ONSAGER'S VARIATIONAL PRINCIPLE FOR SOLVING INVERSE PROBLEMS

RITOBAN ROY-CHOWDHURY*, IRINA TEZAUR†, NATHAN URBAN‡, AND ANTHONY GRUBER§

Abstract. Data driven models in physics face a perennial trade-off between being flexible and expressive, and being closely based on conventional physical models, which possess the advantages of being interpretable and have robust numerical guarantees. We advocate for variational principles as being the appropriate reconciliation between these two competing interests. In particular, the present work considers Onsager's variational principle, which describes thermodynamic systems that are dissipation dominated. The principle describes the dynamics of a system as the minimizer of a functional balancing the decay of free energy and the dissipative force, and admits an unconditionally energy stable time discretization. Within this framework, we allow the free energy and dissipation potential functionals to be parametrized by learnable weights. By optimizing these weights so that the resulting simulation matches training data, we can tackle inverse or system identification problems that can recover features of dynamics that may not be known in advance. We illustrate our method on several problems using synthetic data, with a particular interest towards phase field models used in modelling block copolymers.

*University of California, San Diego, rroychowdhury@ucsd.edu
†Sandia National Laboratories, CA
‡Brookhaven National Laboratory
§Sandia National Laboratories, NM

1. Introduction. A wide variety of dissipative systems play central roles in materials science, physics, and chemistry. Such systems model phenomena ranging from the nano-scale phase separation of polymers in solution to the continent-scale motion of glacial ice sheets. Yet, in many practical settings, the precise structure of the governing equations is only partially known. The precise potentials governing forces must often be inferred from experimental measurements or fine-scale simulations. This motivates the integration of data-driven methods into PDE modeling: by learning unknown components directly from data, one can close the gap between idealized theory and complex real-world behavior. Such hybrid models are particularly appealing in inverse problems, where the goal is to reconstruct hidden physical laws from observed dynamics.

Any data driven model for physics faces two conflicting demands: it should be flexible enough to describe the data provided, while also being constrained enough for the results to be physically reasonable and readily interpretable. Naively inserting neural networks or other flexible function approximators into PDEs often destroys fundamental structural properties of the underlying dynamics. In dissipative systems, the most important such property is energy stability. The total free energy of the system must decay monotonically in time; equivalently, the second law of thermodynamics states that the total entropy of a system must increase monotonically. Violating such physical principles leads to models that may overfit to training data but exhibit physically unreasonable behavior, such as energy growth or instability when evaluated out of distribution. Further, such unphysical behavior may compromise the optimization procedure, making it impossible for the data-driven model to converge in the first place.

Onsager's Variational Principle offers a natural balance between these two demands [13, 1]. The principle formulates dissipative dynamics as a minimization problem that balances the forces due to free energy and dissipation. We describe a natural time discretization of Onsager's Variational Principle for which the energy decreases unconditionally; as a result, we can be confident that regardless of the trajectory taken by the optimizer while solving inverse problems, the simulation problems will remain stable. The principle is sufficiently general so that terms in the free energy or dissipation potential can be represented by trainable, data-driven models. The result is a variationally consistent integration of machine learning with non-equilibrium thermodynamics: one that retains the stability guarantees of Onsager's framework while allowing expressive parameterization of unresolved physics.

2. Background on Onsager's Variational Principle. Before we tackle solving inverse problems, we need to understand the class of forward problems that we will consider. We adopt the class of systems satisfying **Onsager's Variational Principle** (OVP): such systems, emerging from non-equilibrium thermodynamics, are driven by a free energy, but are dissipation dominated (in the sense that they lack any inertia, and the resulting velocity is determined by the strength of the dissipative force) [1]. The class is flexible enough to include a variety of interesting partial differential equations (including phase field models like Allen-Cahn and Cahn-Hilliard, the Stokes" equations for highly viscous fluids, and PDEs emerging from Brownian motion like the diffusion and Fokker-Planck equations) – while being restricted enough that we can design an unconditionally stable time integrator which guarantees robustness during the inverse problem optimization.

We recommend [1] for an introduction to Onsager's principle, and largely adopt their terminology and notation. The principle was introduced originally in [12, 13].

DEFINITION 2.1. *Let Q be the configuration manifold of a system (possibly an infinite dimensional function space). Suppose we are given*

- *The **free energy** $\mathcal{F} : Q \to \mathbb{R}$, which the system seeks to minimize*
- *The **dissipation potential** $\mathcal{D} : E \to \mathbb{R}^+$, defined over E, a vector bundle over Q*

with fibers $E_{\mathbf{q}}$.

- *The* ***process function*** $\mathcal{P} : E \to TQ$*, which converts the process variables* E *to changes in configuration, and is linear in each fiber.*

The ***Rayleighian*** *at a point* $\mathbf{w} \in E_{\mathbf{q}}$ *is*

$$\begin{aligned} \mathcal{R} &: E \to \mathbb{R} \\ \mathcal{R}_{\mathbf{q}}(\mathbf{w}) &= \langle \mathrm{d}\mathcal{F}_{\mathbf{q}} \mid \mathcal{P}_{\mathbf{q}}\mathbf{w} \rangle + \mathcal{D}_{\mathbf{q}}(\mathbf{w}), \end{aligned}$$

and then ***Onsager's Variational Principle*** *defines the following ODE for* $\mathbf{q} : (0, T) \to Q$*,*

$$\dot{\mathbf{q}}(t) = \mathcal{P}_{\mathbf{q}(t)}\mathbf{w}^* \text{ where } \mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w} \in E_{\mathbf{q}(t)}} \mathcal{R}_{\mathbf{q}(t)}(\mathbf{w}) \tag{2.1}$$

When $\mathbf{w}$ achieves the minimum, the first term of the Rayleighian equals rate of change of the free energy

$$\left\langle \mathrm{d}\mathcal{F}_{\mathbf{q}(t)} \mid \mathcal{P}\mathbf{w}^* \right\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}(\mathbf{q}(t)).$$

For many problems, the process variables $E = TQ$, and the process function $\mathcal{P}$ will be the identity. But in several examples, particularly PDEs emerging from a continuity equation, the space of optimization variables $E_{\mathbf{q}}$ will be larger than the space of changes to configurations (for example, we may optimize over the space of all velocity vector fields, instead of simply changes to the concentration profile of a substance).

In order to better understand the class of physical systems described by Onsager's principle, we next explore the optimality conditions of the minimization and how to interpret them.

2.1. Perspectives. How should we understand the systems that Onsager's Principle describes? We present three different perspectives.

