

Кафедра Интеллектуальных Систем

Градиентные потоки Вассерштайна: методы моделирования и применение в приложениях

Студент: *П.В.Мокров*

Научный руководитель: д.ф.-м.н. *Е.В.Бурнаев*

Июнь, 2022

Общая постановка задачи

Моделирование градиентных потоков Вассерштайна

Сферы применения:

- моделирование динамики толпы (*Maury et. al., 2010*)
- генеративное моделирование (*Gao et. al., 2019*)
- обучение с подкреплением (*Zhang et. al., 2018*)
- моделирование динамики популяции (*Hashimoto et. al., 2016*)
- нелинейная фильтрация (*Mokrov et. al., 2021*)
- семплирование из ненормированного распределения (*Mokrov et. al., 2021*)

Существующие методы решения:

- дискретизация пространства → в пространствах низкой размерности
- методы симуляции частиц → нет моделирования вероятностных мер
- прямой метод Эйлера → не приближают истинный град. поток
- ЖКО схема (*Jordan et. al., 1996*)

Введение. Градиентные потоки Вассерштайна

Пусть $\rho_t, t \in [0, T]$ - кривая в вероятностном пространстве $\mathcal{P}_2(\mathbb{R}^N)$. Тогда **метрическая производная** этой кривой есть:

$$|\rho'_t| = \lim_{dt \rightarrow 0} \frac{\mathcal{W}_2(\rho_t, \rho_{t+dt})}{dt}$$

Кривая $\rho_t, t \in [0, T]$ называется **градиентным потоком Вассерштайна** относительно функционала $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^N) \rightarrow \mathbb{R}$ если она образует кривую наискорейшего спуска относительно этого функционала, т.е. $\forall t \in [0, T]$:

$$\left. \frac{d}{dt} \mathcal{F}(\rho_t) \right|_{t=t} = \min_{\substack{\mu_t, \text{ s.t.:} \\ \mu_t = \rho_t \\ |\mu'_t| = |\rho'_t|}} \left. \frac{d}{dt} \mathcal{F}(\mu_t) \right|_{t=t}$$

Введение. Процессы Ланжевена

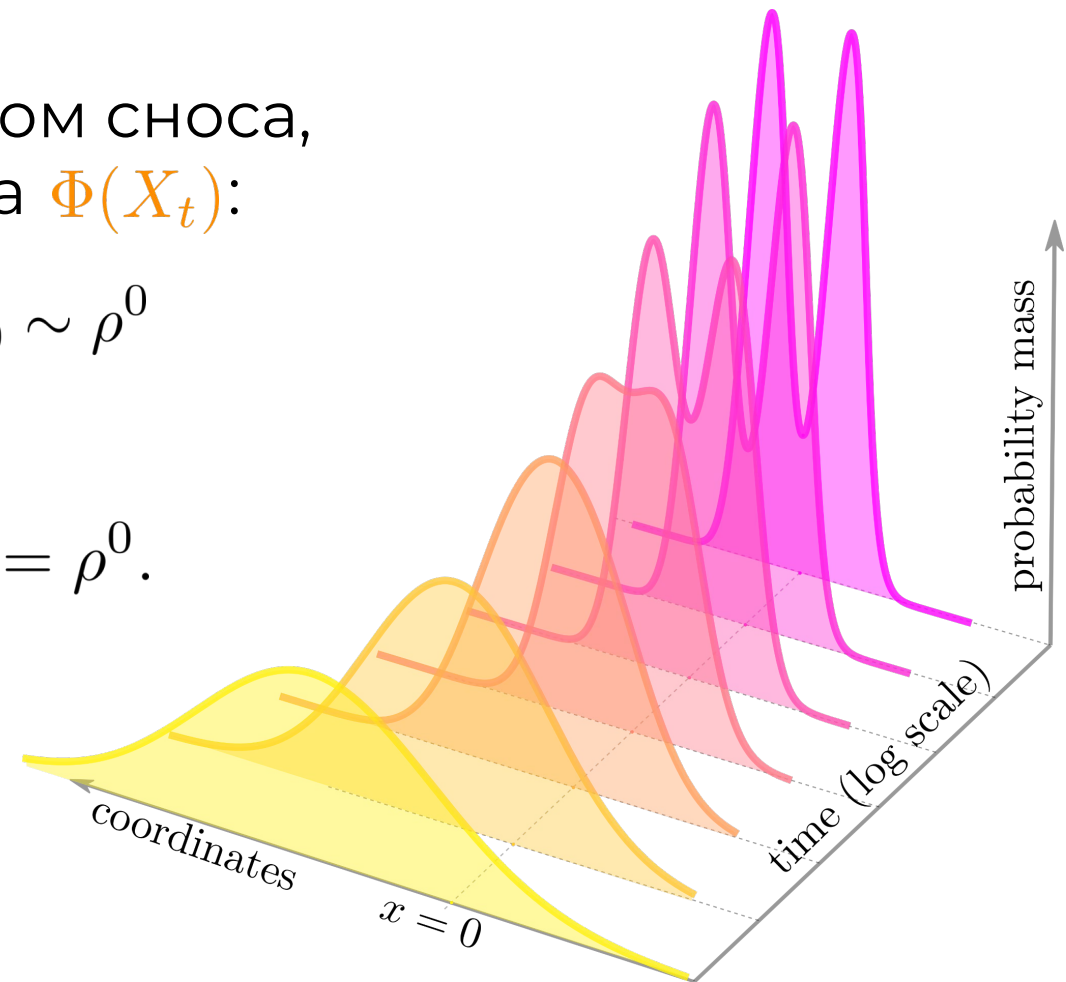
Процесс Ланжевена с коэффициентом сноса, являющимся градиентом потенциала $\Phi(X_t)$:

