (CG) update steps, by showing that DPS gradient update steps can be effectively replaced with the CG steps under assumptions on the data manifold [25]. λ_t is taken to be a constant value across all t . DiffPIR uses a closed-form solution for Equation 3.46, and proposes a schedule for λ_t that is proportional to the signal-to-noise (SNR) ratio of the diffusion at time t . Specifically, one chooses $\lambda_t = \sigma_t \zeta$, where ζ is a constant.

3.2 Variational Inference

These methods approximate the true posterior distribution, $p(\mathbf{x}|\mathbf{y})$, with a simpler, tractable distribution. Variational formulations are then used to optimize the parameters of this simpler distribution.

3.2.1 RED-Diff [94]

Mardani et al. [94] introduce RED-diff, a new approach for solving inverse problems by leveraging stochastic optimization and diffusion models. The core idea is to use variational method by introducing a simpler distribution, $q := \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I_n)$, to approximate the true posterior $p(\mathbf{x}_0|\mathbf{y})$ by minimizing the KL divergence \mathcal{D}_{KL} between them:

$$\min_q \mathcal{D}_{\text{KL}}(q(\mathbf{x}_0|\mathbf{y})\|p(\mathbf{x}_0|\mathbf{y})). \quad (3.47)$$

Here, $\mathcal{D}_{\text{KL}}(q(\mathbf{x}_0|\mathbf{y})\|p(\mathbf{x}_0|\mathbf{y}))$ can be written as follows:

$$\mathcal{D}_{\text{KL}}(q(\mathbf{x}_0|\mathbf{y})\|p(\mathbf{x}_0|\mathbf{y})) = \underbrace{-\mathbb{E}_{q(\mathbf{x}_0|\mathbf{y})}[\log p(\mathbf{y}|\mathbf{x}_0)] + \mathcal{D}_{\text{KL}}(q(\mathbf{x}_0|\mathbf{y})\|p(\mathbf{x}_0))}_{\text{Variational Bound (VB)}} + \text{constant}. \quad (3.48)$$

via classic variational inference argument. The first term in VB can be simplified into reconstruction loss, and the second term can be decomposed as score-matching objective which involves matching the score function of the variational distribution with the score function of the true posterior denoisers at different timesteps:

$$\min_{\mu} \frac{\|\mathbf{y} - \mathcal{A}(\mu)\|^2}{2\sigma_{\mathbf{y}}^2} + \mathbb{E}_{t, \epsilon} [\lambda_t \|\epsilon_{\theta}(\mathbf{x}_t; t) - \epsilon\|^2] \quad (3.49)$$

where μ is the mean of the variational distribution, and σ_v^2 is the noise variance in the observation, $\epsilon_{\theta}(\mathbf{x}_t; t)$ is the score function of the diffusion model at timestep (t) and λ_t is a time-weighting factor.

Sampling as optimization. The goal is then to find an image μ that reconstructs the observation \mathbf{y} given by f , while having a high likelihood under the denoising diffusion prior (regularizer). This score-matching objective is optimized using stochastic gradient descent, effectively turning the sampling problem into an optimization problem. The weighting factor (λ_t) is chosen based on the signal-to-noise ratio (SNR) at each timestep to balance the contribution of different denoisers in the diffusion process.

3.2.2 Blind RED-Diff [6]

In [6] authors introduce blind RED-diff, an extension of the RED-diff framework [94] to solve blind inverse problems. The main idea is to use variational inference to jointly estimate the latent image and the unknown forward model parameters.

Similar to RED-Diff, the key mathematical formulation is the minimization of the KL-divergence between the true posterior distribution $p(\mathbf{x}_0, \gamma|\mathbf{y})$ and a variational approximation $q(\mathbf{x}_0, \gamma|\mathbf{y})$:

$$\min_q \mathcal{D}_{\text{KL}}(q(\mathbf{x}_0, \gamma|\mathbf{y}) \| p(\mathbf{x}_0, \gamma|\mathbf{y})).$$

If we assume the latent image and the forward model parameters are independent, the KL-divergence can be decomposed as:

$$\mathcal{D}_{\text{KL}}(q(\mathbf{x}_0|\mathbf{y}) \| p(\mathbf{x}_0)) + \mathcal{D}_{\text{KL}}(q(\gamma|\mathbf{y}) \| p(\gamma)) - \mathbb{E}_{q(\mathbf{x}_0, \gamma|\mathbf{y})} [\log p(\mathbf{y}|\mathbf{x}_0, \gamma)] + \log p(\mathbf{y}).$$

The minimization with respect to q involves three terms:

- i. $\mathcal{D}_{\text{KL}}(q(\mathbf{x}_0|\mathbf{y}) \| p(\mathbf{x}_0))$ represents the KL divergence between the variational distribution of the image (\mathbf{x}_0) and its prior distribution. This term is approximated using a score-matching loss, which leverages denoising score matching with a diffusion model (as in RED-Diff).
- ii. $\mathcal{D}_{\text{KL}}(q(\gamma|\mathbf{y}) \| p(\gamma))$ is the KL divergence between the variational distribution of the forward model parameters (γ) and their prior distribution. This term acts as a *regularizer* on γ .
- iii. $-\mathbb{E}_{q(\mathbf{x}_0, \gamma|\mathbf{y})} [\log p(\mathbf{y}|\mathbf{x}_0, \gamma)]$ is the expectation of the negative log-likelihood of the observed data \mathbf{y} given the image \mathbf{x}_0 and the forward model parameters γ . This term ensures data consistency.

The resulting optimization can be achieved using alternating stochastic optimization, where the image \mathbf{x}_0 and the forward model parameters γ are updated iteratively.

The formulation assumes conditional independence between \mathbf{x}_0 and γ given the measurement \mathbf{y} , and it also requires a specific form for the prior distribution $p(\gamma)$.

3.2.3 Score Prior [54]

We again start by introducing a variational distribution $q_{\phi}(\mathbf{x}_0)$ that aims to approximate the posterior distribution determined by the diffusion prior. The optimization problem becomes

$$\min_{\phi} \mathcal{D}_{\text{KL}}(q_{\phi}(\mathbf{x}_0) \| p_{\theta}(\mathbf{x}_0|\mathbf{y})) \quad (3.50)$$

$$\min_{\phi} \int q_{\phi}(\mathbf{x}_0|\mathbf{y}) [-\log p(\mathbf{y}|\mathbf{x}_0) - \log p_{\theta}(\mathbf{x}_0) + \log q_{\phi}(\mathbf{x}_0)]. \quad (3.51)$$

One of the most expressive yet tractable proposal distributions is normalizing flows (NF) [112, 44]. Choosing q_ϕ to be an NF, we can transform the optimization problem to

$$\min_{\phi} \mathbb{E}_{\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)} \left[\underbrace{-\log p(\mathbf{y}|G_\phi(\mathbf{z}))}_{\text{Likelihood}} - \underbrace{\log p_\theta(G_\phi(\mathbf{z}))}_{\text{Prior}} + \underbrace{\log \pi(\mathbf{z}) - \log \left| \det \frac{dG_\phi(\mathbf{z})}{d\mathbf{z}} \right|}_{\text{Entropy}} \right], \quad (3.52)$$

where the expectation is over the input latent variable \mathbf{z} , and π is the reference Gaussian distribution. Observe that the likelihood term and the entropy can be efficiently computed with a single forward/backward pass through the NF due to the parametrization of q_ϕ with an NF. All that is left for us is to compute the prior term $\log p_\theta(G_\phi(\mathbf{z}))$. In score prior [54], this is solved by leveraging the instantaneous change-of-variables formula with the diffusion PF-ODE, as originally proposed in [134]

$$\log p_\theta(\mathbf{x}_0) = \log p_T(\mathbf{x}_T) + \int_0^T \nabla \cdot \tilde{\mathbf{f}}_\theta(\mathbf{x}_t, t) dt, \quad (3.53)$$

where $\tilde{\mathbf{f}}_\theta(\mathbf{x}_t, t)$ is the drift term of the reverse SDE in Equation 2.2 with the score replaced by the network approximation. Notice that by plugging in Equation 3.53 to Equation 3.52, we can optimize the NF model in an unsupervised fashion. Notice that while this formulation does not incur approximation errors, it is very costly as every optimization steps involve computing Equation 3.53. Moreover, observe that the training of NF is done for a specific measurement \mathbf{y} . One has to run Equation 3.52 for every different measurement that one wishes to recover.

3.2.4 Efficient Score Prior [53]

As computing Equation 3.53 is costly, Feng *et al.* proposed to optimize q_ϕ with the evidence lower bound (ELBO), originally presented in the work of Score-flow [131] $b_\theta(\mathbf{x}_0) \leq \log p_\theta(\mathbf{x}_0)$

$$b_\theta(\mathbf{x}_0) = \mathbb{E}_{p(\mathbf{x}_T|\mathbf{x}_0)} [\log \pi(\mathbf{x}_T)] - \frac{1}{2} \int_0^T g(t)^2 h(t) dt, \quad (3.54)$$

where

$$h(t) := \mathbb{E}_{p(\mathbf{x}_t|\mathbf{x}_0)} \left[\underbrace{\frac{1}{\sigma_t^4} \|\mathbf{h}_\theta(\mathbf{x}_t) - \mathbf{x}_0\|_2^2}_{\text{Denoising loss}} - \|\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_0)\|_2^2 - \frac{2}{g(t)^2} \nabla_{\mathbf{x}_t} \cdot \mathbf{f}(\mathbf{x}_t, t) \right]. \quad (3.55)$$

Intuitively, the value of b_θ is small when we have a small denoising loss, and large when our diffusion denoiser \mathbf{h}_θ cannot properly denoise the given image. Replacing the exact likelihood Equation 3.53 that requires hundreds to thousands of NFEs to the surrogate denoising likelihood Equation 3.54 that requires only a single NFE makes the method much more efficient and scalable to higher dimensions.

3.3 Asymptotically Exact Methods

These methods aim to sample from the true posterior distribution. Of course, the intractability of the posterior distribution cannot be circumvented but what these methods trade compute for approximation error: as the number of network evaluations increases to infinity, these methods will asymptotically converge to the true posterior (assuming no other approximation errors).

3.3.1 Plug and Play Diffusion Models (PnP-DM) [151]

As explained in the introduction, the end goal is to sample from the distribution $p(\mathbf{x}_0|\mathbf{y}) \propto p(\mathbf{x}_0)p(\mathbf{y}|\mathbf{x})$. The authors of [151] introduce an auxiliary variable \mathbf{z} and an auxiliary distribution:

$$\pi(\mathbf{x}_0, \mathbf{z}|\mathbf{y}) \propto p(\mathbf{x}_0) \cdot p(\mathbf{y}|\mathbf{z}) \cdot \exp \left(-\frac{1}{2\rho^2} \|\mathbf{x}_0 - \mathbf{z}\|^2 \right). \quad (3.56)$$

It is easy to see that as $\rho \rightarrow 0$, the auxiliary distribution converges to the target distribution $p(\mathbf{x}_0|\mathbf{y})$.

