Diffusion models for Gaussian distributions: Exact solutions and Wasserstein errors

Emile Pierret
emile.pierret@univ-orleans.fr

Bruno Galerne
bruno.galerne@univ-orleans.fr

ICML 2025

We study continuous diffusion models applied to Gaussian data.

- We derive exact solutions to the backward equations (Proposition 1).
- We theoretically study the initialization error (Proposition 2).
- We compute exact Wasserstein errors over time.
- We present an empirical observation of the score approximation error.

Forward SDE [1]

$$dx_t = -\beta_t x_t dt + \sqrt{2\beta_t} dw_t, \quad 0 \leq t \leq T, x_0 \sim p_0 .$$

$\hookrightarrow x_t = e^{-B_t} x_0 + \eta_t$ with $B_t = \int_0^t \beta_s ds$ and $\mathrm{Cov}(\eta_t) = (1 - e^{-2B_t})I$.

Gaussian assumption

The data distribution is Gaussian: $p_0 = \mathcal{N}(0,\Sigma)$.

$\Rightarrow \nabla \log p_t(x) = -\Sigma_t^{-1} x$ with $\Sigma_t = e^{-2B_t}\Sigma + (1 - e^{-2B_t})I.$

Proposition 1 (Exact solutions of the backward SDEs under Gaussian assumption)

Backward SDE

$$dy_t = \beta_{T-t}[y_t + 2\nabla \log p_{T-t}(y_t)]dt + \sqrt{2\beta_{T-t}} dw_t .$$

$\hookrightarrow$ Under the Gaussian assumption, the strong solution is

$$y_t = e^{-(B_T - B_{T-t})}\Sigma_{T-t}\Sigma_T^{-1} y_0 + \xi_t \text{ with } \mathrm{Cov}(\xi_t) = \Sigma_{T-t} - e^{-2(B_T - B_{T-t})}\Sigma_{T-t}^2\Sigma_T^{-1} .$$

Probability flow ODE

$$dy_t = \beta_{T-t}[y_t + \nabla \log p_{T-t}(y_t)]dt .$$

$\hookrightarrow$ Under the Gaussian assumption, the unique solution is

$$y_t = \Sigma_T^{-1/2}\Sigma_{T-t}^{1/2} y_0 .$$

Exact 2-Wasserstein errors:

Assuming that the two covariance matrices Σ_1, Σ_2 are simultaneously diagonalizable with respective eigenvalues $(\lambda_{i,1})_i$ and $(\lambda_{i,2})_i$,

$$W_2(\mathcal{N}(0,\Sigma_1), \mathcal{N}(0,\Sigma_2))^2 = \sum_i \left(\sqrt{\lambda_{i,1}} - \sqrt{\lambda_{i,2}}\right)^2 .$$

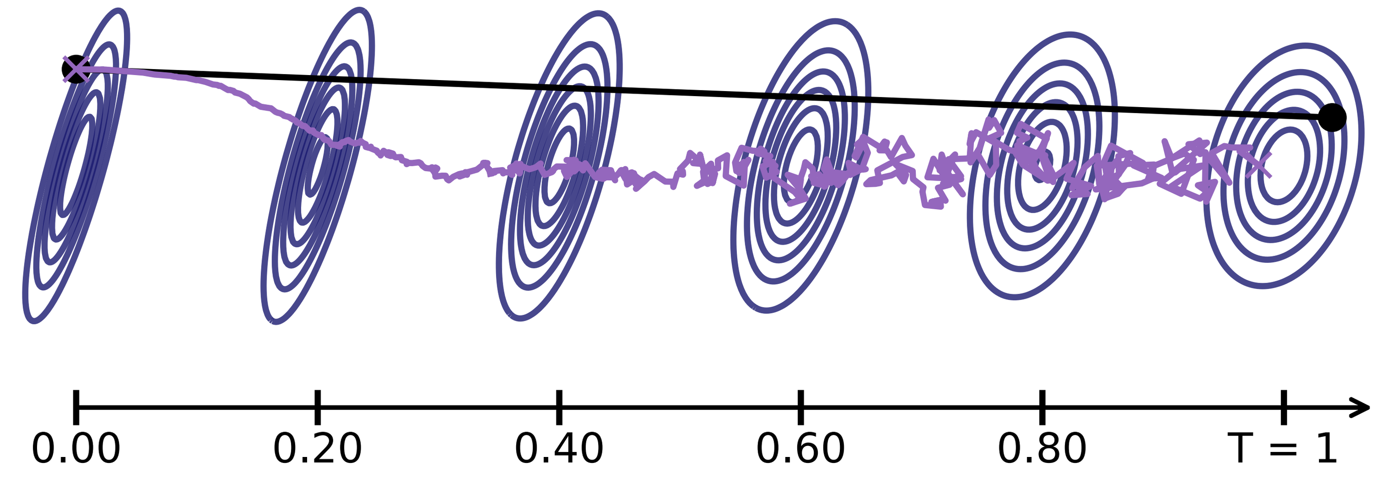
Proposition 2 (Study of the initialization error)

