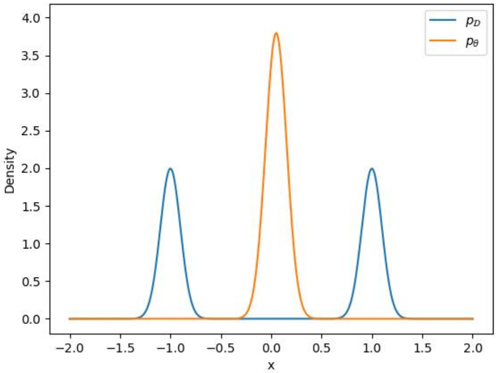
Generative models: Support slides.

Gabriel V. Cardoso.

March 03, 2026

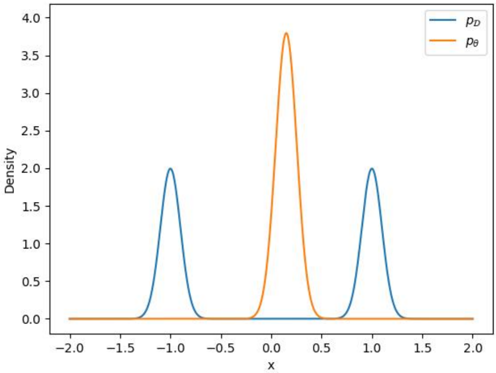
Learning V1



$$p_{\mathcal{D}} = \frac{1}{2}\mathcal{N}(-1,1) + \frac{1}{2}\mathcal{N}(1,1), \quad p_\theta = \mathcal{N}(\mu, \sigma^2)$$

$$\theta \in \arg\min \sum_{i=1}^{n} \log p_\theta\,(X_i), \quad X_i \sim p_{\mathcal{D}}\ .$$

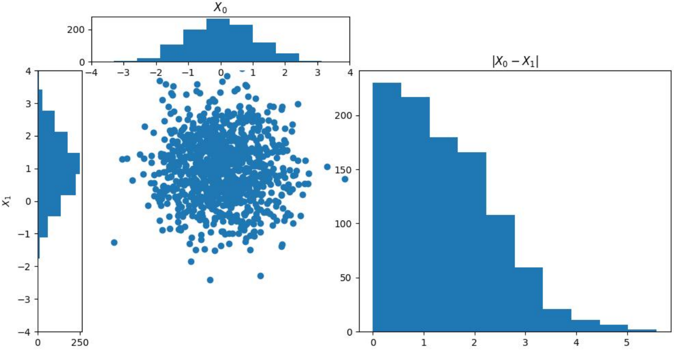
Learning V2



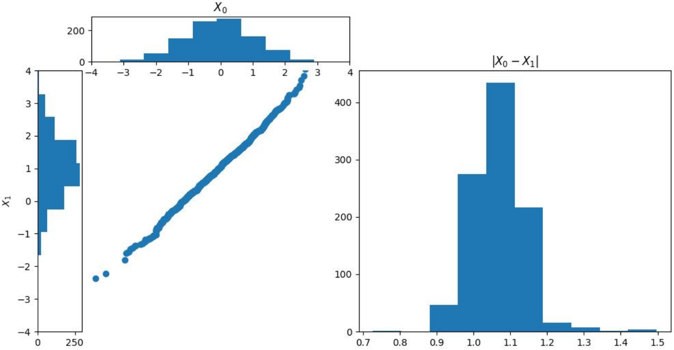
$$p_{\mathcal{D}} = \frac{1}{2}\mathcal{N}(-1,1) + \frac{1}{2}\mathcal{N}(1,1), \quad p_\theta = \mathcal{N}(\mu, \sigma^2)$$

$$\theta \in \arg\min \sum_{i=1}^{n} \log p_{\mathcal{D}}(X_i), \quad X_i \sim p_\theta \ .$$

Coupling $\mathcal{N}(0,1), \mathcal{N}(1,1)$: Independent



Coupling $\mathcal{N}(0,1), \mathcal{N}(1,1)$: Earth Mover distance



Deep Gauss VAE: Loss

Deep Gauss framework:

- $p_\theta(x|z) = \mathcal{N}(x; \mu_\theta(z), \mathrm{Diag}(\exp(\ln \sigma^2_\theta(z))))$,
- $q_\phi(z|x) = \mathcal{N}\left(z; \mu_\phi(x), \mathrm{Diag}\left(\exp(\ln \sigma^2_\phi(x))\right)\right)$,

where the neural networks are such that $\mu_\theta, \ln \sigma^2_\theta : \mathbb{R}^{d_z} \to \mathbb{R}^{d_x}$ and $\mu_\phi, \ln \sigma^2_\phi : \mathbb{R}^{d_x} \to \mathbb{R}^{d_z}$.

