Diffusion Models for Inverse Problems

Brendan Dowling
Supervisor: Deniz Akyildiz

Background

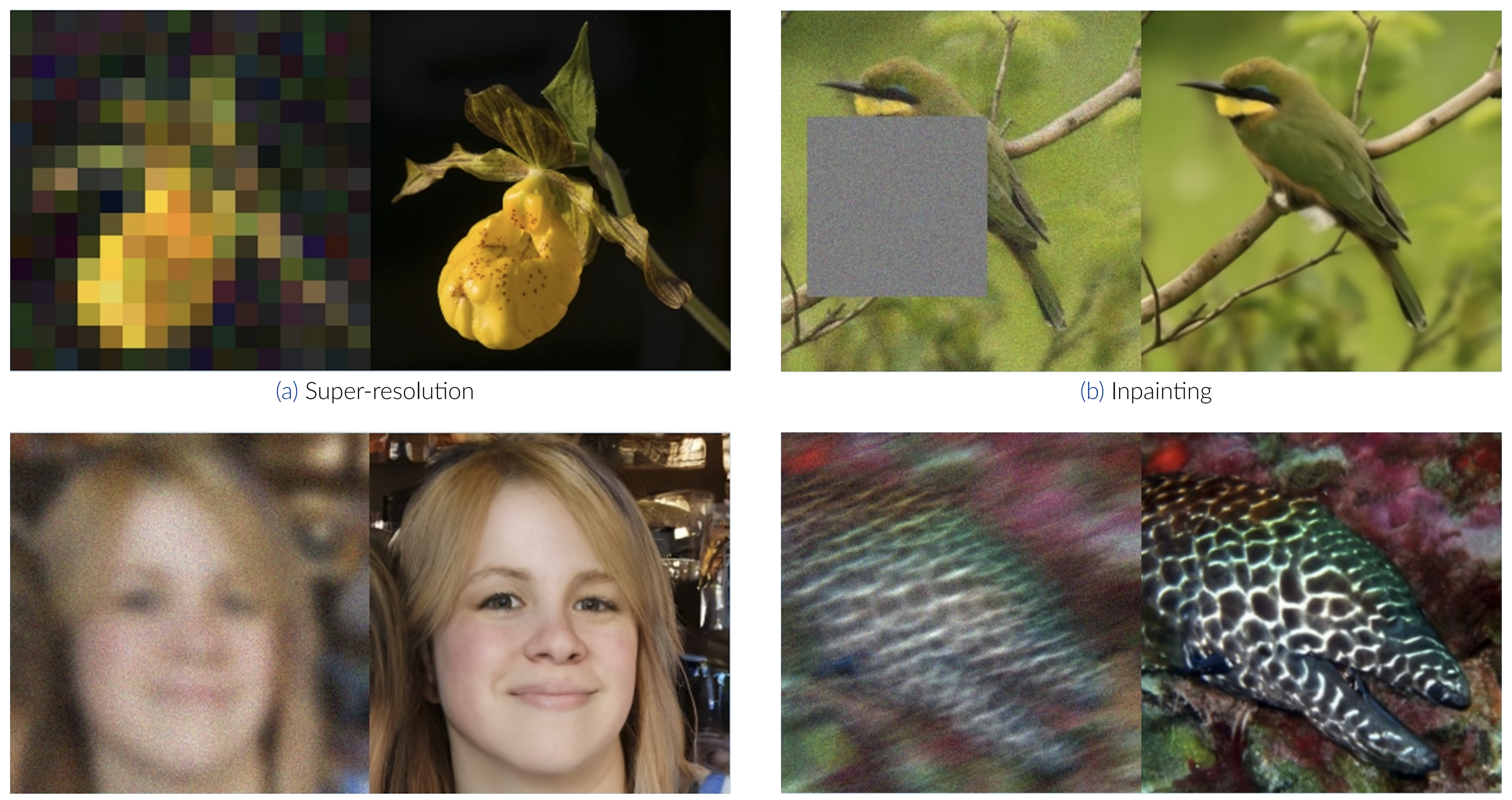
Linear inverse problems: Given a datapoint $\mathbf{x} \in \mathbb{R}^{d_x}$, denote some lossy measurement by

