Check for
updates

Primal Dual Methods for Wasserstein Gradient Flows

José A. Carrillo[1] · Katy Craig[2] · Li Wang[3] · Chaozhen Wei[4]

Abstract

Combining the classical theory of optimal transport with modern operator splitting techniques, we develop a new numerical method for nonlinear, nonlocal partial differential equations, arising in models of porous media, materials science, and biological swarming. Our method proceeds as follows: first, we discretize in time, either via the classical JKO scheme or via a novel Crank–Nicolson-type method we introduce. Next, we use the Benamou–Brenier dynamical characterization of the Wasserstein distance to reduce computing the solution of the discrete time equations to solving fully discrete minimization problems, with strictly convex objective functions and linear constraints. Third, we compute the minimizers by applying a recently introduced, provably convergent primal dual splitting scheme for three operators (Yan in J Sci Comput 1–20, 2018). By leveraging the PDEs' underlying variational structure, our method overcomes stability issues present in previous numerical work built on explicit time discretizations, which suffer due to the equations' strong nonlinearities and degeneracies. Our method is also naturally positivity and mass preserving and, in the case of the JKO scheme, energy decreasing. We prove that minimizers of the fully discrete problem converge to minimizers of the spatially continuous, discrete time problem as the spatial discretization is refined. We conclude with simulations of nonlinear PDEs and Wasserstein geodesics in one and two dimensions that illustrate the key properties of our approach, including higher-order convergence our novel Crank–Nicolson-type method, when compared to the classical JKO method.

Keywords Optimal transport · Optimization schemes · Steepest descent schemes · Gradient flows · Minimizing movements · Primal dual methods

Mathematics Subject Classification 35A15 · 47J25 · 47J35 · 49M29 · 65K10 · 82B21

🌀 Springer

1 Introduction

Gradient flow methods are classical techniques for the analysis and numerical simulation of partial differential equations. Historically, such methods were exclusively based on gradient flows arising from a Hilbert space structure, particularly $L^2(\mathbb{R}^d)$, but since the work of Jordan, Kinderlehrer, and Otto in the late 90's [75,93,94], interest has emerged in a range of nonlinear, nonlocal partial differential equations that are gradient flows in the *Wasserstein metric*,

$$\begin{cases} \partial_t \rho = \nabla \cdot (\rho \nabla V) + \nabla \cdot (\rho \nabla W * \rho) + \alpha \Delta \rho^m, & x \in \Omega \subseteq \mathbb{R}^d, \quad V, W : \Omega \to \mathbb{R}, \\ \rho(x, 0) = \rho_0(x), & m \geq 1, \quad \alpha \geq 0. \end{cases} \tag{1}$$

When $\Omega \neq \mathbb{R}^d$, we consider no-flux boundary conditions.

Equations of this form arise in a number of physical and biological applications, including models in granular media [12,45,46,102], material science [71], and biological swarming [6,39,77]. Furthermore, many well-known equations may be written in this way: when $V = W = 0$ and $\alpha = 1$, Eq. (1) reduces to the heat equation ($m = 1$), porous medium equation ($m > 1$), and fast diffusion equation ($m < 1$) [103]. In the presence of a drift potential V, it becomes a Fokker–Planck equation ($m = 1$) or nonlinear Fokker–Planck equation ($m > 1$), as used in models of tumor growth [96,100]. When the interaction potential W is given by a repulsive–attractive Morse or power-law potential,

$$W(x) = -C_a e^{-|x|/l_a} + C_r e^{-|x|/l_r}, \quad C_r/C_a < (l_r/l_a)^{-d}, 0 < l_r < l_a, 0 < C_a < C_r,$$

$$W(x) = \frac{|x|^a}{a} - \frac{|x|^b}{b}, \quad -d < b < a, \tag{2}$$

we recover a range of nonlocal interaction models, which are repulsive at short length scales and attractive at long length scales [4,5,34,101]. When $W = (\Delta)^{-1}$, the Newtonian potential, we have the Keller–Segel equation and its nonlinear diffusion variants [17,19,25,26,32,41,76]. Finally, as the diffusion exponent $m \to +\infty$, we recover congested aggregation and drift equations arising in models of pedestrian crowd dynamics and shape optimization problems [23,58,67,84,90,91].

In order to describe the gradient flow structure of equation (1), we begin by rewriting it as a continuity equation in $\rho(x, t)$ for a velocity field $v(x, t)$,

$$\begin{cases} \partial_t \rho = -\nabla \cdot (\rho v) := \nabla \cdot \left[\rho \nabla \left(\alpha U'_m(\rho) + V + W * \rho \right) \right], \\ \rho(x, 0) = \rho_0(x), \end{cases}$$

$$U_m(s) = \begin{cases} s \ln(s) & \text{for } m = 1, \\ \frac{s^m}{m-1} & \text{for } m > 1. \end{cases} \tag{3}$$

In this form, two key properties of the equation become evident: it is positivity preserving and conserves mass. In what follows, we will always consider nonnegative initial data, and we will typically renormalize so that the mass of the initial data equals one,

i.e., $\rho_0 \in \mathcal{P}_{ac}(\Omega)$, where $\mathcal{P}_{ac}(\Omega)$ is the set of probability measures on Ω that are absolutely continuous with respect to Lebesgue measure. Furthermore, as our objective is to develop a numerical method for these equations, we will exclusively consider the case when Ω is a bounded domain. Throughout, we commit a mild abuse of notation and identify all such probability measures with their densities, $d\rho(x) = \rho(x)dx$.

As discovered by Otto [93], given an energy $\mathcal{E} : \mathcal{P}_{ac}(\Omega) \to \mathbb{R} \cup \{+\infty\}$, we may formally define its gradient with respect to the Wasserstein metric $d_{\mathcal{W}}$ using the formula

$$\nabla_{d_{\mathcal{W}}}\mathcal{E}(\rho) = -\nabla \cdot \left(\rho \nabla \frac{\delta\mathcal{E}}{\delta\rho}\right) .$$

(See Sect. 2.1 for the definition of the Wasserstein metric $d_{\mathcal{W}}$.) In this way, gradient flows of $\mathcal{E}$, $\partial_t\rho = -\nabla_{d_{\mathcal{W}}}\mathcal{E}(\rho)$, correspond to solutions of the continuity equation with velocity $v = -\nabla\frac{\delta\mathcal{E}}{\delta\rho}$. In particular, Eq. (3) is the gradient flow of the energy

$$\mathcal{E}(\rho) = \int_\Omega [\alpha U(\rho(x)) + V(x)\rho(x)]\,dx + \frac{1}{2}\int_{\Omega\times\Omega} W(x-y)\rho(x)\rho(y)dxdy . \quad (4)$$

Differentiating the energy (4) along solutions of (3), one formally obtains that the energy is decreasing along the gradient flow

$$\frac{d}{dt}\mathcal{E}(\rho)(t) = -\int_{\mathbb{R}^d} |v(t,x)|^2 \rho(t,x)dx , \quad (5)$$

which coincides with the theoretical interpretation of gradient flows as solutions that evolve in the direction of *steepest descent* of an energy, where the notion of steepest descent is induced by the Wasserstein metric structure.

A key feature of equations of the form (3) is the competition between repulsive and attractive effects. For repulsive–attractive interaction kernels W, as in equation (2), these effects can arise purely through nonlocal interactions, leading to rich structure of the steady states [4,13,14,34,65]. For purely attractive interaction kernels W, as in the Keller–Segel equation, the competition instead arises from the combination of nonlocal interaction with diffusion. In this case, different choices of interaction kernel W, diffusion exponent m, and initial data ρ_0 can lead to widely different behavior—from bounded solutions being globally well posed to smooth solutions blowing up in finite time [17,19,25,26,32,41].

1.1 Summary of Numerical Approach

The goal of the present work is to develop new numerical approach for partial differential equations of the form (1) that combine gradient flow methods with modern operator splitting techniques. Our approach applies to equations of this form with any combination of diffusion $\alpha U'_m(\rho)$ ($\alpha \geq 0$), drift V, or interaction $W * \rho$ terms—in particular, it is not necessary for diffusion to be present in order for our scheme to converge.
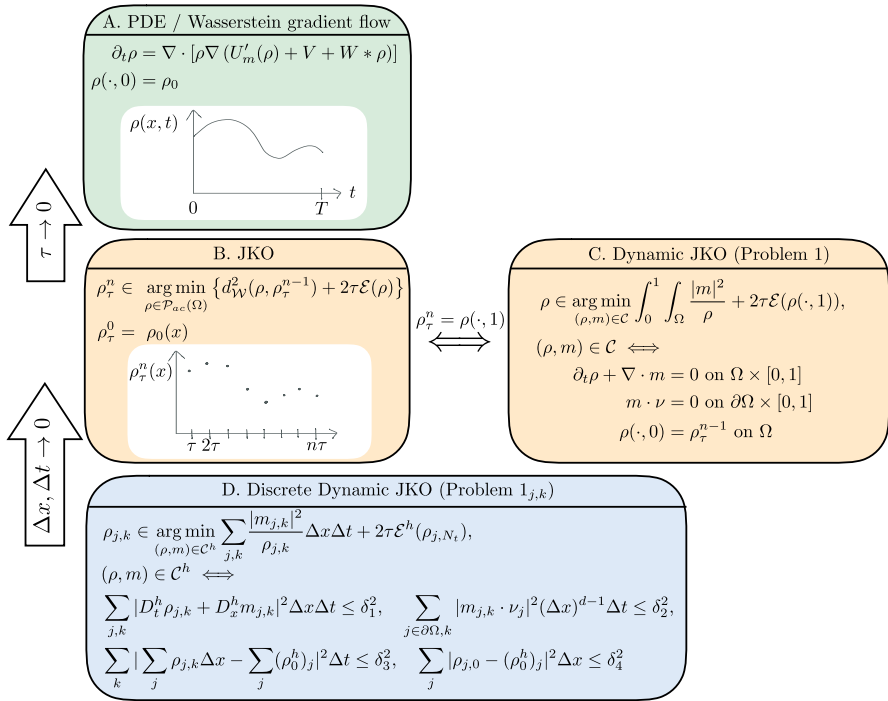
Fig. 1 Levels of discretization: τ is the outer JKO time step, Δt is the inner time step, and Δx is the spatial discretization

The main idea of our approach is to discretize the PDE/Wasserstein gradient flow at two levels. First, we consider a time discretization of the gradient flow with time step τ (see Fig. 1b), either given by the classical JKO scheme (Eq. (6) below) or a new Crank–Nicolson inspired variant (Eq. (7) below). This reduces computation of the gradient flow to solving a sequence of infinite-dimensional minimization problems. Then, we consider a dynamical reformulation of these minimization problems, stemming from Benamou and Brenier's dynamic characterization of the Wasserstein metric, by which the problem becomes the minimization of a strictly convex integral functional subject to a linear PDE constraint (see Fig. 1c). At this level, the problem remains continuous in space and time. We conclude by considering a further discretization of the problem, with inner time step (Δt) and spatial discretization (Δx), by taking piecewise constant approximations of the functions and using a finite difference approximation of the PDE constraint (see Fig. 1d). In this final, fully discrete form, we then compute the minimizer using modern operator splitting techniques, applying Yan's recent extension of the classical primal dual algorithm for minimizing sums of three convex functions [106].

Our paper is organized as follows. In Sect. 1.2, we discuss the relationship between our numerical approach and previous work. In Sect. 1.3, we summarize our contribution. In Sect. 2, we describe the details of our numerical method. Along with numerically simulating Wasserstein gradient flows, our method also provides, as a

special case, a new method for computing Wasserstein geodesics and the Wasserstein distance between probability densities; see Remark 1. In Sect. 3, we prove that, provided a smooth, positive solution of the continuum JKO scheme exists and the energy corresponding to the PDE is sufficiently regular, then minimizers of the fully discrete problem exist (Theorem 1), the objective functions of the discrete problems Γ-converge to the objective function of the continuum problem (Theorem 2), and thus, solutions of the fully discrete scheme converge, up to a subsequence, to a solution of the continuum scheme (Theorem 3). As a special case, we also recover convergence of a numerical method for computing Wasserstein geodesics, similar to that introduced by Papadakis, Péyre, and Oudet [95]. Finally, in Sect. 4, we provide several numerical simulations illustrating our approach in both one and two dimensions, computing Wasserstein geodesics, nonlinear Fokker–Planck equations, aggregation diffusion equations, and other related equations.

1.2 Details of Approach and Comparison with Previous Work

1.2.1 Classical Numerical PDE Methods

We now compare our approach to existing numerical methods. Perhaps the most common numerical approach for equations of the form (1) is to consider the equation as an advection–diffusion equation and apply classical finite difference, finite volume, or Galerkin discretizations [3,29,54,66,85]. However, when such methods are based on explicit time discretizations, they suffer from stability constraints due either to the degeneracy of the diffusion (when $m > 1$) or the nonlocality from the interaction potential W. (See for instance the mesa problem [83].) Implicit time discretizations, on the other hand, are computationally intensive, due to the difficulty of matrix inversion, even when the implicit steps are solved by smart iterative methods to avoid the high computation cost of convolution [3].

Another common approach is to leverage structural similarities between (3) and equations from fluid dynamics to develop particle methods [14,27,30,36,43,48,57, 60,88,92]. Until recently, the key limitation of such methods has been developing approaches to incorporate diffusion. Following the analogy with the Navier–Stokes equations, stochastic particle methods have been proposed in the case of linear diffusion ($m = 1$) [72–74,86]. More recently the first two authors and Patacchini developed a deterministic blob method for linear and nonlinear diffusion ($m \geq 1$) [31]. On the one hand, particle methods naturally conserve mass and positivity, and they can also be designed to respect the underlying gradient flow structure of the equation, including the energy dissipation property (5). On the other hand, a large number of particles are often required to resolve finer properties of solutions.

In contrast with such classical methods, our method introduces an auxiliary momentum variable m and an additional inner layer of time discretization, which enlarges the dimension of the problem. However, as later pointed out in [80], the inner layer of time can be discretized with just one step without violating the overall first-order accuracy, there completely eliminating the additional cost introduced by the inner layer. Another major advantage of our approach is that, by reforming the PDE problem into

an optimization problem, we obtain unconditional stability (for the JKO discretization, see Eq. (6) below) while avoiding the inversion of a full matrix in the general implicit setting, which is extremely expensive, especially in higher dimensions; see for instance [3]. Finally, compared to other implicit methods, such as the backward Euler method, the suboptimization problems can be solved independently at each gridpoint, and therefore are massively parallelizable and suitable for high-dimensional problems.

1.2.2 Variational Methods

Compared to the classical numerical PDE approaches described in the previous section, a more modern class of numerical methods leverages the gradient flow structure of (1) to approximate solutions of the PDE by solving a sequence of minimization problems. This is the approach we take in the present work. Originally introduced by Jordan, Kinderlehrer, and Otto as a technique for computing solutions of the Fokker–Planck equation (Eq. (1), $W = 0$, $m = 1$) [75], this scheme approximates the solution $\rho(x, t)$ at time t by solving the following sequence of n minimization problems with time step $\tau = t/n$,

$$\rho_\tau^n \in \underset{\rho \in \mathcal{P}_{ac}(\Omega)}{\arg\min} \left\{ d_{\mathcal{W}}^2(\rho, \rho_\tau^{n-1}) + 2\tau \mathcal{E}(\rho) \right\}, \quad \rho_\tau^0 = \rho_0(x). \tag{6}$$

The JKO scheme is precisely the analogue of the implicit Euler method in the infinite-dimensional Wasserstein space. The constraint $\rho \in \mathcal{P}_{ac}(\Omega)$ ensures that the method is positivity and mass preserving, and the fact that $d_{\mathcal{W}}^2(\rho, \rho^n) \geq 0$ ensures the energy decreasing along the scheme, $\mathcal{E}(\rho_\tau^{n+1}) \leq \mathcal{E}(\rho_\tau^n)$.

Under sufficient assumptions on the underlying domain Ω, drift potential V, interaction potential W, and initial data ρ_0 (see Sect. 2.1), the solution of the JKO scheme ρ_τ^n converges to the solution $\rho(x, t)$ of the partial differential equation (1), with a first-order rate in terms of the time step $\tau = t/n$ [2, Theorem 4.0.4],

$$d_{\mathcal{W}}(\rho_n(\cdot), \rho(\cdot, t)) \leq C\tau.$$

In our numerical simulations, we observe that this discretization error dominates other errors in our numerical method; see Sects. 4.2.1 and 4.2.2. Consequently, we also introduce a new time discretization, in analogy with the Crank–Nicolson method

$$\rho^{n+1} \in \underset{\rho \in \mathcal{P}_{ac}(\Omega)}{\arg\min} \left\{ d_{\mathcal{W}}^2(\rho, \rho^n) + \tau \mathcal{E}(\rho) + \tau \int_\Omega \frac{\delta \mathcal{E}}{\delta \rho}(\rho^n)\rho \right\}. \tag{7}$$

The connection between the above scheme and the classical Crank–Nicolson discretization can be seen by considering the optimality conditions for (7):

$$\frac{1}{\tau}(\rho^{n+1} - \rho^n) = \frac{1}{2} \nabla \cdot \left(\rho^{n+1} \nabla \left(\frac{\delta \mathcal{E}(\rho^{n+1})}{\delta \rho} + \frac{\delta \mathcal{E}(\rho^n)}{\delta \rho} \right) \right).$$

Like the JKO scheme, our Crank–Nicolson inspired method is also positivity and mass-preserving, though it is not energy decreasing. In Figs. 7, 8, and 10 of our numerics section, we conduct a preliminary analysis of the rate of convergence of this method, which verifies that it is indeed higher order than the JKO scheme. As the goal of the present work is primarily the development of fully discrete numerical schemes, we leave a thorough analysis of the rate of convergence of our Crank–Nicolson inspired method as $\tau \to 0$ to future work.

On the one hand, our Crank–Nicolson inspired method (7) is not the first higher-order method proposed for metric space gradient flows: Matthes and Plazotta developed a provably second-order scheme for general metric space gradient flows by generalizing the backward differentiation formula [89]. The Matthes–Plazotta method, however, requires two evaluations of the Wasserstein distance at each outer time step and thus is less practical for our purpose of numerically computing gradient flows in higher dimensions. Another method was introduced by Legendre and Turinici [79] based on the midpoint method. This method can be reformulated as the classical JKO step with half time step followed by an extrapolation. This extrapolation step could be implemented by solving the corresponding continuity equation either explicitly or implicitly; however, solving the equation explicitly could potentially violate conservation of positivity, while solving it implicitly would require an additional matrix inversion. Another higher-order variational method was also proposed in [78], which resembles explicit Runge–Kutta methods and, again, require two or more evaluations of the Wasserstein distance at each outer time step.

1.2.3 Numerical Methods for the Wasserstein Distance

To use either the classical JKO scheme (6) or our new Crank–Nicolson inspired scheme (7) as a basis for numerical simulations, one must first develop a fully discrete approximation of the minimization problem at each step of the scheme. Here, the main numerical difficulty arises in approximating the Wassserstein distance, and there are several different approaches for dealing with this term. First, one can reformulate the Wasserstein distance in terms of a Monge–Ampére equation with nonstandard boundary conditions [11,68], though difficulties arise due to the lack of a comparison principle [70]. Second, one can reframe the problem as a classical $L^2(\mathbb{R}^d)$ gradient flow at the level of diffeomorphisms [16,37,47,49,69], but to pursue this approach, one has to overcome complications arising from the underlying geometry and the structure of the PDE system for the diffeomorphisms. Third, one can discretize the Wasserstein distance term as a finite-dimensional linear program, overcoming the lack of strict convexity of the objective function by adding a small amount of entropic regularization [8,55,61]. (For a detailed survey of computational optimal transport, we refer the reader to the recent book by Péyre and Cuturi for [97].)

A fourth approach for computing the Wasserstein distance, and the one which we develop in the present work, is to consider a dynamic formulation due to Benamou and Brenier [7]. This reframes the problem as a strictly convex optimization problem with linear PDE constraints, which can be discretized using Benamou and Brenier's original augmented Lagrangian method ALG2 or, more generally, a range of modern proximal splitting methods, as shown by Papadakis, Peyre, and Oudet [95]. (See also [21,22]

for related work on mean field games.) Adding an additional Fisher information term in this dynamic formulation (in analogy with entropic regularization) has also been explored in [82].

Only recently have these above approaches for computing the Wasserstein distance been integrated with the JKO scheme (6) in order to simulate partial differential equations of the form (1). The Monge–Ampére approach extends naturally, though the presence of a diffusion term $\alpha U'_m(\rho)$ for $\alpha > 0$ is required to enforce convexity constraints at the discrete level [10]. Similarly, entropic regularization (or the addition of a Fisher information term) vastly accelerates the computation of gradient flows, but at the level of the partial differential equation, this corresponds to introducing numerical diffusion, which may disrupt the delicate balance between aggregation and diffusion inherent in PDEs of this type [28,55,82]. Finally, Benamou and Brenier's dynamic reformulation of the Wasserstein distance has also been adopted in recent work to approximate gradient flows [9]. A key benefit of this latter approach when compared to entropic regularization is that it leads to an optimization problem in $N_x^d \times N_t$ variables, where N_x and N_t are the number of spatial and temporal gridpoints, whereas the latter leads to an optimization problem in N_x^{2d} variables.

In the present work, we further develop this last approach, using Benamou and Brenier's dynamic reformulation of the Wasserstein distance to simulate Wasserstein gradient flows, via both the classical JKO scheme (6) and our new Crank–Nicolson inspired scheme (7). This leads to a sequence of minimization problems (Fig. 1C), which we discretize (Fig. 1D) and then solve using a modern primal dual three operator splitting scheme due to Yan [106], instead of the classical ALG2 method. See Sect. 2 for a detailed description of our approach.

Due to the fact that we use operator splitting methods to compute the minimizer in Benamou and Brenier's dynamic formulation of the Wasserstein distance, our work can be seen as an extension of previous work by Papadakis, Peyre, and Oudet [95], which applied similar two operator splitting schemes to simulate the Wasserstein distance. However, there are a few key differences between our approach and previous work. First, we are able to implement the primal dual splitting scheme in a manner that does not require matrix inversion of the finite difference operator, which reduces the computational cost. Second, we succeed in obtaining the exact expression for the proximal operator, which allows our method to be truly positivity preserving, while other similar methods are only positivity preserving in the limit as $\Delta x, \Delta t \to 0$; see Remark 5. Third, instead of imposing the linear PDE constraint in Benamou and Brenier's dynamic reformulation *exactly*, via a finite difference approximation, we allow the linear PDE constraint to hold up to an error of order $\delta > 0$, which can be tuned according to the spatial discretization (Δx), the inner temporal discretization (Δt), and the outer time step τ to respect the order of accuracy of the finite difference approximation; see Remark 3. Numerically, this allows our method to converge in fewer iterations, without any reduction in accuracy, as demonstrated in Fig. 3. From a theoretical perspective, the fact that we only require the PDE constraint to hold up to an error of order $\delta > 0$ makes it *possible* to prove convergence of minimizers of the fully discrete problem to minimizers of the JKO scheme (6), since minimizers of the fully discrete problem always exist for $\delta > 0$, which is not the case when the PDE constraint is enforced exactly ($\delta = 0$); see Remark 8 and Theorem 1.

1.3 Contribution

The main components of our numerical method for computing solutions to (1) are:

(a) an outer time discretization, of either JKO (6) or Crank–Nicolson type (7) (Fig. 1B)
(b) a dynamic interpretation of the Wasserstein distance (Fig. 1C), which when discretized via finite difference approximations leads to a sequence of constrained optimization problems (Fig. 1D)
(c) an application of modern three operator splitting schemes for solving these optimization problems.