$$dX_t = -\nabla \Phi(X_t)dt + \sqrt{2\beta^{-1}}dW_t, \quad \text{s.t. } X_0 \sim \rho^0$$

Уравнение **Фоккера-Планка**:

$$\frac{\partial \rho_t}{\partial t} = \text{div}(\nabla \Phi(x)\rho_t) + \beta^{-1}\Delta \rho_t, \quad \text{s.t. } \rho_0 = \rho^0.$$

Пример процесса Ланжевена ►



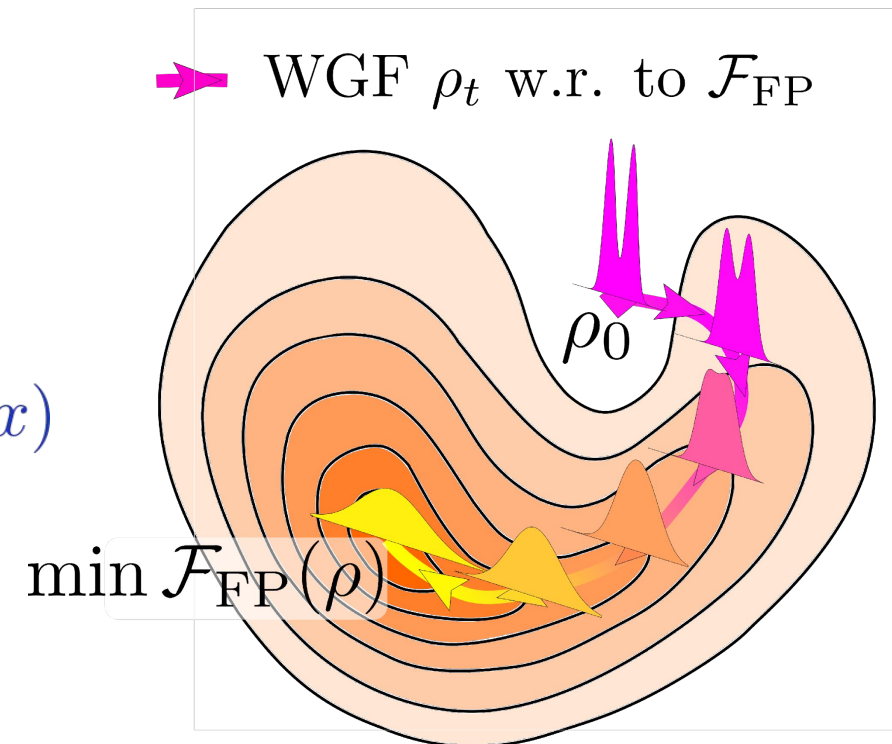
Введение. Процесс Ланжевена как градиентный поток Вассерштайна

Функционал Фоккера-Планка *(Jordan, Kinderlehrer, Otto, 1996)*

Маргинальные вероятностные меры процесса Ланжевена подчиняются градиентному потоку Вассерштайна отн. функционала Фоккера-Планка:

$$\mathcal{F}_{\text{FP}}(\rho) = \int_{\mathbb{R}^N} \Phi(x) d\rho(x) + \beta^{-1} \int_{\mathbb{R}^N} \log \rho(x) d\rho(x)$$

Иллюстрация градиентного потока Вассерштайна ►



Цель

- разработать масштабируемые методы моделирования градиентных потоков Вассерштайна на основе ЖКО схемы и теории оптимального транспорта
- применить полученные методы в приложениях

План работы

1. **Переформулировать ЖКО схему** для WGF относительно функционала Фоккера-Планка:

$$\mathcal{F}_{\text{FP}}(\rho) = \int_{\mathbb{R}^N} \Phi(x) d\rho(x) + \beta^{-1} \int_{\mathbb{R}^N} \log \rho(x) d\rho(x)$$

с помощью ICNNs и теории оптимального транспорта.

