Stochastic Differential Equations for Generative Modeling

Georgiy Malaniya, Anastasia Archangelskaya, Nikolay Kalmykov, Ole Shepelin, Vadim Shirokinsky (Project 5, team 19)

We would like to thank Anton Bolychev for contribution in this work

TASK

I. Reproduce results for Score-based Generative Modelling through stochastic differential equations with CIFAR dataset

II. Train GANs as energy based models (EBM-GAN) — train with comparable architectures

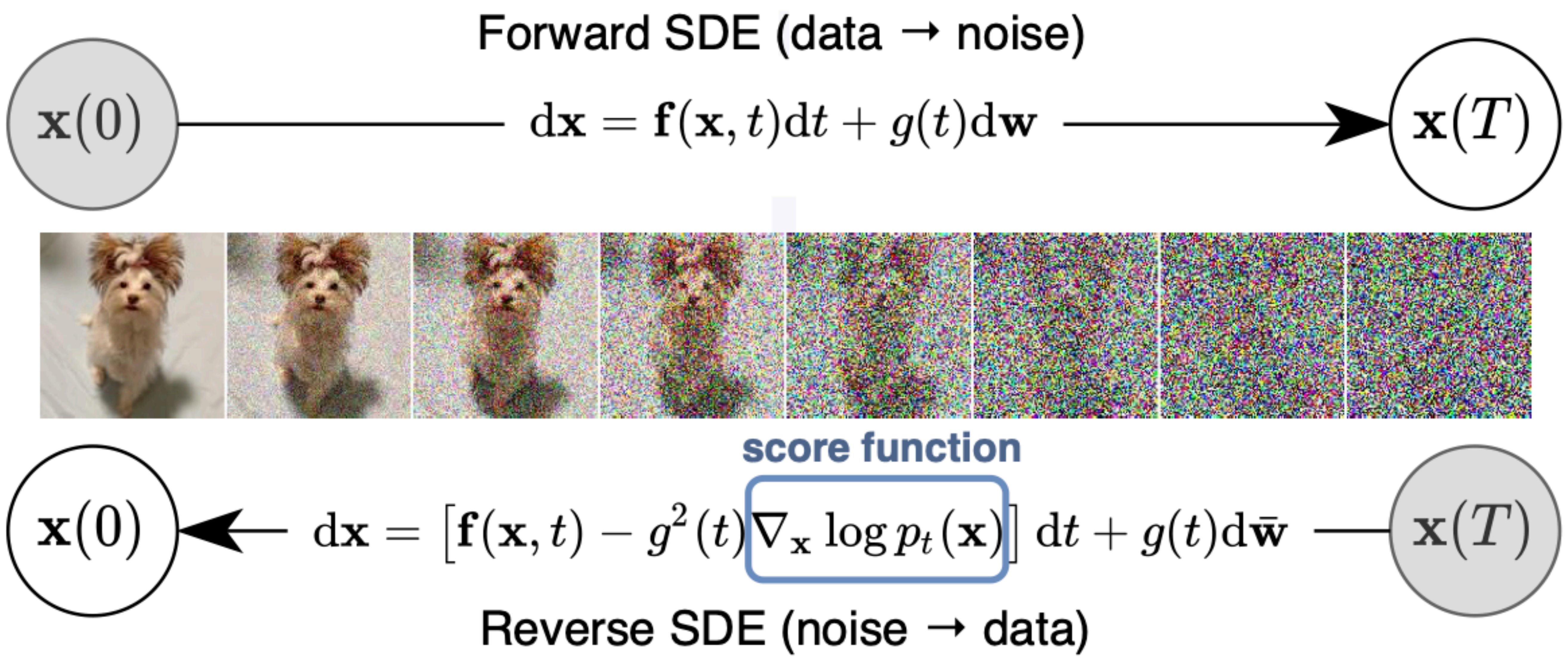
III. Compare DDPM/SMLD experiments with EBM-GAN

PROBLEM TO SOLVE

- **sampling techniques are inefficient**
- **decreased quality of image generation**

Task 1 DDPM/SMLD reproduction

SCORE-BASED GENERATIVE MODELING THROUGH STOCHASTIC DIFFERENTIAL EQUATIONS with CIFAR



01.