To sample from the joint distribution $\pi(\mathbf{x}_0, \mathbf{z}|\mathbf{y})$, the authors use Gibbs Sampling, i.e. the alternate between sampling from the posteriors. Specifically, the sampling algorithm alternates between two steps:

- Likelihood term:

$$\mathbf{z}^{(k)} \sim \pi(\mathbf{z}|\mathbf{y}, \mathbf{x}_0^{(k)}) \propto p(\mathbf{y}|\mathbf{z}) \cdot \exp\left(-\frac{1}{2\rho^2} \|\mathbf{x}_0^{(k)} - \mathbf{z}\|^2\right). \quad (3.57)$$

- Prior term:

$$\mathbf{x}_0^{(k+1)} \sim \pi(\mathbf{x}_0|\mathbf{y}, \mathbf{z}^{(k)}) \propto p(\mathbf{x}_0) \cdot \exp\left(-\frac{1}{2\rho^2} \|\mathbf{x}_0^{(k)} - \mathbf{z}^{(k)}\|^2\right). \quad (3.58)$$

The likelihood term samples a vector that satisfies the measurements and is close to $\mathbf{x}_0^{(k)}$. The prior term samples a vector that is likely under $p(\mathbf{x}_0)$ and is close to $\mathbf{z}^{(k)}$. For most problems of interest, sampling from Equation 3.57 is easy because the distribution is log-concave, e.g. that's the case for linear inverse problems. The interesting observation is that sampling from Equation 3.58 corresponds to a denoising problem, for which diffusion models excel. Indeed, for any \mathbf{x}_t at noise level σ_t , we have that:

$$p(\mathbf{x}_0|\mathbf{x}_t) \propto p(\mathbf{x}_0)p(\mathbf{x}_t|\mathbf{x}_0) = p(\mathbf{x}_0)\exp\left(-\frac{1}{2\sigma_t^2} \|\mathbf{x}_0 - \mathbf{x}_t\|^2\right). \quad (3.59)$$

Hence, to sample from Equation 3.58, one initializes the reverse process at $\mathbf{z}^{(k)}$ and time t such that: $\sigma_t = \rho$.

3.3.2 FPS [47]

FPS connects posterior sampling to Bayesian filtering and uses sequential Monte Carlo methods to solve the filtering problem, avoiding the need to handcraft approximations to the posterior $p(\mathbf{y}|\mathbf{x}_t)$. Given an observation \mathbf{y} , FPS proposes to first construct a sequence $\{\mathbf{y}_t\}_{t=0}^N$ from \mathbf{y} , and then determine a tractable distribution $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_{t-1})$. Starting from $\mathbf{x}_N \sim \mathcal{N}(\mathbf{0}, I_n)$, FPS can then recursively sample \mathbf{x}_t for $t = N-1, \dots, 1$, and finally obtain \mathbf{x}_0 . Specifically, FPS consists of two steps:

Step 1. Generating a sequence of $\{\mathbf{y}_t\}_{t=0}^N$ with an observation \mathbf{y} . This can be done either using the forward process or unconditional DDIM backward sampling.

For the construction via the forward process, we recursively construct \mathbf{y}_t as follows:

$$\mathbf{y}_t = \mathbf{y}_{t-1} + \sigma_t A \mathbf{z}_t, \quad \text{initialized with } \mathbf{y}_0 := \mathbf{y}. \quad (3.60)$$

This arises from $\mathbf{x}_t = \mathbf{x}_{t-1} + \sigma_t \mathbf{z}_t$ and applying the linear operator A to it.

For the construction via backward sampling, FPS uses methods such as unconditional DDIM as in Equation 2.9,

$$\mathbf{y}_{t-1} = \underbrace{u_t \mathbf{y}_0}_{\text{clean}} + \underbrace{v_t \mathbf{y}_t}_{\text{direction to time } t \text{ sample}} + \underbrace{w_t A \mathbf{z}_t}_{\text{noise}}, \quad \text{initialized with } \mathbf{y}_N \sim \mathcal{N}(\mathbf{0}, AA^\top). \quad (3.61)$$

Here, u_t , v_t , and w_t are DDIM coefficients that can be explicitly computed. Note that \mathbf{y}_N is sampled from $\mathcal{N}(\mathbf{0}, AA^\top)$ because the prior distribution of the diffusion model is a standard Gaussian $\mathbf{x}_N \sim \mathcal{N}(\mathbf{0}, I)$, and due to the linearity of the inverse problem, $\mathbf{y}_N = A\mathbf{x}_N$.

Step 2. Generating a backward sequence of $\{\mathbf{x}_t\}_{t=0}^N$ from Step 1's $\{\mathbf{y}_t\}_{t=0}^N$. First, note that $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{y}_{t-1})$ is a tractable normal distribution. This results from applying Bayes' rule and the conditional independence of \mathbf{x}_t and the random vector \mathbf{Y}_{t-1} given \mathbf{x}_{t-1} :

$$p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{Y}_{t-1}) \propto p(\mathbf{x}_{t-1}|\mathbf{x}_t)p(\mathbf{Y}_{t-1}|\mathbf{x}_{t-1}). \quad (3.62)$$

Here, $p(\mathbf{x}_{t-1}|\mathbf{x}_t)$ is approximated via backward diffusion sampling with learned scores, and $p(\mathbf{Y}_{t-1}|\mathbf{x}_{t-1}) = \mathcal{N}(A\mathbf{x}_{t-1}, c_{t-1}^2 I)$, where c_{t-1} , dependent on $\sigma_{\mathbf{y}} > 0$, can be computed explicitly [133]. Thus, with $\{\mathbf{y}_t\}_{t=0}^N$ and initial condition $\mathbf{x}_N \sim \mathcal{N}(\mathbf{0}, I_n)$, FPS recursively samples $\mathbf{x}_{N-1}, \dots, \mathbf{x}_1$ using $p(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{Y}_{t-1} = \mathbf{y}_{t-1})$, ultimately yielding \mathbf{x}_0 .

FPS algorithm is theoretically supported to recover the oracle $p(\mathbf{x}|\mathbf{y})$ once the step size is sufficiently small.

3.3.3 PMC [137]

Plug-and-Play (PnP) [71] and RED [114] are two representative methods of using denoisers as priors for solving inverse problems. Let $g_{\mathbf{y}}(\mathbf{x}) = \frac{1}{2\sigma_{\mathbf{y}}^2} \|\mathbf{y} - A\mathbf{x}\|_2^2$ be the log-likelihood function, $\mathbf{h}_{\boldsymbol{\theta}}^{\sigma}(\cdot)$ an MMSE denoiser from Equation 2.11 conditioned on the noise level σ , and $R_{\boldsymbol{\theta}}^{\sigma}(\cdot) := \text{Id} - \mathbf{h}_{\boldsymbol{\theta}}^{\sigma}(\cdot)$ the residual projector. Note that conditioning on the noise level σ is equivalent to the network being conditioned no t , since the mapping is one-to-one. A single iteration of these methods read

- PnP proximal gradient method [72]:

$$\mathbf{x}_{k+1} = \mathbf{h}_{\boldsymbol{\theta}}^{\sigma}(\mathbf{x}_k - \gamma \nabla_{\mathbf{x}_k} g_{\mathbf{y}}(\mathbf{x}_k)) \quad (3.63)$$

$$= \mathbf{x}_k - \gamma \left(\nabla_{\mathbf{x}_k} g_{\mathbf{y}}(\mathbf{x}_k) + \frac{1}{\gamma} R_{\boldsymbol{\theta}}^{\sigma}(\mathbf{x}_k - \gamma \nabla_{\mathbf{x}_k} g_{\mathbf{y}}(\mathbf{x}_k)) \right). \quad (3.64)$$

- RED gradient descent [114]:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma (\nabla_{\mathbf{x}_k} g_{\mathbf{y}}(\mathbf{x}_k) + \tau (\mathbf{x}_k - \mathbf{h}_{\boldsymbol{\theta}}^{\sigma}(\mathbf{x}_k))) \quad (3.65)$$

$$= \mathbf{x}_k - \gamma (\nabla_{\mathbf{x}_k} g_{\mathbf{y}}(\mathbf{x}_k) + \tau R_{\boldsymbol{\theta}}^{\sigma}(\mathbf{x}_k)). \quad (3.66)$$

Notice that by using Tweedie's formula, we see that $R_{\boldsymbol{\theta}}^{\sigma}(\mathbf{x}) = -\sigma^2 \nabla_{\mathbf{x}} \log p_{\sigma}(\mathbf{x})$. Rearranging Equation 3.64 and Equation 3.66,

- PnP:

$$\frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\gamma} = -P(\mathbf{x}_k), \quad P(\mathbf{x}) := \nabla_{\mathbf{x}} g_{\mathbf{y}}(\mathbf{x}) - \frac{\sigma^2}{\gamma} \nabla_{\mathbf{x}} \log p_{\sigma}(\mathbf{x} - \gamma \nabla_{\mathbf{x}} g_{\mathbf{y}}(\mathbf{x})), \quad (3.67)$$

- RED:

$$\frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\gamma} = -G(\mathbf{x}_k), \quad G(\mathbf{x}) := \nabla_{\mathbf{x}} g_{\mathbf{y}}(\mathbf{x}) - \tau \sigma^2 \nabla_{\mathbf{x}} \log p_{\sigma}(\mathbf{x}). \quad (3.68)$$

Moreover, by setting $\gamma = \sigma^2$ and $\tau = 1/\sigma^2$, one can show that

$$\begin{aligned} \lim_{\sigma \rightarrow 0} P(\mathbf{x}) &= \nabla_{\mathbf{x}} g_{\mathbf{y}}(\mathbf{x}) - \lim_{\sigma \rightarrow 0} \{ \nabla_{\mathbf{x}} \log p(\mathbf{x} - \sigma^2 \nabla_{\mathbf{x}} g_{\mathbf{y}}(\mathbf{x}_k)) \} \\ &= \nabla_{\mathbf{x}} g_{\mathbf{y}}(\mathbf{x}) - \lim_{\sigma \rightarrow 0} \{ \lim_{\sigma \rightarrow 0} \log p_{\sigma}(\mathbf{x}) \} = \lim_{\sigma \rightarrow 0} G(\mathbf{x}) \\ &= -\nabla_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) - \nabla_{\mathbf{x}} \log p(\mathbf{x}) = -\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y}). \end{aligned} \quad (3.69)$$

In other words, we see that the iteration of PnP/RED in Equation 3.64 and Equation 3.66 will converge to sampling from the posterior as $\sigma^2 = \gamma \rightarrow 0$

$$d\mathbf{x}_t = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y}) dt, \quad (3.70)$$

where t indexes the continuous time flow of \mathbf{x} , as opposed to the discrete formulations in Equation 3.64 and Equation 3.66. Note that this notion of t does not match the diffusion time t , where the time index matches a specific noise level. In PMC, the authors propose to incorporate noise level annealing as done in the usual reverse diffusion process by starting from a large noise level σ and gradually reducing the noise level. Solving Equation 3.70 with PMC then boils down to iterative application of Equation 3.64 and Equation 3.66 with the annealing strategy. Moreover, introducing Langevin diffusion yields a stochastic version

$$d\mathbf{x}_t = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y}) dt + \sqrt{2} d\mathbf{W}_t, \quad (3.71)$$

which can be solved in the same way, but with additional stochasticity.

3.3.4 Sequential Monte Carlo-based methods

SMCDiff [141], MCGDiff [17], and TDS [150] belong to the category of sequential Monte Carlo (SMC)-based methods [48]. SMC aims to sample from the posterior by constructing a sequence of distributions $\mathbf{X}_{1:T}$, which terminates at the target distribution. The evolution of the distribution is approximated by K particles. In a high level, SMC can be described with three steps: 1) Transition with a proposal kernel $\{\mathbf{x}_t^{1:K}\} \sim p(\mathbf{X}_t|\mathbf{X}_{t-1})$, 2) computing the weights to re-weight the importance, and 3) resampling from a reweighted multinomial distribution. Methods that belong to this category propose different ways of constructing the proposal distribution and the weighting function.

