2024년 3월 22일 DMQA 연구실 오픈 세미나

The Two Formulations of Diffusion Models

이종현

Introduction

발표자 소개

❖ **이종현 (Jong Hyun Lee)**

- 고려대학교 산업경영공학과 석사 과정 (2022.09 ~ Present)
- Data Mining & Quality Analytics Lab

❖ **Research Interests**

- Computer Vision
- Deep Generative Models
- Diffusion Models

❖ **Contact**

- E-mail: tomtom1103@korea.ac.kr

Introduction

Generative Models

❖ **생성모델 (Generative Models)**

- Text 를 입력하면 그림을 그려주는 AI 로 큰 화제
- 엄청난 속도로 발전하는 Domain

"A drawing of a cat". "Horse eating a cupcake". "A 3D rendering of a temple".

CLIPDraw, 2021

Introduction

Generative Models

❖ 생성모델 (Generative Models)

- Text 를 입력하면 그림을 그려주는 AI 로 큰 화재
- 엄청난 속도로 발전하는 Domain

CLIPDraw, 2021

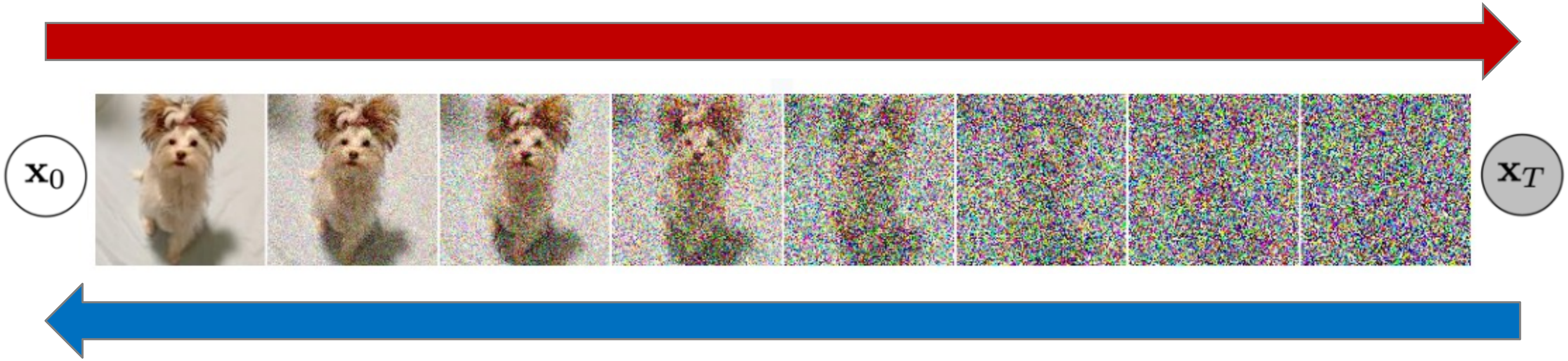
Sora, 2024

Introduction

Generative Models

❖ Diffusion Models

- Generative Models 중 가장 큰 성공을 이룬 model : Diffusion Models
- 비교적 느리지만, 높은 Quality 와 다양한 종류의 샘플을 생성 할 수 있다는 장점
- 그 근간에는 Forward / Reverse Process 가 있음

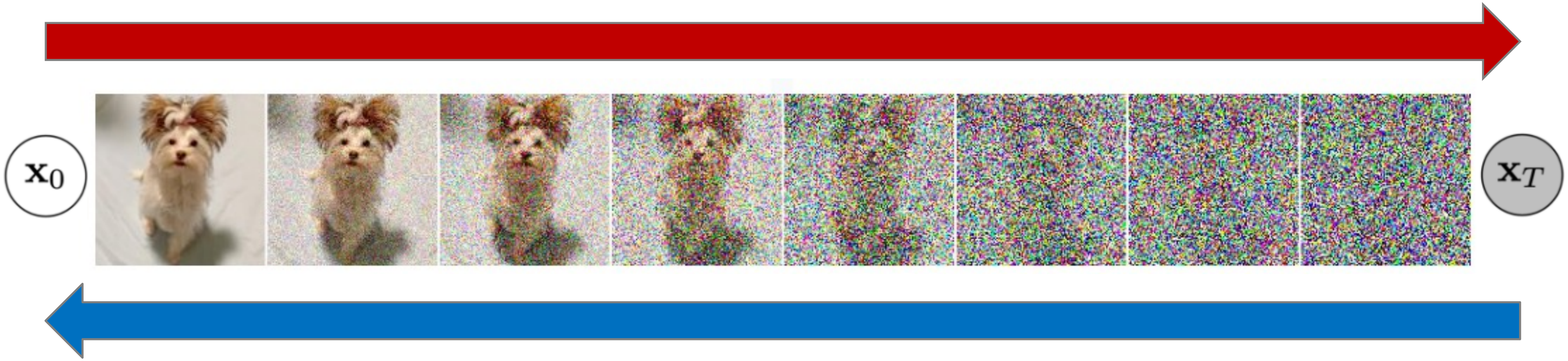


Introduction

Generative Models

❖ Diffusion Models

- Generative Models 중 가장 큰 성공을 이룬 model : Diffusion Models
- 비교적 느리지만, 높은 Quality 와 다양한 종류의 샘플을 생성 할 수 있다는 장점
- 그 근간에는 Forward / Reverse Process 가 있음
- Forward Process : 이미지 (x_0) 가 완전한 Gaussian Noise (x_T) 가 될 때까지 Gaussian Noise 를 점진적으로 추가하는 Markov Process
- Reverse Process : Gaussian Noise (x_T) 에서 점진적으로 Gaussian Noise 를 제거하여 이미지 (x_0) 를 복원하는 Markov Process

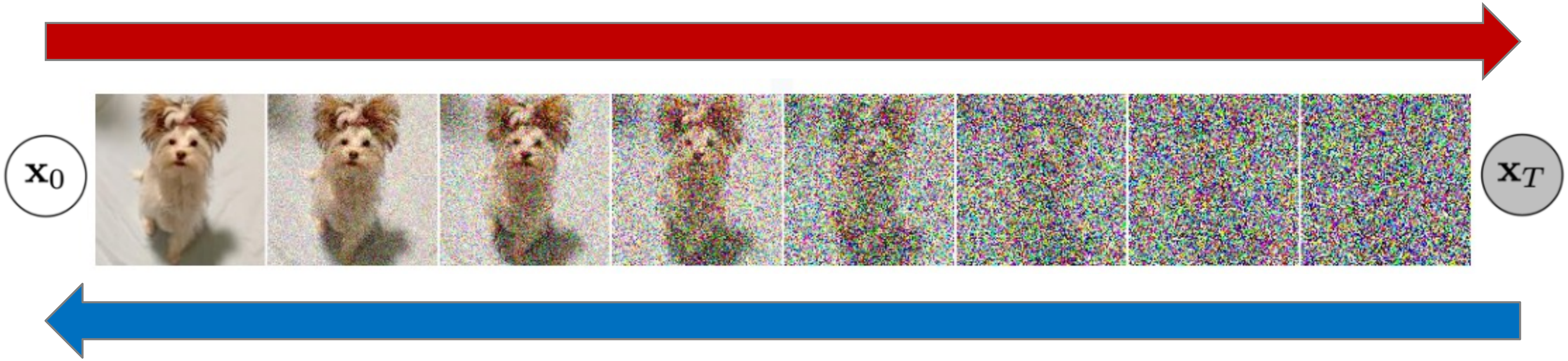


Introduction

Generative Models

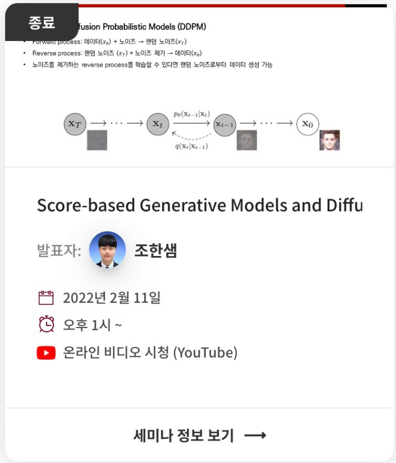
❖ Diffusion Models

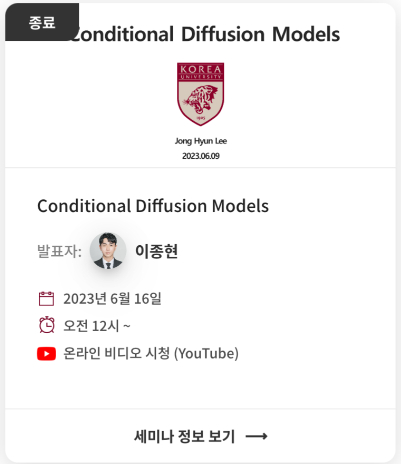
- Generative Models 중 가장 큰 성공을 이룬 model : Diffusion Models
- 비교적 느리지만, 높은 Quality 와 다양한 종류의 샘플을 생성 할 수 있다는 장점
- 그 근간에는 Forward / Reverse Process 가 있음
- Forward Process : 이미지 (x_0) 가 완전한 Gaussian Noise (x_T) 가 될 때까지 Gaussian Noise 를 점진적으로 추가하는 Markov Process
- Reverse Process : Gaussian Noise (x_T) 에서 점진적으로 Gaussian Noise 를 제거하여 이미지 (x_0) 를 복원하는 Markov Process
- **이런 Diffusion Model 을 Formulating 하는 방법은 크게 두가지가 있으며, 하나의 Unified Framework 또한 존재**



