Parametrization and Computation of Wasserstein Hamiltonian flows

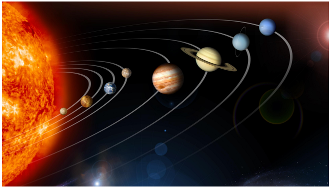
Shu Liu (UCLA)

Joint work with Hao Wu (Georgia Tech), , Xiaojing Ye (Georgia State), and Haomin Zhou (Georgia Tech)

August 1, 2023

Wasserstein Hamiltonian flow

Hamiltonian systems are used to describe the evolution of a physical system. They are ubiquitous in our physical world.



(a) Solar system

(b) Pendulum clock

(c) Lissajou curve as a complex harmonic motion

Figure: Examples of Hamiltonian systems

Wasserstein Hamiltonian (WH) flows

- WH flow [1] is a Hamiltonian system lifted to the probability manifold.
- It is related to classical particle Hamiltonian systems.
- Links to many PDE systems such as Schrödinger equation or Schrödinger Bridge problem.

Our goal: a scalable and sampling-friendly method for computing WH flows.

[1] S. Chow, W. Li, H. Zhou Wasserstein Hamiltonian flows, JDE 2020

Wasserstein Hamiltonian (WH) flow

Consider the following Hamiltonian defined on the cotangent bundle of $\mathcal{P}(\mathbb{R}^d)$:

$$\mathcal{H}(\rho, \Phi) = \int_{\mathbb{R}^d} \frac{1}{2}|\nabla \Phi|^2 \rho dx + \mathcal{F}(\rho)$$

The WHF is:

$$\frac{\partial}{\partial t}\rho_t = \frac{\delta}{\delta \Phi}\mathcal{H}(\rho, \Phi), \quad \frac{\partial}{\partial t}\Phi_t = -\frac{\delta}{\delta \rho}\mathcal{H}(\rho, \Phi).$$

Wasserstein Hamiltonian (WH) flow

Consider the particle Hamiltonian system:

$$\frac{d}{dt}X_t = v_t, \quad , \quad \frac{d}{dt}v_t = -\nabla_{X_t}\frac{\delta}{\delta\rho_t}\mathcal{F}(\rho_t, X_t),$$

with $X_0 \sim \rho_0$, $v_0 = \nabla\Phi_0(X_0)$, and $0 \le t \le T$.

The connection between WHF and particle Hamiltonian is:

$$X_t \sim \rho_t, \quad v_t = \nabla\Phi_t(X_t), \quad \text{for } t \in [0, T].$$

A Lagrangian perspective of WH flow

- Consider the following Lagrangian

$$\mathcal{L}(\rho_t, \partial_t \rho_t) = \frac{1}{2} g^W(\partial_t \rho_t, \partial_t \rho_t) - \mathcal{F}(\rho_t),$$

- Use L in a two-endpoint value problem

$$\mathcal{I}(\rho_t) = \inf_{\{\rho_t\}} \left\{ \int_0^T \mathcal{L}(\rho_t, \partial_t \rho_t) dt : \rho_0 = \rho^0, \rho_T = \rho^T \right\}.$$

and its Euler-Lagrange equation is

$$\partial_t \frac{\delta}{\delta \partial_t \rho_t} \mathcal{L}(\rho_t, \partial_t \rho_t) = \frac{\delta}{\delta \rho_t} \mathcal{L}(\rho_t, \partial_t \rho_t) + C(t).$$

A Lagrangian perspective of WH flow

Applying Legendre transform (with ρ fixed), we obtain

$$\mathcal{L}(\rho, \eta) = \sup_{\Phi} \{\langle \eta, \Phi \rangle - \mathcal{H}(\rho, \Phi)\}$$

where the optimal Φ satisfies $-\nabla \cdot (\rho \nabla \Phi) = \eta$.

