At the Intersection of Probabilistic Inference and Exploration Methods

Dinghuai Zhang 2022.12

Sampling from an unnormalized energy function

Given $p^*(x) \propto R(x) = \exp(-E(x))$

- Traditional: MCMC
- Amortized
  - Variational inference with probabilistic models (e.g., normalizing flows)
    - trained with KL
    - "mean-seeking" / "zero-avoiding" issues

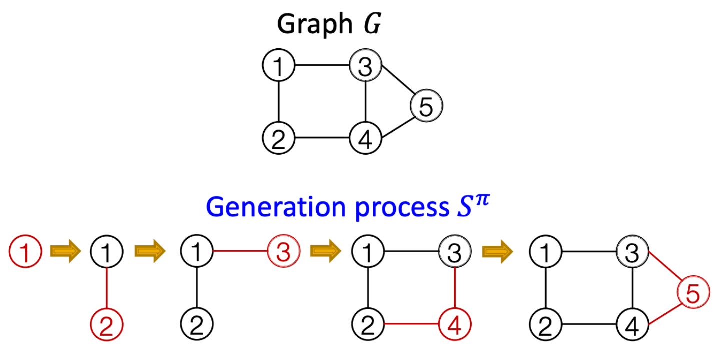
Divergence measures and message passing. Tom Minka, 2005.

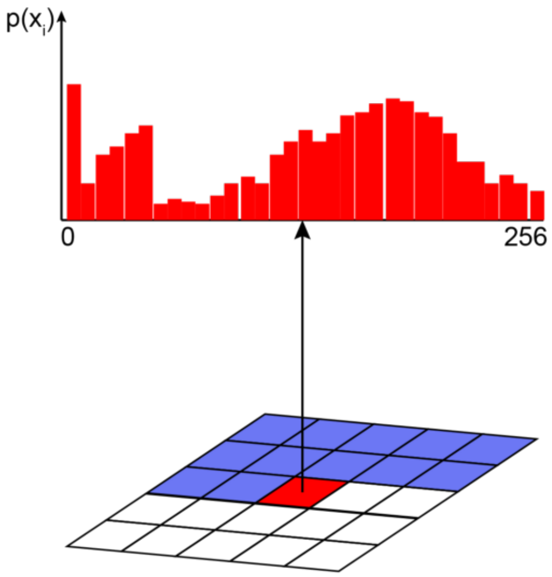
  - GFlowNets (generative flow networks)
    - treat sampling as a (sequential) decision-making process
    - + RL insights: exploration for probabilistic inference!

GFlowNets Foundation. Arxiv 2021.

