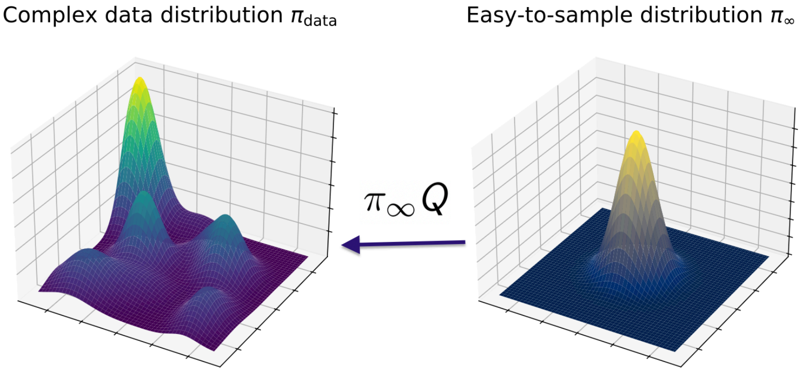
Conditional Sampling with Score-Based Generative Models

Stanislas Strasman, Antonio Ocello, Claire Boyer, Sylvain Le Corff, Vincent Lemaire

SCIENCES SORBONNE UNIVERSITÉ

Generative modeling framework

- $\mathcal{D} = \{u_i\}_{i=1}^n \in (\mathbb{R}^d)^n$ a collection of i.i.d. samples from an unknown distribution π_{data}.
- Goal: **generate new samples from** π_{data} (i.e. find a proba π_∞ and a simulable kernel Q such that $\pi_{\mathrm{data}} \simeq \pi_\infty Q$).

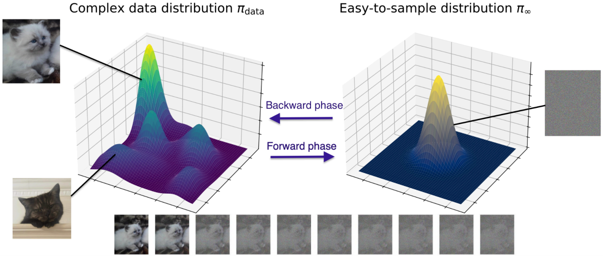


SGMs Philosophy - Forward process

- ⚠ The other way around is easy ($\pi_{\mathrm{data}} \simeq Q'\pi_\infty$)

$$\mathrm{d}\overrightarrow{\mathbf{U}}_t = -\overrightarrow{\mathbf{U}}_t \mathrm{d}t + \sqrt{2}\mathrm{d}B_t\,, \quad \mathbf{U}_0 \sim \pi_{\mathrm{data}}\,. \tag{1}$$

- By the ergodicity of the O.-U. process, the marginal p_T converges to $\mathcal{N}(0, \mathbf{I}_d)$ as $T \to \infty$.



- "Creating noise from data is easy; creating data from noise is generative modeling." (Song et al., 2021)

Time-reversal and the backward process

- Under mild conditions (1) admits a **time-reversed process** (Anderson, 1982), i.e. in law,

$$\left(\overleftarrow{\mathbf{U}}_t\right)_{t\in[0,T]} = \left(\overrightarrow{\mathbf{U}}_{T-t}\right)_{t\in[0,T]} .$$

- The reverse-time process $\left(\overleftarrow{\mathbf{U}}_t\right)_{t\in[0,T]}$ is solution to

$$\mathrm{d}\overleftarrow{\mathbf{U}}_t = \left(\overleftarrow{\mathbf{U}}_t + 2\underbrace{\nabla \log p_{T-t}\left(\overleftarrow{\mathbf{U}}_t\right)}_{\textbf{score function}}\right)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t\,, \quad \overleftarrow{\mathbf{U}}_0 \sim p_T,$$

with p_T the p.d.f. of (1).
- Sampling from the backward SDE yields a **generative model**

$$\overleftarrow{\mathbf{U}}_T \sim \pi_{\mathrm{data}}.$$

Learning the score is as easy as denoising...