2. Предложить и **реализовать алгоритм оптимизации** для решения ЖКО схемы.
3. **Оценить качество подхода** на численных экспериментах, особое внимание уделяя задачам с высокой размерностью.

Методы. Теория

ЖКО схема (*Jordan, Kinderlehrer, Otto, 1996*)

Приближает WGF с функционалом Фоккера-Планка \mathcal{F}_{FP} с помощью последовательности мер $\{\rho_\tau^k\}_{k=0}^K$; $\rho_\tau^0 = \rho^0$ такой, что:

$$\rho_\tau^k \leftarrow \arg \min_{\rho \in \mathcal{P}(\mathbb{R}^N)} \frac{1}{2} \mathcal{W}_2^2(\rho_\tau^{k-1}, \rho) + \tau \mathcal{F}_{\text{FP}}(\rho)$$

Переформулировка ЖКО с помощью **теоремы Бренье** (*Brenier, 1987*):

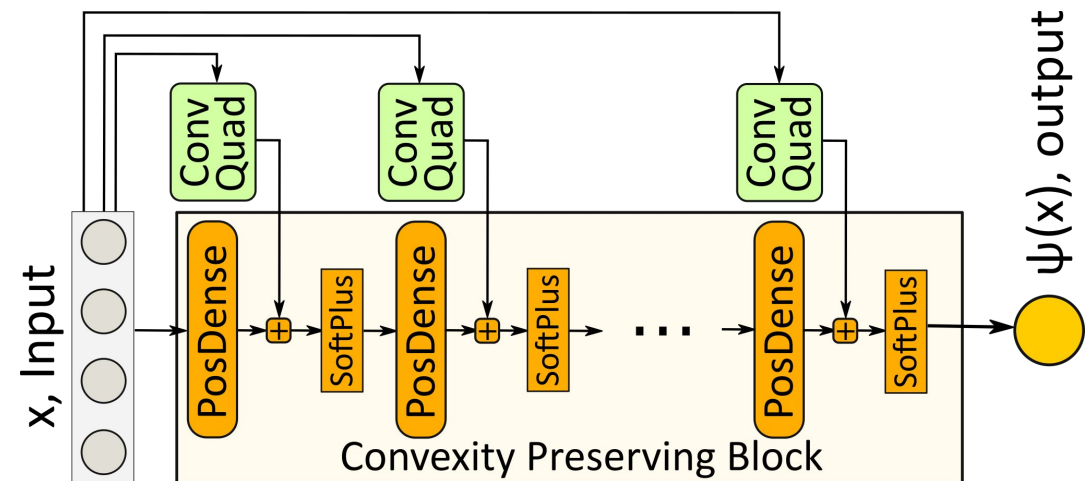
$$\psi_k = \arg \min_{\psi \in \text{Conv}(\mathbb{R}^N)} \tau \mathcal{F}_{\text{FP}}(\rho_\tau^k \circ \nabla \psi^{-1}) + \frac{1}{2} \int_{\mathbb{R}^N} \|x - \nabla \psi(x)\|_2^2 d\rho_\tau^k(x);$$

$$\rho_\tau^{k+1} = \rho_\tau^k \circ \nabla \psi_k^{-1}$$

Методы. Оптимизация

Выпуклые нейронные сети (ICNNs) (Amos, 2017)

Пример структуры ICNN ►. Выпуклость обеспечивается особыми условиями на веса и активации, а также особой топологией слоев сети. Источник рисунка: (Korotin et. al., 2021)



Стохастическая функционал оптимизации ЖКО (Korotin, 2021)

$$\widehat{F}_{\text{FP}}(x_{1:n}) = \frac{1}{n} \sum_{i=1}^n \left\{ \Phi(\nabla \psi_{\theta}(x_i)) - \beta^{-1} \log |\det \text{Hess}(\psi_{\theta})(x_i)| \right\} ; x_i \sim \rho_{\tau}^{k-1}$$