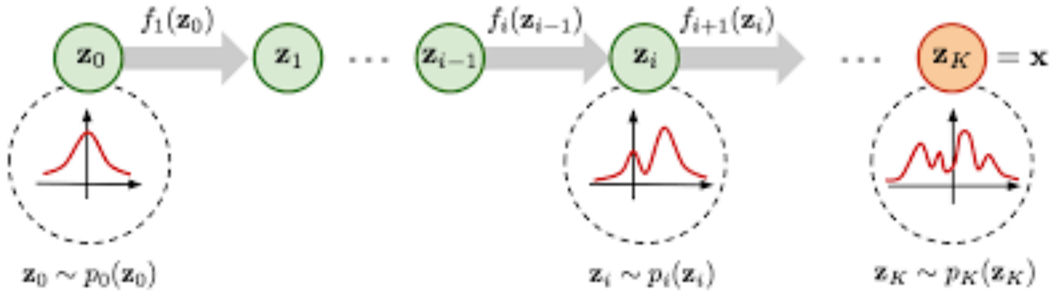
Examples of sequential sampling

Auto-regressive models

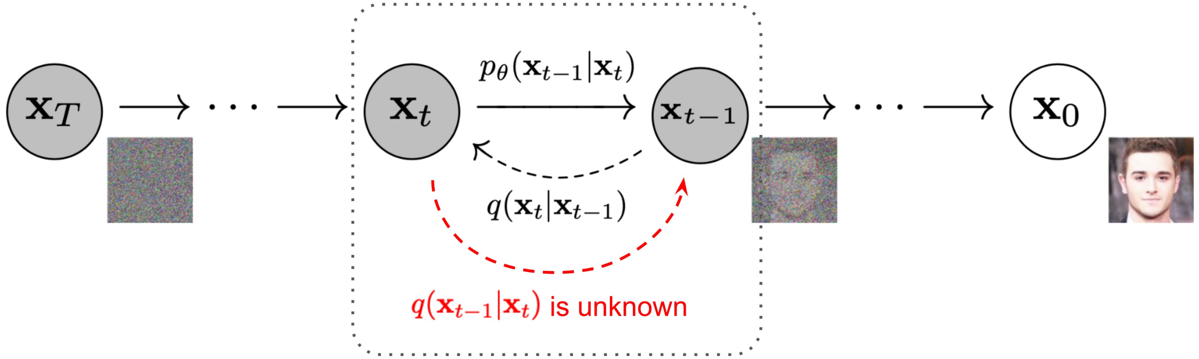




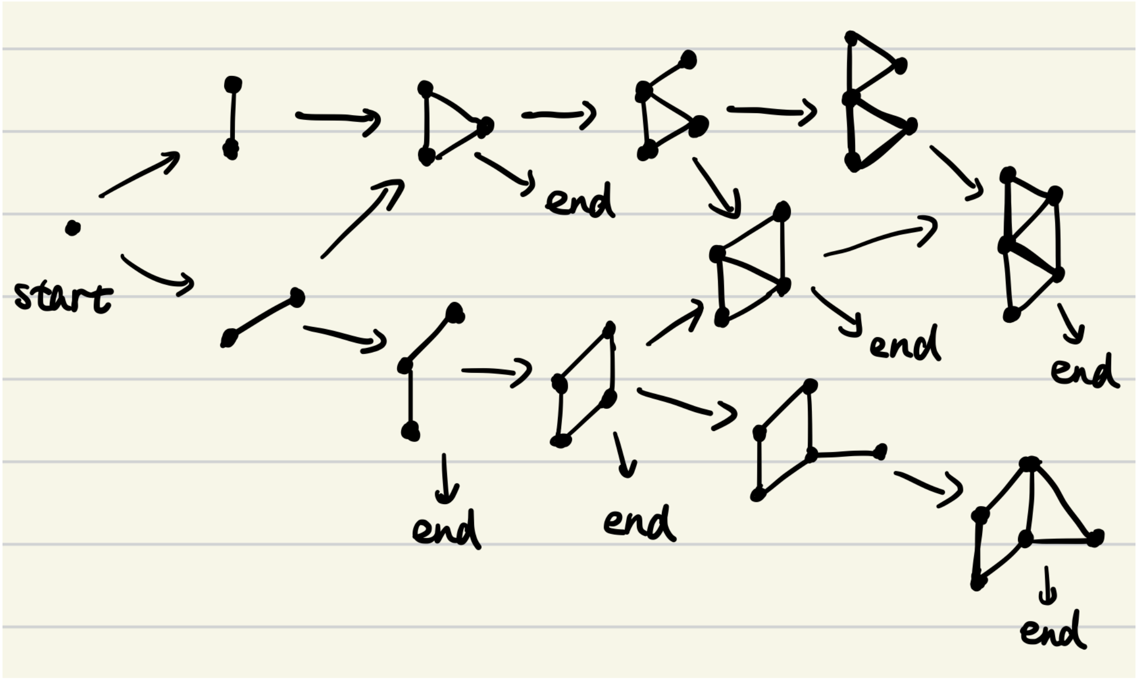
Normalizing flows

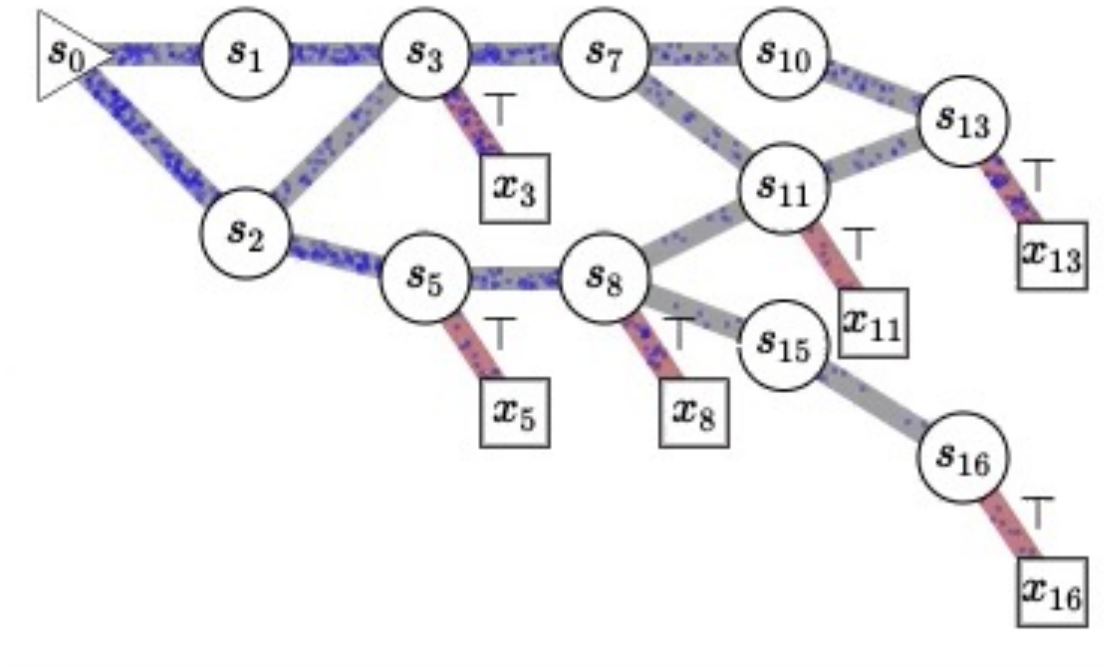


Diffusion models



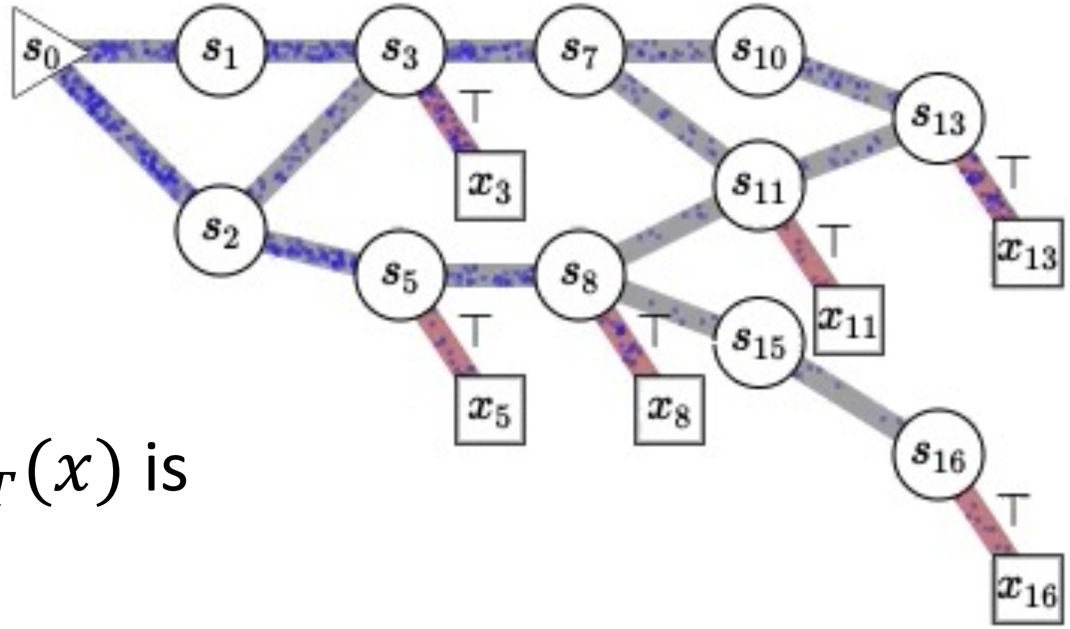
Abstract the graph generation example…





GFlowNets Basics

- To generate $x \in X$
- $\top$ = terminating action
- $P_T(x)$: terminating prob



- Goal: learn a GFlowNet such that $P_T(x)$ is proportional to given reward $R(x)$

$$R(\mathbf{x}) = \sum_{\tau=(\mathbf{s}_0 \to \ldots \to \mathbf{s}_n), \mathbf{s}_n = \mathbf{x}} F(\tau)$$

amount of “water” in τ

GFlowNets Foundation. Arxiv 2021.

Training criterion