Under Gaussian assumption, denoting p_t^{SDE} (resp. p_t^{ODE}) the marginals of the backward SDE (resp. probability flow ODE) with initialization error

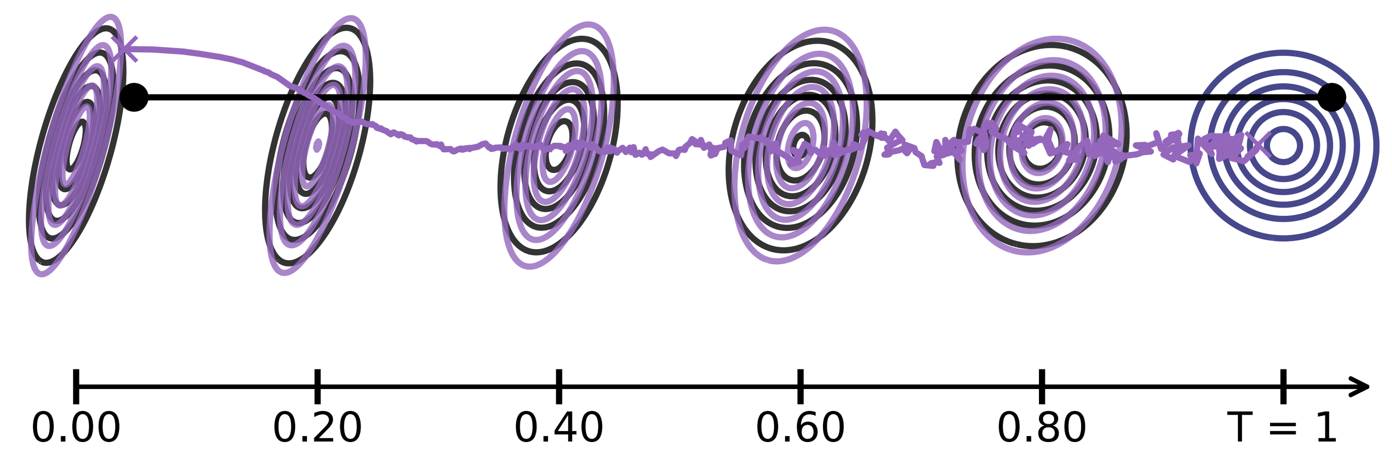
$$W_2(p_t^{\mathrm{SDE}}, p_t) \leq W_2(p_t^{\mathrm{ODE}}, p_t) .$$

