Samsung Research

Robust Generalized Schrödinger Bridge via Sparse Variational Gaussian Processes

Minyoung Kim

(mikim21@gmail.com)

Samsung AI Center Cambridge

ICLR'26 Rio de Janeiro, Brazil

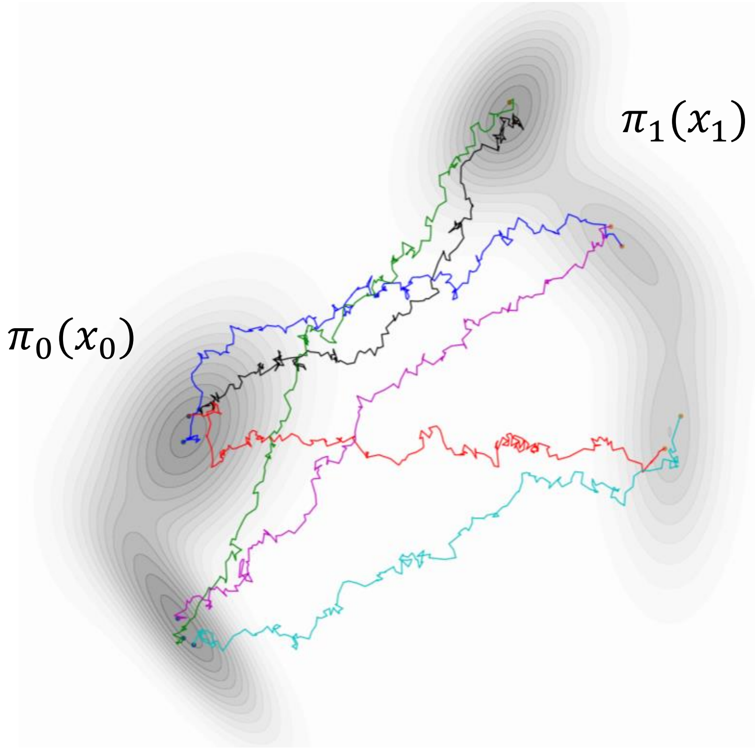
Bridge Problem, Schrödinger Bridge, and More

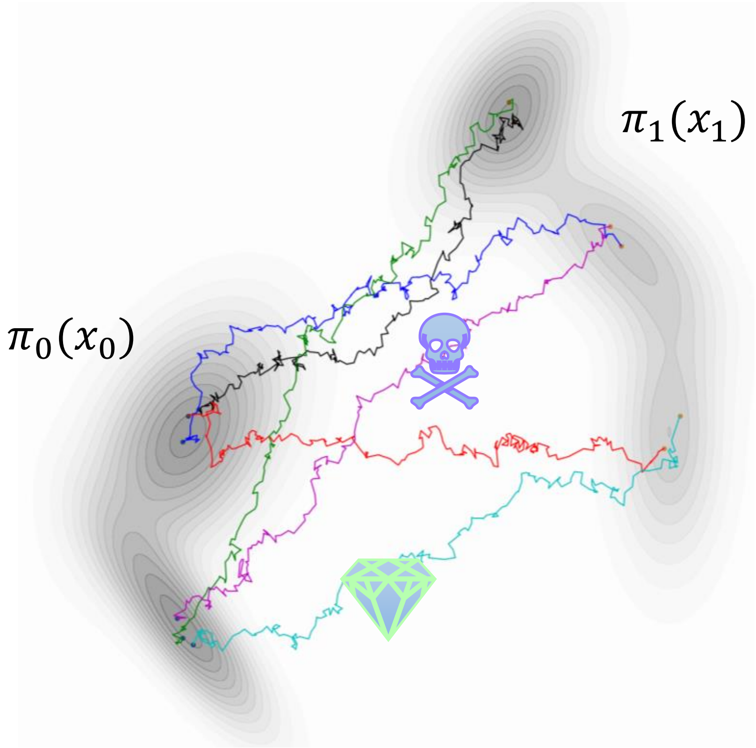
Bridge problem:
Find *any* (NNet) $v_t(x)$ s.t.
$P^v: dx_t = v_t(x_t)dt + \sigma dW_t\ , x_0 \sim \pi_0$
satisfies: $P_1^v(x_1) = \pi_1(x_1)$