- Parameterize the flow of edge / transition: $F(s \rightarrow s')$
  - Flow matching criterion / conservation law: ”in-flow” = “out-flow”
  - $\sum_s F(s \rightarrow s') = \sum_{s''} F(s' \rightarrow s'') + R(s')$
  - We can also define $F\left(s' \rightarrow s_f\right) = R(s')$
- Parameterize: $P_F(s'|s),\ P_B(s|s'), F(s)$
  - Forward and backward policy
  - $P_F(s'|s) = {F(s\rightarrow s')}/{F(s)}, P_B(s|s') = {F(s\rightarrow s')}/{F(s')}$
  - Detailed balance criterion: $F(s)P_F(s'|s) = F(s')P_B(s|s')$
- And others …

Regarding generative modeling

- GFlowNet is a general framework that includes most generative models as special cases
  - Hierarchical VAEs
  - Diffusion models
  - Auto-regressive models
  - Normalizing flows, …

Unifying Generative Models with GFlowNets. Arxiv 2022.

- … and could be further combined with energy-based learning to learn from data (set), rather than target density function

Energy-based model

$$p_{\phi}(\mathbf{x}) = \frac{1}{Z_{\phi}} \exp(-\mathcal{E}_{\phi}(\mathbf{x}))$$

- EBMs are usually trained with contrastive divergence (CD)

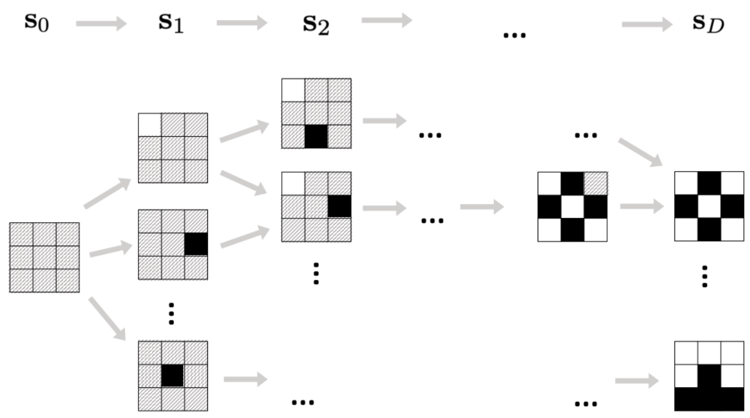
$$\begin{aligned} -\nabla_{\phi} \log p_{\phi}(\mathbf{x}) &= \nabla_{\phi} \mathcal{E}_{\phi}(\mathbf{x}) + \nabla_{\phi} \log Z_{\phi} \\ &= \nabla_{\phi} \mathcal{E}_{\phi}(\mathbf{x}) - \mathbb{E}_{\mathbf{x}' \sim p_{\phi}(\mathbf{x}')}[\nabla_{\phi} \mathcal{E}_{\phi}(\mathbf{x}')] \end{aligned}$$

