Correlational Lagrangian Schrödinger Bridge: Learning Dynamics with Population-Level Regularization

Yuning You[1], Ruida Zhou[2], Yang Shen[1]

[1]Texas A&M University , [2]University of California, Los Angeles

➢ TL;DR

- ❖ A **brand-new training objective** for diffusion generative models
- ❖ termed as population regularization – to enforce the **conservativeness in population** statistics.
- ❖ We name the pipeline as Correlational Lagrangian Schrödinger Bridge (**CLSB**).

Individual-level regularization:

$$\min_{(\pi_t)_{t\in[0,1]}} \int \overbrace{\mathbb{E}_{\pi_t}[}^{\text{Application to all particles}} \overbrace{\left|\frac{\mathrm{d}}{\mathrm{d}t}\underbrace{h(\mathbf{x})}_{\text{Individual state}}\right|^2}^{\text{Regularization}}]\mathrm{d}t, \ \text{s.t.} \ \overbrace{\pi_0=\hat{p}_0, \pi_1=\hat{p}_1}^{\text{Data fitting}}$$

Population-level regularization:

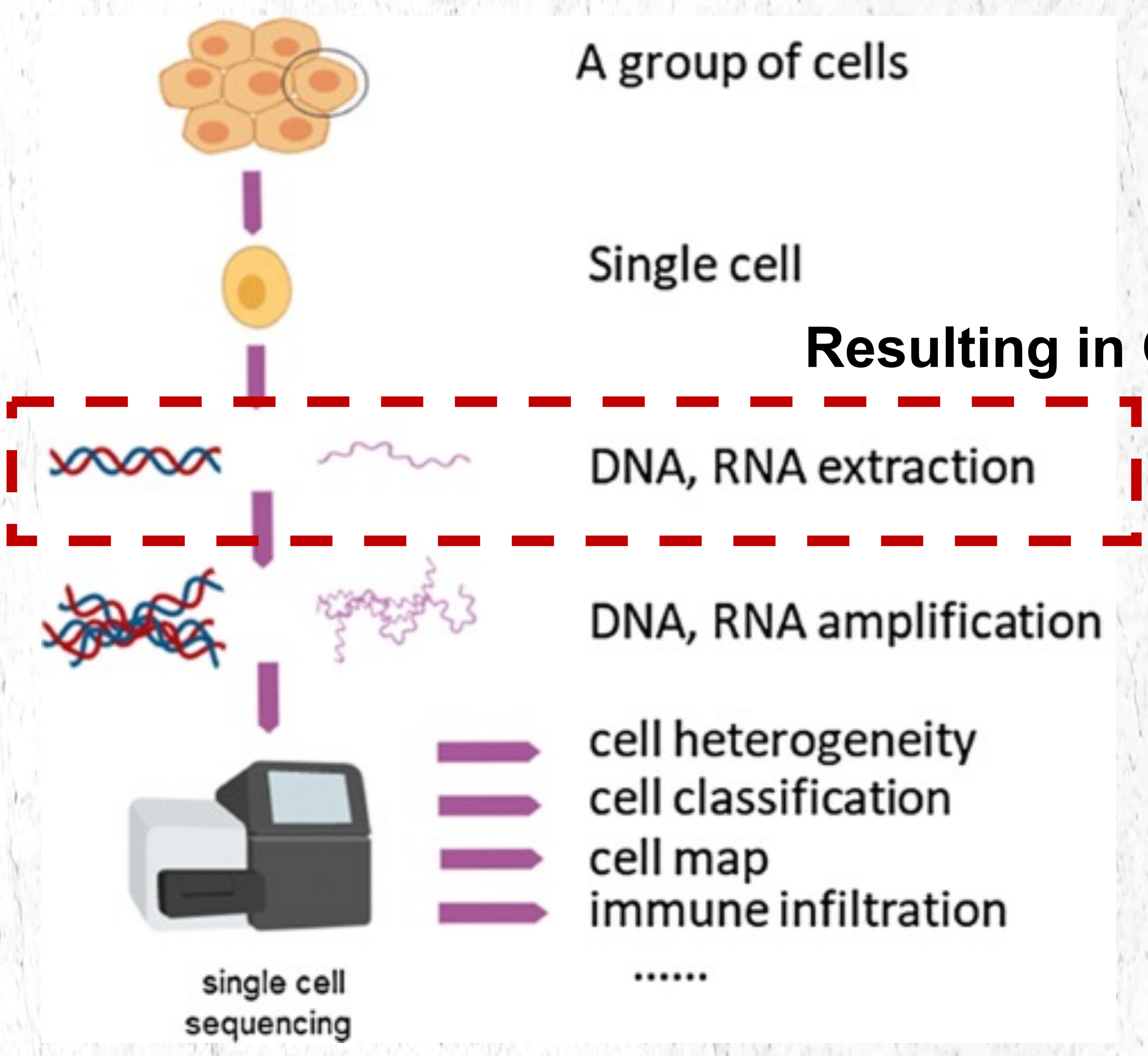
$$\min_{(\pi_t)_{t\in[0,1]}} \int \overbrace{\left|\frac{\mathrm{d}}{\mathrm{d}t}\underbrace{\mathbb{E}_{\pi_t}[h(\mathbf{x})]}_{\text{Population state}}\right|^2}^{\text{Regularization}} \mathrm{d}t, \ \text{s.t.} \ \overbrace{\pi_0=\hat{p}_0, \pi_1=\hat{p}_1}^{\text{Data fitting}}$$