Our main contributions are:

– Unlike classical explicit methods, our JKO-type method is unconditionally stable. Unlike classical implicit methods, it achieves this stability without an expensive matrix inversion.
– In practice, we observe that our Crank–Nicolson-type method performs even better than our JKO-type method, in terms of rate of convergence with respect to the outer time step (see Figs. 7, 8, and 10). We leave a thorough analysis of the rate of this convergence of this method to future work.
– By formulating our optimization problem with a linear *inequality* constraint instead of a linear *equality* constraint, our algorithm converges in fewer iterations when compared to related algorithms for Wasserstein geodesics; see Remark 3 and Fig. 3.
– We prove convergence of our fully discrete method (Fig. 1D) to the JKO scheme (Fig. 1B, C) as the spatial discretization and inner time discretization go to zero.

2 Numerical Method

2.1 Dynamic Formulation of JKO Scheme

As described in the previous section, our numerical method for computing the JKO scheme is based on the following dynamic reformulation of the Wasserstein distance due to Benamou and Brenier [7]:

$$
d_{\mathcal{W}}(\rho_0, \rho_1) = \inf_{(\rho, v) \in \mathcal{C}_0} \left\{ \int_0^1 \int_\Omega |v(x, t)|^2 \mathrm{d}\rho(x, t) \mathrm{d}t \right\}^{1/2}, \tag{8}
$$

where $(\rho, v) \in AC(0, 1; \mathcal{P}(\Omega)) \times L^1(0, 1; L^2(\rho))$ belongs to the constraint set $\mathcal{C}_0$ provided that

$$
\partial_t \rho + \nabla \cdot (\rho v) = 0 \text{ on } \Omega \times [0, 1] \tag{9}
$$

$$
(\rho v) \cdot \nu = 0 \text{ on } \partial\Omega \times [0, 1], \tag{10}
$$

$$
\rho(\cdot, 0) = \rho_0, \ \rho(\cdot, 1) = \rho_1 \text{ on } \Omega, \tag{11}
$$

where ν is the outer unit normal on the boundary of the domain Ω. A curve ρ in $\mathcal{P}(\Omega)$ is *absolutely continuous* in time, denoted $\rho \in AC(0, 1; \mathcal{P}(\Omega))$, if there exists

$w \in L^1(0, 1)$ so that $d_{\mathcal{W}}(\rho(\cdot, t_0), \rho(\cdot, t_1)) \leq \int_{t_0}^{t_1} w(s)\mathrm{d}s$ for all $0 < t_0 \leq t_1 < 1$. The PDE constraint (9 and 10) holds in the duality with smooth test functions on $\mathbb{R}^d \times [0, 1]$, i.e., for all $f \in C_c^\infty(\mathbb{R}^d \times [0, 1])$,

$$\int_0^1 \int_\Omega [\partial_t f(x, t)\rho(x, t) + \nabla f(x, t) \cdot v(x, t)\rho(x, t)]\,\mathrm{d}x\mathrm{d}t$$

$$+ \int_\Omega f(x, 0)\rho_0(x) - f(x, 1)\rho_1(x)\mathrm{d}x = 0.$$

This dynamic reformulation reduces the problem of finding the Wasserstein distance between any two measures to identifying the curve in $\mathcal{P}(\Omega)$ that connects them with minimal kinetic energy. However, the objective function (8) is not strictly convex, and the PDE constraint (9) is nonlinear. For these reasons, in Benamou and Brenier's original work, they restrict their attention to the case $\rho(\cdot, t) \in \mathcal{P}_{ac}(\Omega)$ and introduce the momentum variables $m = v\rho$, in order to rewrite (8) as

$$d_{\mathcal{W}}^2(\rho_0, \rho_1) = \min_{(\rho, m) \in \mathcal{C}_1} \int_0^1 \int_\Omega \Phi(\rho(x, t), m(x, t))\mathrm{d}x\mathrm{d}t, \tag{12}$$

where

$$\Phi(\rho, m) = \begin{cases} \frac{\|m\|^2}{\rho} & \text{if } \rho > 0, \\ 0 & \text{if } (\rho, m) = (0, 0), \\ +\infty & \text{otherwise}, \end{cases} \tag{13}$$

and $(\rho, m) \in AC(0, 1; \mathcal{P}_{ac}(\Omega)) \times L^1(0, 1; L^2(\rho^{-1}))$ belong to the constraint set $\mathcal{C}_1$ provided that

$$\partial_t \rho + \nabla \cdot m = 0 \text{ on } \Omega \times [0, 1]$$
$$m \cdot \nu = 0 \text{ on } \partial\Omega \times [0, 1].$$
$$\rho(\cdot, 0) = \rho_0, \ \rho(\cdot, 1) = \rho_1 \text{ on } \Omega.$$

After this reformulation, the integral functional

$$(\rho, m) \mapsto \int_0^1 \int_\Omega \Phi(\rho, m) \tag{14}$$

is strictly convex along linear interpolations and lower semicontinuous with respect to weak-* convergence [1, Example 2.36], and the PDE constraint is linear. As an immediate consequence, one can conclude that minimizers are unique. Furthermore, for any $\rho_0, \rho_1 \in \mathcal{P}_{ac}(\Omega)$, a direct computation shows that the minimizer $(\bar{\rho}, \bar{m})$ is given by the Wasserstein geodesic from ρ_0 to ρ_1,

$$\bar{\rho}(x, t) = T_t \# \rho_0, \ \bar{v}(x, t) = T \circ T_t^{-1}(x) - T_t^{-1}(x), \ \bar{m}$$
$$= \bar{v}\bar{\rho}, \ \text{for } T_t(x) := (1 - t)x + tT(x), \tag{15}$$

where T is the optimal transport map from ρ_0 to ρ_1. (See [2,98,105] for further background on optimal transport.) Consequently, given any minimizer $(\bar{\rho}, \bar{m})$ of (12), we can recover the optimal transport plan T via the following formula:

$$T(x) = x + \bar{v}(x, 0) = x + \bar{m}(x, 0)/\bar{\rho}(x, 0). \tag{16}$$

Building upon Benamou and Brenier's dynamic reformulation of the Wasserstein distance, one can also consider a dynamic reformulation of the JKO scheme (6). In particular, substituting (12) in (6) leads to the following dynamic JKO scheme:

Problem 1 *(Dynamic JKO) Given $\tau > 0, \mathcal{E}$, and ρ_0, solve the constrained optimization problem,*

$$\inf_{(\rho, m) \in \mathcal{C}} \int_0^1 \int_\Omega \Phi(\rho(x, t), m(x, t)) \mathrm{d}x \mathrm{d}t + 2\tau \mathcal{E}(\rho(\cdot, 1)),$$

where $(\rho, m) \in AC(0, 1; \mathcal{P}_{ac}(\Omega)) \times L^1(0, 1; L^2(\rho^{-1}))$ belong to the constraint set $\mathcal{C}$ provided that

$$\partial_t \rho + \nabla \cdot m = 0 \text{ on } \Omega \times [0, 1], \quad m \cdot \nu = 0 \text{ on } \partial\Omega \times [0, 1], \quad \text{and } \rho(\cdot, 0) = \rho_0 \text{ on } \Omega. \tag{17}$$

We emphasize that the requirement $\rho(x, t) \in \mathcal{P}_{ac}(\Omega)$ for all $t \in [0, 1]$ ensures that $\rho(x, t) \geq 0$.

Remark 1 (Wasserstein geodesics) Note that for any $\rho_1 \in \mathcal{P}_{ac}(\Omega)$, we may take

$$\mathcal{E}(\rho(\cdot, 1)) = \mathcal{G}_{\rho_1}(\rho(\cdot, 1)) := \begin{cases} 0 & \text{if } \rho(\cdot, 1) = \rho_1, \\ +\infty & \text{otherwise,} \end{cases} \tag{18}$$

in which case Problem 1 reduces to the Benamou–Brenier formulation of the Wasserstein distance (12). Consequently, the numerical method we develop for Problem 1 offers, as a particular case, a provably convergent numerical method for computing the Wasserstein geodesic and Wasserstein distance between ρ_0 and ρ_1. On the one hand, there are many alternative methods for computing Wasserstein geodesics in Euclidean space. Indeed, the many algorithms described in the introduction for computing the Wasserstein distance also provide an optimal transport plan, which can be linearly interpolated to give the Wasserstein geodesic [8,11,55,61,68,97]. On the other hand, our method is distinguished because it could be more naturally extended to variants of the Wasserstein metric built on the Benamou–Brenier formulation [33,64,87], as well as to Wasserstein geodesics on non-Euclidean manifolds, where the geodesic equations on the underlying manifold may no longer be explicit, so that one cannot pass directly from the optimal transport plan to the Wasserstein geodesic.

Remark 2 (existence and uniqueness of minimizers) If the underling domain Ω is convex and the energy $\mathcal{E}$ is proper, lower semicontinuous, coercive, and λ-convex along generalized geodesics, and also satisfies $\{\mu : \mathcal{E}(\mu) < +\infty\} \subseteq \mathcal{P}_{ac}(\Omega)$, then,

for $\tau > 0$ sufficiently small, there exists a unique solution to Problem 1 [2, Theorem 4.0.4, Theorem 8.3.1]. In particular, these assumptions are satisfied by the energy $\mathcal{G}_{\rho_1}$ (18), as well as by the drift–diffusion interaction energy from the introduction (4), for U as in Eq. (3), V, $W \in C^2(\Omega)$. (See, for example, [2, Section 9.3] or [56] for more general conditions on U, V, W.)

Thus, if we denote by $(\bar{\rho}, \bar{m})$ the minimizer of Problem 1, then for $\tau > 0$ sufficiently small, the *proximal map*,

$$J_\tau(\rho_0) := \overline{\rho}(\cdot, 1) \,,$$

is well defined for all $\rho_0 \in D(\mathcal{E})$. Furthermore, the energy decreases under the proximal map,

$$\mathcal{E}(J_\tau(\rho_0)) \leq \mathcal{E}(\rho_0), \tag{19}$$

which can be seen by comparing the value of the objective function at the minimizer $(\overline{\rho}, \overline{m})$ to the value of the objective function at $(\rho(x, 0), 0) \in \mathcal{C}$ and using that $\Phi(\rho, m) \geq 0$.

Given $\rho_0 \in D(\mathcal{E})$, if we recursively define the *discrete time gradient flow sequence*

$$\rho_\tau^n = J_\tau(\rho_\tau^{n-1}), \quad \text{for all } n \in \mathbb{N}, \tag{20}$$

then, taking $\tau = t/n$, ρ_τ^n converges to $\rho(x, t)$, the gradient flow of the energy $\mathcal{E}$ with initial data ρ_0 at time t, and under mild regularity assumptions on ρ_0, we have

$$d_\mathcal{W}(\rho_n(\cdot), \rho(\cdot, t)) \leq C\tau. \tag{21}$$

In this way, the classical JKO scheme provides a *first-order* approximation of the gradient flow [2, Theorem 4.0.4]. In our numerical simulations, we observe that this discretization error dominates other errors in our numerical method; see Sects. 4.2.1 and 4.2.2. For this reason, we introduce the following new scheme, inspired by the Crank–Nicolson method.

Problem 2 *(Crank–Nicolson Inspired Dynamic JKO) Given $\tau > 0$, $\mathcal{E}$, and ρ_0, solve the constrained optimization problem,*

$$\inf_{(\rho, m) \in \mathcal{C}} \int_0^1 \int_\Omega \Phi(\rho(x, t), m(x, t)) \mathrm{d}x \mathrm{d}t + \tau \mathcal{E}(\rho(x, 1)) + \tau \int_\Omega \frac{\delta \mathcal{E}}{\delta \rho}(\rho(x, 0))\rho(x, 1)\mathrm{d}x,$$

where $(\rho, m) \in AC(0, 1; \mathcal{P}_{ac}(\Omega)) \times L^1(0, 1; L^2(\rho^{-1}))$ belong to the constraint set $\mathcal{C}$ provided that

$$\partial_t \rho + \nabla \cdot m = 0 \text{ on } \Omega \times [0, 1], \quad m \cdot \nu = 0 \text{ on } \partial\Omega \times [0, 1], \quad \text{and } \rho(\cdot, 0) = \rho_0 \text{ on } \Omega.$$

In Sect. 4.2.2, we provide numerical examples comparing the above method to the classical JKO scheme from Problem 1, illustrating that it achieves a higher-order rate

of convergence in practice (see Figs. 7, 8, and 10), in spite of the fact that that it lacks the energy decay property of Problem 1. Under what conditions a higher-order analogue of inequality (21) holds for the new scheme is an interesting open question that we leave to future work, as the main goal of the present work is the development of fully discrete numerical methods for computing minimizers of Problem 1 and 2. By iterating either of these minimization problems, as in Eq. (20), we obtain a numerical method for simulating Wasserstein gradient flows.

2.2 Fully Discrete JKO

We now turn to the discretization of the dynamic JKO scheme, Problem 1, and the Crank–Nicolson inspired scheme, Problem 2. We begin by noting that the Crank–Nicolson inspired Problem 2 can be rewritten in the same form as Problem 1 by considering the energy

$$\mathcal{H}_{\rho_0}(\rho) := \frac{1}{2}\mathcal{E}(\rho(x,1)) + \frac{1}{2}\int_\Omega \frac{\delta\mathcal{E}}{\delta\rho}(\rho(x,0))\rho(x,1)\mathrm{d}x. \tag{22}$$

Using this observation, we will now describe our discretization of both problems simultaneously.

2.2.1 Discretization of Functions and Domain

Given an n-dimensional hyperrectangle $S = \Pi_{I=1}^n[a_I, b_I] \subseteq \mathbb{R}^n$, we discretize it as a union of cubes Q_i, $i \in \mathbb{N}^n$, where in the lth direction, we suppose there are N_l intervals of spacing $(\mathbf{z})_l = (b_l - a_l)/N_l$:

$$S = \bigcup_{i:Q_i \subseteq S} Q_i, \quad Q_i := \{(z_1, z_2, \ldots, z_n) \in \mathbb{R}^n : z_l \in [(i_l - 1)(\mathbf{z})_l, i_l(\mathbf{z})_l] \,\forall l = 1, \ldots, n\}.$$

Piecewise constant functions with respect to this discretization are given by

$$f^h := \sum_{i:Q_i \subseteq S} f_i 1_{Q_i}, \text{ for } f_i \in \mathbb{R} \text{ and } 1_{Q_i}(z) = \begin{cases} 1 \text{ if } z \in Q_i \\ 0 \text{ otherwise.} \end{cases}$$

To discretize Problem 1, we take $S = \overline{\Omega} \times [0,1] \subseteq \mathbb{R}^{d+1}$, where $\overline{\Omega} = \Pi_{i=1}^n[a_i, b_i]$. For any $i \in \mathbb{N}^{d+1}$, write $i = (j,k)$, for the spatial index $j \in \mathbb{N}^d$ and the temporal index $k \in \mathbb{N}$. We let $N_x \in \mathbb{N}$ denote the number of intervals in each spatial direction and $N_t \in \mathbb{N}$ denote the number of intervals in the temporal direction. Take $\mathbf{z} = (\mathbf{x}, \Delta t)$ for $(\mathbf{x})_l = (\Delta x) > 0$ for all $l = 1, \ldots, d$ and $\Delta t > 0$.

We consider piecewise constant approximations (ρ^h, m^h) of the functions (ρ, m), with coefficients denoted by $(\rho_{j,k}, m_{j,k})$. For any $(\rho, m) \in C(\overline{\Omega} \times [0,1])$, one such approximation is the *pointwise piecewise* approximation $(\hat{\rho}^h, \hat{m}^h)$, obtained by defining the coefficients $(\hat{\rho}_{j,k}, \hat{m}_{j,k})$ to be the value of (ρ, m) on a regular grid of spacing $(\Delta x) \times (\Delta t)$:

$$\hat{\rho}_{j,k} := \rho(x_j, t_k), \quad \hat{m}_{j,k} := m(x_j, t_k),$$

$$(x_j, t_k) = (\hat{\mathbf{x}} + (j - \mathbf{1})(\Delta x), \hat{t} + (k - 1)(\Delta t))$$
$$\hat{\mathbf{x}} \in \Pi_{l=1}^d [0, \Delta x], \ \hat{t} \in [0, \Delta t]. \tag{23}$$

where $\mathbf{1} = [1, 1, \ldots, 1]^t \in \mathbb{N}^d$. Note that, whenever $(\rho, m) \in C(\overline{\Omega} \times [0, 1])$, we have that $(\hat{\rho}^h, \hat{m}^h)$ converges to (ρ, m) uniformly.

2.2.2 Discretization of Energy Functionals

Next, we approximate the energy functionals by discrete energies $\mathcal{E}^h$, beginning with energies of the form (4). Given a piecewise constant function ρ^h with coefficients ρ_j,

$$\mathcal{F}^h(\rho_j) := \sum_j \left(U(\rho_j) + V_j \rho_j \right) (\Delta x)^d + \frac{1}{2} \sum_{j,l} W_{j,l} \rho_j \rho_l (\Delta x)^{2d}, \tag{24}$$

where $V^h(x) = \sum_j V_j 1_{Q_j}(x)$ is a piecewise constant approximation of $V(x)$ and $W^h(x, y) = \sum_{j,l} W_{j,l} 1_{Q_j}(x) 1_{Q_j}(y)$ is a piecewise constant approximation of $W(x - y)$. Here, $W_{j,l} = W(|x_j - x_l|)$ symmetric, i.e., $W_{j,l} = W_{l,j}$.

Likewise, for energies of the form (4), we consider the following discretization of the energy $\mathcal{H}_{\rho_0}$ from Eq. (22) for the Crank–Nicolson inspired scheme, Problem 2,

$$\mathcal{H}_{\rho_0}^h(\rho_j) := \frac{1}{2} \mathcal{F}^h(\rho_j) + \frac{1}{2} \sum_j \left(U'((\rho_0)_j) + V_j + \sum_l W_{j,l} (\rho_0)_l (\Delta x)^d \right) \rho_j (\Delta x)^d. \tag{25}$$

Finally, to compute Wasserstein geodesics between two measures $\rho_0, \rho_1 \in \mathcal{P}_{ac}(\Omega)$, we consider a discretization of the energy $\mathcal{G}_{\rho_1}$ from Eq. (18). Given a piecewise constant approximation ρ_1^h of ρ_1 and $\delta \geq 0$, define

$$\mathcal{G}_{\rho_1}^h(\rho_j) := \begin{cases} 0 & \text{if } \sum_j |\rho_j - (\rho_1^h)_j|^2 (\Delta x)^d \leq \delta^2 \\ +\infty, & \text{otherwise.} \end{cases} \tag{26}$$

2.2.3 Discretization of Derivative Operators

Let $D_t^h \rho^h$ and $D_x^h m^h$ denote the discrete time derivative and spatial divergence on $\Omega \times [0, 1]$ and let ν^h denote the discrete outer unit normal of Ω. (See Hypothesis 3 for the precise requirements we impose on each of these discretizations). For example, in one dimension we may choose a centered difference in space and a forward Euler method in time,

$$D_t^h \rho_{j,k} = \frac{\rho_{j,k+1} - \rho_{j,k}}{\Delta t}, \quad D_x^h m_{j,k} = \frac{m_{j+1,k} - m_{j-1,k}}{2\Delta x}. \tag{27}$$

or a Crank–Nicolson method,

$$D_t^h \rho_{j,k} = \frac{\rho_{j,k+1} - \rho_{j,k}}{\Delta t}, \quad D_x^h m_{j,k} = \frac{m_{j+1,k} - m_{j-1,k} + m_{j+1,k+1} - m_{j-1,k+1}}{4\Delta x}.$$
(28)

We compute these discretizations of the derivatives at the boundary by extending $m_{j,k}$ to be zero in the direction of the outer unit normal vector. As we can only expect these approximations of the temporal and spatial derivatives to hold up to an error term, we relax the equality constraints from (17) in the following discrete dynamic JKO scheme.

2.2.4 Discrete Dynamic JKO

The discretizations described in the previous sections lead to a fully discrete dynamic JKO problem:

Problem $1_{j,k}$ **(Discrete Dynamic JKO)** *Fix* $\tau, \delta_1, \delta_2, \delta_3, \delta_4 > 0, \mathcal{E}^h$, *and* ρ_0^h. *Solve the constrained optimization problem,*

$$\inf_{(\rho_{j,k}, m_{j,k}) \in \mathcal{C}^h} \sum_j \sum_k \Phi(\rho_{j,k}, m_{j,k})(\Delta x)^d \Delta t + 2\tau \mathcal{E}^h(\rho_{j,N_t}),$$
(29)

where $(\rho_{j,k}, m_{j,k})$ *belong to the constraint set* $\mathcal{C}^h$ *provided that for all* j, k,

$$\sum_{j,k} |D_t^h \rho_{j,k} + D_x^h m_{j,k}|^2 (\Delta x)^d (\Delta t) \leq \delta_1^2,$$

$$\sum_{j \in \partial\Omega, k} |m_{j,k} \cdot v_j|^2 (\Delta x)^{d-1} (\Delta t) \leq \delta_2^2,$$
(30)

$$\sum_k |\sum_j \rho_{j,k}(\Delta x)^d - \sum_j (\rho_0^h)_j (\Delta x)^d|^2 (\Delta t) \leq \delta_3^2,$$

$$\sum_j |\rho_{j,0} - (\rho_0^h)_j|^2 (\Delta x)^d \leq \delta_4^2.$$
(31)

The inequalities (30) enforce the PDE constraint and the boundary condition; the inequalities (31) enforce the mass constraint and the initial conditions. Recall that, by definition of Φ in Eq. (13), $\Phi(\rho_{j,k}, m_{j,k}) < +\infty$ only if $\rho_{j,k}$ is nonnegative. Consequently, if a minimizer $\rho_{j,k}$ exists, it must be nonnegative.

Remark 3 (relaxation of PDE constraints) A key element of our numerical method is that we relax the equality constraint (17) at the fully discrete level. This reflects the fact that even an exact solution of the continuum PDE will only satisfy the discrete constraints (30-31) up to an error term depending on the order of the finite difference operators.

We allow the choice of δ_i to vary for each of the above constraints. However, when the desired exact solution is sufficiently smooth, the optimal choice of δ_i for a second-order discretization of the spatial and temporal derivatives is

$$\delta_1 \sim (\Delta x)^2 + (\Delta t)^2 \tau \quad \text{and} \quad \delta_2, \delta_3, \delta_4 \sim (\Delta x)^2,$$

where $\tau > 0$ is the size of the timestep in the outer time discretization; see equations (6-7). As we will demonstrate in Fig. 3 of our numerics section, relaxing the PDE constraint accelerates convergence to a minimizer of the fully discrete Problem $1_{j,k}$ without any loss of accuracy with respect to the exact continuum solution.

Finally, note that while the discrete PDE constraint (30) automatically enforces the mass constraint up to order $\delta_1^2 + \delta_2^2$, we choose to impose the mass constraint separately via the first Eq. in (31). This leads to better performance in examples where the exact solution is not smooth enough to satisfy the discrete PDE constraint up to a high order of accuracy but imposing a stricter mass constraint leads to a higher quality numerical solution; see Fig. 4.