The dual Φ in WHF relates to the velocity $\eta = \partial_t \rho_t$ accordingly:

$$-\nabla \cdot (\rho_t \nabla \Phi_t) = \partial_t \rho_t$$

This links to the Wasserstein metric

$$g^W(\rho)(\eta_1, \eta_2) = \int_{\mathbb{R}^d} \nabla \Phi_1 \cdot \nabla \Phi_2 \rho dx.$$

where Φ_i solves the elliptic equation $-\nabla \cdot (\rho \nabla \Phi_i) = \eta_i$ for $i = 1, 2$.

Parametrization of Push-forward Mapping

- Consider a parametric push-forward map T_θ:

$$T_\theta : \quad z \sim \lambda \quad \mapsto \quad T_\theta(z) \sim \rho_\theta := T_{\theta\#}\lambda$$

  where $\theta \in \Theta \subset \mathbb{R}^d$. Here $\rho_\theta(\cdot) = \lambda(T_\theta^{-1}(\cdot))\mathrm{det}(\nabla T_\theta^{-1}(\cdot))$.
- If θ_t is time-varying, we obtain a trajectory of push-forward densities:

$$\rho_{\theta(t)} = T_{\theta(t)\#}\lambda \quad (\approx \rho(t,\cdot)?)$$

- Parameterized densities form a submanifold of $\mathcal{P} = \mathcal{P}(\mathbb{R}^d)$:

$$\mathcal{P}_\Theta = \{\rho_\theta = T_{\theta\#}\lambda \mid \theta \in \Theta\} \subset (\mathcal{P}, g^W)$$

  For every θ, $T_{\theta\#}$ also induces a push-forward of tangent vector $\dot\theta \in \mathcal{T}_\theta\Theta$ to $(T_{\theta\#})_*\dot\theta \in \mathcal{T}_{\rho_\theta}\mathcal{P}_\Theta$.

Parametrized Wasserstein Hamiltonian flow

Consider $L = \mathcal{L} \circ (T_\sharp, (T_\sharp)_*) : \mathcal{T}\Theta \to \mathbb{R}$. Then:

$$L(\theta, \dot{\theta}) = \mathcal{L}(\rho_\theta, \frac{\partial \rho_\theta}{\partial \theta}\dot{\theta}) = \frac{1}{2}\dot{\theta}^T G(\theta)\dot{\theta} - F(\theta),$$

with $F(\theta) = \mathcal{F}(\rho_\theta)$.

Here the metric tensor $G(\theta)$ is the analogy of g^W on Θ:

$$G(\theta) = \int \nabla\Psi_\theta \nabla\Psi_\theta^\top \rho_\theta \, dx$$

where $-\nabla \cdot (\rho_\theta \nabla \psi_{\theta,i}) = -\nabla \cdot (\rho_\theta \partial_{\theta_i} T_\theta)$ for $i = 1, \ldots, m$ and $\Psi_\theta = (\psi_{\theta,i})_i$.

Parametrized Wasserstein Hamiltonian flow

By taking $p = \nabla_v L(\theta, \dot{\theta}) = G(\theta)\dot{\theta}$, the related Hamiltonian is

$$H(\theta, p) = \dot{\theta} p - L(\theta, \dot{\theta}) = \frac{1}{2} p^\top G(\theta)^{-1} p + F(\theta).$$

The Hamiltonian system of $H(\theta, p)$ is

$$\dot{\theta} = \nabla_p H(\theta, p) = G(\theta)^{-1} p,$$
$$\dot{p} = -\nabla_\theta H(\theta, p) = \frac{1}{2} [p^\top G(\theta)^{-\top} (\partial_{\theta_k} G(\theta)) G(\theta)^{-1} p]_{k=1}^m - \nabla_\theta F(\theta).$$

We call this ODE system the Parametrized Wasserstein Hamiltonian flow.

