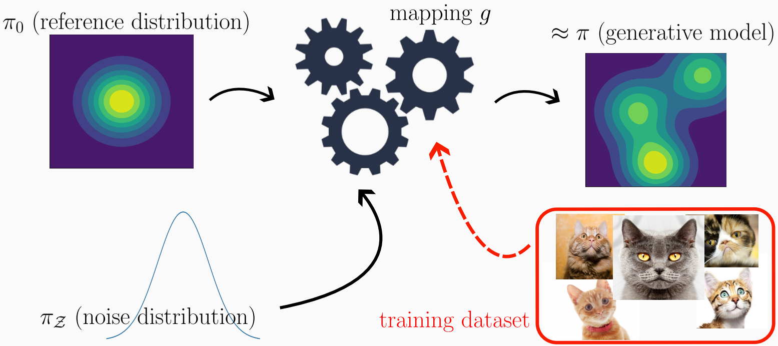
Denoising diffusion models for inverse problems

Valentin De Bortoli

December 8, 2022

What is generative modeling?

- **Generative modeling**: Given a **dataset** of samples from a distribution π how to obtain **new samples** from π?
- **A general approach**:
  - Sample X_0 from π_0 (reference distribution).
  - Sample Z from $\pi_{\mathcal{Z}}$ (noise distribution).
  - Push with $g(X_0, Z) \rightarrow$ approximate sample from π.



Why generative modeling?

- Application in **computational biology**: Senior et al. (2020).
  - **Amino-acid sequence** to **3D structure**.
  - Cryo-Electron Microscopy or crystallography = experimental techniques to determine the shape of the protein.
  - Crystallizing a protein is a real challenge Avanzato et al. (2019).
  - Competition to predict structure: **C**ritical **A**ssessment of protein **S**tructure **P**rediction.
- **Conditional generative modeling**.

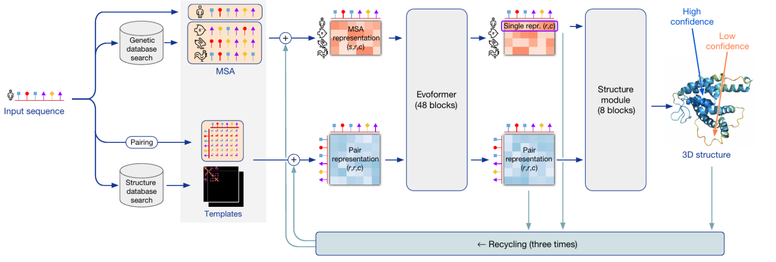
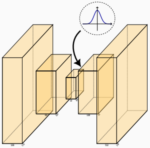


Image extracted from Senior et al. (2020).

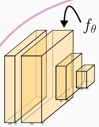
A myriad of models

Variational AutoEncoder



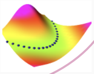
Kingma et al. (2014)
Rezende et al. (2014)
Ranganath et al. (2016)
Vahdat et al. (2021)

Energy-Based Model



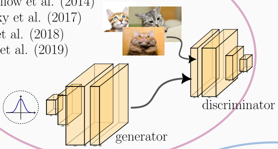
Zhu et al. (1998)
LeCun et al. (2006)
Hinton et al. (2006)
Du et al. (2019)

$$\frac{\exp[-f_\theta(x)]}{\int \exp[-f_\theta(\tilde{x})]\mathrm{d}\tilde{x}}$$



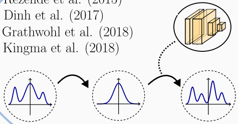
Generative Adversarial Network

Goodfellow et al. (2014)
Arjovsky et al. (2017)
Brock et al. (2018)
Karras et al. (2019)



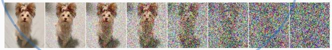
Normalizing Flow

Rezende et al. (2015)
Dinh et al. (2017)
Grathwohl et al. (2018)
Kingma et al. (2018)