2.1.1. Force Balance. If $\mathcal{R}$ is differentiable, the optimality conditions for the minimization above are

$$\mathcal{P}^\top \mathrm{d}\mathcal{F}_{\mathbf{q}(t)} + \partial_{\mathbf{w}}\mathcal{D}_{\mathbf{q}(t)}(\mathbf{w}^*) = 0,$$

where $\mathcal{P}^\top : T^*Q \to E^*$ denotes the transpose of $\mathcal{P}$ and $\partial_{\mathbf{w}}\mathcal{D} : E \to E^*$ denotes the derivative in the fiber direction.

This can be interpreted as as balance law between the force associated to the free energy, and the corresponding dissipative force. For example, if one considers a mass connected to a spring, whose state is given by its position in 1-dimension $x \in Q = \mathbb{R}$, then the spring force emerges from the potential energy $\mathcal{F}(x) = \frac{1}{2}kx^2$, and a dissipative force emerges from the potential $\mathcal{D}(x, \dot{x}) = \frac{1}{2}\dot{x}^2$ (in this case, $\mathcal{P}$ is the identity). Then, optimality conditions are simply the force balance between the spring force and the dissipative force,

$$\dot{x} + kx = 0.$$

As with all systems satisfying OVP, there is no inertia: it is dissipation dominated, and forces from the potential are immediately counteracted by a corresponding dissipative force. The velocity is determined solely by this balance of forces.

2.1.2. Gradient Flow of Free Energy. Denote by $\flat = \partial_{\mathbf{w}} \mathcal{D} : E \to E^*$ (in the case where $\mathcal{D}$ is quadratic in $\mathbf{w}$, this defines an inner product by $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = \langle \mathbf{w}_1 \mid \flat \mathbf{w}_2 \rangle$, motivating the notation), and $\sharp = \flat^{-1}$. Then, the optimality conditions above imply the ODE from Onsager's principle is

$$\dot{\mathbf{q}}(t) = -\mathcal{P} \sharp \mathcal{P}^\top \mathrm{d}\mathcal{F}_{\mathbf{q}(t)}.$$

Therefore, Onsager's principle is a negative gradient flow of the free energy.

If $\flat$ is symmetric positive semi-definite (PSD) in $\mathbf{w}$, the free energy will monotonically decrease along solutions to OVP.

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{F}\left(\mathbf{q}(t)\right) = \left\langle \mathrm{d}\mathcal{F}_{\mathbf{q}(t)} \mid \mathcal{P}\mathbf{w}^* \right\rangle = -\left\langle \flat \mathbf{w}^* \mid \mathbf{w}^* \right\rangle \leq 0.$$

A sufficient condition for $\flat$ to be PSD is the dissipation potential $\mathcal{D}$ being convex in $\mathbf{w}$ and vanishing at zero $\mathcal{D}_{\mathbf{q}}(0) = 0$. Then, for any $\mathbf{w} \in E_{\mathbf{q}}$, $\langle \mathbf{w} \mid \flat \mathbf{w} \rangle \geq \mathcal{D}_{\mathbf{q}}\left(\mathbf{w}\right) - \mathcal{D}_{\mathbf{q}}(0) \geq 0$.

2.1.3. Non-equilibrium Principle of Maximum Entropy. The equilibrium thermodynamics, state variables are functions of a systems internal energy U, volume V, and number of molecules N, or their conjugate quantities. For example, If β denotes inverse temperature, then the thermodynamic free energy

$$\mathcal{F}(\beta, V, N) = \inf_U \{\beta U - S(U, V, N)\},$$

is the Legendre transform (up to reflections) of the entropy S in the internal energy variable. In equilibrium thermodynamics, the fundamental postulate is that systems will settle in configurations that maximize entropy, or equivalently minimize the above free energy [5], [3]. Onsager's principle is thus a non-equilibrium generalization of this principle, that shows which path the system takes to reach the equilibrium configuration: the negative gradient flow under the metric $\mathcal{P}\sharp\mathcal{P}^*$ [13].

2.2. Examples. There are several common partial differential equations that emerge from Onsager's variational principle, that we can describe now and will later use in numerical examples. For a more complete list, we recommend [4], [1].

It is clear from Section 2.1.2 that any gradient flow on an inner product space or Riemannian manifold $\dot{\mathbf{q}}(t) = -\sharp \mathrm{d}\mathcal{F}_{\mathbf{q}(t)}$ is automatically formally an Onsager system: one can simply take $\mathcal{D}\left(\mathbf{w}\right) = \frac{1}{2}\left\langle \mathbf{w}\, \mathbf{w} \right\rangle$ to be the quadratic form associated to the inner product under which the gradient flow is defined, and $\mathcal{P}$ to be the identity.

For example, it is well known that the heat or diffusion equation $\dot{c} = \Delta c$ is the L^2-gradient flow of the Dirichlet energy $\mathcal{F}(c) = \int_{\mathcal{M}} |\nabla c|^2 \mathrm{d}V$, and is thus is a system satisfying Onsager's variational principle. But Onsager's principle is not merely about arbitrary gradient flows: it provides a physical interpretation of each of the quantities involved, from which the particular gradient flow emerges. In the diffusion equation case, the Dirichlet energy is simply not the physically correct free energy. The free energy density of a solute with concentration c should be of the form $c \log c$. A more physically justifiable perspective emerges from treating the diffusion equation as a Wasserstein gradient flow.

2.2.1. Wasserstein Gradient Flows. A more enlightening formulation of the diffusion equation turns out to be as a Wasserstein gradient flow. In this case, the configuration

space $Q = \{c \in C^\infty_+(\mathcal{M}) : \int_\mathcal{M} c\, dV = 1\}$ is the space of smooth and non-negative concentration fields $c >= 0$ integrating to 1, defined on a manifold $\mathcal{M}$ with boundary $\partial\mathcal{M}$. Then, the process variables that we optimize over $E_\mathbf{q} = \{\mathbf{w} \in \Gamma(T\mathcal{M}) : \langle \mathbf{w}\, \mathbf{n}\rangle = 0 \text{ on } \partial\mathcal{M}$ consist of vector fields tangent to the boundary, which affect the concentration by the continuity equation encoded in $\mathcal{P}$,

$$\mathcal{P}(c, \mathbf{w}) = -\operatorname{div} c\mathbf{w}$$

so that the ordinary differential equation (ODE) 2.1 becomes $\dot{c} = \mathcal{P}(c, \mathbf{w})$. This equation describes the conservation of mass for a substance of concentration c. Finally, the dissipation potential is

$$\mathcal{D}(c, \mathbf{w}) = \frac{1}{2}\int_\mathcal{M} c\,|\mathbf{w}|^2\, \mathrm{d}V.$$

THEOREM 2.2. *For any functional $\mathcal{F} : Q \to \mathbb{R}$, in the above setup,*

$$\mathcal{P}\sharp\mathcal{P}^\top \mathrm{d}\mathcal{F} = -\nabla\cdot(c\nabla\mu(c)),$$

where $\mu(c) = \operatorname{grad}_{L^2}\mathcal{F}(c)$ is called the chemical potential of c. Further, at minimizer in OVP, c will satisfy Neumann boundary conditions $\partial_\mathbf{n} c = 0$ on $\partial\mathcal{M}$.

