VECTOR INSTITUTE

Wasserstein Quantum Monte Carlo:

A Novel Approach for Solving the Quantum Many-Body Schrödinger Equation

Kirill Neklyudov[1], Jannes Nys[2], Luca Thiede[1,3],

Juan Felipe Carrasquilla[1,4], Qiang Liu[5], Max Welling[6], Alireza Makhzani[1,3]

1. Vector Institute
2. École Polytechnique Fédérale de Lausanne (EPFL)
3. University of Toronto
4. University of Waterloo
5. UT Austin
6. Microsoft Research AI4Science

Agenda

1. Problem statement and motivation
2. Gradient flows perspective
3. Simulation of the gradient flows in practice
4. Results and discussion

Motivation and fast recap

We want to solve the stationary Schrödinger equation:

$$\left(-\frac{1}{2}\nabla_x^2 + V(x)\right)\psi(x) = E\psi(x)$$

given $V(x)$; find $\psi(x)$ and E

In order to do so, we minimize the energy of the system:

$$E[q] = \mathbb{E}_{q(x)}[E_{\text{loc}}(x)] \qquad E_{\text{loc}}(x) = V(x) - \frac{1}{4}\nabla_x^2 \log q(x) - \frac{1}{8}\|\nabla_x \log q(x)\|^2$$

$$q(x) = |\psi(x)|^2$$

we need samples (from $q(x)$)

and the density ($\log q(x)$)

Agenda

1. Problem statement and motivation
2. Gradient flows perspective
3. Simulation of the gradient flows in practice
4. Results and discussion

Energy minimization as a **non-parametric** gradient flow

The energy defines a functional on the space of distributions:

$$E[q] : \text{distributions} \to \mathbb{R}$$

We can minimize the energy doing the gradient descent:

(informally) $$q_{t+dt} = q_t - dt \cdot \frac{\delta E[q_t]}{\delta q_t}$$

(formally) $$\inf_{q_{t+dt}} E[q_{t+dt}] - E[q_t] + \frac{1}{2dt}\mathrm{KL}(q_{t+dt} \| q_t)$$

$$dt \to 0$$

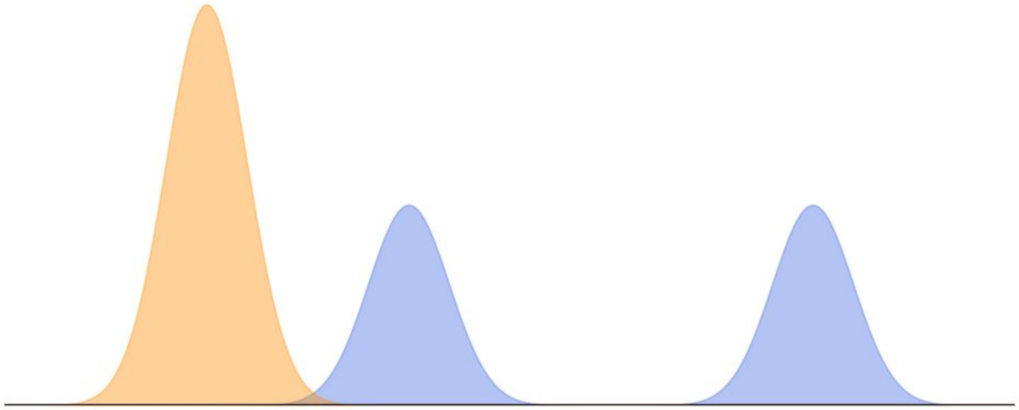
$$\frac{\partial q_t}{\partial t} = -\left(\frac{\delta E[q_t]}{\delta q_t} - \mathbb{E}_{q_t}\left[\frac{\delta E[q_t]}{\delta q_t}\right]\right) q_t$$

Energy minimization as a **non-parametric** gradient flow

$$\frac{\partial q_t}{\partial t} =$$ something

model:

target:

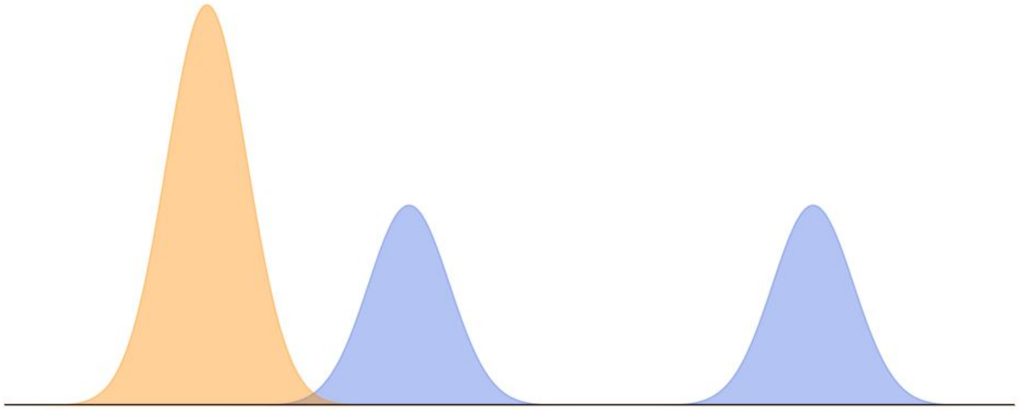


Energy minimization as a **non-parametric** gradient flow

$$\frac{\partial q_t}{\partial t} =$$ **something**

model:

target:



Non-parametric gradient flow under 2-Wasserstein metric

The energy defines a functional on the space of distributions:

$$E[q] : \text{distributions} \to \mathbb{R}$$

We can minimize the energy doing the gradient descent:

(what we had before) $$\inf_{q_{t+dt}} E[q_{t+dt}] - E[q_t] + \frac{1}{2dt}\boxed{\mathrm{KL}}(q_{t+dt} \| q_t)$$

(another gradient flow) $$\inf_{q_{t+dt}} E[q_{t+dt}] - E[q_t] + \frac{1}{2dt}\boxed{W_2}(q_{t+dt}, q_t)$$

$$\boxed{dt \to 0}$$

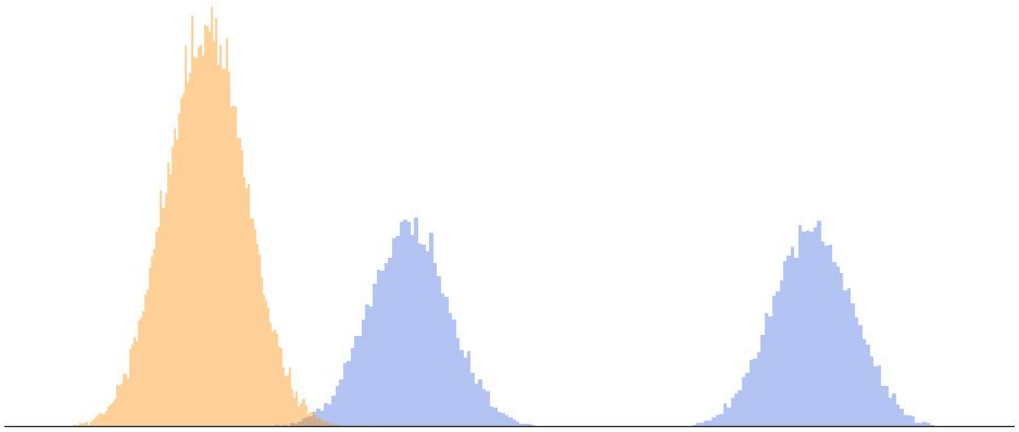
$$\frac{\partial q_t}{\partial t}(x) = -\nabla_x \cdot \left(q_t(x) \left(-\nabla_x \frac{\delta E[q_t]}{\delta q_t}(x) \right) \right)$$