Forward SDE

Transforming data to a simple noise distribution with a continuous-time SDE

02.

Reverse SDE

Generating images from the noise based on denoising diffusion probabilistic models (DDPM)

Source: https://openreview.net/pdf?id=PxTIG12RRHS

Background & approaches

1st approach

Denoising score matching with Langevin dynamics (SMLD)

2nd approach

Denoising diffusion probabalistic models

NOISE CONDITIONAL SCORE

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \boxed{\sigma_i^2} \mathbb{E}_{p_{\text{data}}(\mathbf{x})} \mathbb{E}_{p_{\sigma_i}(\tilde{\mathbf{x}}|\mathbf{x})} \big[\| \mathbf{s}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}}, \sigma_i) - \nabla_{\tilde{\mathbf{x}}} \log p_{\sigma_i}(\tilde{\mathbf{x}} \mid \mathbf{x}) \|_2^2 \big].$$

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \boxed{(1-\alpha_i)} \mathbb{E}_{p_{\text{data}}(\mathbf{x})} \mathbb{E}_{p_{\alpha_i}(\tilde{\mathbf{x}}|\mathbf{x})} [\| \mathbf{s}_{\boldsymbol{\theta}}(\tilde{\mathbf{x}}, i) - \nabla_{\tilde{\mathbf{x}}} \log p_{\alpha_i}(\tilde{\mathbf{x}} \mid \mathbf{x}) \|_2^2].$$

SAMPLES GENERATION

$$\mathbf{x}_i^m = \mathbf{x}_i^{m-1} + \epsilon_i \mathbf{s}_{\boldsymbol{\theta}^*}(\mathbf{x}_i^{m-1}, \sigma_i) + \sqrt{2\epsilon_i}\mathbf{z}_i^m, \quad m = 1, 2, \cdots, M,$$

$$\mathbf{x}_{i-1} = \frac{1}{\sqrt{1-\beta_i}}(\mathbf{x}_i + \beta_i \mathbf{s}_{\boldsymbol{\theta}^*}(\mathbf{x}_i, i)) + \sqrt{\beta_i}\mathbf{z}_i, \quad i = N, N-1, \cdots, 1.$$

$$\mathbf{x}_N \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Source: https://openreview.net/pdf?id=PxTIG12RRHS