Under sufficient hypotheses on the discrete energy $\mathcal{E}^h$ and the initial data ρ_0^h, minimizers of Problem $1_{j,k}$ exist; see Theorem 1. Furthermore, this discrete dynamic JKO scheme preserves the energy decreasing property of the original JKO scheme. To see this, note that, given an energy $\mathcal{E}^h$, time step $\tau > 0$, and initial data $(\rho_0^h)_j$ we may define the fully discrete proximal map by

$$J_\tau^h((\rho_0)_j) := \overline{\rho}_{j,N_t},$$

where $(\overline{\rho}_{j,k}, \overline{m}_{j,k})$ is any minimizer of Problem $1_{j,k}$. Independently of which minimizer is chosen, we have

$$\mathcal{E}^h(J_\tau^h((\rho_0)_j)) \leq \mathcal{E}^h((\rho_0)_j),$$

which can be seen by comparing the value of the objective function at the minimizer $(\overline{\rho}_{j,k}, \overline{m}_{j,k})$ to the value of the objective function at $(\rho_{j,k}, m_{j,k}) = ((\rho_0)_j, 0) \in \mathcal{C}$ and using the fact that $\Phi \geq 0$. Furthermore, by iterating the fully discrete proximal map, we may construct a *fully discrete gradient flow sequence*

$$(\rho_\tau^n)_j = J_\tau^h((\rho_\tau^{n-1})_j) \text{ for all } n \in \mathbb{N}, \quad \rho_\tau^0 = \rho_0^h.$$

In analogy with the continuum case, we will use this fully discrete JKO scheme to simulate gradient flows. (See Algorithm 3.)

2.3 Primal Dual Algorithms for Fully Discrete JKO

In order to find minimizers of Problem $1_{j,k}$, we apply a primal dual operator splitting method. Since the constraints in Problem $1_{j,k}$ are linear inequality constraints, we may rewrite them in the form $\|\tilde{A}_i u - \tilde{b}_i\|_2 \leq \delta_i$ for $i = 1, 2, 3, 4$, where $u = (\text{\ae}, \mathbf{m})$, and $\text{\ae}$ and $\mathbf{m}$ are vector representations of the matrices $\rho_{j,k}$ and $m_{j,k}$. (See the Appendix A

for explicit formulas for $\tilde{A}_i$ and $\tilde{b}_i$, in one spatial dimension). Similarly, we may rewrite the first term of the objective function (29) in terms of u, defining

$$\Phi(u) = \sum_k \sum_j \Phi(\rho_{j,k}, m_{j,k})(\Delta x)^d \Delta t.$$

We consider two cases for the energy term in the objective function. When the energy is of the form $\mathcal{G}_{\rho_1}^h$, as in Eq. (26), we reframe the problem by removing the energy from the objective function and adding $\sum_j |\rho_{j,N_t} - (\rho_1^h)_j|^2 (\Delta x)^d \leq \delta_5^2$ to the constraints (30) and (31), denoting $\|A_i u - b_i\|_2 \leq \delta_i$, for $i = 1, 2, 3, 4, 5$, as the modified constraints. On the other hand, when the energy is of the form (24) or (25), we rewrite it in terms of u as

$$F(u) = \sum_j \left(U(\rho_{j,N_t}) + V_j \rho_{j,N_t} \right)(\Delta x)^d + \frac{1}{2}\sum_{j,l} \left(W_{j,l}\rho_{j,N_t}\rho_{l,N_t} \right)(\Delta x)^{2d},$$

$$H(u) = \frac{1}{2}F(u) + \frac{1}{2}\sum_j \left(U'(\rho_{j,0}) + V_j + \sum_l W_{j,l}\rho_{l,0}(\Delta x)^d \right)\rho_{j,N_t}(\Delta x)^d.$$

$$(32)$$

In particular, if we let S be the selection matrix

$$\mathsf{S} : \mathbb{R}^N \to \mathbb{R}^{N_x} : u \mapsto \rho_{j,N_t},$$

then $F(u) = \mathcal{F}^h(\mathsf{S}u)$ and $H(u) = \mathcal{H}_{\rho_0}^h(\mathsf{S}u)$, where $\mathcal{F}^h$ and $\mathcal{H}_{\rho_0}^h$ are defined in (24) and (25), respectively.

This leads to the following two optimization problems:

Problem 3(a) $\min_u \Phi(u) + \mathrm{i}(Au)$, $\quad \mathrm{i}_\delta(Au) = \begin{cases} 0 & \|A_i u - b_i\|_2 \leq \delta_i, \ i = 1, \ldots, 5 \\ +\infty, & \textit{otherwise.} \end{cases}$

Problem 3(b) $\min_u \Phi(u) + 2\tau E(u) + \mathrm{i}_{\tilde{\delta}}(\tilde{A}u)$, $\quad \mathrm{i}_{\tilde{\delta}}(\tilde{A}u) = \begin{cases} 0 & \|\tilde{A}_i u - \tilde{b}_i\|_2 \leq \tilde{\delta}_i, \ i = 1, \ldots, 4 \\ +\infty, & \textit{otherwise.} \end{cases}$

To compute the Wasserstein distance, we solve Problem 2.3, and to compute the gradient flow of an energy, we iterate Problem 2.3 $O(\frac{1}{\tau})$ times, for either $E(u) = F(u)$ (classical JKO) or $E(u) = H(u)$ (Crank–Nicolson inspired scheme).

Primal-dual methods for solving optimization problems in which the objective function is the sum of two convex functions, as in Problem 2.3, are widely available [52]. However, analogous methods for optimizations problems in which the objective function is the sum of three convex functions, as in Problem 2.3, have only recently emerged [62,106]. In particular, in Algorithm 1, for Problem 2.3, we use Chambolle and Pock's well-known primal dual algorithm, and in Algorithm 2, for Problem 2.3, we use Yan's recent extension of this algorithm to objective functions with three convex terms. Both algorithms offer an extended range of primal and dual step sizes λ and σ and low per-iteration complexity, due to the sparseness of S, A, and $\tilde{A}$. Note specifically that the success of Algorithm 1 depends on the ease of computing the

proximal operators related to ϕ and i_δ, and therefore if we simply group the additional energy term in Problem 2.3 to either ϕ or i_δ, it would violate such property. Instead, we shall consider $E(u)$ as a separate term and take advantage of its smoothness, as shown in Algorithm 2. Finally, in Algorithm 3, we describe how Algorithm 2 can be iterated to approximate the full JKO sequence and, consequently, solutions of a range of nonlinear partial differential equations of Wasserstein gradient flow type.

Algorithm 1: Primal-Dual for Wasserstein distance

Input: $u^0, \phi^0, \text{Iter}_{max}, \lambda, \sigma > 0$
Output: $u^* = (\rho^*, m^*)$ and the Wasserstein distance $\Phi(u^*)^{1/2}$

1 Let $\bar{u}^0 = u^0$ and $l = 0$;
2 **while** $l < \text{Iter}_{max}$ **do**
3   **repeat**
4       $\phi^{(l+1)} = \text{Prox}_{\sigma i_\delta^*}(\phi^{(l)} + \sigma A\bar{u}^{(l)})$,
5       $u^{(l+1)} = \text{Prox}_{\lambda\Phi}(u^{(l)} - \lambda A^T\phi^{(l+1)})$,
6       $\bar{u}^{(l+1)} = 2u^{(l+1)} - u^{(l)}$,
7   **until** *stopping criteria are achieved*;
8 **end**
9 $u^* = u^{(l+1)}$

Algorithm 2: Primal-Dual for one step of dynamic JKO

Input: $u^0, \phi^0, \text{Iter}_{max}, \lambda, \sigma, \tau > 0$
Output: u^*, ϕ^*

1 Let $\bar{u}^0 = u^0$ and $l = 0$;
2 **while** $l < \text{Iter}_{max}$ **do**
3   **repeat**
4       $\phi^{(l+1)} = \text{Prox}_{\sigma i_\delta^*}(\phi^{(l)} + \sigma \tilde{A}\bar{u}^{(l)})$,
5       $u^{(l+1)} = \text{Prox}_{\lambda\Phi}(u^{(l)} - \lambda\nabla E(u^{(l)}) - \lambda\tilde{A}^t\phi^{(l+1)})$,
6       $\bar{u}^{(l+1)} = 2u^{(l+1)} - u^{(l)} + \lambda\nabla E(u^{(l)}) - \lambda\nabla E(u^{(l+1)})$,
7   **until** *stopping criteria is achieved*;
8    $u^* = u^{(l+1)}$
9    $\phi^* = \phi^{(l+1)}$
10 **end**

To initialize both algorithms, we choose ϕ^0 and m^0 to be zero vectors, and for ρ^0, we let its components at the initial time (i.e., $k = 0$) be $\rho_0(x)$ evaluated on an equally spaced grid of width Δx, and other times to be zero. The stopping criteria consist of checking the PDE constraint (30)–(31) along with the convergence monitors:

$$\frac{|F(u^{(l)}) - F(u^{(l-1)})|}{|F(u^{(l)})|} < \epsilon_1, \tag{33}$$

$$\max\left\{\frac{\|u^{(l)} - u^{(l-1)}\|}{\|u^{(l)}\|}, \ \frac{\|\phi^{(l)} - \phi^{(l-1)}\|}{\|\phi^{(l)}\|}\right\} < \epsilon_2. \tag{34}$$

The proximal operator, which appears in Algorithms 1 and 2, is defined by

$$\mathrm{Prox}_h(x) = \mathrm{argmin}_u\left\{\frac{1}{2}\|u - x\|^2 + h(u)\right\}.$$

For both $h = \sigma i_\delta^*$ and $h = \lambda\Phi$, there are explicit formulas for the proximal operators. By Moreau's identity, we may write $\mathrm{Prox}_{\sigma i_\delta^*}(x)$ in terms of projections onto balls of radius δ_i centered at b_i for the ith portion of vector x:

$$\mathrm{Prox}_{\sigma i^*}(x) = x - \sigma\,\mathrm{Proj}_{B_\delta}(x/\sigma)$$

$$\mathrm{Proj}_{B_\delta}(x) = \begin{cases} x_i & \|x_i - b_i\|_2 \le \delta_i \\ \delta\frac{x_i - b_i}{\|x_i - b_i\|_2} + b_i & \text{otherwise,} \end{cases} \quad i = 1, 2, 3, 4. \tag{35}$$

For the proximal operator of Φ, as shown by Peyré, Papadakis, and Oudet [95, Proposition 1],

$$\mathrm{Prox}_{\lambda\Phi}(u) = \big(\mathrm{Prox}_{\lambda\varphi}(\rho_{j,k}, m_{j,k})\big)_{j,k} \quad \text{for} \quad \mathrm{Prox}_{\lambda\varphi}(\rho, m) = \begin{cases} (\rho^*, m^*) & \text{if } \rho^* > 0, \\ (0, 0) & \text{otherwise,} \end{cases} \tag{36}$$

where ρ^* is the largest real root of the cubic polynomial equation $P(x) := (x - \rho)(x + \lambda)^2 - \frac{\lambda}{2}|m|^2 = 0$, and m^* can be obtained by $m^* = \rho^* m/(\rho^* + \lambda)$. By computing the proximal operator exactly, our primal dual method is positivity preserving, respecting a key property of the original Problems 1 and $1_{j,k}$.

As the computations of both proximal operators (35), (36) are *component-wise*, they can easily be *parallelized*. Likewise, the computation of the gradient ∇E is also component-wise:

$$(\nabla_u F(u))_j = (U'(\rho_{j,N_t}) + V_j + \sum_l W_{j,l}\rho_{l,N_t}(\Delta x)^d)(\Delta x)^d,$$

$$(\nabla_u H(u))_j = \frac{1}{2}(\nabla_u F(u))_j + \frac{1}{2}(U'(\rho_{j,0}) + V_j + \sum_l W_{j,l}\rho_{l,0}(\Delta x)^d)(\Delta x)^d.$$

Remark 4 (discrete convolution) As written, the above functionals involves a computation of the convolutions $\sum_l W_{j,l}\rho_{l,N_t}$ and $\sum_l W_{j,l}\rho_{l,0}$, which can be achieved efficiently using the fast Fourier transform. Note that since the product of the discrete Fourier transforms of two vectors is the Fourier transform of the circular convolution and the interaction potential $W_{j-k} = W(x_j - x_k)$ is not a periodic function, we need zero-padding for computing the convolution. For the 1D case, we can first use the fast Fourier transform to compute the circular convolution of $\mathbf{W} = (W_j)_{j=-N_x+2}^{N_x-2}$ and $(\boldsymbol{\rho}, (\mathbf{0})_{N_x-2})$, and then extract the last $N_x - 1$ elements, which are the desired convolution $\sum_k W_{j-k}\rho_k$ for $1 \le j \le N_x - 1$.

Embedding Algorithm 2 to the JKO iteration, we have the following algorithm for Wasserstein gradient flows. Note that line 6 in Algorithm 3 is to construct a better

---

Algorithm 3: Primal-Dual for JKO sequence

Input: $\rho(x, t_0)$, Iter$_{\max}$, $\lambda, \sigma, \tau, n > 0$
Output: $\rho(x, t_k)$ for $0 \leq k \leq n$ and the corresponding energy $\mathcal{E}(\rho(x, t_k))$

1 Given u^0, ϕ^0;
2 **for** $k = 1, 2, \ldots, n$ **do**
3      $u^*, \phi^* = \textbf{Algorithm 2}(u^0, \phi^0, \text{Iter}_{\max}, \lambda, \sigma, \tau)$
4      $\rho(x, t_k) = \mathsf{S}u^*$
5      $\phi^0 = \phi^*$
6      $u^0 = \max\left\{u^* - [1_{N_t+1}, 0_{N_t+1}]^t \otimes \rho(x, t_{k-1}) + [1_{N_t+1}, 0_{N_t+1}]^t \otimes \rho(x, t_k), \quad 0\right\}$.
7 **end**

---

initial guess for ρ at each JKO iteration by applying an extrapolation.

Remark 5 (Comparison of our numerical method to previous work) Our definition of the indicator function in Problems 3(a) and 3 (b) differs from previous work, and as a result, our primal-dual algorithm does not require the inversion of the matrix $\mathsf{A}\mathsf{A}^T$ [7,95], which makes it quite efficient in high dimensions thanks to the sparsity of A. A similar approach is taken in a recent preprint [81] to compute the earth mover's distance W_1, though, in this context, the earth mover's distance is dissimilar from the Wasserstein distance, since it does not require an extra time dimension and is thus a lower-dimensional problem.

A second difference between our method and the approach in previous works is that, since $P(x)$ has at most one strictly positive root, it can be obtained by the general solution formula for cubic polynomials with real coefficients. Therefore, in our numerical simulations, we may compute the proximal operator $\text{Prox}_{\lambda\Phi}(u)$ by using this general solution formula, rather than via Newton iteration [95]. As a consequence, our method is truly positivity preserving, as opposed to positivity preserving in the limit as $\Delta x, \Delta t \to 0$.

We close this section by recalling sufficient conditions on the primal and dual step sizes σ and λ that ensure Algorithms 1 and 2 converge to minimizers of Problems 2.3 and 2.3.

Proposition 1 (Convergence of Algorithm 1, c.f. [52]) *Suppose $\sigma\lambda < 1/\lambda_{max}(AA^t)$ and a minimizer of Problem 2.3 exists. Then, as Iter$_{\max} \to +\infty$, and $\epsilon_1, \epsilon_2 \to 0$ in the stopping criteria* (33)(34)*, the output u^* of Algorithm 1 converges to a minimizer of Problem 2.3.*

Proposition 2 (Convergence of Algorithm 2, c.f. [106]) *Suppose that the discrete energy $E(u)$ defined in Eq.* (32) *is proper, lower semi-continuous, convex, and there exists $\beta > 0$ such that $\langle u_1 - u_2, \nabla_u E(u_1) - \nabla_u E(u_2) \rangle \geq \beta\|\nabla E(u_1) - \nabla E(u_2)\|^2$. Suppose further that $\sigma\lambda < 1/\lambda_{max}(\tilde{A}\tilde{A}^t)$, $\lambda < 2\beta$, and a minimizer of Problem 2.3 exists. Then, as Iter$_{\max} \to +\infty$ and $\epsilon_1, \epsilon_2 \to 0$, the output u^* converges to a minimizer of Problem 2.3.*

Note here that the co-coercivity requirement on ∇E in the above proposition is equivalent to require the Lipschitz continuity of ∇E, i.e., $\|\nabla_u E(u_1) - \nabla_u E(u_2)\| \leq \frac{1}{\beta}\|u_1 - u_2\|$. For the energy of the form (4), this requirement reduces to the boundedness of $U''(\rho)$ and W, which can be satisfied independent of the numerical resolution if we consider bounded solution (no finite time blow up in ρ) and nonsingular interaction kernel. In the case when W is singular, for example when W is a Newtonian interaction potential, we approximate W by a continuous function via convolution with a mollifier; see Remark 7.

3 Convergence

We now prove the convergence of solutions of the fully discrete JKO scheme, Problem $1_{j,k}$, to a solution of the continuum JKO scheme, Problem 1. We begin, in Sect. 3.1, by describing the hypotheses we place on the underlying domain Ω, the energy $\mathcal{E}$, the initial data ρ_0, and the discretization operators. Then, in Sect. 3.2, we show that minimizers of Problem $1_{j,k}$ exist, provided the discretization is sufficiently refined. Finally, in Sect. 3.3, we prove that any sequence of minimizers of Problem $1_{j,k}$ has a subsequence that converges to a minimizer of Problem 1. In order for our finite difference approximation to converge, we assume throughout that a smooth, positive minimizer of the continuum JKO scheme Problem 1 exists. See hypothesis (H6) and Remark 9 for further discussion of this assumption.

3.1 Hypotheses

We impose the following hypotheses on the underlying domain, energy, and discretization operators.

(H1) $\Omega = \Pi_{i=1}^d (a_i, b_i) \subseteq \mathbb{R}^d$, for $a_i < b_i \in \mathbb{R}$. We assume that the spacing of the spatial discretization $(\Delta x) > 0$ and the temporal discretization $(\Delta t) > 0$ are both functions of h satisfying $\lim_{h\to 0}(\Delta x) = \lim_{h\to 0}(\Delta t) = 0$.

(H2) For any piecewise constant function ρ^h on Ω, the discrete energy functional $\mathcal{E}^h$ has one of the following forms, as described in Sect. 2.2.2:

(a) $\mathcal{F}^h(\rho^h) = \sum_j \left(U(\rho_j) + V_j \rho_j\right)(\Delta x)^d + \sum_{j,l} W_{j,l}\rho_j\rho_l(\Delta x)^{2d}$

(b) $\mathcal{H}^h_{\rho_0}(\rho^h) = \frac{1}{2}\mathcal{F}^h(\rho^h) + \frac{1}{2}\sum_j \left(U'((\rho_0)_j) + V_j + \sum_l W_{j,l}(\rho_0)_l(\Delta x)^d\right) \rho_j(\Delta x)^d$

(c) $\mathcal{G}^h_{\rho_1}(\rho_j) := \begin{cases} 0 & \text{if } \sum_j |\rho_j - (\rho_1^h)_j|^2(\Delta x)^d \leq \delta_5^2 \\ +\infty & \text{otherwise.} \end{cases}$

We place the following assumptions on U, V, and W and the target measure ρ_1:

(i) Either $U \equiv 0$ or $U \in C([0, +\infty))$ is convex, $U \in C^1((0, +\infty))$, $\lim_{r \to +\infty} \frac{U(r)}{r} = +\infty$, and $U(0) = 0$;

(ii) $V^h(x) := \sum_{j \in \mathbb{Z}^d} V_j 1_{Q_j}(x)$ and $W^h(x, y) := \sum_{(j,l) \in \mathbb{Z}^d \times \mathbb{Z}^d} W_{j,l} 1_{Q_j}(x) 1_{Q_l}$ (y) are piecewise constant approximations of $V, W \in C(\overline{\Omega})$ converging uniformly on $\overline{\Omega}$.

(iii) $\rho_1 \in C^1(\overline{\Omega})$ and ρ_1^h is a pointwise piecewise constant approximation of ρ_1.

(H3) D_t^h and D_x^h are finite difference approximations of the time derivative and spatial divergence. We assume that D_t^h is a forward Euler method in time, whereas D_x^h can be given by an explicit or implicit scheme of first or higher order. We denote by D_t^{-h} and D_x^{-h} the dual operators with respect to the ℓ^2 inner product, and we assume the following integration by parts formulas hold for all piecewise constant functions $\rho^h, f^h : [0, 1] \to \mathbb{R}$,

$$\int_0^1 D_t^h \rho^h f^h \, dt = \left(\rho^h f^h \Big|_0^1 \right) - \int_0^1 \rho^h D_t^{-h} f^h \, dt$$

and if $m^h : \Omega \to \mathbb{R}^d$, $f^h : \Omega \to \mathbb{R}$,

$$\int_\Omega D_x^h m^h f^h \, dx = \int_{\partial\Omega} f^h m^h \cdot v^h \, dx - \int_\Omega m^h D_x^{-h} f^h \, dx,$$

where $v^h : \overline{\Omega} \to \mathbb{R}^d$ is the discrete outer unit normal of Ω. Finally, we assume there exists $C > 0$ depending on the domain $\overline{\Omega} \times [0, 1]$, so that, for any $f \in C^1(\overline{\Omega} \times [0, 1]; \mathbb{R})$ and $v \in C^1(\overline{\Omega} \times [0, 1]; \mathbb{R}^d)$, if (f^h, v^h) are pointwise piecewise constant approximations,

$$\|D_t^h f^h - \partial_t f\|_\infty \leq C\|\partial_t^2 f\|_\infty (\Delta t), \quad \|D_t^{-h} f^h - \partial_t f\|_\infty \leq C\|\partial_t^2 f\|_\infty (\Delta t)$$
$$\|D_x^h v^h - \nabla \cdot v\|_\infty \leq C\|D^2 v\|_\infty (\Delta x), \quad \|D_x^{-h} f^h - \nabla f\|_\infty \leq C\|D^2 f\|_\infty (\Delta x)$$
$$\|v^h \cdot v^h - v \cdot v\|_\infty \leq C\|v\|_\infty (\Delta x).$$

(See Sect. 2.2.3 for finite difference approximations satisfying these hypotheses.)

(H4) The constraint relaxation parameters $\delta_1, \delta_2, \delta_3, \delta_4 \geq 0$ are functions of h with $\lim_{h \to 0} \delta_i = 0$, for all i. If the energy is of the form (H2c), we require that δ_5 is a function of h satisfying $\lim_{h \to 0} \delta_5 = 0$ and $\lim_{h \to 0} (\Delta x + \Delta t)/\delta_5 = 0$.

(H5) The initial data of the continuum problem satisfy $\rho_0 \in C^1(\overline{\Omega})$ and ρ_0^h is a pointwise piecewise constant approximation of ρ_0.

(H6) Given the domain, energy, and initial data described in the previous hypotheses, there exists a minimizer (ρ, m) of the continuum Problem 1 satisfying $\rho \in C^2([0, 1]; C^1(\overline{\Omega}))$, $\rho > 0$, and $m \in C^1([0, 1]; C^2(\overline{\Omega}))$.

To ease notation in the following convergence proof, we observe that Problem $1_{j,k}$ may be rewritten as follows in terms of (ρ^h, m^h), the piecewise constant functions on $\Omega \times [0, 1]$ corresponding to the coefficients $(\rho_{j,k}, m_{j,k})$.

