# Intern Project

Intern Name

**Abstract**

# Contents

# 1 Second-order Algorithm

Assume we define a field with $x_t = \alpha_t x_0 + \beta_t x_1$, we can manually calculate $\dot{x}_t$ and $\ddot{x}$, representing the first-order and second-order gradient, respectively.

**Definition 1.1.** *The loss function for second order method contain two parts. We define the first part which is trying to using $\dot{x}_t$, $x_t$ and $t$ to learn function $u_{1,t}$, thus the loss is*

$$L_{2,1,\theta_1} := \|\dot{x}_t - u_{1,\theta_1}(x_t, t)\|_2^2 \tag{1}$$

*Next, we define the second part which is trying to use $\ddot{x}_t, u_{1,\theta_1}(x_t, t)$, $x_t$ and $t$ to learn $u_{2,\theta_2}$ function, thus the loss is*

$$L_{2,2,\theta_2,\theta_1} := \|\ddot{x}_t - u_{2,\theta_2}(u_{1,\theta_1}(x_t, t), x_t, t)\|_2^2 \tag{2}$$

*Overall the total loss is*

$$L_{2,\theta} := L_{2,1,\theta_1,\theta_2} + L_{2,2,\theta_2} \tag{3}$$

**Algorithm 1**

**procedure** OurSecondOrderAlg()
    **for** each iteration **do**
        Random sample $x_0$ and time $t$, with target $x_1$
        $x_t \leftarrow \alpha_t \cdot x_0 + \sqrt{1 - \alpha_t^2} \cdot x_1$
        Compute gradient with respect to $L_{2,\theta}$ ▷ see Def. 1.1
    **end for**
    **return** $u_1, u_2$ ▷ Two network functions
**end procedure**
/* Below is an inference algorithm that only use first-order learner $u_1$*/
**procedure** Inf1($u_1$)
    $x_0 \sim \mathcal{N}(0, 1)$
    Initial $x \leftarrow x_0$
    **for** $t$ from 0 to 1 with step $\Delta t = 0.01$ **do**
        $x \leftarrow x + \Delta t \cdot u_1(x, t)$
    **end for**
    **return** $x$
**end procedure**

**Algorithm 2** Second Order Inference Algorithm 3

```
/* Option 3 */
procedure Inf2ndOrderAlg3(u_1, u_2)
    x_0 ~ N(0, 1)
    Initial x ← x_0
    for t from 0 to 1 with step Δt = 0.01 do
        x ← x + Δt · u_1(x, t) + ((Δt)^2 / 2) · u_2(u_1(x, t), x, t)
    end for
    return x
end procedure
```

## 2 Experimental Details

**IMPORTANT!** Please modify this colab for the code task.

### 2.1 Stage 1 instructions

We would like to implement the second order version VP ODE framework from [LGL22], the algorithm is written in Algorithm 2.

Recall that we have defined the original trajectory as

$$z_t = \alpha_t z_0 + \beta_t z_1$$

in VP ODE, we choose $\alpha_t = \exp(-\frac{1}{4}a(1-t)^2 - \frac{1}{2}b(1-t))$ and $\beta_t = \sqrt{1-\alpha_t^2}$, with hyperparameters $a = 19.9$ and $b = 0.1$.

We calculate the first order ground truth by

$$\frac{\mathrm{d}z_t}{\mathrm{d}t} = \frac{\mathrm{d}\alpha_t}{\mathrm{d}t} z_0 + \frac{\mathrm{d}\beta_t}{\mathrm{d}t} z_1$$

We calculate the second order ground truth by

$$\frac{\mathrm{d}^2 z_t}{\mathrm{d}t^2} = \frac{\mathrm{d}^2\alpha_t}{\mathrm{d}t^2} z_0 + \frac{\mathrm{d}^2\beta_t}{\mathrm{d}t^2} z_1$$

The first-order derivatives are given by

$$\frac{d\alpha_t}{dt} = \alpha_t \cdot \frac{1}{2}(a(1-t)+b)$$

and

$$\frac{d\beta_t}{dt} = -\frac{\alpha_t}{\sqrt{1-\alpha_t^2}} \cdot \frac{d\alpha_t}{dt}$$

The second-order derivatives are

$$\frac{d^2\alpha_t}{dt^2} = \frac{1}{2}(\alpha_t \cdot (a(1-t)+b)^2 - a \cdot \alpha_t)$$

and

$$\frac{d^2\beta_t}{dt^2} = -\frac{1}{(1-\alpha_t^2)\sqrt{1-\alpha_t^2}} \cdot \frac{d\alpha_t}{dt} + \frac{d\beta_t}{dt} \cdot \frac{d^2\alpha_t}{dt^2}$$

The **loss** should be calculate as (3)

## 2.2 Stage 1 expected results

You should get a result similar to Figure 1.

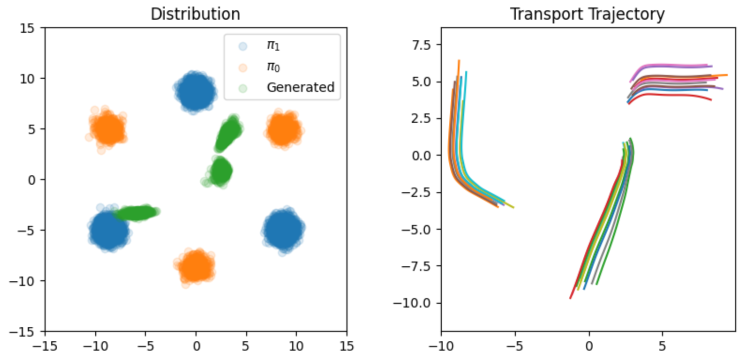


Figure 1: Stage 1 expected results.

## 2.3 Stage 2 instructions

Please observe the loss curve of first order loss (Eq.(1)) and second order loss (Eq. (2)).

It can be easily find that second order loss has a greater magnitude than first order loss. However, we expect first and second order loss should have the same magnitude. So we change the total loss as

$$L_{2,\theta} := \lambda_1 \cdot L_{2,1,\theta_1,\theta_2} + \lambda_2 \cdot L_{2,2,\theta_2}$$

For simplicity, we would like to fix $\lambda_1$ here to 1. Please find the correct $\lambda_2$ to make sure that $\lambda_1 \cdot L_{2,1,\theta_1,\theta_2}$ and $\lambda_2 \cdot L_{2,2,\theta_2}$ are in the same magnitude.

## 2.4 Stage 2 expected results

If you choose the correct $\lambda_2$, you will probably get result like Figure 2.

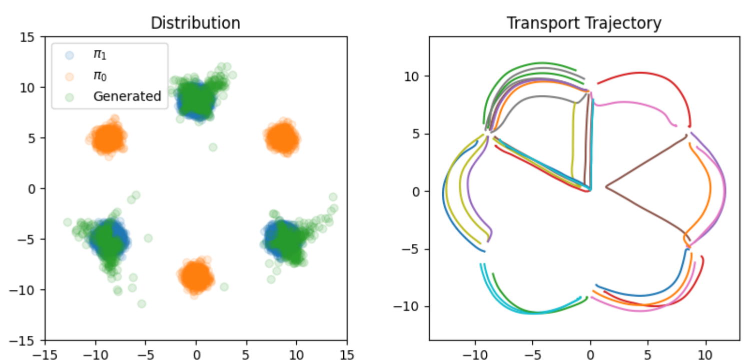


Figure 2: Stage 2 expected results

## References

[LGL22] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.