$$\mathbf{y} = A\mathbf{x} + \sigma_y \epsilon \in \mathbb{R}^{d_y}, \quad \epsilon \sim \mathcal{N}\left(0, \mathrm{I}_{d_y}\right)$$

where $A \in \mathbb{R}^{d_y \times d_x}$ is some *measurement matrix*, and $\sigma_y \in \mathbb{R}^+$ controls the measurement noise. The goal of a linear inverse problem is to recover $\mathbf{x}$ from $\mathbf{y}$. Typically $d_x > d_y$, leading to a many-to-one $\mathbf{x} \to \mathbf{y}$ mapping (Chung et al.), and requiring some *prior* information about $\mathbf{x}$. If we assume some prior distribution of the data, $p(\mathbf{x})$, "solving" the inverse problem amounts to sampling from some posterior:

$$p(\mathbf{x} \mid \mathbf{y}) \propto p(\mathbf{x})g(\mathbf{y} \mid \mathbf{x}) = p(\mathbf{x})\mathcal{N}\left(\mathbf{y} \mid A\mathbf{x}, \sigma_y^2 \mathrm{I}_{d_y}\right)$$

For many problems, $p(\mathbf{x})$ is generally unknown or does not conjugate with $g(\mathbf{y} \mid \mathbf{x})$, necessitating numerical methods to sample from $p(\mathbf{x} \mid \mathbf{y})$.



(a) Super-resolution
(b) Inpainting
(c) Gaussian deblur
(d) Motion deblur

Figure 1. Examples of linear inverse problems for images (Chung et al.)

Diffusion models: Diffusion models provide a means of sampling from intractable data distributions, $p(\mathbf{x})$, and have proven highly effective in high dimensional settings, such as image generation (Dhariwal and Nichol). They can be formulated by considering some forwards noising process and some backwards de-noising process (Dou and Y. Song). In their discrete representation, the forward process can be described by a Markov chain:

$$q(\mathbf{x}_{1:T} \mid \mathbf{x}_0) = \prod_{t=1}^{T} \mathcal{N}\left(a_t \mathbf{x}_{t-1}, b_t^2 \mathrm{I}_{d_x}\right) \tag{1}$$

where $\{(a_t, b_t)\}_{t=1}^T$ differ depending on formulation of the problem (Y. Song et al.). From this, we yield the marginal $q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(c_t \mathbf{x}_0, d_t^2 \mathrm{I}_{d_x})$ where c_t, d_t are derived from a_t and b_t. We assume the backwards process to similarly be a Markov chain and train an ϵ (Ho et al.) or score (Y. Song et al.) approximating neural network, $s_\theta(\mathbf{x}_t, t)$, to estimate the score function $\nabla_{\mathbf{x}_t} \log q(\mathbf{x}_t \mid \mathbf{x}_0)$ so to enable sampling from:

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \mathcal{N}\left(u_t \cdot \frac{\mathbf{x}_t + d_t^2 s_\theta(\mathbf{x}_t, t)}{c_t} + v_t s_\theta(\mathbf{x}_t, t), w_t^2 \mathrm{I}_{d_x}\right)$$

with u_t, v_t some functions of a_t, b_t, and the mean derived from Tweedie's formula (Dou and Y. Song).

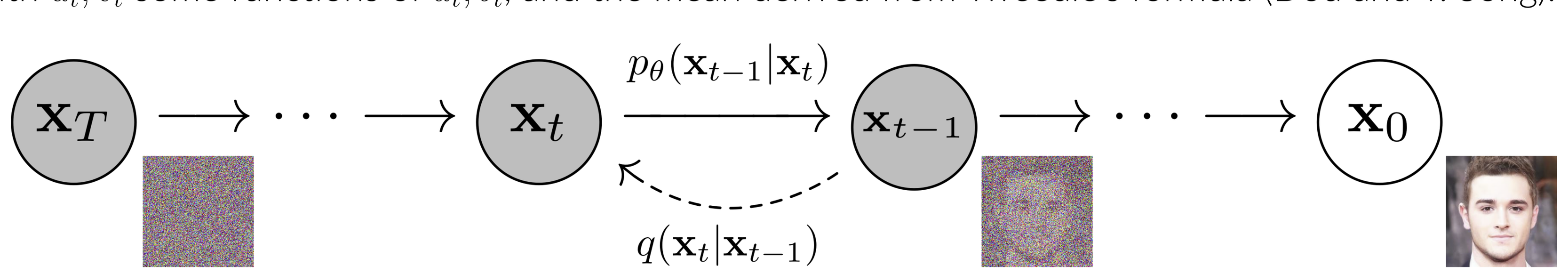


Figure 2. DDPM graphical model (Ho et al.)

