Electronic Transactions on Numerical Analysis.
Volume 57, pp. 193–215, 2022.

ISSN 1068–9613.
DOI: 10.1553/etna_vol57s193

ETNA
Kent State University and
Johann Radon Institute (RICAM)

DATA-DRIVEN GRADIENT FLOWS*

JAN-FREDERIK PIETSCHMANN† AND MATTHIAS SCHLOTTBOM‡

Abstract. We present a framework enabling variational data assimilation for gradient flows in general metric spaces, based on the minimizing movement (or Jordan–Kinderlehrer–Otto) approximation scheme. After discussing stability properties in the most general case, we specialize to the space of probability measures endowed with the Wasserstein distance. This setting covers many non-linear partial differential equations (PDEs), such as the porous-medium equation or general drift–diffusion–aggregation equations, which can be treated by our methods independently of their respective properties (such as finite speed of propagation or blow-up). We then focus on the numerical implementation using a primal–dual algorithm. The strength of our approach lies in the fact that, by simply changing the driving functional, a wide range of PDEs can be treated without the need to adopt the numerical scheme. We conclude by presenting several numerical examples.

Key words. gradient flows, variational data assimilation, primal–dual algorithms, Wasserstein distance

AMS subject classifications. 49N45, 35R30, 49N45, 49Q22

1. Introduction.

The aim of this work is to develop a data assimilation approach for gradient flows in metric spaces. In particular, we are interested in time-dependent non-linear partial differential equations (PDEs) that are gradient flows in the space of probability measures endowed with the 2-Wasserstein metric. Data assimilation is useful in this context if the model is not completely specified; for instance, if the initial datum is unknown or certain model parameters are not exact.

Our approach combines a time-discrete variational scheme, the famous Jordan–Kinderlehrer—Otto (JKO) or minimizing movement scheme, with additional data fidelity terms in order to overcome model uncertainties and to improve the solution in each time step.

Let us briefly describe our idea in the finite-dimensional case. Given $u_0 \in \mathbb{R}^d$ and $F \in C^{1,1}(\mathbb{R}^d;\mathbb{R})$, we consider the corresponding gradient flow

$$\begin{aligned} u'(t) &= -\nabla F(u(t)), \quad t > 0,\\ u(0) &= u_0. \end{aligned}$$

For a fixed time step $\tau > 0$ and initial condition u_0, we may recursively define a sequence $(u_k^\tau)_{k\in\mathbb{N}}$ using the implicit Euler scheme, which amounts to solving the following non-linear system of equations:

$$u_{k+1}^\tau = u_k^\tau - \tau\nabla F(u_{k+1}^\tau), \quad k = 0, 1, \dots. \tag{1.1}$$

A well-known, but crucial, observation is that (1.1) is the first-order optimality condition of the following minimization problem,

$$u_{k+1}^\tau = \operatorname*{arg\,min}_{u\in\mathbb{R}^d} \left(F(u) + \frac{1}{2\tau}\|u - u_k^\tau\|_2^2 \right), \tag{1.2}$$

which is a variational interpretation of the implicit Euler time-stepping scheme.

*Received May 17, 2022. Accepted August 19, 2022. Published online on October 5, 2022. Recommended by Elena Resmerita.

†Fakultät für Mathematik, Technische Universität Chemnitz, Reichenhainer Straße 41, 09126 Chemnitz, Germany (jfpietschmann@math.tu-chemnitz.de).

‡Department of Applied Mathematics, University of Twente, Postbus 217, 7500 AE Enschede, The Netherlands (m.schlottbom@utwente.nl).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

In practically relevant situations, the model might be specified incompletely; that is, the functional F or the initial condition u_0 are not known or are known with uncertainties. In such situations, we may, however, have access to data points v_k that are obtained from the (true) state u_k by applying a measurement operator $\mathcal{B}$ at time $t_k = k\tau$, i.e.,

$$v_k = \mathcal{B}(u(t_k)) + \eta_k, \quad k = 1, 2, \dots,$$

where η_k models (possibly random) measurement errors. Motivated by the methodology of variational data assimilation, we use this information to modify the objective functional in (1.2) by adding an additional *data fidelity term* as follows:

$$u_{k+1}^\tau = \arg\min_u F(u) + \frac{1}{2\tau}\left(\|u - u_k^\tau\|_2^2 + \frac{\|\mathcal{B}(u) - v_{k+1}\|^2}{\theta}\right). \tag{1.3}$$

Here, $\theta \in (0, \infty)$ is a weighting parameter. In the limit $\theta \to 0^+$, we formally obtain u_{k+1}^τ such that $\mathcal{B}(u_{k+1}^\tau) = v_{k+1}$, i.e., the flow is completely data-driven; for $\theta \to +\infty$, the data has no more influence on u_{k+1}^τ. We note that (1.3) is local in time in the sense that only observations for the next time step are required. Such an approach is usually referred to as the *filtering problem* in the data assimilation literature [33].

Equation (1.3) allows for different interpretations. We may think of performing the JKO scheme for the functional F in the modified, data-induced distance measure

$$d_{\mathrm{dat}}^2(u, u_k^\tau) := \|u - u_k^\tau\|_2^2 + \frac{\|\mathcal{B}(u) - v_{k+1}\|^2}{\theta}.$$

Clearly, $d_{\mathrm{dat}}(u, u_k^\tau)$ is no longer a metric, as neither $d_{\mathrm{dat}}(u, u) = 0$ nor does the triangle inequality hold for the first argument. The second interpretation, which we will exploit in the actual computations later, is a JKO scheme in the original metric but for the modified (and now τ- and data-dependent) functional

$$E_{\tau,\mathrm{dat}} = F(u) + \frac{1}{2\tau}\frac{\|\mathcal{B}(u) - v_{k+1}\|^2}{\theta}.$$

Finally, we can also interpret (1.3) as Tikhonov regularization with data term $\|\mathcal{B}(u) - v_{k+1}\|^2$, regularizer $(\tau F(u) + \|u - u_k^\tau\|_2^2/2)$, and θ takes the role of a regularization parameter. The latter interpretation might be useful for the analysis of the minimization problem (1.3). Another interesting remark is that, if one chooses θ such that $\theta\tau = \mathrm{const.}$, one recovers a nudging method [3], which is an example of a four-dimensional data assimilation method; see Remark 3.2 for further details.

The connection to non-linear evolution PDEs is due to the seminal works of Otto [41, 42] and Otto, Jordan and Kinderlehrer [30]. They showed that, on replacing the L^2-norm in (1.2) by the 2-Wasserstein metric W_2 over the space of probability measures $P(\Omega)$ over some set $\Omega \subset \mathbb{R}^d$, the resulting iterates converge to (weak) solutions to equations of the form

$$\begin{aligned} \partial_t u &= \nabla \cdot \left(u \nabla \frac{\delta F}{\delta u}(u)\right), \quad \text{in } \Omega \times (0, T), \\ u(0) &= u_0 \text{ in } \Omega, \end{aligned} \tag{1.4}$$

for some final time T and supplemented with no-flux boundary conditions

$$u \nabla \frac{\delta F}{\delta u}(u) \cdot n = 0, \quad \text{on } \partial\Omega \times (0, T),$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

with n denoting the outward normal vector. Typical functionals F of interest are of the form

$$F(u) = \int_\Omega U\left(\frac{\mathrm{d}u}{\mathrm{d}x}(x)\right)\,\mathrm{d}x + \int_\Omega V(x)\,\mathrm{d}u(x) + \frac{1}{2}\int_\Omega\int_\Omega W(x,y)\,\mathrm{d}u(x)\,\mathrm{d}u(y), \tag{1.5}$$

where $\mathrm{d}u/\mathrm{d}x$ denotes the Radon–Nikodym derivative with respect to the Lebesgue measure, and the convention is to set the first integral to $+\infty$ if u is not absolutely continuous. In (1.5), $U : \mathbb{R} \to \mathbb{R}$ denotes the internal energy, $V : \Omega \to \mathbb{R}$ a potential and $W : \Omega \times \Omega \to \mathbb{R}$ an interaction potential.