Problem 1^h (Discrete Dynamic JKO) *Fix τ, δ_1, δ_2, δ_3, $\delta_4 > 0$, $\mathcal{E}^h$, and ρ_0^h. Solve the constrained optimization problem,*

$$\inf_{(\rho^h, m^h) \in \mathcal{C}^h} \int_0^1 \int_\Omega \Phi(\rho^h, m^h) \mathrm{d}x \mathrm{d}t + 2\tau \mathcal{E}^h(\rho^h(\cdot, 1)),$$

where (ρ^h, m^h) belong to the constraint set $\mathcal{C}^h$ provided that they are piecewise constant functions on $\Omega \times [0, 1]$ and the following inequalities hold

$$\|D_t^h \rho^h + D_x^h m^h\|_{L^2(\Omega \times [0,1])} \le \delta_1, \quad \|m^h \cdot \nu^h\|_{L^2(\partial\Omega \times [0,1])} \le \delta_2, \quad (37)$$

$$\left\| \int_\Omega \rho^h(x, \cdot)\mathrm{d}x - \int_\Omega \rho_0^h(x)\mathrm{d}x \right\|_{L^2([0,1])} \le \delta_3, \quad \|\rho^h(\cdot, 0) - \rho_0^h\|_{L^2(\Omega)} \le \delta_4. \quad (38)$$

Similarly, we may rewrite the definition of the discrete energies in hypothesis (H2) in terms of a piecewise constant functions ρ^h on Ω corresponding to ρ_j,

$$\mathcal{F}^h(\rho^h) = \int_\Omega U(\rho^h(x)) + V^h(x)\rho^h(x)\mathrm{d}x + \frac{1}{2} \iint_{\Omega \times \Omega} W^h(x, y)\rho^h(x)\rho^h(y)\mathrm{d}x\mathrm{d}y,$$

$$\mathcal{H}_{\rho_0}^h(\rho^h) = \frac{1}{2}\mathcal{F}^h(\rho^h) + \frac{1}{2} \int_\Omega \left(U'(\rho_0^h(x)) + V^h(x) + \int_\Omega W^h(x, y)\rho_0^h(y)\mathrm{d}y \right) \rho(x)\mathrm{d}x,$$

$$\mathcal{G}_{\rho_1}^h(\rho^h) = \begin{cases} 0 & \text{if } \|\rho^h - \rho_1^h\|_{L^2(\Omega)} \le \delta_5 \\ +\infty & \text{otherwise.} \end{cases}$$

Recall that, by definition of Φ in equation (13), $\Phi(\rho^h, m^h) < +\infty$ only if ρ^h is nonnegative. Consequently, if a minimizer ρ exists, it must be nonnegative.

We conclude this section with several remarks on the sharpness of the preceding hypotheses.

Remark 6 (assumption on domain Ω) In hypothesis (H1), we assume that Ω is an n-dimensional hyperrectangle. We impose this assumption for simplicity, as it provides an natural interpretation of the discretized outer unit normal ν^h, which is essential in imposing the boundary conditions for our PDE constraint at the discrete level. More generally, our convergence result can be extended to any Lipschitz domain, as long as sufficient care is taken to define the discrete outer unit normal and the corresponding no flux boundary conditions.

Remark 7 (assumption on energy) As described in hypothesis (H2), our convergence result applies to internal U, drift V, and interaction W potentials that are sufficiently regular on $\overline{\Omega}$. Our assumptions on U are classical and ensure that the internal energy is lower semicontinuous with respect to weak-* convergence [2, Remark 9.3.8]. Our assumptions on V and W, on the other hand, are somewhat stronger, and in practice, one often encounters partial differential equations for which the corresponding choices of V and W are not continuous. However, there are robust methods for approximating these potentials by continuous functions that ensure convergence of the gradient flows.

For example, the second author and Topaloglu provide sufficient conditions on discontinuous interaction potentials W for which gradient flows of the regularized interaction potential, $W_\varepsilon := W * \varphi_\varepsilon$ for a smooth mollifier φ_ε, converge to gradient flows of the original interaction potential W, as well as conditions that ensure minimizers of W_ε converge to minimizers of W [59]. (The convergence of general stationary points of W_ϵ that are not global minimizers to stationary point of W remains open.)

Remark 8 (assumption on δ_5) In hypothesis (H4), it is essential that δ_5 not vanish too quickly with respect to other parameters in the discretization. A simple illustration of this fact arises in the case that $\delta_1 \equiv \delta_2 \equiv \delta_3 \equiv \delta_4 \equiv 0$. In this case, we cannot choose $\delta_5 \equiv 0$, since our pointwise piecewise approximation of the initial data ρ_0^h will not generally have the same mass as our pointwise piecewise approximation of the target measure ρ_1^h, and if they do not have the same mass, *minimizers of the discrete problem do not exist*. Consequently, it would be impossible to prove that minimizers of the fully discrete problem converge to minimizers of the continuum problem. On the one hand, this does not greatly impact the performance of our numerical method, as can be seen by considering previous work by Papadakis, Péyre, and Oudet, which numerically implements this approach [95]. On the other hand, our numerical simulation in Fig. 3 indicates that poor choice of the relaxation parameters can cause the method to iterate longer than necessary, without any improvement in accuracy.

Our requirement that $\lim_{h\to 0}(\Delta x + \Delta t)/\delta_5 = 0$ is sufficient to fix this problem and ensure convergence of the method, and this requirement is nearly sharp. To see this, note that, for an arbitrary pointwise piecewise approximation ρ_0^h of a continuous function ρ_0, we cannot in general achieve accuracy of $|\int_\Omega \rho_0^h - \int_\Omega \rho_0|$ better than $O(\Delta x)$. If either δ_1 and δ_3, the parameters for the PDE constraint and the mass constraint, are chosen arbitrarily small, then $|\int_\Omega \rho^h(\cdot, 1) - \int_\Omega \rho_0^h|$ can likewise be made arbitrarily small. Thus, since $\rho_0, \rho_1 \in \mathcal{P}_{ac}(\Omega)$,

$$
\begin{aligned}
O(\Delta x) &\approx \left| \int_\Omega \rho_0^h - \int_\Omega \rho_0 \right| \\
&\approx \left| \int_\Omega \rho_0^h - \int_\Omega \rho_0 \right| - \left| \int_\Omega \rho_0 - \int_\Omega \rho_1 \right| \\
&\quad - \left| \int_\Omega \rho^h(\cdot, 1) - \int_\Omega \rho_0^h \right| \\
&\leq \left| \int_\Omega \rho^h(\cdot, 1) - \int_\Omega \rho_1 \right| \\
&\leq |\Omega|^{1/2} \|\rho^h(\cdot, 1) - \rho_1\|_{L^2(\Omega)} \leq |\Omega|^{1/2} \delta_5,
\end{aligned}
$$

so we much have $\delta_5 \geq O(\Delta x)$. While a CFL-type condition is not necessary for the stability of our discretization of the PDE constraint, since ρ and m indeed become coupled in the continuum limit (see Eqs. (8) and (12)), one should expect $(\Delta t) \leq O(\Delta x)$ to give the best balance between computational accuracy and cost, and we indeed observe this numerically. Combining these facts shows that enforcing that δ_5 cannot decay faster than $O(\Delta x + \Delta t)$ by assuming $\lim_{h\to 0}(\Delta x + \Delta t)/\delta_5 = 0$ is nearly optimal.

Remark 9 (assumption on existence of smooth, positive minimizer) In hypothesis (H6), we suppose that there exists a sufficiently regular minimizer $(\overline{\rho}, \overline{m})$, $\overline{\rho} > 0$, of the continuum problem. Our proof of the existence of minimizers of the fully discrete problem and our proof that minimizers of the discrete problems converge to a minimizer of the continuum problem as $h \to 0$ strongly rely on this assumption. In particular, the smoothness assumption allows us to use convergence of the finite difference operators, described in hypothesis (H3), to construct an element of $\mathcal{C}^h$ in Proposition 3. The positivity assumption allows us to conclude that $\nabla_{\rho, m} \Phi$ is uniformly bounded on the range of $\overline{\rho}$, which we use to prove the lim sup inequality for the recovery sequence in Theorem 2(b).

From the perspective of approximating gradient flows, which are solutions of diffusive partial differential equations (3), such regularity and positivity can be guaranteed as long as the initial data are smooth and positive and either the diffusion is sufficiently strong or the drift and interaction terms do not cause loss of regularity. On the other hand, developing conditions on the energy and initial data that ensure such regularity and positivity holds at the level of the JKO scheme, for minimizers of Problem 1, remains largely open: results on the propagation of $L^p(\mathbb{R}^d)$ or BV bounds along the scheme have only recently emerged [17,50,63].

From the perspective of approximating Wasserstein geodesics, the now classical regularity theory developed by Caffarelli and Urbas ensures that if the source and target measures ρ_0 and ρ_1 are smooth and strictly positive, then the minimizer of Problem 1 $(\overline{\rho}, \overline{m})$ is also smooth and strictly positive. (See, for example, [105, Section 4.3] and [2, Section 8.3].)

Along with this analytical justification for our smoothness and positivity assumptions, our numerical results also indicate that such assumptions are in general necessary. For example in Fig. 4, we observe that if the source and target measure of a Wasserstein geodesic are not sufficiently smooth, the numerical solution introduces artificial regularity. Likewise, even in Fig. 6, we observe that the numerical simulation is strictly positive (though very close to zero in places), while the exact solution is identically zero outside of its support. Still, in spite of the fact that our theoretical convergence result requires smoothness and positivity assumptions, in practice our numerical method still performs well on nonsmooth or nonpositive problems, provided that the spatial and temporal discretization are taken to be sufficiently small; see Figs. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18.

Finally, these types of smoothness and positivity assumptions are typically needed in convergence proofs for numerical methods based on the JKO scheme. For example, in a method based on the Monge Ampére approximation of the Wasserstein distance, the exact solution is required to be uniformly bounded above and below [10]. Likewise, while rigorous convergence results for fully discrete numerical methods based on entropic or Fisher information regularization remain open, since these methods correspond to introducing numerical diffusion at the level of the PDE, they automatically enforce smoothness and positivity [28,55,82].

3.2 Existence of Minimizers

We now show that, under the hypotheses described in the previous section, minimizers of the fully discrete JKO scheme, Problem 1^h, exist for all $h > 0$ sufficiently small. We begin with the following proposition, which constructs a specific element in the constraint set $\mathcal{C}^h$, which we will use both in our proof of existence of minimizers and in our Γ-convergence results in the next section.

Proposition 3 (construction of element in $\mathcal{C}^h$) *Suppose that hypotheses (H1)–(H6) hold, and choose $(\rho, m) \in \mathcal{C}$ satisfying $\rho \in C^2([0, 1]; C^1(\overline{\Omega}))$, $\rho > 0$, and $m \in C^1([0, 1]; C^2(\overline{\Omega}))$. Then for $h > 0$ sufficiently small, there exists $(\tilde{\rho}^h, \tilde{m}^h) \in \mathcal{C}^h$ satisfying $(\tilde{\rho}^h, \tilde{m}^h) \xrightarrow{h \to 0} (\rho, m)$ uniformly on $\Omega \times [0, 1]$ and*

$$\inf_{h > 0, (x,t) \in \Omega \times [0,1]} \tilde{\rho}^h(x, t) > 0. \tag{39}$$

If, in addition, the energy satisfies hypothesis (H2c) and $\mathcal{E}(\rho(\cdot, 1)) < +\infty$, then we have

$$\|\tilde{\rho}^h(\cdot, 1) - \rho_1^h\|_{L^2(\Omega)} \le \delta_5, \tag{40}$$

for all $h > 0$ sufficiently small.

Proof We construct $(\rho^h, m^h) \in \mathcal{C}^h$ as follows. Let $\hat{m}^h$ be a pointwise piecewise constant approximation of m; see Eq. (23). Recall that ν^h is the discrete outer unit normal vector. We define $\tilde{m}^h : \Omega \times [0, 1] \to \mathbb{R}^d$ component-wise to respect the no flux boundary conditions, letting $(\tilde{m}^h)_l$ denote the lth component of the vector for $l = 1, \ldots, d$. If $x \in \partial\Omega$, then we define

$$(\tilde{m}^h(x, t))_l = \begin{cases} (\hat{m}^h(x, t))_l & \text{for } e_l \cdot \nu^h(x) = 0, \\ 0 & \text{for } e_l \cdot \nu^h(x) \ne 0. \end{cases}$$

Otherwise, we take $\tilde{m}^h(x, t) = \hat{m}^h(x, t)$. Define $\tilde{\rho}^h : \Omega \times [0, 1] \to \mathbb{R}$ so that $\tilde{\rho}^h(x, 0) = \rho_0^h$ and $D_t^h \tilde{\rho}^h(x, t) + D_x^h \tilde{m}^h(x, t) \equiv 0$.

We begin by showing that $(\tilde{\rho}^h, \tilde{m}^h) \in \mathcal{C}^h$. By construction, for all $h > 0$,

$$\|D_t^h \tilde{\rho}^h + D_x^h \tilde{m}^h\|_{L^2(\Omega \times [0,1])} = 0$$
$$\|\tilde{m}^h \cdot \nu^h\|_{L^2(\partial\Omega \times [0,1])} = 0$$
$$\|\tilde{\rho}^h(\cdot, 0) - \rho_0^h\|_{L^2(\Omega)} = 0.$$

Taking $f^h \equiv 1$ in Hypothesis (H3) and applying the PDE constraint ensures that, for all $s \in [0, 1]$ and $k \in \mathbb{N}$ so that $k(\Delta t) \le s < (k + 1)\Delta t$,

$$\int_\Omega \tilde{\rho}^h(x, s)\mathrm{d}x - \int_\Omega \tilde{\rho}^h(x, 0)\mathrm{d}x = \int_0^{k(\Delta t)} \int_\Omega D_t^h \tilde{\rho}^h(x, t) f^h(x, t)\mathrm{d}x\mathrm{d}t$$

$$= -\int_0^{k(\Delta t)} \int_\Omega D_x^h \tilde{m}^h(x, t) f^h(x, t) dx dt$$

$$= -\int_0^{k(\Delta t)} \int_{\partial\Omega} \tilde{m}^h(x, t) \cdot v^h(x, t) dx dt = 0.$$

Thus, we also obtain

$$\left\| \int_\Omega \tilde{\rho}^h(x, \cdot) dx - \int_\Omega \rho_0^h(x) dx \right\|_{L^2([0,1])} = 0, \text{ for all } h > 0.$$

This concludes the proof that $(\tilde{\rho}^h, \tilde{m}^h) \in C^h$.

We now show that $(\tilde{\rho}^h, \tilde{m}^h) \to (\rho, m)$ uniformly on $\Omega \times [0, 1]$ as $h \to 0$. We begin by proving convergence of $\tilde{m}^h$ to m. Due to hypothesis (H1) on our domain Ω, whenever $e_i \cdot v^h(x) \neq 0$, there exists $y \in \partial\Omega$ so that $|y - x| \leq 2\sqrt{d}(\Delta x)$ and $v(y) = e_i$. Thus, whenever $e_i \cdot v^h(x) \neq 0$, the continuum boundary condition $m(y, t) \cdot v(y) = 0$ ensures that for all $t \in [0, 1]$,

$$|(\tilde{m}^h(x, t) - m(x, t))_i| = |m(x, t) \cdot e_i| \leq |(m(x, t) - m(y, t)) \cdot e_i|$$
$$+ |m(y, t) \cdot e_i| \leq 2\sqrt{d}(\Delta x) \|Dm\|_\infty.$$

We also have that, for all $(x, t) \in \Omega \times [0, 1]$,

$$|\hat{m}^h(x, t) - m(x, t)| \leq (\Delta x) \|Dm\|_\infty + (\Delta t) \|\partial_t m\|_\infty.$$

Therefore, for all $(x, t) \in \Omega \times [0, 1]$, there exists $C_m = C_m(d, \|Dm\|_\infty, \|\partial_t m\|_\infty) > 0$ so that

$$|\tilde{m}^h(x, t) - m(x, t)| \leq C_m(\Delta t + \Delta x) \xrightarrow{h \to 0} 0.$$

We now prove the convergence of $\tilde{\rho}^h$ to ρ. Since (ρ, m) is a classical solution of the PDE constraint and $\tilde{\rho}^h : \Omega \times [0, 1] \to \mathbb{R}$ is defined by the conditions that $\tilde{\rho}^h(x, 0) = \hat{\rho}_0^h$ and $D_t^h \tilde{\rho}^h(x, t) + D_x^h \tilde{m}^h(x, t) \equiv 0$, for $(x, t) \in \Omega \times [0, 1]$ and $k \in \mathbb{N}$ so that $k(\Delta t) \leq t < (k + 1)(\Delta t)$, we have

$$|\tilde{\rho}^h(x, t) - \rho(x, t)|$$
$$= \left| \tilde{\rho}^h(x, 0) + \int_0^{k(\Delta t)} D_t^h \tilde{\rho}^h(x, s) ds - \rho(x, 0) - \int_0^t \partial_s \rho(x, s) ds \right|$$
$$= \left| \rho_0^h(x) - \int_0^{k(\Delta t)} D_x^h \tilde{m}^h(x, s) ds - \rho(x, 0) + \int_0^t \nabla \cdot m(x, s) ds \right|$$
$$= \left| \rho_0^h(x) - \rho(x, 0) \right| + \left| \int_0^{k(\Delta t)} D_x^h \tilde{m}^h(x, s) ds - \int_0^{k(\Delta t)} \nabla \cdot m(x, s) ds \right|$$

$$+ \left| \int_{k(\Delta t)}^{t} \nabla \cdot m(x, s) \mathrm{d}s \right|$$

$$\leq \|\nabla \rho\|_{\infty}(\Delta x) + C\|D^2 m\|_{\infty}(\Delta x)$$

$$+ \|\nabla \cdot m\|_{\infty}(\Delta t) \xrightarrow{h \to 0} 0. \tag{41}$$

Since $\tilde{\rho}^h \to \rho$ uniformly and $\rho > 0$, we immediately obtain (39).

Finally, suppose the energy satisfies (H2c). Since $\mathcal{E}(\rho(\cdot, 1)) = \mathcal{G}_{\rho_1}(\rho(\cdot, 1)) < +\infty$, we have $\rho(\cdot, 1) = \rho_1$. By inequality (41) and the fact that ρ_1^h is a pointwise piecewise approximation of $\rho(\cdot, 1)$,

$$\|\tilde{\rho}^h(\cdot, 1) - \rho_1^h\|_{L^2(\Omega)} \leq |\Omega|^{1/2} \left(\|\tilde{\rho}^h(\cdot, 1) - \rho(\cdot, 1)\|_{\infty} \right.$$

$$\left. + \|\rho(\cdot, 1) - \rho_1^h\|_{\infty} \right) \leq C_{\rho, m}(\Delta x + \Delta t)$$

where $C_{\rho, m} = C_{\rho, m}(\Omega, \|\nabla \rho\|_{\infty}, \|\nabla \cdot m\|_{\infty}, \|D^2 m\|_{\infty}) > 0$. By hypothesis (H4), $\lim_{h \to 0} \frac{(\Delta x + \Delta t)}{\delta_5} \to 0$. Thus, for h sufficiently small,

$$\|\tilde{\rho}^h(\cdot, 1) - \rho_1^h\|_{L^2(\Omega)} \leq \delta_5,$$

which completes the proof.

$\square$

Theorem 1 (minimizers of discrete dynamic JKO exist) *Suppose that hypotheses (H1)–(H6) hold. Then for all $h > 0$ sufficiently small, a minimizer of Problem 1^h exists.*

Proof First, we note that Proposition 3 ensures that, for $h > 0$ sufficiently small, the constraint set $\mathcal{C}^h$ is nonempty and contains some (ρ^h, m^h) satisfying $\rho^h > 0$. If the energy satisfies (H2a) or (H2b), then we immediately obtain $\mathcal{E}^h(\rho^h(\cdot, 1)) < +\infty$. Similarly, if the energy satisfies (H2c), then inequality (40) in Proposition 3 again ensures that $\mathcal{E}^h(\rho^h(\cdot, 1)) < +\infty$.

Since $\Phi(\rho^h, m^h) < +\infty$ whenever $\rho^h \geq 0$, this ensures that value of the objective function in the discrete minimization problem 1^h is not identically $+\infty$ on the constraint set. Therefore,

$$\inf_{(\rho^h, m^h) \in \mathcal{C}^h} \int_0^1 \int_{\Omega} \Phi(\rho^h(x, t), m^h(x, t)) \mathrm{d}x \mathrm{d}t + 2\tau \mathcal{E}^h(\rho^h(\cdot, 1)) < +\infty, \tag{42}$$

and we may choose a minimizing sequence $(\rho_n^h, m_n^h) \in \mathcal{C}^h$ that converges to the infimum. We may assume, without loss of generality, that

$$\sup_n \int_0^1 \int_{\Omega} \Phi(\rho_n^h(x, t), m_n^h(x, t)) \mathrm{d}x \mathrm{d}t + 2\tau \mathcal{E}^h(\rho_n^h(\cdot, 1)) < +\infty, \tag{43}$$

To conclude the proof of the theorem, we will now show that there exists (ρ_*^h, m_*^h) so that a subsequence of (ρ_n^h, m_n^h) converges to (ρ_*^h, m_*^h) uniformly on $\Omega \times [0, 1]$. Then, since the objective functional $\mathcal{E}^h$ is lower semi-continuous along uniformly convergent sequences [1, Example 2.36] and the constraint set $\mathcal{C}^h$ is closed under uniform convergence for fixed $h > 0$, (ρ_*^h, m_*^h) must be a minimizer of the fully discrete problem.

In order to obtain compactness of (ρ_n^h, m_n^h), first note that (42) ensures $\Phi(\rho^h, m^h) < +\infty$ on $\overline{\Omega} \times [0, 1]$, so $\rho^h \geq 0$ on $\overline{\Omega}$. Furthermore, the mass constraint (38) ensures that there exists $R = R(h) > 0$, depending on $\Omega, (\Delta x), (\Delta t)$, and δ_3 so that $|\rho_n^h(x, t)| \leq R$ for all $(x, t) \in \Omega \times [0, 1]$. Therefore, the vector of coefficients $(\rho_n^h)_{j,k}$ for this piecewise constant function satisfies $(\rho_n^h)_{j,k} \in B_R(0) \subseteq \mathbb{R}^{N_x^d N_t}$. Consequently, by the Heine–Borel theorem, there exists a vector $(\rho_*^h)_{j,k} \in \mathbb{R}^{N_x^d N_t}$ so that, up to a subsequence, $(\rho_n^h)_{j,k} \to (\rho_*^h)_{j,k}$. Therefore, if ρ_*^h denotes the corresponding piecewise constant function, we have that, up to taking a subsequence which we again denote by $\rho_n^h(x, t)$, $\lim_{n \to +\infty} \rho_n^h(x, t) = \rho_*^h(x, t)$ uniformly on $\Omega \times [0, 1]$.

Next, we show that

$$\inf_n \mathcal{E}^h(\rho_n^h(\cdot, 1)) > -\infty. \tag{44}$$

If the energy satisfies (H2c), then $\mathcal{E}^h(\rho_n^h(\cdot, 1)) \geq 0$ for all n, and the above inequality is immediate. If the energy satisfies (H2a) or (H2b), then this follows from the fact that U is bounded below on $[0, +\infty]$, V and W are bounded below on $\overline{\Omega}$ and $\rho_n^h(x, t) \to \rho_*^h(x, t)$ uniformly.

Combining (43) and (44), we obtain

$$\sup_n \int_0^1 \int_\Omega \Phi(\rho_n^h(x, t), m_n^h(x, t)) \mathrm{d}x \mathrm{d}t < +\infty. \tag{45}$$

Furthermore, since $0 \leq \rho_n^h(x, t) \leq R$ for all $(x, t) \in \Omega \times [0, 1]$, $n \in \mathbb{N}$, we have

$$\Phi(\rho_n^h(x, t), m_n^h(x, t)) \geq |m_n^h(x, t)|^2 / R. \tag{46}$$

Therefore, combining (45) and (46), we obtain that there exists $R' = R'(h) > 0$, depending on $\Omega, (\Delta x), (\Delta t)$, and δ_3, so that $|m_n^h(x, t)| \leq R'$ for all $(x, t) \in \Omega \times [0, 1]$. Arguing as before, the Heine–Borel theorem ensures that, up to a subsequence, $\lim_{n \to +\infty} m_n^h(x, t) = m_*^h(x, t)$ uniformly on $\Omega \times [0, 1]$, for some piecewise constant function $m_*^h(x, t)$. This gives the result.

□

3.3 Convergence of Minimizers

We now prove that minimizers of the discrete dynamic JKO scheme, Problem 1^h converge to minimizers of Problem 1 as $h \to 0$. We begin with the following lemma, showing that any $(\rho^h, m^h) \in \mathcal{C}^h$ satisfies a weak form of the PDE constraint, in the limit as $h \to 0$.