COROLLARY 2.3. *For the free energy*

$$\mathcal{F}(c) = \int_\mathcal{M} c\log c + cU\ \mathrm{d}V$$

where $U : \mathcal{M} \to \mathbb{R}$ is a potential, the OVP system is the Fokker-Planck equation

$$\begin{aligned}\dot{c} &= \Delta c + \nabla\cdot(c\nabla U) \text{ in } \mathcal{M}\\ \partial_\mathbf{n} c &= 0 \text{ on } \partial\mathcal{M}.\end{aligned}$$

COROLLARY 2.4. *If further, $V = 0$, then the OVP system is the diffusion equation, $\dot{c} = \Delta c$.*

COROLLARY 2.5. *For the free energy*

$$\mathcal{F}(c) = \int_\mathcal{M} c^m \mathrm{d}V,$$

the OVP system is the Porous medium equation

$$\dot{c} = \Delta c^m \text{ in } \mathcal{M}.$$

Remark: Theorem 2.2 actually shows that $\mathcal{P}\sharp\mathcal{P}^\top\mathrm{d}\mathcal{F}$ equals the Wasserstein gradient of the functional $\mathcal{F}$. Indeed, as first observed by [14], the problem of optimal transport is a Riemannian geodesic problem on the space of probability measures, equipped with the inner product at a probability measure μ, in the directions $\mathring{\mu_0}, \mathring{\mu_0}$

$$\langle \mathring{\mu_0}\, \mathring{\mu_1}\rangle^{W_2}_\mu = \int_{\mathbb{R}^d} \langle \nabla u_0\, \nabla u_1\rangle\, \mathrm{d}\mu \qquad \text{where} \qquad -\nabla\cdot(\mu\nabla u_k) = \mathring{\mu_k}, \quad k = 0, 1.$$

Under this inner product, the gradient of a functional is precisely

$$\operatorname{grad}_{W_2} \mathcal{F}(\rho) = -\nabla \cdot \left(c \nabla \operatorname{grad}_{L^2} \mathcal{F}(c)\right).$$

For a readable introduction to the theory of optimal transport, we recommend [17].

2.2.2. Phase Field Models from Sobolev Gradient Flows. Of particular interest are phase field models, which model the concentration of a solute $c : \mathcal{M} \to [-1, 1]$, and capture phase separation behavior. We are particularly interested in the application of the Cahn-Hilliard equation for modeling di-block copolymers, which are polymer chains consisting of blocks of two distinct species. The two polymer species generally repel each other, causing the solution to separate into distinct phases, while bonding between the species limits the extent of this separation and can cause interesting patterns to form [10].

The Allen-Cahn Equation is the L^2 gradient flow of the Ginzburg-Landau free energy. Therefore, it emerges from Onsager's principle with the free energy

$$\mathcal{F}(c) = \int_{\mathcal{M}} \frac{\alpha}{2} |\nabla c|^2 + F(c) \mathrm{d}V \tag{2.2}$$

where $F : \mathbb{R} \to \mathbb{R}$ is the bulk free energy density depending on c, generally a double well potential such as

$$F(c) = \frac{\theta}{2}\left[(1-c)\log(1-c) + (1+c)\log(1+c)\right] - \frac{\theta_c}{2} c^2,$$

where θ is the absolute temperature of the system, and θ_c is the critical temperature of phase separation. A common approximation is

$$F(c) \approx \frac{\left(1-c^2\right)^2}{4},$$

sometimes called the "shallow quenching" limit [18]. These potentials have an interval $(-a, a)$ on which $F'' \leq 0$. This region is called the "spinodal regime", and concentrations in that interval will tend to separate. The L^2 gradient of free energy is called the chemical potential $\mu(c) \coloneqq \operatorname{grad}_{L^2} \mathcal{F}(c) = -\alpha \Delta c + F'(c)$.

The L^2 gradient structure yields the Allen-Cahn equation; it can be derived from OVP in the setting where $E = TQ$, $\mathcal{P} = \mathrm{Id}$, and

$$\mathcal{D}(c, \dot{c}) = \frac{1}{2} \int_{\mathcal{M}} \dot{c}^2 \mathrm{d}V, \tag{2.3}$$

and the resulting PDE is simply

$$\begin{aligned} \dot{c} &= \alpha \Delta c - F'(c) \text{ in } \mathcal{M} \\ \partial_{\mathbf{n}} c &= 0 \text{ on } \partial\mathcal{M}. \end{aligned}$$

The Cahn-Hilliard equation, on the other hand, is well known to be the H^{-1}-gradient flow of the same Ginzburg-Landau free energy 2.2. It emerges from Onsager's principle similar to the Wasserstein gradient flows above, where E_c is the space of vector fields tangent to the boundary, and $\mathcal{P}(c, \mathbf{w}) = -\operatorname{div} c\mathbf{w}$.

Let the **mobility** $M : T^*\mathcal{M} \to T\mathcal{M}$ be an invertible linear map on the domain $\mathcal{M}$ (it can be interpreted as mapping forces to velocities), and dissipation potential for the Cahn-Hilliard equation is

$$\mathcal{D}(c, \mathbf{w}) = \frac{1}{2} \int_{\mathcal{M}} |c\mathbf{w}|^2_{M^{-1}} \, \mathrm{d}V.$$

One can check that $\mathcal{P}\sharp\mathcal{P}^\top \mathrm{d}\mathcal{F}$ in this setting corresponds to the operator $-\operatorname{div} M\nabla \left(\operatorname{grad}_{L^2} \mathcal{F}\right)$, and thus OVP yields the M-weighted H^{-1} gradient flow of the free energy $\mathcal{F}$. Altogether, Onsager's principle becomes the Cahn-Hilliard equations.

$$\begin{aligned} \dot{c} &= \operatorname{div} M\nabla\mu \text{ in } \mathcal{M} \\ \mu &= -\alpha\Delta c + F'(c) \text{ in } \mathcal{M}. \partial_{\mathbf{n}} c = \partial_{\mathbf{n}} \mu = 0 \text{ on } \partial\mathcal{M} \end{aligned}$$

3. Time Discretization. Next, we present a time discretization of OVP. We particularly seek a discretization that respects features of the continuous structure. Perhaps most crucially, the free energy should decrease at each time step, consistent with the monotonically decreasing free energy in the continuous case.