In this setting, our approach becomes

$$u_{k+1}^\tau = \arg\min_u F(u) + \frac{1}{2\tau}\left(W_2^2(u, u_k^\tau) + \frac{d_D^2(\mathcal{B}(u), v_{k+1}^\delta)}{\theta}\right), \tag{1.6}$$

with d_D a data fidelity metric. We call (1.6) the daJKO scheme in the following. Before stating the contributions of this work, let us briefly review related work.

Related work.

W_2 gradient flows. The interpretation of (1.4) as Wasserstein gradient can be used to show the existence of solutions via the limit $\tau \to 0$ in (1.2) but also to obtain results on the long-time behaviour of solutions; see for example [16, 37]. More recently, there have been efforts to understand a larger class of evolution equations as gradient flows, e.g., equations on discrete state spaces, [18, 36], graph structures [25, 26, 27], and equations featuring non-local terms such as the Boltzmann [23] or aggregation equation [24]. Another popular research direction is equations with reaction terms or non-homogeneous boundary conditions where the total mass is not conserved in time and the Wasserstein distance has to be replaced by Fisher–Rao type distances that are finite even for two measures having different mass [11, 39]. A detailed overview of the general theory in metric spaces can be found in [2].

Computational optimal transport. To implement our daJKO scheme, we need to be able to efficiently compute the 2-Wasserstein distance. In recent years, many algorithms for solving this optimal transport problem have been developed. Different approaches have been taken, for example, based on the Monge–Ampère equation [5], for semi-discrete problems where the target space is finite [32], or scaling algorithms for regularized problems obtained by adding an entropic term to the static Wasserstein functional [19, 44]. We focus on a class of algorithms that is based on the dynamic formulation of optimal transport due to Benamou and Brenier [4]; see also (2.2) below. For the approximation of the Wasserstein distance, the original paper [4] used an augmented Lagrangian (ALG) approach to enforce the divergence constraints, which has been extended in [6] to a convex formulation for the corresponding JKO scheme. The ALG approach has also been adopted to finite volumes in [13]; see also [14] for an application to multiphase problems. In [43] proximal splitting algorithms are combined with staggered grid discretizations to enforce the divergence constraints; proximal methods have also been combined with finite element discretizations [1] for the discretization of L^2-gradient flows. Starting also from the dynamic formulation of the Wasserstein distance, Carrillo et al. [15] employ piecewise-constant approximations and proximal splitting algorithms to discretize the JKO scheme. Notably, in [15], the divergence constraints are enforced only in a relaxed sense. The resulting formulation then involves the minimization of a sum of three convex functionals, which is done by the algorithm developed in [51].

Data assimilation. Our approach is closest to three-dimensional variational data assimilation methods (3DVar) in that it relies on minimizing a functional containing a data term in each time step. Being purely deterministic, it is, however, different from methods which aim to obtain posterior probability distributions; see [33] for an overview. As mentioned

ETNA
Kent State University and
Johann Radon Institute (RICAM)

above, we focus on the filtering problem to improve the quality of the numerical solutions based on measurements at the current time step. The Wasserstein metric has recently started to become an important tool in the analysis of particle filters [45]. There are also several works that use it in a variational setting which is closer to our situation [48]. Performing data assimilation in general metric spaces does not seem to have been considered so far, but see [47] for a related problem. Let us also mention the connection to minimization-based formulations of inverse problems, where both model and data fidelity term are combined in a variational approach [29, 31]. Finally, let us remark that (1.3) is, at least in the Euclidean setting, connected to a closed-loop control technique called *nudging* [3]; see Remark 3.2 for further details.

Contributions. In this work we:

- Introduce a methodology for variational data assimilation that can be applied to gradient flows in metric spaces.
- Consider the particular case of the space of probability measures endowed with the 2-Wasserstein distance, which includes many time-dependent non-linear PDEs with linear mobility. We provide a uniform framework mostly independent of the properties (such as regularity, finite time blow-up, etc.) of the specific PDE.
- Introduce a numerical algorithm based on the recent work [15], but, instead of a piecewise-constant approximation, use a finite difference approximation to discretize the divergence constraints. Similar to [15], we obtain a positivity-preserving scheme that avoids the solution of linear systems. This again yields a uniform method that can be applied to a large class of non-linear PDEs without changing the algorithm.
- Test our method numerically on different examples of non-linear PDEs.

The paper is structured as follows. In Section 2, we present some background on optimal transport, the p-Wasserstein distances and their dynamic formulation, as well as details on the variational scheme in general metric spaces. Section 3 presents the data-driven variational approach and some of its properties, while Section 4 is devoted to a detailed description of the numerical scheme. Finally, Section 5 contains several numerical studies.

2. Preliminaries.

We call $\Omega \subset \mathbb{R}^d$ an open, bounded, and connected set. By $P(\Omega)$ we denote the space of probability measures.

2.1. Optimal transport.

Given $u_0, u_1 \in P(\Omega)$, optimal transport aims at finding a transport plan $\pi \in P(\Omega \times \Omega)$ that realizes the transport of u_0 onto u_1 at minimal cost with respect to a cost functional $c : \Omega \times \Omega \to \mathbb{R}_+$. Formally speaking, $c(x, y)\, \mathrm{d}\pi(x, y)$ is the cost for transporting an infinitesimal amount of mass from x to y. Evaluating the total cost at the optimal transport plan gives rise to a distance between probability measures, the p-Wasserstein distance, which is defined as

$$W_p^p(u_0, u_1) = \inf_{\pi \in \Pi(u_0, u_1)} \int_{\Omega \times \Omega} \frac{1}{p} |x - y|^p \, \mathrm{d}\pi(x, y),$$

with infimum taken over all $\pi \in P(\Omega \times \Omega)$ having marginals u_0 and u_1, respectively, i.e.,

$$\Pi(u_0, u_1) = \{\pi \in P(\Omega \times \Omega) : \pi(A, \Omega) = u_0(A) \text{ and } \pi(\Omega, B) = u_1(B) \\ \text{for all measurable } A,\ B \subset \Omega\}.$$

With this distance, $(P(\Omega), W_p)$ becomes a metric space and convergence with respect to W_p metrizes weak-$*$ convergence of measures.

Let us remark that, in the space $(P(\Omega), W_p)$, constant-speed geodesics are curves $\rho : [0, 1] \to P(\Omega)$ satisfying [2, (7.2.1)]

$$W_p(\rho(s), \rho(t)) = (t - s) W_p(\rho(0), \rho(1)) \quad \forall\, 0 \le s \le t \le 1.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Given two measures $u_0, u_1 \in P(\Omega)$, a constant-speed geodesic connecting them can be constructed as follows: Denoting $\pi \in \Pi(u_0, u_1)$ an optimal transport plan from u_0 to u_1, the constant-speed geodesic is defined by [2, Theorem 2.2.2] as

$$\rho(t) = (r_t)_{\#}\pi, \quad \text{for } r_t : \Omega \times \Omega \to \mathbb{R}^d,\ (x, y) \mapsto (1-t)x + ty, \tag{2.1}$$

with push-forward defined by $((r_t)_{\#}\pi)(B) := \pi(r_t^{-1}(B))$ for all Borel subsets B of $\mathbb{R}^d$.

In the following, we will restrict ourselves to the case $p = 2$, and introduce a dynamic version of the Wasserstein distance due to Benamou and Brenier [4], which will later serve as a basis for our numerical scheme. The idea is to select, among all absolutely continuous (with respect to W_2) curves $\rho : [0, 1] \to P(\Omega)$ connecting two given measures $u_0, u_1 \in P(\Omega)$, the one that has the least kinetic energy. For a rigorous formulation, we introduce the convex lower-semicontinuous function [4]

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

Note that m is a momentum variable, so that m^2/ρ indeed yields a kinetic energy. Given $\rho \in P(\Omega)$ and $m \in P(\Omega^d)$, we then define the *action density*

$$\mathcal{A}(\rho, m) = \int_\Omega \Phi\left(\frac{\mathrm{d}\rho}{\mathrm{d}\sigma}, \frac{\mathrm{d}m}{\mathrm{d}\sigma}\right) \mathrm{d}\sigma(x),$$

where σ is a reference measure such that ρ and m are absolutely continuous with respect to σ. Owing to the one-homogeneity of Φ, the value of $\mathcal{A}$ is independent of the choice of σ. We have

$$W_2^2(u_0, u_1) = \inf_{(\rho,m)\in \mathrm{CE}(u_0,u_1)} \int_0^1 \mathcal{A}(\rho, m)\, \mathrm{d}t. \tag{2.2}$$

Here, $(\rho, m) \in P(\Omega) \times P(\Omega)^d$ belong to the constrained set $\mathrm{CE}(u_0, u_1)$ if

$$\partial_t \rho + \mathrm{div}(m) = 0 \quad \text{on } \Omega \times (0, 1), \tag{2.3}$$

$$m \cdot n = 0 \quad \text{on } \partial\Omega \times (0, 1), \tag{2.4}$$

$$\rho(0) = u_0 \quad \text{on } \Omega, \tag{2.5}$$

$$\rho(1) = u_1 \quad \text{on } \Omega, \tag{2.6}$$

where n denotes the unit normal vector field pointing outwards from Ω. This formulation constitutes a convex optimization problem with linear constraints [4].

