Diffusion Exponential Integral Sampler (DEIS) 제시.

→ ODE를 이산화하고, error를 줄이기 위해 semilinear structure 이용

1. Introduction.

Diffusion은 기본적으로 diffusion과 reverse로 이뤄짐.

→ Time-dependency score function을 배우기 때문에 score-based model 이라고 불림.

→ 다른 model보다 scalability가 좋고, hyperparam에 덜 민감

→ 모든 task에서 좋은 performance

하지만 느리기 때문에 빠르게 하기위한 방법을 연구.

→ 효율적인 reverse를 위한 forward 수정 및 변경 (DDIM)

→ numerical solver or SDE (Score-SDE, ···)

→ 둘다 : PNDM

DDIM은 Score-SDE의 PF의 이산화인데 Euler 같은 일반적인 방법보다 왜 잘되는지 모름

따라서 빠른 sampling을 위해 DM의 reverse의 일반화되어 있는 이산화 방법 설립

Network 과정이 가장 큰 resource를 필요로 한다.

∴ DM과 관련된 ODE/SDE 와 fitting error 와 이산화 error에 대해 조사.

같은 DM 이어도 이산화 방법에 따라 error가 극심히 다름

→ 오류를 조사하고 semilinear structure를 이용한 Exponential Integral (EI)가

가장 작은 error를 보인다는 것을 보임

이산화 에러를 더 줄이기 위해 ODE의 비선형을 근사하기 위한 고차 다항식이나

변형된 ODE에 Runge-kutta를 적용한다.

→ DEIS

Contributions :

1. 빠른 sampling을 위한 marginal ODEs/SDEs 계열의 조사와
   이에 대한 numerical solver에 대한 여러 조사.
2. DM에 일반적으로 적용할 수 있는 제한된 NFE에서 뛰어난 sampling quality를 보이는 DEIS 제안
3. DDIM의 이산화를 정당화하고, DDIM이 DEIS의 한 종류라는 것을 증명
4. DEIS의 성능 실험

2. Background.

- Forward noise diffusion.

SDE :

$$d\boldsymbol{x} = \underbrace{\boldsymbol{F}_t}_{\text{linear drift}}\boldsymbol{x}dt + \underbrace{\boldsymbol{G}_t}_{\text{Diffusion.}}d\boldsymbol{w}, \quad (1)$$

Table 1: Two popular SDEs, variance preserving SDE (VPSDE) and variance exploding SDE (VESDE). The parameter α_t is decreasing with $\alpha_0 \approx 1, \alpha_T \approx 0$, while σ_t is increasing.

SDE	$\boldsymbol{F}_t$	$\boldsymbol{G}_t$	μ_t	Σ_t
VPSDE (Ho et al., 2020)	$\frac{1}{2}\frac{d\log\alpha_t}{dt}\boldsymbol{I}$	$\sqrt{-\frac{d\log\alpha_t}{dt}}\boldsymbol{I}$	$\sqrt{\alpha_t}\boldsymbol{I}$	$(1-\alpha_t)\boldsymbol{I}$
VESDE (Song et al., 2020b)	0	$\sqrt{\frac{d[\sigma_t^2]}{dt}}\boldsymbol{I}$	$\boldsymbol{I}$	$\sigma_t^2\boldsymbol{I}$

- Backward denoising diffusion

Reverse-time diffusion process

$$d\boldsymbol{x} = [\boldsymbol{F}_t\boldsymbol{x}dt - \boldsymbol{G}_t\boldsymbol{G}_t^T\nabla\log p_t(\boldsymbol{x})]dt + \boldsymbol{G}_t d\boldsymbol{w}, \quad (2)$$

- Training

$$\mathcal{L}(\theta) = \mathbb{E}_{t\sim\text{Unif}[0,T]}\mathbb{E}_{p(\boldsymbol{x}_0)p_{0t}(\boldsymbol{x}_t|\boldsymbol{x}_0)}[||\nabla\log p_{0t}(\boldsymbol{x}_t|\boldsymbol{x}_0) - \boldsymbol{s}_\theta(\boldsymbol{x}_t,t)||^2_{\Lambda_t}]. \quad (3)$$

3. Fast Sampling with Learned Score Models

학습된 $S_\theta(x) \approx \nabla \log p_t(x)$가 훈련되면 infinity step으로 backward 가능

$\lambda \geq 0$ 일 때, SDE 군을 고려하면 (일반화)

$$d\hat{\boldsymbol{x}} = [\boldsymbol{F}_t \hat{\boldsymbol{x}} - \frac{1+\lambda^2}{2} \boldsymbol{G}_t \boldsymbol{G}_t^T \boldsymbol{s}_\theta(\hat{\boldsymbol{x}}, t)]dt + \lambda \boldsymbol{G}_t d\boldsymbol{w}, \tag{4}$$

→ $\lambda=0$ 이면 Score-SDE의 probability flow ODE

→ $\lambda=1$ 이면, eq.2.