Derivation of parametrized WH flow

Starting with a given Hamiltonian $\mathcal{H}(\rho,\Phi)$

$\Downarrow$

By taking Legendre Transform of $\mathcal{H}$, we obtain Lagrangian $\mathcal{L}(\rho,\partial_t\rho)$

$\Downarrow$

Using $\mathcal{L}$, we can define L on $\mathcal{T}\Theta$ as $L(\theta,\dot{\theta})=\mathcal{L}((T_{\cdot\sharp}\lambda)(\theta),(T_{\theta\sharp}\lambda)_*\dot{\theta})$

$\Downarrow$

Apply Legendre transform to L, we obtain the Hamiltonian $H(\theta,p)$ on $\mathcal{T}^*\Theta$

$\Downarrow$

We formulate the parameterized Wasserstein Hamiltonian flow as

$$\dot{\theta}(t)=\partial_p H(\theta(t),p(t)),$$

$$\dot{p}(t)=-\partial_\theta H(\theta(t),p(t)).$$

Variant of Parametrized Wasserstein Hamiltonian flow

$G(\theta)$ is often difficult to compute.
We replace it with a simplified version

$$\widehat{G}(\theta) = \int_{\mathbb{R}^d} \partial_\theta T_\theta(z)^\top \partial_\theta T_\theta(z) \, d\lambda(z).$$

Either $G(\theta)$ or $\widehat{G}(\theta)$ could be degenerate for certain choices of T_θ.
We use its pseudo-inverse $\widehat{G}(\theta)^\dagger$ to get a variant of parameterized WH flow:

$$\dot{\theta} = \widehat{G}(\theta)^\dagger p$$
$$\dot{p} = \frac{1}{2}[(\widehat{G}(\theta)^\dagger p)^\top (\partial_{\theta_k} \widehat{G}(\theta)) \widehat{G}(\theta)^\dagger p]_{k=1}^m - \nabla_\theta F(\theta)$$

It is also a *Hamiltonian system* with

$$H(\theta, p) = \frac{1}{2} p^\top \widehat{G}^\dagger p + F(\theta)$$

where p_0 is in the range of $\widehat{G}(\theta_0)$.

Symplectic integrator for PWHF

To solve parameterized WH flow, we use the *symplectic Euler scheme*:

$$\frac{\theta_{k+1} - \theta_k}{h} = \nabla_p H(\theta_{k+1}, p_k) = \widehat{G}(\theta_{k+1})^\dagger p_k,$$
$$\frac{p_{k+1} - p_k}{h} = -\nabla_\theta H(\theta_{k+1}, p_k),$$

The first equation is implicit and we solve it using least squares solver.

Pushforward map

- Since the trajectory of the Hamiltonian flow can intersect in the configuration space $\mathbb{R}^n$, we don't require the pushforward map T_θ to be invertible.
- There are different ways to choose T_θ:
  - T_θ: Affine, $T_\theta(x) = Ux + b$, $\theta = (U, b), U \in GL_d(\mathbb{R}),\ b \in \mathbb{R}^d$;
  - T_θ: Fourier series;
  - T_θ: Invertible neural networks such as normalizing flow;
  - T_θ: Non-invertible neural networks, for example, multi-layer perceptron (MLP)

Harmonic Oscillator as Wasserstein Hamiltonian flow

Take the potential function and initial Φ to be:

$$\begin{aligned} V(x) &= \Sigma_{i=1}^{d} \frac{1}{2}(a_i x_i)^2 \\ \Phi(0,x) &= \Sigma_{i=1}^{d} \frac{1}{2} b_i x_i^2 \end{aligned} \tag{1}$$

Then the i-th component of solution is given as:

$$X_i(t,x) = \sqrt{1+b_i^2} \cdot x_0^i \cdot \cos(|a_i| \cdot t - \arctan(b_i)) \tag{2}$$

In this example the solution ρ_t may develop singularity in finite time t_s, i.e., choose $t_s = \frac{1}{|a_i|}(\frac{\pi}{2} + arctan(b_i))$, we have:

$$X_i(t_s,x) = 0 \text{ and } \rho_{t_s}(x) = \delta_0(x), \quad \forall x \tag{3}$$