2.2. Minimizing movement/JKO scheme. Let us consider the minimizing movement scheme (1.2) in a general metric space (X, d) where it reads as

$$u_{k+1}^\tau \in \arg\min_{u\in X} \left(F(u) + \frac{1}{2\tau} d^2(u_k^\tau, u) \right). \tag{2.7}$$

The existence of iterates in this setting is usually shown via the direct method of calculus of variations, i.e., using the fact that sublevel sets are compact, to show that every minimizing sequence admits a converging subsequence and showing that, as a consequence of lower-semicontinuity, its limit is a minimizer. Here, one can assume compactness either already in the topology induced by d or in a weaker topology in which d is still lower-semicontinuous

ETNA
Kent State University and
Johann Radon Institute (RICAM)

(e.g., weak convergence in a Sobolev space). Before we specialize to the Wasserstein case, we briefly discus the stability properties of iterates. This is relevant for us since we are interested in a setting where both the initial datum as well as the functional F may not be known exactly. The following lemma shows that iterates are stable under perturbations of the functional F or the previous time step.

LEMMA 2.1. *Let $\delta_F > 0$, and let $F_1, F_2 : X \to \mathbb{R}_+ \cup \{+\infty\}$ be two functionals satisfying*

$$\sup_u |F_1(u) - F_2(u)| \leq \delta_F.$$

Furthermore, fix two elements $u_k^{\tau,1}, u_k^{\tau,2} \in X$. Denote by $u_{k+1}^{\tau,1}$ and $u_{k+1}^{\tau,2}$ the solutions of (2.7) with $F = F_1$ and $F = F_2$, and $u_k^\tau = u_k^{\tau,1}$ and $u_k^\tau = u_k^{\tau,2}$, respectively. Then, we have the estimate

$$d^2(u_{k+1}^{\tau,1}, u_{k+1}^{\tau,2}) \leq 9d^2(u_k^{\tau,1}, u_k^{\tau,2}) + 8\tau\delta_F + 4\tau(\|F_1\|_{C^0} + \|F_2\|_{C^0}).$$

Proof. Using the optimality of $u_{k+1}^{\tau,i}$, $i = 1, 2$ in (2.7) yield

$$\begin{aligned} F_1(u_{k+1}^{\tau,1}) + \frac{1}{2\tau}d^2(u_{k+1}^{\tau,1}, u_k^{\tau,1}) &\leq F_1(u_k^{\tau,2}) + \frac{1}{2\tau}d^2(u_k^{\tau,2}, u_k^{\tau,1}), \\ F_2(u_{k+1}^{\tau,2}) + \frac{1}{2\tau}d^2(u_{k+1}^{\tau,2}, u_k^{\tau,2}) &\leq F_2(u_k^{\tau,1}) + \frac{1}{2\tau}d^2(u_k^{\tau,1}, u_k^{\tau,2}), \end{aligned}$$

which implies

$$d^2(u_{k+1}^{\tau,j}, u_k^{\tau,j}) \leq 2\tau(F_j(u_k^{\tau,i}) - F_j(u_{k+1}^{\tau,j})) + d^2(u_k^{\tau,i}, u_k^{\tau,j}), \quad i, j = 1, 2,\ i \neq j.$$

Using the triangle inequality, this allows one to estimate

$$\begin{aligned} d^2(u_{k+1}^{\tau,1}, u_{k+1}^{\tau,2}) &\leq 3(d^2(u_{k+1}^{\tau,1}, u_k^{\tau,1}) + d^2(u_{k+1}^{\tau,2}, u_k^{\tau,2}) + d^2(u_k^{\tau,1}, u_k^{\tau,2})) \\ &\leq 9d^2(u_k^{\tau,1}, u_k^{\tau,2}) + 2\tau(F_1(u_k^{\tau,2}) - F_1(u_{k+1}^{\tau,1}) + F_2(u_k^{\tau,1}) - F_2(u_{k+1}^{\tau,2})). \end{aligned}$$

We further estimate the second term on the right-hand side as

$$\begin{aligned} &F_1(u_k^{\tau,2}) - F_1(u_{k+1}^{\tau,1}) + F_2(u_k^{\tau,1}) - F_2(u_{k+1}^{\tau,2}) \\ &\quad = F_1(u_k^{\tau,2}) - F_2(u_k^{\tau,2}) + F_2(u_{k+1}^{\tau,1}) - F_1(u_{k+1}^{\tau,1}) + F_2(u_k^{\tau,1}) - F_1(u_k^{\tau,1}) \\ &\qquad + F_1(u_{k+1}^{\tau,2}) - F_2(u_{k+1}^{\tau,2}) + F_2(u_k^{\tau,2}) - F_2(u_{k+1}^{\tau,1}) + F_1(u_k^{\tau,1}) - F_1(u_{k+1}^{\tau,2}) \\ &\quad \leq 4\delta_F + 2(\|F_1\|_{C^0} + \|F_2\|_{C^0}), \end{aligned}$$

which completes the proof. ∎

REMARK 2.2. Note that, in general, assuming F to be convex but not strictly convex, minimizers of (2.7) are not unique. Thus, the additional term $4\tau(\|F_1\|_{C^0} + \|F_2\|_{C^0})$ on the right-hand side is natural in this setting. However, because this term remains even if we assume F to be strictly convex (implying uniqueness of minimizers), this shows that our estimate is not optimal. A way to improve this estimate would be to work with geodesic interpolations; see, e.g., [46]. We leave this question for future work.

A natural question is the behaviour of iterates as τ tends to zero. The starting point for the convergence analysis is to derive a priori estimates that, using compactness, guarantee the existence of a limiting curve. The key idea is to exploit the minimization property of the iterates, i.e.,

$$F(u_{k+1}^\tau) + \frac{1}{2\tau}d^2(u_k^\tau, u_{k+1}^\tau) \leq F(u_k^\tau). \tag{2.8}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

This implies the *energy estimate* $\sup_k F(u_k^\tau) \leq F(u_0)$. Summing over k further yields the *total square estimate*

$$\sum_{k\in\mathbb{N}_0} d^2(u_k^\tau, u_{k+1}^\tau) \leq 2\tau \left(F(u_0) - \inf_k F(u_k^\tau) \right). \tag{2.9}$$

Assuming F to be bounded from below and u_0 such that $F(u_0) < +\infty$, this provides a uniform bound. Applying the triangle inequality for any $0 \leq l \leq k$, and using the Cauchy–Schwarz inequality, we obtain

$$d(u_l^\tau, u_k^\tau) \leq \sqrt{2\tau} \sum_{j=l}^{k-1} \frac{1}{\sqrt{2\tau}} d(u_j^\tau, u_{j+1}^\tau) \leq \sqrt{2\tau}\sqrt{(k-l)} \sqrt{F(u_0^\tau) - \inf_{n\in\mathbb{N}} F(u_k^\tau)}.$$