Proposition 1. *When $\boldsymbol{s}_\theta(\boldsymbol{x}, t) = \nabla \log p_t(\boldsymbol{x})$ for all $\boldsymbol{x}, t$, and $\hat{p}_T^* = \pi$, the marginal distribution $\hat{p}_t^*$ of Eq. (4) matches p_t of the forward diffusion Eq. (1) for all $0 \leq t \leq T$.*

Fitting Error: S_θ 와 $\nabla \log p_t(x)$의 차이

Discretization Error: Eq.4의 이산화 Error

∴ $\lambda=0$을 중심으로 두 error를 해결하기 위해 연구.

VPSDE(DDPM)을 기준으로 함

3.1 Can We Learn Globally Accurate Score?

Diffusion의 성공으로 인해 fitting Error가 작다고 생각하지만 실제로 크다.

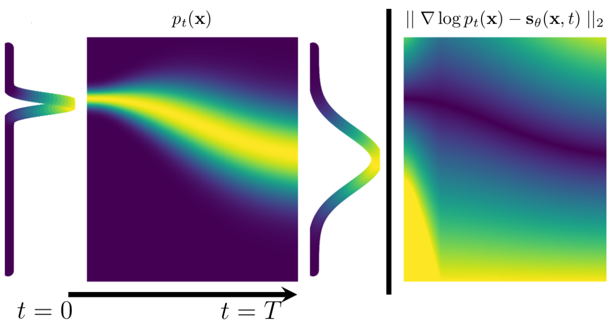

Figure 2: Fitting error on a toy demo. Lighter areas represent higher probability region (left) and larger fitting error (right).

→ $P_t(x)$의 region이 작으면 error↑, 크면 error↓

즉, 낮은 $p_t(x)$ 값 (낮은 확률) 에서는 훈련된 score가 부정확

∴ 이미지 같은 고차원 data는 작은 manifold에 모여있다

∴ $\hat{x}_t$의 품질을 높이려면 모든 t에서 $p_t(x)$가 높아야 한다.

→ 큰 step을 취하는 것이 힘듦

→ 좋은 이산화는 fitting error를 줄일수 있어야 함

3.1 Discretization Error

PF ODE($\lambda=0$)

$$\frac{d\hat{\boldsymbol{x}}}{dt} = \boldsymbol{F}_t\hat{\boldsymbol{x}} - \frac{1}{2}\boldsymbol{G}_t\boldsymbol{G}_t^T\boldsymbol{s}_\theta(\hat{\boldsymbol{x}}, t). \tag{5}$$

이에 대한 정확한 solution은

$$\hat{\boldsymbol{x}}_t = \Psi(t,s)\hat{\boldsymbol{x}}_s + \int_s^t \Psi(t,\tau)[-\frac{1}{2}\boldsymbol{G}_\tau\boldsymbol{G}_\tau^T\boldsymbol{s}_\theta(\hat{\boldsymbol{x}}_\tau,\tau)]d\tau, \tag{6}$$

↳ $\psi(t,s)$는 $\frac{\partial}{\partial t}\psi(t,s) = F_t\psi(t,s)$를 만족하고,
$\psi(s,s) = I$가 time $s \to t$ 까지의 transition matrix로 알려져 있을 때,

Eq.5가 linear term $F_t\hat{x}$와 nonlinear term $s_\theta(\hat{x},t)$로 구성된 semilinear stiff ODE 다.

Eq.6를 근사하기 위해 다양한 이산화 scheme과 관련된 eq.5에 대한 다양한 numerical solver가 있다.

이산화 step size가 0으로 가면, 모든 solution이 eq.5로 수렴

뒤에서, 큰 step size를 위해 eq.5의 discretization error 조사.

그리고 작은수의 Neural Function Evaluation (NFE)가 필요한 algorithm 개발

Ingredient 1. Exponential Integral over Euler method.