$h(\cdot)$ is the domain-specific cost function.

➢ Background & Problem

- ❖ Data: **Cross-sectional** observations:
  - ❖ Data are sampled from unknown SDEs;
  - ❖ **Trajectories are not accessible!**
  - ❖ Populations at different time stamps are accessible.
- ❖ Motivating example: Single-cell sequencing data.
- ❖ Goal: Modeling the temporal evolution of the data.
- ❖ In formulation:

Data:	**Goal:**
$\{\boldsymbol{x}_0 \sim p_0\}, \{\boldsymbol{x}_1 \sim p_1\}$	Modeling
sampled from $(p_t)_{t\in[0,1]}$	$(p_t)_{t\in(0,1)}$

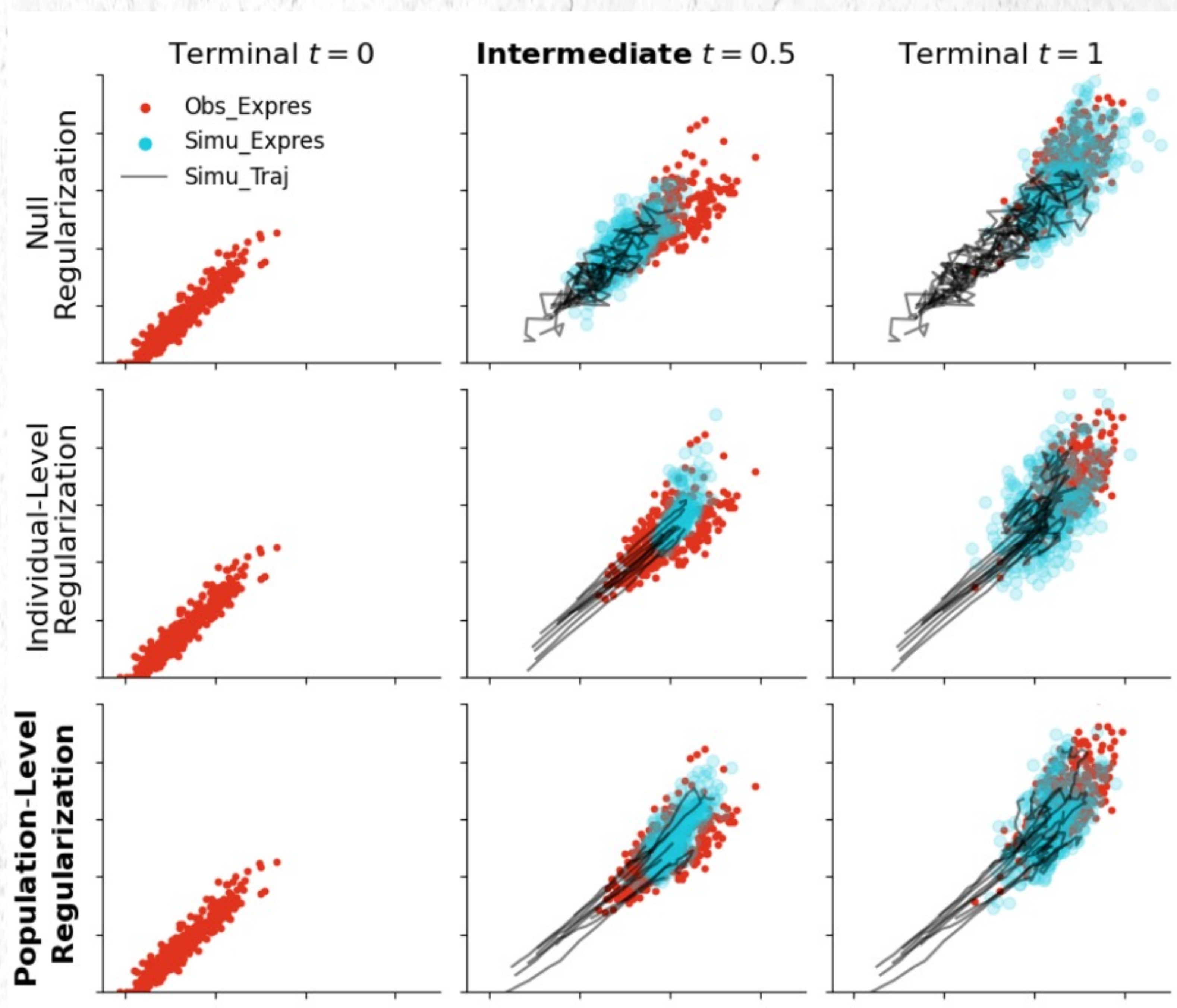


➢ Approach

- ❖ The CLSB pipeline:
  - ❖ A diffusion generative model to parametrize SDEs;
  - ❖ Optimizing models to generate samples to match the marginal observations at varied time stamps;
  - ❖ Regularizing the generated trajectories with priors.
- ❖ In formulation:

Method: Constructing $(\pi_t)_{t\in[0,1]}$ that
1. $\pi_1 = \hat{p}_1$ given $\pi_0 = \hat{p}_0$ (data fitting);
2. $(\pi_t)_{t\in[0,1]}$ adheres to certain criteria (regularization).

- ❖ Innovation: A novel regularization at the population level.
  - ❖ Existing approaches are referred as individual regularization – **Priors are enforced to individuals**.
  - ❖ We propose the novel population regularization – by switching the order of expectation and derivation,
  - ❖ to leverage the more effective and robust conservativeness prior at population – **Priors are enforced to distributions**.
