VARIATIONAL PARTICLE SCHEMES FOR THE POROUS MEDIUM EQUATION AND FOR THE SYSTEM OF ISENTROPIC EULER EQUATIONS

MICHAEL WESTDICKENBERG AND JON WILKENING

Abstract. Both the porous medium equation and the system of isentropic Euler equations can be considered as steepest descents on suitable manifolds of probability measures in the framework of optimal transport theory. By discretizing these variational characterizations instead of the partial differential equations themselves, we obtain new schemes with remarkable stability properties. We show that they capture successfully the nonlinear features of the flows, such as shocks and rarefaction waves for the isentropic Euler equations. We also show how to design higher order methods for these problems in the optimal transport setting using backward differentiation formula (BDF) multi-step methods or diagonally implicit Runge-Kutta methods.

Contents

1. Introduction

It is well-known that the dynamics of many physical systems can be derived from variational principles. For classical mechanics, for example, the first variation of an *action functional* yields the equations of motion, which are also called the Euler-Lagrange equations of the functional. By discretizing these equations one can obtain numerical methods. This strategy is time-tested and reliable. More recently,

however, an alternative approach has attracted considerable interest, which consists of *discretizing the action functional* instead. The numerical method is then given by the Euler-Lagrange equations of this discrete functional. Methods obtained in this way enjoy remarkable stability properties and preserve the "symplectic" structure of the problem. We refer the reader to [13] for further information. In the present paper, we apply a similar strategy for the porous medium equation and the system of isentropic Euler equations in one space dimension, and derive numerical methods by discretizing a suitable variational principle.

The isentropic Euler equations model the dynamics of compressible fluids under the simplifying assumption that the thermodynamical entropy is constant in space and time. These equations form a system of hyperbolic conservation laws for the density ϱ and the Eulerian velocity field $\mathbf{u}$. They take the form

$$\begin{aligned} \partial_t \varrho + \nabla \cdot (\varrho \mathbf{u}) &= 0, \\ \partial_t(\varrho \mathbf{u}) + \nabla \cdot (\varrho \mathbf{u} \otimes \mathbf{u}) + \nabla P(\varrho) &= 0, \end{aligned} \tag{1.1}$$

where $(\varrho, \mathbf{u}) \colon [0, \infty) \times \mathbb{R}^d \longrightarrow \mathbb{U}$ are measurable functions with $\mathbb{U} := [0, \infty) \times \mathbb{R}^d$. Since entropy is assumed to be constant, the pressure $P(\varrho)$ depends on the density only. It is defined in terms of the internal energy $U(\varrho)$ of the fluid by

$$P(\varrho) := U'(\varrho)\varrho - U(\varrho) \quad \text{for all } \varrho \geqslant 0. \tag{1.2}$$

For the important case of polytropic fluids, we have

$$U(\varrho) = \frac{\kappa \varrho^\gamma}{\gamma - 1} \quad \text{and} \quad P(\varrho) = \kappa \varrho^\gamma. \tag{1.3}$$

Here $\gamma > 1$ is the adiabatic coefficient, and we will assume the common normalization $\kappa := \theta^2/\gamma$ with $\theta := (\gamma - 1)/2$. For isothermal flows, we have

$$U(\varrho) = \varrho \log \varrho \quad \text{and} \quad P(\varrho) = \varrho. \tag{1.4}$$

We are interested in the Cauchy problem, so we require that

$$(\varrho, \mathbf{u})(t = 0, \cdot) = (\bar\varrho, \bar{\mathbf{u}})$$

for suitable initial data $(\bar\varrho, \bar{\mathbf{u}})$ with finite mass and total energy.

It is again well-known [2, 10] that the system of isentropic Euler equations can be considered as the Euler-Lagrange equations of a certain action functional, so we might try to derive numerical schemes by following the procedure outlined above. But we want to work in the framework of optimal transport theory. Our discussion in the present paper will therefore not be based on the variation of an action functional, but on the notion that the system of isentropic Euler equations (1.1) can be considered as a *steepest descent* on a suitable abstract manifold. This notion was introduced in [8] and will be explained in detail below. To motivate our approach, recall that any *smooth* solution $(\varrho, \mathbf{u})$ of (1.1) also satisfies

$$\partial_t \left(\frac{1}{2} \varrho |\mathbf{u}|^2 + U(\varrho) \right) + \nabla \cdot \left(\left(\frac{1}{2} \varrho |\mathbf{u}|^2 + Q(\varrho) \right) \mathbf{u} \right) = 0, \tag{1.5}$$

where $Q(\varrho) := U(\varrho) + P(\varrho)$. Therefore the functions

$$\eta(\varrho, \mathbf{u}) := \frac{1}{2} \varrho |\mathbf{u}|^2 + U(\varrho) \quad \text{and} \quad q(\varrho, \mathbf{u}) := \left(\frac{1}{2} \varrho |\mathbf{u}|^2 + Q(\varrho) \right) \mathbf{u}$$

form an entropy/entropy flux pair for (1.1); see [6]. The entropy $\eta(\varrho, \mathbf{u})$ is a convex function of ϱ and $\varrho\mathbf{u}$. As a consequence of (1.5), the total energy

$$E[\varrho, \mathbf{u}](t) := \int_{\mathbb{R}^d} \left(\tfrac{1}{2}\varrho|\mathbf{u}|^2 + U(\varrho) \right)(t, x)\, dx \tag{1.6}$$

is conserved in time as long as the solution remains smooth.

On the other hand, it is known that solutions of (1.1) are generally not smooth: No matter how regular the initial data is, jump discontinuities can form in finite time. These jumps occur along d-dimensional submanifolds in space-time and are called shocks. Across shocks, total energy is dissipated (that is, transformed into forms of energy such as heat that are not accounted for in an isentropic model), so (1.5) cannot hold anymore. It is therefore natural to consider *weak* solutions $(\varrho, \mathbf{u})$ of (1.1) such that $E[\varrho, \mathbf{u}]$ is a non-increasing function of time.

As will be explained below, we consider solutions of the isentropic Euler equations as curves on the space of probability measures, using optimal transport theory. We are interested in curves along which the total energy decreases as fast as possible, under conservation of mass and momentum. This steepest descent interpretation has already proved very fruitful for certain degenerate parabolic equations, such as the porous medium equation; see below. Numerical schemes based on this variational principle have been derived in [11]. As the authors point out, these schemes have remarkable properties: oscillations can be reduced because the optimizations take place in weak topologies; the approximations can be constructed using discontinuous functions because no derivatives are needed; and overall, the schemes are very stable. In this paper, we revisit the discretization of the porous medium equation, and also introduce variational approximations for the isentropic Euler equations. We demonstrate that our schemes capture very well the nonlinear features of the flows. The simplest versions of our schemes are first order accurate.

Designing higher order methods in the optimal transport setting is an interesting challenge that has not previously been addressed. We present schemes that achieve second order convergence (away from shocks) and propose a general strategy for constructing higher order methods in a backward differentiation formula multi-step (BDF) or diagonally implicit Runge-Kutta framework. Again we observe that our schemes, due to their variational nature, are remarkably stable.

To put our work and that in [8] in perspective, let us first recall recent research by various authors on a steepest descent interpretation for certain degenerate parabolic equations. It was shown by Otto [15] that the porous medium equation

$$\partial_t \varrho - \Delta P(\varrho) = 0 \quad \text{in } [0, \infty) \times \mathbb{R}^d \tag{1.7}$$

is a *gradient flow* in the sense enumerated below. This result has been generalized considerably, and we refer the reader to [1, 17] for further information.

(1) We denote by $\mathscr{P}(\mathbb{R}^d)$ the space of all $\mathfrak{L}^d$-measurable, non-negative functions with unit integral and finite second moments, where $\mathfrak{L}^d$ is the Lebesgue measure. The space $\mathscr{P}(\mathbb{R}^d)$ is equipped with the *Wasserstein distance*, defined by

$$\mathbf{W}(\varrho_1, \varrho_2)^2 := \inf \left\{ \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x_2 - x_1|^2\, \gamma(dx_1, dx_2) \colon \pi^i \# \gamma = \varrho_i \mathfrak{L}^d \right\}, \tag{1.8}$$

where the infimum is taken over all transport plans γ, which are probability measures on the product space $\mathbb{R}^d \times \mathbb{R}^d$ with the property that the pushforward $\pi^i \# \gamma$

of γ under the projection $\pi^i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ onto the ith component equals $\varrho_i \mathfrak{L}^d$. The infimum in (1.8) represents the minimal quadratic cost required to transport the measure $\varrho_1 \mathfrak{L}^d$ to the measure $\varrho_2 \mathfrak{L}^d$, and one can show that it is always attained for some optimal transport plan γ. In fact, there is a lower semicontinuous, convex function $\phi : \mathbb{R}^d \longrightarrow \mathbb{R}$, uniquely determined ϱ_1-a.e. up to a constant, such that $\gamma = (\mathrm{id} \times \nabla\phi) \# (\varrho_1 \mathfrak{L}^d)$; see [1, 3, 7, 17]. This identity implies that

$$\mathbf{W}(\varrho_1, \varrho_2)^2 = \int_{\mathbb{R}^d} |\nabla\phi(x) - x|^2 \varrho_1(x)\, dx \quad \text{and} \quad (\nabla\phi)\#(\varrho_1 \mathfrak{L}^d) = \varrho_2 \mathfrak{L}^d.$$

We call $\nabla\phi$ an optimal transport map. It is invertible ϱ_1-a.e. and the Hessian $D^2\phi$ exists and is positive definite ϱ_1-a.e. When $d = 1$, the Wasserstein distance and the transport map can be computed more explicitly; see below.

(2) We introduce a differentiable structure on $\mathscr{P}(\mathbb{R}^d)$ as follows: For any point $\varrho \in \mathscr{P}(\mathbb{R}^d)$, the tangent space $\mathbb{T}_\varrho \mathscr{P}(\mathbb{R}^d)$ is defined as the closure of the space of smooth gradient vector fields in the $\mathscr{L}^2(\mathbb{R}^d, \varrho)$-norm. This definition is motivated by the fact that for any absolutely continuous curve $t \mapsto \varrho_t \in \mathscr{P}(\mathbb{R}^d)$ with $\varrho_0 = \varrho$, there exists a unique $\mathbf{u} \in \mathbb{T}_\varrho \mathscr{P}(\mathbb{R}^d)$ with the property that

$$\partial_t \varrho_t \big|_{t=0} + \nabla \cdot (\varrho\mathbf{u}) = 0 \quad \text{in } \mathscr{D}'(\mathbb{R}^d). \tag{1.9}$$

That is, for all test functions $\phi \in \mathscr{D}(\mathbb{R}^d)$ we have

$$\frac{d}{dt}\bigg|_{t=0} \int_{\mathbb{R}^d} \varrho_t(x)\phi(x)\, dx = \int_{\mathbb{R}^d} \mathbf{u}(x) \cdot \nabla\phi(x)\varrho(x)\, dx,$$

and so a change in density can be accounted for by a flux of mass along the velocity field $\mathbf{u}$. This structure makes $\mathscr{P}(\mathbb{R}^d)$ a Riemannian manifold, and we define

$$\mathbb{T}\mathscr{P}(\mathbb{R}^d) := \bigcup \Big\{ (\varrho, \mathbf{u}) : \varrho \in \mathscr{P}(\mathbb{R}^d), \mathbf{u} \in \mathbb{T}_\varrho \mathscr{P}(\mathbb{R}^d) \Big\}.$$

(3) If $\mathbf{u} = -\varrho^{-1}\nabla P(\varrho)$, then (1.9) yields the porous medium equation (1.7) at one instant in time. This vector field is the "gradient" of the internal energy

$$\mathcal{U}[\varrho] := \int_{\mathbb{R}^d} U(\varrho(x))\, dx \quad \text{with} \quad P(\varrho) = U'(\varrho)\varrho - U(\varrho) \tag{1.10}$$

in the sense that $\mathbf{u}$ is the uniquely determined element of minimal length in the *subdifferential* of $\mathcal{U}[\varrho]$ with respect to the Wasserstein distance. The vector field $\mathbf{u}$ is indeed tangent to $\mathscr{P}(\mathbb{R}^d)$ because $-\varrho^{-1}\nabla P(\varrho) = -\nabla U'(\varrho)$ is a gradient.

The internal energies for the porous medium equation that arise most frequently in physical applications are identical to those of the isentropic Euler equations, namely (1.3) or (1.4). For the latter choice with $\kappa = 1$, equation (1.7) becomes

$$\partial_t \varrho - \Delta\varrho = 0 \quad \text{in } [0, \infty) \times \mathbb{R}^d, \tag{1.11}$$

i.e., the porous medium equation reduces to the heat equation.

The interpretation of (1.7) as an abstract gradient flow suggests a natural time discretization: Given a time step $\tau > 0$ and the value $\varrho^n \in \mathscr{P}(\mathbb{R}^d)$ of the approximate solution at time $t^n := n\tau$, the value at time t^{n+1} is defined as

$$\varrho^{n+1} := \mathrm{argmin} \left\{ \frac{1}{2\tau}\mathbf{W}(\varrho^n, \varrho)^2 + \mathcal{U}[\varrho] : \varrho \in \mathscr{P}(\mathbb{R}^d) \right\}. \tag{1.12}$$

We assume that ϱ^0 has finite internal energy $\mathcal{U}[\varrho^0] < \infty$. It can then be shown that for all $n \geqslant 0$, problem (1.12) admits a unique solution with $\mathcal{U}[\varrho^{n+1}] \leqslant \mathcal{U}[\varrho^n]$. We describe the motivation for the minimization (1.12) in Appendix A.

It is useful to reformulate the porous medium equation (1.7) in Lagrangian terms. We fix a reference density $\bar{\varrho}$ and consider a time-dependent family of transport maps $X(t, \cdot) \colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$ such that $X(t, \cdot) \# \bar{\varrho} = \varrho(t, \cdot) \mathfrak{L}^d$ for all $t \geqslant 0$. If

$$X(t, x) = X(t, y) \quad \text{implies that} \quad \partial_t^+ X(t, x) = \partial_t^+ X(t, y)$$

for $\bar{\varrho}$-a.e. $x, y \in \mathbb{R}^d$, where ∂_t^+ denotes the forward partial derivative in time, then we can define an Eulerian velocity $\mathbf{u}$ by the formula $\partial_t^+ X =: \mathbf{u} \circ X$. For the porous medium flow, the velocity is given by Darcy's law $\mathbf{u} = -\nabla U'(\varrho)$. This vector field is related to the first variation of the functional $\mathcal{V}[X] := \mathcal{U}[X \# \bar{\varrho}]$, which is defined for all $\bar{\varrho}$-measurable maps $X \colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$. Let $\frac{\delta \mathcal{V}[X]}{\delta X} := \nabla U'(\varrho) \circ X$. As explained in Appendix A, we then have

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \mathcal{V}[X + \varepsilon(\zeta \circ X)] = \int_{\mathbb{R}^d} \frac{\delta \mathcal{V}[X]}{\delta X} \cdot (\zeta \circ X)\, d\bar{\varrho} = \int_{\mathbb{R}^d} \nabla U'(\varrho) \cdot \zeta\, \varrho\, dx$$

for all smooth test functions $\zeta \colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$. Recall that $X \# \bar{\varrho} = \varrho \mathfrak{L}^d$.

The minimization (1.12) can now be rewritten in terms of transport maps as

$$X^{n+1} := \operatorname*{argmin}_X \left\{ \frac{1}{2\tau} \int_{\mathbb{R}^d} |X - X^n|^2\, d\bar{\varrho} + \mathcal{V}[X] \right\}, \tag{1.13}$$

where X^0 is chosen such that $X^0 \# \bar{\varrho} = \varrho^0 \mathfrak{L}^d$. We obtain $X^{n+1} = \mathbf{r}^{n+1} \circ X^n$ for all $n \geqslant 0$, where $\mathbf{r}^{n+1}$ is the optimal transport map pushing $\varrho^n \mathfrak{L}^d$ forward to $\varrho^{n+1} \mathfrak{L}^d$. The discrete analogue of Darcy's law is then $\mathbf{u}^{n+1} = -\nabla U'(\varrho^{n+1})$, where

$$\mathbf{u}^{n+1} := \frac{\mathbf{r}^{n+1} - \operatorname{id}}{\tau} \circ (\mathbf{r}^{n+1})^{-1} = \frac{X^{n+1} - X^n}{\tau} \circ (X^{n+1})^{-1}. \tag{1.14}$$

We refer the reader to Appendix A for details. In particular, we have that

$$\frac{X^{n+1} - X^n}{\tau} = -\nabla U'(\varrho^{n+1}) \circ X^{n+1} = -\frac{\delta \mathcal{V}[X]}{\delta X}\bigg|_{X = X^{n+1}},$$

which shows that the minimization problem (1.13) is equivalent to one step of the backward Euler method for the dynamical system $\partial_t X = -\frac{\delta \mathcal{V}[X]}{\delta X}$.