Non-parametric gradient flow under Wasserstein-2 metric

$$\frac{\partial q_t}{\partial t}(x) =$$ **another something**

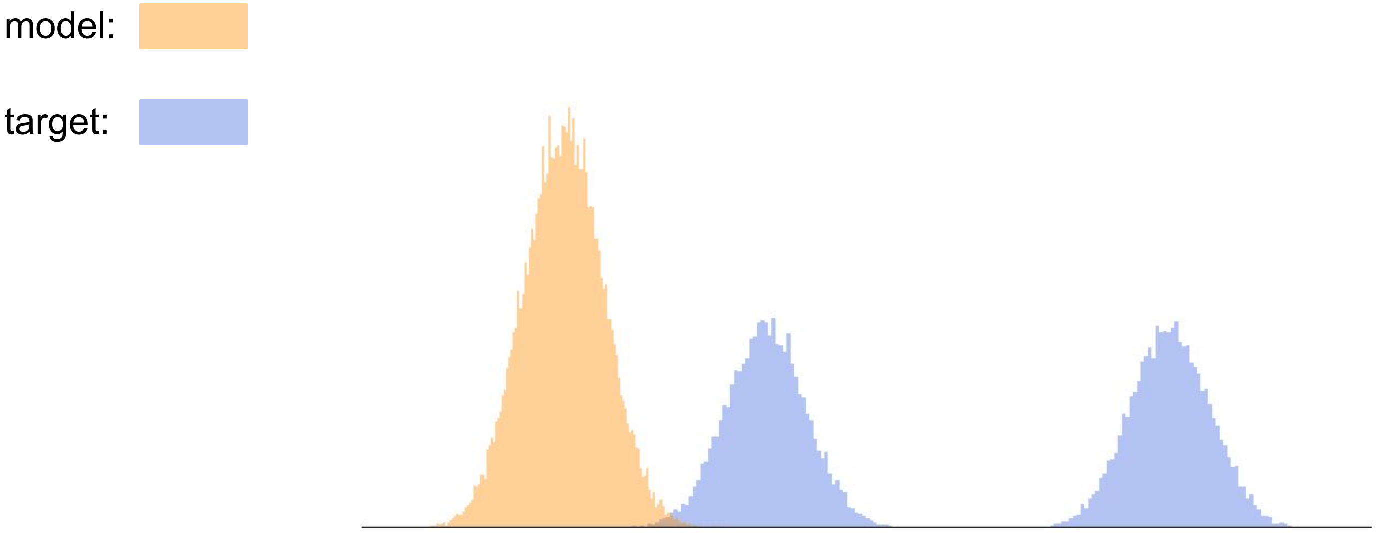
model:

target:



Non-parametric gradient flow under Wasserstein-2 metric

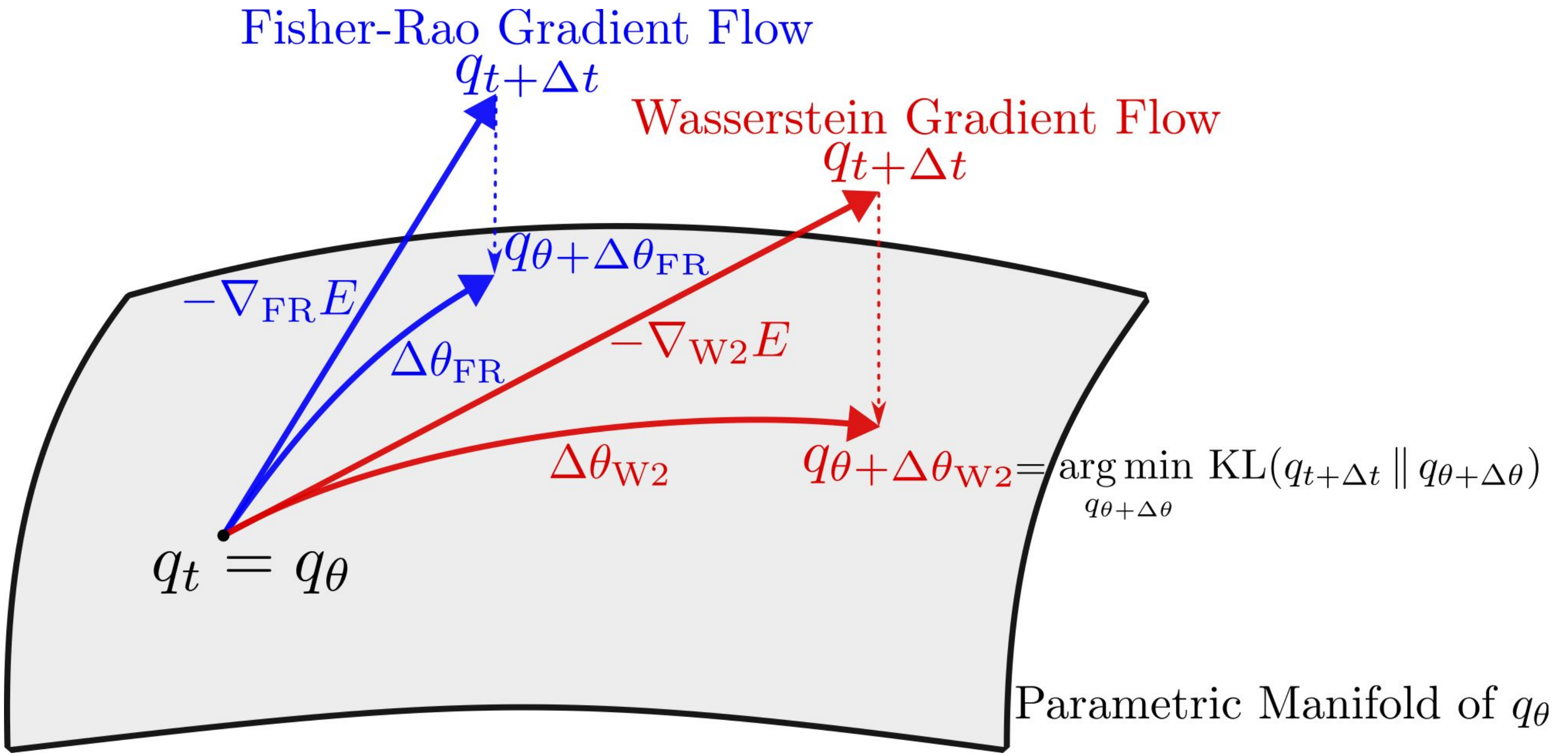
$$\frac{\partial q_t}{\partial t}(x) =$$ another something