3.4 CSGM-Type methods

3.4.1 DMPlug [146], SHRED [21]

Compressed sensing generative model (CSGM) [15, 34] is a general method for solving inverse problems with deep generative models by aiming to find the input latent vector z through

$$z^* = \arg \min_z \|y - AG_\theta(z)\|^2, \quad (3.72)$$

where G_θ is an arbitrary generative model. DMPlug and SHRED can be seen as extensions of CSGM to the case where one uses a diffusion model. Unlike GANs or Flows where the mapping from the latent space to the image space is done through a single NFE, diffusion models require multiple NFE to solve the generative SDE/ODE. One can rewrite Equation 3.72 as

$$z^* = \arg \min_z \|y - A\hat{x}(z)\|^2, \quad (3.73)$$

where $\hat{x} = \hat{x}(z)$ is the solution of the deterministic sampler initialized at z . Essentially, the models in this category optimize over the “latent” space of noises that are fed to the deterministic ODE sampler. One caveat of Equation 3.73 is the exploding memory required for backpropagation through time. To mitigate this, when sampling from $p_\theta(x_0|x_T)$, a few-step sampling (e.g. 3 for DMPlug and 10 for SHRED) is used to approximate the true sampling process.

3.4.2 CSGM with consistent diffusion models [153]

Diffusion models can be distilled into one-step models, known as Consistency Models [130], that solve in one step the Probability Flow ODE. These models can be used in Equation 3.73, replacing the ODE sampling, to reduce the computational requirements [153].

3.4.3 Intermediate Layer Optimization [34, 32]

CSGM has been extended to perform the optimization in some intermediate latent space [34]. The problem is that the intermediate latents need to be regularized to avoid exiting the manifold of realistic images. Score-Guided Intermediate Layer Optimization (Score-ILO) [32] uses diffusion models to regularize the intermediate solutions.

3.5 Latent Diffusion Models

3.5.1 Motivation

In this subsection, we focus on algorithms that have been developed for solving inverse problems with latent diffusion models (see Section 2.3). There are a few additional challenges when dealing with latent diffusion models that have led to a growing literature of papers that are trying to address them.

Loss of linearity. The first challenge in solving inverse problems with latent diffusion models is that linear inverse problems become essentially non-linear. The problem stems from the fact that diffusion happens in the latent space but measurements are in the pixel-space. In order to guide the diffusion there are two potential solutions: i) either project the measurements to the latent space through the encoder, or, ii) project the latents to the pixel space as we diffuse through the decoder. Both approaches depend on non-linear functions (Enc, Dec respectively) and hence even linear inverse problems need a more general treatment.

Decoding is expensive. The other issue that arises is computational. Most of the time, we need to decode the latent to pixel-space to compare with the measurements. The motivation behind latent diffusion models is to accelerate training and sampling. Hence, we want to avoid repeated calls to the decoder as we solve inverse problems.

Decoding-encoding map is not one-to-one. Even if we ignore the computational challenges, it is not straightforward to decode the latent to the pixel-space, compare with the measurements and get meaningful guidance in the latent space since the decoding-encoding map is not an one-to-one function.

Text-conditioning. Finally, latent diffusion models typically get a textual prompt as an additional input. A lot of algorithms that have been developed in the space of using latent diffusion models to solve inverse problems innovate on how they use text conditioning.

3.5.2 Latent DPS

The first algorithm we review in the space of solving inverse problems with latent diffusion models is Latent DPS, i.e. the straightforward extension of DPS for latent diffusion models. The approximation made in this algorithm is:

$$\nabla_{\mathbf{x}_t^E} \log p(\mathbf{y} | \mathbf{X}_t^E = \mathbf{x}_t^E) \approx \nabla_{\mathbf{x}_t^E} \log p(\mathbf{y} | \mathbf{X}_0 = \text{Dec}(\mathbb{E}[\mathbf{X}_0^E | \mathbf{X}_t^E = \mathbf{x}_t^E])). \quad (3.74)$$

The algorithm works by performing one-step denoising in the latent space and measuring how much the decoding of the denoised latent matches the measurements \mathbf{y} .

3.5.3 PSLD [118]

The performance of Latent DPS is hindered by the fact that the decoding-encoding map is not an one-to-one function, as discussed earlier. The approximation made above could pull \mathbf{x}_t^E towards any latent \mathbf{x}_0^E that has a decoding that matches the measurements while the score function is pulling \mathbf{x}_t^E towards a specific \mathbf{x}_0^E , i.e. towards $\mathbb{E}[\mathbf{x}_0^E | \mathbf{x}_t^E]$.

PSLD mitigates this problem by adding an additional term that pulls towards latents that are fixed points of the decoder-encoder map. Concretely, the approximation made in PSLD is:

$$\begin{aligned} \nabla_{\mathbf{x}_t^E} \log p(\mathbf{y} | \mathbf{X}_t^E = \mathbf{x}_t^E) \approx \\ \nabla_{\mathbf{x}_t^E} \log p(\mathbf{y} | (\mathbf{X}_0 = \text{Dec}(\mathbb{E}[\mathbf{X}_0^E | \mathbf{X}_t^E = \mathbf{x}_t^E]))) \\ + \gamma_t \nabla_{\mathbf{x}_t^E} \left\| \mathbb{E}[\mathbf{x}_0^E | \mathbf{x}_t^E] - \text{Enc}(\text{Dec}(\mathbb{E}[\mathbf{x}_0^E | \mathbf{x}_t^E])) \right\|^2, \end{aligned} \quad (3.75)$$

where γ_t is a tunable parameter.

3.5.4 Resample [127]

Resample, a concurrent work with PSLD, proposes an alternative way to improve the performance of Latent DPS. After each clean prediction $\hat{\mathbf{x}}_0(\mathbf{x}_{t+1}^E)$ is obtained from the previous sample \mathbf{x}_{t+1}^E via Tweedie's formula in Equation 2.10, and the unconditional reverse denoising process is updated using, say, DDIM:

$$\mathbf{x}'_t := \text{UnconditionalDDIM}(\hat{\mathbf{x}}_0(\mathbf{x}_{t+1}^E), \mathbf{x}_{t+1}^E) \quad (3.76)$$

the authors project the latent back to a point $\hat{\mathbf{x}}_t$ that satisfies measurements using:

$$\mathcal{N}\left(\hat{\mathbf{x}}_t; \frac{\sigma_t^2 \sqrt{\bar{\alpha}_t} \hat{\mathbf{x}}_0(\mathbf{y}) + (1 - \bar{\alpha}_t) \mathbf{x}'_t}{\sigma_t^2 + (1 - \bar{\alpha}_t)}, \frac{\sigma_t^2 (1 - \bar{\alpha}_t)}{\sigma_t^2 + (1 - \bar{\alpha}_t)} I_k\right). \quad (3.77)$$

Here, σ_t^2 is a hyperparameter used to tune the alignment with measurements, $\bar{\alpha}_t$ is predefined in forward process, and $\hat{\mathbf{x}}_0(\mathbf{y})$ is found by solving:

$$\hat{\mathbf{x}}_0(\mathbf{y}) \in \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathcal{A}(\text{Dec}(\mathbf{x}))\|_2^2 \quad \text{initialized at } \hat{\mathbf{x}}_0(\mathbf{x}_{t+1}^E). \quad (3.78)$$

3.6 MPGD [60]

The MPGD authors note that some methods require expensive computations for measurement alignment during gradient updates, as they involve passing through the gradient (chain rule) of the pre-trained diffusion model $\epsilon_{\theta}(\mathbf{x}_t^E, t)$:

$$\mathbf{x}_t^E \leftarrow \mathbf{x}_t^E - \eta_t \nabla_{\mathbf{x}_t^E} \left\| \mathbf{y} - \mathcal{A}(\text{Dec}(\mathbf{x}_{0|t})) \right\|_2^2, \quad (3.79)$$

where $\mathbf{x}_{0|t} := \frac{1}{\sqrt{\bar{\alpha}_t}} (\mathbf{x}_t^E - \sqrt{1 - \bar{\alpha}_t} \epsilon_{\theta}(\mathbf{x}_t^E, t))$ is a clean estimation via Tweedie's formula in Equation 2.10. This gradient bottleneck slows down the overall inverse problem solving. MPGD proposes bypassing the direct gradient $\nabla_{\mathbf{x}_t^E}$ with theoretical guarantees by updating with $\nabla_{\mathbf{x}_{0|t}}$:

$$\mathbf{x}'_{0|t} \leftarrow \mathbf{x}_{0|t} - \eta_t \nabla_{\mathbf{x}_{0|t}} \left\| \mathbf{y} - \mathcal{A}(\text{Dec}(\mathbf{x}_{0|t})) \right\|_2^2 \quad (3.80)$$

with

$$\mathbf{x}_{0|t} := \frac{1}{\sqrt{\bar{\alpha}_t}} (\mathbf{x}_t^E - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\mathbf{x}_t^E, t)), \quad (3.81)$$

and use the obtained $\mathbf{x}'_{0|t}$ for unconditional reverse denoising process

$$\mathbf{x}'_{t-1} := \text{UnconditionalDDIM}(\mathbf{x}'_{0|t}, \mathbf{x}_t^E). \quad (3.82)$$

3.6.1 P2L [30]

While text conditioning is a viable option for modern latent diffusion models such as Stable diffusion, the actual use was underexplored due to ambiguities on which text to use. P2L addresses this question by proposing an algorithm that optimizes for the text embedding on the fly while solving an inverse problem.

$$\mathbf{c}_t^* = \arg \min_{\mathbf{c}} \|\mathbf{y} - A \text{Dec}(\mathbb{E}[\mathbf{x}_0^E | \mathbf{x}_t^E, \mathbf{c}])\|^2, \quad (3.83)$$

where \mathbf{c} is the text embedding, and one can approximate $\mathbb{E}[\mathbf{x}_0^E | \mathbf{x}_t^E, \mathbf{c}]$ by using the Tweedie's formula with the denoiser conditioned on \mathbf{c} . Using the optimized embedding at each timestep \mathbf{c}_t^* , sampling follows the procedure of Latent DPS

$$\nabla_{\mathbf{x}_t^E} \log p(\mathbf{y} | \mathbf{x}_t^E = \mathbf{x}_t^E, \mathbf{c}) \approx \nabla_{\mathbf{x}_t^E} \log p(\mathbf{y} | \mathbf{X}_0 = \text{Dec}(\mathbb{E}[\mathbf{x}_0^E | \mathbf{x}_t^E = \mathbf{x}_t^E, \mathbf{c}_t^*])) \quad (3.84)$$

In addition to the optimization of the text embedding, P2L further tries to leverage the VAE prior by decoding - running optimization in the pixel space - re-encoding

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{y} - A\mathbf{x}\|_2^2 + \lambda \|\mathbf{x} - \text{Dec}(\mathbb{E}[\mathbf{x}_0^E | \mathbf{x}_t^E = \mathbf{x}_t^E])\|_2^2 \quad (3.85)$$

$$\mathbf{x}^E = \text{Enc}(\mathbf{x}^*) \quad (3.86)$$

3.6.2 TReg [77], DreamSampler [78]

Instead of automatically finding a suitable text embedding to achieve maximal reconstructive performance, another advantage of text conditioning is that it can be used as an additional guiding signal to lead to a specific mode. This may seem trivial, as one has access to a conditional diffusion model. However, in practice, simply using a conditional diffusion model does not induce enough guidance as reported in [41, 64], and naively using classifier free guidance [64] (CFG) does not lead to satisfactory results. In addition to using data consistency imposing steps as in P2L, TReg proposes adaptive negation to update the null text embeddings used for CFG guidance.

$$\mathbf{c}_\emptyset^* = \arg \min_{\mathbf{c}} \text{sim}(\mathcal{T}(\mathbf{x}^*), \mathbf{c}), \quad (3.87)$$

where \mathbf{x}^* comes from Equation 3.85, sim denotes the CLIP similarity [110] score, and \mathcal{T} is the CLIP image encoder. In essence, Equation 3.87 minimizes the similarity between the current estimate of the image and the null text embedding. Hence, when the optimized \mathbf{c}_\emptyset^* is used for CFG with

$$\epsilon_\theta(\mathbf{x}_t^E, \mathbf{c}_\emptyset^*) + \omega (\epsilon_\theta(\mathbf{x}_t^E, \mathbf{c}) - \epsilon_\theta(\mathbf{x}_t^E, \mathbf{c}_\emptyset^*),) \quad (3.88)$$

the conditioning vector direction $\epsilon_\theta(\mathbf{x}_t^E, \mathbf{c}) - \epsilon_\theta(\mathbf{x}_t^E, \mathbf{c}_\emptyset^*)$ is amplified. Later, TReg was further advanced by devising a way to better make use of CFG by combining score distillation sampling [108] into the sampling framework.