$$\mathrm{L}^{mc}_{\mathrm{ELBO}}(\theta, \phi, x) = n_{mc}^{-1} \sum_{i=1}^{n_{mc}} - \log p_\theta\left(x \middle| \mu_\phi(x) + \mathrm{Diag}\left(\exp(\ln \sigma^2_\phi(x))\right) Z_i\right) \\ + \mathrm{D}_{\mathrm{KL}}\left(q_\phi(\cdot|x) \,||\, \mathcal{N}(0, \mathrm{I})\right)$$

with $Z_i \sim \mathcal{N}(0, \mathrm{I})$. And

$$(\theta^\star_n, \phi^\star_n) \in \arg\min_{\theta,\phi} n^{-1} \sum_{i=0}^{n} \mathrm{L}^{mc}_{\mathrm{ELBO}}(\theta, \phi, X_i)$$

Naive Deep Gauss VAE simulation

- $Z_i \sim \mathcal{N}(0, \mathrm{I})$,
- $X_i^\theta \sim p_\theta(\cdot|Z_i)$.

Score and Denoising: [Vin11, 2011]

If $X_0 \sim p_{\mathcal{D}}$ and $X_\sigma = X_0 + \sigma Z$ where $Z \sim \mathcal{N}(0, \mathrm{I})$ and $Z \perp\!\!\!\perp X_0$

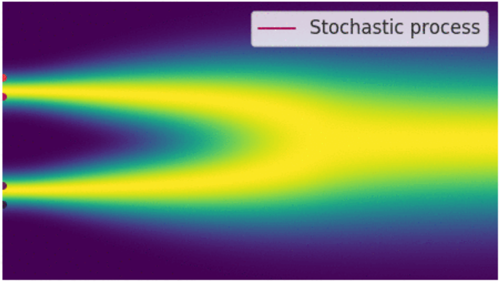
$$\sigma^2 \nabla \log p_\sigma(x) + x = \mathbb{E}\left[X_0 \,|\, X_\sigma = x \right] \,, \tag{1}$$

where p_σ is the density of X_σ.

SDE framework

Forward SDE:

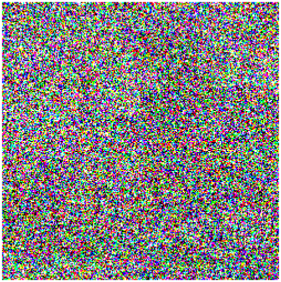
$$\mathrm{d}X_t = -\beta_t X_t \mathrm{d}t + \sqrt{2\beta_t}\mathrm{d}B_t \ ,$$

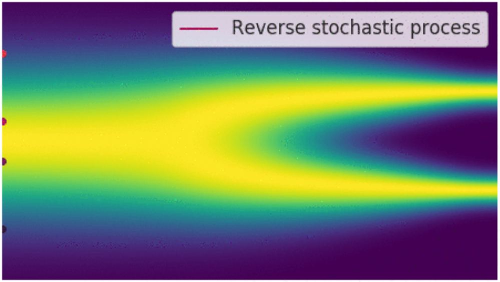


SDE Framework

Backward SDE:

$$\mathrm{d}\overleftarrow{X}_t = (\beta_{T-t}X_t + \beta_{T-t}\nabla \log p_{T-t}\left(\overleftarrow{X}_t\right))\mathrm{d}t + \sqrt{2\beta_{T-t}}\mathrm{d}\tilde{B}_t$$





Denoising Loss

Denoiser network

$$\mathrm{D}_\theta : \mathbb{R}^{d_x} \times \mathbb{R}_+ \ni (x, \sigma) \to \mathrm{D}_\theta(x, \sigma) \in \mathbb{R}^{d_x} \tag{2}$$

Trained on the objective

$$\mathrm{L}_{\text{denoiser}} = \mathbb{E}_{\sigma \sim \lambda_\sigma} \left[\frac{\gamma(\sigma)}{\sigma^2} \mathbb{E} \left[\| X_0 - \mathrm{D}_\theta(X_0 + \sigma Z, \sigma) \|^2 \right] \right] , \tag{3}$$

where λ_σ is an importance distribution (typically $\log \mathcal{N}(-1, 1)$) and $\gamma(\sigma)$ a weighting function (typically $\gamma(\sigma) = 1$) which can be tuned on the dataset.