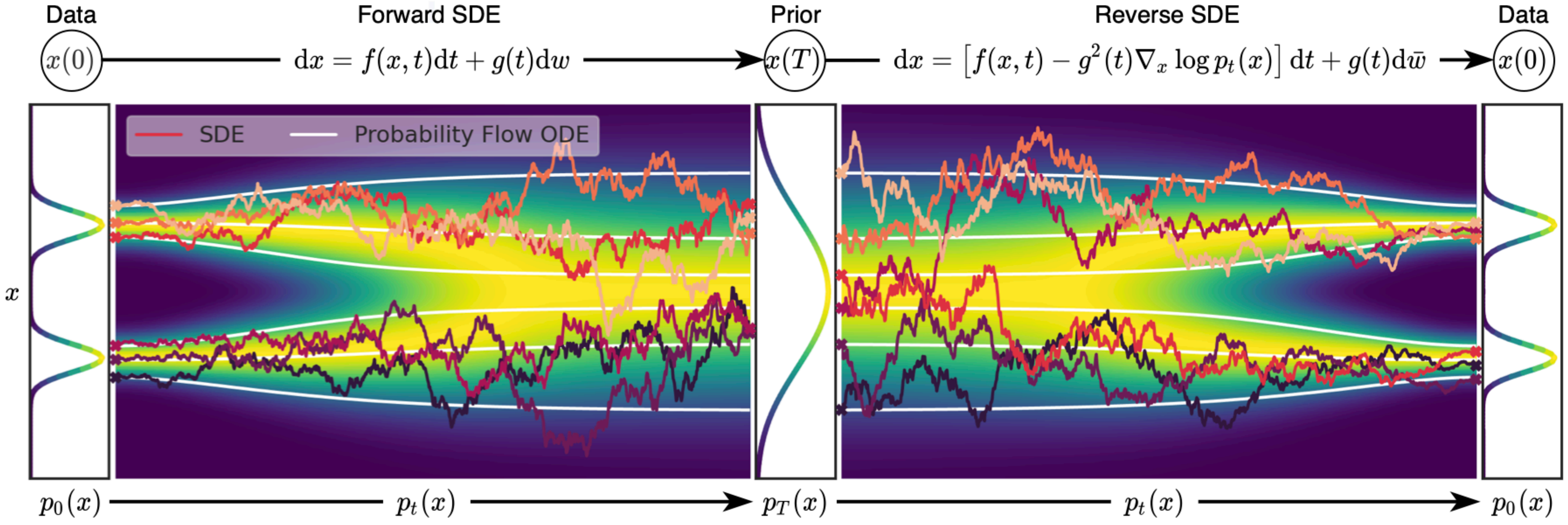
DDPM / SMLD reproduction solution

01. Noise distribution

data mapping with SDE

02. Generative modelling

back reverse SDE



Source: https://openreview.net/pdf?id=PxTIG12RRHS

Combination of approaches

The noise perturbations used in SMLD and DDPM can be regarded as discretizations of two different

1st approach

$$\mathrm{d}\mathbf{x} = \sqrt{\frac{\mathrm{d}\left[\sigma^2(t)\right]}{\mathrm{d}t}}\mathrm{d}\mathbf{w}.$$

2nd approach

$$\mathrm{d}\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}\,\mathrm{d}t + \sqrt{\beta(t)}\,\mathrm{d}\mathbf{w}. \qquad N \to \infty,$$

$$\mathrm{d}\mathbf{x} = -\frac{1}{2}\beta(t)\mathbf{x}\,\mathrm{d}t + \sqrt{\beta(t)(1 - e^{-2\int_0^t \beta(s)\mathrm{d}s})}\mathrm{d}\mathbf{w}.$$

perform particularly well on likelihoods

Source: https://openreview.net/pdf?id=PxTIG12RRHS

Research Task

GANs as energy based models (EBM-GAN)

Task

Implement Lagevin dynamics for generation from latent noise space

Goals:

- Implement Langevin dynamics
- Evaluate it using Fréchet Inception Distance

Solution

Implemented computation of energy for pre-trained DCGAN

Source: https://arxiv.org/pdf/2003.06060.pdf

Research Task Conducted experiments (EBM-GAN)

Steps

- **Langevin dynamics** is implemented into the pre-trained GAN

$$E(z) = -\log p_0(z) - d(G(z))$$

$$x_{i+1} = x_i - \frac{\epsilon}{2}\nabla_x E(x) + \sqrt{\epsilon}n, n \sim N(0, I)$$