Notice that if for each timestep we *reset the reference measure* $\bar{\varrho}$ to be equal to $\varrho^n \mathfrak{L}^d$, then we may choose $X^n = \operatorname{id}$ and obtain $X^{n+1} = \mathbf{r}^{n+1}$, so the minimization (1.13) reduces to (1.12) if the latter is rewritten as a minimization problem for the optimal transport map pushing $\varrho^n \mathfrak{L}^d$ forward to $\varrho^{n+1} \mathfrak{L}^d$; see again Appendix A. This is what we do in our first order schemes detailed below. In designing more complicated multi-step or Runge-Kutta schemes, however, it is useful to describe all the intermediate steps or stages using one single reference density, which requires the generality of the second half of (1.14). We also emphasize that in contrast to the space of probability measures $\mathscr{P}(\mathbb{R}^d)$, the space of transport maps $X \colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$ is *flat*, with a linear structure and a translation invariant metric.

A similar steepest descent interpretation as for the porous medium equation can be given for the system of isentropic Euler equations (1.1). The philosophy here is that among all possible weak solutions of (1.1) we try to pick the one that dissipates the total energy (1.6) as fast as possible, as advocated by Dafermos [6]. Note that

unlike in the porous medium case, for (1.1) the energy dissipation will be singular in the sense that it only occurs when the solution is discontinuous.

At the heart of the variational time discretization introduced in [8] lies an optimization problem similar to (1.12), but with a different homogeneity in the timestep; see (1.17) below. In fact, the relaxation of internal energy is a second order effect only. What dominates the flow is the transport of mass along the integral curves of the velocity field. Let us recall the time discretization introduced in [8].

Definition 1.1 (Variational Time Discretization).

(0) Let initial data $(\varrho^0, \mathbf{u}^0) \in \mathbb{T}\mathscr{P}(\mathbb{R}^d)$ and $\delta > 0$ be given. Put $t^0 := 0$.

Assume that $\varrho^n \in \mathscr{P}(\mathbb{R}^d)$ and $\mathbf{u}^n \in \mathscr{L}^2(\mathbb{R}^d, \varrho^n)$ are given, where $n \geqslant 0$. Then

(1) Choose $\tau \in [\delta/2, \delta]$ in such a way that the push-forward measure

$$(\mathrm{id} + \tau \mathbf{u}^n)\#(\varrho^n \mathfrak{L}^d) =: \hat{\varrho}^n \mathfrak{L}^d$$

is absolutely continuous with respect to the Lebesgue measure. Find

$$\hat{\mathbf{u}}^n \in \mathscr{L}^2(\mathbb{R}^d, \varrho^n) \quad \text{such that} \quad \left\{ \begin{array}{l} (\mathrm{id} + \tau \hat{\mathbf{u}}^n)\#(\varrho^n \mathfrak{L}^d) = \hat{\varrho}^n \mathfrak{L}^d \\[2mm] \displaystyle\int_{\mathbb{R}^d} |\tau \hat{\mathbf{u}}^n|^2 \varrho^n \, dx = \mathbf{W}(\varrho^n, \hat{\varrho}^n)^2 \end{array} \right\}. \qquad (1.15)$$

That is, find the velocity field $\hat{\mathbf{u}}^n$ with minimal $\mathscr{L}^2(\mathbb{R}^d, \varrho^n)$-norm such that the map $\mathrm{id} + \tau \hat{\mathbf{u}}^n$ pushes the measure $\varrho^n \mathfrak{L}^d$ forward to $\hat{\varrho}^n \mathfrak{L}^d$. Such a vector field always exists and is uniquely determined. It is given by the gradient of a semi-convex function, and is therefore a tangent vector. Moreover, we have the estimate

$$\int_{\mathbb{R}^d} |\hat{\mathbf{u}}^n|^2 \varrho^n \, dx \leqslant \int_{\mathbb{R}^d} |\mathbf{u}|^2 \varrho^n \, dx \qquad (1.16)$$

for any velocity $\mathbf{u}$ satisfying the constraint that $(\mathrm{id} + \tau \mathbf{u})\#\varrho^n = \hat{\varrho}^n$. This shows that by replacing $\mathbf{u}^n$ by the optimal transport velocity $\hat{\mathbf{u}}^n$, we decrease the kinetic energy as much as possible. The map $\mathrm{id} + \tau \hat{\mathbf{u}}^n$ is invertible $\hat{\varrho}^n$-a.e.

(2) Update the density by computing the minimizer

$$\varrho^{n+1} := \mathrm{argmin}\left\{ \frac{3}{4\tau^2} \mathbf{W}(\hat{\varrho}^n, \varrho)^2 + \mathcal{U}[\varrho] \colon \varrho \in \mathscr{P}(\mathbb{R}^d) \right\}. \qquad (1.17)$$

Here $\mathcal{U}[\varrho]$ is the internal energy functional from (1.10), which is a part of (1.6). One can show that the density ϱ^{n+1} is uniquely determined and that

$$\left(\mathrm{id} + \frac{2\tau^2}{3} \nabla U'(\varrho^{n+1}) \right)\#\varrho^{n+1} = \hat{\varrho}^n,$$

$$\int_{\mathbb{R}^d} \left| \frac{2\tau^2}{3} \nabla U'(\varrho^{n+1}) \right|^2 \varrho^{n+1} \, dx = \mathbf{W}(\varrho^{n+1}, \hat{\varrho}^n)^2.$$

The density ϱ^{n+1} is thus regular in the sense that $\frac{2\tau^2}{3} \nabla U'(\varrho^{n+1}) \in \mathscr{L}^2(\mathbb{R}^d, \varrho^{n+1})$. In fact, the latter vector field is the gradient of a semi-convex function.

(3) Update the velocity by defining

$$\mathbf{u}^{n+1} := \hat{\mathbf{u}}^n \circ (\mathrm{id} + \tau \hat{\mathbf{u}}^n)^{-1} \circ \left(\mathrm{id} + \frac{2\tau^2}{3} \nabla U'(\varrho^{n+1}) \right) - \tau \nabla U'(\varrho^{n+1}). \qquad (1.18)$$

(4) Let $t^{n+1} := t^n + \tau$, increase n by one, and continue with Step (1).

We refer the reader to [8] for more details. It is shown there that

$$E[\varrho^{n+1}, \mathbf{u}^{n+1}] \leqslant E[\varrho^n, \mathbf{u}^n] \quad \text{for all } n \geqslant 0.$$

One can then define the function

$$(\varrho, \mathbf{u})(t, \cdot) := (\varrho^n, \mathbf{u}^n) \quad \text{for all } t \in [t^n, t^{n+1}) \text{ and } n \geqslant 0,$$

which is piecewise constant in time and approximates a weak, energy-dissipating solution of the isentropic Euler equation (1.1). A formal argument is given in [8] for the convergence of this approximation when the timestep $\delta \to 0$.

We give a derivation of the time discretization of Definition 1.1 in Appendix B. From the point of view of numerical analysis, it is interesting to generalize this scheme to a family of first order implicit methods that also includes the backward Euler method as a special case. To this end, let us temporarily assume the solution is smooth and write the isentropic Euler equations in a Lagrangian frame as

$$\frac{d}{dt}\begin{pmatrix} X \\ V \end{pmatrix} = \begin{pmatrix} V \\ \mathbf{f}[X] \end{pmatrix} \quad \text{with} \quad \mathbf{f}[X] := -\frac{\delta \mathcal{V}[X]}{\delta X} = -\nabla U'(\varrho_X) \circ X, \qquad (1.19)$$

where $X \# \bar\varrho =: \varrho_X \mathcal{L}^d$ and $V := \mathbf{u} \circ X$ is the pull-back of the Eulerian velocity. Again we set $\mathcal{V}[X] := \mathcal{U}[X \# \bar\varrho]$, as in the porous medium case discussed above. Let now $\alpha \in (0, 1]$ be a fixed parameter and consider the implicit method

$$\begin{pmatrix} X^{n+1} \\ V^{n+1} \end{pmatrix} = \begin{pmatrix} X^n \\ V^n \end{pmatrix} + \tau \begin{pmatrix} (1 - \alpha)V^n + \alpha V^{n+1} \\ \mathbf{f}[X^{n+1}] \end{pmatrix}. \qquad (1.20)$$

Using the first equation to eliminate V^{n+1} in the second, we obtain

$$\frac{1}{\alpha\tau^2}\Big(X^{n+1} - (X^n + \tau V^n) \Big) = \mathbf{f}[X^{n+1}], \qquad (1.21)$$

$$V^{n+1} = V^n + \tau \mathbf{f}[X^{n+1}]. \qquad (1.22)$$

Equation (1.21) will be satisfied by the solution of the minimization problem

$$X^{n+1} = \underset{X}{\operatorname{argmin}} \left\{ \frac{1}{2\alpha\tau^2} \int_{\mathbb{R}^d} |X - (X^n + \tau V^n)|^2 \, d\bar\varrho + \mathcal{V}[X] \right\}. \qquad (1.23)$$

Once this is solved, we substitute the left-hand side of (1.21) for $\mathbf{f}[X^{n+1}]$ in (1.22) to obtain V^{n+1}. This is the Lagrangian reformulation of the energy minimization in Step (2) of Definition 1.1. As explained there, we combine this step with a velocity projection, so we replace in each step the velocity V^n by an optimal transport velocity, which in the Lagrangian framework takes the form $\hat{V}^n = \hat{\mathbf{u}}^n \circ X^n$ with $\hat{\mathbf{u}}^n$ given by (1.15). This modification dissipates the maximum amount of kinetic energy without changing the convected distribution of mass.

The three choices of α that seem most natural are $\alpha = 1$, which corresponds to the backward Euler method; $\alpha = \frac{2}{3}$, which corresponds to the original time discretization proposed in [8] (see Appendix B for further information); and $\alpha = \frac{1}{2}$, which causes the acceleration $A^{n+1} = \mathbf{f}[X^{n+1}]$ to agree with its Taylor expansion, $A^{n+1} \approx 2\tau^{-2}\big(X^{n+1} - (X^n + \tau V^n)\big)$. All three of these variants yield first order methods that appear to capture shocks and rarefaction waves correctly in our numerical experiments. We present a second order version in Section 2.2.

2. Variational Particle Scheme

We now introduce a fully discrete version of the variational time discretization discussed in the previous section for the one-dimensional isentropic Euler equations

$$\partial_t \varrho + \partial_x(\varrho u) = 0$$
$$\partial_t(\varrho u) + \partial_x(\varrho u^2 + P(\varrho)) = 0 \qquad \text{in } [0, \infty) \times \mathbb{R}, \tag{2.1}$$

as well as for the one-dimensional porous medium equation

$$\partial_t \varrho - \Delta P(\varrho) = 0 \quad \text{in } [0, \infty) \times \mathbb{R}. \tag{2.2}$$

The pressure P is related to the internal energy by equation (1.2), with energy density U given by either (1.3) in the case of polytropic gases (we put $\kappa = 1$ for the porous medium equation), or by (1.4) in the case of isothermal gases.

2.1. **First Order Scheme.** In the simplest version of the algorithm, we assume the fluid to be composed of N particles of equal mass $m := 1/N$, where $N > 1$. At a given time t^n, the particles are located at positions $\mathbf{x}^n = \{x_1^n, \ldots, x_N^n\}$ with velocities $\mathbf{u}^n = \{u_1^n, \ldots, u_N^n\}$. In the following, we will always assume that the x_i^n are ordered and *strictly increasing*. The total energy is then given by

$$E(\mathbf{x}^n, \mathbf{u}^n) := \sum_{i=1}^{N} \frac{1}{2} m |u_i^n|^2 + \sum_{i=1}^{N-1} U\left(\frac{m}{x_{i+1}^n - x_i^n}\right)(x_{i+1}^n - x_i^n).$$

In accordance with thermodynamics, we think of the density as the inverse of the specific volume, which we define as the distance between two neighboring particles divided by the particle mass. This formula is used in the internal energy.

Definition 2.1 (Variational Particle Scheme, Version 1)**.**

(0) Let initial positions $\mathbf{x}^0$, velocities $\mathbf{u}^0$, and $\tau > 0$ be given. Put $t^0 := 0$.

Assume that positions $\mathbf{x}^n$ and velocities $\mathbf{u}^n$ are given for $n \geqslant 0$. Then

(1) Compute intermediate positions defined by $y_i := x_i^n + \tau u_i^n$ for all $1 \leqslant i \leqslant N$. Find the permutation σ of the set $\{1, \ldots, N\}$ with the property that the $\hat{x}_i^n := y_{\sigma(i)}$ are nondecreasing in i. We assume that if there exist indices $i < j$ with $\hat{x}_i = \hat{x}_j$, then $\sigma(i) < \sigma(j)$. Then define $\hat{u}_i^n := (\hat{x}_i^n - x_i^n)/\tau$ for all i.

(2) Define new positions $\mathbf{x}^{n+1}$ as the minimizer of the functional

$$F(\mathbf{z}) := \sum_{i=1}^{N} \frac{3}{4\tau^2} m |z_i - \hat{x}_i^n|^2 + \sum_{i=1}^{N-1} U\left(\frac{m}{z_{i+1} - z_i}\right)(z_{i+1} - z_i) \tag{2.3}$$

over all families of positions $\mathbf{z} \in \mathbb{R}^N$ such that $z_{i+1} > z_i$ for all i.

(3) Compute the new velocities $u_i^{n+1} := \hat{u}_i^n + \frac{3}{2\tau}(x_i^{n+1} - \hat{x}_i^n)$ for all i.

(4) Increase n by one and continue with Step (1).

Remark 2.2. In Step (1) we perform a free transport of all particles in the direction of their given velocities. We obtain a new distribution of mass that is characterized by the new particle positions y_i. Note that we defined the density in terms of the distance between *neighboring* particles. Since the original ordering can be destroyed during the free transport step, we rearrange the new positions in nondecreasing order. On the discrete level, it is not necessary to make sure that positions do

not coincide because we never compute the internal energy of this intermediate configuration. Therefore we can use the same timestep τ for all updates.

To understand the definition of the new velocities in Step (1), notice that the Wasserstein distance as defined in (1.8) can easily be generalized to pairs of *probability measures* μ_1 and μ_2 with finite second moment; see [1,17] for example. If the two measures are given as convex combinations of Dirac measures:

$$\mu_1 = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i} \quad \text{and} \quad \mu_2 = \frac{1}{N} \sum_{i=1}^{N} \delta_{y_i},$$

where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}^d$ for all i, then the squared Wasserstein distance between the measures μ_1 and μ_2 coincides with the minimum of the functional

$$W(\sigma)^2 := \frac{1}{N} \sum_{i=1}^{N} |x_i - y_{\sigma(i)}|^2$$

among all permutations σ of the index set $\{1, \ldots, N\}$. This amounts to a Linear Assignment Problem, which in the one-dimensional case reduces to sorting the positions y_i. This can be done with worst-case complexity $\mathcal{O}(N \log N)$. In our case, the y_i are actually largely ordered already, so the sorting is quite inexpensive. Step (1) therefore amounts to computing the Wasserstein distance between the old and the intermediate positions of particles, and then finding the optimal velocity that achieves this transport: We have that $x_i^n + \tau \hat{u}_i^n = \hat{x}_i^n$ for all i.

The functional F in (2.3) is the analogue of the one in (1.17). Since we minimize over positions z_i that are increasing, the first term of (2.3) can again be interpreted as a Wasserstein distance squared between convex combinations of Dirac measures located at positions z_i and $\hat{x}_i^n$. Note that the choice $z_i := x_i^n$ for all i is admissible in (2.3) and results in a finite value of the functional. Therefore the new positions x_i^{n+1} must have finite internal energy and thus be strictly increasing.

Remark 2.3. With only minor modifications we obtain a particle scheme for the porous medium equation (2.2), which contains the heat equation as a special case. Since now the only unknown is the density ϱ, Steps (1) and (3) in Definition 2.1 become irrelevant. In Step (2), we substitute the factor $1/(2\tau)$ for $3/(4\tau^2)$ to match (1.12), and the intermediate positions $\hat{x}_i$ must be replaced by x_i. Of course, we can also implement the backward Euler method for the isentropic and isothermal Euler equations by changing $3/(4\tau^2)$ and $3/(2\tau)$ to $1/(2\tau^2)$ and $1/\tau$, respectively.