Score-based sampling

Parameters: T, h. K = T/h

- Initialization: $\overleftarrow{X}_0 \sim \mathcal{N}(0, \mathrm{I})$.
- Inner loop: For $k \in \{1, \cdots, K\}$:
  - Draw $Z_k \sim \mathcal{N}(0, \mathrm{I})$
  - Set $\overleftarrow{X}_{kh} = \beta_{T-(k-1)h}(\overleftarrow{X}_{(k-1)h} + \nabla \log p_{T-t}\left(\overleftarrow{X}_{(k-1)h}\right))h + \sqrt{2h\beta_{T-(k-1)h}}Z_k$

References: Basic theoretical tools

- Basic functional analysis: [Rud87, 1987]
- High dimensional probability: [Ver18, 2018]
- Stochastic processes: [LS01, 2001]
- Information theory: [CT06, 2006]
- Optimal transport: [Vil21, 2021]

References: Normalizing flows

- Original paper: [RM15, 2015]
- RealNVP: [DSB17, 2017]
- Other famous Flow (Glow) [KD18, 2018]
- Cool application [Sam+22, 2022]

VAE

- Original paper: [KW13, 2013]
- Tutorial: [KW+19, 2019]
- Theoretical analysis (estimation error): [TY21, 2021]

Score-based generative models

- Paper that layed out modern foundations: [Son+21b, 2021]
- Paper about training models: [Kar+22, 2022; Kar+24, 2024]
- Theoretical analysis: [Son+21a, 2021; CDS24, 2024]
- Surprising properties of SBGM: [Kad+23, 2023; Zha+24, 2024]

GANs

- Original paper: [Goo+14, 2014]
- Modern way: [ACB17, 2017]
- Training architectures [Kar+20, 2020]

References I

[ACB17] Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks". In: *International conference on machine learning*. PMLR. 2017, pp. 214–223.

[CDS24] Giovanni Conforti, Alain Durmus, and Marta Gentiloni Silveri. *KL Convergence Guarantees for Score Diffusion Models under Minimal Data Assumptions*. Sept. 12, 2024. arXiv: `2308.12240 [math, stat]`. URL: `http://arxiv.org/abs/2308.12240` (visited on 09/17/2024). Pre-published.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. 2nd ed. Hoboken, N.J: Wiley-Interscience, 2006.

References II

[DSB17] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. "Density Estimation Using Real NVP". In: *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. 2017.

[Goo+14] Ian J. Goodfellow et al. "Generative Adversarial Nets". In: *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. NIPS'14. Montreal, Canada: MIT Press, 2014, pp. 2672–2680.

[Kad+23] Zahra Kadkhodaie et al. "Generalization in diffusion models arises from geometry-adaptive harmonic representation". en. In: *The Twelfth International Conference on Learning Representations*. 2023.

[Kar+20] Tero Karras et al. *Analyzing and Improving the Image Quality of StyleGAN*. Mar. 2020. arXiv: `1912.04958 [cs]`.

References III

[Kar+22] Tero Karras et al. “Elucidating the Design Space of Diffusion-Based Generative Models”. In: *Proc. NeurIPS*. 2022.

[Kar+24] Tero Karras et al. *Analyzing and Improving the Training Dynamics of Diffusion Models*. Mar. 2024. arXiv: 2312.02696.

[KD18] Durk P Kingma and Prafulla Dhariwal. “Glow: Generative Flow with Invertible 1x1 Convolutions”. In: *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc., 2018.

[KW+19] Diederik P Kingma, Max Welling, et al. “An introduction to variational autoencoders”. In: *Foundations and Trends® in Machine Learning* 12.4 (2019), pp. 307–392.

[KW13] Diederik P. Kingma and Max Welling. “Auto-Encoding Variational Bayes”. In: (Dec. 23, 2013).

[LS01] R. S. Lipster and A. N. Shiryaev. *Statistics of Random Processes: I. General Theory*. 2nd. Springer, 2001.

References IV

[RM15] Danilo Rezende and Shakir Mohamed. “Variational Inference with Normalizing Flows”. In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. PMLR. Lille, France: PMLR, July 2015, pp. 1530–1538.

[Rud87] W. Rudin. *Real and Complex Analysis*. Singapore: McGraw-Hill, 1987.

[Sam+22] Sergey Samsonov et al. “Local-Global MCMC Kernels: The Best of Both Worlds”. In: *Advances in Neural Information Processing Systems* 35 (Dec. 2022), pp. 5178–5193.

[Son+21a] Yang Song et al. “Maximum likelihood training of score-based diffusion models”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 1415–1428.

References V

[Son+21b] Yang Song et al. “Score-Based Generative Modeling through Stochastic Differential Equations”. In: *International Conference on Learning Representations*. 2021.

[TY21] Rong Tang and Yun Yang. “On Empirical Bayes Variational Autoencoder: An Excess Risk Bound”. In: *Proceedings of Thirty Fourth Conference on Learning Theory*. PMLR, July 2021, pp. 4068–4125.

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge university press, 2018.

[Vil21] Cédric Villani. *Topics in optimal transportation*. Vol. 58. American Mathematical Soc., 2021.

[Vin11] Pascal Vincent. “A Connection Between Score Matching and Denoising Autoencoders”. In: *Neural Computation* 23.7 (2011), pp. 1661–1674.

References VI

[Zha+24] Huijie Zhang et al. "The Emergence of Reproducibility and Consistency in Diffusion Models". In: *Proceedings of the 41st International Conference on Machine Learning*. PMLR, July 2024, pp. 60558–60590.