Problem Overview

Using diffusion models for solving inverse problems amounts to guiding each step of the backwards process towards a region on the data manifold which is sensibly explains the measurement. This is achieved by considering $p(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{y})$ transitions, and replacing the unconditional score with the conditional one (Chung et al.):

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t \mid \mathbf{y}) = \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y} \mid \mathbf{x}_t) \tag{2}$$

The first term in Equation 2 is exactly what the unconditional score network, $s_\theta(\mathbf{x}_t, t)$, estimates. The second term, however, is generally difficult to compute due to the time t dependence (Chung et al.); many different methods (Chung et al., J. Song et al., Boys et al.) aim to approximate it.

Significant recent research (Cardoso et al., Trippe et al., Wu et al., Janati et al.) has considered using SMC methods to instead more directly target the posterior and circumvent the need to compute this term. Below, we describe a very simple SMC algorithm which aligns closely with the MCGdiff algorithm of Cardoso et al. except with a much simpler covariance structure.

We note here that the focus is on taking pre-trained, unconditional score models and using them for general linear inverse problems. One could train a model specific to each inverse problem, with such a model then providing conditional scores out of the box. However, such a model will not be usable on other inverse tasks. Furthermore, training such a conditional model requires paired/labeled data which may not be easily available; this also has implications for generalization/few-shot learning. That said, when score is limited and compute (for training) is not an issue, such conditional models provide state-of-the-art performance.

A Simple SMC Approach

Consider the state space model (SSM):

$$\mathbf{X}_t \sim p_{t|t+1}(\cdot \mid \mathbf{X}_{t+1}), \quad \mathbf{X}_T \sim \mathcal{N}(0_{d_x}, \mathrm{I}_{d_x})$$
$$\mathbf{Y}_t \mid \mathbf{X}_t \sim \mathcal{N}(H\mathbf{X}_t, \sigma \mathrm{I}_{d_y})$$

where $p_{t|t+1}(\cdot \mid \mathbf{X}_{t+1})$ is the diffusion prior model. The diffusion model begins at $\mathbf{X}_T$, near 0_{d_x}, and iteratively moves towards some $\mathbf{x}_0$ on the data manifold. However, a measurement, $\mathbf{y}$, is only available at $t = 0$, and is likely to live far from $\mathbf{y}_t$ (according to the SSM) for large t. Hence, we construct an auxiliary sequence $\{\mathbf{y}_t\}_{t=1}^T$ according to

$$\mathbf{y}_t = \rho_t \mathbf{y}_{t-1}, \quad y_0 = y$$

with $0 < \rho_{t-1} < \rho_t < 1$, so to align the initial and intermediate measurements with the prior model. Loosely, this enables "smoother" guidance of diffusion process towards $p(\mathbf{x}_0 \mid \mathbf{y})$ samples. A sensible choice for $\{\rho_t\}_{t=1}^T$ are the coefficients $\{a_t\}_{t=1}^T$ that define the forward transition in Equation 1.

Define a Feynman-Kac model (Chopin and Papaspiliopoulos) with the $p_{t|t+1}$ transition of the SSM as the proposal but with weights for each i particle set according to the ratio of likelihoods:

$$\omega_t^{(i)} = \frac{g\left(\mathbf{y}_t \mid \tilde{\mathbf{x}}_t^{(i)}\right)}{g\left(\mathbf{y}_{t+1} \mid \mathbf{x}_{t+1}^{i}\right)} = \frac{g\left(\overline{\alpha}_t^{\frac{1}{2}} \mathbf{y} \mid \tilde{\mathbf{x}}_t^{(i)}\right)}{g\left(\overline{\alpha}_{t+1}^{\frac{1}{2}} \mathbf{y} \mid \mathbf{x}_{t+1}^{i}\right)} = \frac{\mathcal{N}\left(\overline{\alpha}_t^{\frac{1}{2}} \mathbf{y};\ H\tilde{\mathbf{x}}_t^{(i)}, \sigma_y I_{d_y}\right)}{\mathcal{N}\left(\overline{\alpha}_{t+1}^{\frac{1}{2}} \mathbf{y};\ H\mathbf{x}_{t+1}^{(i)}, \sigma_y I_{d_y}\right)}$$

Running a particle filter will then provide approximate posterior samples. Note here that we can easily use a DDIM (J. Song et al.) sampler to reduce the number of timesteps/iterations.