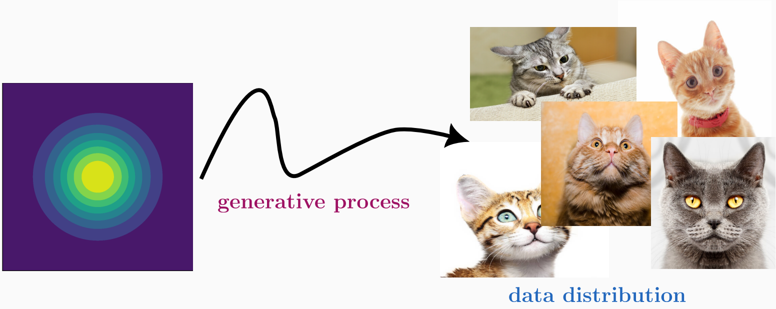


Denoising Diffusion Model

Song et al. (2019)
Ho et al. (2020)
Vahdat et al. (2021)



Some challenges in generative modeling



Theoretical understanding

- Convergence of generative models?

Properties of the data

- Riemannian data.
- Inverse problems.

Properties of the process

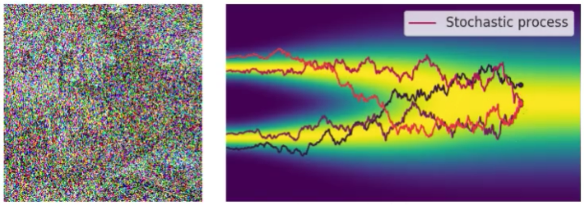
- Optimal transport.
- Stochastic control.

Focus on denoising diffusion models.

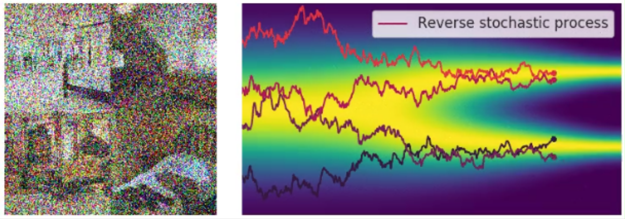
Generative Modeling: the rise of diffusion models

Time-reversal of diffusions

- **Forward decomposition**: $p(x_{0:N}) = p_0(x_0) \prod_{k=0}^{N-1} p_{k+1|k}(x_{k+1}|x_k)$.



- **Backward decomposition**: $p(x_{0:N}) = p_N(x_N) \prod_{k=0}^{N-1} p_{k|k+1}(x_k|x_{k+1})$.



Images extracted from Song and Ermon (2019).

Approximate time reversal

¿How to approximate the backward decomposition?

- **Backward decomposition**: $p(x_{0:N}) = p_N(x_N) \prod_{k=0}^{N-1} p_{k|k+1}(x_k|x_{k+1})$.
  - How to compute $p_{k|k+1}(x_k|x_{k+1}) = p_{k+1|k}(x_{k+1}|x_k)p_k(x_k)/p_{k+1}(x_{k+1})$?
  - In practice $p_{k+1|k} = \mathrm{N}(x_k - \gamma x_k, \sqrt{2\gamma}\,\mathrm{Id})$ is **Gaussian**.
  - (**Discretization** of $\mathrm{d}\mathbf{X}_t = -\mathbf{X}_t\mathrm{d}t + \sqrt{2}\mathrm{d}\mathbf{B}_t$ (**Ornstein-Ulhenbeck**))
  - $p_{k|k+1}$ is approximately Gaussian

$$\boxed{p_{k|k+1} = \mathrm{N}(x_{k+1} + \gamma x_{k+1} + 2\gamma \nabla \log p_{k+1}(x_{k+1}), \sqrt{2\gamma}\,\mathrm{Id}).}$$

¿How to compute the score term?

- **Score matching** techniques: Vincent (2011); Hyvärinen (2005)

$$\nabla \log p_{k+1}(x_{k+1}) = \mathbb{E}_{p_{0|k+1}}[\nabla \log p_{k+1|0}(x_{k+1}|X_0)].$$

  - **Loss function**: $\boxed{\ell(\mathbf{s}_{k+1}) = \mathbb{E}[\|\mathbf{s}_{k+1}(X_{k+1}) - \nabla \log p_{k+1|0}(X_{k+1}|X_0)\|^2].}$
  - Algorithm: replace $\nabla \log p_{k+1}$ by $\mathbf{s}_{k+1}$.