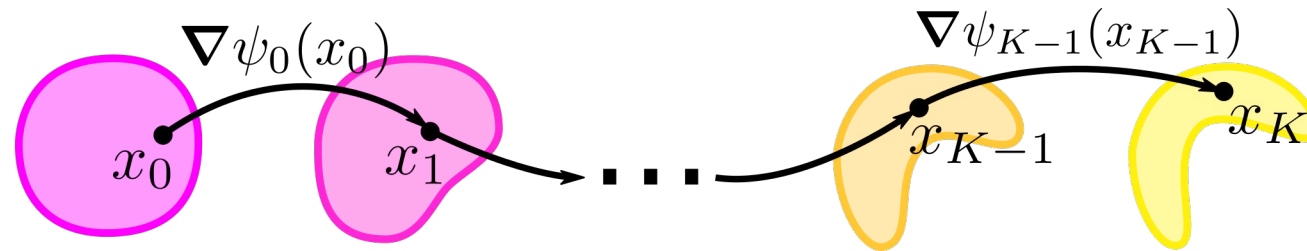
Процедура семплирования ЖКО:

$$\text{Sample batch } Z \sim \rho^0; \Rightarrow X \leftarrow \nabla \psi_{K-1} \circ \dots \circ \nabla \psi_0(Z) \sim \rho_{\tau}^K$$

Методы. Приложения

Оценка плотности ЖКО (Korotin, 2021)

Формула замены переменной $\rho_\tau^K(x_K) = \rho^0(x_0) \cdot [\prod_{i=0}^{K-1} \det \nabla^2 \psi_i(x_i)]^{-1}$
 как посл. выпуклых задач оптимизации $x_{i+1} = \arg \max_x (x^T x_i - \psi_i(x))$



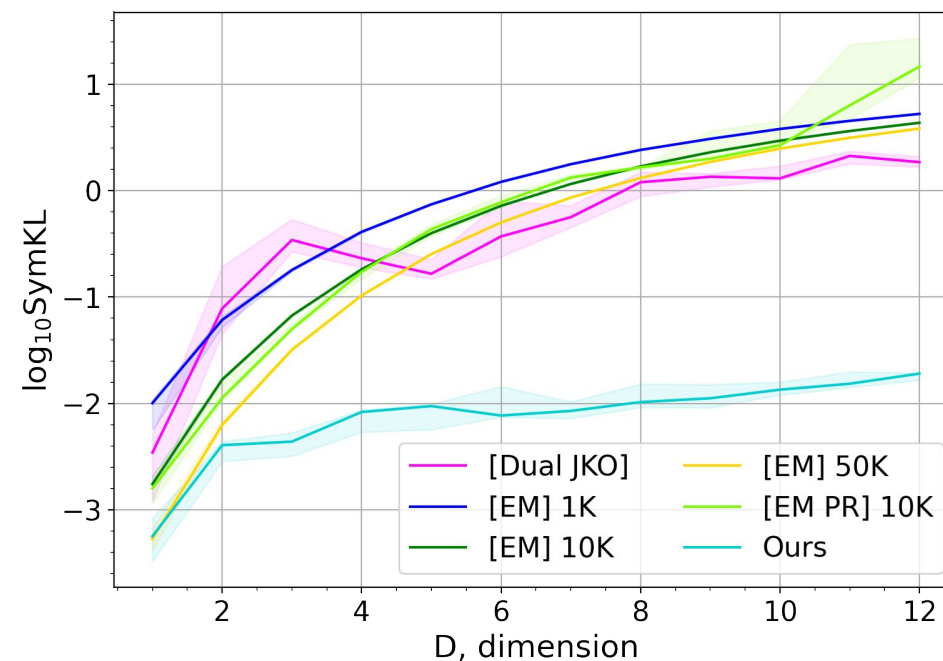
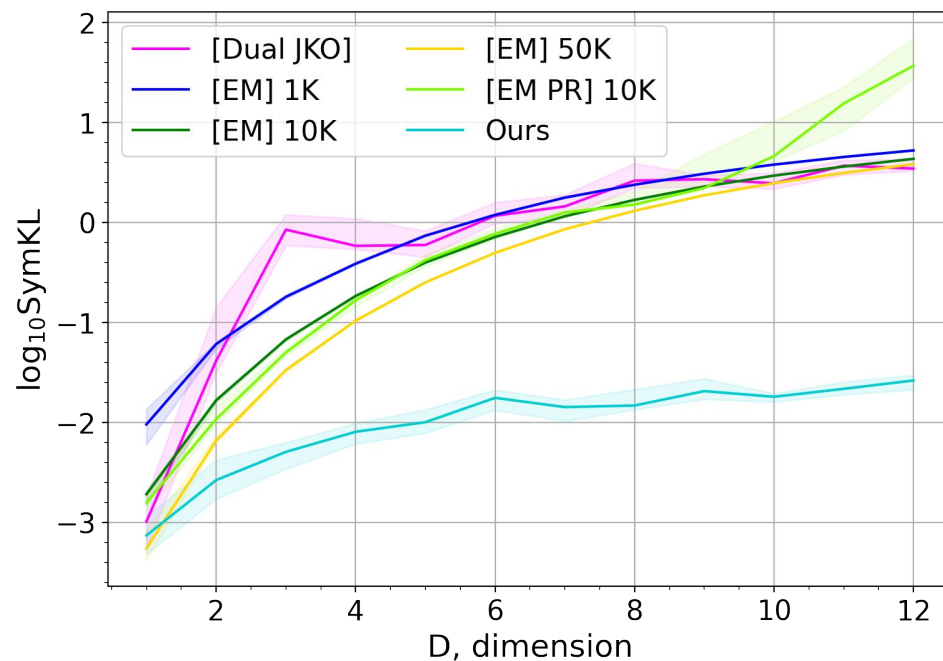
Алгоритм Метрополиса-Гастингса на основе ЖКО (Mokrov, 2021)