2.2. **Second Order Scheme.** In this section, we derive a variant of the scheme of Definition 2.1 that is formally second order in space and time. Before doing so, let us collect a few facts about the Wasserstein distance in one space dimension. For any probability measure μ in $\mathbb{R}$ we define the distribution function

$$F_\mu(t) := \mu\big((-\infty, t)\big) \quad \text{for all } t \in \mathbb{R}.$$

The function F_μ is nondecreasing, but not necessarily strictly, and can be discontinuous if the measure μ contains atoms (Dirac measures). We define

$$F_\mu^{-1}(s) := \sup \left\{ t \in \mathbb{R} \colon F_\mu(r) \leqslant s \right\} \quad \text{for all } s \in [0, 1],$$

which is the generalized inverse of F_μ. For any pair of probability measures (μ, ν) on $\mathbb{R}$ with finite second moment, the Wasserstein distance between μ and ν can

then be computed in terms of the inverse distribution functions as

$$\mathbf{W}(\mu,\nu)^2 := \int_{[0,1]} |F_\mu^{-1}(s) - F_\nu^{-1}(s)|^2 \, ds. \tag{2.4}$$

If μ does not contain any atoms, then F_μ is continuous, and the composition $F_\nu^{-1} \circ F_\mu$ is an optimal transport map pushing μ forward to ν; see Theorem 6.0.2 in [1]. Note that the inverse function F_μ^{-1} pushes the measure $\mathbf{1}_{[0,1]} \mathfrak{L}^1$ forward to μ.

Assume now that the measure μ is absolutely continuous with respect to the Lebesgue measure, with a piecewise constant density: Let numbers $x_0 < \ldots < x_N$ and $m_i \geqslant 0$ be given such that $\sum_{i=1}^N m_i = 1$. Let $\mu = \varrho \mathfrak{L}^1$ with

$$\varrho(x) := \begin{cases} 0 & \text{if } x < x_0 \text{ or } x \geqslant x_N, \\ \dfrac{m_i}{x_i - x_{i-1}} & \text{if } x \in [x_{i-1}, x_i) \text{ and } 1 \leqslant i \leqslant N. \end{cases} \tag{2.5}$$

Then the inverse distribution function of μ is a piecewise linear function that can be computed explicitly. Defining $s_k := \sum_{i=1}^k m_i$ for $0 \leqslant k \leqslant N$, we have

$$F_\mu^{-1}(s) = x_{k-1} \frac{s_k - s}{s_k - s_{k-1}} + x_k \frac{s - s_{k-1}}{s_k - s_{k-1}} \quad \text{for all } s \in [s_{k-1}, s_k]$$

and $1 \leqslant k \leqslant N$. If the second measure ν is piecewise constant as well, it is possible to compute the Wasserstein distance between μ and ν using formula (2.4). This is particularly simple if both μ and ν have the same number of intervals, and if the same mass is assigned to each interval, because then the inverse distribution functions F_μ^{-1} and F_ν^{-1} are piecewise linear on the same intervals.

Assume now that $\nu = \hat{\varrho} \mathfrak{L}^1$ with

$$\hat{\varrho}(x) := \begin{cases} 0 & \text{if } x < \hat{x}_0 \text{ or } x \geqslant \hat{x}_N, \\ \dfrac{\hat{m}_i}{\hat{x}_i - \hat{x}_{i-1}} & \text{if } x \in [\hat{x}_{i-1}, \hat{x}_i) \text{ and } 1 \leqslant i \leqslant N, \end{cases}$$

for suitable numbers $\hat{x}_0 < \ldots < \hat{x}_N$ and $\hat{m}_i \geqslant 0$ with $\sum_{i=0}^N \hat{m}_i = 1$. We want to project ν onto the space of measures of the form $\mu = \varrho \mathfrak{L}^d$, where the density ϱ is given by (2.5) and the masses m_i are fixed. More precisely, we want to choose the positions $\mathbf{x} := (x_0, \ldots, x_N)$ in (2.5) so that the Wasserstein distance

$$\mathbf{W}(\mu,\nu)^2 = \int_{[0,1]} |F_\mu^{-1}(s) - F_\nu^{-1}(s)|^2 \, ds$$

is minimal. This amounts to a quadratic minimization problem that can be solved easily. Let $\{\varphi_k\}_{k=0}^N$ be the standard finite element hat functions with vertices s_k as defined above. Then the minimizer is given by $\mathbf{x} = A^{-1}b$, where

$$A_{kl} := \int_{[0,1]} \varphi_k(s)\varphi_l(s) \, ds \quad \text{and} \quad b_k := \int_{[0,1]} \varphi_k(s) F_\nu^{-1}(s) \, ds$$

for all $0 \leqslant k, l \leqslant N$. The calculation of A_{kl} and b_k is straightforward. Notice that the function F_ν^{-1} is piecewise linear on the intervals $[\hat{s}_{k-1}, \hat{s}_k]$ with $\hat{s}_k := \sum_{i=1}^k \hat{m}_i$, and the $\hat{s}_k$ are different from s_k (otherwise there is nothing to do). The computation of b_k therefore requires a partition of $[0,1]$ that uses both sets of positions. Even if the $\hat{x}_i$ are nondecreasing, the components x_i of the minimizer may not be.

The projection outlined in the previous paragraph works also in the degenerate situation where $\hat{x}_{i-1} = \hat{x}_i$ for one or more i. In this case ν has a Dirac measure at position $\hat{x}_i$ and $\hat{m}_i$ denotes the mass of that measure. The inverse distribution

function F_ν^{-1} is then no longer continuous, but it is still straightforward to compute the numbers b_i. The minimizer $\mathbf{x}$ can be obtained as above.

We can now define our second order method for the isentropic Euler equations. Let us first discuss the accuracy in space. Instead of working with point masses, we now approximate the density by a function that is piecewise constant on intervals of variable length. Specifically, we fix numbers $m_i \geqslant 0$ with $\sum_{i=1}^N m_i = 1$. At any time t^n, the density ϱ^n is then determined by a vector $\mathbf{x}^n = (x_0^n, x_1^n, \ldots, x_N^n) \in \mathbb{R}^{N+1}$ of positions with $x_{i-1}^n < x_i^n$ for all $1 \leqslant i \leqslant N$, through the formula

$$\varrho^n(x) := \begin{cases} 0 & \text{if } x < x_0^n \text{ or } x \geqslant x_N^n, \\ \dfrac{m_i}{x_i^n - x_{i-1}^n} & \text{if } x \in [x_{i-1}^n, x_i^n) \text{ and } 1 \leqslant i \leqslant N. \end{cases} \tag{2.6}$$

We approximate the velocity at time t^n by a piecewise linear function on the intervals $[x_{i-1}^n, x_i^n)$. It is determined by $\mathbf{u}^n = (u_0^n, u_1^n, \ldots, u_N^n) \in \mathbb{R}^{N+1}$ as

$$u^n(x) := \begin{cases} 0 & \text{if } x < x_0^n \text{ or } x \geqslant x_N^n, \\ \dfrac{x_i^n - x}{x_i^n - x_{i-1}^n} u_{i-1}^n + \dfrac{x - x_{i-1}^n}{x_i^n - x_{i-1}^n} u_i^n & \text{if } x \in [x_{i-1}^n, x_i^n) \text{ and } 1 \leqslant i \leqslant N. \end{cases}$$

For any $n \geqslant 0$ consider now the push-forward measure $\nu := (\mathrm{id} + \tau u^n)\#(\varrho^n \mathcal{L}^1)$. Let $y_i := x_i^n + \tau u_i^n$ for all $0 \leqslant i \leqslant N$, and let σ be a permutation of $\{0, 1, \ldots, N\}$ such that the positions $\hat{x}_i := y_{\sigma(i)}$ are nondecreasing in i. We assume that if there are indices $i < j$ with $\hat{x}_i = \hat{x}_j$, then $\sigma(i) < \sigma(j)$. The measure ν is piecewise constant on intervals $[\hat{x}_{i-1}, \hat{x}_i)$ for which $\hat{x}_{i-1} < \hat{x}_i$, and we denote the mass carried by this interval by $\hat{m}_i$ and the density by $\hat{\varrho}_i := \hat{m}_i / (\hat{x}_i - \hat{x}_{i-1})$. If $\hat{x}_{i-1} = \hat{x}_i$, then the measure ν has a Dirac measure located at position $\hat{x}_i$ and $\hat{m}_i$ denotes the mass of this measure. By construction, we have $\hat{m}_i \geqslant 0$ and $\sum_{i=1}^N \hat{m}_i = 1$.

To compute the numbers $\hat{m}_i$ we first initialize them to zero. We then run through the original intervals $[x_{i-1}, x_i)$ and assign their masses to the new intervals they are mapped onto, in proportion to the lengths of these new intervals:

for $i = 1$ **to** N
    $k = \min\{\sigma^{-1}(i-1), \sigma^{-1}(i)\}$
    $l = \max\{\sigma^{-1}(i-1), \sigma^{-1}(i)\}$
    for $j = k+1$ **to** l
        if $\hat{x}_k^n = \hat{x}_l^n$
            $\hat{m}_j = \hat{m}_j + \dfrac{m_i}{l - k}$ (point mass case)
        else
            $\hat{m}_j = \hat{m}_j + m_i \dfrac{\hat{x}_j^n - \hat{x}_{j-1}^n}{\hat{x}_l^n - \hat{x}_k^n}$ (distribute according to length).

As we explained above, the computation of the Wasserstein distance in one space dimension is very simple if the pair of measures is induced by two piecewise constant densities that have the same number N of intervals, and that assign the same mass m_i to the ith interval. We therefore project the push-forward $\nu = (\mathrm{id} + \tau u^n)\#(\varrho^n \mathcal{L}^1)$ onto the space of measures of this form, before proceeding with the variational time discretization (VTD) algorithm of Definition 1.1. The minimization in (1.12) is only performed over densities ϱ of the same form; see below for more details. Otherwise this version of the fully discrete algorithm is identical to the VTD algorithm.

Next we show how to modify the algorithm to achieve second order accuracy in time. Consider again the Lagrangian formulation (1.19) of the system of isentropic Euler equations. The key feature of the scheme (1.20) that allows X^{n+1} to be determined by solving a minimization problem of the form (1.23) is that the functional $\mathbf{f}$ is only evaluated at X^{n+1} in (1.20). It is also desirable to choose an L-stable method to prevent oscillations from growing near shocks. The second order backward differentiation formula method [9] satisfies both of these conditions. Therefore it is a natural candidate to try to adapt to our variational particle scheme framework. Applied to problem (1.19), BDF2 takes the form

$$\frac{3}{2}\begin{pmatrix} X^{n+1} \\ V^{n+1} \end{pmatrix} - 2\begin{pmatrix} X^n \\ V^n \end{pmatrix} + \frac{1}{2}\begin{pmatrix} X^{n-1} \\ V^{n-1} \end{pmatrix} = \tau\begin{pmatrix} V^{n+1} \\ \mathbf{f}[X^{n+1}] \end{pmatrix}. \tag{2.7}$$

Solving for $X := X^{n+1}$ and V^{n+1} in turn, we obtain

$$\frac{3}{\tau^2}\left[\left(X - (X^n + \tau V^n)\right) - \frac{1}{4}\left(X - \left(X^{n-1} + 2\tau(\tfrac{2}{3}V^n + \tfrac{1}{3}V^{n-1}))\right)\right)\right] = \mathbf{f}[X],$$

$$V^{n+1} = 2V^n - (\tfrac{2}{3}V^n + \tfrac{1}{3}V^{n-1}) + \tfrac{2\tau}{3}\mathbf{f}[X^{n+1}]. \tag{2.8}$$

The first equation characterizes the solution of the minimization problem (2.9) below, which is similar to (1.23). As before, we handle the formation of shocks by dissipating the maximum amount of energy possible without changing the convected distribution of mass from t^{n-1} to t^n to t^{n+1}. Notice that the convex combination $\tfrac{2}{3}V^n + \tfrac{1}{3}V^{n-1}$ (rather than V^{n-1}) is used to transport mass from time t^{n-1} to time t^{n+1}. As V and $\mathbf{u}$ are related via $V = \mathbf{u} \circ X$, we will use $\mathbf{u}$ to denote velocities below. Here is our second order method for the isentropic Euler equations (1.1):

Definition 2.4 (Variational Particle Scheme, Version 2).

Fix once and for all masses $m_i \geqslant 0$ with $\sum_{i=1}^N m_i = 1$.

(0) Let initial positions $\mathbf{x}^0$, velocities $\mathbf{u}^0$, and $\tau > 0$ be given. Put $t^0 := 0$.

Assume that positions $\mathbf{x}^n$ and velocities $\mathbf{u}^n$ are given for $n \geqslant 0$. Then

(1a) Compute the intermediate positions $y_i := x_i^n + \tau u_i^n$ for all $0 \leqslant i \leqslant N$. Sort them and redistribute the masses to obtain $\hat{x}_i^n$ and $\hat{m}_i$. Project back onto the space of densities that are constant on N intervals and assign mass m_i to the ith interval. Let $\mathbf{x} = (x_0', \dots, x_N')$ denote the positions of the minimizer obtained from the quadratic minimization problem induced by this projection. See the beginning of this section for more details. Then define the velocities $u_i' := (x_i' - x_i^n)/\tau$.

(1b) Compute the intermediate positions $y_i := x_i^{n-1} + 2\tau(\tfrac{2}{3}u_i^n + \tfrac{1}{3}u_i^{n-1})$ for all $0 \leqslant i \leqslant N$. Sort them and redistribute the masses to obtain $\hat{x}_i^n$ and $\hat{m}_i$ as before. Project back onto the space of densities that are constant on N intervals and assign mass m_i to the ith interval. Let $\mathbf{x}'' = (x_0'', \dots, x_N'')$ denote the positions of the minimizer obtained from the quadratic minimization problem induced by this projection. Then define the velocities $u_i'' := (x_i'' - x_i^{n-1})/(2\tau)$, which play the role of $\tfrac{2}{3}V^n + \tfrac{1}{3}V^{n-1}$ in (2.8).
    Note: Skip Step (1b) on the first iteration.

(2) Minimize the functional

$$F(\mathbf{z}) := \frac{3}{2\tau^2}\|\mathbf{z} - \mathbf{x}'\|_m^2 - \frac{3}{8\tau^2}\|\mathbf{z} - \mathbf{x}''\|_m^2 + \sum_{i=1}^{N} U\left(\frac{m_i}{z_i - z_{i-1}}\right)(z_i - z_{i-1}) \quad (2.9)$$

over all $\mathbf{z} = (z_0, \ldots, z_N)$ such that $z_{i-1} < z_i$ for all $1 \leqslant i \leqslant N$. We define

$$\|\mathbf{z}\|_m^2 := \int_0^1 \left| \sum_{i=0}^{N} z_i \varphi_i(s) \right|^2 ds,$$

where φ_i is the piecewise linear function satisfying $\varphi_i(s_k) = \delta_{ik}$ for all $0 \leqslant k \leqslant N$, with $s_k := \sum_{i=1}^{k} m_i$. In fact, we have $\|\mathbf{z}\|_m^2 = \mathbf{z}^T A \mathbf{z}$ with the matrix A defined by

$$A_{ij} := \int_{[0,1]} \varphi_i(s)\varphi_j(s) \, ds$$

for all $0 \leqslant i, j \leqslant N$. We denote by $\mathbf{x}^{n+1} = (x_0^{n+1}, \ldots, x_N^{n+1})$ the minimizer of this convex optimization problem. Notice that the Euler-Lagrange equations of (2.9) are of the form (2.8) in the second order backwards differentiation scheme.

Note: On the first iteration, minimize instead the functional

$$F(\mathbf{z}) := \frac{3}{4\tau^2}\|\mathbf{z} - \mathbf{x}'\|_m^2 + \sum_{i=1}^{N} U\left(\frac{m_i}{z_i - z_{i-1}}\right)(z_i - z_{i-1}).$$

(3) Compute the new velocities

$$u_i^{n+1} := 2u_i' - u_i'' + \frac{2}{\tau}\left(x_i^{n+1} - x_i'\right) - \frac{1}{2\tau}\left(x_i^{n+1} - x_i''\right).$$

for all $0 \leqslant i \leqslant N$. Notice that this formula is of the form (2.8), where we recall that u_i'' represents $\frac{2}{3}u_i^n + \frac{1}{3}u_i^{n-1}$, not u_i^{n-1}.

Note: On the first iteration, define instead

$$u_i^{n+1} := u_i' + \frac{3}{2\tau}\left(x_i^{n+1} - x_i'\right).$$

(4) Increase n by one and continue with Step (1).

Remark 2.5. While the $\hat{x}_j^n$ are nondecreasing by construction, the projection can cause the components of the minimizer to be slightly out of order. This is acceptable as the minimization (1.17) over densities ϱ of the form (2.6) gives the same result if we replace the push-forward measure $\nu := (\mathrm{id} + \tau u^n)\#(\varrho^n \mathcal{L}^1)$ by its projection onto the space of densities of the form (2.6). Thus, in spite of the projection, we still minimize the distance from ν to compute ϱ^{n+1}. The values x_i^{n+1} will again be strictly increasing since otherwise the internal energy would be infinite.