As a consequence, for any $0 \leq s \leq t$, denoting by $u^\tau(t)$ the constant-speed geodesic interpolation introduced in (2.1), we have

$$\begin{aligned} d(u^\tau(s), u^\tau(t)) &\leq d(u^\tau(s), u^\tau_{[s/\tau+1]\tau}) + d(u^\tau_{[s/\tau+1]\tau}, u^\tau_{[t/\tau]\tau}) + d(u^\tau_{[t/\tau]\tau}, u^\tau(t)) \\ &\leq \left(\left[\frac{s}{\tau} + 1\right] - \frac{s}{\tau} \right) d(u^\tau_{[s/\tau]\tau}, u^\tau_{[s/\tau+1]\tau}) \\ &\quad + \sqrt{2\tau \left(\left[\frac{t}{\tau}\right] - \left[\frac{s}{\tau} + 1\right] \right) \left(F(u_0^\tau) - \inf_{n\in\mathbb{N}} F(u_k^\tau) \right)} \\ &\quad + \left(\frac{t}{\tau} - \left[\frac{t}{\tau}\right] \right) d(u^\tau_{[t/\tau]\tau}, u^\tau_{[t/\tau+1]\tau}) \\ &\leq \sqrt{6 \left(F(u_0) - \inf_{n\in\mathbb{N}} F(u_n) \right)} (t-s)^{1/2}. \end{aligned} \tag{2.10}$$

Ultimately, this $1/2$-Hölder estimate allows the application of the Arzelá–Ascoli theorem to conclude relative compactness of the family $(u^\tau)_{\tau>0}$, i.e., the existence of a limit curve. It then remains to show that this limit indeed satisfies the chosen gradient flow formulation, which depends on the problem at hand.

REMARK 2.3. While the above estimate is only of order $\sqrt{\tau}$, it can be shown [2, Ch. 4] that, under mild regularity assumptions on the initial point, convergence with order τ holds. In order to obtain second-order convergence, one has to modify the variational scheme. In the Euclidean case, replacing (1.2) by

$$u_{k+1}^\tau = \arg\min_u \left(2F\left(\frac{u + u_k^\tau}{2} \right) + \frac{1}{2\tau} \|u - u_k^\tau\|_2^2 \right)$$

is sufficient. Using the notion of geodesics, this idea carries over to the Wasserstein setting [34]; see also [15, 38] for Crank–Nicolson and backward differentiation formula (BDF) type higher-order schemes.

3. Data-driven minimizing movement scheme. We first introduce our approach in the most general setting in metric spaces, generalizing the Euclidean setting described in the introduction. To this end we denote by (X, d_S) the metric space containing the state, while the data lives in the space (Y, d_D). Finally we introduce a measurement operator $\mathcal{B} : X \to Y$. In this work, we will always assume $\mathcal{B}$ to be d_D–d_S Lipschitz continuous with constant $L_\mathcal{B}$, i.e., for all $x_1, x_2 \in X$, the following holds:

$$d_D(\mathcal{B}(x_1), \mathcal{B}(x_2)) \leq L_\mathcal{B} d_S(x_1, x_2).$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

In this set-up, we extend (2.7) by introducing the data-driven minimizing movement scheme as follows:

$$u_{k+1}^{\tau,\delta} \in \arg\min_u \left(F(u) + \frac{1}{2\tau} \left(d_S^2(u, u_k^{\tau,\delta}) + \frac{d_D^2(\mathcal{B}(u), v_{k+1}^\delta)}{\theta} \right) \right), \tag{3.1}$$

where the $v_k^\delta \in Y$ are given noisy measurements and $\theta \in (0, \infty)$ is a weighting parameter. For the particular choice $d_S = W_2$, we call (3.1) the data-driven JKO (daJKO) scheme.

REMARK 3.1 (Minimizers without noise). Let us briefly discuss the case of exact measurements. In the limit $\theta \to 0^+$ and for noise-free data v_k, we obtain a minimizer $u_{k+1}^{\tau,0}$ of (3.1) such that $\mathcal{B}(u_{k+1}^{\tau,0}) = v_{k+1}$, i.e., the flow is completely data-driven, while for $\theta \to +\infty$, the data has no more influence on $u_{k+1}^{\tau,0} = u_{k+1}^\tau$. If the data is generated by solutions u_k^τ to (2.7), i.e., $v_k = \mathcal{B}(u_k^\tau)$, then solving (3.1) yields the same minimizers as (2.7). If, on the other hand, the data is generated from the time-continuous limit curve $u(t)$ evaluated at discrete times $t_k = t\tau$, we will always have an approximation error of order τ.

REMARK 3.2 (Connection to nudging). Returning to the Euclidean setting for a moment, and assuming $\mathcal{B}$ to be a linear operator, we see that, if we choose θ as a function of τ, we obtain, formally in the limit $\tau \to 0$, an interpretation as a closed-loop control problem/data assimilation problem called nudging introduced in [3]. We start with

$$u_{k+1}^\tau = \arg\min_{u \in \mathbb{R}^d} \left(F(u) + \frac{1}{2\tau} \|u - u_k^\tau\|_2^2 + \frac{1}{2\tau\theta} \|\mathcal{B}u - v_{k+1}\|_2^2 \right).$$

Calculating the Euler–Lagrange equation results in

$$u_{k+1}^\tau = u_k^\tau - \tau \nabla F(u_{k+1}^\tau) + \frac{1}{\theta} \mathcal{B}^*(\mathcal{B}u_{k+1}^\tau - v_{k+1}).$$

If, for a given constant $c > 0$, we choose $\theta = c/\tau$ and (formally) pass to the limit $\tau \to 0$, we recover the equation

$$\begin{aligned} u'(t) &= -\nabla F(u(t)) + \frac{1}{c} \mathcal{B}^*(\mathcal{B}u(t) - v(t)), \quad t > 0, \\ u(0) &= u_0, \end{aligned}$$

where $v(t)$ is a suitable interpolation of the data points in time (e.g., piecewise-linear).

We do not argue on the question of the existence of minimizers (usually proven by means of the direct method of calculus of variations) which depends on specific choices of (X, d_S) and (Y, d_D). Instead, we comment on convergence as $\tau \to 0$.

REMARK 3.3 (Convergence as $\tau \to 0$). For the JKO scheme without data, the estimates in (2.9) and (2.10) provide the necessary compactness to obtain a limiting curve as $\tau \to 0$. They are shown by comparing the functional values of u_{k+1}^τ and u_k^τ, using optimality of the former; see (2.8). In the data-driven case, this inequality becomes

$$\begin{aligned} &F(u_{k+1}^{\tau,\delta}) + \frac{1}{2\tau} d^2(u_k^{\tau,\delta}, u_{k+1}^{\tau,\delta}) \\ &\quad \le F(u_k^{\tau,\delta}) + \frac{1}{2\theta\tau} \left[d^2(\mathcal{B}(u_k^{\tau,\delta}), v_{k+1}) - d^2(\mathcal{B}(u_{k+1}^{\tau,\delta}), v_{k+1}) \right]. \end{aligned}$$

In order to proceed, we assume the error bound

$$d_D(\mathcal{B}(u_{k+1}^\tau), v_{k+1}) \le \delta_k, \tag{3.2}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

as well as d_S–d_D Lipschitz continuity of $\mathcal{B}$ with constant $L_{\mathcal{B}}$ satisfying $L_{\mathcal{B}}^2 < \theta/2$. This allows us to further estimate, using the reverse triangle inequality $|d(x,z) - d(y,z)| \le d(x,y)$, that

$$\begin{aligned} &d^2(\mathcal{B}(u_k^\tau), v_{k+1}) - d^2(\mathcal{B}(u_{k+1}^\tau), v_{k+1}) \\ &\quad = \big[d(\mathcal{B}(u_k^\tau), v_{k+1}) - d(\mathcal{B}(u_{k+1}^\tau), v_{k+1})\big]\big[d(\mathcal{B}(u_k^\tau), v_{k+1}) + d(\mathcal{B}(u_{k+1}^\tau), v_{k+1})\big] \\ &\quad \le d(\mathcal{B}(u_k^\tau), \mathcal{B}(u_{k+1}^\tau))\big[d(\mathcal{B}(u_k^\tau), v_{k+1}) + d(\mathcal{B}(u_{k+1}^\tau), v_{k+1})\big]. \end{aligned}$$

Using the triangle inequality once more,

$$d(\mathcal{B}(u_k^\tau), v_{k+1}) \le d(\mathcal{B}(u_{k+1}^\tau), v_{k+1}) + d(\mathcal{B}(u_k^\tau), \mathcal{B}(u_{k+1}^\tau)),$$

we obtain

$$d^2(\mathcal{B}(u_k^\tau), v_{k+1}) - d^2(\mathcal{B}(u_{k+1}^\tau), v_{k+1}) \le 2d^2(\mathcal{B}(u_k^\tau), \mathcal{B}(u_{k+1}^\tau)) + d^2(\mathcal{B}(u_{k+1}^\tau), v_{k+1}).$$

Employing the Lipschitz continuity of $\mathcal{B}$ and $L_{\mathcal{B}}^2 < \theta/2$ as well as the bound on the measurement error (3.2) results in

$$F(u_{k+1}^\tau) + \frac{1}{2\tau}\left(1 - \frac{2L_{\mathcal{B}}^2}{\theta}\right) d^2(u_k^\tau, u_{k+1}^\tau) \le F(u_k^\tau) + \frac{1}{2\theta\tau}\delta_k^2.$$

Thus, in order to have convergence, both δ_k as well as θ must depend on τ in a suitable way to ensure that the last term converges to zero in the final estimate. A similar situation occurs when studying (2.7) with a perturbed functional; see [9, 10].