- ⚠ How to train $s_\theta : [0, T] \times \mathbb{R}^d \mapsto \mathbb{R}^d$ to learn $\nabla \log p_t\left(\overrightarrow{\mathbf{U}}_t\right)$ when $p_t(x)$ is **unknown** ?
- 💡 **Conditional score matching** (Vincent, 2011):

$$\mathcal{L}_{\mathrm{score}}(\theta) = \mathbb{E}\left[\|s_\theta\left(\tau, \overrightarrow{\mathbf{U}}_\tau\right) - \nabla \log p_\tau\left(\overrightarrow{\mathbf{U}}_\tau | \overrightarrow{\mathbf{U}}_0\right)\|^2\right] ,$$

with $\tau \sim \mathcal{U}(0, T)$ independent of the forward process $(\overrightarrow{\mathbf{U}}_t)_{t\geq 0}$.

- Training target is **explicit**:

$$\nabla \log \pi_\tau(\overrightarrow{\mathbf{U}}_\tau | \overrightarrow{\mathbf{U}}_0) = \frac{m_\tau \overrightarrow{\mathbf{U}}_0 - \overrightarrow{\mathbf{U}}_\tau}{\sigma_\tau^2} = -\frac{Z}{\sigma_\tau} ,$$

with $Z \sim \mathcal{N}(0, I_d)$ and $Z \perp \overrightarrow{\mathbf{U}}_0$.

but denoising is not so cheap...

- **Tweedie's formula** (Gaussian denoising): if $\overrightarrow{\mathbf{U}}_t = \mathbf{U}_0 + \sqrt{2}Z$, then the MMSE estimator of $\mathbf{U}_0$ given $\overrightarrow{\mathbf{U}}_t$ is

$$\hat{\mathbf{U}}_0 = \overrightarrow{\mathbf{U}}_t + 2\nabla \log p_t(\overrightarrow{\mathbf{U}}_t) .$$

- In practice, training high-quality score models requires:
  - Large-scale datasets (e.g., ImageNet, Celeb-A, CIFAR-10),
  - High-capacity architectures (e.g., U-Nets with attention),
  - **Extensive compute**: tens or hundreds of thousands of GPU hours.
- Stable Diffusion v1 :
  - training consumed 150,000 A100 GPU-hours,
  - estimated cost of $\sim$ \$600,000,
  - 860 million parameters.

Results are breathtaking...

Conditional sampling: the example of inpainting

- In many applications (*e.g.*, inpainting), one want to sample from a **conditional distribution**.
- Let $\mathbf{U} = (X, Y) \in \mathbb{R}^d$, where:
  - $Y \in \mathbb{R}^{d_y}$ is observed,
  - $X \in \mathbb{R}^{d_x}$ is missing.
- Let $M \in \{0, 1\}^d$ be a binary mask: $M_i = 1$ if the i-th component is observed.
- The goal is to reconstruct the full signal $\mathbf{U} = (X, Y)$ given Y, i.e.,

$$\hat{\mathbf{U}} = (\hat{X}, Y)\,.$$

Option 1: conditional training

- Score-based models can handle this via **specific training methods** (*e.g.* incorporate masking information M) to get an approximation of the conditional score function $\nabla \log p_t(U_t \mid Y, M)$.
- Requires **additional training cost**.
- Generalization to arbitrary masks is not guaranteed unless explicitly trained for them.
- What if one only have access to **unconditional** score models?

Option 2: Constrained sampling with unconditional score

- In practice, backward sampling is **sequential** (Euler-Maruyama) with $\Delta = T/N$ and $0 = t_0 < t_1 < \ldots < t_N = T$:

$$p^{\theta}_{0:T}(x_{0:T}, y_{0:T}) = p_{\infty}(x_0, y_0) \prod_{i=1}^{N} \bar{p}_{\theta, t_i | t_{i-1}} \left(x_{t_i}, y_{t_i} \middle| x_{t_{i-1}}, y_{t_{i-1}} \right) \,,$$

with

$$\bar{p}_{\theta, t_k | t_{k-1}} \left(x_{t_k}, y_{t_k} \middle| x_{t_{k-1}}, y_{t_{k-1}} \right) := \mathcal{N}\left(x_{t_k}, y_{t_k}; \bar{\mu}_{k-1}, 2\Delta I_d \right) \,,$$

$$\bar{\mu}_{k-1} = 2\Delta \left\{ \begin{pmatrix} \bar{x}_{t_{k-1}} \\ \bar{y}_{t_{k-1}} \end{pmatrix} + s_{\theta} \left(T - t_{k-1}, \begin{pmatrix} \bar{x}_{t_{k-1}} \\ \bar{y}_{t_{k-1}} \end{pmatrix} \right) \right\} \,.$$