With these considerations in mind, we present the following time discretization of systems obeying Onsager's principle.

Definition 3.1. *Given an Onsager system $(Q, E, \mathcal{F}, \mathcal{D}, \mathcal{P})$ for which Q is an affine space, we define the time-discrete Rayleighian $\mathcal{R}^d : E \to \mathbb{R}$ by*

$$\mathcal{R}^d(\mathbf{q}, \mathbf{w}) = \mathcal{F}(\mathbf{q} + \Delta t \mathcal{P}_{\mathbf{q}} \mathbf{w}) + \Delta t \mathcal{D}(\mathbf{q}, \mathbf{w})), \tag{3.1}$$

and a time-discrete Onsager system $\mathbf{q}^\bullet : \mathbb{N} \to Q$ by the difference equation

$$\mathit{Step}_{\mathcal{R}}\left(\mathbf{q}^k\right) \coloneqq \mathbf{q}^{k+1} = \mathbf{q}^k + \Delta t \mathcal{P}\left(\mathbf{q}^k, \mathbf{w}^*\right) \qquad \mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w} \in E_{\mathbf{q}^k}} \mathcal{R}^d\left(\mathbf{q}^k, \mathbf{w}\right). \tag{3.2}$$

Note that the Rayleighian (3.1) differs from the continuous setting by a factor of Δt (merely for convenience), and the fact that the directional derivative $\langle \mathrm{d}\mathcal{F}_{\mathbf{q}} \mid \mathcal{P}(\mathbf{w}) \rangle$ is approximated by the first order expansion

$$\langle \mathrm{d}\mathcal{F}_{\mathbf{q}} \mid \mathcal{P}(\mathbf{w}^*) \rangle \approx \frac{\mathcal{F}(\mathbf{q}^{k+1}) - \mathcal{F}(\mathbf{q}^k)}{\Delta t},$$

and then the $-\frac{\mathcal{F}(\mathbf{q}^k)}{\Delta t}$ term is ignored (because it does not depend on $\mathbf{w}$, it does not affect the argmin). This approximation is particularly attractive because it results in an integrator that is unconditionally energy stable.

Theorem 3.2. *For a discrete-time Onsager system satisfying $\mathcal{D}_{\mathbf{q}}(0) = 0$, free energy decreases over time $\mathcal{F}(\mathbf{q}^{k+1}) \leq \mathcal{F}(\mathbf{q}^k)$.*

Proof. Because the dissipation potential $\mathcal{D}$ is non-negative, $\mathbf{w}^*$ minimizes $\mathcal{R}^d$, and because $\mathcal{D}(\cdot, 0) = 0$,

$$\mathcal{F}(\mathbf{q}^{k+1}) = \mathcal{F}(\mathbf{q}^k + \Delta t \mathcal{P}(\mathbf{q}^k, \mathbf{w}^*)) \leq \mathcal{R}^d\left(\mathbf{q}^k, \mathbf{w}^*\right) \leq \mathcal{R}^d\left(\mathbf{q}^k, 0\right) = \mathcal{F}(\mathbf{q}^k).$$

∎

We illustrate the evolution of three different OVP systems in Fig. 3.1, simulating them using the time discretization above, and a spatial discretization described in the appendix. One can observe from these simulations that they are energy stable. We plot the free energy over time for each in Fig. 3.2; it is clear that the energy decreases monotonically.

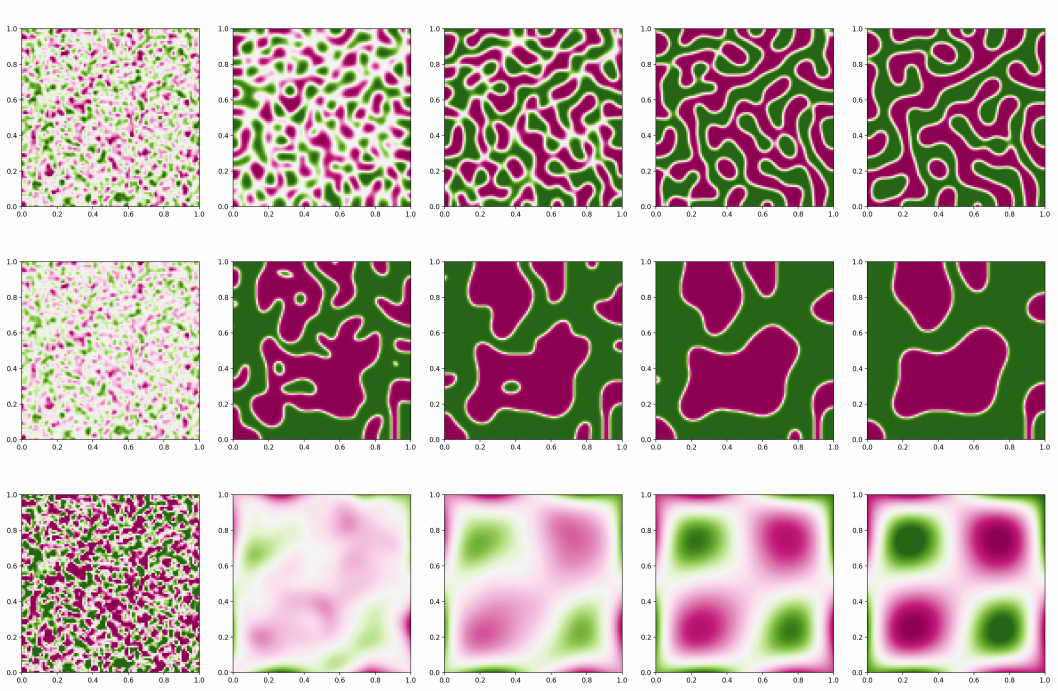

FIG. 3.1. *Frames from a simulations of the Cahn-Hilliard (top), Allen-Cahn (midddle), and Fokker-Planck (with $U(x, y) = \sin 3x \sin 3y$) (bottom) equations.*

