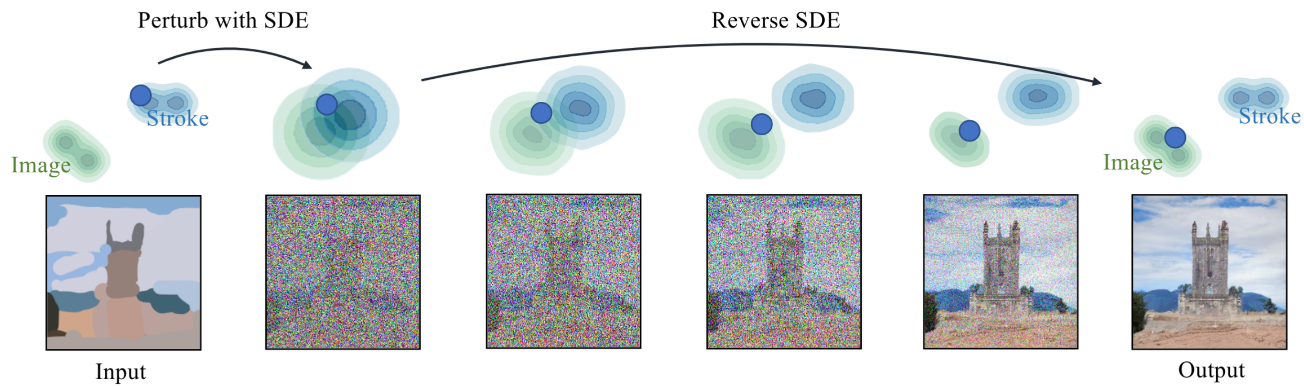
@mengSDEditGuidedImage2022

SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, Stefano Ermon
2022
http://arxiv.org/abs/2108.01073
Zotero
Tags:

- [x] ~~handwritten notes created~~



Notes

Theoretical foundation:
Denoising is possible because reverse SDE is well defined. For example, the forward Langevin of an Orstein-Uhlenbeck process is:

$$dx = -x dt + \sqrt{2}\, dW$$

we can convert it to Fokker Planck using @StochasticMethodsSpringerLink, chapter 3.8.4 (set $k = 1$ and $D = \sigma^2 = 2$):

$$\frac{\partial(p(x,t))}{\partial t} = \frac{\partial}{\partial x}[x\, p(x,t)] + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2} p(x,t)$$

solve to obtain:

$$x(t) = \mathbf{x(0)} e^{-t} + \mathbf{z}\sqrt{1 - e^{-2t}} \text{ where } \mathbf{z} \sim N(0,1)$$

The diffusion model community prefers to directly parametrize $x(t)$ as:

$$x(t) = \alpha(t)\mathbf{x(0)} + \sigma(t)\mathbf{z}$$

There are two common ways of injecting noise:

1. Orstein-Uhlenbeck noise (**variance-preserving**)

$$\alpha(t)^2 + \sigma(t)^2 = 1$$

If the data distribution $p(x, t=0)$ has unit variance, then the total variance of data + noise is preserved, using relation: $Var[X+Y] = Var[X] + Var[Y]$

2. $\alpha(t) = 1, \forall\, t$,and $\sigma(t)$ tends to some large but finite value (**variance-exploding**)

Example: Gaussian noise: $dx = dW$, with solution:

$$\alpha(t) = 1, \sigma(t) = \sqrt{t}$$

For the VE noise injecting scheme, one can define reverse SDE [1]:

$$dx(t) = d(\sigma(t)^2)\frac{1}{p}\nabla_x P(x,t)dt + \sqrt{d(\sigma(t)^2)}\,\mathbf{z}$$

In the gaussian case, this becomes:

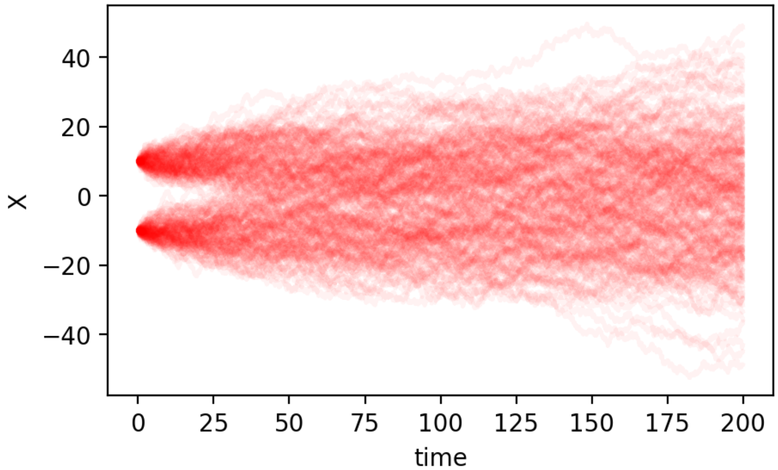
$$x(t) = x(t+dt) + dt\frac{1}{p}\nabla_x P(x,t) + \sqrt{dt}\,\mathbf{z}$$