Agenda

1. Problem statement and motivation
2. Gradient flows perspective
3. Simulation of the gradient flows in practice
4. Results and discussion

How do we run them in practice?



One iteration of QVMC

1. Given the model of the density and samples:

$$q(x) = q(x, \theta)$$
$$\{x^{(i)}\}_{i=1}^N \sim q(x, \theta)$$

2. Update the model minimizing the energy:

$$E_{\text{loc}}(x) = V(x) - \frac{1}{4}\nabla_x^2 \log q(x) - \frac{1}{8}\|\nabla_x \log q(x)\|^2$$
$$\Delta\theta^* = -\mathbb{E}_{q_t(x)}\left[\Big(E_{\text{loc}}(x) - \mathbb{E}_{q_t(x)}[E_{\text{loc}}(x)]\Big)\nabla_\theta \log q(x, \theta)\right]$$
$$\theta' = \theta + (\text{learning rate}) \cdot \Delta\theta^*$$

Projection onto the parametric family

3. Update the samples to match the new density using MCMC:

$$\{x^{(i)}\}_{i=1}^N \sim q(x, \theta')$$

One iteration of WQMC

1. Given the model of the density and samples:

$$q(x) = q(x, \theta)$$
$$\{x^{(i)}\}_{i=1}^N \sim q(x, \theta)$$

2. Update the model minimizing the energy:

$$E_{\text{loc}}(x) = V(x) - \frac{1}{4}\nabla_x^2 \log q(x) - \frac{1}{8}\|\nabla_x \log q(x)\|^2$$
$$\Delta\theta^* = -\mathbb{E}_{q_t(x)} \nabla_\theta \left\langle \nabla_x E_{\text{loc}}(x), \nabla_x \log q(x, \theta) \right\rangle$$

← new parameters update

$$\theta' = \theta + (\text{learning rate}) \cdot \Delta\theta^*$$

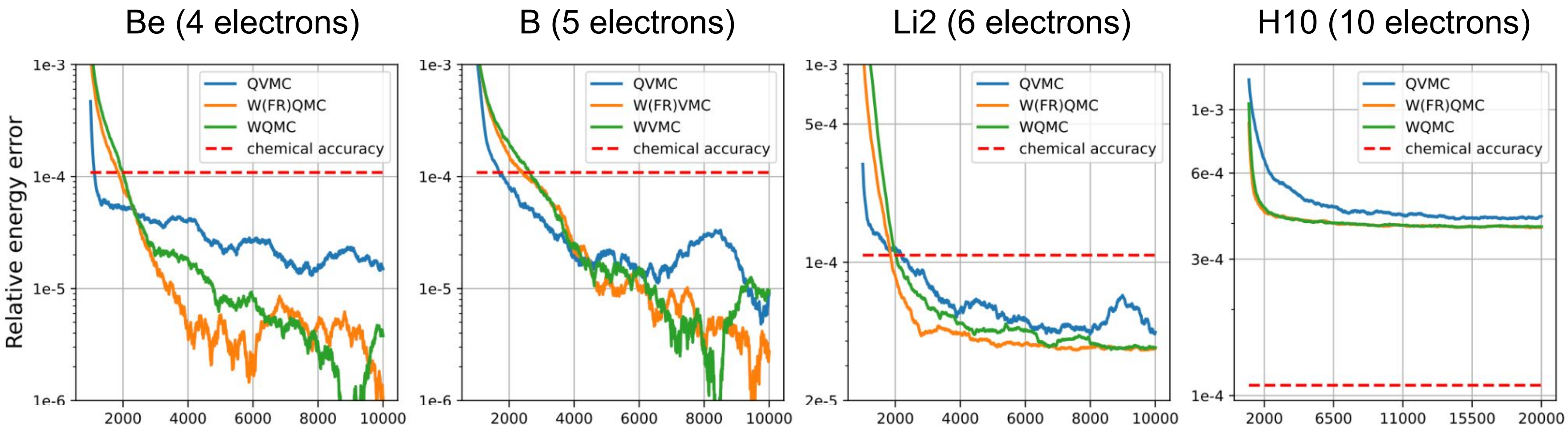
3. Update the samples to match the new density using MCMC:

$$\{x^{(i)}\}_{i=1}^N \sim q(x, \theta')$$

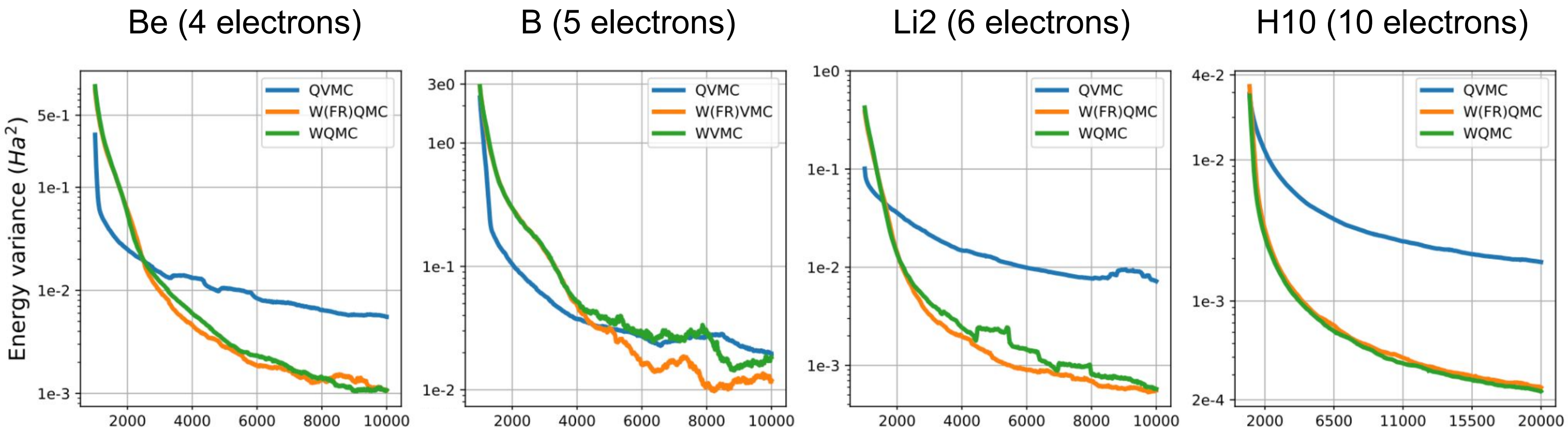
Agenda

1. Problem statement and motivation
2. Gradient flows perspective
3. Simulation of the gradient flows in practice
4. Results and discussion

Results (energy)

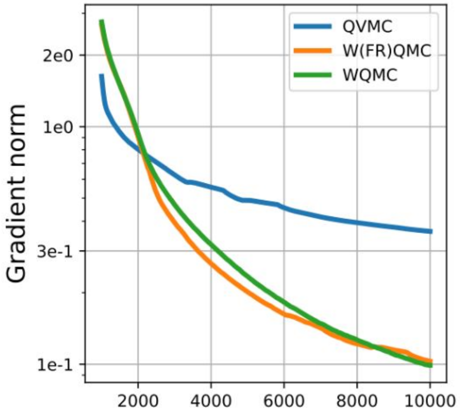


Results (energy variance)