Remark 2.6. As before, the same procedure works for the porous medium equation if we eliminate Steps (1) and (3) and replace the functional F in step (2) by

$$F(\mathbf{z}) := \frac{1}{\tau}\|\mathbf{z} - \mathbf{x}^n\|_m^2 - \frac{1}{4\tau}\|\mathbf{z} - \mathbf{x}^{n-1}\|_m^2 + \sum_{i=1}^{N} U\left(\frac{m_i}{z_i - z_{i-1}}\right)(z_i - z_{i-1}).$$

On the first iteration, one should use

$$F(\mathbf{z}) := \frac{1}{2\tau}\|\mathbf{z} - \mathbf{x}^n\|_m^2 + \sum_{i=1}^{N} U\left(\frac{m_i}{z_i - z_{i-1}}\right)(z_i - z_{i-1}).$$

Remark 2.7. We have also constructed second (and higher) order variational particle schemes in the diagonally implicit Runge-Kutta framework [9]. The underlying Runge-Kutta method must be "stiffly accurate", i.e., the numerical solution at t^{n+1} must agree with the last internal stage of the Runge-Kutta procedure. We illustrate the basic idea using the two-stage second order DIRK scheme with Butcher array

$$
\begin{array}{c|cc}
1/4 & 1/4 & \\
1 & 2/3 & 1/3 \\
\hline
 & 2/3 & 1/3
\end{array} \ . \tag{2.10}
$$

The system we need to solve is

$$
\begin{pmatrix} X^{n+\frac{1}{4}} \\ V^{n+\frac{1}{4}} \end{pmatrix} = \begin{pmatrix} X^n \\ V^n \end{pmatrix} + \frac{\tau}{4} \begin{pmatrix} V^{n+\frac{1}{4}} \\ \mathbf{f}[X^{n+\frac{1}{4}}] \end{pmatrix}, \tag{2.11}
$$

$$
\begin{pmatrix} X^{n+1} \\ V^{n+1} \end{pmatrix} = \begin{pmatrix} X^n \\ V^n \end{pmatrix} + \frac{2\tau}{3} \begin{pmatrix} V^{n+\frac{1}{4}} \\ \mathbf{f}[X^{n+\frac{1}{4}}] \end{pmatrix} + \frac{\tau}{3} \begin{pmatrix} V^{n+1} \\ \mathbf{f}[X^{n+1}] \end{pmatrix}. \tag{2.12}
$$

Equation (2.11) is simply the backward Euler method with steplength $\tau/4$. Therefore it can be solved as in (1.21)–(1.23) above with α, τ, X^{n+1} and V^{n+1} replaced by 1, $\tau/4$, $X^{n+\frac{1}{4}}$ and $V^{n+\frac{1}{4}}$, respectively. Next we choose a suitable linear combination of (2.11) and (2.12) to eliminate the term $\mathbf{f}[X^{n+\frac{1}{4}}]$. We find

$$
3 \begin{pmatrix} X^{n+1} \\ V^{n+1} \end{pmatrix} - 8 \begin{pmatrix} X^{n+\frac{1}{4}} \\ V^{n+\frac{1}{4}} \end{pmatrix} + 5 \begin{pmatrix} X^n \\ V^n \end{pmatrix} = \tau \begin{pmatrix} V^{n+1} \\ \mathbf{f}[X^{n+1}] \end{pmatrix}. \tag{2.13}
$$

The scheme coefficients (2.10) were chosen to minimize the magnitude of the coefficient 8 in (2.13). Since $X^{n+\frac{1}{4}}$ and $V^{n+\frac{1}{4}}$ are already known from solving (2.11), this equation is structurally identical to the multi-step scheme (2.7). As before, the unknown $X := X^{n+1}$ can be found independently from V^{n+1}. We have

$$
\frac{24}{\tau^2} \left[\left(X - (X^{n+\frac{1}{4}} + \tfrac{3\tau}{4} V^{n+\frac{1}{4}}) \right) - \frac{5}{8} \left(X - (X^n + (\tfrac{\tau}{3} V^n + \tfrac{2\tau}{3} V^{n+\frac{1}{4}})) \right) \right] = \mathbf{f}[X],
$$

$$
V^{n+1} = 8(\tfrac{3}{4} V^{n+\frac{1}{4}}) - 5(\tfrac{1}{3} V^n + \tfrac{2}{3} V^{n+\frac{1}{4}}) + \tfrac{\tau}{3} \mathbf{f}[X^{n+1}].
$$

The first of these is solved by a minimization problem similar to (1.23) and (2.9) above. The left-hand side is then substituted for $\mathbf{f}[X^{n+1}]$ in the second equation. We leave the details of spatial discretization and energy dissipation (through a choice of new velocities that dissipate the maximum amount of energy without changing the convected distribution of mass) to the reader as they are similar to the BDF2 case. The generalization to higher order multi-step and diagonally implicit Runge-Kutta methods is also straightforward, although we have not tested them.

2.3. **Implementation details.** The schemes of Definitions 2.1 and 2.4 are easy to implement once a solver for convex optimization is available. We experimented with our own implementation of a trust-region method described below, and with the convex optimization solvers cvxopt [5] and Knitro [12]. In order to illustrate the key issues, let us consider the problem of minimizing the functional

$$
F(\mathbf{z}) := \sum_{i=1}^{N} \frac{3}{4\tau^2} m|z_i - \hat{x}_i^n|^2 + \sum_{i=1}^{N-1} U\left(\frac{m}{z_{i+1} - z_i} \right)(z_{i+1} - z_i) \tag{2.14}
$$

among all $\mathbf{z} \in \mathbb{R}^N$ such that $z_{i+1} > z_i$ for all i. The first term in (2.14) penalizes the difference between the positions z_i and the intermediate positions $\hat{x}_i^n$, and since $1/\tau^2$

is very large for small timesteps, it is natural to choose the intermediate positions as the starting vector for the minimization process. To have a bounded internal energy, we spread the $\hat{x}_i^n$ slightly apart so that the minimal distance between neighboring positions of the starting vector is bigger than some number $dmin$. We define

$$z_i^{(0)} := \hat{x}_i^n + \tfrac{1}{2}\big(l_i + r_i\big), \tag{2.15}$$

where the negative displacements l_i are built up by

> **for** $i = 2$ **to** N
>> **for** $j = i - 1$ **downto** 1
>>> $d = \hat{x}_{j+1}^n + l_{j+1} - (\hat{x}_j^n + l_j) - dmin$
>>> **if** $d < 0$
>>>> $l_j = l_j + d$ (decrease l_j)
>>> **else**
>>>> **break**   (continue with outer loop)

and a similar algorithm is used to obtain the r_i. Note that this procedure preserves the ordering. In practice, only a few points $\hat{x}_i^n$ need to be adjusted and most of the resulting positions $z_i^{(0)}$ are equal to $\hat{x}_i^n$. We use $dmin := \tfrac{1}{2}\min_i(x_{i+1}^n - x_i^n)$.

From this initial guess, we proceed with a trust region Newton method to minimize the functional $F(\mathbf{z})$. Our implementation is based on Chapter 4 of [14]. For a given $\mathbf{z} \in \mathbb{R}^N$, the gradient $g = \nabla_{\mathbf{z}} F(\mathbf{z})$ and Hessian $H = \nabla_{\mathbf{z}}^2 F(\mathbf{z})$ are trivial to compute for the first term in (2.14), and are easily assembled element by element (one interval $[z_i, z_{i+1})$ at a time) for the second term. The Hessian is positive definite (as it is strictly diagonally dominant) and tridiagonal. At each step of the minimization algorithm, we seek the minimizer of the quadratic approximation

$$Q(\mathbf{p}) := F(\mathbf{z}) + g^T\mathbf{p} + \tfrac{1}{2}\mathbf{p}^T H\mathbf{p} \approx F(\mathbf{z} + \mathbf{p})$$

over all $\mathbf{p} \in \mathbb{R}^d$ satisfying the trust region constraint $\|D^{-1}\mathbf{p}\| \leqslant \Delta$. Here D is the diagonal scaling matrix with entries $D_{ii} := \tfrac{1}{3}\min\big\{z_i - z_{i-1}, z_{i+1} - z_i\big\}$ that prevents the positions z_i from crossing. The trust region radius Δ is not allowed to exceed one. By defining $\hat{\mathbf{p}} := D^{-1}\mathbf{p}$, we map the elliptical trust region to a sphere. We then consider the following modified problem: Minimize the functional

$$\hat{Q}(\hat{\mathbf{p}}) := F(\mathbf{z}) + \hat{g}^T\hat{\mathbf{p}} + \tfrac{1}{2}\hat{\mathbf{p}}^T\hat{H}\hat{\mathbf{p}}, \tag{2.16}$$

over all $\hat{\mathbf{p}} \in \mathbb{R}^N$ with $\|\hat{\mathbf{p}}\| \leqslant \Delta$. Here $\hat{g} := Dg$ and $\hat{H} := DHD$. Notice that this minimization problem reduces to the previous one if $D = \mathrm{id}$.

The main theorem governing the design of trust region methods is

Theorem 2.8 (Trust region optimality criterion). *The vector* $\mathbf{p}_* \in \mathbb{R}^N$ *minimizes the quadratic functional* (2.16) *over all* $\hat{\mathbf{p}} \in \mathbb{R}^N$ *with* $\|\hat{\mathbf{p}}\| \leqslant \Delta$ *if and only if* $\mathbf{p}_*$ *is feasible and there exists a scalar* $\lambda_* \geqslant 0$ *with the following properties:*

$$(\hat{H} + \lambda_* \,\mathrm{id})\mathbf{p}_* = -\hat{g}, \tag{2.17}$$

$$\lambda_*(\Delta - \|\mathbf{p}_*\|) = 0, \tag{2.18}$$

$$(\hat{H} + \lambda_* \,\mathrm{id}) \ \textit{is positive semidefinite.} \tag{2.19}$$

In our case, the matrix $\hat{H}$ is positive definite, so (2.19) is satisfied automatically. Moreover, equation (2.17) has a solution for any $\lambda_* \geqslant 0$, which eliminates the need to search for the most negative eigenvalue of $\hat{H}$ and also rules out the "hard case"

of Moré and Sorensen; see [14]. We use a variant of the Cholesky factorization algorithm to find the Lagrange multiplier λ_*. In pseudo-code it is given as

> Given $\lambda^{(0)}$, $\Delta > 0$
> **for** $k = 0, 1, 2, \ldots$
> > Factorize $\hat{H} + \lambda^{(k)}\text{id} = LL^T$
> > Solve $LL^T p_k = -\hat{g}$ and $Lq_k = p_k$
> > Define $\lambda^{(k+1)} = \lambda^{(k)} + \left(\dfrac{\|p_k\|}{\|q_k\|}\right)\left(\dfrac{\|p_k\| - \Delta}{\Delta}\right)$
> > **if** $\lambda^{(k)} = 0$ and $\lambda^{(k+1)} \leq 0$
> > > **return**
> > **if** $|(\|p_k\| - \Delta)/\Delta| < 10^{-12}$
> > > compute p_{k+1}
> > > **return**
> > **if** $\lambda^{(k+1)} < 0$
> > > set $\lambda^{(k+1)} = 0$
> **end for**

This algorithm is equivalent to Newton's method for finding the zeros of the function $\Delta^{-1} - \|p(\lambda)\|^{-1}$, which usually converges to machine precision in 4–6 iterations. We vary the size Δ of the trust region in the standard way [14] by comparing the actual reduction of the function to the predicted reduction of the quadratic model. For any given Δ, we use the Lagrange multiplier λ_* of the previous trust region step as starting value $\lambda^{(0)}$ for the iteration above. Since we use the exact Hessian, the method converges quadratically (usually requiring 4-20 trust region steps). Moreover, as $\hat{H}$ is tridiagonal, the Cholesky factorization is of complexity $O(N)$ only, and so the convex optimization procedure can be solved very efficiently.

The minimization problem for the functional (2.9) is almost identical to the one we just described. The first term in (2.14) must be replaced by

$$\sum_{i,j=0}^{N}\left(\frac{3}{2\tau^2}(z_i - x_i')A_{ij}(z_j - x_j') - \frac{3}{8\tau^2}(z_i - x_i'')A_{ij}(z_j - x_j'')\right).$$

The Hessian matrix for this part of the objective function is $\frac{9}{4\tau^2}A$, which is again positive definite and tridiagonal, and so the problem can be solved efficiently.

3. Numerical Experiments

In this section, we report on numerical experiments we performed with the variational particle methods VPS1 and VPS2 introduced in Definitions 2.1 and 2.4. We studied both the porous medium and heat equations, as well as the isentropic and isothermal Euler equations, for various choices of parameters and initial data.

3.1. **Porous Medium and Heat Equations.** For both the porous medium equation (1.7) and the heat equation (1.11), solutions converge to self-similar profiles as $t \to \infty$. These profiles can be computed explicitly, which makes it possible to measure the error of the numerical approximation exactly.

3.1.1. *Porous Medium Equation.* In the case of the porous medium equation the self-similar solution are called Barenblatt profiles and given by the formula

$$\varrho_*(t, x) = t^{-\alpha}\left(C^2 - kt^{-2\beta}|x|^2\right)_+^{1/(\gamma-1)}, \tag{3.1}$$

where $(s)_+ := \max\{s, 0\}$, and

$$\alpha = \frac{d}{d(\gamma - 1) + 2}, \quad \beta = \frac{\alpha}{d}, \quad k = \frac{\beta(\gamma - 1)}{2\gamma}.$$

see [16]. The constant C is chosen in such a way that $M(t) := \int_{\mathbb{R}^d} \varrho_*(t, x)\, dx$ equals the total mass. Notice that $M(t)$ is in fact independent of $t > 0$, and for simplicity we assume that $M(t) = 1$. Then $\varrho_*(t, \cdot)$ converges to a Dirac measure as $t \to 0$. In the one-dimensional case under consideration, the constants simplify to

$$\alpha = \frac{1}{\gamma + 1}, \quad \beta = \alpha, \quad k = \frac{\gamma - 1}{2\gamma(\gamma + 1)},$$

and

$$C = \left(\left(\frac{\gamma - 1}{2\pi\gamma(\gamma + 1)}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{3}{2} + \frac{1}{\gamma-1}\right)}{\Gamma\left(\frac{\gamma}{\gamma-1}\right)}\right)^{\frac{\gamma-1}{\gamma+1}},$$

where $\Gamma(z) := \int_0^\infty t^{z-1}e^{-t}\, dt$ for all $z \geqslant 0$. To obtain an approximate solution of the porous medium equation (1.7) we used the variational particle schemes VPS1 and VPS2 of the previous section with internal energy $U(\varrho) = \varrho^\gamma/(\gamma - 1)$.

Figure 1 shows the solution of (1.7) with $\gamma = 5/3$ at different times. We started out with asymmetric and discontinuous initial data

$$\bar{\varrho}(x) := \begin{cases} 0.5 & \text{if } x \in (-1, 0), \\ 0.25 & \text{if } x \in (0, 2), \\ 0 & \text{otherwise.} \end{cases}$$

We used the VPS1 scheme with $N = 1000$ points and $\tau = 0.0016$ to generate these plots. The running time was 8 seconds on a 2.4 GHZ laptop. Note that the solution at time $T = 1.6$ is close to a (shifted) Barenblatt profile.