REMARK 3.4 (Relation to Tikhonov regularization). As mentioned in the introduction, (1.3) can also be seen as a Tikhonov regularization with a non-linear forward operator. Thus, one might expect stability and convergence results as in [22, Ch. 10]. However, in the most general metric setting, neither minimizers nor limiting curves are unique (not even locally). In the 2-Wasserstein case, combined with additional convexity assumptions on F, similar results might be possible, which is, however, beyond the scope of this work.

4. Numerical realization. We now introduce our numerical scheme, which is based on the recent work by Carrillo et al. [15], yet using a finite difference approximation to discretize the divergence constraints instead of a piecewise-constant approximation. This approach yields a clear interpretation of the initial and boundary conditions appearing in (2.2). We provide a detailed description of the numerical scheme to highlight the influence of weighted norms and inner products that arise from the finite difference discretization. Let us remark that higher-order schemes have been developed [15, 38], and that our data assimilation approach can be combined with these schemes in a similar fashion.

4.1. Finite difference approximation. To convey the main ideas, we consider the one-dimensional case, i.e., $\Omega = (L, R)$ with $L < R$ being real numbers. The case of multiple dimensions can be handled similarly if the domain is a Cartesian product of intervals by employing the usual tensor product construction for the discretization. We partition Ω into N_x intervals with mesh size $\delta_x = (R-L)/N_x$. Similarly, we partition the time interval $(0,1)$ into N_t intervals with mesh size δ_t, and introduce the grid points

$$x_j = (j-1)\delta_x + L, \quad t_k = (k-1)\delta_t, \quad j \in \{1, \dots, N_x + 1\}, \quad k \in \{1, \dots, N_t + 1\}.$$

To any continuous function $v : [L, R] \times [0, 1] \to \mathbb{R}$ we associate a grid function $v_h = (v_{j,k})_{j,k}$ via

$$v_{j,k} = v(x_j, t_k), \quad j \in \{1, \dots, N_x + 1\}, \quad k \in \{1, \dots, N_t + 1\}.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Similar notation is used for grid functions associated to the spatial and temporal partitions, respectively.

The discrete energy approximating (1.5) is obtained by using a composite trapezoidal rule,

$$F_h(u_h) = \sum_{j=1}^{N_x+1} w_j^x \left(U(u_j) + V_j u_j + \frac{1}{2} \sum_{i=1}^{N_x+1} w_i^x W_{i,j} u_i u_j \right), \tag{4.1}$$

where $V_j = V(x_j)$ and $W_{i,j} = W(x_i, x_j)$, and $u_h = (u_j)$ is a grid function associated with the spatial partition; and the weight functions are given by

$$w_j^x = \begin{cases} \delta_x/2 & j \in \{1, N_x + 1\}, \\ \delta_x & \text{otherwise.} \end{cases}$$

We denote by w_k^t a similar weight function for the composite trapezoidal rule associated with the temporal partition, and note that $\sum_{k=1}^{N_t+1} w_k^t = 1$.

Using these rules, we define the following inner product and norm for grid functions $v_h = (\rho_{j,k}^v, m_{j,k}^v)$ and $w_h = (\rho_{j,k}^w, m_{j,k}^w)$,

$$\langle v_h, w_h \rangle = \sum_{k=1}^{N_k+1} \sum_{j=1}^{N_x+1} w_k^t w_j^x (\rho_{j,k}^v \rho_{j,k}^w + m_{j,k}^v m_{j,k}^w), \quad \|v_h\| = \sqrt{\langle v_h, v_h \rangle}, \tag{4.2}$$

which makes the linear space of grid functions a Hilbert space.

Before we state the discrete optimization problem corresponding to (1.6), we discretize the constraints (2.3)–(2.5). For the divergence constraint (2.3), we use centred differencing in space and backward differencing in time for the interior grid points $j \in \{2, \dots, N_x\}$, $k \in \{2, \dots, N_t + 1\}$, i.e.,

$$\frac{\rho_{j,k} - \rho_{j,k-1}}{\delta_t} + \frac{m_{j+1,k} - m_{j-1,k}}{2\delta_x} = 0. \tag{4.3}$$

For the boundary, we use a one-sided finite difference approximation to approximate m_x, i.e., for $j \in \{1, N_x + 1\}$ we use

$$\frac{\rho_{1,k} - \rho_{1,k-1}}{\delta_t} + \frac{m_{2,k} - m_{1,k}}{\delta_x} = 0 \tag{4.4}$$

and

$$\frac{\rho_{N_x+1,k} - \rho_{N_x+1,k-1}}{\delta_t} + \frac{m_{N_x+1,k} - m_{N_x,k}}{\delta_x} = 0. \tag{4.5}$$

The boundary condition (2.4) becomes

$$m_{1,k} = 0 = m_{N_x+1,k} \quad \text{for all } k \in \{1, \dots, N_t + 1\}. \tag{4.6}$$

Since the discretization does not depend on $m_{j,1}$, we set $m_{j,1} = 1$ for $j = 1, \dots, N_x + 1$. Furthermore, the initial condition (2.5) and the mass constraint are discretized as follows:

$$\rho_{j,1} = \rho_0(x_j), \quad j \in \{1, \dots, N_x + 1\} \tag{4.7}$$

and

$$\sum_{j=1}^{N_x+1} w_j^x (\rho_{j,k} - \rho_0(x_j)) = 0, \quad k \in \{2, \dots, N_t + 1\}. \tag{4.8}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

4.2. Discrete daJKO scheme. Using the dynamic formulation of the Wasserstein distance (2.2) and the notation $\rho_h^1 = \rho_h(\cdot, 1)$ for a grid function associated with the spatial partition, the discrete daJKO scheme, i.e., the minimization problem (1.6) – but without the data term – is then

$$\inf_{(\rho_h, m_h)} \sum_{k=1}^{N_t+1} \sum_{j=1}^{N_x+1} w_j^x w_k^t \left(\Phi(\rho_{j,k}, m_{j,k}) + \tau \left(U(\rho_j^1) + V_j \rho_j^1 + \frac{1}{2} \sum_{i=1}^{N_x+1} w_i^x W_{i,j} \rho_i^1 \rho_j^1 \right) \right).$$

Here, (ρ_h, m_h) are grid functions associated with the space–time grid such that they satisfy the constraints (4.3)–(4.8) in the following relaxed form, that is,

$$\begin{aligned} &\sum_{k=2}^{N_t+1} w_k^t \left(w_1^x \left(\frac{\rho_{1,k} - \rho_{1,k-1}}{\delta_t} + \frac{m_{2,k} - m_{1,k}}{\delta_x} \right)^2 \right. \\ &\quad + w_{N_x+1}^x \left(\frac{\rho_{N_x+1,k} - \rho_{N_x+1,k-1}}{\delta_t} + \frac{m_{N_x+1,k} - m_{N_x,k}}{\delta_x} \right)^2 \\ &\quad \left. + \sum_{j=2}^{N_x} w_j^x \left(\frac{\rho_{j,k} - \rho_{j,k-1}}{\delta_t} + \frac{m_{j+1,k} - m_{j-1,k}}{2\delta_x} \right)^2 \right) \le \delta_1^2, \end{aligned} \tag{4.9}$$

$$\sum_{k=1}^{N_t+1} w_k^t (m_{0,k}^2 + m_{N_x+1,k}^2) \le \delta_2^2, \tag{4.10}$$

$$\sum_{k=1}^{N_t+1} w_k^t \left(\sum_{j=1}^{N_x+1} w_j^x (\rho_{j,k} - \rho_0(x_j)) \right)^2 \le \delta_3^2, \tag{4.11}$$

$$\sum_{j=1}^{N_x+1} w_j^x (\rho_{j,1} - \rho_0(x_j))^2 \le \delta_4^2 \tag{4.12}$$

for some tolerances δ_i, $i = 1, 2, 3, 4$.