Future Work

- Consider more complex covariance structures, such as those of Cardoso et al.
- Consider more complex likelihood models and observation noise scheduling.
- Examine performance (quality and efficiency) on image-based task.
- Consider link with twisting functions (Wu et al.).

Example

Setup

Consider the mixture model, q_{data}, example of Cardoso et al., with 25 equally weighted d_x-dimensional Gaussian random variables with means $\mu_{i,j} := (8i, 8j, \ldots, 8i, 8j) \in \mathbb{R}^{d_x}$ for $(i, j) \in \{-2, \ldots, 2\}^2$ and unit variance. We generate some d_y-dimensional measurement, $\mathbf{y}$, according to the following process:

- Draw $\tilde{A} \sim \mathcal{N}(0,1)^{d_y \times d_x} \in \mathbb{R}^{d_y \times d_x}$ and compute its SVD decomposition, $USV^\top$.
- Sample $s_{i,j} \sim \mathcal{U}[0,1]$ for $(i, j) \in \{-2, \ldots, 2\}^2$.
- Set $A := U\, diag(\{s_{i,j}\})\, V^\top$.
- Draw $\mathbf{x}_* \sim q_{\text{data}}$ and set $\mathbf{y} := A\mathbf{x}_* + \sigma_y \epsilon$, $\epsilon \sim \mathcal{N}(\mathbf{0}_{d_y}, \mathrm{I}_{d_y})$

For this model, the posterior, $p(\mathbf{x}_* \mid \mathbf{y})$, is available in closed form (another GMM; see Cardoso et al.) which enables us to benchmark numerical methods. The backwards marginal is also available in closed form allowing us to avoid training a score network (but still using DDPM sampling).

Results

We run our simple SMC algorithm for this model, using 1000 particles and adaptive systematic resampling when the ESS dropped below 90%. We considered many different values for d_x and d_y; Figure 3 corresponds to $d_x = d_y = 2$. From 3, we see it has well targeted the true posterior distribution. Empirically (via sliced Wasserstein distance) the algorithm retained performance in higher dimensional settings, in spite of its simplified covariance structure, which provides some promise for applications in image-based settings.