To estimate the accuracy of the VPS1 scheme, we used initial data

$$\bar{\varrho}(x) := \begin{cases} 50 & \text{if } x \in (-0.01, 0.01), \\ 0 & \text{otherwise,} \end{cases} \tag{3.2}$$

which approximates a Dirac mass, and computed the solution at time $T = 10$ for both the porous medium equation (1.7) with different values of $\gamma > 1$, and for the heat equation, which corresponds to the limiting case $\gamma = 1$. The particle mass was chosen as $m = 0.001$ and the timestep $\tau = 0.01$. The results are shown in Figure 2. Note that the solution is not continuously differentiable for $\gamma \geqslant 2$: The contact angle is strictly positive for $\gamma = 2$, is vertical for $\gamma > 2$, and vanishes for $\gamma < 2$. For the heat equation, the self-similar solution is given by a Gauss function

$$\varrho_*(t, x) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{|x|^2}{4t}\right). \tag{3.3}$$
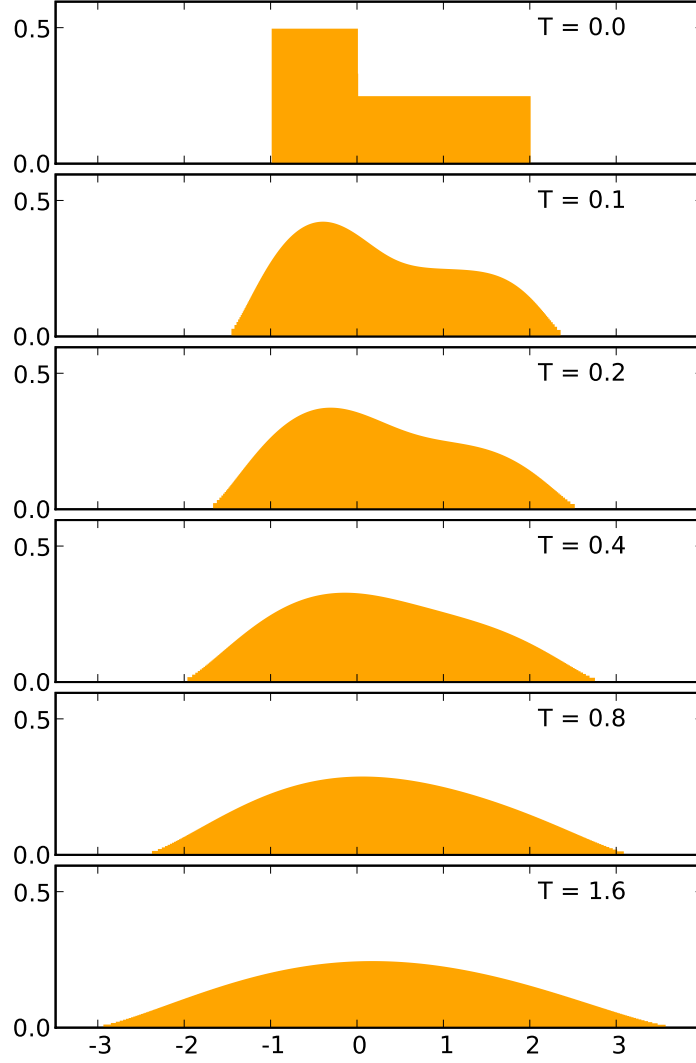
FIGURE 1. Porous medium equation with $\gamma = 5/3$.

The following table shows the $\mathscr{L}^\infty$-error of the approximate solution:

Problem	$\mathscr{L}^\infty$-error
heat eq.	9.45e-5
$\gamma = 5/3$	1.53e-4
$\gamma = 3$	8.63e-4
$\gamma = 5$	2.62e-3

This error is computed, with $n := T/\tau$, according to the formula

$$\mathscr{L}^\infty\text{-error} := \max_{1 \leqslant i < N} \left| \frac{m}{x_{i+1}^n - x_i^n} - \varrho_* \left(10, \tfrac{1}{2}(x_i^n + x_{i+1}^n) \right) \right|. \tag{3.4}$$
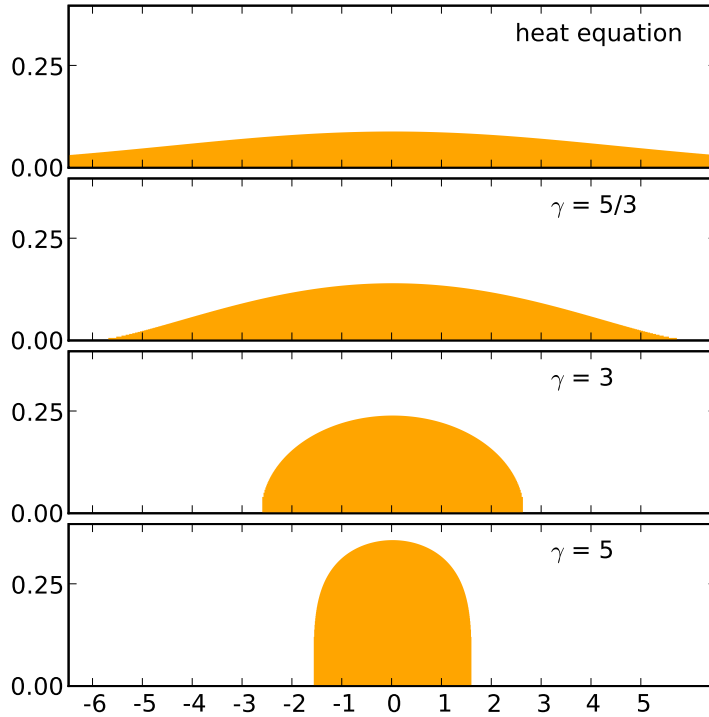
FIGURE 2. Porous medium/heat equation at time $T = 10.0$.

Since the Barenblatt profile has steep gradients for large exponents γ, the infinity norm of the error grows as we increase γ. On the other hand, the support of the solution of (1.7) decreases as we increase γ. For the heat equation, which has an infinite speed of propagation, the support is unbounded. The two effects largely cancel each other out when we estimate the $\mathscr{L}^1$-error as follows:

Problem	$\mathscr{L}^1$-error
heat eq.	1.11e-3
$\gamma = 5/3$	1.08e-3
$\gamma = 3$	9.62e-4
$\gamma = 5$	7.26e-4

Here the error is computed, with $n := T/\tau$, according to the formula

$$\mathscr{L}^1\text{-error} := \sum_{i=1}^{N-1} \left| \frac{m}{x_{i+1}^n - x_i^n} - \varrho_*\left(10, \tfrac{1}{2}(x_i^n + x_{i+1}^n)\right) \right| (x_{i+1}^n - x_i^n). \qquad (3.5)$$

To compute the convergence rate of the VPS1 method, we computed the solution at time $T = 10$ of the porous medium equation with initial data (3.2) and $\gamma = 5/3$,

for several choices of particle mass m and timestep τ. We have

m	τ	Error at $x = 0$	Rate	$\mathscr{L}^\infty$-error	Rate	$\mathscr{L}^1$-error	Rate
0.01	0.1	1.15e-3		1.15e-3		6.79e-3	
0.004	0.04	5.21e-4	0.86	5.21e-4	0.86	3.35e-3	0.77
0.001	0.01	1.53e-4	0.88	1.53e-5	0.88	1.08e-3	0.82
0.0004	0.004	6.73e-5	0.90	6.73e-5	0.90	4.98e-4	0.84
0.0001	0.001	1.90e-5	0.91	3.34e-5	0.50	1.47e-4	0.88

Thus, the VPS1 method performs slightly worse than a first order method for these initial conditions. The last data point for the $\mathscr{L}^\infty$-norm is anomalous because the location of the largest error jumps from the center $x = 0$ to the boundary of the support of the solution, which is not a smooth transition.

In this experiment, the initial data was rather singular. To estimate the effect of this irregularity, we also performed a computation starting off from the Barenblatt profile at time $t = 1$, which is much more regular, and continued to $t = 2$. More specifically, we partitioned the support $[-C/\sqrt{k}, C/\sqrt{k}]$ of $\varrho_*(1, \cdot)$ into intervals $I_i = [\tilde{x}_{i-1}, \tilde{x}_i]$ such that $\int_{I_i} \varrho_*(1, x)\, dx = 1/N$ for all $1 \leqslant i \leqslant N$, and define x_i to be the center of mass of $\varrho_*(1, \cdot)$ over I_i. Again we used $\gamma = 5/3$. We have

m	τ	Error at $x = 0$	Rate	$\mathscr{L}^\infty$-error	Rate	$\mathscr{L}^1$-error	Rate
0.01	0.05	4.32e-4		7.99e-4		1.58e-3	
0.004	0.02	1.88e-4	0.91	5.69e-4	0.37	7.88e-4	0.76
0.001	0.005	5.00e-5	0.96	2.89e-4	0.49	2.38e-4	0.86
0.0004	0.002	2.04e-5	0.98	1.76e-4	0.54	1.03e-4	0.91
0.0001	0.0005	5.17e-6	0.99	8.03e-5	0.57	2.80e-5	0.94

In the interior of the support of the solution, the method converges at first order, but near the boundary the order deteriorates. The $\mathscr{L}^1$-error involves a mixture of the two convergence rates, and is therefore slightly worse than first order.

We also tested the convergence rates of the VPS2 method. For initial data of the form (3.2) with $\gamma = 5/3$, we chose $N + 1$ knots x_i uniformly spaced over the interval $[-0.01, 0.01]$ and set $m_i = 1/N$ for all $1 \leqslant i \leqslant N$. The errors are smaller than they were for VPS1, but the method still converges at first order only due to the discontinuity in the initial density. We have

m	τ	Error at $x = 0$	Rate	$\mathscr{L}^\infty$-error	Rate	$\mathscr{L}^1$-error	Rate
0.01	0.1	1.13e-4		1.34e-3		2.34e-3	
0.004	0.04	4.37e-5	1.04	8.46e-4	0.50	1.02e-3	0.90
0.001	0.01	1.06e-5	1.02	3.97e-4	0.55	2.84e-4	0.93
0.0004	0.004	4.18e-6	1.01	2.36e-4	0.57	1.19e-4	0.95
0.0001	0.001	1.03e-6	1.01	1.05e-4	0.58	3.12e-5	0.97

To estimate the influence of the singularity of the initial data, we tried the VPS2 method starting again off from the Barenblatt profile at time $t = 1$ and continuing to $t = 2$. We partitioned the support $[-C/\sqrt{k}, C/\sqrt{k}] =: [a, b]$ into equal subintervals with knots $x_i = a + i(b - a)/N$ for all $1 \leqslant i \leqslant N$ and set

$$m_i := \int_{x_{i-1}}^{x_i} \varrho_*(1, x)\, dx \quad \text{for all } 1 \leqslant i \leqslant N. \tag{3.6}$$

We used $\gamma = 5/3$. When comparing the result to $\varrho_*(2, \cdot)$, we have

N	τ	Error at $x = 0$	Rate	$\mathscr{L}^\infty$-error	Rate	$\mathscr{L}^1$-error	Rate
100	0.1	1.64e-5		3.44e-4		1.02e-4	
250	0.04	2.64e-6	1.99	9.82e-5	1.37	1.71e-5	1.95
1000	0.01	1.66e-7	2.00	1.32e-5	1.45	1.10e-6	1.98
2500	0.004	2.66e-8	2.00	3.40e-6	1.48	1.78e-7	1.99
10000	0.001	1.66e-9	2.00	4.30e-7	1.49	1.13e-8	1.99

As expected, we obtain second order convergence throughout most of the interval when we start with continuously differentiable initial data. A few points next to the boundary of the support, however, seem to converge at a lower rate of $3/2$. This raises the question of what happens if the initial data is continuous but not $\mathscr{C}^1$. The Barenblatt profile with $\gamma = 3$ has a square-root singularity at the edges of its support. Repeating the above procedure for this case, we obtain

N	τ	Error at $x = 0$	Rate	$\mathscr{L}^\infty$-error	Rate	$\mathscr{L}^1$-error	Rate
100	0.1	2.36e-5		5.49e-4		1.22e-4	
250	0.04	3.78e-6	2.00	3.43e-4	0.51	3.16e-5	1.48
1000	0.01	2.37e-7	2.00	1.67e-4	0.52	4.02e-6	1.49
2500	0.004	3.80e-8	2.00	1.04e-4	0.51	1.02e-6	1.50
10000	0.001	2.37e-9	2.00	5.17e-5	0.51	1.28e-7	1.50

It appears that the infinite slope of the exact solution near the boundary of the support creates large errors there that dominate the overall $\mathscr{L}^1$-error of the method. The situation can be improved if we redistribute the mass to better resolve the solution near these endpoints. Specifically, we choose the knot positions as

$$x_i := f\left(-1 + \frac{2i}{N}\right)\frac{C}{\sqrt{k}} \quad \text{for all } 0 \leqslant i \leqslant N,$$

with weight function

$$f(x) := \frac{\displaystyle\int_0^x \sqrt{1-y^2}\, dy}{\displaystyle\int_0^1 \sqrt{1-y^2}\, dy} \quad \text{for all } x \in [-1, 1],$$

and we assign to each interval $[x_{i-1}, x_i)$ the mass m_i defined in (3.6). We have

N	τ	Error at $x = 0$	Rate	$\mathscr{L}^\infty$-error	Rate	$\mathscr{L}^1$-error	Rate
100	0.1	2.36e-5		2.84e-3		1.20e-4	
250	0.04	3.79e-6	2.0	1.58e-3	0.64	2.22e-5	1.85
1000	0.01	2.38e-7	2.0	6.07e-4	0.69	1.62e-6	1.89
2500	0.004	3.80e-8	2.0	3.14e-4	0.72	2.79e-7	1.92
10000	0.001	2.38e-9	2.0	1.14e-4	0.73	1.89e-8	1.94

In spite of the singularity in slope near the endpoints of the support, we achieve second order accuracy over enough of the interval to converge at second order in the $\mathscr{L}^1$-norm. The $\mathscr{L}^\infty$-norm is larger because the first and last intervals are smaller than the interior intervals, so their midpoints are closer to the singularity. The

first function we tried for the mass redistribution, namely $f(x) = \frac{3}{2}x - \frac{1}{2}x^3$, caused difficulties for the convex optimization solver for $N = 10000$.

3.1.2. *Heat Equation.* Both schemes perform similarly for the heat equation (1.11) as they did for the porous medium equation. Here we report only the VPS2 results. Using the internal energy $U(\varrho) = \varrho \log \varrho$ and initial data (3.2), and comparing the solution at $t = 10$ to the Gaussian $\varrho_*(10, \cdot)$ defined in (3.3), we find

m	τ	Error at $x = 0$	Rate	$\mathscr{L}^\infty$-error	Rate	$\mathscr{L}^1$-error	Rate
0.01	0.1	1.42e-4		5.86e-4		5.12e-3	
0.004	0.04	5.42e-5	1.05	2.73e-4	0.83	2.24e-3	0.90
0.001	0.01	1.27e-5	1.05	8.07e-5	0.88	6.13e-4	0.93
0.0004	0.004	4.78e-6	1.06	3.52e-5	0.91	2.54e-4	0.96
0.0001	0.001	1.02e-6	1.12	9.78e-6	0.92	6.47e-5	0.99

As before, we obtain first order convergence for the VPS2 scheme for discontinuous initial data. Note that we could obtain a more precise error estimate by comparing the numerical solution to the exact solution of (1.11) with data (3.2), which can be computed explicitly. But the $\mathscr{L}^\infty$-distance between this exact solution and $\varrho_*(10, \cdot)$ is around 7.4×10^{-8}, which is small in comparison to the numerical errors.

Finally, we use the VPS2 scheme to evolve the heat kernel $\varrho_*(1, \cdot)$ to time $t = 2$ via the heat equation (1.11). As the support of $\varrho_*(1, \cdot)$ is infinite, instead of fixing the knot positions first, we start by distributing the mass according to

$$m_i := \frac{f\left(\dfrac{i}{N+1}\right)}{\displaystyle\sum_{i=1}^{N} f\left(\dfrac{i}{N+1}\right)} \quad \text{for all } 1 \leqslant i \leqslant N,$$

with weight function

$$f(x) := q(x)q(1-x) \quad \text{and} \quad q(x) := 10x^2 + x/10.$$

We then initialize the knot positions x_i^0 for all $1 \leqslant i \leqslant N - 1$ as

$$x_i^0 := -2\,\mathrm{erfc}^{-1}(2s_i) \quad \text{with} \quad s_i := \sum_{j=1}^{i} m_j.$$

This construction implies that $m_i = \int_{x_{i-1}}^{x_i} \varrho_*(1, x)\,dx$ for all $2 \leq i \leq N - 1$. Since the mass of the first and last interval are small, any reasonable choice of x_0 and x_N works well. We used $x_0^0 := 3x_1^0 - 2x_2^0$ and $x_N^0 := 3x_{N-1}^0 - 2x_{N-2}^0$. We have

N	τ	Error at $x = 0$	Rate	$\mathscr{L}^\infty$-error	Rate	$\mathscr{L}^1$-error	Rate
100	0.1	7.02e-6		1.87e-5		1.82e-4	
250	0.04	1.10e-6	2.02	3.04e-6	1.98	3.13e-5	1.92
1000	0.01	6.82e-8	2.01	1.91e-7	2.00	2.10e-6	1.95
2500	0.004	1.09e-8	2.00	3.05e-8	2.00	3.50e-7	1.96
10000	0.001	6.80e-10	2.00	1.90e-9	2.00	2.31e-8	1.96

Thus, for the heat equation, even the $\mathscr{L}^\infty$-norm converges at second order as there is no singularity in the slope of the exact solution over the course of this simulation.

The $\mathscr{L}^1$-norm converges slightly slower than second order due to the fact that as we add points, the support of the numerical solution grows.

In conclusion, we have found our variational particle scheme to be an effective method of solving the porous medium and heat equations. Moreover, the second order version of the method does indeed give second order accuracy as long as the initial conditions are not too singular.