Eg, Flow-matching is one way to solve it

Schrödinger bridge:
$\min KL(P^v||P^0)$ s.t.
P^v transports π_0 to π_1

Generalized Schrödinger bridge:
Additional obstacles to avoid, or preferred regions along the paths





GSBM[1] Algorithm

(Step 1)
Define coupling

$Q(x_0, x_1) = P^v(x_1|x_0)\pi_0(x_0)$ (from previous model $v = v_\theta$)

(Step 2)
Find pinned SDEs
for pinned marginals

"CondSOC"

$$\min_{\{P_t(x_t|x_0,x_1)\}_t} \int_0^1 \mathbb{E}_{P_t(x_t|x_0,x_1)}[||\alpha_t(x_t|x_0,x_1)||^2 + V_t(x_t)]dt$$

$$P_t(x|x_0,x_1) = \mathcal{N}(x;\mu_t,\gamma_t^2 I) \quad \alpha_t(x|x_0,x_1) = \dot{\mu}_t + (\frac{\dot{\gamma}_t}{\gamma_t} - \frac{\sigma^2}{2{\gamma_t}^2})(x-\mu_t)$$

Task-specific path cost expressed in $V_t(x_t)$ (Eg, no path cost ($V = 0$) makes DSBM[2])

(Step 3)
Reciprocal-Markovian
projection

$$\theta \leftarrow \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{t,Q(x_0,x_1)P_t(x_t|x_0,x_1)}||\alpha_t(x_t|x_0,x_1) - v_\theta(t,x_t)||^2$$

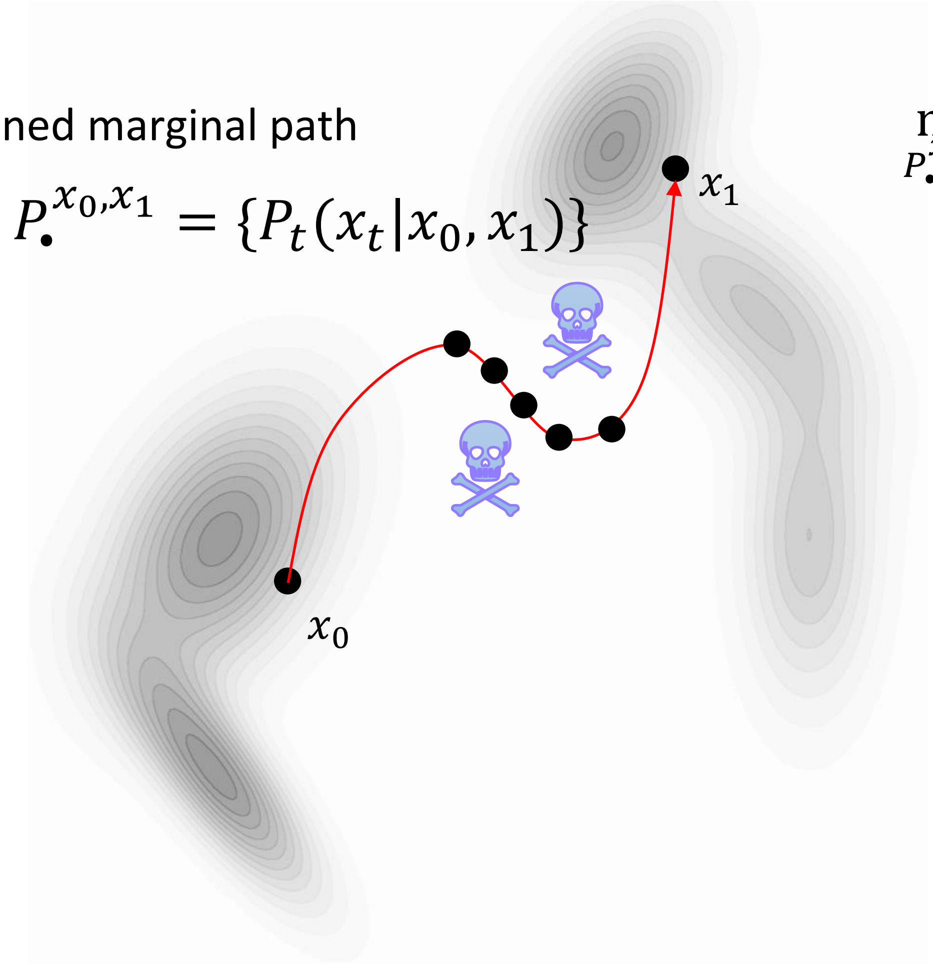
[1] Generalized Schrödinger Bridge Matching, Liu et al. ICLR 2024
[2] Diffusion Schrödinger Bridge Matching, Shi et al. NeurIPS 2023