Алгоритм Метрополиса-Гастингса в задаче нелинейной фильтрации с особыми $\{\psi_i\}_{i=1}^K$ модельно - зависимыми вспомогательными функциями распределения для быстрой оценки отношения плотностей

Результаты. Синтетические эксперименты

Моделирование процессов Орнштейна - Уленбека

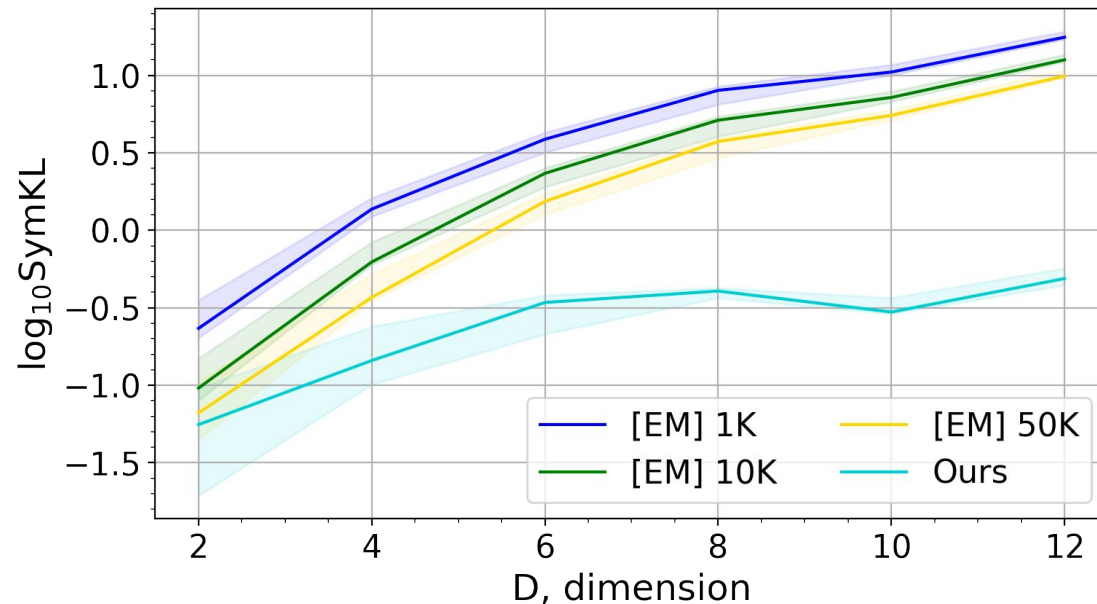
$$\Phi(x) = \frac{1}{2}(x - b)^T A(x - b), A \text{ is SPD}$$



Расхождения между истинными и предсказанными вероятностными мерами при $t = 0.5$ (слева) и $t = 0.9$ (справа) ▲