Introduction

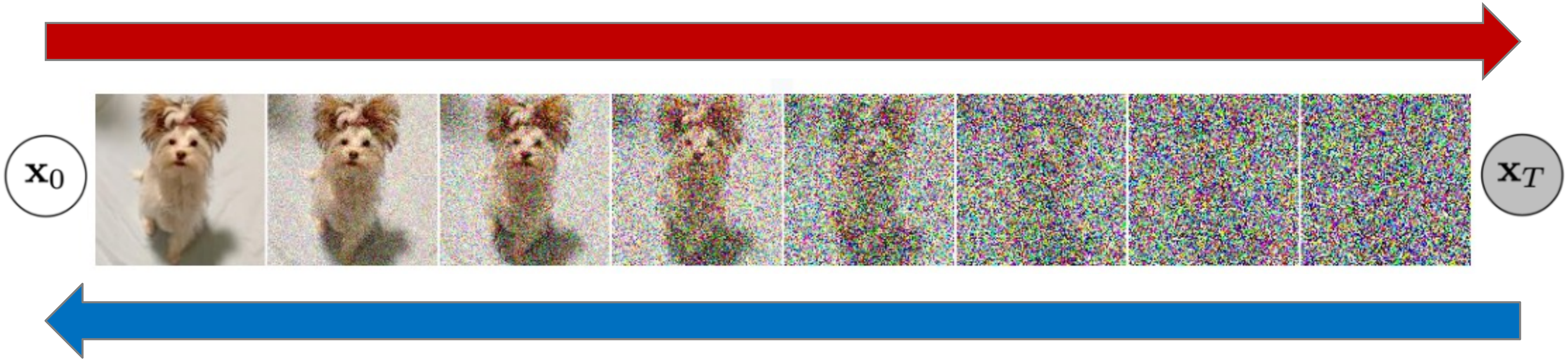
Generative Models





❖ Diffusion Models

- Generative Models 중 가장 큰 성공을 이룬 model : Diffusion Models
- 비교적 느리지만, 높은 Quality 와 다양한 종류의 샘플을 생성 할 수 있다는 장점
- 그 근간에는 Forward / Reverse Process 가 있음
- Forward Process : 이미지 (x_0) 가 완전한 Gaussian Noise (x_T) 가 될 때까지 Gaussian Noise 를 점진적으로 추가하는 Markov Process
- Reverse Process : Gaussian Noise (x_T) 에서 점진적으로 Gaussian Noise 를 제거하여 이미지 (x_0) 를 복원하는 Markov Process
- **이런 Diffusion Model 을 Formulating 하는 방법은 크게 두가지가 있으며, 하나의 Unified Framework 또한 존재**



Introduction

Generative Models

VP Formulation	VE Formulation	SDE/PF-ODEs	CM
DDPM, VDM	DSM,NCSN	Score SDE	Consistency Models

Two Formulations

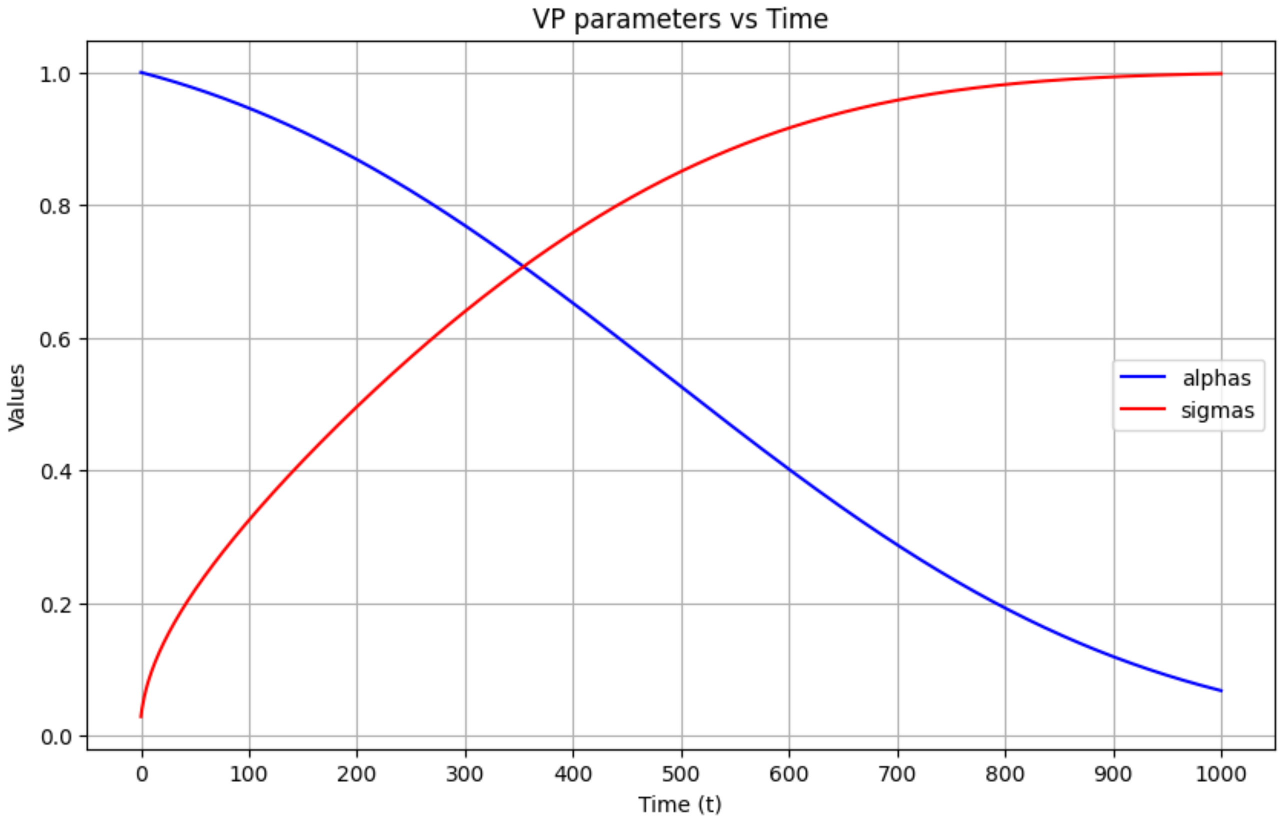
VP

❖ The Variance Preserving Formulation (VP)

- 대부분의 사람들이 익숙한 Diffusion Models (DDPM, Stable Diffusion, DALL-E, etc..) 는 이 Variance Preserving Formulation 을 따름
- Variance Preserving : 분산을 보존하는 Formulation
- Forward Process 에서 이미지가 점진적으로 Gaussian Noise 가 될 때 확인 가능

$$q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)=\mathcal{N}\left(\alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}\right)$$

$$\mathbf{x}_t=\alpha_t \mathbf{x}_0+\sigma_t \epsilon$$



Two Formulations

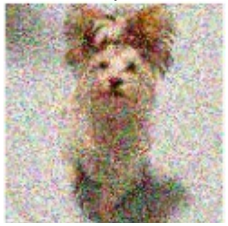
VP

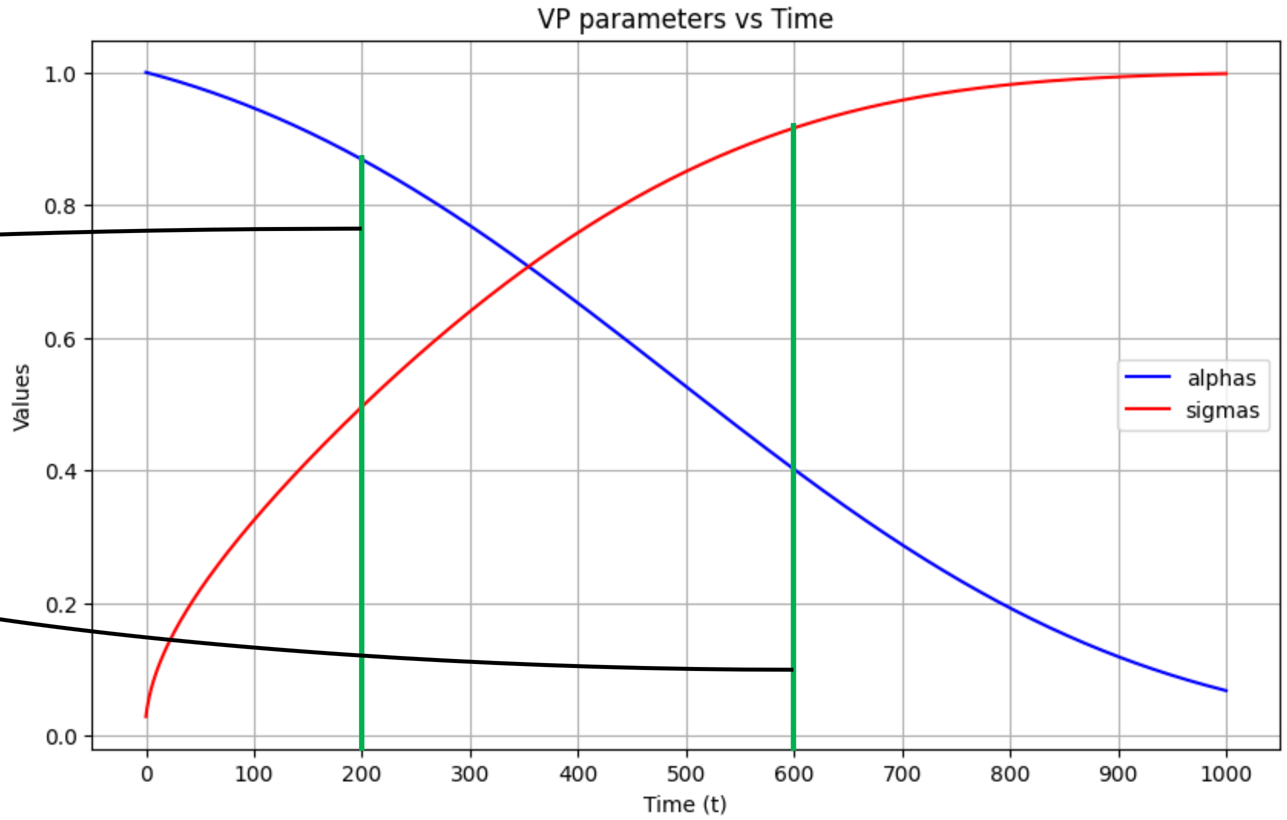
❖ The Variance Preserving Formulation (VP)