3.2. **Isentropic Euler Equations.** As few smooth solutions of the isentropic Euler equations (2.1) are known explicitly, we test our numerical methods for self-consistency using smooth initial data for a short enough time that shocks do not occur. To study their ability to capture shocks and rarefaction waves, we test our schemes using piecewise constant initial data, for which the Riemann problems can be solved analytically. In general, we find that shocks cause the second order method to perform as a first order method, and the first order method to perform slightly worse. We mention in passing that these schemes also work well for the isothermal equations, but we did not include examples for the sake of brevity.

3.2.1. *Smooth solutions.* We begin by testing the convergence rates of the VPS1 and VPS2 schemes for the isentropic Euler equations with initial data

$$\bar{\varrho}(x) = \frac{3}{8}\left(1 - \frac{x^2}{4}\right)_+ \quad \text{and} \quad \bar{u}(x) = 0 \quad \text{for all } x \in \mathbb{R},$$

for $\gamma = 5$ and $\gamma = 7/5$. The solutions are shown in Figure 3 at times $t = 0, 1, 2, 3, 4$. For the VPS1 scheme, we choose initial positions x_i^0 such that

$$\int_{-2}^{x_i^0} \bar{\varrho}(x)\,dx = \frac{i - 1/2}{N} \quad \text{for all } 1 \leqslant i \leqslant N.$$

For the VPS2 scheme with $\gamma = 5$, we choose initial position x_i^0 uniformly spaced in the support $[-2, 2]$ of the initial density, and define masses

$$m_i := \int_{x_{i-1}^0}^{x_i^0} \bar{\varrho}(x, 0)\,dx \quad \text{for all } 1 \leqslant i \leqslant N. \tag{3.7}$$

For the VPS2 scheme with $\gamma = 7/5$, it is necessary to distribute more points near the ends of the support to achieve full second order accuracy. We found that

$$x_i := 2f(-1 + 2i/N) \quad \text{for all } 0 \leqslant i \leqslant N,$$

with weight function

$$f(x) := \frac{\displaystyle\int_0^x \sqrt{1 - y^2}\,dy}{\displaystyle\int_0^1 \sqrt{1 - y^2}\,dy} \quad \text{for all } x \in [-1, 1],$$
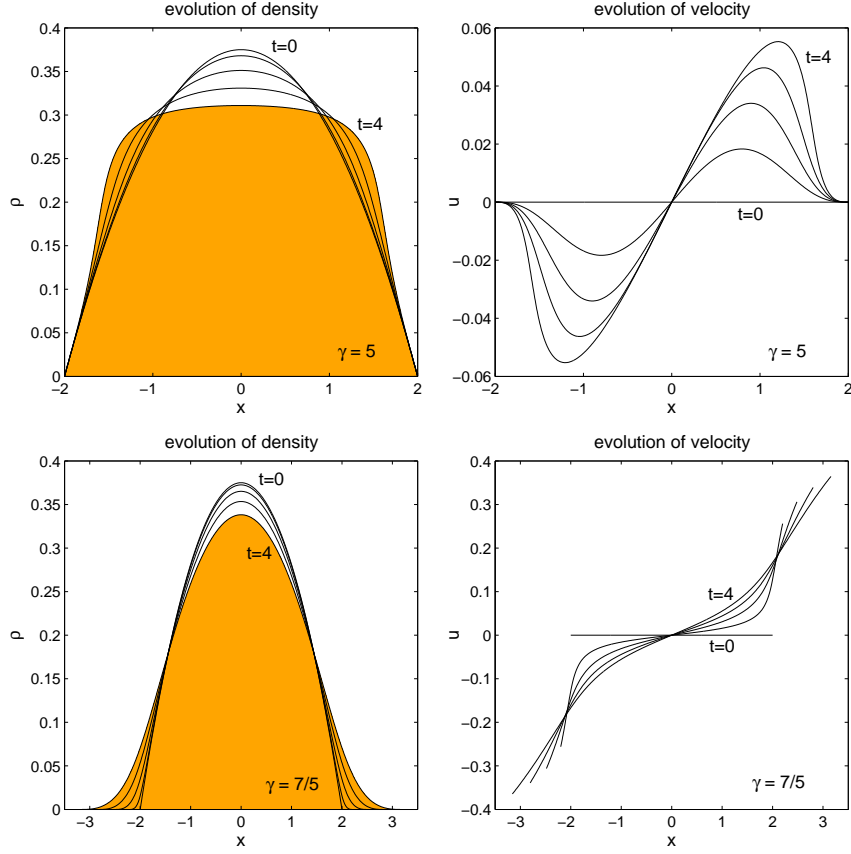
FIGURE 3. Solution of isentropic Euler equations with a parabolic initial distribution of mass, starting from rest. ($N = 800$, VPS2)

and with m_i defined in (3.7) works well. We used the VPS2 solution with $N = 4000$ and $\tau = 0.004$ as the "exact" solution to compute the errors for both schemes.

N	τ	$\gamma = 5$		$\gamma = 7/5$	
		VPS1	VPS2	VPS1	VPS2
50	0.4	1.12e-2	1.46e-3	7.75e-3	3.07e-3
100	0.2	8.83e-3	3.90e-4	4.68e-3	8.38e-4
200	0.1	4.57e-3	1.01e-4	2.51e-3	2.18e-4
400	0.05	2.37e-3	2.61e-5	1.24e-3	5.54e-5
800	0.025	1.19e-3	6.58e-6	5.84e-4	1.37e-5
1600	0.0125	5.96e-4	1.54e-6	3.08e-4	3.14e-6

The errors reported for the VPS1 scheme are estimates of the $\mathscr{L}^\infty$-error of ϱ alone, measured at midpoints of the numerical solution against the "exact" solution, interpolated linearly between its own midpoints. The errors reported for the VPS2 scheme involve the Wasserstein distance of the densities and the difference in velocities. Notice that the finiteness of the total energies only implies that

$$u^n \in \mathscr{L}^2(\mathbb{R}, \varrho^n) \quad \text{and} \quad u^{\text{exact}} \in \mathscr{L}^2(\mathbb{R}^d, \varrho^{\text{exact}}),$$

so in general the two velocities are in different spaces. Let $\mathbf{r}^n$ and $\mathbf{r}^{\mathrm{exact}}$ denote the inverse distribution functions of the measures $\varrho^n \mathfrak{L}^1$ and $\varrho^{\mathrm{exact}} \mathfrak{L}^1$ resp. As explained in Section 2.2, we then have $\mathbf{r}^n \# (\mathbf{1}_{[0,1]} \mathfrak{L}^1) = \varrho^n \mathfrak{L}^1$, which implies that

$$\int_{\mathbb{R}} |u^n|^2 \varrho^n \, dx = \int_{\mathbb{R}} |u^n|^2 \left(\mathbf{r}^n \# (\mathbf{1}_{[0,1]} \mathfrak{L}^1) \right)(dx) = \int_{[0,1]} |u^n \circ \mathbf{r}^n|^2 \, ds.$$

That is, the pull-back $u \circ \mathbf{r}^n$ of the velocity field u^n under the map $\mathbf{r}^n$ is an element in $\mathscr{L}^2([0,1])$. The same is true for $u^{\mathrm{exact}} \circ \mathbf{r}^{\mathrm{exact}}$. We therefore define

$$E_{\mathbf{W}} := \sqrt{\mathbf{W}\big(\varrho^n, \varrho^{\mathrm{exact}}(\cdot, t^n)\big)^2 + \frac{1}{2} \int_{[0,1]} \left| u^n \circ \mathbf{r}^n - u^{\mathrm{exact}} \circ \mathbf{r}^{\mathrm{exact}}(\cdot, t^n) \right|^2 ds}, \quad (3.8)$$

where $t^n = 4$ is the final time. Since convergence in the Wasserstein space implies only weak* convergence in the sense of mesaures (see [1, 17]), our error estimate is weaker that one that involves the $\mathscr{L}^1$-norm. The $\mathscr{L}^\infty$-error of the densities of the VPS2 solutions are about three times larger than the $E_{\mathbf{W}}$-errors when $\gamma = 5$, and about three times smaller than the $E_{\mathbf{W}}$-errors when $\gamma = 7/5$.

In the following three sections, we study the performance of our scheme on problems involving various combinations of shocks and rarefaction waves emanating from an initial discontinuity at the origin. In the presence of shocks, it no longer makes sense to use the $\mathscr{L}^\infty$-norm to measure the error.

3.2.2. *Shock/Shock.* We use $\gamma = 5/3$ and initial data

$$(\bar{\varrho}, \bar{u}) = \begin{cases} (0.25, 1) & \text{if } x \in (-2, 0), \\ (0.25, 0) & \text{if } x \in (0, 2), \\ (0, 0) & \text{otherwise.} \end{cases} \quad (3.9)$$

Note that this amounts to solving three Riemann problems: one at $x = 0$, and two others at $x = \pm 2$. Figure 4 shows the evolution at different times, computed via a first order version of the VPS2 method (with variable mass

$$m_i := \frac{6}{N} \int_{i-1}^{i} \left(\frac{x}{N} \right) \left(1 - \frac{x}{N} \right) dx \quad \text{for all } 1 \leqslant i \leqslant N,$$

but a first order time-stepper), with $N = 1000$ and timestep $\tau = 0.01$. We will refer to this hybrid method as VPS1a. Since there cannot be any shocks connected to the vacuum, the solution immediately forms two rarefaction waves at the boundaries of the support that grow in width in a self-similar fashion. The Riemann problem at $x = 0$ evolves into a self-similar, double shock structure with constant intermediate state $(\varrho_m, u_m) = (1.16641, 0.5)$ with large density. For simplicity, we only plot the density here, omitting the velocity. As time moves on, the different waves eventually interact and the shock strengths decrease. At time $t = 7$, the solution has developed into a continuous profile (not unlike the Barenblatt profiles of the porous medium equation), which seems to evolve in a self-similar way while moving to the right with a constant background speed.

The VPS1 and VPS2 schemes give similar results to this hybrid method, but the former is less accurate inside the rarefaction waves (due to lack of resolution) while the latter suffers from small oscillations near shocks. These oscillations are rather interesting as they do not grow exponentially in time nor prevent the scheme from converging as we refine the mesh; see Figure 5. Instead, they arise because once we
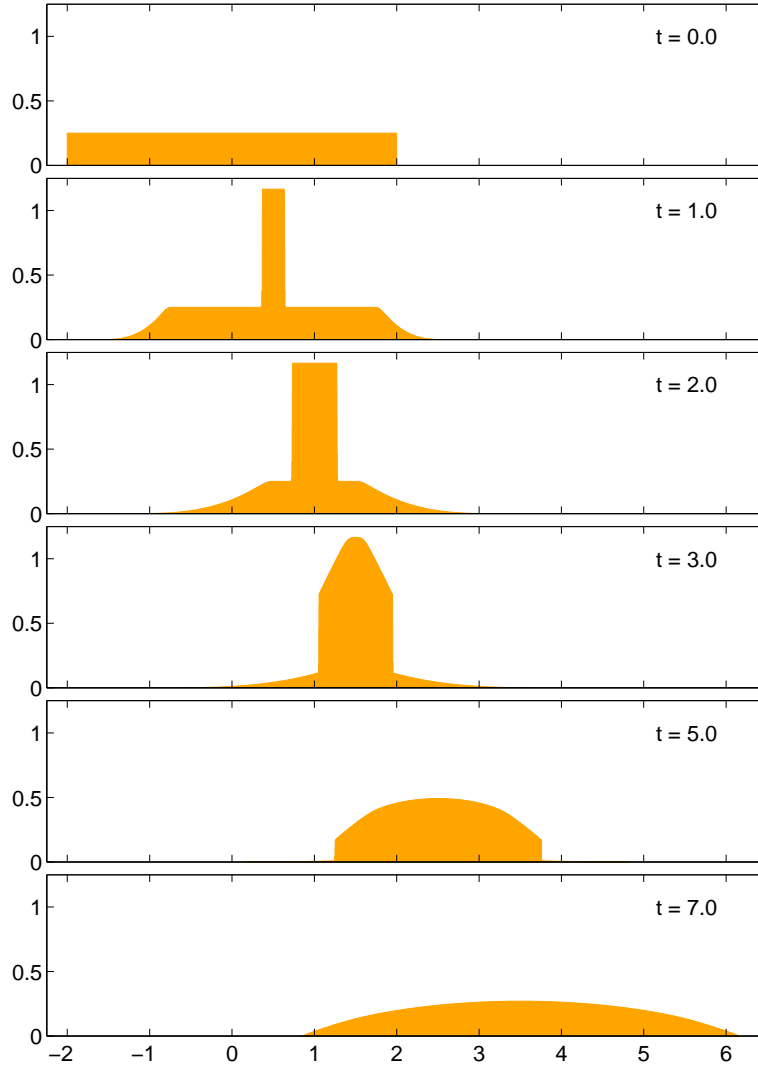
FIGURE 4. Isentropic Euler equations with $\gamma = 5/3$.

choose a spatial discretization (N together with the mass m_i of each interval), the time-stepper is effectively solving a coupled system of masses connected by springs. Indeed, except for the step in which mass is re-distributed when the particles cross, we are solving Lagrange's equations of motion for the particle trajectories $x_i(t)$, with the kinetic and potential energy of the system defined in terms of the x_i and $u_i = \dot{x}_i$ in the obvious way (depending on the scheme). If we were to take the time-step τ to zero holding N constant, the particles would never cross during the transport step and the scheme would convert the energy that is supposed to be dissipated by the shock into lattice vibrations. However, by simultaneously refining the mesh as we decrease the time-step (with $\tau \sim N^{-1}$), the backward-Euler and BDF2 methods seem to damp out these vibrations without significantly smoothing
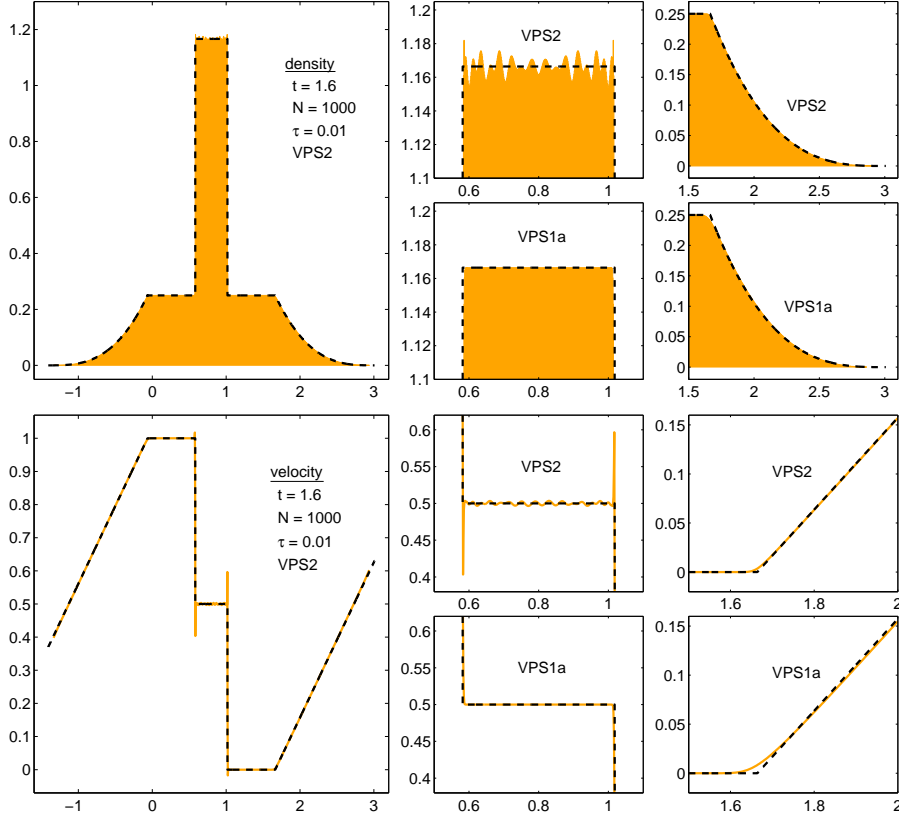
FIGURE 5. Comparison of the VPS2 and VPS1a results with the exact solution of the Riemann problems (dashed lines).

the shocks, which is somewhat amazing. We emphasize that although the VPS1a scheme does not suffer from oscillations for the timestep used in Figure 5, they will eventually appear if we take $\tau \to 0$ holding N constant.