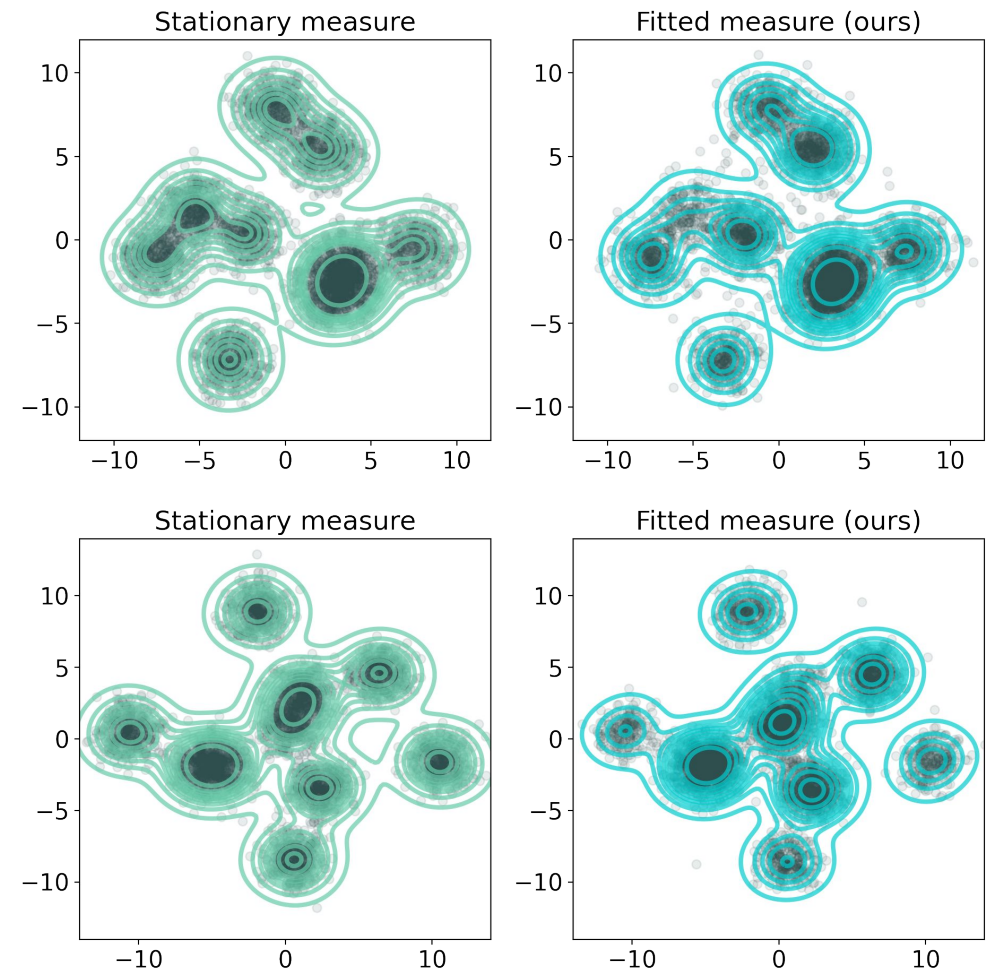
Результаты. Синтетические эксперименты

Сходимость к стационарному распределению $\rho^*(x) = Z^{-1} \exp(-\beta \Phi(x))$



Сравнение сходимости к стационарному распределению в разных размерностях ▲

Визуальное расхождение между истинными и предсказанными стационарными распределениями для размерностей $N = 32$ (сверху) и $N = 13$ (снизу). Проекция на 2 главные PC ►



Результаты. Приложения

Семплирование из ненормированного апостериорного распределения в задаче байесовской логистической регрессии