- 대부분의 사람들이 익숙한 Diffusion Models (DDPM, Stable Diffusion, DALL-E, etc..) 는 이 Variance Preserving Formulation 을 따름
- Variance Preserving : 분산을 보존하는 Formulation
- Forward Process 에서 이미지가 점진적으로 Gaussian Noise 가 될 때 확인 가능

$$q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)=\mathcal{N}\left(\alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}\right)$$

$$\mathbf{x}_t=\alpha_t \mathbf{x}_0+\sigma_t \epsilon$$





Two Formulations

VE

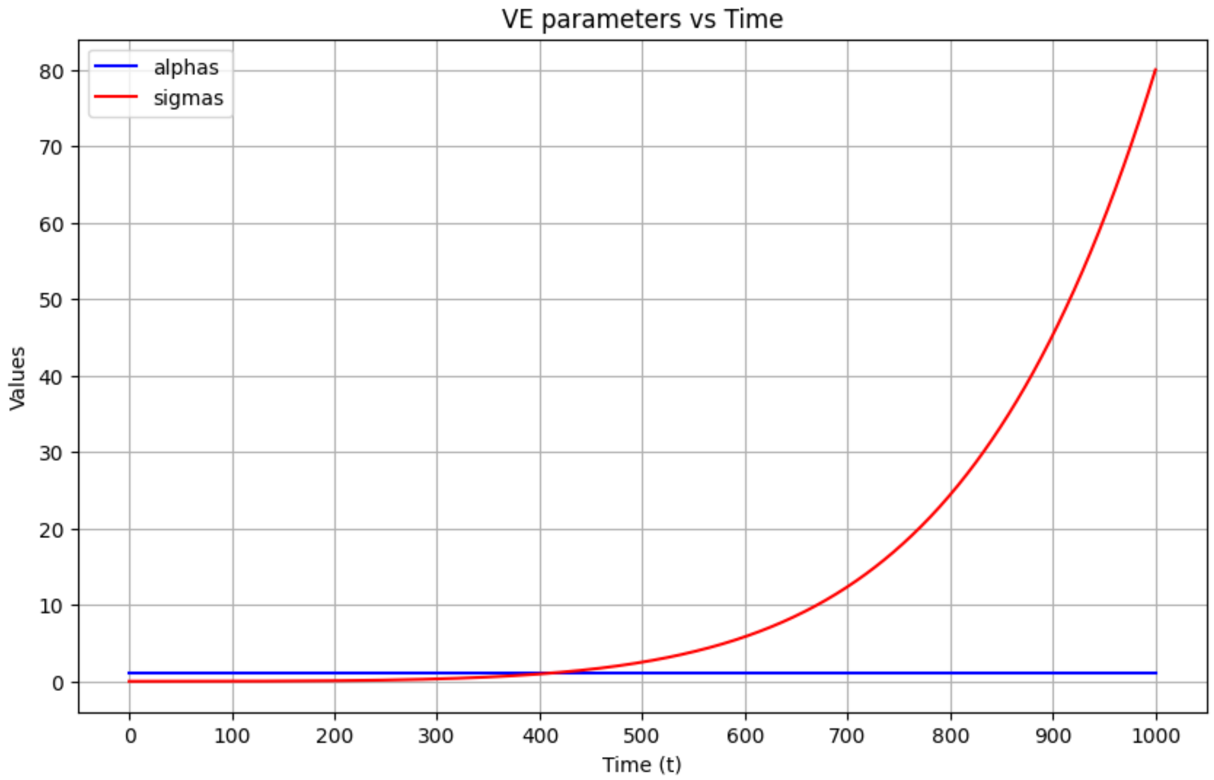
❖ The Variance Exploding Formulation (VE)

- 반면 Variance Exploding Formulation 도 존재 (NCSN, EDM, etc…)
- Variance Exploding : VP sigma 의 max 값은 1이지만, **VE 는 그 이상으로 설정**
- VP alpha 값은 timestep 이 증가할수록 작아지지만, **VE 에선 constant 값 1로 설정**

$$q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)=\mathcal{N}\left(\alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}\right)$$

$$\mathbf{x}_t=\alpha_t \mathbf{x}_0+\sigma_t \epsilon$$

$$\alpha_t=1, \sigma_{\max }=80$$



Two Formulations

VE

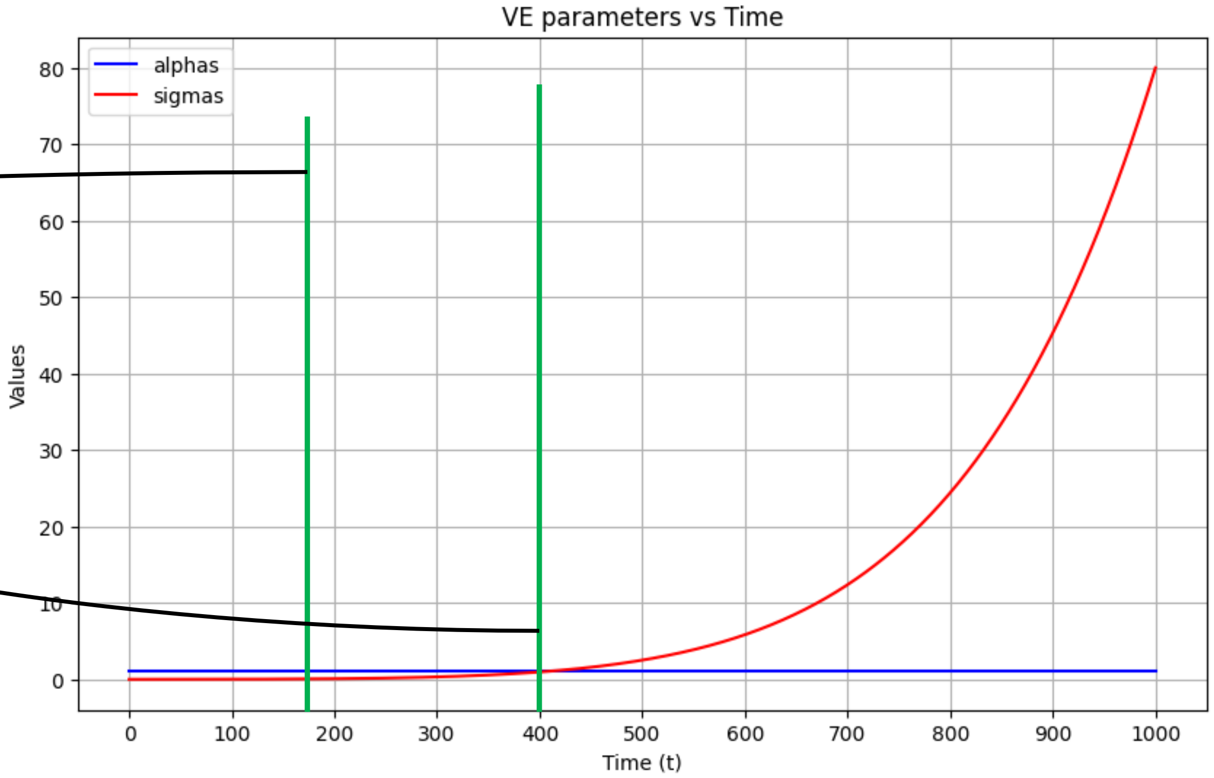
❖ **The Variance Exploding Formulation (VE)**

- 반면 Variance Exploding Formulation 도 존재 (NCSN, EDM, etc…)
- Variance Exploding : VP sigma 의 max 값은 1이지만, **VE 는 그 이상으로 설정**
- VP alpha 값은 timestep 이 증가할수록 작아지지만, **VE 에선 constant 값 1로 설정**

$$q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)=\mathcal{N}\left(\alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}\right)$$

$$\mathbf{x}_t=\alpha_t \mathbf{x}_0+\sigma_t \epsilon$$

$$\alpha_t=1, \sigma_{\max}=80$$



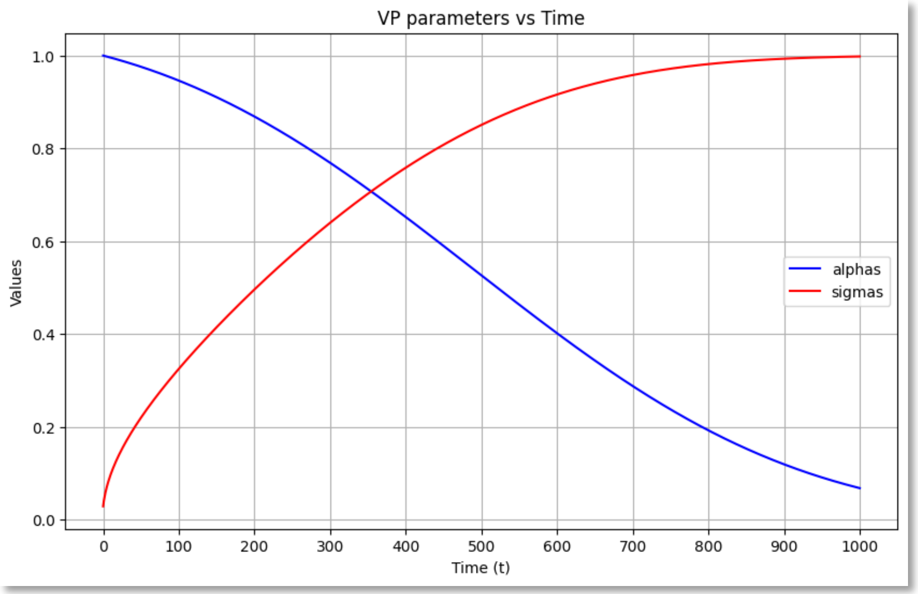
Data Mining
Quality Analytics

Two Formulations

VP vs. VE

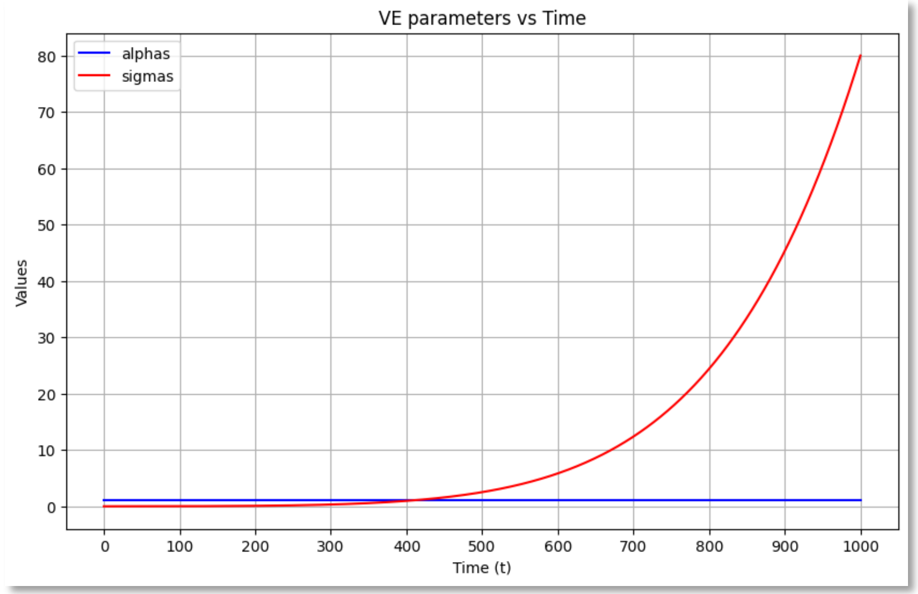
❖ VP vs. VE Formulations

- VP Formulation 은 Text to image Diffusion models 에서 많이 사용
- VE Formulation 은 Theoretical 연구에서 많이 사용



$$q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)=\mathcal{N}\left(\alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}\right)$$
$$\mathbf{x}_t=\alpha_t \mathbf{x}_0+\sigma_t \epsilon$$