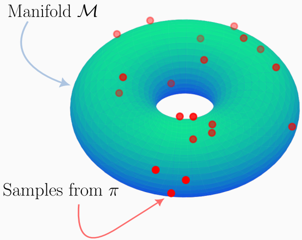
Convergence of diffusion models ($\hat{\pi}$)

Under dissipativity conditions (D.B et al., 2021[1])

- $\|\mathbf{s}_t(x) - \nabla \log p_t(x)\| \leq \mathtt{M}$.
- π admits a density p and $\langle \nabla \log p(x), x \rangle \leq -\mathtt{m}\|x\|^2 + \mathtt{c}$.

Then, there exists $A \geq 0$ such that

$$\|\pi - \hat{\pi}\|_{\mathrm{TV}} \leq A(\underbrace{\exp[-T]}_{\text{forward convergence}} + \exp[T](\underbrace{\gamma^{1/2}}_{\text{discretization}} + \underbrace{\mathtt{M}}_{\text{score approximation}}))$$



Under the manifold hypothesis (D.B., 2022[2])

- π is supported on a compact manifold $\mathcal{M}$.

Then there exists $A \geq 0$ such that

$$\mathbf{W}_1(\pi, \hat{\pi}) \leq A(\exp[-T] + \gamma^{1/2} + \mathtt{M}).$$

[1] **D.B.**, Thornton, Heng, Doucet – Diffusion Schrödinger Bridge – NeurIPS 2021

[2] **D.B.** – Convergence of diffusion models under manifold hypotheses – under review (2022)

Inverse problems with denoising diffusion models

Diffusion models for inverse problems

- **Question:** how to use denoising diffusion models for inverse problems?
- We present several techniques:
  - **Amortization**
  - **Replacement** (with or without correction)
  - **Conditional guidance**

- The goals:
  - **Inpainting, deblurring**
  - **Class conditional generative modelling**
  - **Text-to-image**

Amortization

- The simplest technique: **amortize** everything.
- **Score matching** techniques: Vincent (2011); Hyvärinen (2005)

$$\nabla \log p_{k+1}(x_{k+1}|y) = \mathbb{E}_{p_{0|k+1,y}}[\nabla \log p_{k+1|0}(x_{k+1}|X_0)].$$

  - **Loss function**:
    $$\boxed{\ell(\mathbf{s}_{k+1}) = \mathbb{E}[\|\mathbf{s}_{k+1}(X_{k+1}, Y) - \nabla \log p_{k+1|0}(X_{k+1}|X_0)\|^2].}$$
  - Algorithm: replace $\nabla \log p_{k+1}$ by $\mathbf{s}_{k+1}$.
- Same algorithm as before but instead of sampling X_0 and then noise it, sample (X_0, Y) and then noise it.
- **Advantages:**
  - Straightforward to implement (just another input to the network).
  - Works for generic data.
- **Problems:**
  - What if I only want to train one generative model?
  - What if at inference size y has a different size than the training samples?

The replacement method

- Second technique: **replacement** technique.
- We only train **one** diffusion model.
- Example of inpainting:
  - Train a denoising diffusion model.
  - At inference time, we observe part of the image (y with a mask m)
  - Diffuse y forward in time $Y_{0:N}$
  - Sample $X_N \sim \mathrm{N}(0, \mathrm{Id})$
  - Apply the backward diffusion step:
    $\hat{X}_n = X_{n+1} + \gamma X_{n+1} + 2\gamma \mathbf{s}_\theta(X_n) + \sqrt{2\gamma} Z_n$
  - **Replace** using $X_n = m\hat{X}_n + (1 - m) Y_n$ (pointwise multiplication)
  - Go back to the backward diffusion step and iterate.
- **Advantages:**
  - Only one generative model to train
  - Straightforward to implement
  - Very useful in protein modeling
- **Problems:**
  - Only work on specific problems (mask)
  - No guarantee of convergence

Repaint

Figure extracted from Lugmayr et al. (2022)

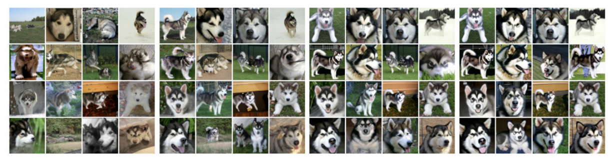
- Additional tricks:
  - Particle filtering at each step to ensure convergence Trippe et al. (2022).
  - Iterating the replacement step Lugmayr et al. (2022)