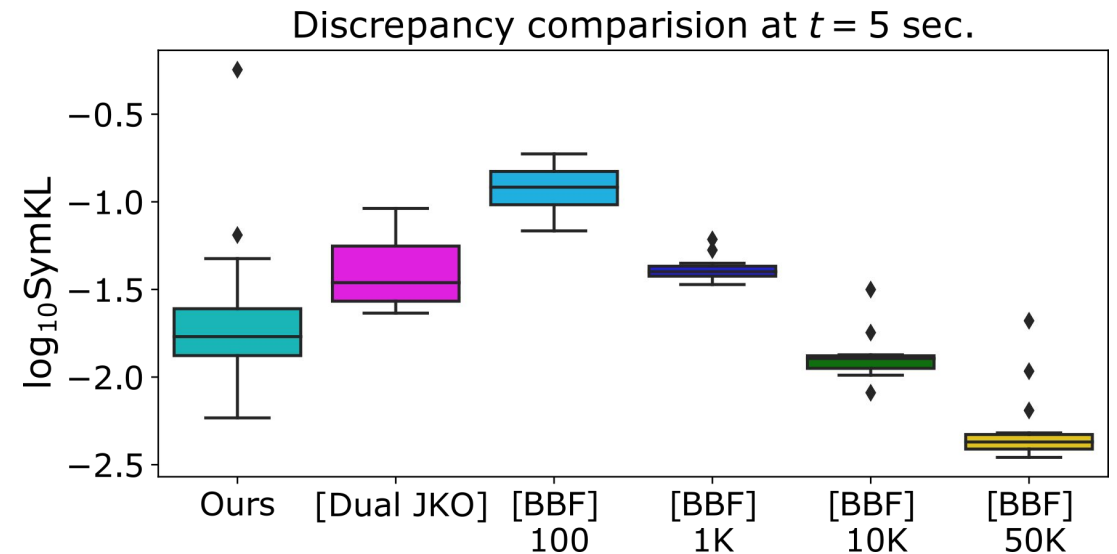
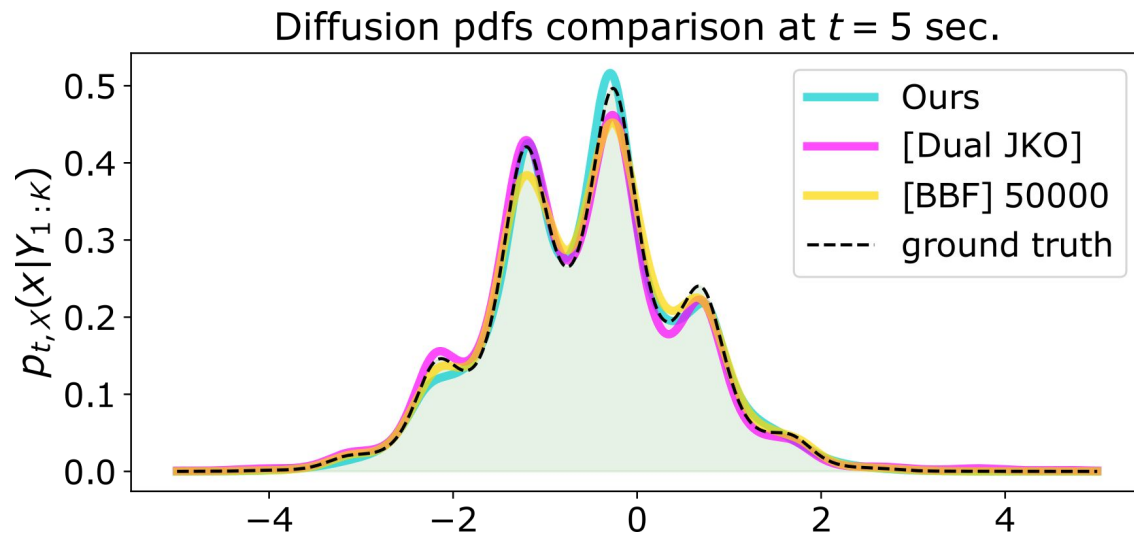
Dataset	Accuracy		Log-Likelihood	
	Ours	[SVGD]	Ours	[SVGD]
covtype	0.75	0.75	-0.515	-0.515
german	0.67	0.65	-0.6	-0.6
diabetis	0.775	0.78	-0.45	-0.46
twonorm	0.98	0.98	-0.059	-0.062
ringnorm	0.74	0.74	-0.5	-0.5
banana	0.55	0.54	-0.69	-0.69
splice	0.845	0.85	-0.36	-0.355
waveform	0.78	0.765	-0.485	-0.465
image	0.82	0.815	-0.43	-0.44

Сравнение с методом SVGD (*Liu et. al., 2016*) в задаче байесовской логистической регрессии на 9 стандартных датасетах

Результаты. Приложения

Нелинейная фильтрация

Апостериорное распределение процесса Ланжевена X_t на основе зашумленных наблюдений из процесса: $p(X_t | Y_{t_1}, Y_{t_2}, \dots, Y_{t_k})$



Визуальное (слева) и boxplot (справа) сравнение апостериорных распределений для 1D нелинейной диффузии Фоккера-Планка с $\Phi(x) = \frac{1}{\pi} \sin(2\pi x) + \frac{1}{4}x^2$

Обсуждение результатов

- Лучшее качество в экспериментах с синтетическими данными (процессы Орнштейна-Уленбека, сходимость к стационарному распределению).
- Сравнимое качество с SVGD в задаче семплирования из ненормированного распределения, более быстрая генерация выборки.
- Конкурентоспособное качество в задаче 1D нелинейной фильтрации. Возможно, лучшее качество в задаче нелинейной фильтрации в больших размерностях.