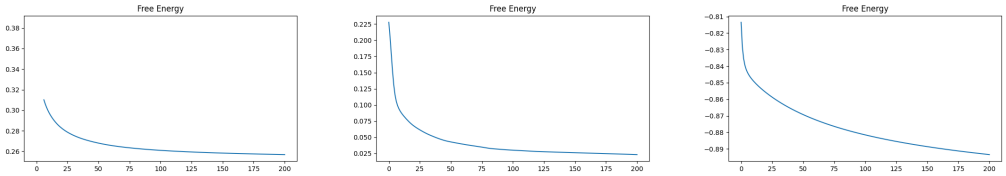

FIG. 3.2. *The free energy over time for each of the simulations in Figure 3.1.*

3.1. Relation to Other Formulations.

3.1.1. Constraints on E. One particularly recent work that explores variational integration for systems described by OVP is [4]. Their description of the principle is quite similar to ours, as is their discrete-time version. One minor difference is that their implementations compute the argmin in Eqn. 3.2 by using it's optimality conditions, instead of performing gradient-based optimization with auto-differentiation like we do. More interestingly, in their exposition, several systems have the minimization in Eqn. 2.1 performed over the set of $\mathbf{w}$ that satisfy some linear constraints $C_{\mathbf{q}}\mathbf{w} = 0$, which are then enforced using Lagrange multipliers. Formally, these linear constraints can simply be understood as part of the definition of the bundle E in our description. Interestingly, in the Wasserstein gradient flow cases, they consider the minimization to occur over both $\dot{c} \in T_c Q$ and vector fields $\mathbf{w}$), subject to the continuity equation $\dot{c} + \operatorname{div} c\mathbf{w} = 0$ as a linear constraint. We argue that treating this as a constraint is redundant: one can completely recover $\dot{c}$ from the vector field $\mathbf{w}$ using the continuity equation, and correspondingly only minimize over $\mathbf{w}$ in all of

our examples.

3.1.2. Jordan-Kinderleherer-Otto Scheme. Another interesting similarity arises from the optimal transport context. For Wasserstein gradient flows, our time discretization can be understood as a approximate version of the Jordan-Kinderlehrer-Otto (JKO) scheme [7]. The JKO scheme is generally written

$$\mathbf{q}^{k+1} = \underset{\mathbf{q}' \in Q}{\operatorname{argmin}} \mathcal{F}(\mathbf{q}') + \frac{1}{2\Delta t} W_2\left(\mathbf{q}', \mathbf{q}^k\right)^2,$$

in which $W_2\left(\mathbf{q}', \mathbf{q}^k\right)$ is the 2-Wasserstein distance between the density functions $\mathbf{q}', \mathbf{q}^k$. The most obvious difference between this and Eqn. 3.2 is that this is an optimization of the next state $\mathbf{q}$, rather than over the velocity $\mathbf{w}$. But this is no real difference: in the case where $\mathcal{P}_{\mathbf{q}}$ is surjective, one could rewrite 3.2 as a minimization over all possible $\mathbf{q}' \in Q$ and $\mathbf{w} \in E_{\mathbf{q}^k}$ for which $\mathcal{P}(\mathbf{q}^k, \mathbf{w}) = \frac{\mathbf{q}' - \mathbf{q}^k}{\Delta t}$ as follows.

$$\text{Step}_{\mathcal{R}}\left(\mathbf{q}^k\right) = \underset{\mathbf{q}' \in Q}{\operatorname{argmin}} \left[\mathcal{F}(\mathbf{q}') + \Delta t \inf_{\mathbf{w}} \mathcal{D}(\mathbf{q}^k, \mathbf{w})\right]. \tag{3.3}$$

More substantive is the difference in the dissipation term. The Wasserstein distance between two density functions $\mathbf{q}_0, \mathbf{q}_1$ can be computed using the Benamou-Brenier formula as an infimum over time-dependent velocity fields $\mathbf{w}_t$ that transport $\mathbf{q}_0$ onto $\mathbf{q}_1$ under the continuity equation (stated informally here, see [17] for precise regularity conditions):

$$W_2\left(\mathbf{q}_0, \mathbf{q}_1\right)^2 = \inf\left\{\int_0^1 \|\mathbf{w}_t\|^2_{L^2(\mathbf{q}_t)} \, \mathrm{d}t : \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{q}_t + \operatorname{div}\left(\mathbf{w}_t \mathbf{q}_t\right) = 0\right\}.$$

Then, the infimum in Eqn. 3.3 can be understood as a one step approximation of this quantity, because $\mathcal{D}(\mathbf{q}, \mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2_{L^2(\mathbf{q})}$ and $\mathcal{P}(\mathbf{q}^k, \mathbf{w}) = \frac{\mathbf{q}' - \mathbf{q}^k}{\Delta t}$ is an approximation of the continuity equation. Therefore, our time discretization of Onsager's principle can be understood as a approximate version of the JKO scheme for Wasserstein gradient flows, which approximates the Benamou-Brenier integral by one point evaluation of the dissipation potential.

4. Solving Inverse Problems. Describing systems by Onsager's Variational Principle readily lends itself to solving inverse problems. In many real world scenarios, the precise forms of the free energy $\mathcal{F}$ and dissipation potential $\mathcal{D}$ are not necessarily known. While it's common to make approximations (for example, the shallow and deep quenching limits in the Cahn-Hilliard equation [18]), it would be valuable to be able to fit the model to data obtained from a laboratory or higher fidelity simulations. Then, one could use the simulation generated by Onsager's principle as an accurate but computationally efficient low order model, that still respects the thermodynamic properties that one would hope for. Our goal is then to learn the underlying physics – the potentials $\mathcal{F}$ and $\mathcal{D}$ from data.

We approach this goal by parametrizing the potentials by some learnable collection of parameters $\theta \in \mathbb{R}^p$, which we denote $\mathcal{F}_\theta$ and $\mathcal{D}_\theta$, which define a Rayleighian $\mathcal{R}_\theta$, as Eqn. 3.1. The inverse problem then becomes an optimization problem: find the parameters θ that minimize the discrepancy between the observed data and the predictions of OVP model.

In particular, given a dataset of n trajectories $\left\{\left(\mathbf{q}_i^1, \ldots, \mathbf{q}_i^T\right)\right\}_{i=1}^n$ optimize the θ to minimize the loss function

$$\text{Loss}(\theta) = \mathbb{E}_{\substack{i \in (1,\ldots n) \\ l \in (1,\ldots,4) \\ k \in (1,\ldots,T-l)}} \left[\sum_{t=0}^{l} \left\|\mathbf{q}_i^{k+t} - \text{Step}_{\mathcal{R}_\theta}^{\circ t}\left(\mathbf{q}_i^k\right)\right\|^2\right],$$

where $\text{Step}^{\circ t}$ means the t-fold composition of Step.

Remark 1. To perform this optimization, we need to be able to differentiate the output of the variational integrator step $\mathbf{q}^{k+1} = \text{Step}_{\mathcal{R}_\theta}\left(\mathbf{q}^k\right)$ with respect to the parameters θ. That requires us to be able to differentiate through the minimization $\mathbf{w}^* = \text{argmin}_{\mathbf{w}\in E_\mathbf{q}} \mathcal{R}^d_\theta(\mathbf{q}, \mathbf{w})$. We can do this by applying implicit differentiation to the optimality conditions of the minimization as in [2].