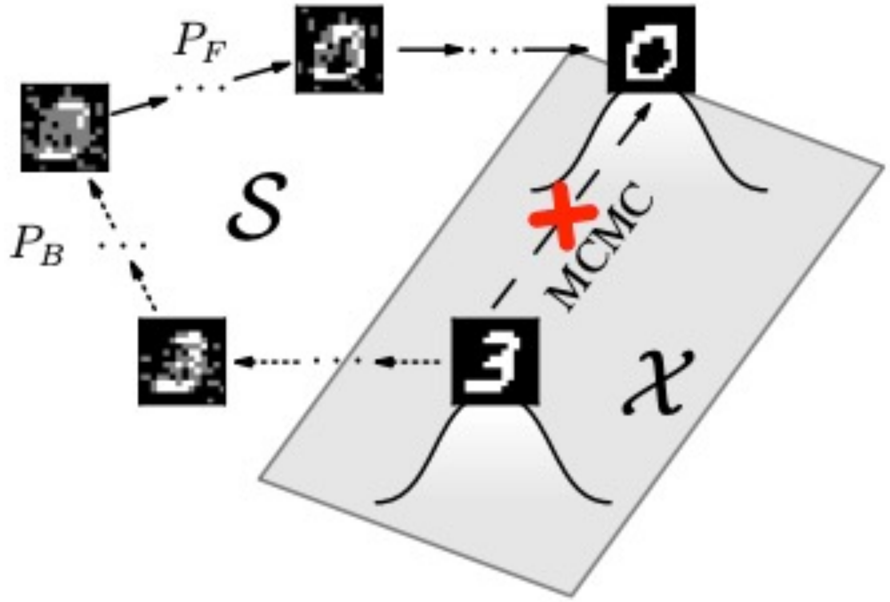
Simulated with truncated MCMC chains for negative samples

- This MCMC could be computationally expensive, and suffer from slow mixing under multi-modal settings.

Energy-based GFlowNets

- We propose to jointly train an EBM and a GFlowNet
  - EBM serves as the reward for GFlowNet
  - GFlowNet provides negative samples for CD-like training





Generative Flow Networks for Discrete Probabilistic Modeling. ICML 2022.

Exploration in Probabilistic inference

With such a “inference as control” framework, we could pour our RL expertises into probabilistic inference tasks…

- Policy = Sampler
  - Explore the target distribution landscape to cover all the modes
- Off-policy training
  - Training data do not necessary come from current model distribution
  - Amortized inference perspective
    - A special case of GFlowNet achieves the same expected gradient with standard variational inference
    - In general, GFlowNet provides additional off-policy learning capability

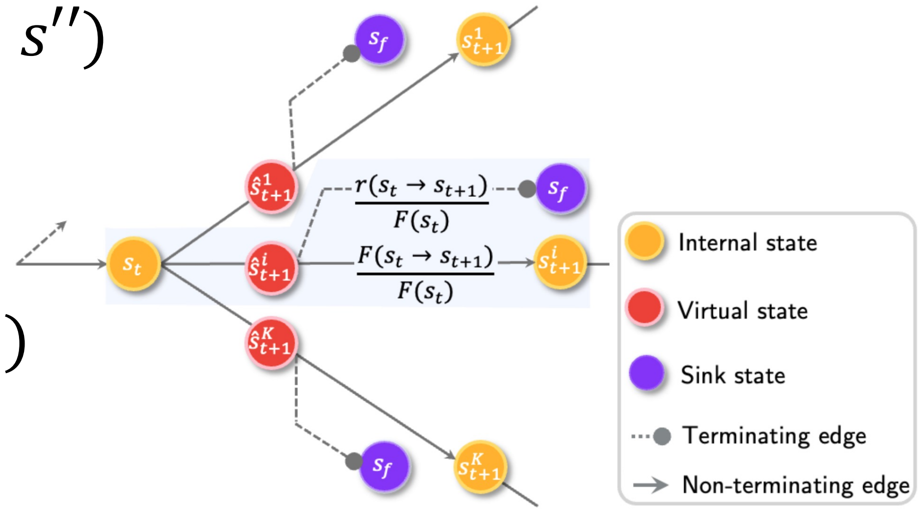
GFlowNets and Variational Inference. Arxiv 2022.