	Variance Preserving (VP)	Variance Exploding (VE)
Models	DDPM, Stable Diffusion, DALL-E, VDM, etc.	NCSN, Score SDE, EDM, CM
Sigma	$\lim_{t\to 1} \sigma_t = 1$	$\lim_{t\to 1} \sigma_t = \sigma_{\max}$
Alpha	$\lim_{t\to 1} \alpha_t = 0$	$\text{const.}\alpha_t = 1$



Two Formulations

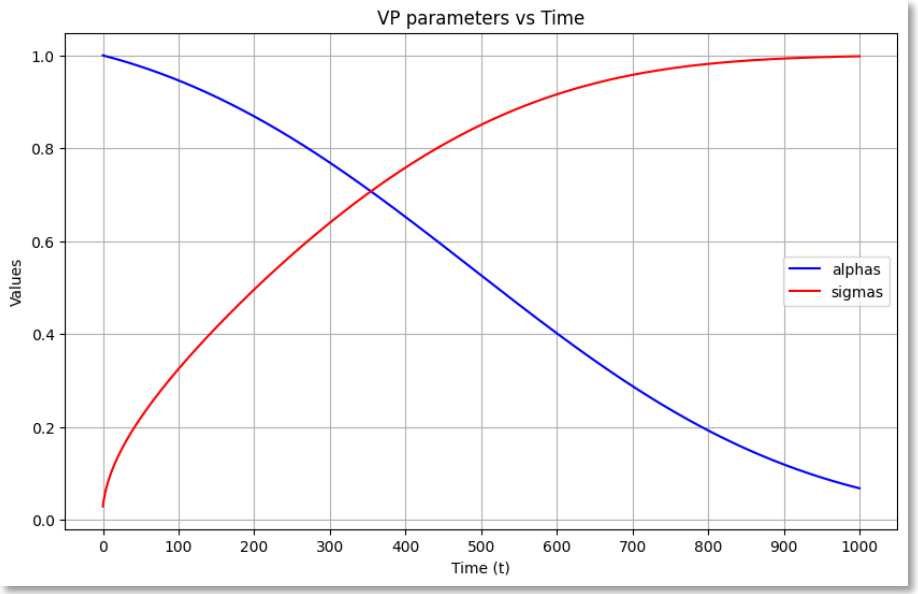
VP vs. VE

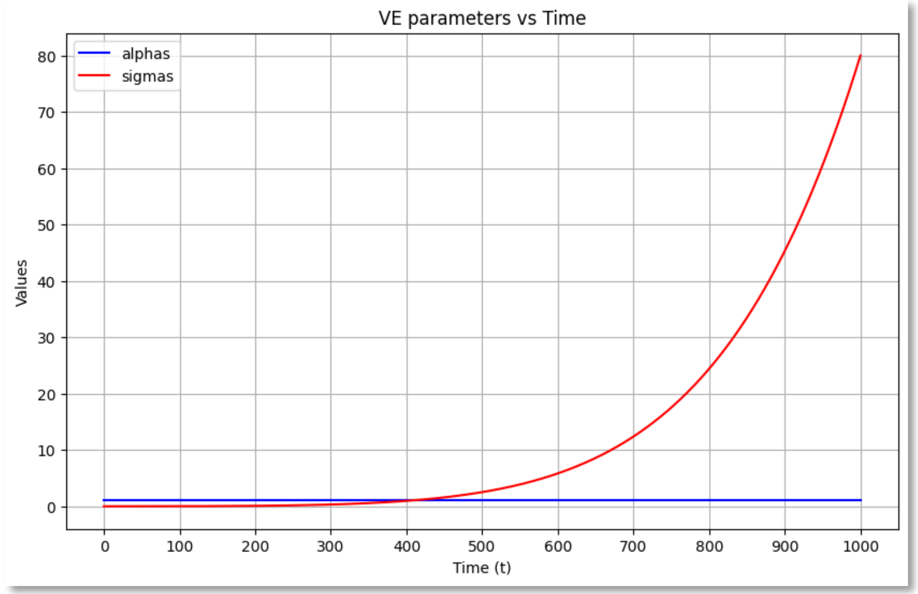
❖ VP vs. VE Formulations

- VP Formulation 은 Text to image Diffusion models 에서 많이 사용
- VE Formulation 은 Theoretical 연구에서 많이 사용
- **왜 두가지 Formulation 이 존재할까?**

$$q\left(\mathbf{x}_t \mid \mathbf{x}_0\right)=\mathcal{N}\left(\alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I}\right)$$
$$\mathbf{x}_t=\alpha_t \mathbf{x}_0+\sigma_t \epsilon$$

	Variance Preserving (VP)	Variance Exploding (VE)
Models	DDPM, Stable Diffusion, DALL-E, VDM, etc.	NCSN, Score SDE, EDM, CM
Sigma	$\lim_{t\to 1} \sigma_t = 1$	$\lim_{t\to 1} \sigma_t = \sigma_{\max}$
Alpha	$\lim_{t\to 1} \alpha_t = 0$	$\mathrm{const.}\alpha_t = 1$



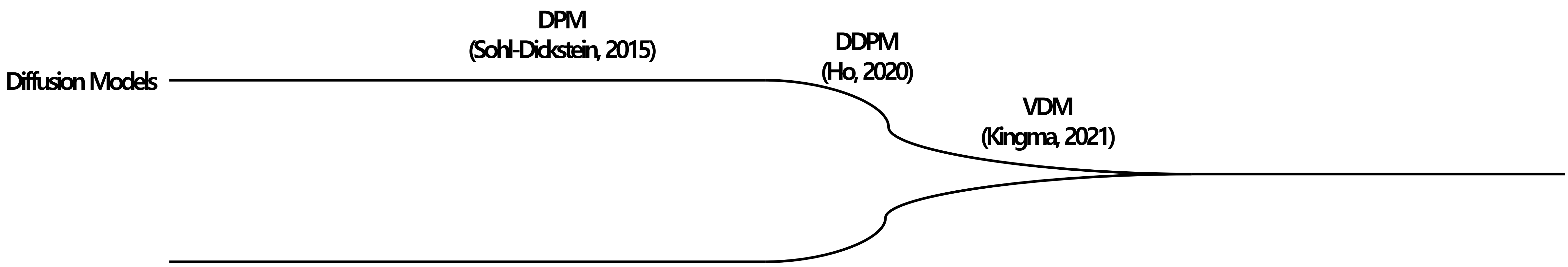


Two Formulations

VP vs. VE

❖ **Two branches of Diffusion Models**

- Diffusion Model 은 사실상 다른 이름으로 동시다발적으로 연구된 분야
- **VP Formulation : Diffusion Models**

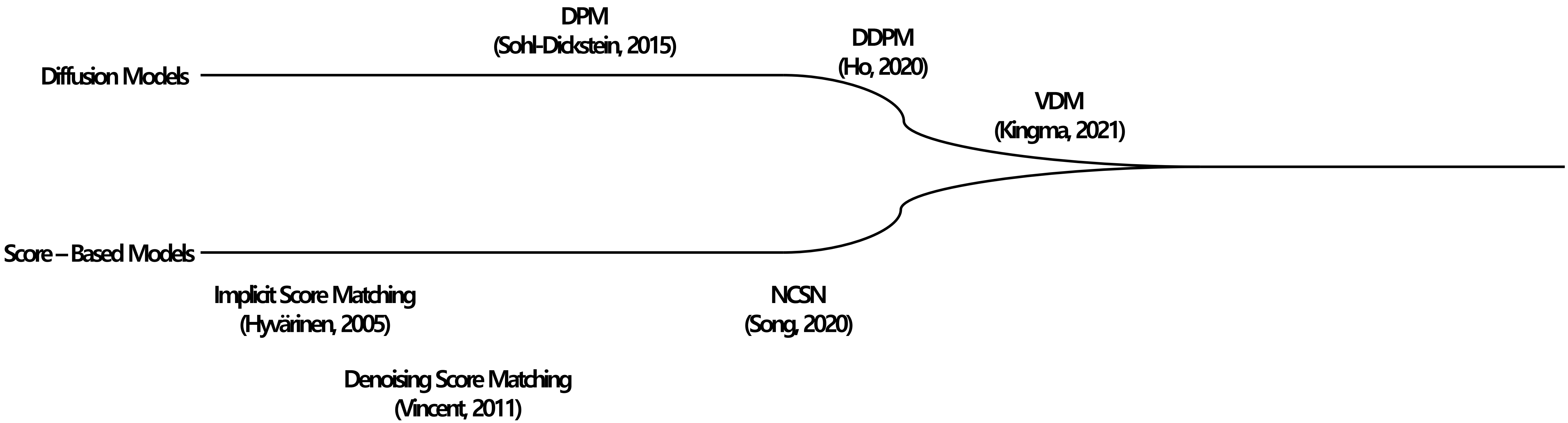


Two Formulations

VP vs. VE

❖ **Two branches of Diffusion Models**

- Diffusion Model 은 사실상 다른 이름으로 동시다발적으로 연구된 분야
- **VP Formulation : Diffusion Models**
- **VE Formulation : Score – Based Models**