Explicit guidance

- Third technique: **conditional guidance**
- just guide the diffusion with an extra term in the drift

$$\mathbf{s}_\theta(x) \to \mathbf{s}_\theta(x) + \nabla \omega \log p_\phi(y|x)$$

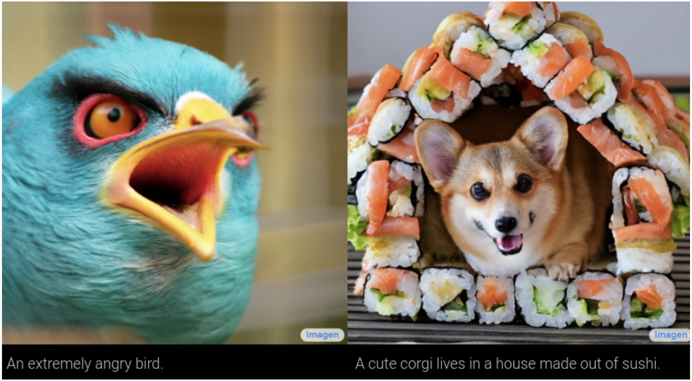
  - ω is the **guidance strength**.
- What is p_ϕ?
  - Classifier in the case of **class conditional sampling** Dhariwal and Nichol (2021).
  - Can be an amortized score model, i.e. (classifier free) $\nabla \log p_\phi(y|x) \to \mathbf{s}_\theta(x, y) - \mathbf{s}_\theta(x)$
  - Push the samples towards $p(x|y)$ and away from $p(x)$.



Increasing amount of guidance on the class “malamute” in ImageNet.

An application: text-to-image

- From prompt to images: Imagen, DALL-E 2, Stable Diffusion, Midjourney.

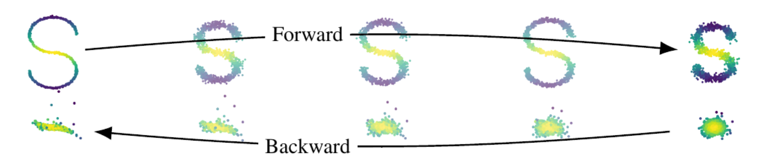


- **CLIP** (**C**ontrastive **L**anguage–**I**mage **P**re-training) guidance.

Schrödinger Bridges: a new generative modeling framework

Shorter generative processes?

- **Not enough stepsizes** leads to poor approximation (the Ornstein-Ulhenbeck process does not mix fast enough).



- Illustration of failure: N is too small so p_N is very different from p_{prior}. This harms the quality of the reconstruction for the time-reversal.

- **Trade-off**:
  - Large $N \to$ improvement in **quality** (fidelity).
  - Large $N \to$ **model is slow** at sampling time.

Challenge: how to “shorten” the diffusion process?

The trilemma of generative modeling

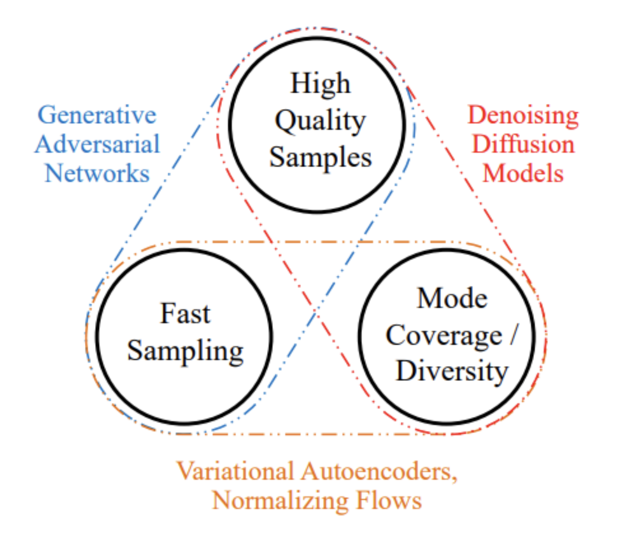


Image extracted from Xiao et al. (2021).