Examples 1: 2-D Harmonic Oscillator with linear pushforward map

We take $d = 2$, $V(x) = \frac{1}{2}(0.7x_1)^2 + \frac{1}{2}(0.6x_2)^2$, we choose the affine transformation as the pushforward map:

$$T_\theta(z) = \Gamma z + b, \theta = (\Gamma, b), \Gamma \in R^{2\times 2} \tag{4}$$

Assume the initial condition is:

$$\begin{aligned} &\rho_0 = \mathcal{N}(\vec{0}, I), \quad \Phi_0(x) = -\frac{1}{2}x_1^2 \\ &\theta_0 = (diag(0.7, 0.6), \vec{0}),\ \ p_0 = (diag(-0.7, 0), \vec{0}) \end{aligned} \tag{5}$$

Examples 1: 2-D Harmonic Oscillator with linear pushforward map

We randomly choose a test sample, compare its trajectory with our network solution:

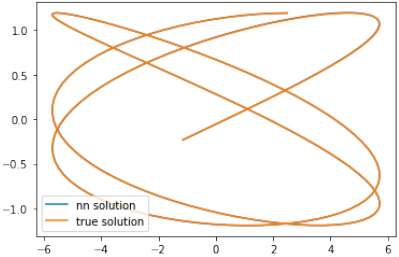


Figure: Trajectory under the Hamiltonian dynamics, $t_1 = 20$

Observation: Our method can compute beyond the singular time of WHF.

Examples 1: Verification of linear error against stepsize

We verify the solution error against stepsize on this example. We change the stepsize from 0.002 to 0.01 and plotted the error-stepsize curve. For convenience, we only consider the trajectory error here:

$$error = sup_{0<t<1} \int ||X_t(z) - X_{\theta_t}(z)||_{l^2} p(z) dz \tag{6}$$

We can see the linear dependence in the results:

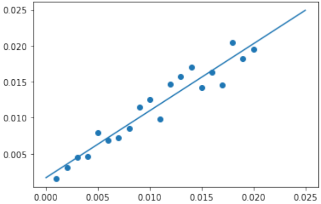


Figure: error vs stepsize Δt

Examples 1: Symplectic preservation of the symplectic Euler scheme

Here we compare our symplectic scheme with explicit Euler discretization on the harmonic oscillator example:

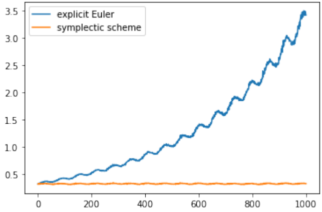


Figure: Hamiltonian, stepsize=0.1

Examples 2: 2-dim Harmonic Oscillator with MLP pushforward map

We also run the experiments for a 2-dimensional harmonic oscillator. Take $V(x) = \frac{1}{2}x_1^2 + \frac{1}{2}(0.75x_2)^2, \Phi(0, x) = -\frac{1}{2}x_1^2$.

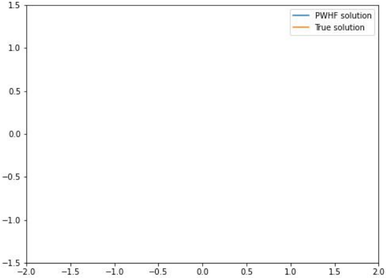


Figure: Trajectory of a random initial point

Examples 2: 10-dim Harmonic Oscillator with MLP pushforward map

And here is an example of 10-dim harmonic oscillator . The following is projection of pushforward distribution compared to the true density function:

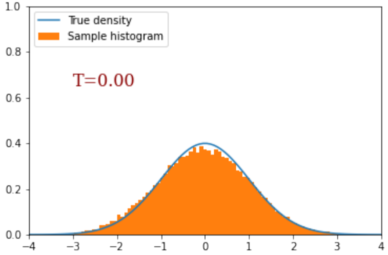


Figure: histogram: projection of pushforward distribution

Summary

PWHF naturally builds the connection between Eulerian (PDE) formulation and Lagrangian (particle) formulation of physical systems:

Parametric

$\Theta: \ (\theta_t, \dot{\theta}_t, p_t)$

Parametric Hamiltonian flow:

$\dot{\theta}_t = \nabla_p H^{\Theta}(\theta_t, p_t), \ \dot{p}_t = -\nabla_\theta H^{\Theta}(\theta_t, p_t).$

$$\boldsymbol{X}_t^{\Theta} = T_{\theta_t}(\boldsymbol{Z}) \ (\boldsymbol{Z} \sim \lambda) \ \approx \ \boldsymbol{X}_t;$$
$$\dot{\boldsymbol{X}}_t^{\Theta} = \frac{\partial T_{\theta_t}(\boldsymbol{Z})}{\partial \theta} \dot{\theta}_t \ \approx \ \dot{\boldsymbol{X}}_t;$$
$$\boldsymbol{P}_t^{\Theta} = \nabla_{\boldsymbol{P}} H(\boldsymbol{X}_t^{\Theta}, \cdot)^{-1} \left(\frac{\partial T_{\theta_t}(\boldsymbol{Z})}{\partial \theta} \nabla_p H^{\Theta}(\theta_t, p_t) \right) \ \approx \ \boldsymbol{P}_t.$$

$$\rho_{\theta_t} = T_{\theta_t \sharp} \lambda \ \approx \ \rho_t;$$
$$\frac{d\rho_{\theta_t}}{dt} = -\nabla \cdot \left(\rho_{\theta_t} \frac{\partial T_{\theta_t}}{\partial \theta} \circ T_{\theta_t}^{-1}(\cdot) \dot{\theta}_t\right) \ \approx \ \dot{\rho}_t;$$
$$\vec{\mathrm{u}}_{\Theta}(x,t) = \nabla_{\boldsymbol{P}} H(x, \cdot)^{-1} \left(\frac{\partial T_{\theta_t}}{\partial \theta} \circ T_{\theta_t}^{-1}(x) \nabla_p H^{\Theta}(\theta_t, p_t) \right) \ \approx \ \nabla \psi(x,t).$$

$\mathbb{R}^d: \ (\boldsymbol{X}_t, \dot{\boldsymbol{X}}_t, \boldsymbol{P}_t)$

Hamiltonian system:

$\dot{\boldsymbol{X}}_t = \nabla_{\boldsymbol{P}} H(\boldsymbol{X}_t, \boldsymbol{P}_t), \ \dot{\boldsymbol{P}}_t = -\nabla_{\boldsymbol{X}} H(\boldsymbol{X}_t, \boldsymbol{P}_t).$

$$\mathrm{Law}(\boldsymbol{X}_t) = \rho_t; \ \ \boldsymbol{P}_t = \nabla \psi(\boldsymbol{X}_t, t).$$

$\mathcal{P}(\mathbb{R}^d): \ (\rho_t, \dot{\rho}_t, \psi_t)$

Hamiltonian flow:

$\dot{\rho}_t = \frac{\delta}{\delta \psi} \mathcal{H}(\rho_t, \psi_t), \ \dot{\psi}_t = -\frac{\delta}{\delta \rho} \mathcal{H}(\rho_t, \psi_t).$

Lagrangian

Eulerian

Summary and Future work

Possible future directions

- Experiments on other types of T_θ such as Neural ODE;
- Apply the method to Schrödinger equation;
- Apply the method to more general Hamiltonian flows with non-quadratic kinetic energy. (In this case, the MINRES algorithm used in the symplectic Euler scheme should be replaced by a nonlinear solver, which could make the computation more challenging.)

Preprint at `https://arxiv.org/pdf/2306.00191.pdf`.

We welcome any comments and suggestions.

Thank you!