Two Formulations

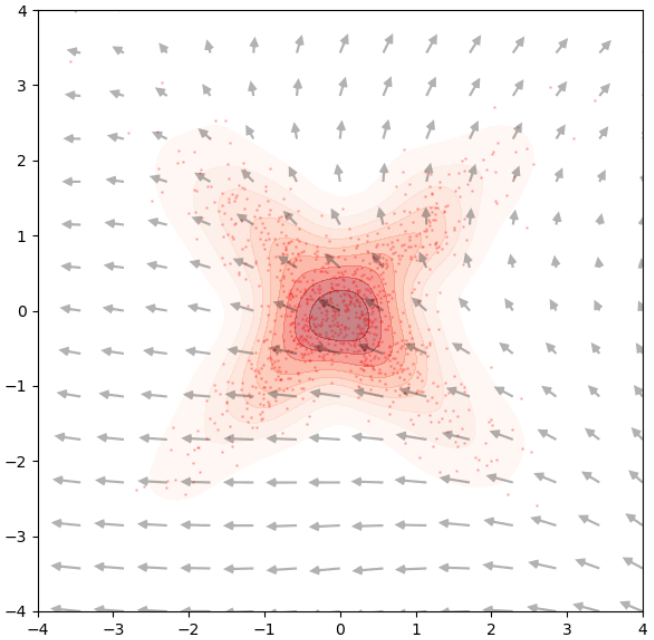
Score – Based Models

❖ **Score – Based Models**

- Score 란? Gradient of the log density of a distribution. 즉, 분포에 로그를 취한 값의 미분값.
- 1935년에 Ronald Fisher (Fisher Divergence) 가 처음 발견
- 특정 Random variable X 가 있고, 이 RV 를 Score 만큼 미세하게 움직이면 새로운 RV 는 이전 RV 보다 Likelihood 가 더 높음

$$\nabla_{\boldsymbol{x}} \log p\left(\boldsymbol{x}\right)$$

$$\text{likelihood}(\boldsymbol{x}) < \text{likelihood}(\boldsymbol{x} + \delta \nabla_{\boldsymbol{x}} \log p\left(\boldsymbol{x}\right))$$



Two Formulations

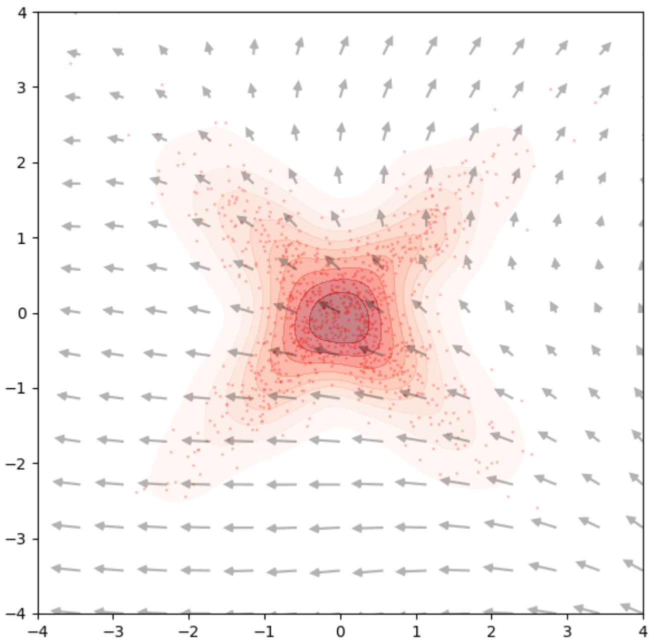
Score – Based Models

❖ Score – Based Models

- Score 란? Gradient of the log density of a distribution. 즉, 분포에 로그를 취한 값의 미분값.
- 1935년에 Ronald Fisher (Fisher Divergence) 가 처음 발견
- 특정 Random variable X 가 있고, 이 RV 를 Score 만큼 미세하게 움직이면 새로운 RV 는 이전 RV 보다 Likelihood 가 더 높음
- 즉, 특정 RV 에 대해 Likelihood 가 높은 방향으로 pointing 을 해주는 vector field
- **아무런 Random variable 값에서 시작해도 Score 를 따라 움직이면 높은 Likelihood 의 샘플을 얻을 수 있다**

$$\nabla_{\boldsymbol{x}} \log p\left(\boldsymbol{x}\right)$$

$$\text{likelihood}(\boldsymbol{x}) < \text{likelihood}(\boldsymbol{x} + \delta \nabla_{\boldsymbol{x}} \log p\left(\boldsymbol{x}\right))$$



Two Formulations

Score – Based Models

❖ Score – Based Models and Diffusion Models

- 아무런 Random variable 값에서 시작해도 Score 를 따라 움직이면 높은 Likelihood 의 샘플을 얻을 수 있다
- 이런 특징을 이용해서, Random Variable 에 대해 Score 을 예측 할 수 있는 모델들이 등장 – Score – Based Models
- 하지만 그냥 Score 을 아무런 Random Variable 에 대해 예측하기보단, 이미지에 Gaussian noise 를 주입해 Score 을 예측

$$J_{\mathrm{D}}(\theta) = \mathbb{E}_{x\sim q_{\text{data}}\ (x),\epsilon\sim\mathcal{N}(0,I)}\left[\tfrac{1}{2}\left\| s_\theta(\underbrace{x+\sigma\epsilon}_{\tilde{x}}) - \left(-\tfrac{\epsilon}{\sigma}\right)\right\|^2\right]$$

Obj. Func. of Denoising Score Matching

Two Formulations

Score – Based Models

❖ Score – Based Models and Diffusion Models

- 아무런 Random variable 값에서 시작해도 Score 를 따라 움직이면 높은 Likelihood 의 샘플을 얻을 수 있다
- 이런 특징을 이용해서, Random Variable 에 대해 Score 을 예측 할 수 있는 모델들이 등장 – Score – Based Models
- 하지만 그냥 Score 을 아무런 Random Variable 에 대해 예측하기보단, 이미지에 Gaussian noise 를 주입해 Score 을 예측
- Modern Day Diffusion model 의 시초인 DDPM 에선 Diffusion/Score based 에 대한 연관성을 인지하지만, 완벽하게 formulate 하진 않음

$$J_{\mathrm{D}}(\theta) = \mathbb{E}_{x \sim q_{\text{data}}\ (x), \epsilon \sim \mathcal{N}(0,I)} [\tfrac{1}{2} \| s_\theta(\underbrace{x + \sigma\epsilon}_{\tilde{x}}) - (-\tfrac{\epsilon}{\sigma})\ \|^2]$$

Obj. Func. of Denoising Score Matching

$$L_{\text{simple}}\ (\theta) := \mathbb{E}_{t, \mathbf{x}_0, \boldsymbol{\epsilon}} \left[\left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta \left(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, t \right) \right\|^2 \right]$$

Obj. Func. of DDPM

Two Formulations

Variational Diffusion Models

❖ Variational Diffusion Models

- VDM 이란 논문에선 Score based 와 Diffusion Model 의 output 들이 서로 상호교환 가능하다 것을 정리
- 즉, Score based models 의 목적함수로 모델을 학습시켰든, DDPM 의 목적함수로 모델을 학습시켰든 상관 없이

  Diffusion Model 로도 Score 을 예측 할 수 있음

$$\hat{\mathbf{x}}_\theta(\mathbf{x}_t, t) = \frac{\mathbf{x}_t - \sigma_t \epsilon_\theta(\mathbf{x}_t, t)}{\alpha_t}$$

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t\right)\right\|^2\right]$$

$$s_\theta(\mathbf{x}_t) = \frac{\alpha_t \hat{\mathbf{x}}_\theta(\mathbf{x}_t, t) - \mathbf{x}_t}{\sigma_t^2}$$

$$s_\theta(\mathbf{x}_t) = \frac{-\epsilon_\theta(\mathbf{x}_t, t)}{\sigma_t}$$

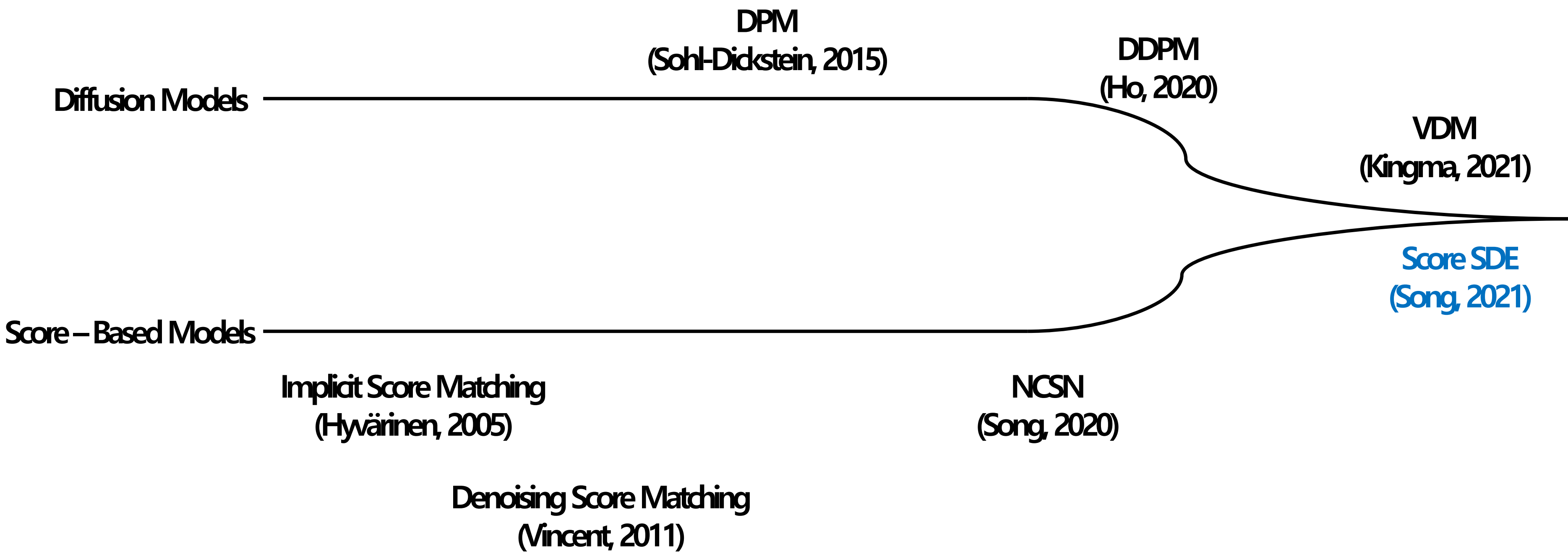
$$J_{\mathrm{D}}(\theta) = \mathbb{E}_{x\sim q_{\text{data}}(x), \epsilon\sim\mathcal{N}(0,I)}\left[\frac{1}{2}\left\| s_\theta(\underbrace{x + \sigma\epsilon}_{\tilde{x}}) - \left(-\frac{\epsilon}{\sigma}\right)\right\|^2\right]$$

Two Formulations

VP vs. VE

❖ **Two branches of Diffusion Models**

- Diffusion Model 은 사실상 다른 이름으로 동시다발적으로 연구된 분야
- VP Formulation : Diffusion Models
- VE Formulation : Score – Based Models
- **A unified Framework : Score SDE**