Similar to [15], we observe that the weakened constraints (4.9)–(4.12) are quadratic and can be written in the form

$$Au \in \mathcal{C}_\delta = \{x : \|x_i - b_i\|_2 \le \delta_i,\ i = 1, 2, 3, 4\},$$

where the vector u contains the coefficients of the grid functions (ρ_h, m_h). Note that the weights w_j^x and w_k^t are included in the definition of A_i and b_i, respectively, and the vectors x_i are slices of the vector x corresponding to the number of rows in A_i. We define the matrix A by vertically concatenating the matrices A_i, $i = 1, \ldots, 4$. We note that A is the matrix of a linear map from the Hilbert space of grid functions with inner product defined in (4.2) to a Euclidean space. In order to enforce the constraints, we introduce the indicator function of the set $\mathcal{C}_\delta$ as

$$\mathfrak{i}_\delta(\phi) = \begin{cases} 0 & \text{if } \phi \in \mathcal{C}_\delta, \\ \infty & \text{otherwise.} \end{cases}$$

Summarizing, given $u_h^{(n)}$ as an approximation of u_n^τ and suitable data $v^{(n+1)}$, one step of the discrete daJKO scheme is to compute $u_h^{(n+1)} = \rho_h^*(\cdot, 1)$, where, for $\rho_h^0 = u_h^{(n)}$, the grid

ETNA
Kent State University and
Johann Radon Institute (RICAM)

function (ρ_h^*, m_h^*) is the minimizer of

$$\inf_{(\rho_h, m_h)} \left(\sum_{k=1}^{N_t+1} \sum_{j=1}^{N_x+1} w_j^x w_k^t \Phi(\rho_{j,k}, m_{j,k}) \right) + \tau E_h^{(n+1)}(\rho_h^1) + \mathfrak{i}_\delta(Au),$$

and where the combined energy is defined by

$$E_h^{(n+1)}(\rho_h^1) = F_h(\rho_h^1) + \frac{1}{2\theta} \| B_h(\rho_h^1) - v^{(n+1)} \|_d^2.$$

Here, B_h denotes a suitable discretization of the measurement operator $\mathcal{B}$ and $\| \cdot \|_d$ a suitable ℓ_2-like norm. One may choose different metrics for the data term, but we do not go in this direction here. Similarly, the initialization $\rho_h^0(x_j) = u_h^{(n)}(x_j)$ can be modified, e.g., to incorporate different data, and we employ such a modification in one of the examples in Section 5.1.2 below.

4.3. Implementation of the daJKO scheme. In order to perform one step of the daJKO algorithm computationally, we employ the algorithm developed in [51] for the minimization of a sum of three convex functionals; see Algorithm 1. Iterating the daJKO steps yields our overall data assimilation scheme, as described in Algorithm 2. We note that, if no data terms are present, Algorithm 2 reduces to the algorithm used in [15]. Here, however, we use a different differencing scheme to discretize the constraints. Next, let us discuss the building blocks used in Algorithm 1.

Algorithm 1: Primal–dual algorithm for one daJKO step; cf. [51].

Input : $u^{(0)}$, $\phi^{(0)}$, $it_{\max}$, E, ∇E, λ, σ, A, b, δ
Output : u^*, ϕ^*

Initialize $\bar{u}^{(0)} = u^{(0)}$ *and* $l = 0$;
for $i = 0$ **to** $it_{\max}$ **do**
  $\phi^{(i+1)} = \operatorname{prox}_{\sigma \mathfrak{i}_\delta^*}(\phi^{(i)} + \sigma A \bar{u}^{(i)})$;
  $u^{(i+1)} = \operatorname{prox}_{\lambda \Phi}(u^{(i)} - \lambda \nabla E(u^{(i)}) - \lambda A^* \phi^{(i+1)})$;
  $\bar{u}^{(i+1)} = 2u^{(i+1)} - u^{(i)} + \lambda \nabla E(u^{(i)}) - \lambda \nabla E(u^{(i+1)})$;
  if *convergence* **then**
    $u^* = u^{(i+1)}$;
    $\phi^* = \phi^{(i+1)}$;
    break

The adjoint A^* of A is defined via the relation

$$\langle \psi, Au \rangle_2 = \langle A^* \psi, u \rangle,$$

where the weighted inner product (4.2) is used in the right-hand side. Denoting by W the Gramian of that weighted inner product, we obtain for the matrix representation of the adjoint the identity $A^* \psi = W^{-1} A^{\mathrm{T}} \psi$, where A^{T} is the transpose matrix of the matrix A.

For a convex, lower-semicontinuous and proper functional Ψ defined on a Hilbert space, the proximal operator is defined by

$$\operatorname{prox}_\Psi(u) = \arg\min_v \Psi(v) + \frac{1}{2} \| v - u \|^2.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Algorithm 2: Primal–dual algorithm for the daJKO scheme, where DAJKOStep refers to Algorithm 1. Line 5 ensures that the initial condition for the next iteration is set correctly.

Input : ρ_0
Output : $(\rho^{(k)})_{k=1}^{N_{JKO}}$

1 *Initialize* $\rho^{(1)} = \rho_0$, $\phi^{(1)} = 0$, $u^{(1)} = 0$, $u^{(1)}_{1:N_x+1,1} = \rho_0$;
2 **for** $n = 1$ **to** $N_{JKO} - 1$ **do**
3 $\quad u^{(n+1)}, \phi^{(n+1)} = \texttt{DAJKOStep}(u^{(n)}, \phi^{(n)}, it_{\max}, \lambda, \sigma, A, b, \delta)$;
4 $\quad \rho^{(n+1)} = u^{(n+1)}_{1:N_x+1,N_t+1}$;
5 $\quad$ *Update* b;

Note that the norm is the one from the corresponding Hilbert space.

Next, we show that the proximal operators used in Algorithm 1 can be evaluated efficiently. By definition of the proximity operator and the weighted norm in (4.2), we need to consider the following minimization problem for a given $u_h = (\rho_{j,k}, m_{j,})$:

$$\begin{aligned}&\inf_{v_h=(\rho^v_{j,k},m^v_{j,k})} \left(\sum_{k=1}^{N_t+1} \sum_{j=1}^{N_x+1} w_k^t w_j^x \Phi(\rho^v_{j,k}, m^v_{j,k}) \right) + \frac{1}{2\lambda} \|v_h - u_h\|^2 \\ &= \inf_{(\rho^v_{j,k},m^v_{j,k})} \sum_{k=1}^{N_t+1} \sum_{j=1}^{N_x+1} w_k^t w_j^x \left(\Phi(\rho^v_{j,k}, m^v_{j,k}) + \frac{1}{2\lambda} (|m^v_{j,k} - m_{j,k}|^2 + (\rho^v_{j,k} - \rho_{j,k})^2) \right).\end{aligned}$$

Therefore, the minimizer is given per grid point by (cf. [43, Proposition 1])

$$\operatorname{prox}_{\lambda\Phi}(\rho_{j,k}, m_{j,k}) = \begin{cases} (\rho^*_{j,k}, m^*_{j,k}) & \text{if } \rho^*_{j,k} > 0, \\ (0,0) & \text{otherwise,} \end{cases} \tag{4.13}$$

with $m^*_{j,k} = \rho^*_{j,k} m_{j,k}/(\rho^*_{j,k}+\lambda)$, and $\rho^*_{j,k}$ the largest positive real root of the cubic polynomial

$$P_{j,k}(x) = (x - \rho_{j,k})(x+\lambda)^2 - \frac{\lambda}{2}|m_{j,k}|^2,$$

which can be computed using Cardano's formula. (We refer the reader to the appendix for details on the calculation of the largest real root.) Therefore, every iterate $u^{(i)} = (\rho^{(i)}_{j,k}, m^{(i)}_{j,k})$ in Algorithm 1 satisfies $\rho^{(i)}_{j,k} \geq 0$. The computation of $\operatorname{prox}_{\lambda\Phi}(\cdot)$ per grid point allows for an efficient implementation and a straightforward parallelization.