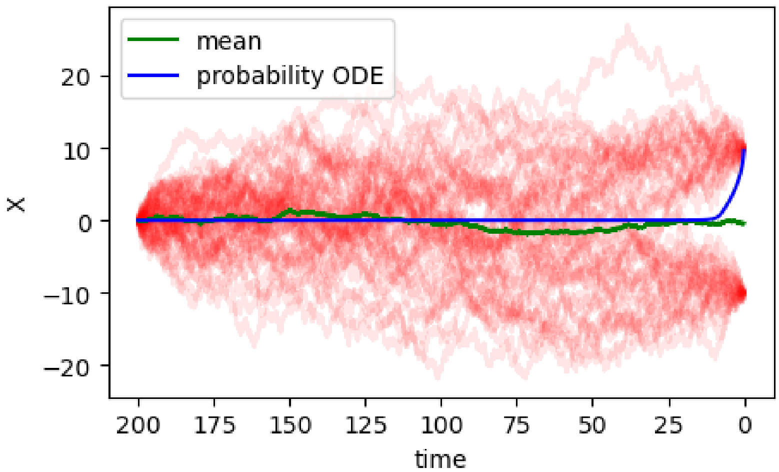
Which samples the same P(x, t) defined by the forward SDE, given that the "final distribution" $P(x,t)$ is consistent with the one we would obtain with the forward SDE. Notice that the gradient of probability landscape is required to construct the reverse process. Different from in maximum likelihood estimation, here the gradient is w.r.t diffused data $\vec{x}$, not model parameter θ.

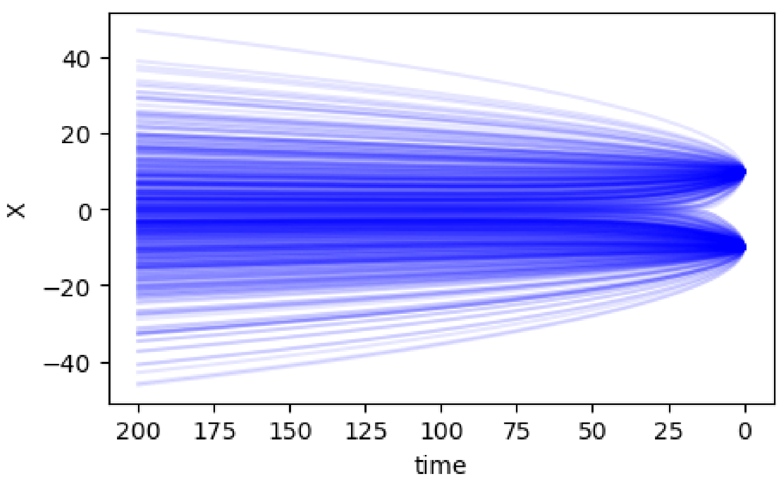
Alternatively, we can also define the reverse ODE, which samples the same distribution (marginal over final distribution) deterministically.
In the gaussian case, this is

$$x(t) = x(t+dt) + dt\frac{1}{2}\frac{1}{p}\nabla_x P(x,t)$$

Toy examples: Gaussian diffusion with bimodal initial distribution.
Jupyter lab
Plots: Forward SDE / Reverse SDE (use delta as final distribution) / Reverse ODE flow (use the correct Gaussian as final distribution)







Where $\frac{1}{p}\nabla_x P(x,t) = S_\theta(x,t)$ is the score function [2], and is parametrized by a neural network (Noise Conditional Score Network [3]).

References

1. The general existence of such a reverse SDE is proven by @andersonReversetimeDiffusionEquation1982 Detailed description of score-matching: @songScoreBasedGenerativeModeling2021.↩
2. List of Score SDE techniques: @yangDiffusionModelsComprehensive2022↩
3. @songGenerativeModelingEstimating2019↩