Illustration of three error types

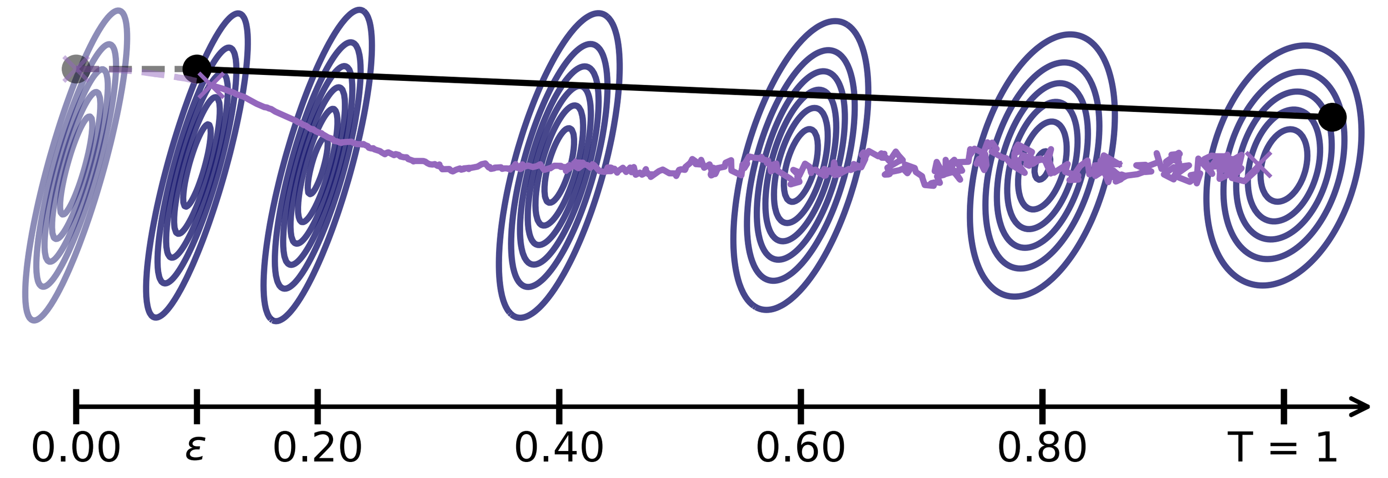
Theoretical backward process



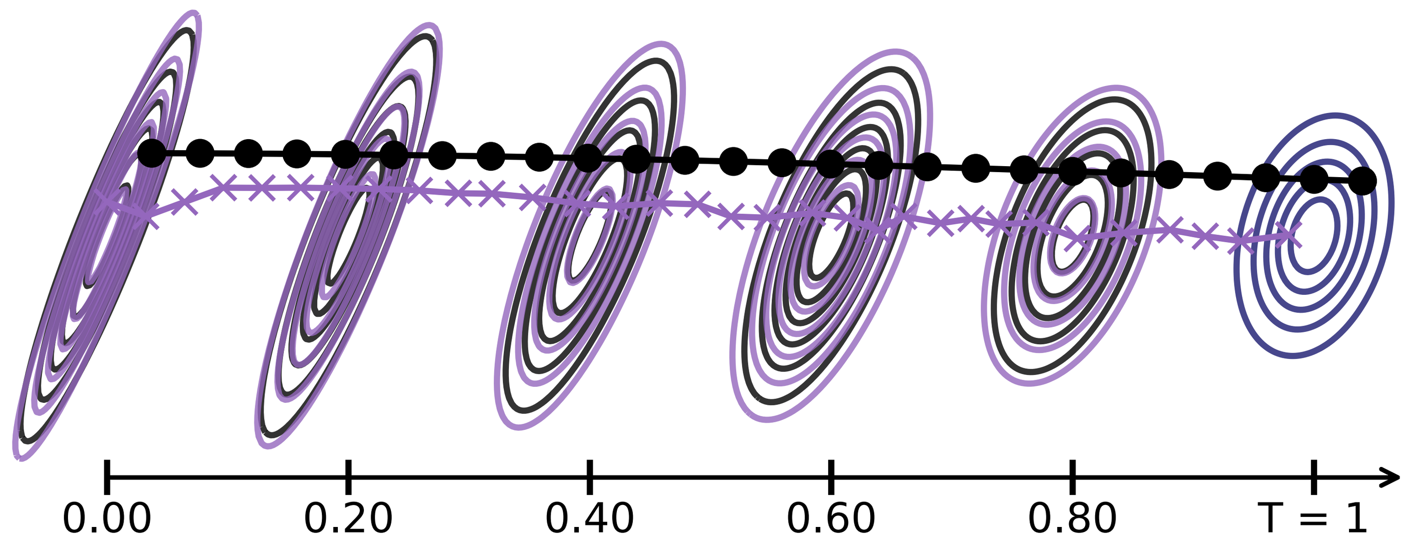
Initialization error



Truncation error



Discretization error



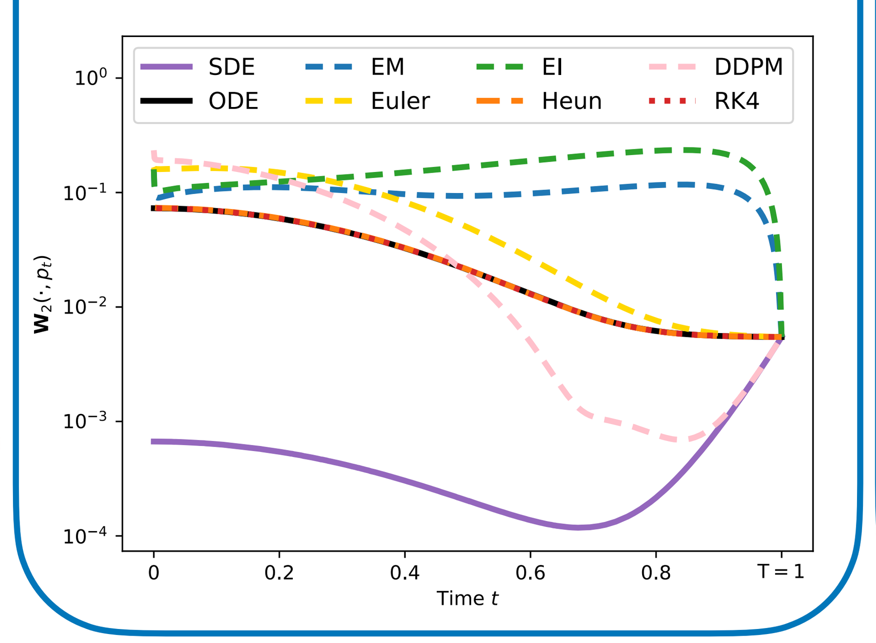
Exact Wasserstein errors for the truncation, the initialization and the discretization errors