The functional i^*_δ showing up in Algorithm 1 is the Legendre–Fenchel transform of i_δ, which is defined as

$$\mathrm{i}^*_\delta(\phi) = \max_\psi \langle \psi, \phi \rangle_2 - \mathrm{i}_\delta(\psi),$$

where the maximum is taken over all vectors ψ with dimension corresponding to the number of rows of A, and the inner product $\langle \cdot, \cdot \rangle_2$ is the standard Euclidean inner product on that space. Moreau's identity [40] implies that

$$\operatorname{prox}_{\sigma \mathrm{i}^*_\delta}(\phi) = \phi - \sigma \operatorname{prox}_{\mathrm{i}_\delta}(\phi/\sigma).$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

We note that

$$\operatorname{prox}_{\mathrm{i}_\delta}(\phi) = \arg\min_{\psi} \mathrm{i}_\delta(\psi) + \frac{1}{2}\|\psi - \phi\|_2^2 = \operatorname{proj}_{B_\delta}(\phi),$$

where the i-th slice of the projection onto B_δ is given by

$$\big(\operatorname{proj}_{B_\delta}(x)\big)_i = \begin{cases} x_i & \|x_i - b_i\|_2 \le \delta_i, \\ \delta_i \dfrac{x_i - b_i}{\|x_i - b_i\|_2} + b_i & \text{otherwise,} \end{cases} \qquad i \in \{1,2,3,4\}.$$

Summarizing, both proximal operators used in Algorithm 1 can be applied efficiently.

Next, let us compute the gradient of the discrete energy functional F_h, defined in (4.1), which is defined via

$$dF_h(u_h)[v_h] = \langle \nabla_u F_h(u_h), v_h \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of the grid function induced by the composite trapezoidal rule defined in (4.2). For $v_h = (\rho^v_{j,k}, m^v_{j,k})_{j,k}$, we have that

$$\begin{aligned} &dF_h(u_h)[v_h] \\ &\quad = \sum_{j=1}^{N_x+1} w^x_j \left(U'(\rho^1_j)\rho^v_{j,N_t+1} + V_j \rho^v_{j,N_t+1} + \sum_{i=1}^{N_x+1} w^x_i W_{i,j} \rho^1_i \rho^v_{j,N_t+1} \right) \\ &\quad = \sum_{k=1}^{N_t+1} \sum_{j=1}^{N_x+1} w^x_j w^t_k \left(\left(U'(\rho^1_j) + V_j + \sum_{i=1}^{N_x+1} w^x_i W_{i,j} \rho^1_i \right) \Big/ w^t_{N_t+1} \right) \rho^v_{j,k} \delta_{N_t+1,k}. \end{aligned}$$

Hence, the gradient is given by

$$(\nabla_u F_h(u_h))_{j,k} = \left(\frac{\delta_{k,N_t+1}}{w^t_k} \left(U'(\rho^1_j) + V_j + \sum_{i=1}^{N_x} w^x_i W_{i,j} \rho^1_i \right),\ 0 \right), \tag{4.14}$$

where the zero entry corresponds to the variation in $m_{j,k}$. As a stopping criterion in line 6 of Algorithm 1, we require that

$$\max\left\{ \frac{\|u_h^{(n+1)} - u_h^{(n)}\|}{\|u_h^{(n)}\|}, \frac{\|\phi^{(n+1)} - \phi^{(n)}\|_2}{\|\phi^{(n)}\|_2}, \frac{|E_h(u^{(n+1)}) - E_h(u^{(n)})|}{E_h(u^{(n)})} \right\} < tol.$$

5. Examples. We will illustrate our method on the following two non-linear PDEs: the porous-medium equation (PME) [49] and a variant of the Patlak–Keller–Segel model for the motion of bacteria under the influence of a chemical signal introduced in [7].

5.1. Porous-medium equation. The PME has a number of physical applications, examples being the description of the flow of an isentropic gas through a porous medium or the study of groundwater infiltration. Given $m \ge 1$, it reads as

$$\partial_t u = \Delta u^m, \quad \text{in } \Omega \times (0,T). \tag{5.1}$$

As shown in the seminal work of Otto [42], this is indeed a gradient flow with respect to the Wasserstein distance for the energy functional

$$F(u) = \int_\Omega U(u)\,\mathrm{d}x,$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

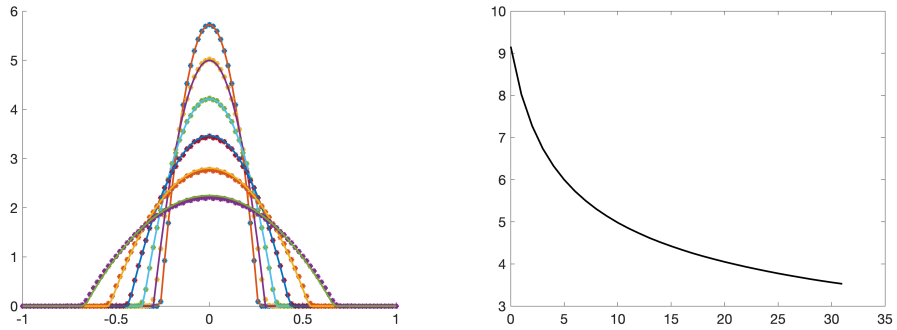

FIG. 5.1. *Left: Barenblatt profiles (solid) and their approximations (dotted) generated by the JKO scheme for different times $t = (k-1)\tau$ with $k = 2^i$, $i \in \{0, \dots, 5\}$. Right: Energy decay during the JKO scheme for simulating a Barenblatt profile, i.e., without data terms. On average, Algorithm 1 required 6000 steps to converge.*

where the internal energy is defined as

$$U(u) = \begin{cases} [1/(m-1)]u^m, & m > 1, \\ u\log(u), & m = 1. \end{cases}$$

We supplement (5.1) with the following initial and boundary conditions:

$$u(x,0) = u_0(x) \quad \text{in } \Omega \quad \text{and} \quad \nabla u \cdot n = 0 \quad \text{on } \partial\Omega \times (0,T).$$

For $m > 1$, the PME has finite speed of propagation, so that for compactly supported initial data the solution remains compactly supported also for later times [49]. The solutions show a self-similar behaviour and in spatial dimension one an explicit form, the Barenblatt solution, is known. It is given as

$$u_b(x,t) = (t+t_0)^{-1/(m+1)} \left(C - \frac{m-1}{2m(m+1)} x^2 (t+t_0)^{-2/(m+1)} \right)_+^{1/(m-1)}, \tag{5.2}$$

for $C, t_0 > 0$.

In Figure 5.1 we show some Barenblatt profiles with $m = 2$, $t_0 = 10^{-3}$, and $C = (3/16)^{1/3}$, together with their numerical approximations generated by the daJKO scheme using no data terms. For the numerical algorithm, we used $\tau = 5 \times 10^{-4}$, $N_t = 10$, $N_x = 100$, $\delta_i = 10^{-5} = tol$, $\lambda = 0.2$, and $\sigma = 4 \times 10^{-4}$ such that $\lambda\sigma\|AA^*\|_2 = 0.9 < 1$, which is required for the convergence theory [51]. We observe a good match between the exact solution and its numerical approximation. Furthermore, the energy is decaying monotonically. Here, we do not aim for a full convergence study of the JKO scheme, but refer to [15] for results in this direction, which might be carried over to our discretization. Instead, let us discuss the use of data to correct the gradient flow in the presence of uncertainties.