To compute $\frac{\partial \mathbf{w}^*}{\partial \theta}$, note that $\mathbf{w}^*$ is implicitly defined by

$$A(\mathbf{w}^*, \theta) \coloneqq \left(\partial_\mathbf{w} \mathcal{R}^d_\theta\right)(\mathbf{q}, \mathbf{w}^*) = 0.$$

where $A : E \times \mathbb{R}^p \to E^*$. Then, by the implicit function theorem,

$$\frac{\partial \mathbf{w}^*}{\partial \theta} = -\left(\partial_\mathbf{w} A\right)^{-1}\left(\partial_\theta A\right).$$

In practice, we use the implementation of implicit differentiation from TorchOpt [16], using the Adam optimizer (for both the "inner" and "outer" optimizations), and conjugate gradient for computing $\left(\partial_\mathbf{w} A\right)^{-1}$ (which is symmetric, as the Hessian of $\mathcal{R}^d$).

Remark 2. In our examples, we always treat the process function $\mathcal{P}$ as fixed: it is either trivial (in which case, OVP is an optimization over all possible tangent vectors $\dot{\mathbf{q}} \in T_\mathbf{q} Q$), or enforces the continuity equation, so $\dot{c} = \mathcal{P}(c, \mathbf{w}) = -\operatorname{div} c\mathbf{w}$ (in which case, OVP is an optimization over all possible vector fields $\mathbf{w}$ on $\mathcal{M}$ that are tangent to the boundary). In principle, there is no reason why these are the only two possibilities: $\mathcal{P}$ could be learned. We are, however, not presently considering any systems where other choices of $\mathcal{P}$ would be physically interesting.

5. Numerical Results. We present an array of inverse problems below, all involving learning various features of the free energy in different PDE settings, from synthetic data (generated from simulations based on Onsager's principle). The spatial discretization for all of the problems is discussed in the Appendix A – overall, they are the natural finite-difference based discretizations of the functionals from Section 2.2.

Our goal in studying toy problems with synthetic data is not for the learned models to be a surrogate or more computationally efficient version of the traditional model. Instead, the goal is to illustrate our approach's ability to converge to parameters that generate simulations close to to the training data. That convergence, in turn, would give us the confidence to deploy the method on data that does not come from an OVP-derived simulation, expecting that the learned potentials are physically reasonable.

In future work, we plan on learning from both experimental data and data from more complex simulations (such as molecular dynamics simulations that capture the effects of the particular chemistry of the substances involved, rather than the mere macroscopic evolution of their concentration profiles). In those settings, the true free energy or dissipation potential are not necessarily known: they are emergent phenomena from some underlying physics. However, we would expect that the learned model that our approach provides would be flexible enough to precisely capture that behavior, while also obeying the underlying physical principles.

5.1. Cahn-Hilliard with a Polynomial Free Energy. As a preliminary inverse problem, we demonstrate that our method can recover a known polynomial free energy from simulations of the Cahn-Hilliard equation. Recall the Cahn-Hilliard (C-H) equation

TABLE 5.1
Parameters and Final Train/Validation Loss for each numerical experiment.

Experiment	Grid Size	LR	Batch Size	Δt	Train Loss	Validation Loss
C-H w/ Poly. Free Energy	100^2	0.05	128	10^{-4}	2.18×10^{-9}	2.33×10^{-9}
C-H w/ Nonlocal Term	100^2	0.05	128	10^{-4}	1.87×10^{-7}	3.77×10^{-7}
Diffusion Eqn.	100^2	0.02	128	10^{-4} – 10^{-2}	3.30×10^{-5}	2.71×10^{-4}

TABLE 5.2
Coefficients of the learned free energy polynomial.

1	x	x^2	x^3	x^4	x^5	x^6
1	1.0003	-5.008×10^{-1}	-1.656×10^{-6}	2.511×10^{-1}	9.869×10^{-7}	-6.974×10^{-4}

from Section 2.2.2

$$\dot{c} = \operatorname{div} M \nabla \mu \text{ in } \mathcal{M}$$
$$\mu = -\alpha \Delta c + F'(c) \text{ in } \mathcal{M}. \partial_{\mathbf{n}} c = \partial_{\mathbf{n}} \mu = 0 \text{ on } \partial \mathcal{M}.$$

We generate synthetic data by simulating the Cahn-Hilliard equation, using our variational time integrator on a 100×100 grid for 50 time steps, with $F(c) = \frac{\left(1-c^2\right)^2}{4}$ and $\alpha = 1$, from 20 different initial conditions that we generate by performing a Gaussian blur on a random concentration field.

We then parametrize the bulk free energy density F as a an arbitrary degree 6 polynomial with learnable coefficients, and also treat the Dirichlet energy coefficient α as a learnable parameter: this is a small learning problem, with only 8 free parameters that should be capable of reproducing the dynamics present in the training data.

We find that this is the case: the recovered coefficients are shown in Table 5.2, with $\alpha = 1.0012$. The recovered polynomial is almost exactly equal to $1 + x - \frac{1}{2}x^2 + \frac{1}{4}x^4$, which is the ground truth energy up to an affine offset, which does not affect the dynamics (note that only F' appears in the PDE, and even then under a Laplacian which contains constants in it's kernel).

5.2. Non-local Free Energy for Block Copolymers. We are particularly interested in a variant of the Cahn-Hilliard equation applied to model di-block copolymers. Block copolymers are conventionally modelled by a variant of the Cahn-Hilliard equation. This variant includes an extra term penalizing deviation from the mean in the interior,

$$\dot{c} = \Delta \mu - \sigma \left(c - \overline{c}\right)$$
$$\mu = -\alpha \Delta c + F'(c).$$

for some coefficient $\sigma \in \mathbb{R}$, and where $\overline{c}$ denotes the spatial average of c (which is a constant preserved by the evolution).

In the Cahn-Hilliard case, the choice of $\mathcal{D}$ and $\mathcal{P}$ yield an H^{-1}-gradient flow of the free energy. This modified version of the Cahn-Hilliard equation emerges from the following free energy

$$\mathcal{F}_{\mathrm{OK}}(c) = \int_{\mathcal{M}} F(c) + \frac{\alpha}{2} |\nabla c|^2 + \frac{\sigma}{2} \left(c - \overline{c}\right) (-\Delta)^{-1} \left(c - \overline{c}\right) \mathrm{d}V,$$

in which Δ^{-1} is interpreted as the inverse of the Laplacian with homogeneous Neumann boundary conditions. This free energy was first proposed by Ohta and Kawasaki to describe the strength of long-range interactions in block copolymers [11]. Correspondingly, it is highly nonlocal.