3.6.3 STSL [116]

Most methods leverage the mean of the reverse diffusion distribution $p(\mathbf{X}_0 | \mathbf{X}_t)$, and take a single gradient step with Equation 3.74. To further leverage the covariance of $p(\mathbf{X}_0 | \mathbf{X}_t)$, Rout *et al.* [116] propose to use the following fidelity loss

$$\begin{aligned} \mathcal{L}(\mathbf{x}_t^E, \mathbf{y}) &= \nabla_{\mathbf{x}_t^E} \log p(\mathbf{y} | \mathbf{X}_0 = \text{Dec}(\mathbb{E}[\mathbf{x}_0^E | \mathbf{x}_t^E = \mathbf{x}_t^E])) \\ &\quad + \gamma \nabla_{\mathbf{x}_t^E} \left(\text{Trace} \left(\nabla_{\mathbf{x}_t^E}^2 \log p(\mathbf{x}_t^E) \right) \right), \end{aligned} \quad (3.89)$$

where γ is a constant. To effectively compute the trace, one can further use the following approximation

$$\text{Trace} \left(\nabla_{\mathbf{x}_t^E}^2 \log p(\mathbf{x}_t^E) \right) \approx \mathbb{E}_{\epsilon \sim \pi} \left[\epsilon^\top \left(\nabla_{\mathbf{x}_t^E} \log p(\mathbf{x}_t^E + \epsilon) - \nabla_{\mathbf{x}_t^E} \log p(\mathbf{x}_t^E) \right) \right], \quad (3.90)$$

where π can be a Gaussian or a Rademacher distribution. Using the loss in Equation 3.89 with Equation 3.90, STSL uses multiple steps of stochastic gradient updates per timestep.

4 Thoughts from the authors

In the previous section, we presented several works in the space of using diffusion models to solve inverse problems. A natural question that both experts and newcomers to the field might have is, eventually, “*which approach works the best?*”. Unfortunately, we cannot provide a conclusive answer to this question within the scope of this survey, but we can share a few thoughts.

Thoughts about Explicit Approximations. In this survey we tried to express seemingly very different works, such as DPS and DDRM, under a common mathematical language that contains the explicit approximations made for the measurements score. We observed that all the methods compute an error metric that matches consistency with the measurement and then lift the error back to the image space dimensions to perform the gradient update. Some of the methods used noised versions of the measurements to compute the error while others use the clean measurements. To the best of our knowledge, it is not clear which one works the best and one can derive new approximation algorithms by simply making the dual change to any of the methods that already exist, e.g. one can propose Score-ALD++ by using the noisy measurements to compute the error. By looking at Figure 1, it is also evident that methods propose increasingly more complex “lifting” matrices. Some of these approximations require increased computation, e.g. the Moments Matching method. We strongly believe that the field would benefit from a standardized benchmark for diffusion models and inverse problems to understand better the computational performance trade-offs of different methods. We also believe that under certain distributional assumptions, it should be possible to characterize analytically the propagation of the approximation errors induced by the different methods.

Thoughts about Variational Methods. Variational Methods try to estimate the parameters of a simpler distribution. The benefit here is that one can employ well-known optimization techniques to better solve the optimization problem at hand. A potential drawback of this approach is that the proposed distribution might not be able to capture the complexity of the real posterior distribution.

Thoughts about CSGM-type Methods. CSGM-type frameworks can benefit from the plethora of techniques that have been previously developed to solve inverse problems with GANs and other deep generative modeling frameworks. The main issue here is computational since the generative model to be inverted here is the Probability Flow ODE mapping that requires several calls to the diffusion model. Consistency Models [130, 76] and other approaches such as Intermediate Layer Optimization could mitigate this issue.

Thoughts about Asymptotically Exact Methods. Asymptotically Exact Methods, usually based on Monte Carlo, could be useful when sampling from the true posterior is really important. However, the theoretical guarantees of these methods only hold under the setting of infinite computation and it remains to be seen if they can scale to more practical settings.

5 Conclusion

In this survey, we discussed different types of inverse problems and different approaches that have been developed to solve them using diffusion priors. We identified four distinct families: methods that propose explicit approximations for the measurement score, variational inference methods, CSGM-type frameworks and finally approaches that asymptotically guarantee exact sampling (at the cost of increased computation). The different frameworks and the works therein are all trying to address the fundamental problem of the intractability of the posterior distribution. In this survey, we tried to unify seemingly different approaches and explain the trade-offs of different methods. We hope that this survey will serve as a reference point for the vibrant field of diffusion models for inverse problems.

Acknowledgments

This research has been supported by NSF Grants AF 1901292, CNS 2148141, Tripods CCF 1934932, IFML CCF 2019844 and research gifts by Western Digital, Amazon, WNCG IAP, UT Austin Machine Learning Lab (MLL), Cisco and the Stanley P. Finch Centennial Professorship in Engineering. Giannis

Daras has been supported by the Onassis Fellowship (Scholarship ID: F ZS 012-1/2022-2023), the Bodossaki Fellowship and the Leventis Fellowship. The authors would like to thank our colleagues Viraj Shah, Miki Rubinstein, Murata Naoki, Yutong He, and Stefano Ermon for helpful discussions.

References

- [1] Asad Aali, Marius Arvinte, Sidharth Kumar, and Jonathan I Tamir. “Solving Inverse Problems with Score-Based Generative Priors learned from Noisy Data”. In: *arXiv preprint arXiv:2305.01166* (2023) (pages 2, 9).
- [2] Asad Aali, Giannis Daras, Brett Levac, Sidharth Kumar, Alexandros G Dimakis, and Jonathan I Tamir. “Ambient Diffusion Posterior Sampling: Solving Inverse Problems with Diffusion Models trained on Corrupted Data”. In: *arXiv preprint arXiv:2403.08728* (2024) (pages 1, 9).
- [3] Kazunori Akiyama, Antxon Alberdi, Walter Alef, Keiichi Asada, Rebecca Azulay, Anne-Kathrin Baczko, David Ball, Mislav Baloković, John Barrett, Dan Bintley, et al. “First M87 event horizon telescope results. IV. Imaging the central supermassive black hole”. In: *The Astrophysical Journal Letters* 875.1 (2019), p. L4 (pages 4, 9).
- [4] Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. “Stochastic interpolants: A unifying framework for flows and diffusions”. In: *arXiv preprint arXiv:2303.08797* (2023) (page 4).
- [5] Michael Samuel Albergo and Eric Vanden-Eijnden. “Building Normalizing Flows with Stochastic Interpolants”. In: *The Eleventh International Conference on Learning Representations*. 2023. URL: <https://openreview.net/forum?id=li7qeBbCR1t> (page 4).
- [6] Cagan Alkan, Julio Oscanoa, Daniel Abraham, Mengze Gao, Aizada Nurdinova, Kawin Setsompop, John M Pauly, Morteza Mardani, and Shreyas Vasanaawala. “Variational Diffusion Models for Blind MRI Inverse Problems”. In: *NeurIPS 2023 Workshop on Deep Learning and Inverse Problems*. 2023 (pages 2, 17).
- [7] Brian D.O. Anderson. “Reverse-time diffusion equation models”. In: *Stochastic Processes and their Applications* 12.3 (1982), pp. 313–326 (page 6).
- [8] Weimin Bai, Yifei Wang, Wenzheng Chen, and He Sun. “An Expectation-Maximization Algorithm for Training Clean Diffusion Models from Corrupted Observations”. In: *arXiv preprint arXiv:2407.01014* (2024) (page 9).
- [9] Harrison H Barrett and Kyle J Myers. *Foundations of image science*. John Wiley & Sons, 2013 (page 4).
- [10] Robert Basset and Julio Deride. “Maximum a posteriori estimators as a limit of Bayes estimators”. In: *Mathematical Programming* 174 (2019), pp. 129–144 (page 4).
- [11] Joshua Batson and Loic Royer. “Noise2self: Blind denoising by self-supervision”. In: *International Conference on Machine Learning*. PMLR. 2019, pp. 524–533 (page 9).
- [12] Amir Beck and Marc Teboulle. “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems”. In: *IEEE transactions on image processing* 18.11 (2009), pp. 2419–2434 (page 4).
- [13] Gregory Beylkin. “The inversion problem and applications of the generalized Radon transform”. In: *Communications on pure and applied mathematics* 37.5 (1984), pp. 579–599 (page 2).
- [14] Yochai Blau and Tomer Michaeli. “The Perception-Distortion Tradeoff”. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE. 2018 (page 4).
- [15] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. “Compressed sensing using generative models”. In: *International conference on machine learning*. PMLR. 2017, pp. 537–546 (page 21).
- [16] Emmanuel J Candès, Justin Romberg, and Terence Tao. “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information”. In: *IEEE Transactions on information theory* 52.2 (2006), pp. 489–509 (page 4).
- [17] Gabriel Cardoso, Sylvain Le Corff, Eric Moulines, et al. “Monte Carlo guided Denoising Diffusion models for Bayesian linear inverse problems.” In: *The Twelfth International Conference on Learning Representations*. 2023 (pages 2, 5, 20).
- [18] Stanley H Chan, Xiran Wang, and Omar A Elgendy. “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications”. In: *IEEE Transactions on Computational Imaging* 3.1 (2016), pp. 84–98 (page 4).