Lemma 1 (properties of $\mathcal{C}^h$) *Suppose that hypotheses (H1)–(H6) hold, and fix $(\rho^h, m^h) \in \mathcal{C}^h$ so that $\int_0^1 \int_\Omega \Phi(\rho^h, m^h) < +\infty$ for each $h > 0$. Then $\rho^h(\cdot, 0) \to \rho_0$ in $L^2(\Omega)$, and there exist $\rho \in \mathcal{P}(\Omega \times [0, 1])$ and $\mu \in \mathcal{P}(\Omega)$ so that, up to a subsequence, $\rho^h \overset{*}{\rightharpoonup} \rho$ and $\rho^h(\cdot, 1) \overset{*}{\rightharpoonup} \mu$. Furthermore, for any piecewise constant function f^h with $\sup_{h>0} \|f^h\|_{L^2(\Omega \times [0,1])} + \|f^h\|_{L^2(\partial\Omega \times [0,1])} < +\infty$, we have*

$$\int_0^1 \int_\Omega \left(D_t^{-h} f^h \rho^h + D_x^{-h} f^h \cdot m^h \right) \mathrm{d}x \mathrm{d}t$$
$$+ \int_\Omega \left(f^h(\cdot, 0)\rho^h(\cdot, 0) - f^h(\cdot, 1)\rho^h(\cdot, 1) \right) \mathrm{d}x \mathrm{d}t \xrightarrow{h \to 0} 0. \qquad (47)$$

Proof By hypothesis (H6), $\rho_0^h \to \rho_0$ uniformly on $\overline{\Omega}$. Likewise, the constraint on the initial data (38) and (H4) ensure $\lim_{h\to 0} \|\rho^h(\cdot, 0) - \rho_0^h\|_{L^2(\Omega)} \leq \lim_{h \to 0} \delta_4 = 0$. Thus, $\rho^h(\cdot, 0) \to \rho_0$ in $L^2(\Omega)$.

We now turn to Eq. (47). By the PDE constraint and boundary conditions (37) and summation by parts, via hypotheses (H3),

$$\left| \int_0^1 \int_\Omega \left(D_t^{-h} f^h \rho^h + D_x^{-h} f^h \cdot m^h \right) \mathrm{d}x \mathrm{d}t \right.$$
$$\left. + \int_\Omega \left(f^h(\cdot, 0)\rho^h(\cdot, 0) - f^h(\cdot, 1)\rho^h(\cdot, 1) \right) \mathrm{d}x \right|$$
$$= \left| \int_0^1 \int_\Omega \left(f^h D_t^h \rho^h + f^h D_x^h m^h \right) \mathrm{d}x \mathrm{d}t - \int_0^1 \int_{\partial\Omega} f^h m^h \cdot v^h \mathrm{d}x \right|$$
$$\leq \|f^h\|_{L^2(\Omega \times [0,1])} \|D_t^h \rho^h + D_x^h m^h\|_{L^2(\Omega \times [0,1])}$$
$$+ \|f^h\|_{L^2(\partial\Omega \times [0,1])} \|m^h \cdot v^h\|_{L^2(\partial\Omega \times [0,1])}$$
$$\leq \delta_4 \|f^h\|_{L^2(\Omega \times [0,1])} + \delta_2 \|f^h\|_{L^2(\partial\Omega \times [0,1])} \xrightarrow{h \to 0} 0,$$

where, in the last line, we use that (H4) ensures $\delta_2, \delta_4 \to 0$ and the fact that f^h is bounded uniformly in h in $L^2(\Omega \times [0, 1]$ and $L^2(\partial\Omega \times [0, 1])$.

Next, we show that there exist $\rho \in \mathcal{P}(\Omega \times [0, 1])$ and $\mu \in \mathcal{P}(\Omega)$ so that, up to a subsequence, $\rho^h \overset{*}{\rightharpoonup} \rho$ and $\rho^h(\cdot, 1) \overset{*}{\rightharpoonup} \mu$. By Hölder's inequality and the mass constraint (38),

$$\left\| \int_\Omega \rho^h(x, \cdot) \mathrm{d}x - \int_\Omega \rho_0^h(x) \mathrm{d}x \right\|_{L^1([0,1])}$$
$$\leq \left\| \int_\Omega \rho^h(x, \cdot) \mathrm{d}x - \int_\Omega \rho_0^h(x) \mathrm{d}x \right\|_{L^2([0,1])} \leq \delta_3 \xrightarrow{h \to 0} 0,$$

where, in the last line, we use that (H4) ensures $\delta_3 \to 0$. Since hypothesis (H6) ensures $\rho_0^h \to \rho_0$ uniformly and $\int_\Omega \rho_0 = 1$, we obtain,

$$\int_0^1 \int_\Omega \rho^h(x, s) \mathrm{d}x \mathrm{d}s \to 1.$$

Furthermore, since $\int_0^1 \int_\Omega \Phi(\rho^h, m^h) < +\infty$ for each $h > 0$, we must have $\rho^h \geq 0$ on $\Omega \times [0, 1]$, and the above equation ensures $\sup_{h>0} \|\rho^h\|_{L^1(\Omega \times [0,1))} < +\infty$. Thus, classical functional analysis results ensure there exists a subsequence that converges to some $\rho \in \mathcal{P}(\Omega \times [0, 1])$ in the weak-* topology (see, e.g., [20, Section 3]).

Finally, taking $f^h \equiv 1$ in Eq. (47) gives,

$$\lim_{h \to 0} \int_\Omega \rho^h(\cdot, 0) - \rho^h(\cdot, 1) \mathrm{d}x = 0 \quad \Longrightarrow \quad \lim_{h \to 0} \int_\Omega \rho^h(\cdot, 1) \mathrm{d}x = 1$$
$$\Longrightarrow \quad \sup_{h>0} \|\rho^h(\cdot, 1)\|_{L^1(\Omega)} < +\infty.$$

Arguing as above, we obtain that, up to a further subsequence, $\rho^h(\cdot, 1) \overset{*}{\rightharpoonup} \mu(\cdot)$ for $\mu \in \mathcal{P}(\Omega)$. □

We now prove that the discrete energies $\mathcal{E}^h$ are lower semicontinuous along weak-* convergent sequences.

Proposition 4 (Lower semicontinuity of energies along weak-* convergent sequences)
Suppose that hypotheses (H1)–(H6) hold. Then, for any sequence of piecewise constant functions $\rho^h : \Omega \to \mathbb{R}$ such that $\rho^h \overset{*}{\rightharpoonup} \rho$, we have $\liminf_{h \to 0} \mathcal{E}^h(\rho^h) \geq \mathcal{E}(\rho)$.

Proof First, suppose the energy satisfies (H2a). Since the piecewise constant approximations $\hat{V}^h$ and $\hat{W}^h$ converge to V and W uniformly, for any sequence $\rho^h \overset{*}{\rightharpoonup} \rho$,

$$\lim_{h \to 0} \int_\Omega V^h \rho^h \mathrm{d}x = \int (V^h - V)\rho^h \mathrm{d}x + \int V \rho^h \mathrm{d}x = \int V \rho \, \mathrm{d}x, \tag{48}$$

$$\lim_{h \to 0} \int_{\Omega \times \Omega} W^h(x, y) \rho^h(x) \rho^h(y) \mathrm{d}x \mathrm{d}y = \int_{\Omega \times \Omega} W(x - y) \mathrm{d}\rho(x) \mathrm{d}\rho(y). \tag{49}$$

Furthermore, our assumptions on U guarantee that the internal energy term is lower semicontinuous with respect to weak-* convergence [2, Remark 9.3.8], so $\liminf_{h \to 0} \int_\Omega U(\rho^h(x)) \mathrm{d}x \geq \int_\Omega U(\rho(x)) \mathrm{d}x$. Combining this with equations (48-49) gives the result.

Next, suppose the energy satisfies (H2b). Since $\rho_0 > 0$ on the compact set $\overline{\Omega}$ and U' is uniformly continuous on $\rho_0(\overline{\Omega}) \subset (0, +\infty)$, the fact that hypothesis (H6) ensures $\hat{\rho}_0^h \to \rho_0$ uniformly ensures $U'(\hat{\rho}_0^h) \to U'(\rho_0)$ uniformly. Therefore,

$$\lim_{h \to 0} \int_\Omega U'(\hat{\rho}_0^h) \rho^h \mathrm{d}x = \int_\Omega U'(\rho_0) \rho \mathrm{d}x.$$

Likewise, since $\hat{V}^h$ and $\hat{W}^h$ converge to V and W uniformly, we also have

$$\lim_{h \to 0} \int_{\Omega} \left(\hat{V}^h(x) + \int_{\Omega} \hat{W}^h(x, y)\rho_0^h(y)dy \right) \rho^h(x)dx$$
$$= \int_{\Omega} \left(V(x) + \int_{\Omega} W(x, y)\rho_0(y)dy \right) \rho(x)dx. \tag{50}$$

Combining these limits with the lim inf inequality for energies of the form (H2a) gives the result.

Finally, suppose the energy satisfies (H2c). Without loss of generality, we may assume that $\liminf_{h \to 0} \mathcal{G}_{\rho_1}^h(\rho^h) < +\infty$, so that up to a subsequence, $\mathcal{G}_{\rho_1}^h(\rho^h) \equiv 0$ and $\lim_{h \to 0} \|\rho^h - \rho_1^h\|_{L^2(\Omega)} = 0$. By uniqueness of limits, $\rho = \rho_1$. Thus, since $\mathcal{G}_{\rho_1}^h \geq 0$, we have $\liminf_{h \to 0} \mathcal{G}_{\rho_1}^h(\rho^h) \geq 0 = \mathcal{G}_{\rho_1}(\rho)$.
□

We now apply Proposition 4 to prove the Γ-convergence of Problem 1^h to Problem 1.

Theorem 2 (Γ-convergence of discrete to continuum JKO) *Suppose hypotheses (H1)–(H6) hold.*

(a) *If $(\rho^h, m^h) \in \mathcal{C}^h$ with $(\rho^h, m^h) \overset{*}{\rightharpoonup} (\rho, m)$, then $(\rho, m) \in \mathcal{C}$ and*

$$\liminf_{h \to 0} \int_0^1 \int_{\Omega} \Phi(\rho^h, m^h)dxdt + 2\tau \mathcal{E}^h(\rho^h(\cdot, 1))$$
$$\geq \int_0^1 \int_{\Omega} \Phi(\rho, m)dxdt + 2\tau \mathcal{E}(\rho(\cdot, 1)).$$

(b) *For any $(\rho, m) \in \mathcal{C}$ satisfying $\rho \in C^2([0, 1]; C^1(\overline{\Omega}))$, $\rho > 0$, and $m \in C([0, 1]; C^2(\overline{\Omega}))$, there exists a sequence $(\tilde{\rho}^h, \tilde{m}^h) \in \mathcal{C}^h$ so that $(\tilde{\rho}^h, \tilde{m}^h) \to (\rho, m)$ uniformly and*

$$\limsup_{h \to 0} \int_0^1 \int_{\Omega} \Phi(\tilde{\rho}^h, \tilde{m}^h)dxdt + 2\tau \mathcal{E}^h(\tilde{\rho}^h(\cdot, 1))$$
$$\leq \int_0^1 \int_{\Omega} \Phi(\rho, m)dxdt + 2\tau \mathcal{E}(\rho(\cdot, 1)).$$

Proof We first prove part (a). Suppose $(\rho^h, m^h) \in \mathcal{C}^h$, with $\rho^h \overset{*}{\rightharpoonup} \rho$ and $m^h \overset{*}{\rightharpoonup} m$. We begin by showing that the limit (ρ, m) belongs to $\mathcal{C}$. Fix $f \in C^\infty(\Omega \times [0, 1])$ and let f^h be a pointwise piecewise constant approximation of f. (See Eq. (23).) By Lemma 1 and hypothesis (H3),

$$\int_0^1 \int_{\Omega} (f_t \rho + \nabla f \cdot m) \, dxdt + \int_{\Omega} f(\cdot, 0)\rho(\cdot, 0) - f(\cdot, 1)\mu dx = 0.$$

We conclude that (ρ, m) satisfies the PDE constraint in the sense of distributions (17), which gives $\rho \in AC([0, 1], \mathcal{P}(\Omega))$ [2, Lemma 8.1.2]. In particular, since ρ is continuous in time, we have that the μ defined in Lemma 1 satisfies $\mu = \rho(\cdot, 1)$.

We now consider the inequality in part (a). Since the integral functional $(\rho, m) \mapsto \int_0^1 \int_\Omega \Phi(\rho, m)$ is lower semicontinuous with respect to weak-* convergence of measures [1, Example 2.36], we immediately obtain

$$\liminf_{h \to 0} \int_0^1 \int_\Omega \Phi(\rho^h, m^h) \mathrm{d}x \mathrm{d}t \geq \int_0^1 \int_\Omega \Phi(\rho, m) \mathrm{d}x \mathrm{d}t.$$

This ensures $m \in L^1([0, 1], L^2(\rho^{-1}))$ and completes the proof that $(\rho, m) \in \mathcal{C}$. Finally, since Lemma 1 ensures $\rho^h(\cdot, 1) \overset{*}{\rightharpoonup} \mu = \rho(\cdot, 1)$, applying Proposition 4 gives

$$\liminf_{h \to 0} \mathcal{E}^h(\rho^h(\cdot, 1)) \geq \mathcal{E}(\rho(\cdot, 1)),$$

which completes the proof of part (a).

We now turn to part (b). Let $(\tilde{\rho}^h, \tilde{m}^h) \in \mathcal{C}^h$ be the sequence constructed in Proposition 3, so $(\tilde{\rho}^h, \tilde{m}^h) \to (\rho, m)$ uniformly. By inequality (39), there exists $c > 0$ so that $\rho^h(x, t) \geq c$ for h sufficiently small. Therefore,

$$\left| \int_0^1 \int_\Omega \Phi(\tilde{\rho}^h, \tilde{m}^h) - \int_0^1 \int_\Omega \Phi(\rho, m) \right|$$
$$\leq |\Omega| \|\nabla_{\rho, m} \Phi\|_{L^\infty(\{\rho \geq c\})} \left(\|\tilde{m}^h - m\|_\infty + \|\tilde{\rho}^h - \rho\|_\infty \right) \xrightarrow{h \to 0} 0.$$

It remains to show that

$$\limsup_{h \to 0} \mathcal{E}^h(\tilde{\rho}^h(\cdot, 1)) \leq \mathcal{E}(\rho(\cdot, 1)).$$

First, suppose the energy satisfies either (H2a) or (H2b). By Eqs. (48)–(50), which hold for any weak-* convergent sequence, and the fact that $U'(\tilde{\rho}^h(\cdot, 0)) \to U'(\rho(\cdot, 0))$ uniformly, it suffices to show

$$\limsup_{h \to 0} \int_\Omega U(\tilde{\rho}^h(\cdot, 1)) \mathrm{d}x \leq \int_\Omega U(\rho(\cdot, 1)) \mathrm{d}x.$$

Since $U \in C([0, +\infty])$, $\rho(\cdot, 1) \in L^\infty(\mathbb{R}^d)$, and $\tilde{\rho}^h(\cdot, 1) \to \rho(\cdot, 1)$ uniformly, $U(\tilde{\rho}^h(\cdot, 1)) \to \int U(\rho(\cdot, 1))$ uniformly, which gives the result.

Finally, suppose the energy satisfies (H2c). Without loss of generality, suppose $\mathcal{E}(\rho(\cdot, 1)) = \mathcal{G}_{\rho_1}(\rho(\cdot, 1)) < +\infty$. Inequality (40) ensures that, for h sufficiently small,

$$\|\tilde{\rho}^h(\cdot, 1) - \rho_1^h\|_{L^2(\Omega)} \leq \delta_5.$$

By definition of $\mathcal{G}_{\rho_1}^h$, this implies $\mathcal{G}_{\rho_1}^h(\tilde{\rho}^h(\cdot, 1)) \equiv 0$. Therefore,

$$\limsup_{h \to 0} \mathcal{G}_{\rho_1}^h(\tilde{\rho}^h(\cdot, 1)) = 0 \le \mathcal{G}_{\rho_1}(\rho_1),$$

which gives the result. □

We conclude this section by applying the Γ-convergence proof from Theorem 2 to prove that any sequence of minimizers of the discrete Problem 1^h converges, up to a subsequence, to a minimizer of the continuum Problem 1.

Theorem 3 (Convergence of minimizers) *Suppose that hypotheses (H1)–(H6) hold. Then, for any sequence of minimizers (ρ^h, m^h) of Problem 1^h, we have, up to a subsequence, $\rho^h \overset{*}{\rightharpoonup} \rho$ and $m^h \overset{*}{\rightharpoonup} m$, where (ρ, m) is a minimizer of Problem 1.*

Note that, if the minimizer of the continuum Problem 1 is unique, then this theorem ensures that any sequence of minimizers of the discrete Problem $1_{j,k}$ has a further subsequence that converges to this minimizer. Therefore, the sequence itself must converge to the unique minimizer of the continuum problem. (See Remark 2 for sufficient conditions that ensure the minimizer of the continuum problem is unique.)

Proof of Theorem 3 First, note that Lemma 1 ensures that there exist $\rho \in \mathcal{P}(\Omega \times [0, 1])$ and $\mu \in \mathcal{P}(\Omega)$ so that, up to a subsequence, $\rho^h \overset{*}{\rightharpoonup} \rho$ and $\rho^h(\cdot, 1) \overset{*}{\rightharpoonup} \mu$. In order to prove an analogous weak-* compactness result for m^h we first prove that, up to a subsequence,

$$\sup_{h>0} \int_0^1 \int_\Omega \Phi(\rho^h, m^h) < +\infty. \tag{51}$$

By (H6), there exists a minimizer $(\bar{\rho}, \bar{m})$ of the continuum Problem 1 satisfying $\bar{\rho} \in C^2([0, 1]; C^1(\overline{\Omega}))$, $\bar{\rho} > 0$, and $\bar{m} \in C^1([0, 1]; C^2(\overline{\Omega}))$. Comparing the recovery sequence $(\tilde{\rho}^h, \tilde{m}^h) \in \mathcal{C}^h$ from Theorem 2(b) for $(\bar{\rho}, \bar{m})$ with the discrete minimizer $(\rho^h, m^h) \in \mathcal{C}^h$, we obtain

$$\limsup_{h \to 0} \int_0^1 \int_\Omega \Phi(\rho^h, m^h) + 2\tau \mathcal{E}^h(\rho^h(\cdot, 1))$$

$$\le \limsup_{h \to 0} \int_0^1 \int_\Omega \Phi(\tilde{\rho}^h, \tilde{m}^h) + 2\tau \mathcal{E}^h(\tilde{\rho}^h(\cdot, 1))$$

$$\le \int_0^1 \int_\Omega \Phi(\bar{\rho}, \bar{m}) + 2\tau \mathcal{E}(\bar{\rho}(\cdot, 1)). \tag{52}$$

Furthermore, Proposition 4 ensures that

$$\liminf_{h \to 0} 2\tau \mathcal{E}^h(\rho^h(\cdot, 1)) \ge 2\tau \mathcal{E}(\mu),$$

which is bounded below by some constant, since hypothesis (H2c) ensures $\mathcal{E} \geq 0$ and hypotheses (H2a) or (H2b) ensures $\mathcal{E}(\mu) > -\infty$, since U, V, and W are bounded below and U' is bounded below on the range of the strictly positive density ρ_0. Therefore, up to a subsequence, we obtain (51).

We now deduce weak-* convergence of m^h. By Hölder's inequality, the fact that $\rho^h \overset{*}{\rightharpoonup} \rho$, and the definition of Φ, we have

$$\sup_{h>0} \|m^h\|_{L^1(\Omega \times [0,1])} \leq \sup_{h>0} \left(\int_0^1 \int_\Omega \Phi(\rho^h, m^h) \right) \left(\int_0^1 \int_\Omega 1^2 \rho^h \right)^{1/2} < +\infty.$$

Thus, up to another subsequence, $m^h \overset{*}{\rightharpoonup} m$ on $\Omega \times [0, 1]$.

It remains to show that the limit (ρ, m) of (ρ^h, m^h) is a minimizer of Problem 1. By Theorem 2, part (a), we have $(\rho, m) \in \mathcal{C}$ and

$$\int_0^1 \int_\Omega \Phi(\rho, m) + 2\tau \mathcal{E}(\rho(\cdot, 1)) \leq \liminf_{h \to 0} \int_0^1 \int_\Omega \Phi(\rho^h, m^h) + 2\tau \mathcal{E}^h(\rho^h(\cdot, 1)).$$

Combining this with inequality (52) above, we conclude that $(\rho, m) \in \mathcal{C}$ is also a minimizer of Problem 1, which completes the proof. $\qquad\square$

4 Numerical Results

In this section, we provide several examples demonstrating the efficiency and accuracy of our algorithms. We begin by using Algorithm 1 to compute Wasserstein geodesics between given source and target measures, and we then turn to Algorithm 3 to compute solutions of nonlinear gradient flows. In the following simulations, we take our computational domain Ω to be a square, imposing the no flux boundary conditions on m dimension by dimension. In practice, unless otherwise specified, we always impose the discrete PDE constraint via the Crank–Nicolson finite difference operators (28), and we choose $\epsilon_1 = \epsilon_2 = \epsilon$ in the stopping criteria to be 10^{-5} unless otherwise specified. For the relaxation of the constraints in (30) and (31), we choose $\delta_1 = \delta_2 = \delta_4 = \delta_5 = \delta$, and δ_3 differently, as specified in each example.