It also presents an interesting challenge to test our framework. We generate synthetic data from simulations of the Cahn-Hilliard equation including the non-local term (which is implemented explicitly; not as part of the OVP minimization). Then, we add a term to the free energy of the form $\int_{\mathcal{M}} \frac{\sigma}{2} (c - \bar{c}) (-K) (c - \bar{c}) \, \mathrm{d}V$, where K is represented by a spectral convolution layer: the building block of a Fourier Neural Operator [8].

We expect that this should be sufficiently expressive parametrization: the Laplace inverse operator is a linear operator, and is diagonalized in frequency space. Therefore, even one linear spectral convolution layer should be capable of representing it. Empirically, we find this to work reasonably well: training converges to a solution with low validation error (see Table 5.1), and visually the evolution under the learned free energy for simulations with parameters and like those in the training set match the ground truth simulations closely.

Nonetheless, when we test this on data that is sufficiently far from the training distribution, or with choices of Δt far from the training distribution, we are able to find examples for which the learned simulation differs from the ground truth simulation. This suggests that some regularization may be needed on the learned operators. One observation that may aid in regularization is that the learned potentials are often spatial discretization invariant. Therefore, we could augment the training data with data at a variety of resolutions. Motivated by this, we sought to explore other resolution invariant terms that we could add to the potentials, and decided to explore treating the diffusivity coefficient as an implicit neural field.

5.3. A Classical Inverse Problem: Diffusivity Coefficient. Finally, we investigate a classical inverse problem: given simulations of the diffusion equations, we seek to recover a spatially varying the diffusivity coefficient. Specifically, we simulate the initial-boundary value problem for a function $c : \mathcal{M} \to \mathbb{R}$ $\dot{c} = \nabla \dot{(} \alpha \nabla c)$ in $\mathcal{M}$ $\partial_{\mathbf{n}} c = 0$ on $\partial\mathcal{M}$. This emerges from Onsager's principle with the same setup as the usual diffusion equation, except with $\mathcal{D}_{\mathbf{q}}(\mathbf{w}) = \int_{\mathcal{M}} \alpha^{-1} c |\mathbf{w}|^2$ Our goal is to recover the unknown value of α, a spatially varying scalar field, from data that satisfies these equations. This is a classical inverse problem, and various existence and uniqueness results for it are well known [6].

We generate data by simulating the diffusion equation, setting the diffusivity coefficient (the reciprocal of which appears in the dissipation potential) to a fixed pattern: we pick the Sandia Logo, scaled to values between 1 and 50.

In the interest of keeping our learned potentials discretization invariant, and following recent work on implicit neural fields [15], we then choose to parametrize the diffusivity coefficient as a neural network that takes in coordinates (x, y), passes them through a 5 hidden layer MLP with tanh activation functions, and then produces a scalar (the reciprocal of the diffusivity coefficient) as an output. This is then used in our usual discretization of the diffusion equation.

Unfortunately, our optimizer seems to struggle substantially with this problem, and it seldom converges. This is in part because the effect of spatially varying diffusivity coefficients is quite small between adjacent frames, and also because diffusion causes the simulation to quickly become very uniform. Nonetheless, we still feel this is a promising avenue, and that there are a variety of scalar fields in the PDEs we've looked at may be more amenable to such an implicit representation (for example, the potential from the Fokker-Planck equation, or the mobility from the Cahn-Hilliard equation).

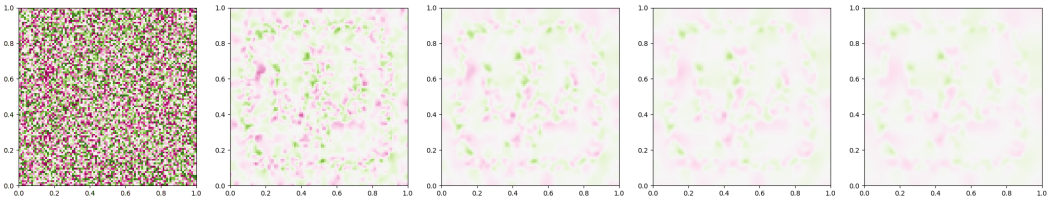

FIG. 5.1. *Frames from a simulation of the diffusion equation, in which the diffusion coefficient is based on the Sandia Thunderbird logo.*

6. Future Work & Conclusion. There are several additional experiments that we hope to investigate within this framework in the future, as we gain confidence about the robustness of the method. We'd like to apply the method to experimental data or higher fidelity simulations (perhaps molecular dynamics simulations for block copolymers), effectively using the parametrized Cahn-Hilliard based model as an efficient reduced order model that hopefully still captures features not captured by the Cahn-Hilliard model alone. We'd also like to further investigate the role of boundary conditions in the framework. The present work always uses the natural boundary conditions implied by the minimization, but there is much work on dynamic boundary conditions in the Cahn-Hilliard literature [18], and they are particularly interesting in the BCP context, where imposing various boundary conditions can be a means of controlling the outcome of the BCP melt.

A broader question this work raises in the realm of machine learning methods for physical problems, is where one should draw the line between physical assumptions and "black-box" machine learning models. On one side of this spectrum, one could make the entire solution operator of the PDE a black-box machine learned model with very few physical assumptions baked in [8], [9]. On the other side, one can adopt an entirely traditional PDE solver with only a small number of parameters left to be inferred from physical data. We would suggest that variational principles provide a natural middle ground between these two approaches. By allowing machine learned models to describe the functionals that give rise to a PDE, we can ensure that important physical properties (such as energy conservation or dissipation) are guaranteed to always hold, while benefiting from the increased flexibility and expressive power of machine learning methods.

References.

[1] M. Arroyo, N. Walani, A. Torres-Sánchez, and D. Kaurin, *Onsager's variational principle in soft matter: introduction and application to the dynamics of adsorption of proteins onto fluid membranes*, in The role of mechanics in the study of lipid bilayers, Springer, 2017, pp. 287–332.

[2] M. Blondel, Q. Berthet, M. Cuturi, R. Frostig, S. Hoyer, F. Llinares-López, F. Pedregosa, and J.-P. Vert, *Efficient and modular implicit differentiation*, Advances in neural information processing systems, 35 (2022), pp. 5230–5242.

[3] H. B. Callen and H. Scott, *Thermodynamics and an introduction to thermostatistics*, 1998.

[4] H. Chen, H. Liu, and X. Xu, *The onsager principle and structure preserving numerical schemes*, Journal of Computational Physics, 523 (2025), p. 113679.

[5] S. Friedli and Y. Velenik, *Statistical mechanics of lattice systems: a concrete mathematical introduction*, Cambridge University Press, 2018.