- Intrinsic exploration as intermediate reward
  - Add unsupervised RL reward into the “out-flow” of GFlowNets to encourage exploration

Generative Augmented Flow Networks, 2022.

Intermediate rewards

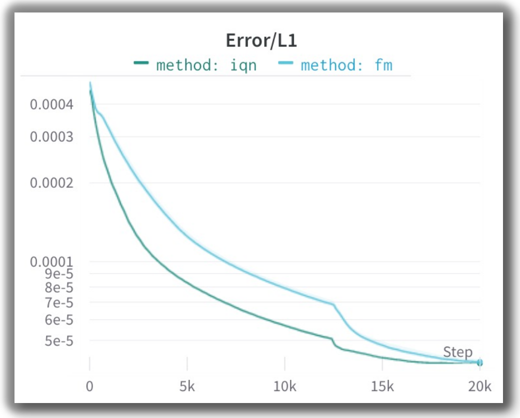
- Original: $\sum_{s} F(s \to s') = \sum_{s''} F(s' \to s'')$
- Augmented flow matching
  - $\sum_{s} F(s \to s') = \sum_{s''} F(s' \to s'') + r(s' \to s'')$
- Augmented detailed balance
  - $P_F(s'|s) = {}^{F(s \to s') + r(s \to s')}/_{F(s)}$
  - $F(s)P_F(s'|s) = F(s')P_B(s|s') + r(s \to s')$
- Others …



Generative Augmented Flow Networks, 2022.

Ongoing directions

- Stochastic transition environment
  - Current GFlowNet formulation only supports deterministic transition
  - Generalizing (stochastic) detailed balance with transition model
  - $F(s)P_F(a|s)P(s'|s,a) = F(s')P_B(s,a|s')$
- Distributional flow matching
  - Distributional RL generalizes scalar Q-function to be a distribution
  - Richer learning signal: Q-learning becomes divergence minimization
  - We could parameterize flow's different quantiles
    - Quantile regression version of flow matching
  - Preliminary result in hypergrid



Exploration problems

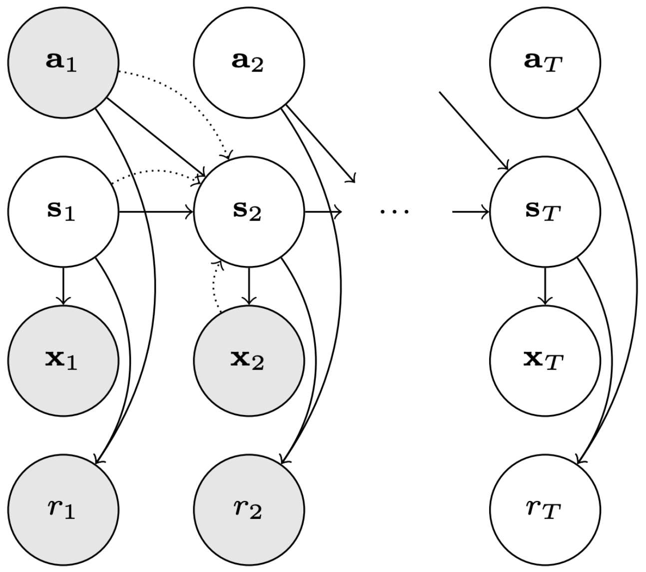
Trade-off exploration vs. exploitation

- Bandits / Online learning
- Reinforcement learning
- Black-box optimization
- Active learning
- …

Next I would talk about examples of our work on using probabilistic methods to achieve better exploration.