- [19] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. “Learning to see in the dark”. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 3291–3300 (page 4).
- [20] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. “Simple baselines for image restoration”. In: *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*. Springer. 2022, pp. 17–33 (page 4).
- [21] Hamadi Chihaoui, Abdelhak Lemkhenter, and Paolo Favaro. *Zero-shot Image Restoration via Diffusion Inversion*. 2024. URL: <https://openreview.net/forum?id=ZnmofqLWMQ> (pages 2, 5, 21).
- [22] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. “ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021, pp. 14367–14376 (pages 2, 5, 12).
- [23] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. “Parallel diffusion models of operator and image for blind inverse problems”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 6059–6069 (pages 2, 13).
- [24] Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. “Diffusion Posterior Sampling for General Noisy Inverse Problems”. In: *The Eleventh International Conference on Learning Representations*. 2023. URL: <https://openreview.net/forum?id=OnD9zGAGT0k> (pages 2, 5).
- [25] Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. “Decomposed Diffusion Sampler for Accelerating Large-Scale Inverse Problems”. In: *arXiv preprint arXiv:2303.05754* (2023) (pages 2, 5, 16).
- [26] Hyungjin Chung, Dohoon Ryu, Michael T McCann, Marc L Klasky, and Jong Chul Ye. “Solving 3d inverse problems using pre-trained 2d diffusion models”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 22542–22551 (page 1).
- [27] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. “Improving diffusion models for inverse problems using manifold constraints”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 25683–25696 (pages 2, 12).
- [28] Hyungjin Chung and Jong Chul Ye. “Deep Diffusion Image Prior for Efficient OOD Adaptation in 3D Inverse Problems”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2024 (page 5).
- [29] Hyungjin Chung and Jong Chul Ye. “Score-based diffusion models for accelerated MRI”. In: *Medical image analysis* 80 (2022), p. 102479 (page 1).
- [30] Hyungjin Chung, Jong Chul Ye, Peyman Milanfar, and Mauricio Delbracio. “Prompt-tuning latent diffusion models for inverse problems”. In: *International Conference on Machine Learning*. PMLR. 2014 (pages 2, 23).
- [31] Regev Cohen, Michael Elad, and Peyman Milanfar. “Regularization by denoising via fixed-point projection (RED-PRO)”. In: *SIAM Journal on Imaging Sciences* 14.3 (2021), pp. 1374–1406 (page 4).
- [32] Giannis Daras, Yuval Dagan, Alex Dimakis, and Constantinos Daskalakis. “Score-Guided Intermediate Level Optimization: Fast Langevin Mixing for Inverse Problems”. In: *Proceedings of the 39th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato. Vol. 162. Proceedings of Machine Learning Research. PMLR, 17–23 Jul 2022, pp. 4722–4753. URL: <https://proceedings.mlr.press/v162/daras22a.html> (pages 2, 21).
- [33] Giannis Daras, Yuval Dagan, Alexandros G Dimakis, and Constantinos Daskalakis. “Consistent diffusion models: Mitigating sampling drift by learning to be consistent”. In: *arXiv preprint arXiv:2302.09057* (2023) (page 9).
- [34] Giannis Daras, Joseph Dean, Ajil Jalal, and Alex Dimakis. “Intermediate Layer Optimization for Inverse Problems using Deep Generative Models”. In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, 18–24 Jul 2021, pp. 2421–2432. URL: <https://proceedings.mlr.press/v139/daras21a.html> (page 21).

- [35] Giannis Daras, Mauricio Delbracio, Hossein Talebi, Alex Dimakis, and Peyman Milanfar. “Soft Diffusion: Score Matching with General Corruptions”. In: *Transactions on Machine Learning Research* (2023). ISSN: 2835-8856. URL: <https://openreview.net/forum?id=W98rebBx1Q> (page 36).
- [36] Giannis Daras, Alex Dimakis, and Constantinos Costis Daskalakis. “Consistent Diffusion Meets Tweedie: Training Exact Ambient Diffusion Models with Noisy Data”. In: *Proceedings of the 41st International Conference on Machine Learning*. Ed. by Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp. Vol. 235. Proceedings of Machine Learning Research. PMLR, 2024, pp. 10091–10108. URL: <https://proceedings.mlr.press/v235/daras24a.html> (page 9).
- [37] Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alex Dimakis, and Adam Klivans. “Ambient Diffusion: Learning Clean Distributions from Corrupted Data”. In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine. Vol. 36. Curran Associates, Inc., 2023, pp. 288–313. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/012af729c5d14d279581fc8a5db975a1-Paper-Conference.pdf (page 9).
- [38] Ingrid Daubechies, Michel Defrise, and Christine De Mol. “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint”. In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 57.11 (2004), pp. 1413–1457 (page 4).
- [39] Mauricio Delbracio and Peyman Milanfar. “Inversion by Direct Iteration: An Alternative to Denoising Diffusion for Image Restoration”. In: *Transactions on Machine Learning Research* (2023). Featured Certification. ISSN: 2835-8856. URL: <https://openreview.net/forum?id=VmyFF51L3F> (page 4).
- [40] Arjun D Desai, Andrew M Schmidt, Elka B Rubin, Christopher M Sandino, Marianne S Black, Valentina Mazzoli, Kathryn J Stevens, Robert Boutin, Christopher Ré, Garry E Gold, Brian A Hargreaves, and Akshay S Chaudhari. “SKM-TEA: A Dataset for Accelerated MRI Reconstruction with Dense Image Labels for Quantitative Clinical Evaluation”. In: (2022). arXiv: 2203.06823 [eess.IV] (page 2).
- [41] Prafulla Dhariwal and Alexander Nichol. “Diffusion models beat gans on image synthesis”. In: *Advances in neural information processing systems* 34 (2021), pp. 8780–8794 (page 23).
- [42] Daniel J Diaz, Chengyue Gong, Jeffrey Ouyang-Zhang, James M Loy, Jordan Wells, David Yang, Andrew D Ellington, Alexandros G Dimakis, and Adam R Klivans. “Stability Oracle: a structure-based graph-transformer framework for identifying stabilizing mutations”. In: *Nature Communications* 15.1 (2024), p. 6170 (page 2).
- [43] Martin Dietz, Lars Liljeryd, Kristofer Kjorling, and Oliver Kunz. “Spectral Band Replication, a novel approach in audio coding”. In: *Audio Engineering Society Convention 112*. Audio Engineering Society. 2002 (page 2).
- [44] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. “Density estimation using real nvp”. In: *arXiv preprint arXiv:1605.08803* (2016) (page 18).
- [45] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. “Image super-resolution using deep convolutional networks”. In: *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2015), pp. 295–307 (page 4).
- [46] David L Donoho. “Compressed sensing”. In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306 (page 4).
- [47] Zehao Dou and Yang Song. “Diffusion posterior sampling for linear inverse problem solving: A filtering perspective”. In: *The Twelfth International Conference on Learning Representations*. 2023 (pages 2, 5, 19).
- [48] Arnaud Doucet, Nando De Freitas, and Neil Gordon. “An introduction to sequential Monte Carlo methods”. In: *Sequential Monte Carlo methods in practice* (2001), pp. 3–14 (page 20).
- [49] Jacques Dubochet, Marc Adrian, Jiin-Ju Chang, Jean-Claude Homo, Jean Lepault, Alasdair W McDowell, and Patrick Schultz. “Cryo-electron microscopy of vitrified specimens”. In: *Quarterly reviews of biophysics* 21.2 (1988), pp. 129–228 (page 2).
- [50] Bradley Efron. “Tweedie’s formula and selection bias”. In: *Journal of the American Statistical Association* 106.496 (2011), pp. 1602–1614 (page 7).

- [51] Yonina C. Eldar. “Generalized SURE for Exponential Families: Applications to Regularization”. In: *IEEE Transactions on Signal Processing* 57.2 (2009), pp. 471–481. DOI: 10.1109/TSP.2008.2008212 (page 9).
- [52] Linwei Fan, Fan Zhang, Hui Fan, and Caiming Zhang. “Brief review of image denoising techniques”. In: *Visual Computing for Industry, Biomedicine, and Art* 2.1 (2019), p. 7 (page 1).
- [53] Berthy T Feng and Katherine L Bouman. “Efficient bayesian computational imaging with a surrogate score-based prior”. In: *arXiv preprint arXiv:2309.01949* (2023) (pages 2, 5, 18).
- [54] Berthy T Feng, Jamie Smith, Michael Rubinstein, Huiwen Chang, Katherine L Bouman, and William T Freeman. “Score-Based diffusion models as principled priors for inverse imaging”. In: *arXiv preprint arXiv:2304.11751* (2023) (pages 2, 5, 17, 18).
- [55] James R Fienup. “Phase retrieval algorithms: a comparison”. In: *Applied optics* 21.15 (1982), pp. 2758–2769 (page 3).
- [56] Mário AT Figueiredo and Robert D Nowak. “An EM algorithm for wavelet-based image restoration”. In: *IEEE Transactions on Image Processing* 12.8 (2003), pp. 906–916 (page 4).
- [57] Martin Genzel, Ingo Gühling, Jan Macdonald, and Maximilian März. “Near-exact recovery for tomographic inverse problems via deep learning”. In: *International Conference on Machine Learning*. PMLR. 2022, pp. 7368–7381 (page 2).
- [58] Shivam Gupta, Ajil Jalal, Aditya Parulekar, Eric Price, and Zhiyang Xun. “Diffusion Posterior Sampling is Computationally Intractable”. In: *arXiv preprint arXiv:2402.12727* (2024) (page 8).
- [59] Elaine T Hale, Wotao Yin, and Yin Zhang. “A fixed-point continuation method for ℓ_1 -regularized minimization with applications to compressed sensing”. In: *CAAM TR07-07, Rice University* 43.44 (2007), p. 2 (page 4).
- [60] Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J Zico Kolter, Ruslan Salakhutdinov, et al. “Manifold Preserving Guided Diffusion”. In: *The Twelfth International Conference on Learning Representations*. 2023 (page 22).
- [61] Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J Zico Kolter, Ruslan Salakhutdinov, et al. “Manifold preserving guided diffusion”. In: *arXiv preprint arXiv:2311.16424* (2023) (pages 2, 12).
- [62] Carlos Hernandez-Olivan, Koichi Saito, Naoki Murata, Chieh-Hsin Lai, Marco A Martínez-Ramirez, Wei-Hsiang Liao, and Yuki Mitsufuji. “Vrdmg: Vocal restoration via diffusion posterior sampling with multiple guidance”. In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2024, pp. 596–600 (page 1).
- [63] Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising diffusion probabilistic models”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 6840–6851 (pages 5, 6).
- [64] Jonathan Ho and Tim Salimans. “Classifier-free diffusion guidance”. In: *arXiv preprint arXiv:2207.12598* (2022) (page 23).
- [65] Sixun Huang, Jie Xiang, Huadong Du, and Xiaoqun Cao. “Inverse problems in atmospheric science and their application”. In: *Journal of Physics: Conference Series*. Vol. 12. 1. IOP Publishing. 2005, p. 45 (page 1).
- [66] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. “Image-to-image translation with conditional adversarial networks”. In: *IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134 (page 4).
- [67] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G Dimakis, and Jon Tamir. “Robust compressed sensing mri with deep generative priors”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 14938–14954 (pages 2, 4, 5, 10, 11).
- [68] Zahra Kadhodaie and Eero P Simoncelli. “Solving linear inverse problems using the prior implicit in a denoiser”. In: *arXiv preprint arXiv:2007.13640* (2020) (pages 2, 5, 11).
- [69] Zahra Kadhodaie and Eero P Simoncelli. “Stochastic Solutions for Linear Inverse Problems using the Prior Implicit in a Denoiser”. In: *Thirty-Fifth Conference on Neural Information Processing Systems*. 2021 (page 4).
- [70] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001 (page 2).