4.1 Wasserstein Geodesics

As described in Remark 1, a particular case of our numerical scheme provides a method for computing the Wasserstein geodesic between two probability densities. We begin by computing the Wasserstein geodesic between rescaled Gaussians in one dimension:

$$g_{\mu,\theta}(x) = \frac{1}{(2\pi\theta^2)^{d/2}} e^{-\frac{(x-\mu)^2}{\theta^2}}. \tag{53}$$
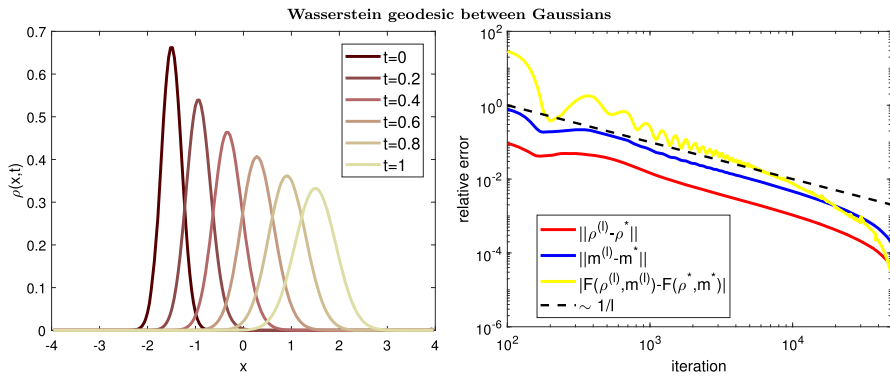
Fig. 2 We compute the Wasserstein geodesic between two Gaussians on the domain $\Omega = [-4, 4]$, with $N_t = 20$ temporal grid points ($\Delta t = \frac{1}{20}$) and $N_x = 200$ spatial points ($\Delta x = \frac{8}{200}$). We choose $\sigma = 0.1$ and $\sigma\lambda = 1.5/\lambda_{max}(AA^t)$ and compute 10^5 iterations. Left: evolution of geodesic from time $t = 0$ to $t = 1$. Right: rate of convergence of numerical solution to exact solution, as a function of the number of iterations in Algorithm 1

The target measure is simply a translation and dilation of the initial measure, $\rho_0(x) = (0.5)g_{\mu_0,\theta_0}(x)$ and $\rho_1(x) = (0.5)g_{\mu_1,\theta_1}(x)$. The optimal transport map $T(x)$ from $\rho_0(x)$ to $\rho_1(x)$ is given explicitly by[1]

$$T(x) = \frac{\theta_1}{\theta_0}(x - \mu_0) + \mu_1.$$

Rewriting Eq. (15) for the geodesic $\rho(x, t)$ and velocity $v(x, t)$ induced by this transport map, via the definition of the push forward, we obtain

$$\rho(x, t) = \rho_0(T_t^{-1}(x))\det(\nabla_x T_t^{-1}) \quad \text{and} \quad m(x, t)$$
$$= \rho(x, t)v(x, t) = \rho(x, t)(T \circ T_t^{-1}(x) - T_t^{-1}(x))),$$
$$T_t^{-1}(x) = \frac{x + (\frac{\theta_1}{\theta_0}\mu_0 - \mu_1)t}{1 - t + t\frac{\theta_1}{\theta_0}}, \quad \det(\nabla_x T_t^{-1}) = \frac{1}{1 - t + t\frac{\theta_1}{\theta_0}}.$$

In Fig. 2, we apply Algorithm 1 to compute the Wasserstein geodesic $\rho(x, t)$ between the initial and target densities (53), with means and variances $\mu_0 = -1.5$, $\theta_0 = 0.3$, $\mu_1 = 1.5$, and $\theta_1 = 0.6$. On the left, we plot the evolution of the geodesic at various times. On the right, we plot the ℓ^1 error of the densities, momenta, and Wasserstein distance as a function of the number of iterations, l, observing a rate of convergence of order $\mathcal{O}(1/l)$ (dashed black line). Here, the error is defined as

$$\|\rho^{(l)} - \rho^*\| = \frac{1}{N_x(N_s + 1)} \sum_{k=0}^{N_s} \sum_{j=1}^{N_x-1} |\rho_{j,k}^{(l)} - \rho_{j,k}^*|. \tag{54}$$

---

[1] One way to see that this is the unique optimal transport map from ρ_0 to ρ_1 is to note that $T\#\rho_0 = \rho_1$ and $T(x)$ is the gradient of a convex function; see, for example, [2, Section 6.2.3].
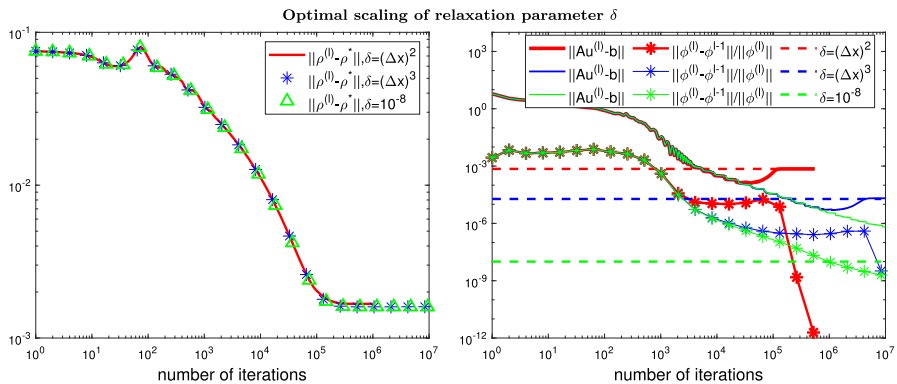
Fig. 3 Analysis of how the scaling relationship between the relaxation parameter δ and the spatial discretization (Δx) affects the accuracy of the numerical method and the number of iterations required to converge. We contrast the choices $\delta = (\Delta x)^2$, $\delta = (\Delta x)^3$ and $\delta = 10^{-8}$ for the example of the Wasserstein distance between geodesics, illustrated in Fig. 2. We take $N_t = 30$, $N_x = 300$, $\sigma = 1$, $\sigma\lambda = 0.99/\lambda_{max}(AA^t)$ and $\delta_3 = \delta$

In Fig. 3, we illustrate how choosing the optimal scaling relationship between the relaxation parameter δ and the spatial and temporal discretizations (Δx), (Δt) allows the method to converge in fewer iterations. We contrast the choices $\delta = (\Delta x)^2$, $\delta = (\Delta x)^3$, and $\delta = 10^{-8}$, for the example of the Wasserstein distance between geodesics, illustrated in Fig. 2, where the outer time step $\tau = 1$, (Δx) $\sim$ (Δt), and $\delta_3 = \delta$. Based on the order of accuracy of our Crank–Nicolson approximation of the PDE constraint, we expect that $\delta = (\Delta x)^2$ should give the optimal balance between accuracy and computational efficiency. (See Remark 3.)

In the plot on the left, we observe that for all choices of δ, the error between the numerical solution $\rho^{(l)}$ and the exact solution ρ^* is identical, with the error saturating after 10^5 iterations. Thus, all three choices of δ provide the same level of accuracy, and the best way to distinguish between them is to identify which choice of δ causes the stopping criteria (33 and 34) to be satisfied in the least number of excess iterations after 10^5. The behavior of two key stopping criteria is shown in the plot on the right— the PDE constraint $\|Au^{(l)} - b\|$ and the convergence monitor for the relative error of the dual variables $\|\phi^{(l)} - \phi^{(l-1)}\|/\|\phi^{(l)}\|$. Of the four stopping criteria we consider (PDE constraint and three convergence monitors), these two are the last to be satisfied in all of the numerical simulations contained in this manuscript, hence these determine when our method terminates its iterations.

For the case of $\delta = (\Delta x)^2$ (red lines), we indeed observe that the PDE constraint (solid line) satisfies its stopping criteria (dashed line) by 10^4 iterations and the dual variables (starred line) satisfy their stopping criteria of 10^{-5} by 10^5 iterations. On the other hand, for the cases of $\delta = (\Delta x)^3$ (blue lines) and $\delta = 10^{-8}$ (green lines), we see that while the dual variables (starred lines) have satisfied their stopping criteria of 10^{-5} by 10^4 iterations, the PDE constraints (solid lines) do not satisfy their stopping criteria (dashed lines) until later—it takes more than 10^5 iterations for $\delta = (\Delta x)^3$ and more than 10^7 iterations for $\delta = 10^{-8}$. This example shows that choosing δ without respecting the order of accuracy of the finite difference approximation in the PDE

Evolution of geodesic between translations of British Parliament
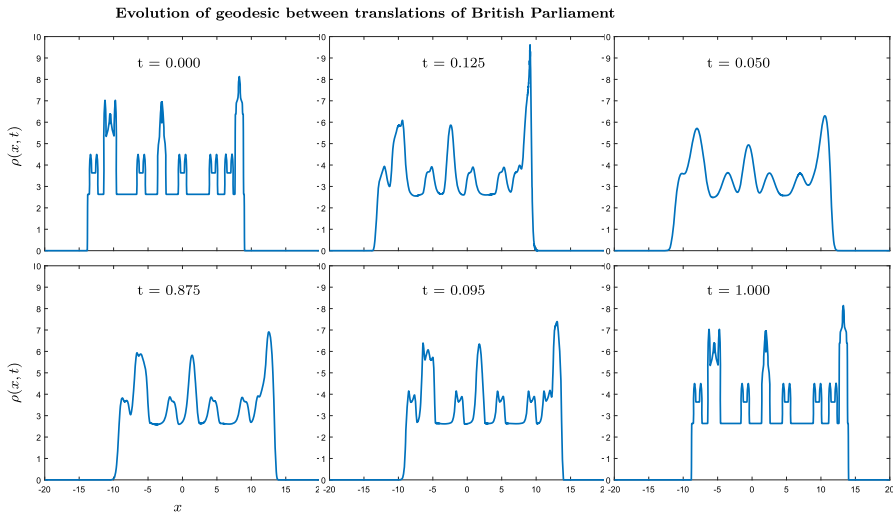
Fig. 4 Computation of the Wasserstein geodesic between two translations of British parliament on the domain, with $N_t = 40$ temporal grid points ($\Delta t = \frac{1}{40}$) and $N_x = 2000$ spatial grids ($\Delta x = \frac{40}{2000}$). Here, $\sigma = 0.1$, $\sigma\lambda = 0.99/\lambda_{max}(AA^t)$ and then $\lambda = 0.9727$, $\delta = 10^{-5}$, and $\delta_3 = 10^{-8}$

constraint, one wastes computational effort without improving the accuracy of the numerical solution.

Next, we compute Wasserstein geodesics between initial and target measures when neither are smooth nor strictly positive. In Fig. 4, we compute the geodesic between a profile of the British Parliament and its translation. We do not observe convergence to the exact geodesic, which would be a constant speed translation, and instead observe degradation of the parliamentary building at intermediate times, due to numerical smoothing. Similarly, in Fig. 5, we compute the geodesic between Pac-Man and a ghost, visualized as characteristic functions on sets in two dimensions. Again, we observe numerical smoothing around the edges of discontinuity. Both of these examples offer a numerical justification for the smoothness assumption we impose in our main convergence Theorem 3. In the absence of such smoothness, it appears that the method does not converge. Similar smoothness assumptions are required in the other numerical methods for Wasserstein geodesics for which rigorous convergence has been analyzed, including Monge Ampére-type methods [11,68].

4.2 Wasserstein Gradient Flows: One Dimension

In this and the next section, we consider several examples of Wasserstein gradient flows, including some which have appeared in previous numerical studies [3,29,49,99], to demonstrate the performance of our method for simulating solutions of nonlinear partial differential equations.

Evolution of geodesic between Pac-Man and a ghost
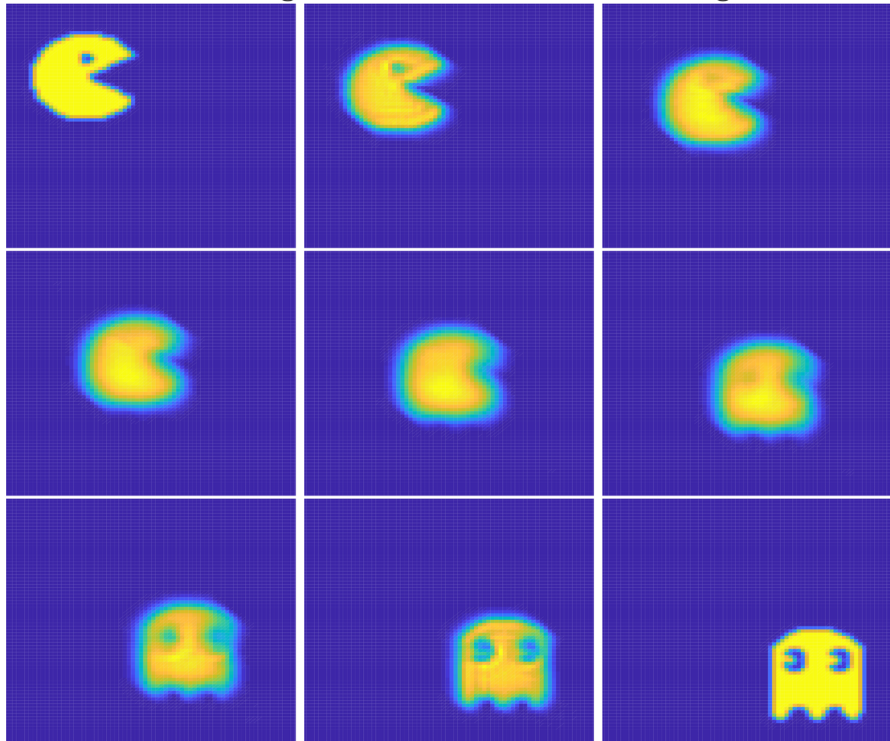
Fig. 5 Computation of the Wasserstein geodesic between Pac-Man and a ghost, with $N_t = 40$ temporal points ($\Delta t = \frac{1}{40}$) and $N_x = N_y = 120$ spatial grids ($\Delta x = \Delta y = 0.0458$). From left to right, up to down, the plots correspond to $t = 0$, $t = 0.15$, $t = 0.275$, $t = 0.4$, $t = 0.525$, $t = 0.65$, $t = 0.775$, $t = 0.9$, and $t = 1$. Here, $\lambda = 40$, $\sigma\lambda = 1.2/\lambda_{max}(AA^t)$ then $\sigma = 0.0036$, $\delta = 10^{-5}$, and $\delta_3 = 10^{-5}$

4.2.1 Porous Medium Equation

The porous medium equation

$$\partial_t \rho = \Delta \rho^m, \quad m > 1, \tag{55}$$

is the Wasserstein gradient flow of the energy (4), with $U(\rho) = \frac{1}{m-1}\rho^m$ and $V = W = 0$. A well-known family of exact solutions is given by Barenblatt profiles (c.f. [104]), which are densities of the form

$$\rho(x, t) = (t + t_0)^{-\frac{1}{m+1}}\left(C - \alpha\frac{m-1}{2m(m+1)}x^2(t+t_0)^{-\frac{2}{m+1}}\right)_+^{\frac{1}{m-1}}, \quad \text{for } C, t_0 > 0 \tag{56}$$

We now apply Algorithm 3 to simulate solutions of the $m = 2$ porous medium equation with Barenblatt initial data, $t_0 = 10^{-3}$ and $C = (3/16)^{1/3}$. Here, the Euler discretization (27) is used. In Fig. 6, we plot the evolution of the numerical solution

Fig. 6 Evolution of the solution $\rho(x, t)$ to the one-dimensional porous medium equation, with $m = 2$ on the domain $\Omega = [-1, 1]$. We choose $\tau = 0.5 \times 10^{-3}$, $\Delta x = 0.02$, $\Delta t = 0.1$, $\lambda = 0.2$, $\sigma\lambda = 1.1/\lambda_{max}(AA^t)$ then $\sigma = 0.1954$, $\delta = 10^{-5}$, and $\delta_3 = 10^{-5}$
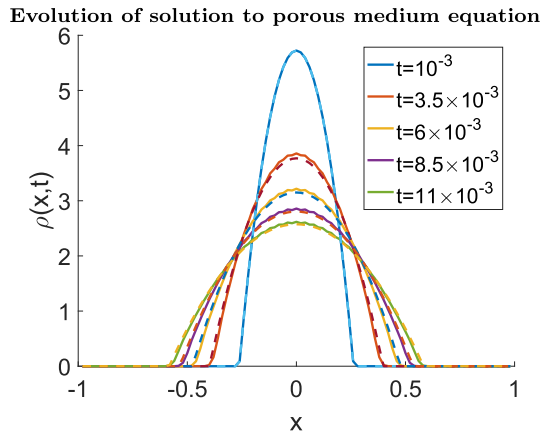
over time, and we observe good agreement with the exact solution of the form (56), which is displayed in dashed curve.

In Fig. 7, we analyze how the rate of convergence depends on the inner time step Δt, the spatial discretization Δx, and outer time step of the JKO scheme τ. We compute the error between the exact solution and the numerical solution in the ℓ^1 norm, i.e.,

$$\|\rho - \rho^*\|_{\ell^1} = \frac{1}{N_x(n+1)} \sum_{k=0}^{N_t} \sum_{j=1}^{N_x-1} |\rho_j(k\tau) - \rho_j^*(k\tau)|.$$

In the plot on the left of Fig. 7, we consider two fixed values of τ and examine how the error depends on N_t and $N_x = 10N_t$. In both cases, the error quickly saturates, indicating that the outer time step τ dominates the error. In the plot on the right, we fix $N_t = 20$ and $N_x = 200$ and consider how the error depends on τ. We observe slightly less than first-order convergence in τ for the classical JKO scheme ($\mathcal{E}^h = \mathcal{F}^h$) and higher-order convergence for the Crank–Nicolson inspired scheme ($\mathcal{E}^h = \mathcal{H}^h$). We believe these slower rates of convergence are due to the lower regularity of solutions to the porous medium equation with compactly supported initial data, which are merely Hölder continuous.

In Fig. 8, we consider the case of smooth, strictly positive initial data, given by a Gaussian with mean $\mu = 0$ and variance $\theta = 0.2$ (53), in which case solutions of the PDE remain smooth over time. On the left, we show the evolution of the solutions over time, and on the right, we illustrate that the classical JKO scheme indeed attains first-order accuracy, though the Crank–Nicolson inspired scheme is still less than second-order accurate.

### 4.2.2 Nonlinear Fokker–Planck Equation

We now consider a nonlinear variant of the Fokker–Planck equation,

$$\partial_t \rho = \nabla \cdot (\rho \nabla V) + \Delta \rho^m, \quad V : \mathbb{R}^d \to \mathbb{R}, \quad m > 1,$$
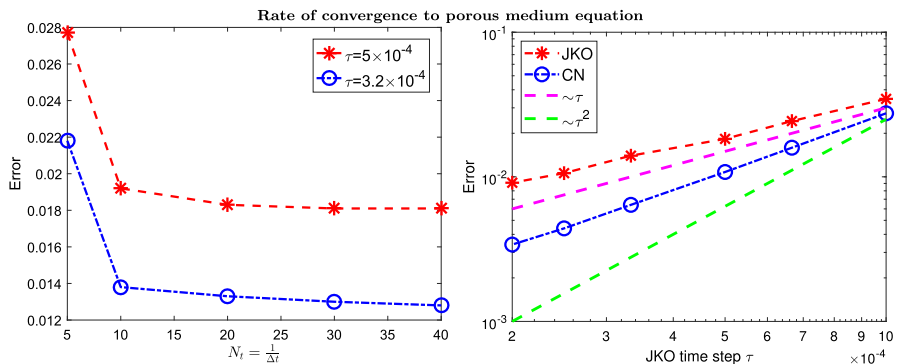
Fig. 7 Analysis of rate of convergence for a solution of porous medium equation, as in Fig. 6. Left: Convergence to exact solution for $N_x/N_t = 10$ for choices of τ. Right: Convergence to exact solution for $N_t = 20$ and $N_x = 200$ and various choices of τ, contrasting the traditional first-order JKO scheme with the new Crank–Nicolson inspired scheme
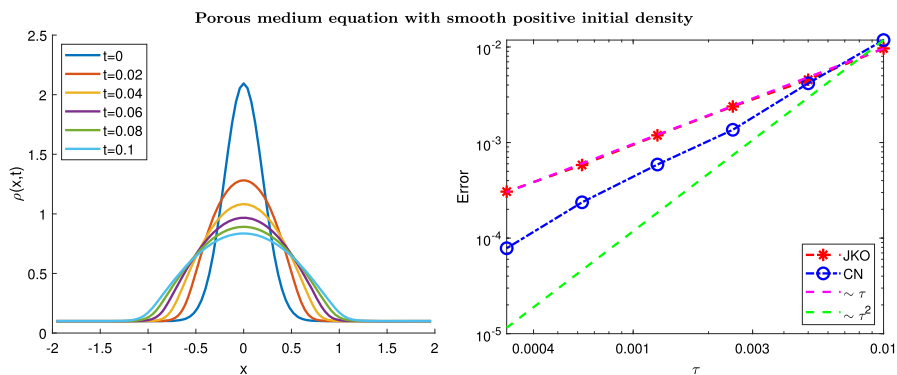

Fig. 8 Evolution and the rate of convergence for a solution of porous medium equation with smooth positive initial density. We choose $N_t = 10$, $N_x = 100$, $\sigma = 10$, $\lambda = 0.0148$. Left: Evolution of the solution $\rho(x, t)$ to the one-dimensional porous medium equation, with $m = 2$ on the domain $\Omega = [-2, 2]$ for $\tau = 0.005$. Right: The rate of convergence for various choices of τ, contrasting the traditional first-order JKO scheme with the new Crank–Nicolson inspired scheme. For each choice of τ in our computation of the higher-order method, we choose our stopping criteria $\epsilon = 10^{-4} * 2^{-0.01/\tau}$

inspired by the porous medium equation described in the previous section (55). When V is a confining drift potential, all solutions approach the unique steady state

$$\rho_\infty(x) = \left(C - \frac{m-1}{m} V(x) \right)_+^{\frac{1}{m-1}},$$

where $C > 0$ depends on the mass of the initial data, so that $\int \rho_\infty dx = \int \rho_0 dx$, see [44,51].

In Fig. 9, we simulate the evolution of solutions to the nonlinear Fokker–Planck equation with $V(x) = x^2$, $m = 2$, and initial data given by a Gaussian with mean $\mu = 0$ and variance $\theta = 0.2$ (53). On the left, we plot the evolution of the density
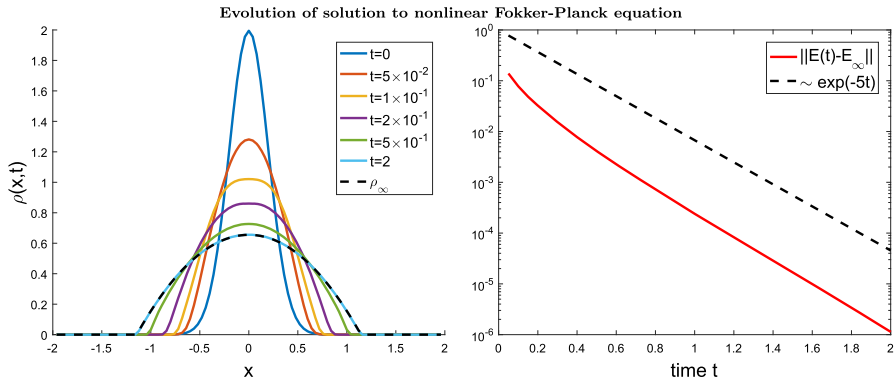
Fig. 9 Evolution of the solution $\rho(x,t)$ to the one-dimensional nonlinear Fokker–Planck equation, with $m=2$ and $V(x)=x^2$. We choose $\tau=0.05$, $\Delta x=0.04$, $\Delta t=0.1$, $\lambda=0.1641$, $\sigma=1$, $\delta=10^{-5}$, and $\delta_3=10^{-5}$. Left: evolution of density $\rho(x,t)$ toward equilibrium $\rho_\infty(x)$. Right: Rate of decay of corresponding energy with respect to time
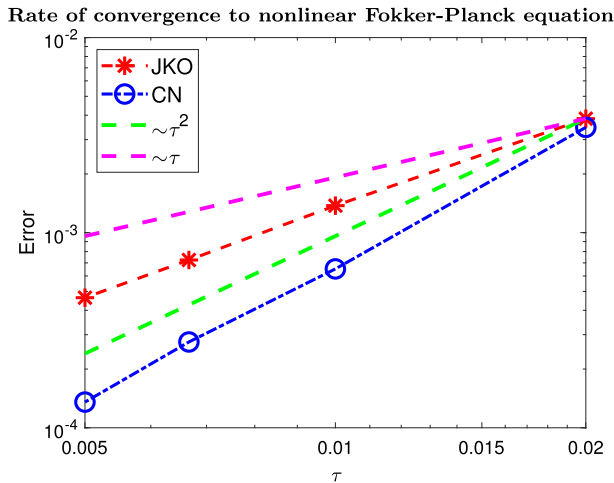

Fig. 10 Analysis of rate of convergence for a solution of the nonlinear Fokker–Planck equation, as in Fig. 9. We choose $\Delta t=0.1$, $\Delta x=0.04$ and consider the error (57) for various choices of τ, contrasting the traditional first-order JKO scheme with the new Crank–Nicolson inspired scheme

$\rho(x,t)$ toward the steady state $\rho_\infty(x)$. On the right, we compute the rate of decay of the corresponding energy (4) as a function of time, observing exponential decay as the solution approaches equilibrium. In this way, our method recovers analytic results on convergence to equilibrium of Carrillo, DiFrancesco, and Toscani [35,51].