CondSOC as Pinned Path Marginal Optimization

For each (x_0, x_1) ~current $Q(x_0, x_1)$

Pinned marginal path

$$P_\bullet^{x_0,x_1} = \{P_t(x_t|x_0, x_1)\}$$



(CondSOC optim = Find the best Gaussian pinned path marginal)

$$\min_{P_\bullet^{x_0,x_1}} J(P_\bullet^{x_0,x_1}; V_\bullet) = \int_0^1 \mathbb{E}_{P_t(x_t|x_0,x_1)}[||\alpha_t(x_t|x_0,x_1)||^2 + V_t(x_t)]dt$$

$\alpha_t(x_t|x_0, x_1)$ derived from $P_\bullet^{x_0,x_1}$ as:

$dx_t = \alpha_t(x_t|x_0, x_1)dt + \sigma dW_t$ has marginals $P_t^{x_0,x_1}(x_t)$

Known fact:

With $\alpha_t(x|x_0, x_1) = \dot{\mu}_t + (\frac{\dot{\gamma}_t}{\gamma_t} - \frac{\sigma^2}{2{\gamma_t}^2})(x - \mu_t)$,

the marginal becomes $P_t(x|x_0, x_1) = \mathcal{N}(x; \mu_t, \gamma_t^2 I)$

Motivation

- In GSBM, CondSOC optimization as "deterministic optimization”
  - Pinned marginal (μ_t, γ_t) parametrized by splines

- Possible extensions?
  - $V_t(x_t)$ is deterministic, noise-free. What if noisy, stochastic $V_t(x_t)$?
  - More flexible path modeling beyond splines?

- Our proposal
  - Gaussian process path modeling: (μ_t, γ_t) as random functions of time
  - GP can deal with uncertainty/noise in stage cost $V_t(x_t)$
  - Denoted by GP-GSBM

GP-GSBM: Key Idea

- Impose a GP prior on pinned path marginal $P_\bullet$ (ie, $GP(\mu_\bullet)$, $GP(\gamma_\bullet)$)
  - Express prior preference on $P_\bullet$ (eg, smooth, close to SB path)

- CondSOC objective as "likelihood" of path $P_\bullet$
  - How compatible a path $P_\bullet$ is with the stage penalty/preference $V_\bullet$

$$\underset{P_\bullet}{\operatorname{argmin}}\ J(P_\bullet; V_\bullet) - \tau \log \mathcal{P}_{prior}(P_\bullet) \equiv \underset{P_\bullet}{\operatorname{argmax}}\ \underbrace{\mathcal{P}_{prior}(P_\bullet)}_{Prior} \cdot \underbrace{\exp(-J(P_\bullet; V_\bullet)/\tau)}_{Likelihood}$$

- MAP -> Bayesian Posterior

$$\mathcal{P}_{post}(P_\bullet) \propto \mathcal{P}_{prior}(P_\bullet) \cdot \exp(-J(P_\bullet; V_\bullet)/\tau)$$

GP-GSBM: Variational Posterior and VI

- Sparse variational free-energy posterior model $\mathcal{Q}(P_\bullet) = \int \mathcal{Q}(P_Z)\mathcal{Q}(P_\bullet|P_Z)d\mu_Z$

$$\mathcal{Q}(\mu_\bullet) = \int \mathcal{Q}(\mu_Z)\mathcal{Q}(\mu_\bullet|\mu_Z)d\mu_Z, \quad \mathcal{Q}(\mu_Z) = \mathcal{N}(\mu_Z; C^\mu, S^\mu)$$