VP Formulation

DDPM,
VDM

VE Formulation

DSM,NCSN

SDE/PF-ODEs

Score SDE

CM

Consistency Models

SDE/PF-ODE

Score SDEs

❖ **Score SDEs**

- 먼저, ODE/SDE 란?
- ODE (Ordinary Differential Equation) : 미지의 함수와 도함수의 관계를 나타낸 방정식. 보통 시간에 따라 변하는 변수의 변화량을 수식화

$$\mathrm{ODE} : \mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t$$

SDE/PF-ODE

Score SDEs

❖ Score SDEs

- 먼저, ODE/SDE 란?
- ODE (Ordinary Differential Equation) : 미지의 함수와 도함수의 관계를 나타낸 방정식. 보통 시간에 따라 변하는 RV 의 변화량을 수식화
- SDE (Stochastic Differential Equation) : ODE 에 랜덤성이 포함된 미분방정식

$$\text{ODE} : \mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t$$

$$\text{SDE} : \mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$$

SDE/PF-ODE

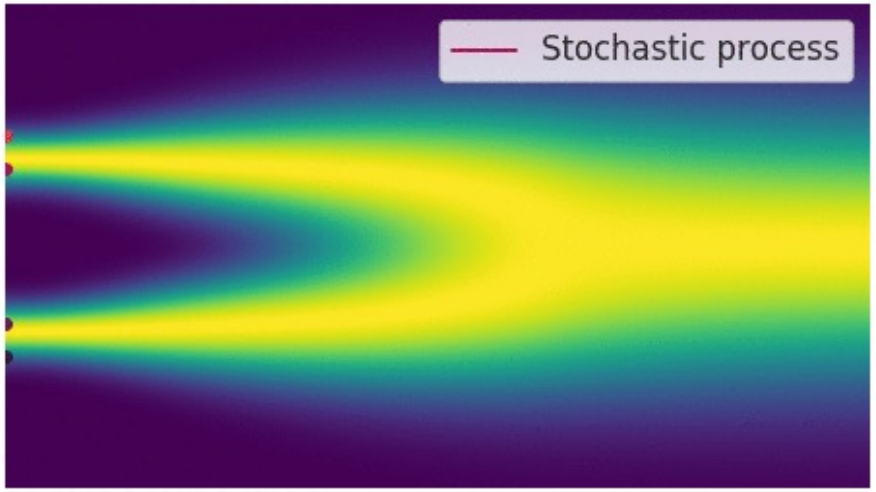
Score SDEs

❖ Score SDEs

- 먼저, ODE/SDE 란?
- ODE (Ordinary Differential Equation) : 미지의 함수와 도함수의 관계를 나타낸 방정식. 보통 시간에 따라 변하는 RV 의 변화량을 수식화
- SDE (Stochastic Differential Equation) : ODE 에 랜덤성이 포함된 미분방정식
- **Score SDE 논문에선 Diffusion Model 과 Score Based Model 을 하나의 SDE 로 표현 할 수 있다는 것을 발견**

$$\mathrm{ODE} : \mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t$$

$$\mathrm{SDE} : \mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$$

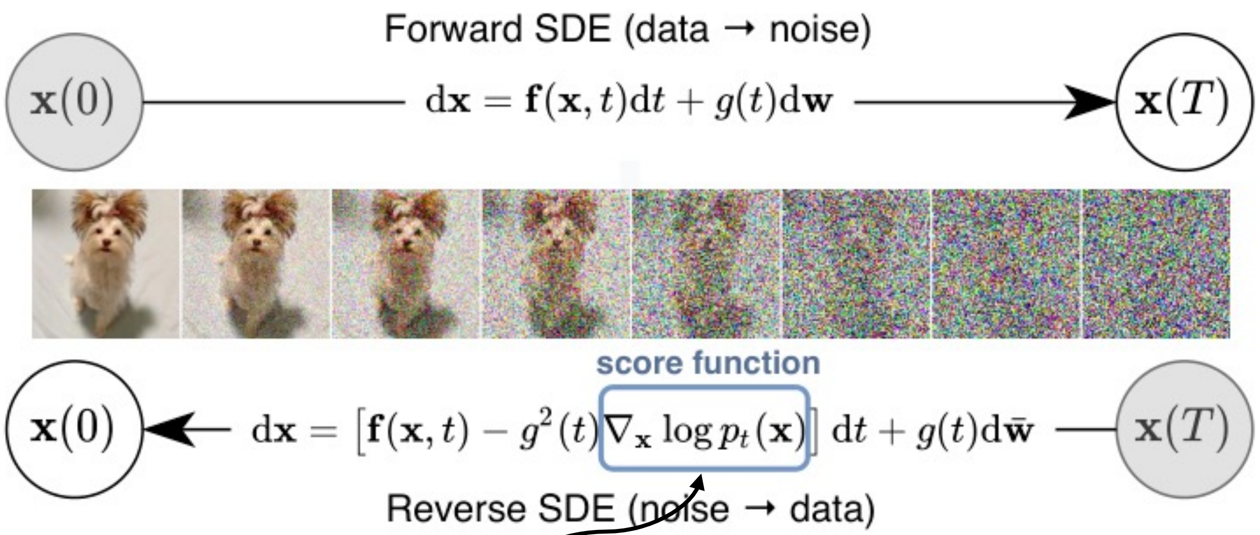


SDE/PF-ODE

Score SDEs

❖ Score SDEs

- Diffusion SDE 는 연속적인 시간의 변화에 따른 이미지의 변화를 모델링
- Forward SDE : 아주 작은 시간의 변화 dt 와 아주 작은 노이즈의 변화 dw 에 대한 이미지의 변화 dx
- Reverse SDE : Reverse SDE: 아주 작은 시간의 변화 dt 와 아주 작은 노이즈의 변화 dw 에 대한 이미지의 변화 dx
- 이때 Reverse SDE 를 계산하려면 Score 값이 필요 : **Diffusion Model 과 Score Based Model 둘 다 활용 가능**



$$s_\theta(\mathbf{x}_t) = \frac{-\epsilon_\theta(\mathbf{x}_t, t)}{\sigma_t}$$

SDE/PF-ODE

Score SDEs

❖ Score SDEs

- 이때 Reverse SDE 를 계산하려면 Score 값이 필요 : **Diffusion Model 과 Score Based Model 둘 다 활용 가능**
- Score SDE 논문에선 이 Framework 를 구축하면서 기존 Diffusion model 은 VP SDE (Variance Preserving) ,
  Score Based model 은 VE SDE (Variance Exploding) 이라 명명
- 이유는 각 모델의 목적함수를 보면 알 수 있음 :
  Diffusion model 에선 이미지에 노이즈를 추가하면서 (Sigma) 동시에 이미지의 픽셀값을 Scaling 함 (Alpha)
  반면 Score Based model 에선 이미지를 따로 Scaling 하지 않고 노이즈만 추가함

$$J_{\mathrm{D}}(\theta) = \mathbb{E}_{x\sim q_{\text{data}}(x),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{1}{2}\left\|s_\theta(\underbrace{x+\sigma\epsilon}_{\tilde{x}}) - \left(-\frac{\epsilon}{\sigma}\right)\right\|^2\right]$$

Obj. Func. of Denoising Score Matching

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t\right)\right\|^2\right]$$

Obj. Func. of DDPM

Data Mining
Quality Analytics

SDE/PF-ODE

Score SDEs

❖ Score SDEs

- 이때 Reverse SDE 를 계산하려면 Score 값이 필요 : Diffusion Model 과 Score Based Model 둘 다 활용 가능
- Score SDE 논문에선 이 Framework 를 구축하면서 기존 Diffusion model 은 VP (Variance Preserving),
  Score Based model 은 VE (Variance Exploding) 이라 명명
- 이유는 각 모델의 목적함수를 보면 알 수 있음 :
  Diffusion model 에선 이미지에 노이즈를 추가하면서 (Sigma) 동시에 이미지의 픽셀값을 Scaling 함 (Alpha)
  반면 Score Based model 에선 이미지를

디퓨전을 SDE 로 재해석하면 장점이 있나?

기존 방법대로 잘만 작동했는데?

$$J_{\mathrm{D}}(\theta) = \mathbb{E}_{x\sim q_{\mathrm{data}}(x),\epsilon\sim\mathcal{N}(0,I)}\left[\frac{1}{2}\left\|s_\theta(\underbrace{x+\sigma\epsilon}_{\tilde{x}}) - \left(-\frac{\epsilon}{\sigma}\right)\right\|^2\right]$$

Obj. Func. of Denoising Score Matching

$$L_{\mathrm{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\epsilon}\left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right)\right\|^2\right]$$

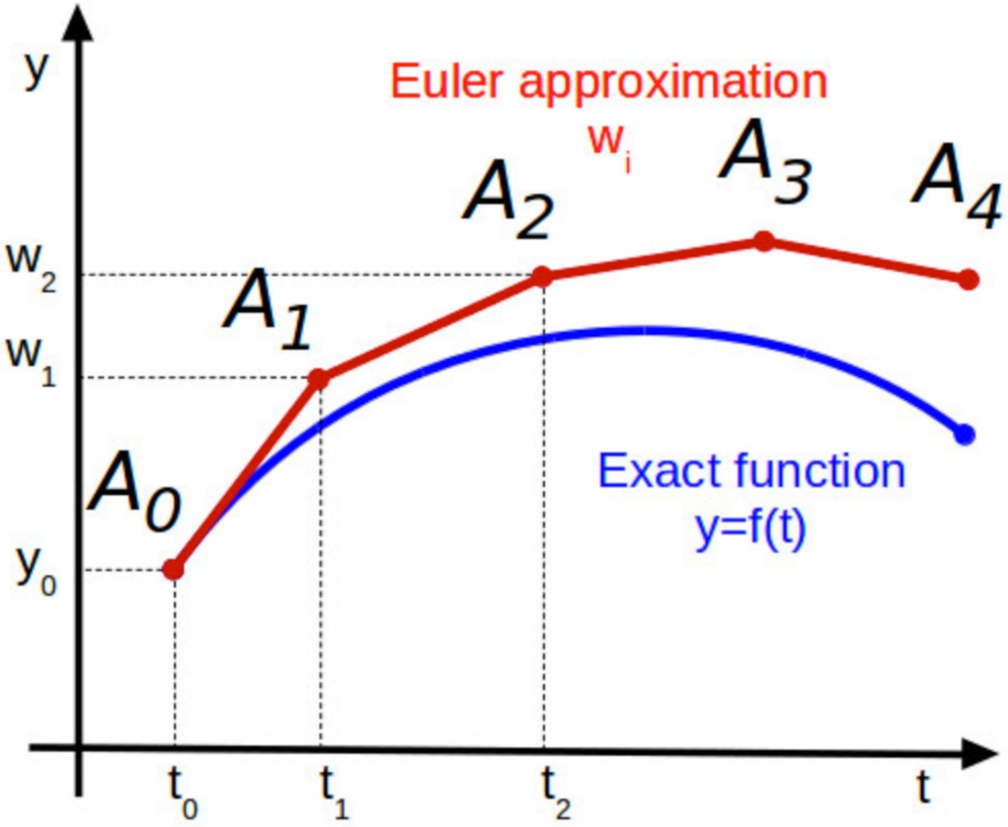
Obj. Func. of DDPM

SDE/PF-ODE

Score SDEs

❖ Score SDEs

- ODE (Ordinary Differential Equations) 를 푼다 : 시작점이 주어지면 끝점과 그 사이의 변화량을 알아내는 것 (Trajectory)
- Euler Method 처럼 ODE 를 위한 Solving 알고리즘을 통해 풀 수 있음