In Fig. 10, we analyze how the rate of convergence depends on the outer time step τ of the scheme, for sufficiently small inner time step $\Delta t=0.1$ and spatial discretization $\Delta x=0.04$. We compute the error

$$e_\tau = \|\rho_\tau(x,t) - \rho_{\tau/2}(x,t)\|_{\ell^1} \tag{57}$$
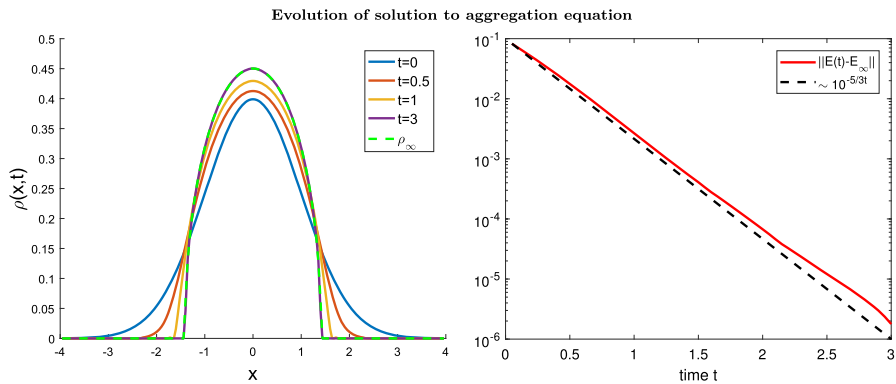
Fig. 11 Evolution of the solution $\rho(x, t)$ to the one-dimensional aggregation equation, with $W(x) = x^2/2 - \ln(|x|)$, $x \in [-4, 4]$. We choose $\tau = 0.05$, $\Delta x = 0.04$, $\Delta t = 0.05$, $\lambda = 0.01$, $\sigma\lambda = 0.99/\lambda_{max}(AA^t)$ then $\sigma = 18.8$, $\delta = 10^{-6}$, and $\delta_3 = 10^{-6}$. Left: evolution of density $\rho(x, t)$ toward equilibrium $\rho_\infty(x)$. Right: Rate of decay of corresponding energy with respect to time

We observe slightly faster than first-order convergence for the traditional JKO scheme ($\mathcal{E}^h = \mathcal{F}^h$) and higher-order convergence for the new Crank–Nicolson inspired scheme ($\mathcal{E}^h = \mathcal{H}^h$). We believe this improvement in the rate of convergence as compared to our previous example for the porous medium equation, Fig. 7, is due to the rapid convergence to the steady state ρ_∞.

4.2.3 Aggregation Equation

In this section, we consider a nonlocal partial differential equation of Wasserstein gradient flow type, known as the aggregation equation

$$\partial_t \rho = \nabla \cdot (\rho \nabla W * \rho), \quad W : \mathbb{R}^d \to \mathbb{R}. \tag{58}$$

In recent years, there has been significant interest in interaction kernels W that are repulsive at short length scales and attractive at longer distances, such as the kernel with logarithmic repulsion and quadratic attraction

$$W(x) = \frac{|x|^2}{2} - \ln(|x|). \tag{59}$$

For this particular choice of W, there exists a unique equilibrium profile [38], given by

$$\rho_\infty(x) = \frac{1}{\pi}\sqrt{(2 - x^2)_+}.$$

In Fig. 11, we simulate the solution to the aggregation equation with Gaussian initial data (53) with mean $\mu = 0$ and variance $\theta = 1$, analyzing convergence to equilibrium. On the left, we plot the evolution of the density $\rho(x, t)$ at varying times, observing
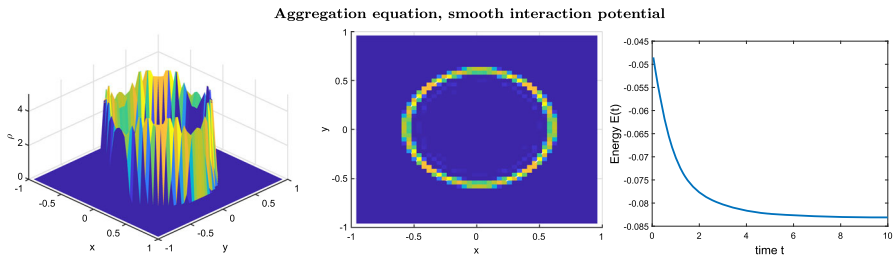
Fig. 12 We compute the steady state of a solution to the two-dimensional aggregation equation with interaction potential $W(x) = |x|^4/4 - |x|^2/2$, which is a Dirac ring of radius 0.5, centered at the origin. The computational domain is [-1,1]×[-1,1]. We choose $\tau = 0.05$, $\Delta x = \Delta y = 0.04$, $\Delta t = 0.1$, $\lambda = 20$, $\sigma = 0.0052$, and $\epsilon_1 = \epsilon_2 = 10^{-6}$. The steady state shown is the solution at time $t = 10$. Left: side view of equilibrium. Center: bird's eye view of equilibrium. Right: rate of decay of energy as solution approaches equilibrium

convergence to the equilibrium profile $\rho_\infty(x)$. On the right, we compute the rate of the decay of the energy as a function of time, observing exponential decay as obtained by Carrillo, Ferreira, and Precioso [38] with a slightly slower numerical rate.

As the interaction potential W defined in Eq. (59) is not continuous, we make the following modifications to our discretization of the JKO scheme. To avoid evaluation of $W(x)$ at $x = 0$, we set $W(0)$ to equal the average value of W on the cell of width $2h$ centered at 0, i.e., $W(0) = \frac{1}{2h}\int_{-h}^{h} W(x)\mathrm{d}x$, where we apply Gauss-Legendre quadrature rule with four grid points to evaluate the integral. In addition to modifying the interaction kernel in this way, we also introduce an artificial diffusion term of the form $\epsilon\partial_x(\rho\partial_x\rho)$ with $\epsilon = 1.6 \times (\Delta x)^2$ to the right-hand side of (58), to avoid the possible overshoot at the boundary. (See also [29] for a similar treatment.)

## 4.3 Wasserstein Gradient Flows: Two Dimensions

In the following, we consider a few gradient flows in two dimensions. Here, the constraint relaxation parameters are always chosen as $\delta = \delta_3 = 10^{-6}$.

4.3.1 Aggregation Equation

We now continue our study of the aggregation equation (58) with repulsive–attractive interaction potentials in two dimensions, with interaction kernels of the form

$$W(x) = \frac{|x|^a}{a} - \frac{|x|^b}{b}, \quad a > b \geq 0, \tag{60}$$

using the convention that $\frac{|x|^0}{0} = \ln(|x|)$. It is well known that the repulsion near the origin of the potential determines the dimension of the support of the steady state measure, see [4,34]. In the following simulations, we take the initial data to be a gaussian (53) with mean $\mu = 0$ and variance $\theta = 0.25$.
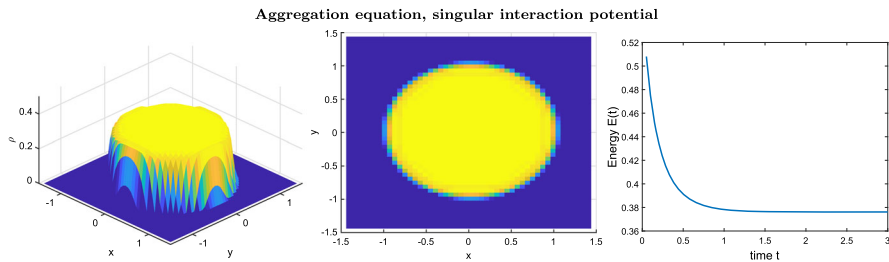
Fig. 13 We compute steady state of a solution to the two-dimensional aggregation equation with interaction potential $W(x) = |x|^2/2 - \ln(|x|)$, which is the characteristic function on a disk of radius 1, centered at the origin. The computational domain is [-1.5,1.5]×[-1.5,1.5]. We choose $\tau = 0.05$, $\Delta x = \Delta y = 0.06$, $\Delta t = 0.05$, $\lambda = 50$, and $\sigma = 0.0037$. The steady state is plotted at time t=3. Left: side view of equilibrium. Center: bird's eye view of equilibrium. Right: rate of decay of energy as solution approaches equilibrium

In Fig. 12, we simulate the evolution of solutions to the aggregation equation, with $a = 4$ and $b = 2$ in the interaction potential W, defined in Eq. (60). We observe that the solution concentrates on a Dirac ring with radius 0.5 centered at the origin, recovering analytical results on the existence of a stable Dirac ring equilibrium for these values of a and b [5,13].

In Fig. 13, we simulate the evolution of solutions to the aggregation equation, with $a = 2$ and $b = 0$. We observe that the solution converges to a characteristic function on the disk of radius 1, centered at the origin, recovering analytic results on solutions of the aggregation equation with Newtonian repulsion [14,34,65]. We follow the same strategy described in Sect. 4.2.3 with $\epsilon = 1.6 \times (\Delta x^2 + \Delta y^2)$ to overcome the singularity of the interaction potential at $x = 0$ and potential overshooting.
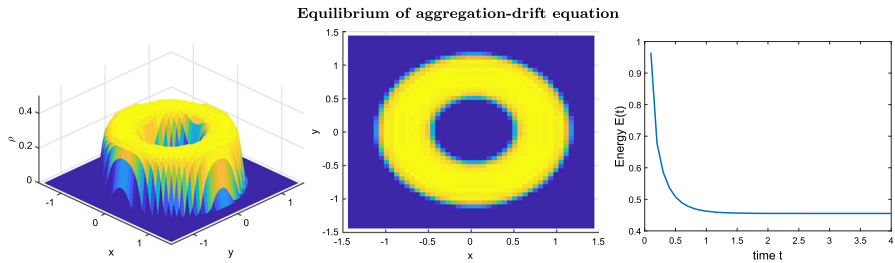
4.3.2 Aggregation Drift Equation

Next, we compute solutions of aggregation-drift equations

$$\partial_t \rho = \nabla \cdot (\rho \nabla W * \rho) + \nabla \cdot (\rho \nabla V),$$

where $W(x) = |x|^2/2 - \ln(|x|)$ and $V(x) = -\frac{\alpha}{\beta} \ln(|x|)$. As shown in several analytical and numerical results [29,42,53], the steady state is a characteristic function on a torus or "milling profile", with inner and outer radius given by

$$R_i = \sqrt{\frac{\alpha}{\beta}}, \quad R_o = \sqrt{\frac{\alpha}{\beta} + 1}.$$

In Fig. 14, we simulate the long time behavior of a solution of the aggregation-drift equation with $\alpha = 1$ and $\beta = 4$ and Gaussian initial data (53), $\mu = 0$, $\theta = 0.25$, as well as the rate of the decay of the entropy as the solution converges to equilibrium. In Fig. (15), we plot the evolution of the density from a nonradially symmetric initial data, given by five Gaussians to the same equilibrium profile. We follow the same strategy described in Sect. 4.2.3 to overcome the singularity of the interaction potential at $x = 0$

Fig. 14 We compute steady state of a solution to the two-dimensional aggregation-drift equation with interaction potential $W(x) = |x|^2/2 - \ln(|x|)$ and drift potential $V(x) = -(1/4)\ln(|x|)$, which is the characteristic function on a torus, centered at the origin. The computational domain is $[-1.5,1.5]\times[-1.5,1.5]$. We choose $\tau = 0.1$, $\Delta x = \Delta y = 0.06$, $\Delta t = 0.05$, $\lambda = 40$, and $\sigma = 0.0046$. The steady state is the solution at time t=4. Left: side view of equilibrium. Center: bird's eye view of equilibrium. Right: rate of decay of energy as solution approaches equilibrium
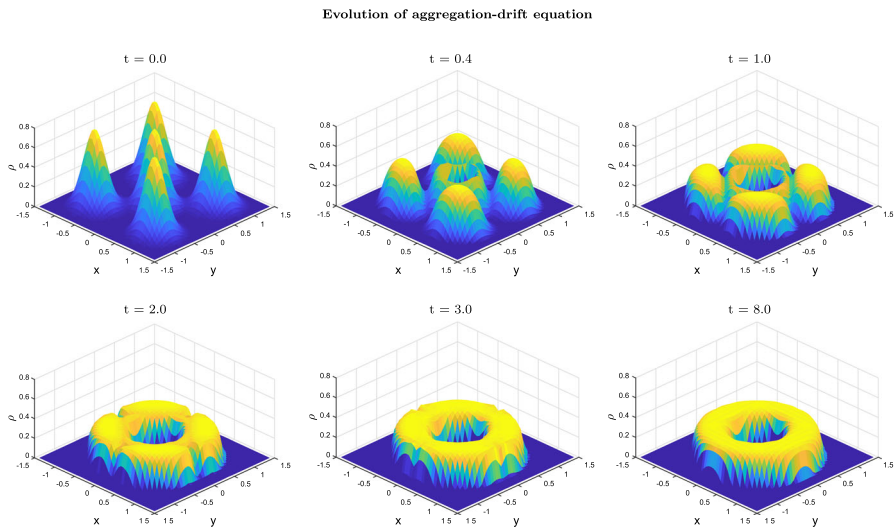
Fig. 15 Evolution of the solution $\rho(x, y, t)$ to the two-dimensional aggregation-drift equation, with $W(x) = x^2/2 - \ln(|x|)$ and $V(x) = -(1/4)\ln(|x|)$. The computational domain is $[-1.5,1.5]\times[-1.5,1.5]$. We choose $\tau = 0.2$, $\Delta x = \Delta y = 0.06$, $\Delta t = 0.1$, $\lambda = 10$, and $\sigma = 0.0244$. We observe convergence to the characteristic function on a torus centered at the origin

and potential overshooting ($\epsilon = 2\times(\Delta x^2 + \Delta y^2)$ in Fig. 14 and $\epsilon = 2.6\times(\Delta x^2 + \Delta y^2)$ in Fig. 15.)

4.3.3 Aggregation–Diffusion Equation

We close by simulating several examples of aggregation–diffusion equations

$$\partial_t \rho = \nabla \cdot (\rho \nabla W * \rho) + \nu \Delta \rho^m, \quad W : \mathbb{R}^d \to \mathbb{R}, \quad m \geq 1. \tag{61}$$
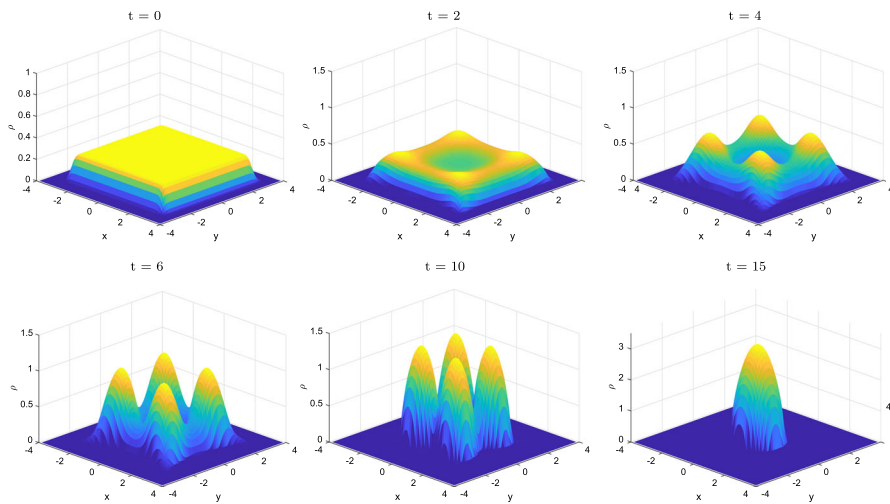
Fig. 16 Evolution of solution $\rho(x, y, t)$ to the two-dimensional aggregation–diffusion equation, with $W(x) = -e^{-|x|^2}/\pi$, $\nu = 0.1$, and $m = 3$ on the domain $\Omega = [-4, 4] \times [-4, 4]$. We choose $\tau = 0.5$, $\Delta x = \Delta y = 0.1$, $\Delta t = 0.1$, $\sigma = 0.1144$, and $\lambda = 0.5$. The total iteration number for 40 JKO time steps is 197852. We observe convergence to the a single bump centered at the origin

In recent years, there has been significant activity studying equations of this form, both analytically and numerically. When the interaction kernel W is attractive, the competition between the nonlocal aggregation $\nabla \cdot (\rho \nabla W * \rho)$ and nonlinear diffusion $\nu \Delta \rho^m$ causes solutions to blow up in certain regimes and exist globally in time in others, see for example [18,19,25,26,41] and the survey [32]. With fixed m, and in the presence of nonlocal interaction, the equation has a unique steady state which is radially decreasing up to a translation [15,40].

In Fig. 16, we simulate a solution of the aggregation–diffusion equation with $W(x) = -e^{-|x|^2}/\pi$, $\nu = 0.1$, and $m = 3$, and initial data given by a rescaled characteristic function on the square,

$$\rho_0(x, y) = \frac{1}{4}\chi_{[-3,3] \times [-3,3]}(x, y),$$

Diffusion dominates both the short and long ranges, and the medium range aggregation leads to the formation of four bumps, which ultimately approach a single bump equilibrium. (See also [29].)

In Fig. 17, we simulate solutions of the Keller–Segel equation, which is an aggregation–diffusion equation (61) with a Newtonian interaction kernel, i.e., $W(x) = \frac{1}{2\pi}\ln(|x|)$ in two dimensions for $\nu = 1$ and both $m = 1$ and $m = 2$, illustrating the role of the diffusion exponent in blowup or global existence of solutions. We choose the initial data to be given by a rescaled gaussian, obtained by multiplying equation (53) by a mass $M = 9\pi$, with mean $\mu = 0$ and variance $\theta = 0.5$. On the left, we
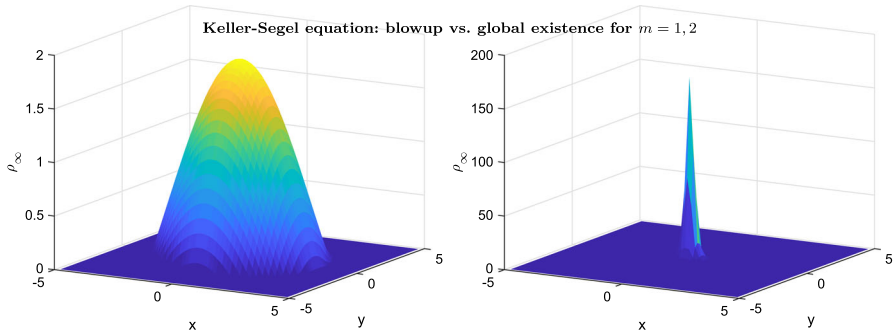
Fig. 17 Plot of solution $\rho(x, y, t)$ to the two-dimensional Keller–Segel equation at $t = 2$. Left: When $m = 2$, the solution approaches a bounded, continuous equilibrium profile. Here, the computational domain is [-5,5]×[-5,5]. Right: When $m = 1$, the solution blows up, becoming sharply peaked. The computational domain here is [-2,2]×[-2,2]. For both we choose $\tau = 0.05$, $\Delta x = \Delta y = 0.067$, $\Delta t = 0.1$, $\lambda = 0.5$, $\sigma = 0.042$
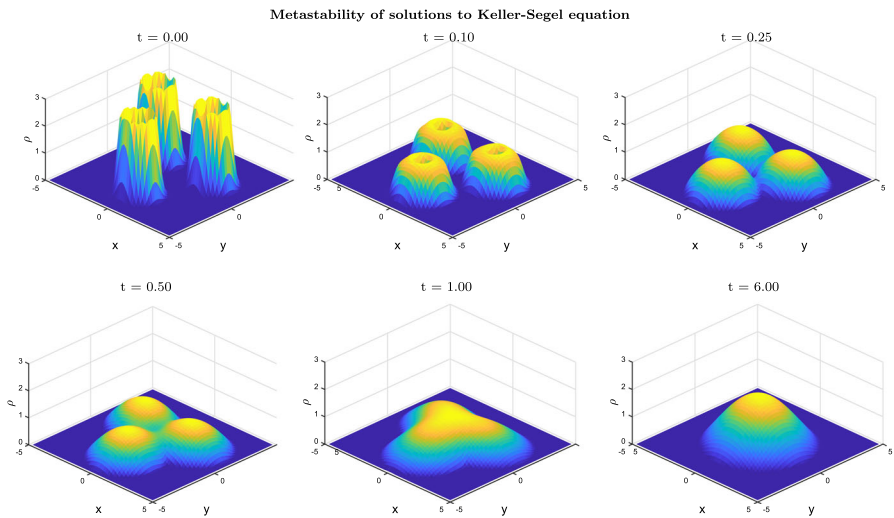
Fig. 18 The evolution of $\rho(x, y, t)$ for the Keller–Segel equation with $U(x) = x^2$. Here, $\Delta t = 0.1$, $h_x = h_y = 0.167$, $\tau = 0.05$

take $m = 2$ and simulate the steady state of the Keller–Segel equation, which is a single bump. On the right, we simulate the long-time behavior of solutions for $m = 1$, in which case we are in the blow up regime. Indeed, at time $t = 2$, we observe the formation of a blowup profile, with the solution becoming sharply peaked at the origin.

In Fig. 18, we again simulate solutions of the Keller–Segel equation with $m = 2$, but in this case we take the initial data to be given by three localized bumps (Gaussian rings, i.e., the radial cut of the ring is a Gaussian with a center on the circle.) We observe a two-stage evolution in which the each of the bumps converges to a localized quasi-stationary state, and then interact and merge into one single bump in the long time limit. This is a manifestation of the typical metastability phenomena, which is likely present in the majority of the diffusion dominated Keller–Segel models [24,29,32].

A Further Details of Numerical Implementation

In this section, we provide explicit formulas for the matrix $\tilde{A}$ and the vectors u and $\tilde{b}$ introduced in Problem 3(b) in Sect. 2.3, which play a key role in the implementation of Algorithms 1, 2, and 3. For simplicity, we consider the case of one space dimension, and the discretization takes the form (28). The constructions of A and b in Problem 3(a) are very similar except a slightly different treatment of ρ at final time. From now on, for simplicity of notation, we will drop the tildes for the matrix $\tilde{A}$ and vector $\tilde{b}$.