[6] V. Isakov, *Inverse problems for partial differential equations*, Springer, 2006.

[7] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the fokker–planck equation*, SIAM journal on mathematical analysis, 29 (1998), pp. 1–17.

[8] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, *Fourier neural operator for parametric partial differential equations*, arXiv preprint arXiv:2010.08895, (2020).

[9] L. Lu, P. Jin, and G. E. Karniadakis, *Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators*, arXiv preprint arXiv:1910.03193, (2019).

[10] D. Luo, L. Cao, P. Chen, O. Ghattas, and J. T. Oden, *Optimal design of chemoepitaxial guideposts for the directed self-assembly of block copolymer systems using an inexact newton algorithm*, Journal of Computational Physics, 485 (2023), p. 112101.

[11] T. Ohta and K. Kawasaki, *Equilibrium morphology of block copolymer melts*, Macromolecules, 19 (1986), pp. 2621–2632.

[12] L. Onsager, *Reciprocal relations in irreversible processes. i.*, Physical review, 37 (1931), p. 405.

[13] ——, *Reciprocal relations in irreversible processes. ii.*, Physical review, 38 (1931), p. 2265.

[14] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, (2001).

[15] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, Journal of Computational physics, 378 (2019), pp. 686–707.

[16] J. Ren, X. Feng, B. Liu, X. Pan, Y. Fu, L. Mai, and Y. Yang, *Torchopt: An efficient library for differentiable optimization*, Journal of Machine Learning Research, 24 (2023), pp. 1–14.

[17] C. Villani, *Topics in optimal transportation*, vol. 58, American Mathematical Soc., 2021.

[18] H. Wu, *A review on the cahn-hilliard equation: classical results and recent advances in dynamic boundary conditions*, arXiv preprint arXiv:2112.13812, (2021).

Appendix A. Spatial Discretization.

For all of our numerical experiments, we use a finite difference framework on a staggered grid of $n_x \times n_y$ cells, on the domain $\mathcal{M} = [0,1]^2$. We represent scalar fields, like concentration, as functions on cell centers, i.e. $c \in Q = \mathbb{R}^{n_x \times n_y}$, and vector fields on the faces of cells

$\mathbf{w} \in \mathbb{R}^{(n_x+1)\times n_y} \times \mathbb{R}^{n_x \times (n_y+1)}$.

We only need to describe the discrete functionals $\mathcal{F}$, $\mathcal{D}$, and $\mathcal{P}$ – all other discrete differential operators emerge automatically through auto-differentiation and the optimization in Eqn. 3.2. This is a particular advantage of the OVP framework, and makes implementing the forward simulation problems quite straightforward.

We begin with the Fokker-Planck equation, as an illustrative case, in which E_c consists of vector fields $\mathbf{w}$ represented on a staggered grid. Then, each continuous potential from Section 2.2 has a discrete spatially counterpart,

$$\mathcal{F}(c) = \int_{\mathcal{M}} c \log c + cU \, \mathrm{d}V \qquad \Rightarrow \qquad \mathcal{F}(c) = \sum_{i,j} \left(c_{i,j} \log c_{i,j} + c_{i,j} U_{i,j}\right) \Delta x \Delta y$$

$$\mathcal{D}\left(c, \mathbf{w}\right) = \frac{1}{2} \int_{\mathcal{M}} c \left|\mathbf{w}\right|^2 \mathrm{d}V \qquad \Rightarrow \qquad \mathcal{D}(c, \mathbf{w}) = \frac{1}{2} \sum_{i,j} c_{i,j} |\mathbf{w}_{i,j}|^2 \Delta x \Delta y$$

$$\mathcal{P}(c, \mathbf{w}) = -\operatorname{div} c\mathbf{w} \qquad \Rightarrow \qquad \mathcal{P}(c, \mathbf{w})_{i,j} = -\left[\frac{\mathbf{J}_{i+1/2,j} - \mathbf{J}_{i-1/2,j}}{\Delta x} + \frac{\mathbf{J}_{i,j+1/2} - \mathbf{J}_{i,j-1/2}}{\Delta y}\right]$$

where we need to use define $\mathbf{w}_{i,j} = 1/2\left(w_{i+1/2,j} + w_{i-1/2,j}, w_{i,j+1/2} + w_{i,j-1/2}\right)$, averaging $\mathbf{w}$ to cell centers, and define the discrete mass flux $\mathbf{J}_{i+1/2,j} = \frac{c_i + c_{i+1}}{2} w_{i+1/2,j}$ (averaging c to edges), and its other component is similar.

For the other PDEs we explore that are also Wasserstein gradient flows, we use the same choices of $\mathcal{D}$ and $\mathcal{P}$. To simulate the diffusion equation, we simply set $V = 0$ above, while the porous medium equation with exponent m emerges from OVP with the choice of

$$\mathcal{F}(c) = \int_{\mathcal{M}} c^m \mathrm{d}V \qquad \Rightarrow \qquad \mathcal{F}(c) = \sum_{i,j} c_{i,j}^m \Delta x \Delta y.$$

The Allen-Cahn system is the L^2 gradient of the free energy, and can be straightforwardly spatially discretized where $\mathcal{P} =$ Id. The discrete versions of 2.2, 2.3 respectively become

$$\mathcal{F}(c) = \textstyle\sum_{i,j} \left[\frac{c_{i+1/2,j} - c_{i-1/2,j}}{\Delta x}\right]^2 + \left[\frac{c_{i,j+1/2} - c_{i,j-1/2}}{\Delta y}\right]^2 + f\left(c_{i,j}\right) \Delta x \Delta y$$
$$\mathcal{D}(c, \dot{c}) = \tfrac{1}{2} \textstyle\sum_{i,j} \dot{c}_{i,j}^2 \Delta x \Delta y.$$

Finally, for the Cahn-Hilliard equation, the discrete free energy $\mathcal{F}$ is the same as for the Allen-Cahn equation, the process function $\mathcal{P}$ encodes the continuity equation, same as for Wasserstein gradient flows, and the dissipation potential with mobility matrix M becomes

$$\mathcal{D}(c, \mathbf{w}) = \int_{\mathcal{M}} |c\mathbf{w}|_{M^{-1}}^2 \, \mathrm{d}V \qquad \Rightarrow \qquad \mathcal{D}(c, \mathbf{w}) = \sum_{i,j} |c_{i,j} \mathbf{w}_{i,j}|_{M^{-1}}^2 \Delta x \Delta y.$$

Altogether, the above expressions are all straightforward discretizations of their contiuous counterparts. One can check that their derivatives yield finite-difference discrete differential operators. The result is an unconditionally energy stable discretization of a variety of PDEs.