Научная новизна

работа

*Benamou et. al.,
2014*

*Frogner et. al.,
2020*

*Bunne et. al., 2022;
Alvarez-Melis et. al.,
2021*

сходства

Похожая формулировка
JKO с помощью теоремы
Бренье

Похожий набор числен-
ных экспериментов

Похожий метод
моделирования JKO,
параметризация с
помощью ICNNs

различия

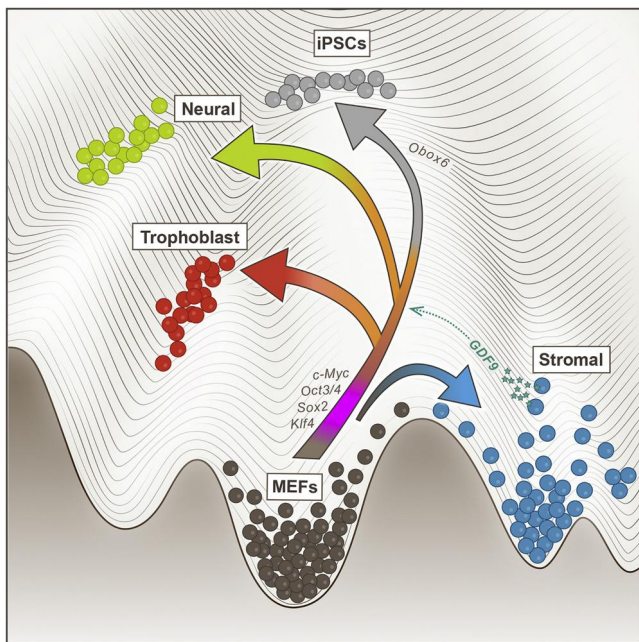
Сложная дискретиза-
ция пространства
выпуклых функций

Моделирование WGFs
с помощью двойст-
венной задачи JKO

Появились
параллельно с нашей
работой

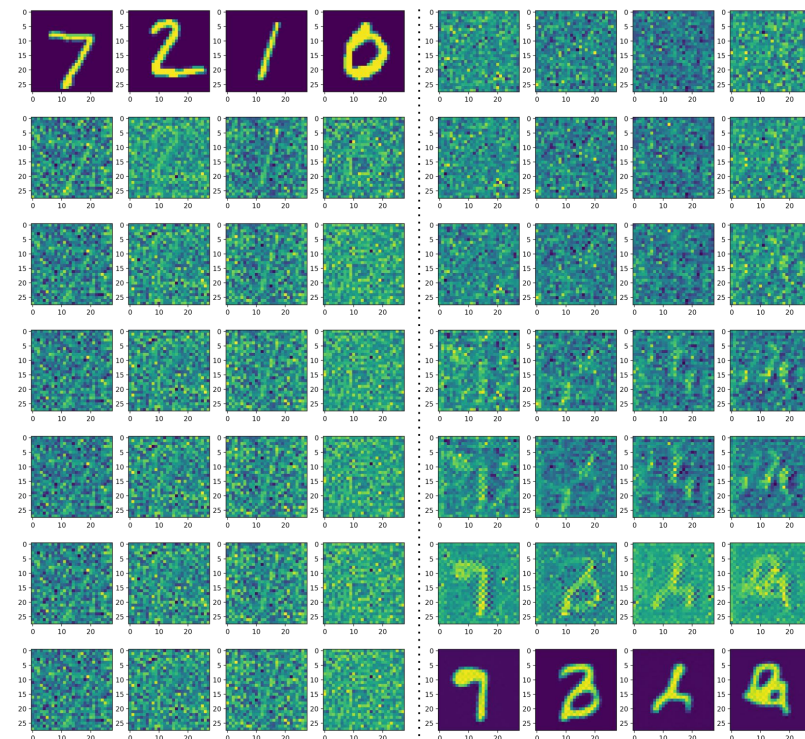
Потенциальные приложения

Популяционная динамика.
Single-cell RNA sequencing



Демонстрация пролиферации клеток ▲.
источник рисунка: (Schiebinger et. al., 2019)

Диффузионные
генеративные модели



forward diffusion: $\rho_{\text{im.}} \rightarrow \mathcal{N}(0, I)$ backward diffusion: $\mathcal{N}(0, I) \rightarrow \rho_{\text{im.}}$

Предварительные эксперименты
генерации MNIST ▲

Выводы

1. Предложена переформулировка ЖКО схемы с использованием теории оптимального транспорта и параметризации выпуклых функций с помощью ICNNs
2. Предложен и реализован алгоритм для решения ICNN-параметризованного ЖКО с помощью градиентного спуска
3. Изучено качество предложенного подхода в нескольких численных экспериментах и в приложениях (семплирование из ненормированного распределения, нелинейная фильтрация).

Публикации по теме диплома

P. Mokrov, A. Korotin, L. Li, A. Genevay, J. Solomon, E. Burnaev.
Large-Scale Wasserstein Gradient Flows; in ***Advances in Neural Information Processing Systems***, 2021

Благодарности

Хочу выразить благодарность Александру Коротину, который помог мне с этим проектом. В частности, он поделился со мной множеством полезных идей и помог с написанием и подачей нашей статьи на конференцию.