Define $N = (N_x + 1)(N_t + 1)$. Let $\otimes$ denote the Kronecker tensor product, I_{N_x+1} the identity matrix of size $N_x + 1$, and $(\mathbf{x})_M$ the column vector in $\mathbb{R}^M$ with all components equal to x. Then we define

$$u = \left[(\text{æ}_{.,k})_{k=0}^{N_t};\ (\mathbf{m}_{.,k})_{k=0}^{N_t} \right] \in \mathbb{R}^N, \quad \text{æ}_{.,k} = (\rho_{j,k})_{j=0}^{N_x}, \quad \mathbf{m}_{.,k} = (m_{j,k})_{j=0}^{N_x}$$

and the matrix $A \in \mathbb{R}^{M \times 2N}$ takes the form

$$A = \left[\begin{array}{c|c} A_\rho & A_m \\ \hline A_{\text{mass}} & 0 \end{array} \right].$$

Here, $A_\rho \in \mathbb{R}^{N \times N}$ reads

$$A_\rho = D_t^{(1)} \otimes I_x^{(1)} + D_t^{(2)} \otimes I_x^{(1)} := A_\rho^{(1)} + A_\rho^{(2)},$$

where $D_t^{(1)}, D_t^{(2)} \in \mathbb{R}^{(N_t+1) \times (N_t+1)}$, and $I_x^{(1)} \in \mathbb{R}^{(N_x+1) \times (N_x+1)}$ are

$$I_x^{(1)} = \begin{bmatrix} 0 & \\ & I_{N_x-1} \\ & & 0 \end{bmatrix}, \quad D_t^{(1)} = \begin{bmatrix} 0 & \\ -1 & 1 \\ & \ddots & \ddots \\ & & -1 & 1 \end{bmatrix}, \quad D_t^{(2)} = \begin{bmatrix} 1 & \\ & \mathbf{0} \end{bmatrix}.$$

Here, $D_t^{(1)}$ and $D_t^{(2)}$ correspond to the temporal discretization and initial condition for ρ. Likewise,

$$A_m \in \mathbb{R}^{N \times N} = B_t^{(1)} \otimes D_x^{(1)} + I_{N_t+1} \otimes D_x^{(2)} := A_m^{(1)} + A_m^{(2)},$$

where $D_x^{(1)}, D_x^{(2)} \in \mathbb{R}^{(N_x+1)}$, and $B_t^{(1)} \in \mathbb{R}^{(N_t+1) \times (N_t+1)}$:

$$D_x^{(1)} = \frac{\Delta t}{4\Delta x}\begin{bmatrix} 0 & & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & & 0 \end{bmatrix}, \quad D_x^{(2)} = \begin{bmatrix} 1 & & \\ & \mathbf{0} & \\ & & 1 \end{bmatrix}, \quad B_t^{(1)} = \begin{bmatrix} 0 & & & \\ 1 & 1 & & \\ & \ddots & \ddots & \\ & & 1 & 1 \end{bmatrix}.$$

For mass conservation, let $S_\rho = (\mathbf{x})_{N_x+1}^t$, then $A_{\text{mass}} = I_{N_t+1} \otimes S_\rho$. In sum, different A_i can be written as

$$A_1 = \begin{bmatrix} A_\rho^{(1)} & A_m^{(1)} \\ 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & A_m^{(2)} \\ 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0 \\ A_{\text{mass}} & 0 \end{bmatrix}, \quad A_4(+A_5) = \begin{bmatrix} A_\rho^{(2)} & 0 \\ 0 & 0 \end{bmatrix}.$$

Accordingly, $b \in \mathbb{R}^{N+N_t+1}$ collects all the initial conditions for ρ and boundary conditions for m. More specifically, it writes

$$b = [(0; (\rho_{j,0})_{j=1}^{N_x-1}; 0); \mathbf{0}_{(N_t-1)(N_x+1)}; \mathbf{0}_{N_x+1}; \mathbf{0}_{N_t+1}]$$
$$+ [(m_{0,0}; \mathbf{0}; m_{N_x,0}); \cdots ; (m_{0,N_t}; \mathbf{0}; m_{N_x,N_t}); \mathbf{0}_{N_t+1}]$$
$$+ [\mathbf{0}_N; \mathbf{1}_{N_t+1}]$$
$$:= b_4 + (b_5) + b_2 + b_3$$

and $b_1 = 0$.

References

1. L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of bounded variation and free discontinuity problems*, vol. 254, Clarendon Press Oxford, 2000.
2. L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient flows: in metric spaces and in the space of probability measures*, Springer Science & Business Media, 2008.
3. R. BAILO, J. A. CARRILLO, AND J. HU, *Fully discrete positivity-preserving and energy-dissipating schemes for aggregation–diffusion equations with a gradient-flow structure*, Commun. Math. Sci., 18 (2020), pp. 1259–1303.
4. D. BALAGUÉ, J. A. CARRILLO, T. LAURENT, AND G. RAOUL, *Dimensionality of local minimizers of the interaction energy*, Arch. Ration. Mech. Anal., 209 (2013), pp. 1055–1088.
5. D. BALAGUÉ, J. A. CARRILLO, T. LAURENT, AND G. RAOUL, *Nonlocal interactions by repulsive–attractive potentials: radial ins/stability*, Phys. D, 260 (2013), pp. 5–25.
6. A. B. T. BARBARO, J. A. CAÑIZO, J. A. CARRILLO, AND P. DEGOND, *Phase transitions in a kinetic flocking model of Cucker–Smale type*, Multiscale Model. Simul., 14 (2016), pp. 1063–1088.
7. J.-D. BENAMOU AND Y. BRENIER, *A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.
8. J.-D. BENAMOU, G. CARLIER, M. CUTURI, L. NENNA, AND G. PEYRE, *Iterative Bregman projections for regularized transportation problems*, SIAM. J. Sci. Comput., 37 (2015), pp. A111–A1138.

9. J.- D. BENAMOU, G. CARLIER, AND M. LABORDE, *An augmented Lagrangian approach to Wasserstein gradient flows and applications*, ESAIM: PROCEEDINGS AND SURVEYS, 54 (2016), pp. 1–17.

10. J.- D. BENAMOU, G. CARLIER, Q. MÉRIGOT, AND E. OUDET, *Discretization of functionals involving the Monge–Ampère operator*, Numer. Math., 134 (2016), pp. 611–636.

11. J.- D. BENAMOU, B. FROESE, AND A. OBERMAN, *Numerical solution of the optimal transportation problem using the Monge–Ampère equation*, J. Comput. Phys., 260 (2014), pp. 107–126.

12. D. BENEDETTO, E. CAGLIOTI, AND M. PULVIRENTI, *A kinetic equation for granular media*, RAIRO Modél. Math. Anal. Numér., 31 (1997), pp. 615–641.

13. A. L. BERTOZZI, T. KOLOKOLNIKOV, H. SUN, D. UMINSKY, AND J. VON BRECHT, *Ring patterns and their bifurcations in a nonlocal model of biological swarms*, Commun. Math. Sci., 13 (2015), pp. 955–985.

14. A. L. BERTOZZI, T. LAURENT, AND F. LÉGER, *Aggregation and spreading via the Newtonian potential: the dynamics of patch solutions*, Math. Models Methods Appl. Sci., 22 (2012), pp. 1140005, 39.

15. S. BIAN AND J.- G. LIU, *Dynamic and steady states for multi-dimensional Keller–Segel model with diffusion exponent $m > 0$*, Comm. Math. Phys., 323 (2013), pp. 1017–1070.

16. A. BLANCHET, V. CALVEZ, AND J. A. CARRILLO, *Convergence of the mass-transport steepest descent scheme for the subcritical Patlak–Keller–Segel model*, SIAM J. Numer. Anal., 46 (2008), pp. 691–721.

17. A. BLANCHET, E. A. CARLEN, AND J. A. CARRILLO, *Functional inequalities, thick tails and asymptotics for the critical mass Patlak–Keller–Segel model*, Journal of Functional Analysis, 262 (2012), pp. 2142–2230.

18. A. BLANCHET, J. A. CARRILLO, AND P. LAURENÇOT, *Critical mass for a Patlak–Keller–Segel model with degenerate diffusion in higher dimensions*, Calc. Var. Partial Differential Equations, 35 (2009), pp. 133–168.

19. A. BLANCHET, J. DOLBEAULT, AND B. PERTHAME, *Two dimensional Keller–Segel model in $\mathbb{RR}^2$: optimal critical mass and qualitative properties of the solution*, Electron. J. Differential Equations, 2006 (2006), pp. 1–33 (electronic).

20. H. BREZIS, *Functional analysis, Sobolev spaces and partial differential equations*, Springer Science & Business Media, 2010.

21. L. BRICEÑO- ARIAS, D. KALISE, Z. KOBEISSI, M. LAURIÈRE, A. M. GONZALEZ, AND F. J. SILVA, *On the implementation of a primal-dual algorithm for second order time-dependent mean field games with local couplings*, ESAIM: Proceedings and Surveys, 65 (2019), pp. 330–348.

22. L. M. BRICENO- ARIAS, D. KALISE, AND F. J. SILVA, *Proximal methods for stationary mean field games with local couplings*, SIAM Journal on Control and Optimization, 56 (2018), pp. 801–836.

23. A. BURCHARD, R. CHOKSI, AND I. TOPALOGLU, *Nonlocal shape optimization via interactions of attractive and repulsive potentials*, Indiana Univ. Math. J., 67 (2018), pp. 375–395.

24. M. BURGER, R. FETECAU, AND Y. HUANG, *Stationary states and asymptotic behavior of aggregation models with nonlinear local repulsion*, SIAM J. Appl. Dyn. Syst., 13 (2014), pp. 397–424.

25. V. CALVEZ, J. A. CARRILLO, AND F. HOFFMANN, *Equilibria of homogeneous functionals in the fair-competition regime*, Nonlinear Anal., 159 (2017), pp. 85–128.

26. V. CALVEZ, J. A. CARRILLO, AND F. HOFFMANN, *The geometry of diffusing and self-attracting particles in a one-dimensional fair-competition regime*, 2186 (2017), pp. 1–71.

27. M. CAMPOS- PINTO, J. A. CARRILLO, F. CHARLES, AND Y.- P. CHOI, *Convergence of a linearly transformed particle method for aggregation equations*, to appear in Numer. Math., (2018).

28. G. CARLIER, V. DUVAL, G. PEYRÉ, AND B. SCHMITZER, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM Journal on Mathematical Analysis, 49 (2017), pp. 1385–1418.

29. J. A. CARRILLO, A. CHERTOCK, AND Y. HUANG, *A finite-volume method for nonlinear nonlocal equations with a gradient flow structure*, Commun. Comput. Phys., 17 (2015), pp. 233–258.

30. J. A. CARRILLO, Y.- P. CHOI, AND M. HAURAY, *The derivation of swarming models: mean-field limit and Wasserstein distances*, in Collective Dynamics from Bacteria to Crowds: An Excursion Through Modeling, Analysis and Simulation, vol. 553 of CISM Courses and Lect., Springer Vienna, 2014, pp. 1–46.

31. J. A. CARRILLO, K. CRAIG, AND F. S. PATACCHINI, *A blob method for diffusion*, Calculus of Variations and Partial Differential Equations, 58 (2019), pp. 1–53.

32. J. A. CARRILLO, K. CRAIG, AND Y. YAO, *Aggregation–diffusion equations: dynamics, asymptotics, and singular limits*, (2019), pp. 65–108.

33. J. A. CARRILLO, M. G. DELGADINO, L. DESVILLETTES, AND J. WU, *The landau equation as a gradient flow*, arXiv preprint arXiv:2007.08591, (2020).

34. J. A. CARRILLO, M. G. DELGADINO, AND A. MELLET, *Regularity of local minimizers of the interaction energy via obstacle problems*, Comm. Math. Phys., 343 (2016), pp. 747–781.

35. J. A. CARRILLO, M. DI FRANCESCO, AND G. TOSCANI, *Strict contractivity of the 2-Wasserstein distance for the porous medium equation by mass-centering*, Proc. Amer. Math. Soc., 135 (2007), pp. 353–363.

36. J. A. CARRILLO, M. DIFRANCESCO, A. FIGALLI, T. LAURENT, AND D. SLEPCEV, *Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations*, Duke Math. J., 156 (2011), pp. 229–271.

37. J. A. CARRILLO, B. DUERING, D. MATTHES, AND D. S. MCCORMICK, *A Lagrangian scheme for the solution of nonlinear diffusion equations using moving simplex meshes*, to appear in J. Sci. Comp., (2018).

38. J. A. CARRILLO, L. C. F. FERREIRA, AND J. C. PRECIOSO, *A mass-transportation approach to a one dimensional fluid mechanics model with nonlocal velocity*, Adv. Math., 231 (2012), pp. 306–327.

39. J. A. CARRILLO, M. FORNASIER, G. TOSCANI, AND F. VECIL, *Particle, kinetic, and hydrodynamic models of swarming*, Modeling and Simulation in Science, Engineering and Technology, (2010), pp. 297–336.

40. J. A. CARRILLO, S. HITTMEIR, B. VOLZONE, AND Y. YAO, *Nonlinear aggregation–diffusion equations: radial symmetry and long time asymptotics*, Invent. Math., 218 (2019), pp. 889–977.

41. J. A. CARRILLO, F. HOFFMANN, E. MAININI, AND B. VOLZONE, *Ground states in the diffusion-dominated regime*, Calc. Var. Partial Differ. Equ., 57 (2018), p. 127.

42. J. A. CARRILLO, Y. HUANG, AND S. MARTIN, *Explicit flock solutions for Quasi–Morse potentials*, European J. Appl. Math., 25 (2014), pp. 553–578.

43. J. A. CARRILLO, Y. HUANG, F. S. PATACCHINI, AND G. WOLANSKY, *Numerical study of a particle method for gradient flows*, Kinet. Relat. Models, 10 (2017), pp. 613–641.

44. J. A. CARRILLO, A. JÜNGEL, P. A. MARKOWICH, G. TOSCANI, AND A. UNTERREITER, *Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities*, Monatsh. Math., 133 (2001), pp. 1–82.

45. J. A. CARRILLO, R. MCCANN, AND C. VILLANI, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Revista Matematica Iberoamericana, 19 (2003), pp. 971–1018.

46. J. A. CARRILLO, R. J. MCCANN, AND C. VILLANI, *Contractions in the 2-Wasserstein length space and thermalization of granular media*, Arch. Ration. Mech. Anal., 179 (2006), pp. 217–263.

47. J. A. CARRILLO AND J. MOLL, *Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms*, SIAM J. Sci. Comput., 31 (2009), pp. 4305–4329.

48. J. A. CARRILLO, F. S. PATACCHINI, P. STERNBERG, AND G. WOLANSKY, *Convergence of a particle method for diffusive gradient flows in one dimension*, SIAM J. Math. Anal., 48 (2016), pp. 3708–3741.

49. J. A. CARRILLO, H. RANETBAUER, AND M. WOLFRAM, *Numerical simulation of nonlinear continuity equations by evolving diffeomorphisms*, J. Comput. Phys., 326 (2016), pp. 186–202.

50. J.- A. CARRILLO AND F. SANTAMBROGIO, *l^∞ estimates for the JKO scheme in parabolic–elliptic Keller–Segel systems*, Quarterly of Applied Mathematics, 76 (2018), pp. 515–530.

51. J. A. CARRILLO AND G. TOSCANI, *Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity*, Indiana Univ. Math. J., 49 (2000), pp. 113–142.

52. A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with application to imaging*, J. Math. Imaging Vis., 40 (2011), pp. 120–145.

53. Y. CHEN AND T. KOLOKOLNIKOV, *A minimal model of predator–swarm interactions*, Journal of The Royal Society Interface, 11 (2014).

54. A. CHERTOCK AND A. KURGANOV, *A second-order positivity preserving central-upwind scheme for chemotaxis and haptotaxis models*, Numer. Math., 111 (2008), pp. 169–205.

55. L. CHIZAT, G. PEYRÉ, B. SCHMITZER, AND F.- X. VIALARD, *Scaling algorithms for unbalanced optimal transport problems*, Math. Comp., 87 (2018), pp. 2563–2609.

56. K. CRAIG, *Nonconvex gradient flow in the Wasserstein metric and applications to constrained non-local interactions*, Proceedings of the London Mathematical Society, 114 (2017), pp. 60–102.
57. K. CRAIG AND A. BERTOZZI, *A blob method for the aggregation equation*, Math. Comp., 85 (2016), pp. 1681–1717.
58. K. CRAIG, I. KIM, AND Y. YAO, *Congested aggregation via newtonian interaction*, Arch. Ration. Mech. Anal., to appear (2017).
59. K. CRAIG AND I. TOPALOGLU, *Convergence of regularized nonlocal interaction energies*, SIAM Journal on Mathematical Analysis, 48 (2016), pp. 34–60.
60. K. CRAIG AND I. TOPALOGLU, *Aggregation-diffusion to constrained interaction: minimizers & gradient flows in the slow diffusion limit*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 37 (2020), pp. 239–279.
61. M. CUTURI, *Sinkhorn distances: Lightspeed computation of optimal transport*, in Advances in neural information processing systems, 2013, pp. 2292–2300.
62. D. DAVIS AND W. YIN, *A three-operator splitting scheme and its optimization applications*, Set-valued and variational analysis, 25 (2017), pp. 829–858.
63. G. DE PHILIPPIS, A. R. MÉSZÁROS, F. SANTAMBROGIO, AND B. VELICHKOV, *BV estimates in optimal transportation and applications*, Archive for Rational Mechanics and Analysis, 219 (2016), pp. 829–860.
64. M. ERBAR, *A gradient flow approach to the Boltzmann equation*, arXiv preprint arXiv:1603.00540, (2016).
65. R. C. FETECAU, Y. HUANG, AND T. KOLOKOLNIKOV, *Swarm dynamics and equilibria for a nonlocal aggregation model*, Nonlinearity, 24 (2011), pp. 2681–2716.
66. F. FILBET, *A finite volume scheme for the Patlak–Keller–Segel chemotaxis model*, Numer. Math., 104 (2006), pp. 457–488.
67. R. L. FRANK AND E. H. LIEB, *A" liquid-solid" phase transition in a simple model for swarming, based on the" no flat-spots" theorem for subharmonic functions*, Indiana University Mathematics Journal, (2018), pp. 1547–1569.
68. W. GANGBO AND R. MCCANN, *The geometry of optimal transportation*, Acta. Math., 177 (1996), pp. 113–161.
69. L. GOSSE AND G. TOSCANI, *Lagrangian numerical approximations to one-dimensional convolution–diffusion equations*, SIAM J. Sci. Comput., 28 (2006), pp. 1203–1227.
70. B. F. HAMFELDT, *Viscosity subsolutions of the second boundary value problem for the Monge–Ampére equation*, arXiv preprint arXiv:1807.04216, (2018).
71. D. HOLM AND V. PUTKARADZE, *Aggregation of finite-size particle with variable mobility*, Phys. Rev. Lett., (2005), p. 95: 226106.
72. H. HUANG AND J.-G. LIU, *Error estimate of a random particle blob method for the Keller–Segel equation*, Math. Comp., 86 (2017), pp. 2719–2744.
73. P.-E. JABIN, *A review of the mean field limits for Vlasov equations*, Kinet. Relat. Models, 7 (2014), pp. 661–711.
74. P.-E. JABIN AND Z. WANG, *Mean field limit for stochastic particle systems*, in Active Particles. Vol. 1. Advances in Theory, Models, and Applications, Model. Simul. Sci. Eng. Technol., Birkhäuser/Springer, Cham, 2017, pp. 379–402.
75. R. JORDAN, D. KINDERLEHRER, AND F. OTTO, *The variational formulation of the Fokker–Plank equation*, SIAM. J. Math. Anal., 29 (1998), pp. 1–17.
76. E. KELLER AND L. SEGEL, *Traveling bands of chemotactic bacteria: a theoretical analysis*, J. Theoret. Biol., 30 (1971), pp. 6420–6437.
77. T. KOLOKOLNIKOV, J. A. CARRILLO, A. BERTOZZI, R. FETECAU, AND M. LEWIS, *Emergent behavior in multi-particle systems with non-local interactions*, Phys. D, 260 (2013), pp. 1–4.
78. L. LAGUZET, *High order variational numerical schemes with application to Nash-MFG vaccination games*, Ric. Mat., 67 (2018), pp. 247–269.
79. G. LEGENDRE AND G. TURINICI, *Second-order in time schemes for gradient flows in Wasserstein and geodesic metric spaces*, C. R. Math. Acad. Sci. Paris, 355 (2017), pp. 345–353.
80. W. LI, J. LU, AND L. WANG, *Fisher information regularization schemes for wasserstein gradient flows*, Journal of Computational Physics, (2020), p. 109449.
81. W. LI, S. OSHER, AND W. GANGBO, *A fast algorithm for earth mover's distance based on optimal transport and l_1 type regularization*, arXiv:1609.07092v3, (preprint).

82. W. Li, P. Yin, and S. Osher, *Computations of optimal transport distance with fisher information regularization*, Journal of Scientific Computing, 75 (2018), pp. 1581–1595.

83. J.- G. Liu, M. Tang, L. Wang, and Z. Zhou, *An accurate front capturing scheme for tumor growth models with a free boundary limit*, Journal of Computational Physics, 364 (2018), pp. 73 – 94.

84. J.- G. Liu, M. Tang, L. Wang, and Z. Zhou, *Analysis and computation of some tumor growth models with nutrient: from cell density models to free boundary dynamics*, DCDS-B, accepted (2018).

85. J.- G. Liu, L. Wang, and Z. Zhou, *Positivity-preserving and asymptotic preserving method for 2d Keller–Segel equations*, Mathematics of Computation, 87 (2018), pp. 1165–1189.

86. J.- G. Liu and R. Yang, *A random particle blob method for the Keller–Segel equation and convergence analysis*, Math. Comp., 86 (2017), pp. 725–745.

87. J. Maas, *Gradient flows of the entropy for finite markov chains*, Journal of Functional Analysis, 261 (2011), pp. 2250–2292.

88. D. Matthes and H. Osberger, *A convergent Lagrangian discretization for a nonlinear fourth-order equation*, Found. Comput. Math., 17 (2017), pp. 73–126.

89. D. Matthes and S. Plazotta, *A variational formulation of the bdf2 method for metric gradient flows*, ESAIM: Mathematical Modelling and Numerical Analysis, 53 (2019), pp. 145–172.

90. B. Maury, A. Roudneff- Chupin, and F. Santambrogio, *A macroscopic crowd motion model of gradient flow type*, Math. Models Methods Appl. Sci., 20 (2010), pp. 1787–1821.

91. B. Maury, A. Roudneff- Chupin, F. Santambrogio, and J. Venel, *Handling congestion in crowd motion modeling*, Netw. Heterog. Media, 6 (2011), pp. 485–519.

92. H. Osberger and D. Matthes, *Convergence of a fully discrete variational scheme for a thin-film equation*, in Topological optimization and optimal transport, vol. 17 of Radon Ser. Comput. Appl. Math., De Gruyter, Berlin, 2017, pp. 356–399.

93. F. Otto, *Double degenerate diffusion equations as steepest descent*, Sonderforschungsbereich, 256 (1996).

94. F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.

95. N. Papadakis, G. Peyre, and E. Oudet, *Optimal transport with proximal splitting*, SIAM. J. Image. Sci., 7 (2014), pp. 212–238.

96. B. Perthame, F. Quiros, and J. Vazquez, *The Hele–Shaw asymptotics for mechanical models of tumor growth*, Arch. Ratio. Mech. Anal., 212 (2014), pp. 93–127.

97. G. Peyré and M. Cuturi, *Computational Optimal Transport*, book in preparation, personal communication, 2018.

98. F. Santambrogio, *Optimal transport for applied mathematicians*, Birkäuser, NY, (2015), pp. 99–102.

99. Z. Sun, J. A. Carrillo, and C.- W. Shu, *A discontinuous Galerkin method for nonlinear parabolic equations and gradient flow problems with interaction potentials*, J. Comput. Phys., 352 (2018), pp. 76–104.

100. M. Tang, N. Vauchelet, I. Cheddadi, I. Vigon- Clementel, D. Drasdo, and B. Perthame, *Composite waves for a cell population system modeling tumor growth and invasion*, Chin. Ann. Math. Ser. B, 34 (2013), pp. 295–318.

101. C. Topaz, A. Bertozzi, and M. Lewis, *A nonlocal continuum model for biological aggregation*, Bull. Math. Bio., 68 (2006), pp. 1601–1623.

102. G. Toscani, *One-dimensional kinetic models of granular flows*, Math. Model. Numer. Anal., 34 (2000), pp. 1277–1291.

103. J. Vazquez, *The Porous Medium Equation*, Oxford Mathematical Monographs, Oxford University Press, 2007. Oxford, UK.

104. J. L. Vázquez, *The porous medium equation*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, Oxford, 2007. Mathematical theory.

105. C. Villani, *Topics in Optimal Transport*, 58 AMS, Grad. Stud. Math., 2003. Providence, RI.

106. M. Yan, *A new primal–dual algorithm for minimizing the sum of three functions with a linear operator*, Journal of Scientific Computing, (2018), pp. 1–20.

Springer

Affiliations

José A. Carrillo[1] · Katy Craig[2] · Li Wang[3] · Chaozhen Wei[4]

✉ José A. Carrillo
carrillo@maths.ox.ac.uk

Katy Craig
kcraig@math.ucsb.edu

Li Wang
wang8818@umn.edu

Chaozhen Wei
cwei4@wpi.edu

[1] Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK

[2] Department of Mathematics, University of California, Santa Barbara, CA 93117, USA

[3] School of Mathematics, University of Minnesota, Twin Cities, MN 55455, USA

[4] Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609, USA