- [71] Ulugbek S Kamilov, Charles A Bouman, Gregory T Buzzard, and Brendt Wohlberg. “Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications”. In: *IEEE Signal Processing Magazine* 40.1 (2023), pp. 85–97 (pages 4, 20).
- [72] Ulugbek S Kamilov, Hassan Mansour, and Brendt Wohlberg. “A plug-and-play priors approach for solving nonlinear imaging inverse problems”. In: *IEEE Signal Processing Letters* 24.12 (2017), pp. 1872–1876 (page 20).
- [73] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. “Denoising Diffusion Restoration Models”. In: *Advances in Neural Information Processing Systems*. 2022 (pages 2, 5, 14).
- [74] Bahjat Kawar, Gregory Vaksman, and Michael Elad. “SNIPS: Solving noisy inverse problems stochastically”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 21757–21769 (pages 2, 5, 9, 14).
- [75] Bahjat Kawar, Gregory Vaksman, and Michael Elad. “Stochastic image denoising by sampling from the posterior distribution”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 1866–1875 (page 14).
- [76] Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. “Consistency Trajectory Models: Learning Probability Flow ODE Trajectory of Diffusion”. In: *The Twelfth International Conference on Learning Representations*. 2023 (page 24).
- [77] Jeongsol Kim, Geon Yeong Park, Hyungjin Chung, and Jong Chul Ye. “Regularization by Texts for Latent Diffusion Inverse Solvers”. In: *arXiv preprint arXiv:2311.15658* (2023) (pages 2, 23).
- [78] Jeongsol Kim, Geon Yeong Park, and Jong Chul Ye. “DreamSampler: Unifying Diffusion Sampling and Score Distillation for Image Manipulation”. In: *arXiv preprint arXiv:2403.11415* (2024) (pages 2, 23).
- [79] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. “Noise2void-learning denoising from single noisy images”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 2129–2137 (page 9).
- [80] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. “Deblurgan: Blind motion deblurring using conditional adversarial networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 8183–8192 (page 4).
- [81] Patrick Lailly and J Bednar. “The seismic inverse problem as a sequence of before stack migrations”. In: *Conference on inverse scattering: theory and application*. Vol. 1983. Philadelphia, Pa. 1983, pp. 206–220 (page 1).
- [82] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. “Noise2Noise: Learning image restoration without clean data”. In: *arXiv preprint arXiv:1803.04189* (2018) (page 9).
- [83] Jean-Marie Lemerrier, Julius Richter, Simon Welker, Eloi Moliner, Vesa Välimäki, and Timo Gerkmann. “Diffusion Models for Audio Restoration”. In: *arXiv preprint arXiv:2402.09821* (2024) (page 1).
- [84] Haoying Li, Yifan Yang, Meng Chang, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. “SRDiff: Single Image Super-Resolution with Diffusion Probabilistic Models”. In: *arXiv preprint arXiv:2104.14951* (2021) (page 4).
- [85] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. “Enhanced deep residual networks for single image super-resolution”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017, pp. 136–144 (page 4).
- [86] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. “Flow Matching for Generative Modeling”. In: *The Eleventh International Conference on Learning Representations*. 2023. URL: <https://openreview.net/forum?id=PqvMRDCJT9t> (page 4).
- [87] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. “I²SB: Image-to-Image Schrödinger Bridge”. In: *arXiv preprint arXiv:2302.05872* (2023) (page 4).

- [88] Xingchao Liu, Chengyue Gong, and qiang liu. “Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow”. In: *The Eleventh International Conference on Learning Representations*. 2023. URL: <https://openreview.net/forum?id=XVjTT1nw5z> (page 4).
- [89] 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*. 2022, pp. 11461–11471 (pages 5, 12).
- [90] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. “Image restoration with mean-reverting stochastic differential equations”. In: *arXiv preprint arXiv:2301.11699* (2023) (page 4).
- [91] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. “Refusion: Enabling large-size realistic image restoration with latent-space diffusion models”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 1680–1691 (page 4).
- [92] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. “Compressed sensing MRI”. In: *IEEE signal processing magazine* 25.2 (2008), pp. 72–82 (page 2).
- [93] Dimitra Maoutsa, Sebastian Reich, and Manfred Opper. “Interacting particle solutions of fokker–planck equations through gradient–log–density estimation”. In: *Entropy* 22.8 (2020), p. 802 (page 6).
- [94] 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 (pages 2, 5, 16, 17).
- [95] Peyman Milanfar and Mauricio Delbracio. “Denoising: A Powerful Building-Block for Imaging, Inverse Problems, and Machine Learning”. In: *arXiv preprint arXiv:2409.06219* (2024) (page 4).
- [96] Eloi Moliner, Filip Elvander, and Vesa Välimäki. “Blind audio bandwidth extension: A diffusion-based zero-shot approach”. In: *arXiv preprint arXiv:2306.01433* (2023) (page 1).
- [97] Eloi Moliner, Jaakko Lehtinen, and Vesa Välimäki. “Solving audio inverse problems with a diffusion model”. In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, pp. 1–5 (page 1).
- [98] Eloi Moliner and Vesa Välimäki. “Diffusion-based audio inpainting”. In: *arXiv preprint arXiv:2305.15266* (2023) (page 1).
- [99] Naoki Murata, Koichi Saito, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Yuki Mitsufuji, and Stefano Ermon. “Gibbsddrm: A partially collapsed gibbs sampler for solving blind inverse problems with denoising diffusion restoration”. In: *International Conference on Machine Learning*. PMLR. 2023, pp. 25501–25522 (pages 2, 15).
- [100] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012 (page 4).
- [101] Tomohiro Nakatani, Takuya Yoshioka, Keisuke Kinoshita, Masato Miyoshi, and Biing-Hwang Juang. “Speech dereverberation based on variance-normalized delayed linear prediction”. In: *IEEE Transactions on Audio, Speech, and Language Processing* 18.7 (2010), pp. 1717–1731 (page 3).
- [102] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. 6th ed. Berlin: Springer Science & Business Media, 2010. ISBN: 9783642143946 (pages 7, 8).
- [103] Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. “Deep learning techniques for inverse problems in imaging”. In: *IEEE Journal on Selected Areas in Information Theory* 1.1 (2020), pp. 39–56 (page 4).
- [104] Jeffrey Ouyang-Zhang, Daniel J Diaz, Adam Klivans, and Philipp Krähenbühl. “Predicting a Protein’s Stability under a Million Mutations”. In: *NeurIPS* (2023) (page 2).
- [105] Xiaochuan Pan, Emil Y Sidky, and Michael Vannier. “Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?” In: *Inverse problems* 25.12 (2009), p. 123009 (page 2).
- [106] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. “Super-resolution image reconstruction: a technical overview”. In: *IEEE signal processing magazine* 20.3 (2003), pp. 21–36 (page 3).
- [107] Marcelo Pereyra. “Revisiting maximum-a-posteriori estimation in log-concave models”. In: *SIAM Journal on Imaging Sciences* 12.1 (2019), pp. 650–670 (page 4).

- [108] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. “DreamFusion: Text-to-3D using 2D Diffusion”. In: *The Eleventh International Conference on Learning Representations*. 2023. URL: <https://openreview.net/forum?id=FjNys5c7VyY> (page 23).
- [109] Weize Quan, Jiaxi Chen, Yanli Liu, Dong-Ming Yan, and Peter Wonka. “Deep learning-based image and video inpainting: A survey”. In: *International Journal of Computer Vision* 132.7 (2024), pp. 2367–2400 (page 2).
- [110] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. “Learning transferable visual models from natural language supervision”. In: *International conference on machine learning*. PMLR. 2021, pp. 8748–8763 (page 23).
- [111] Sriram Ravula, Brett Levac, Ajil Jalal, Jonathan I Tamir, and Alexandros G Dimakis. “Optimizing sampling patterns for compressed sensing MRI with diffusion generative models”. In: *arXiv preprint arXiv:2306.03284* (2023) (page 15).
- [112] Danilo Rezende and Shakir Mohamed. “Variational inference with normalizing flows”. In: *International conference on machine learning*. PMLR. 2015, pp. 1530–1538 (page 18).
- [113] Alejandro Ribes and Francis Schmitt. “Linear inverse problems in imaging”. In: *IEEE Signal Processing Magazine* 25.4 (2008), pp. 84–99 (page 4).
- [114] Yaniv Romano, Michael Elad, and Peyman Milanfar. “The Little Engine That Could: Regularization by Denoising (RED)”. In: *SIAM Journal on Imaging Sciences* 10.4 (2017), pp. 1804–1844. DOI: 10.1137/16M1102884 (pages 4, 20).
- [115] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. “High-resolution image synthesis with latent diffusion models”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 10684–10695 (pages 7, 10).
- [116] Litu Rout, Yujia Chen, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. “Beyond first-order tweedie: Solving inverse problems using latent diffusion”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, pp. 9472–9481 (pages 2, 23).
- [117] Litu Rout, Advait Parulekar, Constantine Caramanis, and Sanjay Shakkottai. “A theoretical justification for image inpainting using denoising diffusion probabilistic models”. In: *arXiv preprint arXiv:2302.01217* (2023) (pages 5, 12).
- [118] Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, and Sanjay Shakkottai. “Solving linear inverse problems provably via posterior sampling with latent diffusion models”. In: *Advances in Neural Information Processing Systems* 36 (2024) (pages 2, 22).
- [119] François Rozet, G r me Andry, Fran ois Lanusse, and Gilles Louppe. “Learning Diffusion Priors from Observations by Expectation Maximization”. In: *arXiv preprint arXiv:2405.13712* (2024) (pages 2, 5, 9, 13).
- [120] Leonid I Rudin, Stanley Osher, and Emad Fatemi. “Nonlinear total variation based noise removal algorithms”. In: *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268 (page 4).
- [121] 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*. 2022, pp. 1–10 (page 4).
- [122] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. “Image super-resolution via iterative refinement”. In: *arXiv preprint arXiv:2104.07636* (2021) (pages 3, 4).
- [123] Koichi Saito, Naoki Murata, Toshimitsu Uesaka, Chieh-Hsin Lai, Yuhta Takida, Takao Fukui, and Yuki Mitsufuji. “Unsupervised vocal dereverberation with diffusion-based generative models”. In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, pp. 1–5 (page 1).
- [124] Yifei Shen, Xinyang Jiang, Yezhen Wang, Yifan Yang, Dongqi Han, and Dongsheng Li. “Understanding Training-free Diffusion Guidance: Mechanisms and Limitations”. In: *arXiv preprint arXiv:2403.12404* (2024) (page 5).
- [125] Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. “Diffusion Schr dinger Bridge Matching”. In: *arXiv preprint arXiv:2303.16852* (2023) (page 4).