$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t$$

1. 시작 지점을 설정 (A_0)
2. 미분방정식 dx 계산
3. 이 dx 값은 해당 시점의 기울기 (slope)
4. 1 Step 증가 : $A_1 = A_0 + \delta dx$
5. 반복

SDE/PF-ODE

Score SDEs

❖ Score SDEs

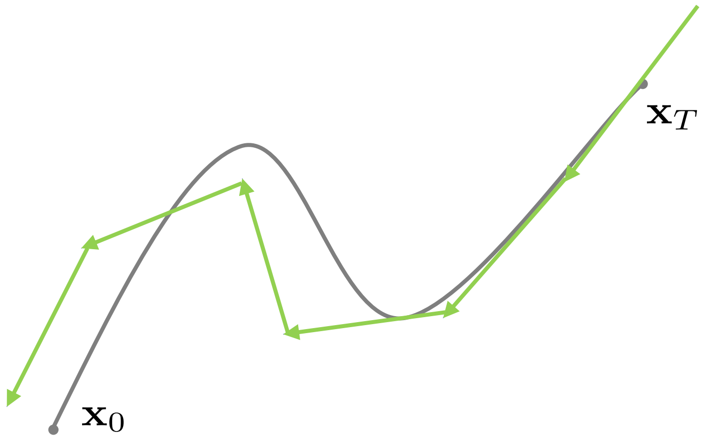
- 놀라운 점은 Diffusion Reverse SDE 와 동일한 Marginal 을 갖는 ODE 가 존재 (Stochasticity 가 없음)
- Score SDE 논문에선 이 ODE 를 Probability Flow ODE 라고 명명
- 즉, 일반적인 ODE Solver 로 디퓨전 샘플링을 할 수 있음

$$\text{Reverse SDE} : \mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] \mathrm{d}t + g(t)\mathrm{d}\overline{\mathbf{w}}$$

$$\text{PF-ODE} : \mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - \tfrac{1}{2} g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] \mathrm{d}t$$

$$\nabla_{\mathbf{x}} \log p_t(\mathbf{x}) = -\epsilon_\theta / \sigma_t$$

일반적으로 학습된
디퓨전 모델의 Output



Euler Method

1. 시작 지점을 설정 (X_T)
2. 미분방정식 dx 계산
3. 이 dx 값은 해당 시점의 기울기 (slope)
4. 1 Step 증가 : $X_{T-1} = X_T + dx$
5. 반복

Data Mining
Quality Analytics

SDE/PF-ODE

Score SDEs

❖ Score SDEs

- 놀라운 점은 Diffusion Reverse SDE 와 동일한 Marginal 을 갖는 ODE 가 존재 (Stochasticity 가 없음)
- Score SDE 논문에서 이 ODE 를 Probability Flow ODE 라고 명명
- 즉, 일반적인 ODE Solver 로 디퓨전 샘플링을 할 수 있음

$$\text{Reverse SDE : } \mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] \mathrm{d}t + g(t)\, \mathrm{d}\bar{\mathbf{w}}$$

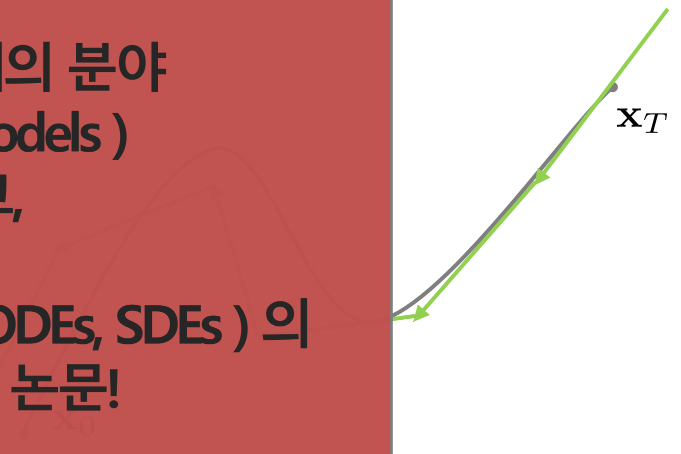
$$\text{PF-ODE : } \mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - \frac{1}{2} g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] \mathrm{d}t$$

$$\nabla_{\mathbf{x}} \log p_t(\mathbf{x}) = -$$

일반적으로 학습된
디퓨전 모델의 Output

Score SDE 논문은 다른 줄 알았던 두개의 분야
(Diffusion Models, Score – Based Models)
를 하나의 Framework 로 합치고,

아무런 관련이 없을 줄 알았던 수치해석학 (ODEs, SDEs) 의
발전을 그대로 적용 할 수 있게 해준 논문!



Euler Method

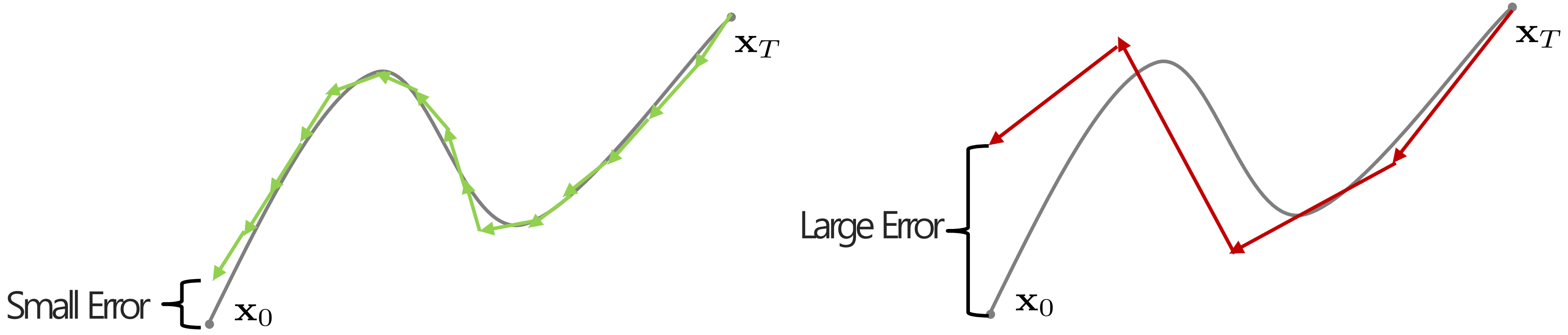
1. 시작 지점을 설정 (X_T)
2. 미분방정식 dx 계산
3. 이 dx 값은 해당 시점의 기울기 (slope)
4. 1 Step 증가 : $X_{T-1} = X_T + dx$
5. 반복

VP Formulation	VE Formulation	SDE/PF-ODEs	CM
DDPM, VDM	DSM,NCSN	Score SDE	Consistency Models

Consistency Models

❖ **Diffusion / Score based models 의 단점**

- Score SDE 의 발견으로 SDE/ODE Solving 을 통한 Few step sampling 으로 이미지를 생성 할 수 있게 됨
- 하지만 SDE/ODE Solving 은 step 수가 작아질 수록 Error 가 증가
- **결국 Quality vs. Sampling speed 의 tradeoff 는 그대로 존재**



Euler Method, 15 Steps

Euler Method, 4 Steps

Consistency Models

❖ PF - ODE Solving and the Denoiser Output

- 일반적인 PF-ODE 를 통한 Sampling 을 다시 살펴보면:
- For Solver Steps, do :
  - Denoised output 계산
  - Denoised output 을 통해 Score 계산
  - Score 을 통해 Derivative 계산
  - Solver Step

$$1.\ D_\phi(\mathbf{x}_t, t) = \frac{\mathbf{x}_t - \sigma_t \epsilon_\theta(\mathbf{x}_t, t)}{\alpha_t}$$

$$2.\ s_\theta(\mathbf{x}_t) = \frac{\alpha_t D_\phi(\mathbf{x}_t, t) - \mathbf{x}_t}{\sigma_t^2} = \frac{-\epsilon_\theta(\mathbf{x}_t, t)}{\sigma_t}$$

$$3.\ \frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = f(t)\mathbf{x}_t - \frac{1}{2} g^2(t) \nabla_{\mathbf{x}} \log p(\mathbf{x}_t)$$

$$4.\ \mathbf{x}_{t-1} = \mathbf{x}_t + \Delta t \frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t}\ \text{(Euler Step)}$$

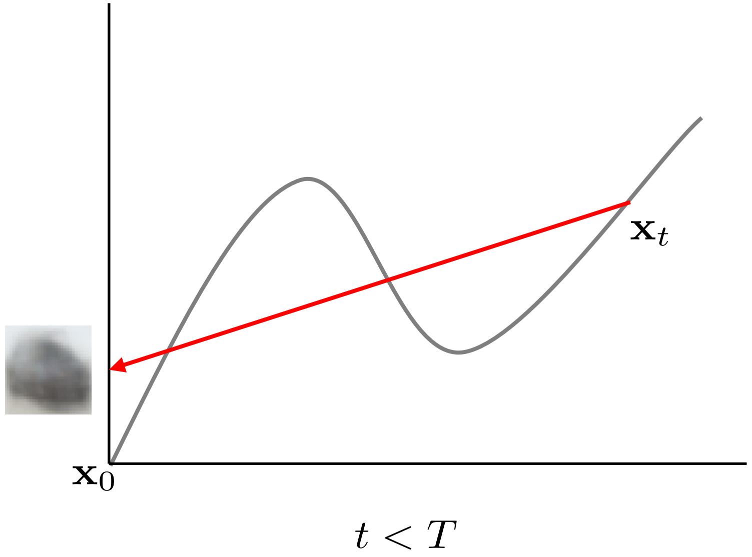
- 이때 Denoised output 이란 모델이 해당 Timestep 에 대해 예측한 Clean Image !
- DDIM 에서도 등장