Structured exploration in RL

- In realistic settings such as partially observed MDP (POMDP), the true states of environment is not observed
- Previous works:
  - Extract deterministic feature: $s = f(x)$, making decision conditioned on s: $\pi(a|s)$
  - Model the belief of true state $p(s|x)$ with a world model, but only use one sample or take mean of $p(s|x)$

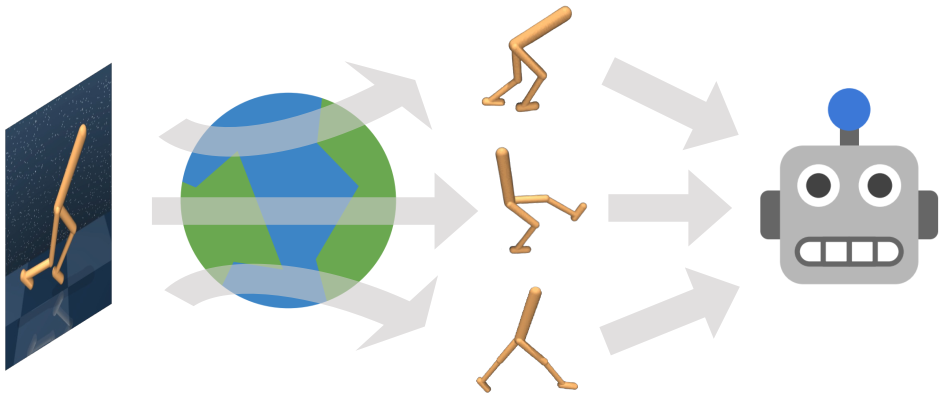


Information is lost! Need to take the whole distribution into account

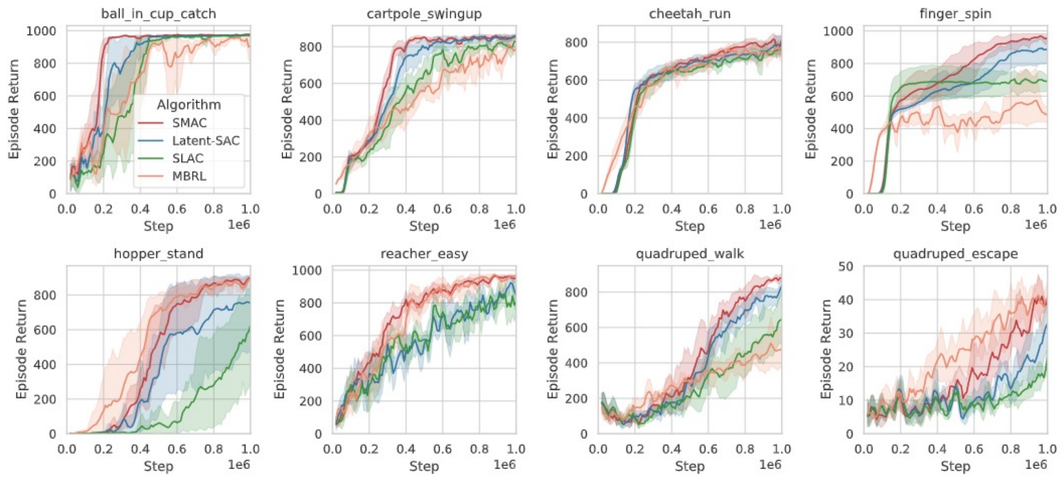
Latent State Marginalization as a Low-cost Approach for Improving Exploration. Arxiv 2022.

Latent State Marginalization

- We propose to marginalize out all the possible latents in belief distribution: $\pi(a|x) = \int \pi(a|s)p(s|x)ds$
  - $p(s|x)$ is from a world model, or unstructured prior



- Address entropy lower bound estimation for MaxEnt RL training
- Conduct experiments on various control tasks



Latent State Marginalization as a Low-cost Approach for Improving Exploration. Arxiv 2022.