Inducing inputs & variables: $Z = (t_1, \ldots, t_n)$ $0 < t_1 < \cdots < t_n < 1$, $\mu_Z = [\mu_{t_1}, \ldots, \mu_{t_n}]^\top$

$\mathcal{Q}(\mu_\bullet)$ is GP:
$$\begin{cases} \mathbb{E}_\mathcal{Q}[\mu_t] = M_t + L_{t,Z}L_{Z,Z}^{-1}(C - M_Z) \\ \mathrm{Cov}_\mathcal{Q}(\mu_s, \mu_t) = L_{s,t} - L_{s,Z}L_{Z,Z}^{-1}SL_{Z,Z}^{-1}L_{Z,t} - L_{s,Z}L_{Z,Z}^{-1}L_{Z,t} \end{cases}$$

(Similarly for $\gamma_\bullet$)

- ELBO Learning: $\min_{\Lambda,\eta,\tau} \mathbb{E}_{\mathcal{Q}(P_\bullet)}[J(P_\bullet; V_\bullet)/\tau] + KL(\mathcal{Q}(P_\bullet)||\mathcal{P}_{prior}(P_\bullet))$

$\Lambda = \{C^\mu, S^\mu, C^\gamma, S^\gamma\}$ = variational params, η = prior/kernel params, τ = likelihood params

Our GP-GSBM Algorithm

Our GP-GSBM

(Step 1)

$$Q(x_0, x_1) = P^{v_\theta}(x_1|x_0)\pi_0(x_0)$$

(Step 2)

$$\min_{\Lambda,\eta,\tau} \mathbb{E}_{Q(P_\bullet)}[J(P_\bullet; V_\bullet)/\tau] + KL(Q(P_\bullet)||\mathcal{P}_{prior}(P_\bullet))$$

$$\alpha_t(x|x_0, x_1) = \dot{\mu}_t + (\frac{\dot{\gamma}_t}{\gamma_t} - \frac{\sigma^2}{2{\gamma_t}^2})(x - \mu_t),$$

Sample $\mu_\bullet, \gamma_\bullet \sim Q(P_\bullet)$

(Step 3)

$$\min_\theta \mathbb{E}_{\substack{t, Q(x_0, x_1) \\ x_t \sim \mathcal{N}(x; \mu_t, \gamma_t^2 I)}} ||\alpha_t(x_t|x_0, x_1) - v_\theta(t, x_t)||^2$$

GSBM[1]

$$Q(x_0, x_1) = P^{v_\theta}(x_1|x_0)\pi_0(x_0)$$

$$\min_{\mu_\bullet, \gamma_\bullet} J(P_\bullet; V_\bullet) = \int_0^1 \mathbb{E}[||\alpha_t(x_t|x_0, x_1)||^2 + V_t(x_t)]dt$$

$$\alpha_t(x|x_0, x_1) = \dot{\mu}_t + (\frac{\dot{\gamma}_t}{\gamma_t} - \frac{\sigma^2}{2{\gamma_t}^2})(x - \mu_t),$$

$$\min_\theta \mathbb{E}_{\substack{t, Q(x_0, x_1) \\ x_t \sim \mathcal{N}(x; \mu_t, \gamma_t^2 I)}} ||\alpha_t(x_t|x_0, x_1) - v_\theta(t, x_t)||^2$$