- Under the Gaussian assumption, we compute exact Wasserstein errors with respect to different error types along the time.
- We illustrate these errors in the two graphs on the right, using the Gaussian fitted to the CIFAR-10 dataset.

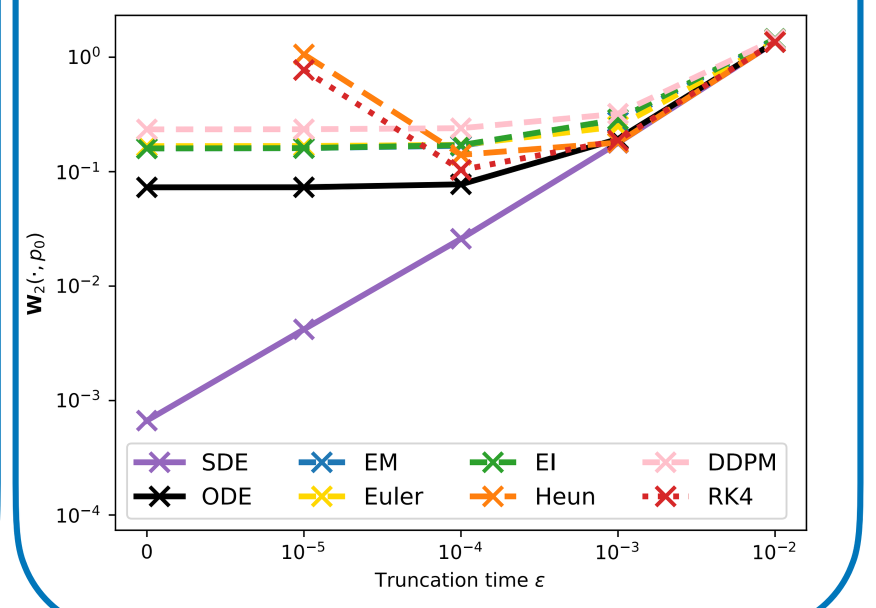
SDE discretization
- Euler Maruyama (EM)
- Exponential integrator (EI)
- Denoising Diffusion Probabilistic Model (DDPM)

ODE discretization
- Euler
- Heun
- Runge-Kutta 4 (RK4)