Treating Black-box Opt in a Bayesian way

- Optimization is the limit of sampling
-  $\mathbf{m}^* = \arg\max_{\mathbf{m} \in \mathcal{M}} f(\mathbf{m})$ <=> $p(m) \propto \exp\left(\frac{f(m)}{T}\right),\ T \cong 0$

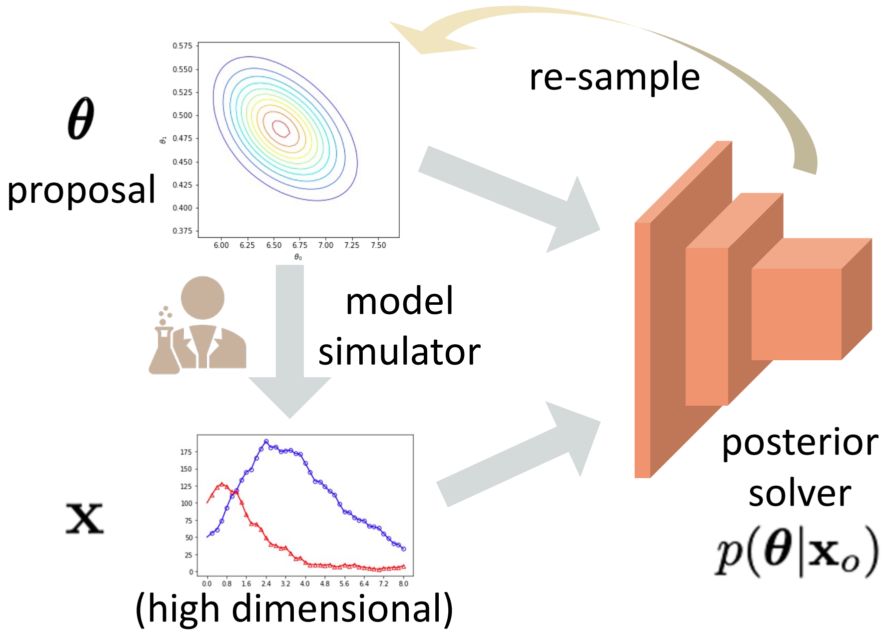
- Special constraint:
  - Limited number of query for each round
  - Black-box oracle

Treating Black-box Opt in a Bayesian way

- We show that it is closely related to likelihood-free Bayesian inference (LFI)

$$p(\boldsymbol{\theta}|\mathbf{x}_o) \propto p(\boldsymbol{\theta})\underbrace{p(\mathbf{x}_o|\boldsymbol{\theta})}_{?}$$

("o" means observation)



- Limited number of samples from likelihood $\mathbf{x} \sim p(\mathbf{x}|\boldsymbol{\theta})$ for each round
- Intractable likelihood function

Unifying Likelihood-free Inference with Black-box Optimization and Beyond, ICLR2021 spotlight

Unifying LFI and Black-Box Opt

- Assume $\mathcal{E}$ denotes a Boolean event:
  - “generated drug $\mathbf{m}$ has good property”
- Then we have a intriguing connection between the two fields! We then bridge / design (more than ten) algorithms from the two worlds

	Likelihood-free inference	Black-box optimization
Element	$(\boldsymbol{\theta}, \mathbf{x})$	$(\mathbf{m}, s)$
Target	$p(\boldsymbol{\theta}\|\mathbf{x}_o)$	$p(\mathbf{m}\|\mathcal{E})$
Constraint	limited simulation: $\mathbf{x} \sim p(\mathbf{x}\|\boldsymbol{\theta})$ intractable likelihood: $p(\mathbf{x}\|\boldsymbol{\theta})$	limited query: $s \sim f(\mathbf{m})$ black-box oracle: $f(\mathbf{m})$

Table 1: Correspondence between likelihood-free inference and black-box optimization.

Unifying Likelihood-free Inference with Black-box Optimization and Beyond, ICLR2021 spotlight

Thank you for listening!

- Also, huge thanks to all my collaborators and advisors!
- Questions?
- More info at my personal website