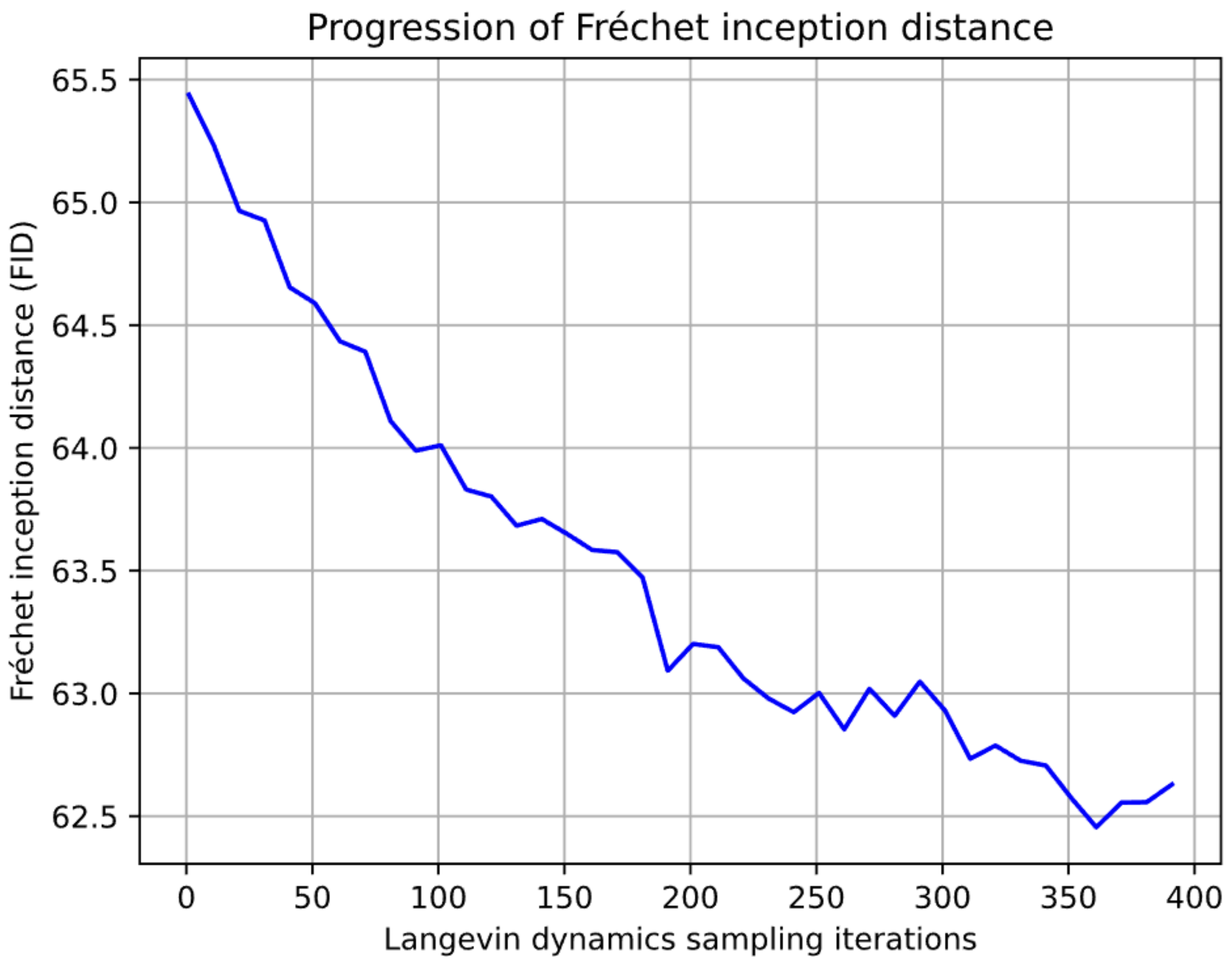
- Obtained results were tasted with Fréchet inception distance

$$d_F(\mathcal{N}(\mu, \Sigma), \mathcal{N}(\mu', \Sigma'))^2 = \|\mu - \mu'\|_2^2 + \operatorname{tr}\left(\Sigma + \Sigma' - 2\left(\Sigma^{\frac{1}{2}} \cdot \Sigma' \cdot \Sigma^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)$$

Results

- Langevin dynamics was implemented
- Expected behaviour of FID curve was observed

Github repo with code and experiments



Source: https://arxiv.org/pdf/2003.06060.pdf

Replication Task

Langevin dynamics for score-based models

Task

Replicate and compare results for predictor and predictor-corrector sampling approaches using different SDE solvers

Solution

Build a pipeline for ancestral sampling and reverse diffusion probability flow

Source: https://arxiv.org/pdf/2003.06060.pdf

Thank you for you attention!

Georgiy Malaniya, Anastasia Archangelskaya, Nikolay Kalmykov, Ole Shepelin, Vadim Shirokinsky (Project 5, team 19)