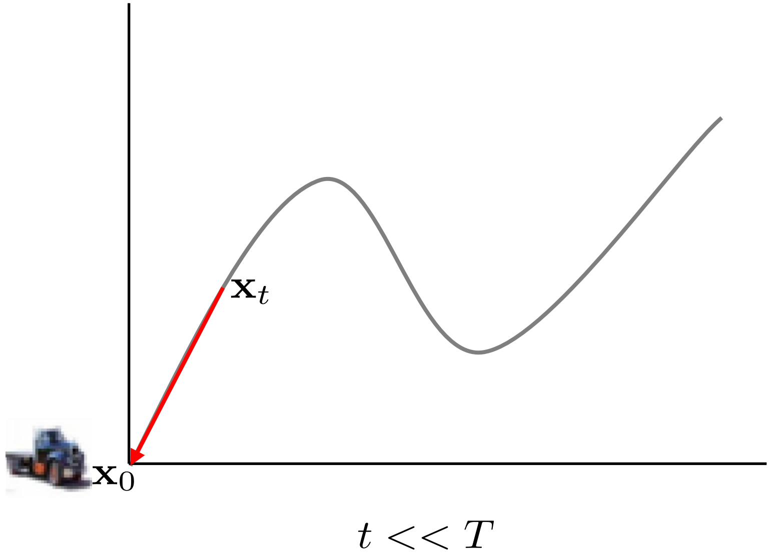
$$\underbrace{\left(\frac{\boldsymbol{x}_t - \sqrt{1 - \alpha_t}\,\epsilon_\theta^{(t)}(\boldsymbol{x}_t)}{\sqrt{\alpha_t}} \right)}_{\text{“ predicted } \boldsymbol{x}_0\text{”}}$$

Consistency Models

❖ **PF - ODE Solving and the Denoiser Output**

- Denoiser output 에 대한 특징 : Large Timestep 에 대해서는 성능이 낮지만, Small Timesteps 에선 성능이 좋음

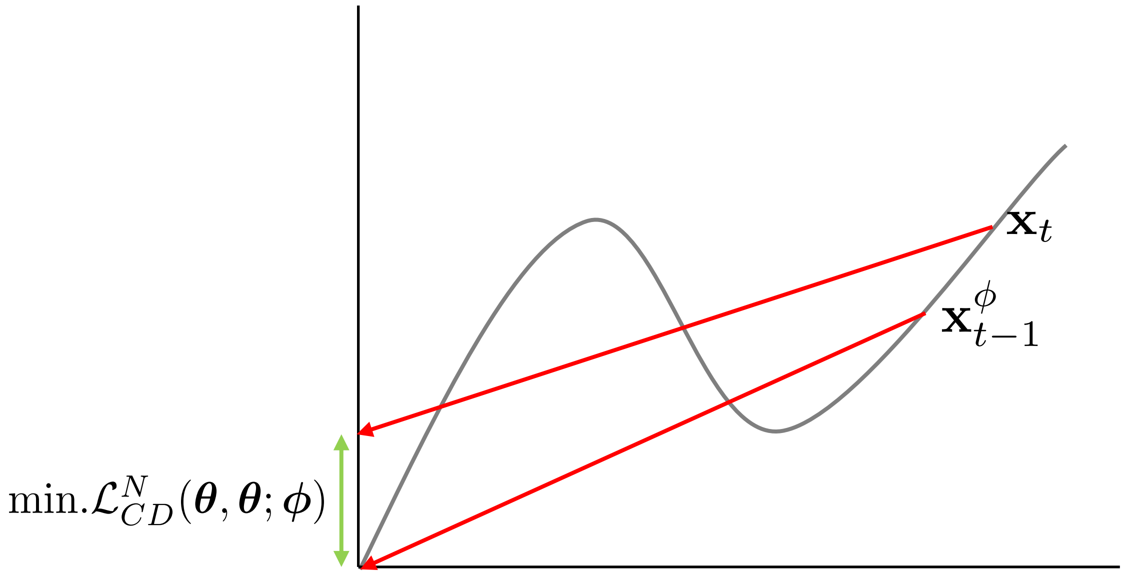




Consistency Models

❖ **Consistency Models**

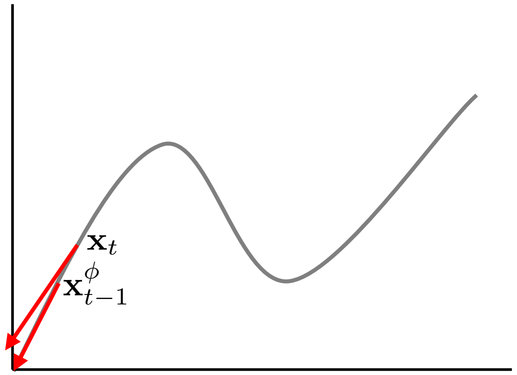
- Denoiser output 에 대한 특징 : Large Timestep 에 대해서는 성능이 낮지만, Small Timesteps 에선 성능이 좋음
- 이러한 특징을 이용한 것이 Consistency Models : Diffusion Model 이 PF – ODE 를 따라간다는 점과 Denoiser Output 또한 구할 수 있다는 점을 활용 기존 모델을 Distillation 하는 방법론 : 1 Step Generation
- 동일한 PF – ODE 위 adjacent RV 의 Denoiser Output 이 가까워 지도록 학습

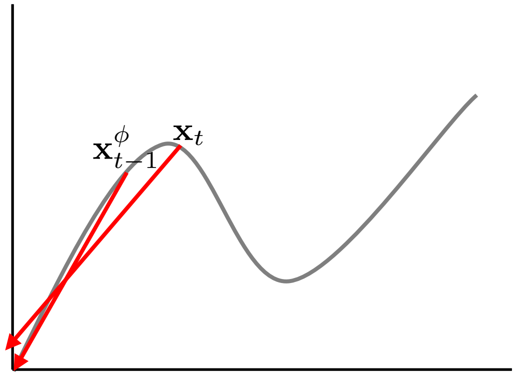


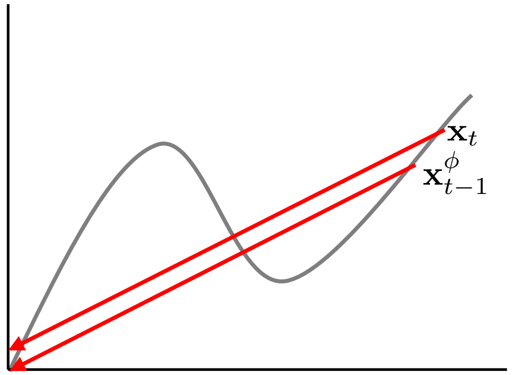
Consistency Models

❖ Consistency Models : Consistency Distillation

- 먼저 Training 이미지를 Forward Process 를 통해 x_t 까지 Noise
- Teacher Model 이 x_t 에서 One step PF-ODE Solving 진행 : x_{t-1}
- Student Model 은 두 RV 에 대한 Denoiser output 예측
- 두 Denoiser Output 에 대한 distance 를 최소화 하도록 Student Model 학습
- **즉, 어느 시점에서 Denoiser output 을 구해도 잘 생성하도록 학습하는 것**



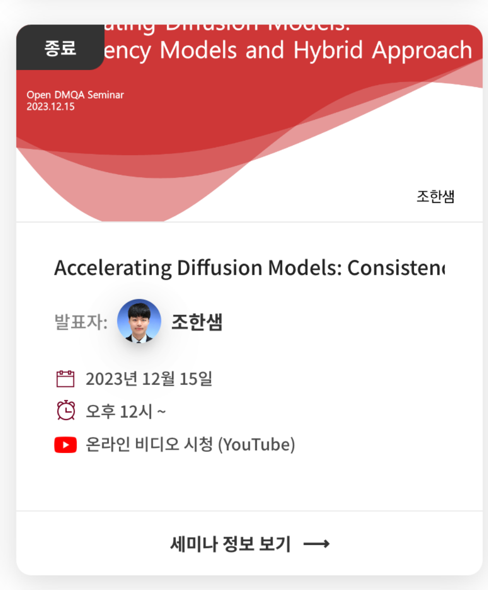


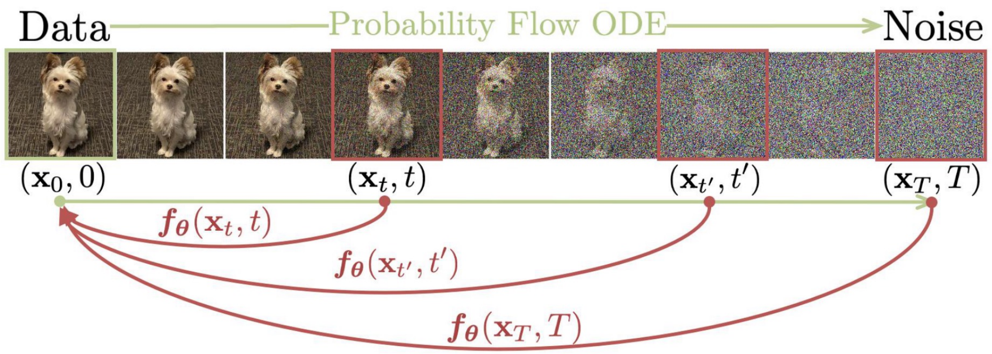


Consistency Models

❖ **Consistency Models : Consistency Distillation**

- 학습이 완료된 Consistency Model 은 1 Step Generation 을 수행 할 수 있게 됨
- PF-ODE Trajectory 위 어떤 Random Variable 의 Denoiser Output 이 성능이 좋기 때문





CM 1 Step

Conclusion

❖ **VP, VE Formulation**

- VP, VE Formulation 의 차이, 기원
- Diffusion Models 와 Score Based Models 의 관계

❖ **Score SDE**

- Diffusion Models 와 Score Based Models 의 Unified Framework
- SDEs, PF-ODEs

❖ **Consistency Models**

- PF-ODE 와 Denoiser 를 활용한 1 Step Generation 방법론

“Photorealistic image of a cute cat wearing a bachelor's cap, holding a sign that says 'Thank You!’”