- [126] Nir Shlezinger, Jay Whang, Yonina C Eldar, and Alexandros G Dimakis. “Model-based deep learning”. In: *Proceedings of the IEEE* 111.5 (2023), pp. 465–499 (page 4).
- [127] Bowen Song, Soo Min Kwon, Zecheng Zhang, Xinyu Hu, Qing Qu, and Liyue Shen. “Solving Inverse Problems with Latent Diffusion Models via Hard Data Consistency”. In: *The Twelfth International Conference on Learning Representations*. 2024. URL: <https://openreview.net/forum?id=j8hdRqOUhN> (pages 2, 22).
- [128] Jiaming Song, Chenlin Meng, and Stefano Ermon. “Denoising diffusion implicit models”. In: *arXiv preprint arXiv:2010.02502* (2020) (page 6).
- [129] Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. “Pseudoinverse-guided diffusion models for inverse problems”. In: *International Conference on Learning Representations*. 2022 (pages 2, 5, 13).
- [130] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. “Consistency Models”. In: *International Conference on Machine Learning*. PMLR. 2023, pp. 32211–32252 (pages 21, 24).
- [131] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. “Maximum likelihood training of score-based diffusion models”. In: *Advances in neural information processing systems* 34 (2021), pp. 1415–1428 (page 18).
- [132] Yang Song and Stefano Ermon. “Generative modeling by estimating gradients of the data distribution”. In: *Advances in Neural Information Processing Systems* 32 (2019) (page 5).
- [133] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. “Solving inverse problems in medical imaging with score-based generative models”. In: *arXiv preprint arXiv:2111.08005* (2021) (pages 1, 19).
- [134] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. “Score-based generative modeling through stochastic differential equations”. In: *arXiv preprint arXiv:2011.13456* (2020) (pages 2, 5, 6, 8, 11, 12, 18).
- [135] Suhas Sreehari, S Venkat Venkatakrishnan, Brendt Wohlberg, Gregory T Buzzard, Lawrence F Drummy, Jeffrey P Simmons, and Charles A Bouman. “Plug-and-play priors for bright field electron tomography and sparse interpolation”. In: *IEEE Transactions on Computational Imaging* 2.4 (2016), pp. 408–423 (page 4).
- [136] Charles M Stein. “Estimation of the mean of a multivariate normal distribution”. In: *The annals of Statistics* (1981), pp. 1135–1151 (page 9).
- [137] Yu Sun, Zihui Wu, Yifan Chen, Berthy T Feng, and Katherine L Bouman. “Provable probabilistic imaging using score-based generative priors”. In: *IEEE Transactions on Computational Imaging* (2024) (pages 2, 20).
- [138] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. “Scale-Recurrent Network for Deep Image Deblurring”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018 (page 4).
- [139] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005 (page 4).
- [140] U Tariq, P Lai, M Lustig, M Alley, M Zhang, G Gold, and Vasanawala Shreyas S. “MRI Data: Undersampled Knees”. In: *Undersampled Knees | MRI Data ()*. URL: <http://old.mridata.org/undersampled/knees> (page 2).
- [141] Brian L. Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi S. Jaakkola. “Diffusion Probabilistic Modeling of Protein Backbones in 3D for the motif-scaffolding problem”. In: *The Eleventh International Conference on Learning Representations*. 2023. URL: <https://openreview.net/forum?id=6TxBxqNME1Y> (pages 2, 20).
- [142] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. “Maxim: Multi-axis mlp for image processing”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 5769–5780 (page 4).
- [143] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. “Plug-and-play priors for model based reconstruction”. In: *2013 IEEE Global Conference on Signal and Information Processing*. IEEE. 2013, pp. 945–948 (page 4).
- [144] Pascal Vincent. “A connection between score matching and denoising autoencoders”. In: *Neural computation* 23.7 (2011), pp. 1661–1674 (pages 7, 36).

- [145] Jean Virieux and Stéphane Operto. “An overview of full-waveform inversion in exploration geophysics”. In: *Geophysics* 74.6 (2009), WCC1–WCC26 (page 1).
- [146] 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: *arXiv preprint arXiv:2405.16749* (2024) (pages 2, 5, 21).
- [147] Yifei Wang, Weimin Bai, Weijian Luo, Wenzheng Chen, and He Sun. “Integrating Amortized Inference with Diffusion Models for Learning Clean Distribution from Corrupted Images”. In: *arXiv preprint arXiv:2407.11162* (2024) (page 9).
- [148] Yinhuai Wang, Jiwen Yu, and Jian Zhang. “Zero-shot image restoration using denoising diffusion null-space model”. In: *arXiv preprint arXiv:2212.00490* (2022) (pages 2, 5, 15).
- [149] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. “Deblurring via stochastic refinement”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 16293–16303 (page 4).
- [150] Luhuan Wu, Brian L. Trippe, Christian A Naesseth, John Patrick Cunningham, and David Blei. “Practical and Asymptotically Exact Conditional Sampling in Diffusion Models”. In: *Thirty-seventh Conference on Neural Information Processing Systems*. 2023. URL: <https://openreview.net/forum?id=eWKqr1zcRv> (pages 2, 5, 20).
- [151] Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L. Bouman. *Principled Probabilistic Imaging using Diffusion Models as Plug-and-Play Priors*. 2024. arXiv: 2405.18782 [eess.IV] (pages 2, 5, 18).
- [152] Carl Wunsch. *The ocean circulation inverse problem*. Cambridge University Press, 1996 (page 1).
- [153] Tongda Xu, Ziran Zhu, Jian Li, Dailan He, Yuanyuan Wang, Ming Sun, Ling Li, Hongwei Qin, Yan Wang, Jingjing Liu, and Ya-Qin Zhang. “Consistency Model is an Effective Posterior Sample Approximation for Diffusion Inverse Solvers”. In: (2024). arXiv: 2403.12063 [cs.CV] (pages 2, 21).
- [154] Yuting Xu, Deeptak Verma, Robert P Sheridan, Andy Liaw, Junshui Ma, Nicholas M Marshall, John McIntosh, Edward C Sherer, Vladimir Svetnik, and Jennifer M Johnston. “Deep dive into machine learning models for protein engineering”. In: *Journal of chemical information and modeling* 60.6 (2020), pp. 2773–2790 (page 2).
- [155] Kevin K Yang, Zachary Wu, and Frances H Arnold. “Machine-learning-guided directed evolution for protein engineering”. In: *Nature methods* 16.8 (2019), pp. 687–694 (page 2).
- [156] Lingxiao Yang, Shutong Ding, Yifan Cai, Jingyi Yu, Jingya Wang, and Ye Shi. “Guidance with Spherical Gaussian Constraint for Conditional Diffusion”. In: *arXiv preprint arXiv:2402.03201* (2024) (page 12).
- [157] G Alastair Young, Richard L Smith, and Richard L Smith. *Essentials of statistical inference*. Vol. 16. Cambridge University Press, 2005 (page 4).
- [158] Tao Yu, Runseng Feng, Ruoyu Feng, Jinming Liu, Xin Jin, Wenjun Zeng, and Zhibo Chen. “Inpaint Anything: Segment Anything Meets Image Inpainting”. In: *arXiv preprint arXiv:2304.06790* (2023) (page 2).
- [159] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. “Restormer: Efficient transformer for high-resolution image restoration”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 5728–5739 (page 4).
- [160] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. “Multi-stage progressive image restoration”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pp. 14821–14831 (page 4).
- [161] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. “fastMRI: An open dataset and benchmarks for accelerated MRI”. In: *arXiv preprint arXiv:1811.08839* (2018) (pages 2, 9).
- [162] Tao Zhang, John Pauly, Shreyas Vasanawala, and Michael Lustig. “MRI Data: Undersampled Abdomens”. In: *Undersampled Abdomens | MRI Data* (). URL: <http://old.mridata.org/undersampled/abdomens> (page 2).

- [163] Yanzhi 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: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 1219–1229 (pages 2, 5, 16).

A Proofs

Lemma A.1 (Conditional Expectation and MMSE). *Let \mathbf{X}_0 and \mathbf{X}_t be two random variables, and $h_\theta(\mathbf{x}_t, t)$ be a function parameterized by θ . Then:*

$$\operatorname{argmin}_\theta \mathbb{E} [||h_\theta(\mathbf{x}_t, t) - \mathbf{x}_0||^2] = \operatorname{argmin}_\theta \mathbb{E} [||h_\theta(\mathbf{x}_t, t) - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]||^2] \quad (\text{A.1})$$

That is, the function $h_\theta(\mathbf{x}_t, t)$ that minimizes the mean squared error with respect to \mathbf{x}_0 is the one that best approximates the conditional expectation $\mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]$.

Proof.

$$\operatorname{argmin}_\theta \mathbb{E} [||h_\theta(\mathbf{x}_t, t) - \mathbf{x}_0||^2] \quad (\text{A.2})$$

$$= \operatorname{argmin}_\theta \mathbb{E} [||h_\theta(\mathbf{x}_t, t) - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t] + \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t] - \mathbf{x}_0||^2] \quad (\text{A.3})$$

$$= \operatorname{argmin}_\theta \mathbb{E} \left[||h_\theta(\mathbf{x}_t, t) - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]||^2 - 2(h_\theta(\mathbf{x}_t, t) - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t])^\top (\mathbf{x}_0 - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]) + ||\mathbf{x}_0 - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]||^2 \right] \quad (\text{A.4})$$

$$= \operatorname{argmin}_\theta \mathbb{E} [||h_\theta(\mathbf{x}_t, t) - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]||^2 - 2h_\theta(\mathbf{x}_t, t)^\top (\mathbf{x}_0 - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t])]. \quad (\text{A.5})$$

Now, for the second term, we have:

$$\mathbb{E}_{\mathbf{x}_0, \mathbf{x}_t} [h_\theta(\mathbf{x}_t, t)^\top (\mathbf{x}_0 - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t])] = \mathbb{E}_{\mathbf{x}_t} \mathbb{E}_{\mathbf{x}_0|\mathbf{x}_t} [h_\theta(\mathbf{x}_t, t)^\top (\mathbf{x}_0 - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t])] \quad (\text{A.6})$$

$$= \mathbb{E}_{\mathbf{x}_t} [h_\theta(\mathbf{x}_t, t)^\top (\mathbb{E}_{\mathbf{x}_0|\mathbf{x}_t} [\mathbf{x}_0 - \mathbb{E}[\mathbf{x}_0|\mathbf{x}_t]])] = 0, \quad (\text{A.7})$$

which concludes the proof. \square

A.1 Tweedie's Formula

Lemma A.2 (Tweedie's Formula). *Let:*

$$\mathbf{X}_t = \mathbf{X}_0 + \sigma_t \mathbf{Z}, \quad (\text{A.8})$$

for $\mathbf{X}_0 \sim p_{\mathbf{X}_0}$ and $\mathbf{Z} \sim \mathcal{N}(0, I)$. Then,

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) = \frac{\mathbb{E}[\mathbf{X}_0|\mathbf{x}_t] - \mathbf{x}_t}{\sigma_t^2}. \quad (\text{A.9})$$

Proof.

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) = \frac{1}{p_t(\mathbf{x}_t)} \nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t) = \frac{1}{p_t(\mathbf{x}_t)} \nabla_{\mathbf{x}_t} \int p_t(\mathbf{x}_t, \mathbf{x}_0) d\mathbf{x}_0 \quad (\text{A.10})$$

$$= \frac{1}{p_t(\mathbf{x}_t)} \nabla_{\mathbf{x}_t} \int p_t(\mathbf{x}_t|\mathbf{x}_0) p_0(\mathbf{x}_0) d\mathbf{x}_0 \quad (\text{A.11})$$

$$= \frac{1}{p_t(\mathbf{x}_t)} \int \nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t|\mathbf{x}_0) p_0(\mathbf{x}_0) d\mathbf{x}_0 \quad (\text{A.12})$$

$$= \frac{1}{p_t(\mathbf{x}_t)} \int p_t(\mathbf{x}_t|\mathbf{x}_0) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0) p_0(\mathbf{x}_0) d\mathbf{x}_0 \quad (\text{A.13})$$

$$= \int p_0(\mathbf{x}_0|\mathbf{x}_t) \frac{\mathbf{x}_0 - \mathbf{x}_t}{\sigma_t^2} d\mathbf{x}_0 \quad (\text{A.14})$$

$$= \frac{\mathbb{E}[\mathbf{X}_0|\mathbf{x}_t] - \mathbf{x}_t}{\sigma_t^2}. \quad (\text{A.15})$$

\square

A.2 Denoising Score Matching

By leveraging the MMSE interpretation of the conditional expectation and Tweedie's formula, one can approximate the score function by training a model to predict the clean image from a corrupted observation (via supervised learning). At inference time, the trained network can be converted to a model that approximates the score through Tweedie's formula. This training procedure is typically known as \mathbf{x}_0 -prediction loss. An alternative, but equivalent, way is to train for the score directly. Vincent [144] independently discovered Denoising Score Matching, which has as a unique minimizer the score function. DSM and the \mathbf{x}_0 -prediction objective are the same up to a simple network reparametrization.