Next we wish to use the exact solution of the Riemann problems to compute the errors of the schemes and check convergence rates. For sufficiently small $t > 0$, the exact density is given by

$$\varrho(x,t) = \begin{cases} \left(\dfrac{(-u_l + \varrho_l^\theta) + (x - x_l)/t}{\theta + 1} \right)^{1/\theta} & \text{if } x \in (x_1, x_2), \\ \varrho_l & \text{if } x \in (x_2, x_3), \\ \varrho_m & \text{if } x \in (x_3, x_4), \\ \varrho_r & \text{if } x \in (x_4, x_5), \\ \left(\dfrac{(u_r + \varrho_r^\theta) - (x - x_r)/t}{\theta + 1} \right)^{1/\theta} & \text{if } x \in (x_5, x_6), \\ 0 & \text{otherwise,} \end{cases}$$
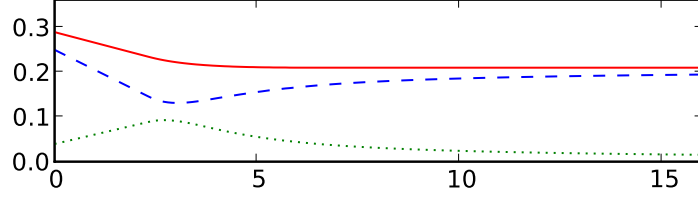
FIGURE 6. Energy (solid: total, dashed: kinetic, dotted: internal).

with constants

$$x_1 = x_l + t(u_l - \varrho_l^\theta), \qquad x_3 = ts_l, \qquad x_5 = x_r + t(u_r - \theta\varrho_r^\theta),$$
$$x_2 = x_l + t(u_l + \theta\varrho_l^\theta), \qquad x_4 = ts_r, \qquad x_6 = x_r + t(u_r + \varrho_r^\theta);$$

see [4]. For the case under consideration, we have

$$x_l = -2, \qquad \varrho_l = 0.25, \qquad u_l = 1, \qquad (3.10)$$
$$x_r = 2, \qquad \varrho_r = 0.25, \qquad u_r = 0, \qquad (3.11)$$

which gives an intermediate density $\varrho_m = 1.16641$, and shock speeds $s_l = 0.36360$ and $s_r = 0.63640$. The exact velocity profile can be computed as well, but we refer the reader to the literature. Using a final time of $T = 1.6$, we obtain

VPS1a									
N	τ	$\mathbf{W}$	Rate	$E_{\mathbf{W}}$	Rate	$\mathscr{L}^1$	Rate	E_{tot}	Rate
100	.1	7.85e-3		3.39e-2		4.60e-2		1.79e-3	
250	.04	4.14e-3	0.70	2.14e-2	0.50	2.18e-2	0.82	7.62e-4	0.93
1000	.01	1.55e-3	0.71	1.03e-2	0.53	6.10e-3	0.92	2.10e-4	0.93
2500	.004	7.81e-4	0.75	5.99e-3	0.60	2.63e-3	0.92	8.87e-5	0.94
10000	.0001	2.66e-4	0.78	2.07e-3	0.77	6.52e-4	1.01	2.33e-5	0.96

VPS2									
N	τ	$\mathbf{W}$	Rate	$E_{\mathbf{W}}$	Rate	$\mathscr{L}^1$	Rate	E_{tot}	Rate
100	.1	3.50e-3		2.28e-2		3.05e-2		8.05e-4	
250	.04	1.27e-3	1.11	1.60e-2	0.38	1.36e-2	0.88	3.77e-4	0.83
1000	.01	2.89e-4	1.07	7.76e-3	0.52	4.48e-3	0.80	1.06e-4	0.92
2500	.004	1.13e-4	1.02	3.67e-3	0.82	1.97e-3	0.90	3.38e-5	1.24
10000	.0001	2.84e-5	1.00	1.29e-3	0.75	3.83e-4	1.18	7.94e-6	1.04

Here $\mathbf{W}$ stands for the Wasserstein distance between the numerical and the exact densities, while $\mathscr{L}^1$ indicates their $\mathscr{L}^1$-difference. The functional $E_{\mathbf{W}}$ was defined in (3.8), and we write $E_{\text{tot}} := |E_{\text{tot}}^n - E_{\text{tot}}^{\text{exact}}|$ for the difference in the total energies. The $\mathscr{L}^\infty$-error of the densities does not approach zero because of the shocks. We see that in spite of the high frequency oscillations, the VPS2 scheme is more accurate than the VPS1a method, with the density converging at first order and the velocity converging somewhat slower. For both schemes, the total energy E_{tot}^n at the final time converges to $E_{\text{tot}}^{\text{exact}}$ from above as $N \to \infty$ and $\tau \to 0$, thus the exact solution dissipates energy (slightly) faster than any of the numerical solutions.

In Figure 6, we plot the energy of the numerical solution (beyond the point that we are able to compute the exact solution). As long as the solution is discontinuous,

the total energy decreases. Moreover, the total energy is a linear function at the beginning of the evolution (up to time $t = 2$, say,) since the strength of the two shocks remains constant. During this time, the internal energy increases along with the width of the intermediate state (ϱ_m, u_m), which carries more energy because of its high density. As the shocks interact with the rarefaction waves at the boundary of the support, their strengths decrease and finally vanish (at about $t = 6.7$), after which the total energy remains essentially constant. This is in agreement with the theory since for smooth solutions of the isentropic Euler equations, the total energy is a conserved quantity. Our schemes capture this behavior remarkably well.

3.2.3. *Shock/Rarefaction.* We also considered the case $\gamma = 5/3$ with initial data

$$(\bar{\varrho}, \bar{u}) = \begin{cases} (0.5, 0) & \text{if } x \in (-1, 0), \\ (0.25, 0) & \text{if } x \in (0, 2), \\ (0, 0) & \text{otherwise.} \end{cases} \tag{3.12}$$

The exact solution of the Riemann problem at $x = 0$ is given by a pattern involving a shock followed a rarefaction wave. Figure 7 shows the approximation at different times. The computation was done using VPS1 with $m = 0.001$ and $\tau = 0.01$.

Figure 8 shows the approximate solution of (2.1) for the same parameters and initial data (3.12), at time $T = 0.6$ for different values of γ. For sufficiently small $t > 0$, the exact density is given by the following formula:

$$\varrho(t, x) = \begin{cases} \left(\dfrac{(-u_l + \varrho_l^\theta) + (x - x_l)/t}{\theta + 1}\right)^{1/\theta} & \text{if } x \in (x_1, x_2), \\ \varrho_l & \text{if } x \in (x_2, x_3), \\ \left(\dfrac{(u_l + \varrho_l^\theta) - x/t}{\theta + 1}\right)^{1/\theta} & \text{if } x \in (x_3, x_4), \\ \varrho_m & \text{if } x \in (x_4, x_5), \\ \varrho_r & \text{if } x \in (x_5, x_6), \\ \left(\dfrac{(u_r + \varrho_r^\theta) - (x - x_r)/t}{\theta + 1}\right)^{1/\theta} & \text{if } x \in (x_6, x_7), \\ 0 & \text{otherwise,} \end{cases} \tag{3.13}$$
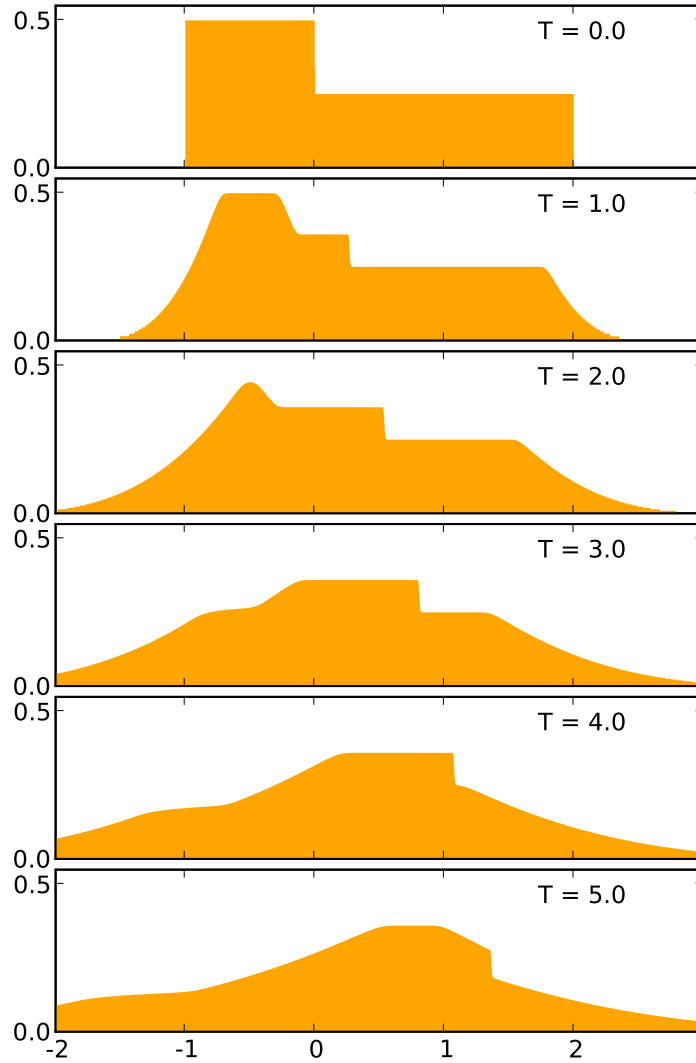
with constants

$$\begin{aligned} x_1 &= x_l + t(u_l - \varrho_l^\theta), & x_3 &= t(u_l - \theta\varrho_l^\theta), & x_6 &= x_r + t(u_r - \theta\varrho_r^\theta), \\ x_2 &= x_l + t(u_l + \theta\varrho_l^\theta), & x_4 &= t(u_m - \theta\varrho_m^\theta), & x_7 &= x_r + t(u_r + \varrho_r^\theta), \\ & & x_5 &= ts_r; \end{aligned}$$

see [4]. For the case under consideration, we have

$$\begin{aligned} x_l &= -1, & \varrho_l &= 0.5, & u_l &= 0, \\ x_r &= 2, & \varrho_r &= 0.25, & u_r &= 0, \end{aligned}$$

which implies an intermediate density $\varrho_m = 0.3601$ and velocity $u_m = 0.0823$, and a shock speed $s_r = 0.3636$. The exact velocity profile can be computed as well, but

FIGURE 7. Isentropic Euler equations with $\gamma = 5/3$.

we refer the reader to the literature. The following table shows the $\mathscr{L}^1$-error:

Problem	$\mathscr{L}^1$-error
$\gamma = 5/3$	4.46e-3
$\gamma = 3$	5.57e-3
$\gamma = 5$	3.87e-3

Notice that the rarefaction waves in (3.13) that connect the profile to the vacuum, are convex for $\gamma < 3$, linear if $\gamma = 3$, and concave if $\gamma > 3$. To get a convergence rate, we computed the approximate solution at time $t = 0.6$ with initial data (3.12)
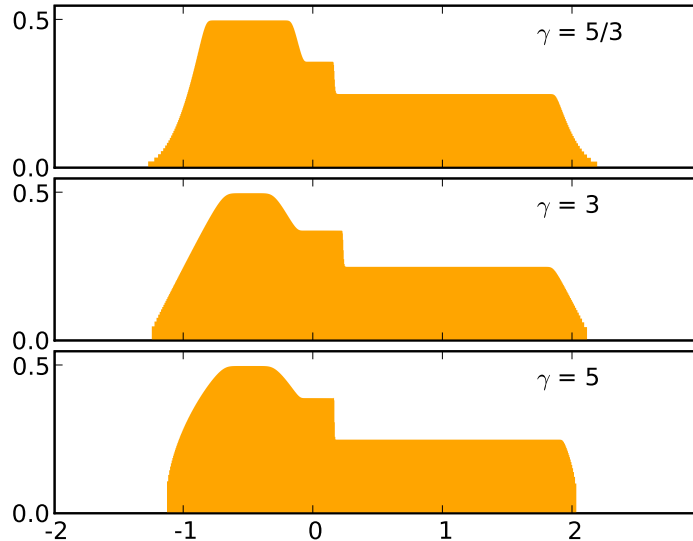
FIGURE 8. Isentropic Euler equations at time $T = 0.6$.

and $\gamma = 5/3$ for different values of particle mass/timestep. We have

m	τ	$\mathscr{L}^1$-Error	Rate
0.01	0.1	2.146e-2	
0.005	0.05	1.397e-2	0.620
0.001	0.01	4.464e-3	0.709
0.0005	0.005	2.615e-3	0.772
0.0001	0.001	7.318e-4	0.791

As with VPS1a in the previous section, the rate is somewhat less than one. We did not study the performance of the VPS1a or VPS2 schemes on this initial data or implement the other measures of error in our VPS1 code.

3.2.4. *Rarefaction/Rarefaction.* Finally, we use the VPS1 method to compute the approximate solution of (2.1) for $\gamma = 5/3$ and initial data

$$(\bar{\varrho}, \bar{u}) = \begin{cases} (0.25, -0.5) & \text{if } x \in (-2, 0), \\ (0.25, 0.5) & \text{if } x \in (0, 2), \\ (0, 0) & \text{otherwise.} \end{cases}$$

Figure 9 shows the result at different times. The fluid splits into two parts that travel in opposite directions. Between the blocks, two rarefaction waves form that are separated by vacuum. In our approximation, the density is defined as particle mass divided by the specific volume, which is the distance between neighboring particles. We therefore do not get a perfect vacuum. As the two blocks move further apart, however, the density between them does approach zero. By construction, the density can never become negative.
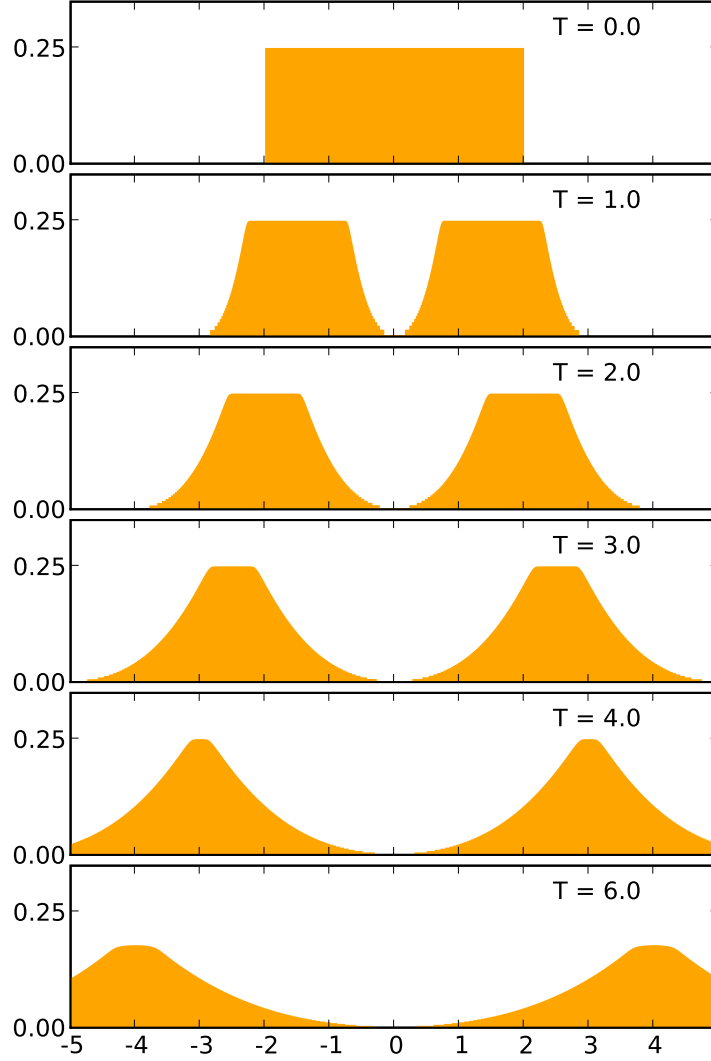
FIGURE 9. Isentropic Euler equations with $\gamma = 5/3$.