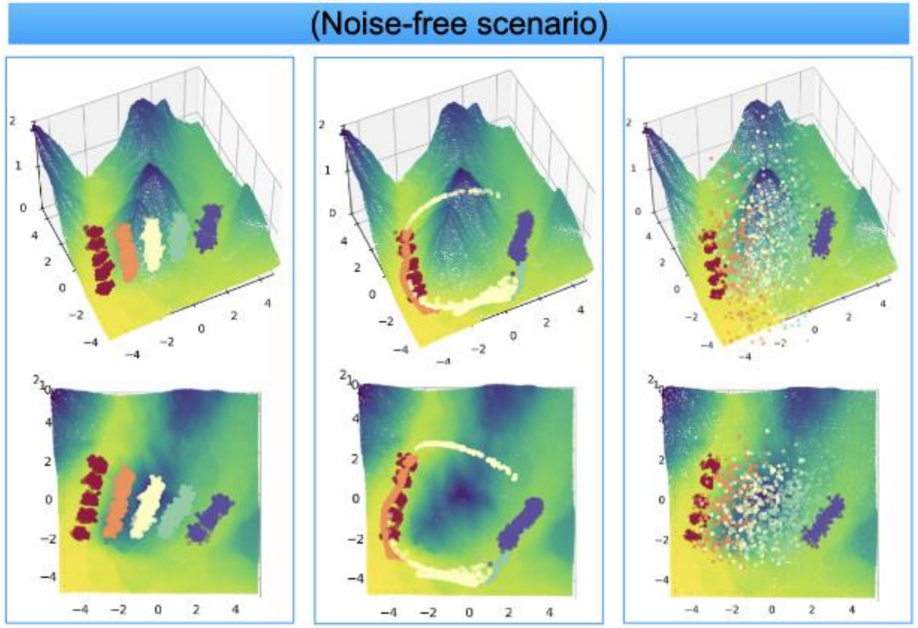
Experiments: LiDAR

Summary:
- Our GP-GSBM: treat $V_t(x)$ as noisy stochastic
- GSBM: treat $V_t(x)$ as noise-free deterministic
- DSBM: Ignores $V_t(x)$

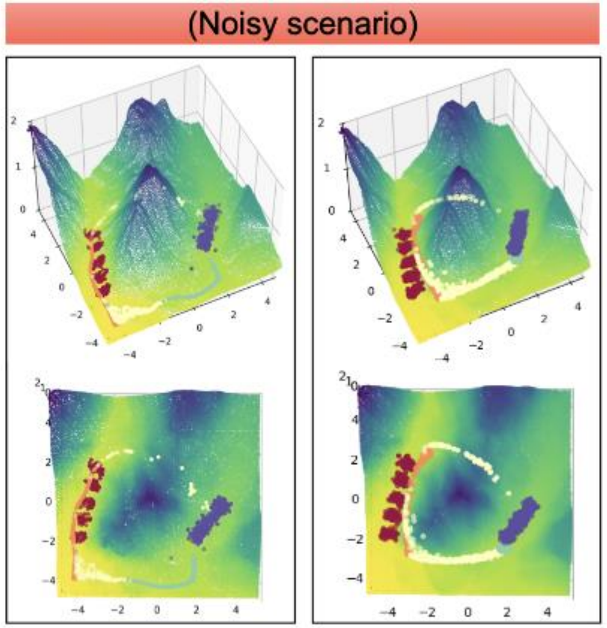
Goal: Transport red points $\pi_0(x_0)$ to blue ones $\pi_1(x_1)$, staying close to surfaces and as low altitudes as possible

$$V_t(x) = \|\Pi_M(x) - x\|^2 + \exp(\Pi_M(x)_{[z]})$$

Noisy setting:
Random noise injected to projected points



DSBM GP-GSBM (Ours) Stream-level GP



GSBM GP-GSBM (Ours)

CondSOC objective attained
$W_2(P_1, \pi_1)$ in parentheses

	DSBM	GSBM	GP-GSBM (Ours)
Noise-free obs.	7747.0 ± 76.4 (0.04)	6199.3 ± 47.3 (0.04)	**5925.0** ± 65.4 (0.03)
Noisy obs.	12686.9 ± 150.9 (0.04)	8506.1 ± 65.6 (0.04)	**8300.0** ± 67.6 (0.04)