Theorem A.3 (Denoising Score Matching [144]). *Let p_0, p_t be two distributions in \mathbb{R}^n . Assume that all the conditional distributions, $p_t(\mathbf{x}_t|\mathbf{x}_0)$, are supported and differentiable in \mathbb{R}^n . Let:*

$$J_1(\theta) = \frac{1}{2} \mathbb{E}_{\mathbf{x}_t \sim p_t} \left[\|\mathbf{s}_\theta(\mathbf{x}_t) - \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\|^2 \right], \quad (\text{A.16})$$

$$J_2(\theta) = \frac{1}{2} \mathbb{E}_{(\mathbf{x}_0, \mathbf{x}_t) \sim p_0(\mathbf{x}_0)p_t(\mathbf{x}_t|\mathbf{x}_0)} \left[\|\mathbf{s}_\theta(\mathbf{x}_t) - \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0)\|^2 \right]. \quad (\text{A.17})$$

Then, J_1 and J_2 have the same minimizer.

We include the proof listed in [35] for completeness.

Proof.

$$J_1(\theta) = \frac{1}{2} \mathbb{E}_{\mathbf{x}_t \sim p_t} \left[\|\mathbf{s}_\theta(\mathbf{x}_t)\|^2 - 2\mathbf{s}_\theta(\mathbf{x}_t)^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \|\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\|^2 \right] \quad (\text{A.18})$$

$$= \frac{1}{2} \mathbb{E}_{\mathbf{x}_t \sim p_t} [\|\mathbf{s}_\theta(\mathbf{x}_t)\|^2] - \mathbb{E}_{\mathbf{x}_t \sim p_t} [\mathbf{s}_\theta(\mathbf{x}_t)^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)] + C_1. \quad (\text{A.19})$$

Similarly,

$$J_2(\theta) = \frac{1}{2} \mathbb{E}_{\mathbf{x}_t \sim p_t} [\|\mathbf{s}_\theta(\mathbf{x}_t)\|^2] - \mathbb{E}_{(\mathbf{x}_0, \mathbf{x}_t)} [\mathbf{s}_\theta(\mathbf{x}_t)^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0)] + C_2. \quad (\text{A.20})$$

It suffices to show that:

$$\begin{aligned} & \mathbb{E}_{\mathbf{x}_t \sim p_t} [\mathbf{s}_\theta(\mathbf{x}_t)^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)] \\ &= \mathbb{E}_{(\mathbf{x}_0, \mathbf{x}_t) \sim p_0(\mathbf{x}_0)p_t(\mathbf{x}_t|\mathbf{x}_0)} [\mathbf{s}_\theta(\mathbf{x}_t)^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0)]. \end{aligned} \quad (\text{A.21})$$

We start with the second term.

$$\begin{aligned} & \mathbb{E}_{(\mathbf{x}_0, \mathbf{x}_t) \sim p_0(\mathbf{x}_0)p_t(\mathbf{x}_t|\mathbf{x}_0)} [\mathbf{s}_\theta(\mathbf{x}_t)^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0)] \\ &= \int_{\mathbf{x}_0} \int_{\mathbf{x}_t} p_0(\mathbf{x}_0)p_t(\mathbf{x}_t|\mathbf{x}_0) \mathbf{s}_\theta(\mathbf{x}_t)^\top \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0) d\mathbf{x}_t d\mathbf{x}_0 \end{aligned} \quad (\text{A.22})$$

$$= \int_{\mathbf{x}_0} \int_{\mathbf{x}_t} \mathbf{s}_\theta^\top(\mathbf{x}_t) (p_0(\mathbf{x}_0)p_t(\mathbf{x}_t|\mathbf{x}_0) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0)) d\mathbf{x}_t d\mathbf{x}_0 \quad (\text{A.23})$$

$$= \int_{\mathbf{x}_0} \int_{\mathbf{x}_t} \mathbf{s}_\theta^\top(\mathbf{x}_t) \left(p_0(\mathbf{x}_0)p_t(\mathbf{x}_t|\mathbf{x}_0) \frac{1}{p_t(\mathbf{x}_t|\mathbf{x}_0)} \nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t|\mathbf{x}_0) \right) d\mathbf{x}_t d\mathbf{x}_0 \quad (\text{A.24})$$

$$= \int_{\mathbf{x}_0} \int_{\mathbf{x}_t} \mathbf{s}_\theta^\top(\mathbf{x}_t) (p_0(\mathbf{x}_0) \nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t|\mathbf{x}_0)) d\mathbf{x}_t d\mathbf{x}_0 \quad (\text{A.25})$$

$$= \int_{\mathbf{x}_t} \int_{\mathbf{x}_0} \mathbf{s}_\theta^\top(\mathbf{x}_t) (p_0(\mathbf{x}_0) \nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t|\mathbf{x}_0)) d\mathbf{x}_0 d\mathbf{x}_t \quad (\text{A.26})$$

$$= \int_{\mathbf{x}_t} \mathbf{s}_\theta^\top(\mathbf{x}_t) \left(\int_{\mathbf{x}_0} p_0(\mathbf{x}_0) \nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t|\mathbf{x}_0) d\mathbf{x}_0 \right) d\mathbf{x}_t \quad (\text{A.27})$$

$$= \int_{\mathbf{x}_t} \mathbf{s}_\theta^\top(\mathbf{x}_t) \left(\int_{\mathbf{x}_0} \nabla_{\mathbf{x}_t} (p_0(\mathbf{x}_0)p_t(\mathbf{x}_t|\mathbf{x}_0)) d\mathbf{x}_0 \right) d\mathbf{x}_t \quad (\text{A.28})$$

$$= \int_{\mathbf{x}_t} \mathbf{s}_\theta^\top(\mathbf{x}_t) \left(\nabla_{\mathbf{x}_t} \left(\int_{\mathbf{x}_0} p_0(\mathbf{x}_0)p_t(\mathbf{x}_t|\mathbf{x}_0) d\mathbf{x}_0 \right) \right) d\mathbf{x}_t \quad (\text{A.29})$$

$$= \int_{\mathbf{x}_t} \mathbf{s}_\theta^\top(\mathbf{x}_t) \nabla_{\mathbf{x}_t} p_t(\mathbf{x}_t) d\mathbf{x}_t \quad (\text{A.30})$$

$$= \int_{\mathbf{x}_t} p_t(\mathbf{x}_t) \mathbf{s}_\theta^\top(\mathbf{x}_t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) d\mathbf{x}_t \quad (\text{A.31})$$

$$= \mathbb{E}_{\mathbf{x}_t \sim p_t(\mathbf{x}_t)} [\mathbf{s}_\theta^\top(\mathbf{x}_t) \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)]. \quad (\text{A.32})$$

□

A.3 Jacobian of the score

Lemma A.4 (Jacobian of score-function). *Let:*

$$\mathbf{X}_t = \mathbf{X}_0 + \sigma_t \mathbf{Z}, \quad (\text{A.33})$$

for $\mathbf{X}_0 \sim p_{\mathbf{X}_0}$ and $\mathbf{Z} \sim \mathcal{N}(0, I)$. Then,

$$\mathbf{H}(\log p_{\mathbf{X}_t})(\mathbf{x}_t) = \frac{\mathbb{E}[\mathbf{X}_0 \mathbf{X}_0^\top | \mathbf{x}_t] - \mathbb{E}[\mathbf{X}_0 | \mathbf{x}_t] \mathbb{E}^\top[\mathbf{X}_0 | \mathbf{x}_t]}{\sigma_t^4} - \frac{I}{\sigma_t^2}. \quad (\text{A.34})$$

Proof.

$$\nabla_{\mathbf{x}_t} \log p_{\mathbf{X}_t}(\mathbf{x}_t) = \frac{\mathbb{E}[\mathbf{X}_0 | \mathbf{x}_t] - \mathbf{x}_t}{\sigma_t^2} \quad (\text{A.35})$$

$$\Rightarrow \sigma_t^2 \mathbf{H}(\log p_{\mathbf{X}_t})(\mathbf{x}_t) = \text{Jacob}(\mathbb{E}[\mathbf{X}_0 | \mathbf{x}_t]) - I. \quad (\text{A.36})$$

We will now analyze the Jacobian.

$$\text{Jacob}(\mathbb{E}[\mathbf{X}_0|\mathbf{x}_t]) = \int \nabla_{\mathbf{x}_t} (p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t)\mathbf{x}_0) d\mathbf{x}_0 \quad (\text{A.37})$$

$$= \int \mathbf{x}_0 \nabla_{\mathbf{x}_t}^\top p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t) d\mathbf{x}_0 \quad (\text{A.38})$$

$$= \int \mathbf{x}_0 p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t) \nabla_{\mathbf{x}_t}^\top \log \left(\frac{p_{\mathbf{X}_t}(\mathbf{x}_t|\mathbf{x}_0)p_{\mathbf{X}_0}(\mathbf{x}_0)}{p_{\mathbf{X}_t}(\mathbf{x}_t)} \right) d\mathbf{x}_0 \quad (\text{A.39})$$

$$= \int \mathbf{x}_0 p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t) \nabla_{\mathbf{x}_t}^\top \log \left(\frac{p_{\mathbf{X}_t}(\mathbf{x}_t|\mathbf{x}_0)}{p_{\mathbf{X}_t}(\mathbf{x}_t)} \right) d\mathbf{x}_0 \quad (\text{A.40})$$

$$= \int \mathbf{x}_0 p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t) \nabla_{\mathbf{x}_t}^\top \log p_{\mathbf{X}_t}(\mathbf{x}_t|\mathbf{x}_0) d\mathbf{x}_0 - \int \mathbf{x}_0 p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t) \nabla_{\mathbf{x}_t}^\top \log p_{\mathbf{X}_t}(\mathbf{x}_t) d\mathbf{x}_0 \quad (\text{A.41})$$

$$= \int \mathbf{x}_0 p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t) \frac{\mathbf{x}_0^\top - \mathbf{x}_t^\top}{\sigma_t^2} d\mathbf{x}_0 - \int \mathbf{x}_0 p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t) \nabla_{\mathbf{x}_t}^\top \log p_{\mathbf{X}_t}(\mathbf{x}_t) d\mathbf{x}_0 \quad (\text{A.42})$$

$$= \int \mathbf{x}_0 p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t) \frac{\mathbf{x}_0^\top - \mathbf{x}_t^\top}{\sigma_t^2} d\mathbf{x}_0 - \int \mathbf{x}_0 p_{\mathbf{X}_0}(\mathbf{x}_0|\mathbf{x}_t) \frac{\mathbb{E}^\top[\mathbf{x}_0|\mathbf{x}_t] - \mathbf{x}_t^\top}{\sigma_t^2} d\mathbf{x}_0 \quad (\text{A.43})$$

$$= \frac{1}{\sigma_t^2} (\mathbb{E}[\mathbf{x}_0 \mathbf{x}_0^\top | \mathbf{x}_t] - \mathbb{E}[\mathbf{x}_0 | \mathbf{x}_t] \mathbb{E}[\mathbf{x}_0 | \mathbf{x}_t]^\top). \quad (\text{A.44})$$

□

Corollary A.5. *Let:*

$$\mathbf{X}_t = \mathbf{X}_0 + \sigma_t \mathbf{Z}, \quad (\text{A.45})$$

for $\mathbf{X}_0 \sim p_{\mathbf{X}_0}$, $\mathbf{X}_0 \in \mathbb{R}^n$ and $\mathbf{Z} \sim \mathcal{N}(0, I)$. Then,

$$\nabla_{\mathbf{x}_t}^2 \log p_t(\mathbf{x}_t) = \frac{\mathbb{E}[||\mathbf{X}_0||^2 | \mathbf{x}_t] - ||\mathbb{E}[\mathbf{X}_0|\mathbf{x}_t]||^2}{\sigma_t^4} - \frac{n}{\sigma_t^2}. \quad (\text{A.46})$$