APPENDIX A. POROUS MEDIUM EQUATION

We derive the Euler-Lagrange equations for the minimization problem

$$\varrho^{n+1} := \operatorname{argmin} \left\{ \frac{1}{2\tau} \mathbf{W}(\varrho^n, \varrho)^2 + \mathcal{U}[\varrho] : \varrho \in \mathscr{P}(\mathbb{R}^d) \right\} \tag{A.1}$$

to show that it indeed gives an approximation for the porous medium equation. More precisely, we will show that (A.1) can be interpreted as a backward Euler method applied to Darcy's law $\mathbf{u} = -\nabla U'(\varrho)$ in a Lagrangian formulation of the problem. We recall first that if ϱ^n and ϱ are Lebesgue measurable functions, then there exists a Borel map $\mathbf{r} : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ defined ϱ^n-a.e., with the property that

$$\mathbf{W}(\varrho^n, \varrho)^2 = \int_{\mathbb{R}^d} |\mathbf{r}(x) - x|^2 \varrho^n(x) \, dx \quad \text{and} \quad \mathbf{r}\#(\varrho^n \mathfrak{L}^d) = \varrho \mathfrak{L}^d. \qquad (A.2)$$

The map $\mathbf{r}$ is called an optimal transport map. It is the gradient of a lower semi-continuous, convex function and invertible ϱ^n-a.e. The second identity in (A.2) now implies that for all test functions $\phi \in \mathscr{C}_{\mathrm{b}}(\mathbb{R}^d)$ we have

$$\int_{\mathbb{R}^d} \varphi\big(\mathbf{r}(x)\big) \varrho^n(x) \, dx = \int_{\mathbb{R}^d} \varphi(z) \varrho(z) \, dz = \int_{\mathbb{R}^d} \varphi\big(\mathbf{r}(x)\big) \varrho\big(\mathbf{r}(x)\big) \det D\mathbf{r}(x) \, dx.$$

Here we used the change of variables formula, which can be justified because $\mathbf{r}$ is a monotone function. Since $\mathbf{r}$ is invertible ϱ^n-a.e. we conclude that

$$\varrho^n(x) = \varrho\big(\mathbf{r}(x)\big) \det D\mathbf{r}(x) \qquad \text{for a.e. } x \in \mathbb{R}^d,$$

which allows us to express the internal energy $\mathcal{U}[\varrho]$ in the form

$$\mathcal{U}[\varrho] = \int_{\mathbb{R}^d} U\big(\varrho(z)\big) \, dz = \int_{\mathbb{R}^d} U\left(\frac{\varrho^n(x)}{\det D\mathbf{r}(x)}\right) \det D\mathbf{r}(x) \, dx. \qquad (A.3)$$

That is, we can consider the internal energy as a functional of $\mathbf{r}$ instead of ϱ.

Let now ϱ^{n+1} be the minimizer of (1.12) and let $\mathbf{r}^{n+1}$ denote the optimal transport map that pushes $\varrho^n \mathfrak{L}^d$ forward to the measure $\varrho^{n+1} \mathfrak{L}^d$. For any smooth vector field $\zeta \colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and any $\varepsilon \geqslant 0$ we now define the functions

$$\mathbf{r}_\varepsilon := (\mathrm{id} + \varepsilon \zeta) \circ \mathbf{r}^{n+1} \quad \text{and} \quad \varrho_\varepsilon \mathfrak{L}^d := \mathbf{r}_\varepsilon \#(\varrho^n \mathfrak{L}^d).$$

Note that the map $\mathbf{r}_\varepsilon$ is not an optimal transport map if $\varepsilon > 0$, but we have

$$\mathbf{W}(\varrho^n, \varrho_\varepsilon)^2 \leqslant \int_{\mathbb{R}^d} |\mathbf{r}_\varepsilon - \mathrm{id}|^2 \varrho^n \, dx \quad \text{for all } \varepsilon \geqslant 0.$$

This implies the estimate

$$\limsup_{\varepsilon \to 0} \frac{\mathbf{W}(\varrho^n, \varrho_\varepsilon)^2 - \mathbf{W}(\varrho^n, \varrho^{n+1})^2}{\varepsilon} \leqslant 2 \int_{\mathbb{R}^d} (\zeta \circ \mathbf{r}^{n+1}) \cdot (\mathbf{r}^{n+1} - \mathrm{id}) \varrho^n \, dx \qquad (A.4)$$

$$= 2 \int_{\mathbb{R}^d} \zeta \cdot \left((\mathbf{r}^{n+1} - \mathrm{id}) \circ (\mathbf{r}^{n+1})^{-1} \right) \varrho^{n+1} \, dz.$$

On the other hand, a straightforward computation using (1.2) and (A.3) shows that

$$\lim_{\varepsilon \to 0} \frac{\mathcal{U}[\varrho_\varepsilon] - \mathcal{U}[\varrho^{n+1}]}{\varepsilon} = \int_{\mathbb{R}^d} \left\{ -U'\left(\frac{\varrho^n}{\det D\mathbf{r}^{n+1}}\right) \frac{\varrho^n}{\det D\mathbf{r}^{n+1}} + U\left(\frac{\varrho^n}{\det D\mathbf{r}^{n+1}}\right) \right\}$$

$$\times \operatorname{tr}\left((D\zeta) \circ \mathbf{r}^{n+1} \right) \det D\mathbf{r}^{n+1} \, dx$$

$$= -\int_{\mathbb{R}^d} P(\varrho^{n+1}) \nabla \cdot \zeta \, dz. \qquad (A.5)$$

We used the change of variables formula again. Now (A.1) implies that

$$0 \leqslant \frac{1}{\tau} \int_{\mathbb{R}^d} \zeta \cdot \left((\mathbf{r}^{n+1} - \mathrm{id}) \circ (\mathbf{r}^{n+1})^{-1} \right) \varrho^{n+1} \, dz - \int_{\mathbb{R}^d} P(\varrho^{n+1}) \nabla \cdot \zeta \, dz$$

for all test functions ζ. Note that by changing the sign of ζ we obtain the converse inequality, so we actually have equality here. Since ζ was arbitrary we obtain

$$\left(\frac{\mathbf{r}^{n+1} - \mathrm{id}}{\tau} \circ (\mathbf{r}^{n+1})^{-1} \right) \varrho^{n+1} + \nabla P(\varrho^{n+1}) = 0$$

in the sense of distributions. If we define $\mathbf{u}^{n+1} := \tau^{-1}(\mathbf{r}^{n+1} - \mathrm{id}) \circ (\mathbf{r}^{n+1})^{-1}$, then we obtain Darcy's law $\mathbf{u}^{n+1} = -\nabla U'(\varrho^{n+1})$ a.e. in $\{\varrho^{n+1} > 0\}$. Since the new density ϱ^{n+1} can be computed according to the formula $\varrho^n = (\varrho^{n+1} \circ \mathbf{r}^{n+1}) \det D\mathbf{r}^{n+1}$ a.e., we conclude that the minimization problem (1.12) in fact amounts to a backward Euler method for the transport map $\mathbf{r}^{n+1}$ pushing $\varrho^n \mathfrak{L}^d$ forward to $\varrho^{n+1} \mathfrak{L}^d$.

Appendix B. Isentropic Euler Equations

The time discretization of Definition 1.1 is motivated by Dafermos' entropy rate criterion [6]: dissipate the total energy as fast as possible. This objective is achieved through two different mechanisms. First, the kinetic energy is minimized by replacing the given velocity by the corresponding optimal transport velocity. Second, the internal energy is minimized by solving an optimization problem similar to the one in (1.12) for the porous medium equation, but with a different penalty term. The intuition for the latter is the following: We think of the fluid as a collection of particles, each one determined by a position and a velocity. The particles prefer to stay on their free flow trajectories (minimizing the acceleration), but the pressure from the surrounding particles forces them to deviate from their characteristic paths.

To measure the acceleration we introduce a functional similar to the Wasserstein distance. Consider first a single particle that initially is located at position $x_1 \in \mathbb{R}^d$ with initial velocity $\xi_1 \in \mathbb{R}^d$. Specify some timestep $\tau > 0$. During the time interval of length τ we allow the particle to move to a new position $x_2 \in \mathbb{R}^d$ and to change its velocity to $\xi_2 \in \mathbb{R}^d$. What is the minimal acceleration required to achieve this change? We are looking for a curve $c\colon [0, \tau] \longrightarrow \mathbb{R}^d$ such that

$$(c, \dot{c})(0) = (x_1, \xi_1) \quad \text{and} \quad (c, \dot{c})(\tau) = (x_2, \xi_2),$$

that minimizes the $\mathscr{L}^2([0, \tau])$-norm of the second derivative $\ddot{c}$ along the curve. This minimization problem has a unique solution; the minimizer is given by a cubic polynomial; and the minimal (averaged) acceleration is given by the formula

$$\frac{1}{\tau} \int_0^\tau |\ddot{c}(s)|^2 \, ds = 12 \left| \frac{1}{\tau} \left(\frac{x_2 - x_1}{\tau} - \frac{\xi_2 + \xi_1}{2} \right) \right|^2 + \left| \frac{\xi_2 - \xi_1}{\tau} \right|^2.$$

Consider now a fluid characterized by a density ϱ and a velocity field $\mathbf{u}$. Assume that we are given a transport map $\mathbf{r}\colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$ that is invertible ϱ-a.e. and determines where each particle travels to during a time interval of length $\tau > 0$. Which velocity field $\hat{\mathbf{u}}$ minimizes the total acceleration cost, defined as the integral

$$\mathscr{A}_{\tau, \mathbf{r}}[\hat{\mathbf{u}}]^2 := \int_{\mathbb{R}^d} \left(3 \left| \frac{\mathbf{r} - \mathrm{id}}{\tau} - \frac{\hat{\mathbf{u}} \circ \mathbf{r} + \mathbf{u}}{2} \right|^2 + \frac{1}{4} |\hat{\mathbf{u}} \circ \mathbf{r} - \mathbf{u}|^2 \right) \varrho \, dx? \tag{B.1}$$

Note that since $\mathbf{r}$ is assumed to be invertible, there is only a single particle located at position $\mathbf{r}(x)$ for ϱ-a.e. $x \in \mathbb{R}^d$, and so the velocity of the transported particles can be described by an Eulerian velocity field $\hat{\mathbf{u}}$. Let $\mathbf{u}^+$ denote the minimizer of

the functional $\mathscr{A}_{\tau,\mathbf{r}}$ (we refer the reader to [8] for further details). For any smooth vector field $\zeta\colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and $\varepsilon \geqslant 0$ define $\mathbf{u}_\varepsilon := \mathbf{u}^+ + \varepsilon\zeta$. Then we compute

$$\lim_{\varepsilon\to 0} \frac{\mathscr{A}_{\tau,\mathbf{r}}[\mathbf{u}_\varepsilon]^2 - \mathscr{A}_{\tau,\mathbf{r}}[\mathbf{u}^+]^2}{\varepsilon} = \int_{\mathbb{R}^d} \left(-3\frac{\mathbf{r}-\mathrm{id}}{\tau} + 2\mathbf{u}^+ \circ \mathbf{r} + \mathbf{u} \right) \cdot (\zeta \circ \mathbf{r})\varrho\, dx,$$

which vanishes because $\mathbf{u}^+$ is a minimizer. Since ζ was arbitrary, we obtain that

$$\mathbf{u}^+ \circ \mathbf{r} = \mathbf{u} + \frac{3}{2\tau}\Big(\mathbf{r} - (\mathrm{id}+\tau\mathbf{u})\Big) \tag{B.2}$$

ϱ-a.e. Using this identity in (B.1) we have

$$\mathscr{A}_{\tau,\mathbf{r}}[\mathbf{u}^+]^2 = \frac{3}{4\tau^2}\int_{\mathbb{R}^d} |\mathbf{r} - (\mathrm{id}+\tau\mathbf{u})|^2\varrho\, dx =: \mathscr{W}_\tau[\mathbf{r}]^2.$$

Assume now that the push-forward of $\varrho\mathfrak{L}^d$ under the map $\mathrm{id}+\tau\mathbf{u}$ is absolutely continuous with respect to the Lebesgue measure and define $\hat{\varrho}\mathfrak{L}^d := (\mathrm{id}+\tau\mathbf{u})\#(\varrho\mathfrak{L}^d)$. If we fix a density $\varrho^+ \in \mathscr{P}(\mathbb{R}^d)$, then there are many transport maps $\mathbf{r}\colon \mathbb{R}^d \longrightarrow \mathbb{R}^d$ that push $\varrho\mathfrak{L}^d$ forward to the measure $\varrho^+\mathfrak{L}^d$. Minimizing in $\mathbf{r}$, we obtain

$$\inf\left\{ \mathscr{W}_\tau[\mathbf{r}]^2 \colon \mathbf{r}\#(\varrho\mathfrak{L}^d) = \varrho^+\mathfrak{L}^d \right\} = \frac{3}{4\tau^2}\mathbf{W}(\hat{\varrho}, \varrho^+)^2,$$

so the minimal total acceleration cost is proportional to the Wasserstein distance between ϱ^+ and the density $\hat{\varrho}$ obtained by freely transporting the initial density ϱ along the initial velocity field $\mathbf{u}$. We refer again to [8] for further details.

Note that Step 2 of the time discretization in Definition 1.1 amounts to minimizing the internal energy $\mathcal{U}[\varrho]$ while penalizing the deviation of particles from their characteristic trajectories (which requires accelerating the particles). Computing the Euler-Lagrange equations for the minimization problem (1.12), as we did above for the time step for the porous medium equation, we find that the map

$$\mathbf{t}^{n+1} := \mathrm{id} + \frac{2\tau^2}{3}\nabla U'(\varrho^{n+1})$$

is an optimal transport map between $\varrho^{n+1}\mathfrak{L}^d$ and $\hat{\varrho}^n\mathfrak{L}^d$, thus invertible ϱ^{n+1}-a.e.

The update of the velocity (1.18) follows from formula (B.2) if we set

$$\mathbf{u} := \hat{\mathbf{u}}^n, \quad \mathbf{u}^+ := \mathbf{u}^{n+1}, \quad \mathbf{r} := \left(\mathrm{id} + \frac{2\tau^2}{3}\nabla U'(\varrho^{n+1})\right)^{-1} \circ (\mathrm{id}+\tau\mathbf{u}^n).$$

Indeed we have that

$$\frac{3}{2\tau}\left(\left(\mathrm{id} + \frac{2\tau^2}{3}\nabla U'(\varrho^{n+1})\right)^{-1} - \mathrm{id} \right) \circ \left(\mathrm{id} + \frac{2\tau^2}{3}\nabla U'(\varrho^{n+1})\right)$$

$$= \frac{3}{2\tau}\left(\mathrm{id} - \left(\mathrm{id} + \frac{2\tau^2}{3}\nabla U'(\varrho^{n+1})\right) \right) = -\tau\nabla U'(\varrho^{n+1})$$

ϱ^{n+1}-a.e. The only caveat is that (B.2) was derived under the assumption that the map $\mathbf{r}$ is essentially invertible. But this condition is satisfied because $\hat{\mathbf{u}}^n$ was chosen in such a way that the map $\mathrm{id} + \tau\hat{\mathbf{u}}^n$ is an optimal transport map pushing $\varrho^n\mathfrak{L}^d$ forward to $\hat{\varrho}^n\mathfrak{L}^d := (\mathrm{id}+\tau\mathbf{u}^n)\#(\varrho^n\mathfrak{L}^d)$; see Step 1 in Definition 1.1. As mentioned there, this step decreases the kinetic energy in an optimal way.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.

[2] V. I. Arnold and B. A. Khesin, *Topological methods in hydrodynamics*, Applied Mathematical Sciences, vol. 125, Springer-Verlag, New York, 1998.

[3] L. A. Caffarelli, *Allocation maps with general cost functions*, Partial differential equations and applications, 1996, pp. 29–35.

[4] G.-Q. Chen and D. Wang, *The Cauchy problem for the Euler equations for compressible fluids*, Handbook of mathematical fluid dynamics, Vol. I, 2002, pp. 421–543.

[5] `http://abel.ee.ucla.edu/cvxopt`.

[6] C. M. Dafermos, *The entropy rate admissibility criterion for solutions of hyperbolic conservation laws*, J. Differential Equations **14** (1973), 202–212.

[7] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Math. **177** (1996), no. 2, 113–161.

[8] W. Gangbo and M. Westdickenberg, *Optimal Transport for the system of Isentropic Euler equations* (2008). FIXME.

[9] E. Hairer, S. P. Norsett, and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems*, 2nd, Springer, Berlin, 2000.

[10] D. D. Holm, J. E. Marsden, and T. S. Ratiu, *The Euler-Poincaré equations and semidirect products with applications to continuum theories*, Adv. Math. **137** (1998), no. 1, 1–81.

[11] D. Kinderlehrer and N. J. Walkington, *Approximation of parabolic equations using the Wasserstein metric*, M2AN Math. Model. Numer. Anal. **33** (1999), no. 4, 837–852.

[12] `http://www.ziena.com/knitro.htm`.

[13] J. E. Marsden and M. West, *Discrete mechanics and variational integrators*, Acta Numer. **10** (2001), 357–514.

[14] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[15] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[16] J. L. Vázquez, *Perspectives in nonlinear diffusion: between analysis, physics and geometry*, International Congress of Mathematicians. Vol. I, 2007, pp. 609–634.

[17] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, R.I., 2003.

Michael Westdickenberg, School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, GA 30332-0160, U.S.A.
    E-mail address: `mwest@math.gatech.edu`

Jon Wilkening, Department of Mathematics, University of California, 1091 Evans Hall #3840, Berkeley, CA 94720-3840, U.S.A.
    E-mail address: `wilken@math.berkeley.edu`