Revisiting Generative Modeling using Schrödinger Bridges

- The **Schrödinger Bridge (SB) problem** is a classical problem appearing in applied mathematics, optimal transport and probability.
  - Consider a **reference density** $p(x_{0:N})$, find $\pi^\star(x_{0:N})$ such that

    $\pi^\star$ distribution on $(\mathbb{R}^d)^{N+1}$

    $$\boxed{\pi^\star = \arg\min\{\mathrm{KL}(\pi|p) : \pi_0 = p_{\mathrm{data}},\ \pi_N = p_{\mathrm{prior}}\}.}$$

  - **Goal:** If $\pi^\star$ is available: $X_N \sim p_{\mathrm{prior}}$ and $X_k \sim \pi^\star_{k|k+1}(\cdot|X_{k+1})$.

- **Static formulation**: $\pi^\star(x_{0:N}) = \pi^{\mathrm{s},\star}(x_0, x_N)p_{|0,N}(x_{1:N-1}|x_0, x_N)$ where
  - Variational form:

    $\pi^{s,\star}$ distribution on $(\mathbb{R}^d)^2$

    $$\boxed{\pi^{\mathrm{s},\star} = \arg\min\{\mathrm{KL}(\pi^{\mathrm{s}}|p_{0,N}) : \pi_0^{\mathrm{s}} = p_{\mathrm{data}},\ \pi_N^{\mathrm{s}} = p_{\mathrm{prior}}\}.}$$

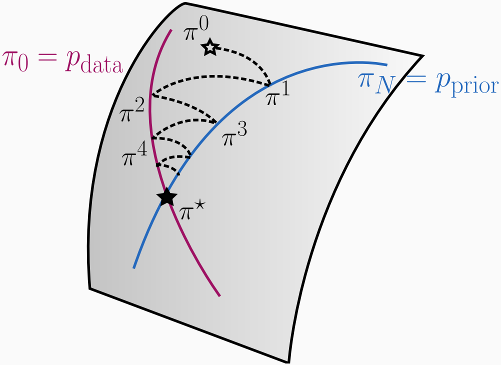
  - In its static form the Schrödinger Bridge is a special case of **entropic optimal transport**, see Mikami (2004).

The Iterative Proportional Fitting algorithm

- The SB problem can be solved using **Iterative Proportional Fitting (IPF)** Sinkhorn and Knopp (1967); Fortet (1940), i.e. set $\pi^0 = p$ and for $n \in \mathbb{N}$

$$\pi^{2n+1} = \arg\min\{\mathrm{KL}(\pi|\pi^{2n}), \;\; \pi_N = p_{\mathrm{prior}}\},$$
$$\pi^{2n+2} = \arg\min\{\mathrm{KL}(\pi|\pi^{2n+1}), \;\; \pi_0 = p_{\mathrm{data}}\}.$$

- This is akin to **alternative projection** in a Euclidean setting.
- $\lim_{n\to+\infty} \pi^n = \pi^\star$ under regularity conditions.



Solving the Schrödinger Bridge

- **Explicit solution** of the first IPF step

$$\mathrm{KL}(\pi|\pi^0) = \mathrm{KL}(\pi_N|p_N) + \mathbb{E}_{\pi_N}[\mathrm{KL}(\pi_{|N}|p_{|N})].$$

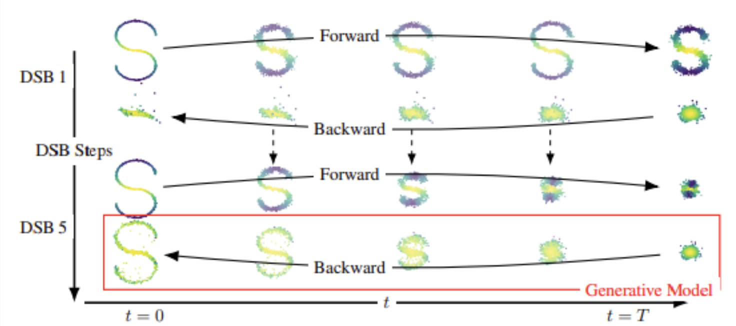
Therefore,

$$\pi^1(x_{0:N}) = p_{\text{prior}}(x_N)p(x_{0:N-1}|x_N)$$

$$\boxed{\pi^1(x_{0:N}) = p_{\text{prior}}(x_N)\prod_{k=0}^{N-1} p_{k|k+1}(x_k|x_{k+1}).}$$

- **Take-home message:** Approximation to first iteration of IPF corresponds to current **denoising diffusion models**.
- The IPF is a **refinement** on denoising diffusion models.