Discretization + initialization errors



Discretization + initialization + truncation errors

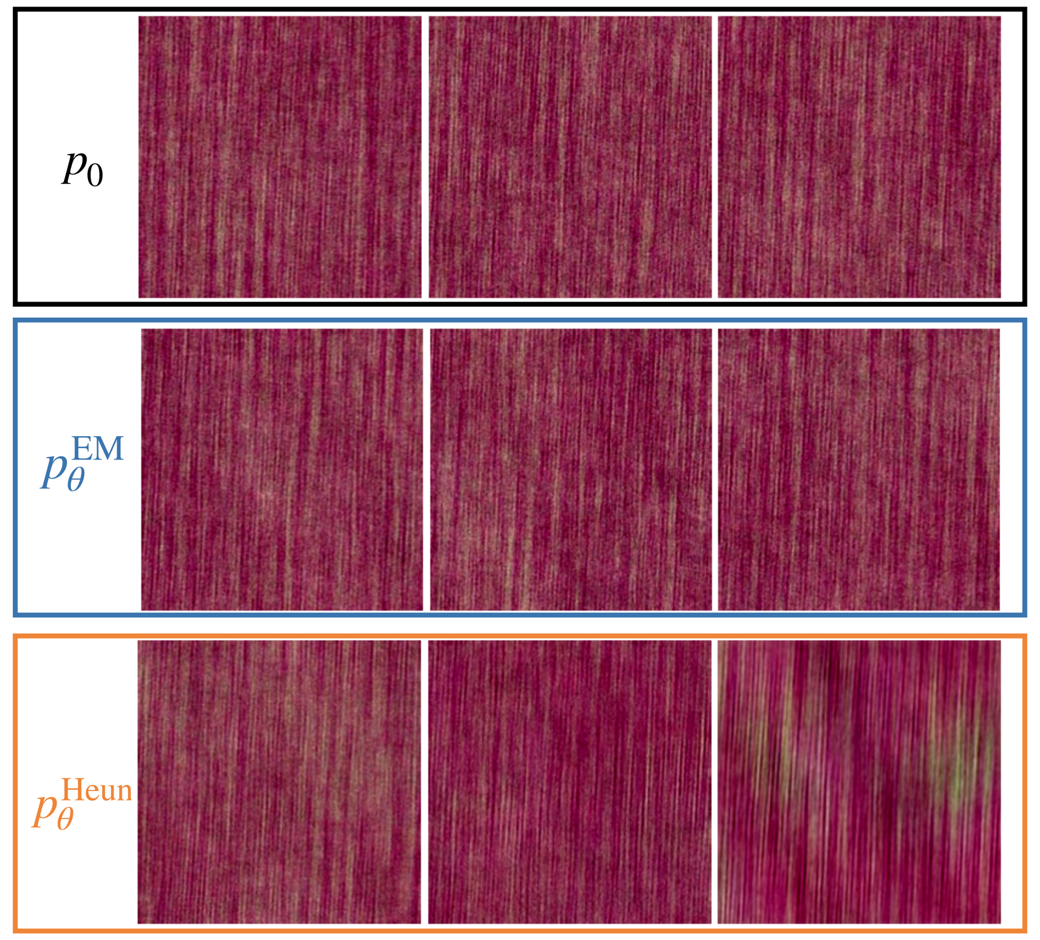


Study of the score approximation: the most crucial practical aspect

- We study the score approximation with a Gaussian distribution of microtextures to produce realistic samples [2].
- The backward equations are discretized with a score learned by a neural network (U-NET).
- We compute empirical Wasserstein errors.
- We provide a comparison based on the FID metric.
- We observe that Heun's scheme is more sensitive to score approximation errors than EM scheme.

	Exact score distribution			Learned score distribution	
p	$\mathbf{W}_2(p,p_{\mathrm{data}}) \downarrow$	$\mathbf{W}_2^{\mathrm{emp.}}(p^{\mathrm{emp.}},p_{\mathrm{data}}) \downarrow$	$\mathrm{FID}(p^{\mathrm{emp.}},p_{\mathrm{data}}^{\mathrm{emp.}}) \downarrow$	$\mathbf{W}_2^{\mathrm{emp.}}(p_\theta^{\mathrm{emp.}},p_{\mathrm{data}}^{\mathrm{emp.}}) \downarrow$	$\mathrm{FID}(p_\theta^{\mathrm{emp.}},p_{\mathrm{data}}^{\mathrm{emp.}}) \downarrow$
EM	5.16	5.16	0.089	15.6	1.02
Heun	3.73	3.73	0.045	56.7	19.4

Various generated samples



Conclusion

- The probability flow ODE is more affected by the initialization error than the backward SDE.
- The backward SDE is more affected by the discretization error than the probability flow ODE.
- Without score approximation, Heun's scheme is theoretically the go-to method, as already observed empirically [3].
- However, stochastic schemes are more resilient with respect to the score approximation error than deterministic ones.

References

[1] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, Ben Poole. **Score-Based Generative Modeling through SDEs.** *ICLR (2021).*

[2] Bruno Galerne, Yann Gousseau, Jean-Michel Morel. **Random Phase Textures: Theory and Synthesis.** *IEEE Transactions on Image Processing (2010).*.

[3] Tero Karras, Miika Aittala, Timo Aila, Samuli Laine. **Elucidating the Design Space of Diffusion-Based Generative Models.** *Neurips (2022).*