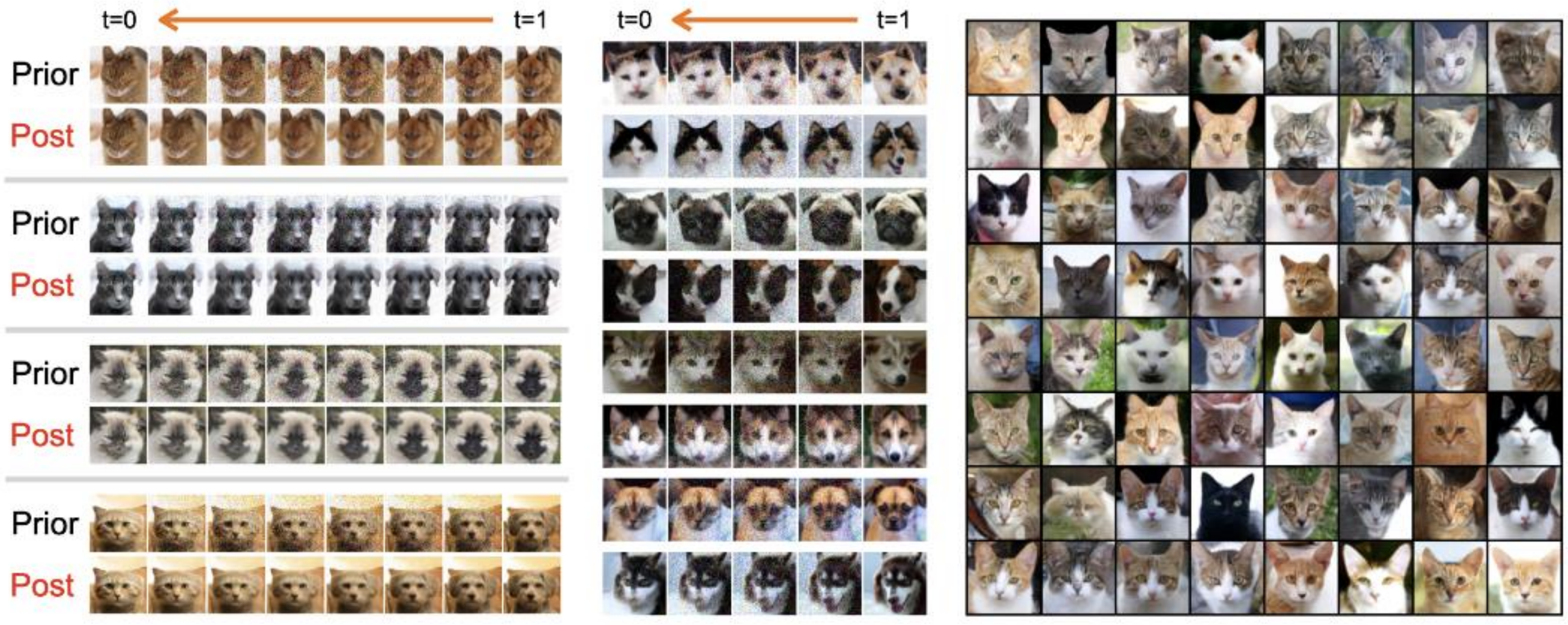
Experiments: Unpaired Img-to-Img Translation

Goal: Transport dog images $\pi_1(x_1)$ to cat images $\pi_0(x_0)$, while preserving smooth latent structures

$$V_t(x_t) = \|x_t - \text{dec}(z_t)\|_1, \quad z_t = \text{slerp}(t, \text{enc}(x_0), \text{enc}(x_1))$$

with a pre-trained VAE

Inherently noisy $V_t(x)$; Sources of stochasticity:
- Imperfect VAE model
- Slerp is only approximation to optimal path



(a) Prior and posterior of $P_t(x \mid x_0, x_1)$ (b) Generation by SDE (c) Generated images

FID scores

DSBM	GSBM
14.16	12.39

STREAM-LEVEL GP	GP-GSBM (OURS)
18.77	**10.21**

Conclusion

- A novel GP approach to pinned path marginal modelling for GSB
- Formulate CondSOC of GSBM into GP inference problem
- More flexible and robust solutions than GSBM
- Benefits & limitations
  - Robust to potential noise in stage costs
  - A principled way to deal with uncertainty in GSB problems
  - Higher computational complexity arising from GP inference (kernel inversion)