- **Diffusion Schrödinger Bridge**[3]:
  - Use **diffusion models** to solve IPF at each step.
  - Alternate between updating the **forward** and **backward dynamics**.
  - (One network parameterizing the forward, one parameterizing the backward).



[3] **D.B.**, Thornton, Heng, Doucet – Diffusion Schrödinger Bridge – NeurIPS 2021

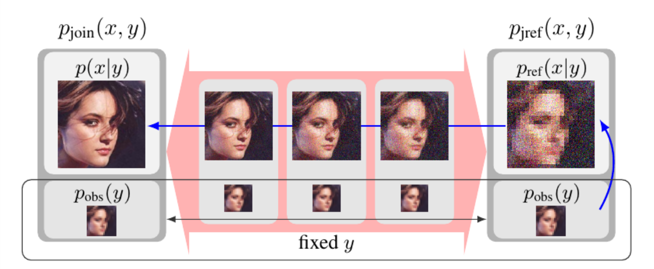
Inverse problems

- Extension to **conditional** generative modeling (inverse problems):
  - Inpainting, deblurring, colorization, class-conditional sampling...
- **Amortization** of generative models (w.r.t. the **observation**):
  - Denoising diffusion models.
  - **Schrödinger bridges**[4]

$$\pi^\star = \arg\min_\pi \{\mathrm{KL}(\pi|p) \ : \ \pi_0 = p(\cdot|y_{\mathrm{obs}}), \ \pi_N = p_{\mathrm{ref}}\}$$

$$\bar{\pi}^\star = \mathrm{argmin}_{\bar{\pi}} \{\mathrm{KL}(\bar{\pi}|\bar{p}) \ : \ \bar{\pi}_0 = p_{\mathrm{join}}, \ \bar{\pi}_N = p_{\mathrm{jref}}\}.$$

$\pi^\star$ is a conditional version of $\bar{\pi}^\star$



[4] Shi, **D.B.**, Deligiannidis, Doucet – Conditional Diffusion Schrödinger Bridge – UAI 2022

Conclusion

Conclusion

- Fruitful interaction between **stochastic processes** and **generative modeling**.
- For specific problems: use the structure of the likelihood Kawar et al. (2022) (DDRM), Kadkhodaie and Simoncelli (2021) (SNIPS), Come-closer-Diffuse-faster Chung et al. (2022a,b).
- Promising developments of **control** and **optimal transport** techniques for generative models (and vice-versa).

"Thank you" generated with the text-to-prompt model Stable diffusion.

References

Victoria A Avanzato, Kasopefoluwa Y Oguntuyo, Marina Escalera-Zamudio, Bernardo Gutierrez, Michael Golden, Sergei L Kosakovsky Pond, Rhys Pryce, Thomas S Walter, Jeffrey Seow, Katie J Doores, et al. A structural basis for antibody-mediated neutralization of nipah virus reveals a site of vulnerability at the fusion glycoprotein apex. *Proceedings of the National Academy of Sciences*, 116(50):25057–25067, 2019.

Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *arXiv preprint arXiv:2206.00941*, 2022a.

Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12413–12422, 2022b.

References ii

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34, 2021.

Robert Fortet. Résolution d'un système d'équations de M. Schrödinger. *Journal de Mathématiques Pures et Appliqués*, 1:83–105, 1940.

Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.

Zahra Kadkhodaie and Eero Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. *Advances in Neural Information Processing Systems*, 34:13242–13254, 2021.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *arXiv preprint arXiv:2201.11793*, 2022.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11461–11471, 2022.

Toshio Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probability theory and related fields*, 129(2):245–260, 2004.

Andrew W Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander WR Nelson, Alex Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020.

Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32, 2019.

Brian L Trippe, Jason Yim, Doug Tischer, Tamara Broderick, David Baker, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. *arXiv preprint arXiv:2112.07804*, 2021.