- ❖ **New theoretical results** are provided on its analytical expression (please refer to main text Section 3.2).



➢ Experiments

- ❖ **Unconditional generation** on developmental modeling of embryonic stem cells;
- ❖ **Conditional generation** on dose-dependent cellular response prediction to perturbations (please refer to main text Section 4.2).

Methods	All-Step Prediction t_1	All-Step Prediction t_2 (Most Challenging)	All-Step Prediction t_3	One-Step Prediction t_1	One-Step Prediction t_2	One-Step Prediction t_3	**A.R.**
Random	1.873±0.014	2.082±0.011	1.867±0.011	1.870±0.013	2.084±0.010	1.868±0.012	10.0
SimpleAvg	1.670±0.019	1.801±0.014	1.749±0.016	1.872±0.014	2.085±0.011	1.868±0.012	9.3
OT-Flow	1.921	2.421	1.542	1.921	1.151	1.438	9.0
OT-Flow+OT	1.726	2.154	1.397	1.726	1.186	1.240	7.6
TrajectoryNet	1.774	1.888	1.076	1.774	1.178	1.315	6.8
TrajectoryNet+OT	1.134	1.336	**1.008**	1.134	1.151	1.132	3.6
DMSB	1.593	2.591	2.058	–	–	–	10.3
NeuralSDE	1.507±0.014	1.743±0.031	1.586±0.038	1.504±0.013	1.384±0.016	0.962±0.014	6.1
NLSB(E)	1.128±0.007	1.432±0.022	1.132±0.034	1.130±0.007	**1.099**±0.010	0.839±0.012	2.6
NLSB(E+D+V)	1.499±0.005	1.945±0.006	1.619±0.016	1.498±0.005	1.418±0.009	0.966±0.016	6.8
CLSB($\alpha_{\text{ind}} > 0$)	1.099±0.019	1.419±0.028	1.132±0.038	1.098±0.018	1.117±0.009	**0.826**±0.010	2.5
CLSB($\alpha_{\text{ind}} = 0$)	**1.074**±0.009	**1.244**±0.016	1.255±0.022	**1.095**±0.009	1.106±0.014	0.842±0.012	**2.1**