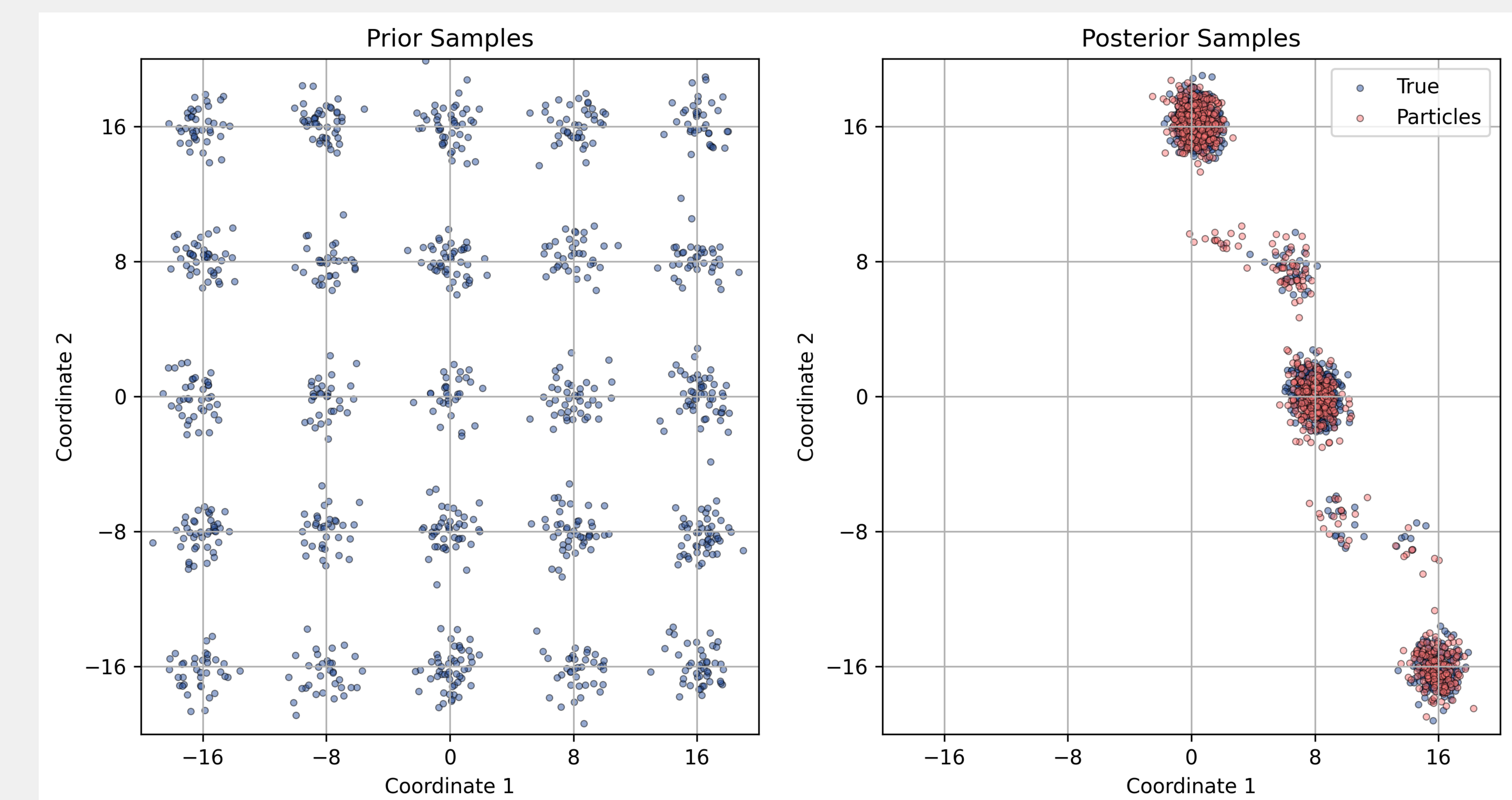


Figure 3. SMC performance on GMM example

References

Boys, Benjamin, et al., "Tweedie Moment Projected Diffusions For Inverse Problems". *arXiv*, 22 Nov. 2023.
Cardoso, Gabriel, et al., "Monte Carlo Guided Diffusion for Bayesian Linear Inverse Problems". *arXiv*, 25 Oct. 2023.
Chopin, Nicolas, and Omiros Papaspiliopoulos, *An Introduction to Sequential Monte Carlo*. Springer, 2020. Springer Series in Statistics.
Chung, Hyungjin, et al., "Diffusion Posterior Sampling for General Noisy Inverse Problems". 2022.
Dhariwal, Prafulla, and Alex Nichol, "Diffusion Models Beat GANs on Image Synthesis". *arXiv*, 1 June 2021.
Dou, Zehao, and Yang Song, "Diffusion Posterior Sampling for Linear Inverse Problem Solving: A Filtering Perspective". The Twelfth International Conference on Learning Representations, 13 Oct. 2023.
Ho, Jonathan, et al., "Denoising Diffusion Probabilistic Models". 2020.
Janati, Yazid, et al., "Divide-and-Conquer Posterior Sampling for Denoising Diffusion Priors". *arXiv*, 17 Mar. 2024.
Song, Jiaming, et al., "Denoising Diffusion Implicit Models". 2020.
Song, Jiaming, et al., "Pseudoinverse-Guided Diffusion Models for Inverse Problems". *International Conference on Learning Representations*. 2023.
Song, Yang, et al., "Score-Based Generative Modeling through Stochastic Differential Equations". *arXiv*, 10 Feb. 2021.
Trippe, Brian L., et al., "Diffusion Probabilistic Modeling of Protein Backbones in 3D for the Motif-Scaffolding Problem". *arXiv*, 19 Mar. 2023.
Wu, Luhuan, et al., "Practical and Asymptotically Exact Conditional Sampling in Diffusion Models". *arXiv*, 30 June 2023.