- But we have noisy samples from the observed parts $y_{0:T}$ can we use them to drive the flow towards

$$\overleftarrow{X}_T | \overleftarrow{Y}_T \sim X | Y \,?$$

Option 2: Constrained Sampling with unconditional score II

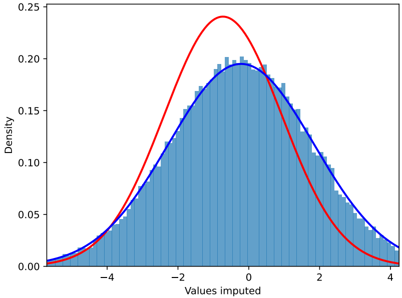
- Zhang et al. (2025) propose a plug-and-play method: run the reverse diffusion discretization, but at each step, **overwrite the known pixels** using:

$$X_{t_k}^{\mathrm{input}} \leftarrow M \odot \overrightarrow{Y}_{t_k} + (1 - M) \odot \bar{X}_{t_k} \,.$$

- **No retraining** is required.
- But **no theoretical guarantees**.
- For Gaussian targets sampling is biased...

Option 2: Constrained Sampling with unconditional score III

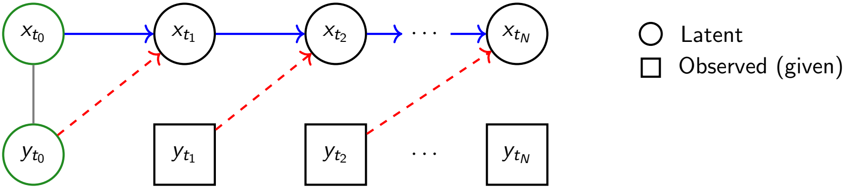
- $\begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1.5 \\ 1.5 & 5 \end{pmatrix} \right)$.
- Exact solution is Gaussian (red line).
- Theoretical and empirical inputation are biased (blue).



SMC and diffusion

- Conditional on $(x_{t_{k-1}}, y_{t_{k-1}})$, x_{t_k} and y_{t_k} are independent.

$$
\begin{aligned}
&p^{\theta}_{0:t_N}(x_{0:T}, y_{0:T}) \\
&= \underbrace{p_{\infty}\left(x_0, y_0\right)}_{\text{Initial sampling}} \underbrace{\bar{p}_{\theta,t_1|t_0}\left(y_{t_1} \middle| y_0, x_0\right)}_{\text{Observation likelihood}} \\
&\quad \prod_{k=1}^{N-1} \underbrace{\bar{p}_{\theta,t_{k+1}|t_k}\left(y_{t_{k+1}} \middle| y_{t_k}, x_{t_k}\right)}_{\text{Observation likelihood}} \underbrace{\bar{p}_{\theta,t_k|t_{k-1}}\left(x_{t_k} \middle| y_{t_{k-1}}, x_{t_{k-1}}\right)}_{\text{Propagation sampling}} \\
&\quad \underbrace{\bar{p}_{\theta,T|t_{N-1}}\left(x_T \middle| y_{t_{N-1}}, x_{t_{N-1}}\right)}_{\text{Propagation sampling}} .
\end{aligned}
$$



SMC sampling Algorithm.

- **Initialization:** For $i = 1, \ldots, M$, sample $\tilde{x}_0^{(i)} \sim p_\infty(\cdot)$ and sample and store a forward trajectory $y_{T:0} \sim \overrightarrow{p}(y_{T:0})$.
- **For each time step $k = 1, \ldots, N$:**
  - Compute the weights $w_k^{(i)} \propto \bar{p}_{\theta, t_k | t_{k-1}}(y_{t_k} \mid \tilde{x}_{t_{k-1}}^{(i)}, y_{t_{k-1}})$.
  - Normalize the weights and resample the particles $\{\tilde{x}_{t_{k-1}}^{(i)}\}$ according to $\{w_k^{(i)}\}$.
  - Propagate each particle i by sampling:

$$\tilde{x}_{t_k}^{(i)} \sim \bar{p}_{\theta, t_k | t_{k-1}}(\cdot \mid \tilde{x}_{t_{k-1}}^{(i)}, y_{t_{k-1}}).$$