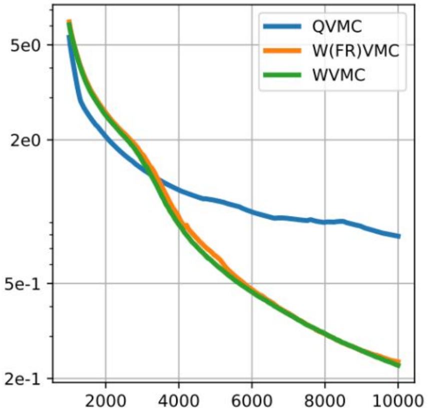


Results (norm of the energy gradient)

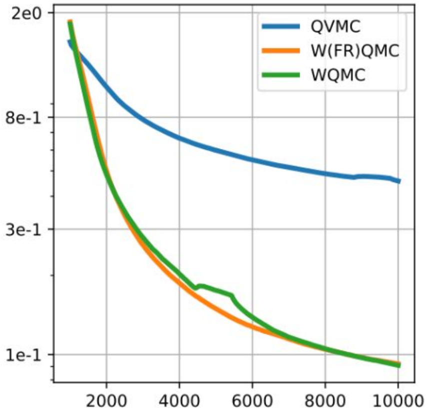
Be (4 electrons)



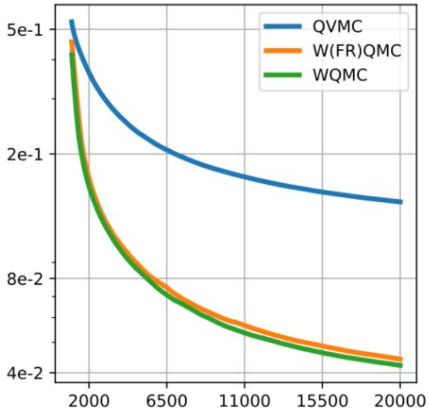
B (5 electrons)



Li2 (6 electrons)



H10 (10 electrons)



Why WQMC is better than QVMC?

Fisher-Rao Gradient Flow

2-Wasserstein Gradient Flow

Definitions:

$$\inf_{q_{t+dt}} E[q_{t+dt}] - E[q_t] + \frac{1}{2dt}\mathrm{KL}(q_{t+dt}\|q_t)$$

$$\inf_{q_{t+dt}} E[q_{t+dt}] - E[q_t] + \frac{1}{2dt}W_2(q_{t+dt}, q_t)$$

different notion of closeness

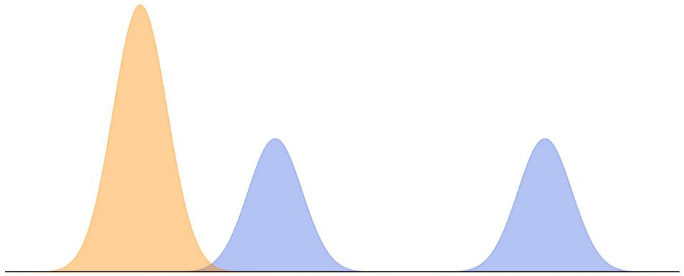
PDEs:

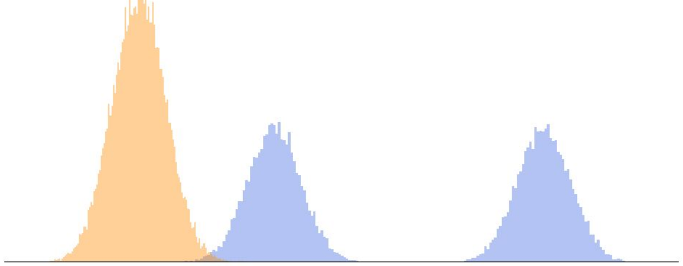
$$\frac{\partial q_t}{\partial t} = -\left(\frac{\delta E[q_t]}{\delta q_t} - \mathbb{E}_{q_t}\left[\frac{\delta E[q_t]}{\delta q_t}\right]\right)q_t$$

$$\frac{\partial q_t}{\partial t} = -\nabla_x \cdot \left(q_t\left(-\nabla_x \frac{\delta E[q_t]}{\delta q_t}\right)\right)$$

reweights density by teleporting mass

propagates samples along this vector field





You will find more in the paper!

- Relation to imaginary-time evolution
- C-Wasserstein metric, where c is any convex function
- Interpolation between Fisher-Rao and C-Wasserstein
- Detailed derivations of the gradient flows
- Code and experiments details