- **Output:**

$$\sum_{i=1}^{M} w_T^{(i)} \delta_{\tilde{x}_T^{(i)}} .$$

This has proven to be effective empirically

Figure: Figure 16 from Cardoso et al. (2024)

Theoretical convergence result

For some function h bounded measurable,

$$
\begin{aligned}
&\left\| \mathbb{E}\left[h(X_T) \mid Y_{0:T}\right] - \sum_{i=1}^{M} w_T^{(i)} \delta_{\tilde{x}_T^{(i)}} \right\| \\
&\leq \underbrace{\left\| \mathbb{E}\left[h(X_T) \mid Y_{0:T}\right] - \mathbb{E}\left[h(\bar{X}_T^{\theta}) \mid Y_{0:T}\right] \right\|}_{\text{SGM bias}} \\
&+ \underbrace{\left\| \mathbb{E}\left[h(\bar{X}_T^{\theta}) \mid Y_{0:T}\right] - \sum_{i=1}^{M} w_T^{(i)} \delta_{\tilde{x}_T^{(i)}} \right\|}_{\text{SMC error}}
\end{aligned}
$$

where $\bar{X}_T^{\theta}$ is the parametric approximation of X_T and $\bar{X}_T^{\theta,M}$ is its Monte Carlo approximation using M particles. The first term encompasses the three standard SGM errors (Strasman et al., 2025). The second term comes from the Monte Carlo approximation.

SMC error

The Monte Carlo error is upper bounded in works from Cardoso et al. (2024); Wu et al. (2024) by

$$\left\| \mathbb{E}\left[h(\bar{X}_T^\theta) \mid Y_{0:T} \right] - \sum_{i=1}^{M} w_T^{(i)} \delta_{\tilde{x}_T^{(i)}} \right\| \leq \frac{C_T}{\sqrt{M}} .$$

SGM error bias

H1 There exists $C > 0$ such that for all $h > 0$, $0 \leq k \leq n-1$, and all $x_{t_k}, y_{t_{k+1}}, y_{t_k} \in \mathbb{R}^{d_\mathrm{x}} \times \mathbb{R}^{d_\mathrm{y}} \times \mathbb{R}^{d_\mathrm{y}}$ and all bounded and measurable functions ϕ,

$$\Big\| \mathbb{E}\left[\phi(X_{t_{k+1}}) \mid X_{t_k} = x_{t_k}, Y_{t_k} = y_{t_k}, Y_{t_{k+1}} = y_{t_{k+1}}\right] - \mathbb{E}\left[\phi(\bar{X}^\theta_{t_{k+1}}) \mid X_{t_k} = x_{t_k}, Y_{t_k} = y_{t_k}, Y_{t_{k+1}} = y_{t_{k+1}}\right] \Big\| \leq hC \left\|\phi\right\|_\infty .$$

H2 $U \in L^2(\Omega)$.

Then we have that, there exists $C_1, C_2 > 0$ such that,

$$\left\| \mathbb{E}\left[h(X_T) \mid Y_{0:T}\right] - \mathbb{E}\left[h(\bar{X}^\theta_T) \mid Y_{0:T}\right] \right\| \leq \left(\mathrm{e}^{-T} C_1 \left\|\mathbf{U}\right\|_{L^2} + C_2 T\right) \left\|\phi\right\|_\infty$$

B. D. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

G. Cardoso, Y. J. el idrissi, S. L. Corff, and E. Moulines. Monte carlo guided denoising diffusion models for bayesian linear inverse problems. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=nHESwXvxWK.

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations (ICLR)*, 2021.

S. Strasman, A. Ocello, C. Boyer, S. L. Corff, and V. Lemaire. An analysis of the noise schedule for score-based generative models. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=BlYIPa0Fx1.

P. Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011. doi: 10.1162/NECO_a_00142.

L. Wu, B. L. Trippe, C. A. Naesseth, D. M. Blei, and J. P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models, 2024. URL https://arxiv.org/abs/2306.17775.

H. Zhang, L. Fang, Q. Wu, and P. S. Yu. Diffputer: An EM-driven diffusion model for missing data imputation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=3fl1SENSYO.