5.1.1. Unknown initial condition. As a first test of our data assimilation framework, we assume that the true, but unknown, initial condition is given by (5.2) with $t = 0$, $m = 2$, $t_0 = 10^{-3}$, and $C = (3/16)^{1/3}$. As a perturbed initial condition, we employ a shifted Barenblatt solution

$$u(x,0) = u_b(x - \bar{x}, \bar{t}), \quad \text{for some } \bar{x}, \bar{t} > 0. \tag{5.3}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

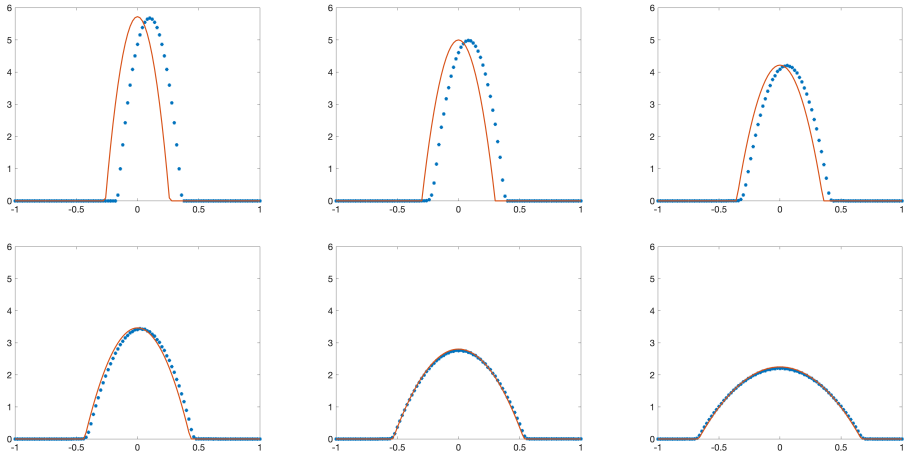

FIG. 5.2. *Barenblatt solution (solid red) and its approximation (dotted blue) using a shifted initial condition* (5.3) *with* $\bar{x} = 0.1$ *and* $\bar{t} = 0$ *and expectation* $B_h^1(u_h)$ *for the data term* ($\vartheta = 0$)*. The results are shown for* $k = 2^i$ *for* $i = 0, \dots, 5$ *from top left to bottom right.*

In order to correct for these perturbations, we assume a measurement operator $\mathcal{B}$, that employs the expected value and the variance, respectively,

$$\mathcal{B}^1(u) = \int_\Omega x u(x)\,\mathrm{d}x,$$
$$\mathcal{B}^2(u) = \int_\Omega (x - \mathcal{B}_1(u))^2 u(x)\,\mathrm{d}x,$$

which, after employing the trapezoidal rule, are discretized as follows:

$$B_h^1(u_h) = \sum_{j=1}^{N_x+1} w_j^x x_j \rho_j^1, \tag{5.4}$$

$$B_h^2(u_h) = \sum_{j=1}^{N_x+1} w_j^x (x_j - B_1(u_h))^2 \rho_j^1. \tag{5.5}$$

Here, as above, we write $\rho_j^1 = \rho_{j,N_t+1}$ for $u_h = (\rho_{j,k}, m_{j,k})$.

The measurement operator is then given by $B_h(u_h) = (B_h^1(u_h), \vartheta B_h^2(u_h))$ with either $\vartheta = 0$ or $\vartheta = 1$, depending on whether or not variance data is used. Similar to (4.14), we compute the gradients of the measurement operator as follows:

$$(\nabla_u B_h^1(u_h))_{j,k} = \frac{\delta_{k,N_t+1}}{w_k^t}(x_j, 0),$$
$$(\nabla_u B_h^2(u_h))_{j,k} = \frac{\delta_{k,N_t+1}}{w_k^t}\left((x_j - B_h^1[u_h])^2 - 2x_j \sum_{j'=0}^{N_x+1} w_{j'}^x (x_{j'} - B_h^1[u_h])\rho_{j'}^1, 0\right).$$

The penalty parameter is chosen $\theta = 1/200$ in the following experiments.

In our first experiment, we use a perturbed initial condition for the daJKO scheme, given by a spatially shifted profile as defined in (5.3) with $\bar{x} = 0.1$ and $\bar{t} = 0$. In order to correct

ETNA
Kent State University and
Johann Radon Institute (RICAM)

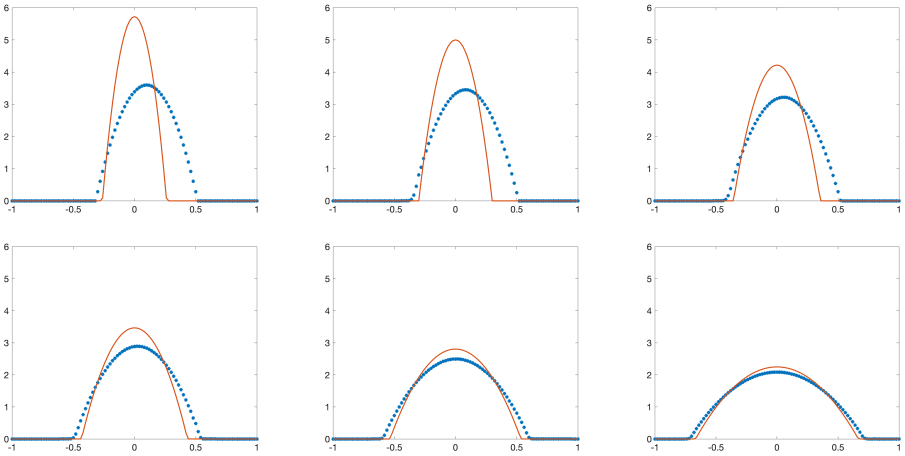

FIG. 5.3. *The same results as shown in Figure 5.2 but with a shifted initial condition (5.3) with* $\bar{x} = 0.1$ *and* $\bar{t} = 6\tau$.

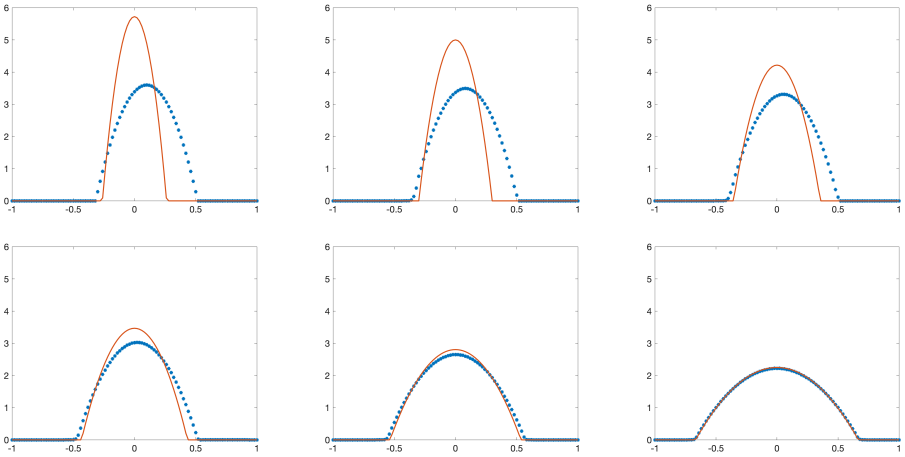

FIG. 5.4. *The same results as shown in Figure 5.3 but using also variance data* $B_h^2(u_h)$ *as data* ($\vartheta = 1$).

for this shift, we employ the expected value of the ground truth, i.e., $\vartheta = 0$. Figure 5.2 shows that the daJKO scheme is able to steer the numerical solution to the exact solution, thereby correcting for the perturbation in the initial condition. This might be expected, because the expected values effectively determines the maximum of the Barenblatt profile.

In the next experiment, we add to the spatial shift $\bar{x} = 0.1$ also a temporal perturbation $\bar{t} = 6\tau$ for the initial condition (5.3) of the daJKO scheme. In this situation, using only the expected value as data, we expect that the daJKO scheme can again compensate the shift. This is in agreement with the results shown in Figure 5.3. It seems that the numerical approximation generated by the daJKO scheme "waits" for the true solution to arrive while aligning its peak with that of the ground truth. The matching is, however, not as good as in the previous experiment, in which no time shift has been applied.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

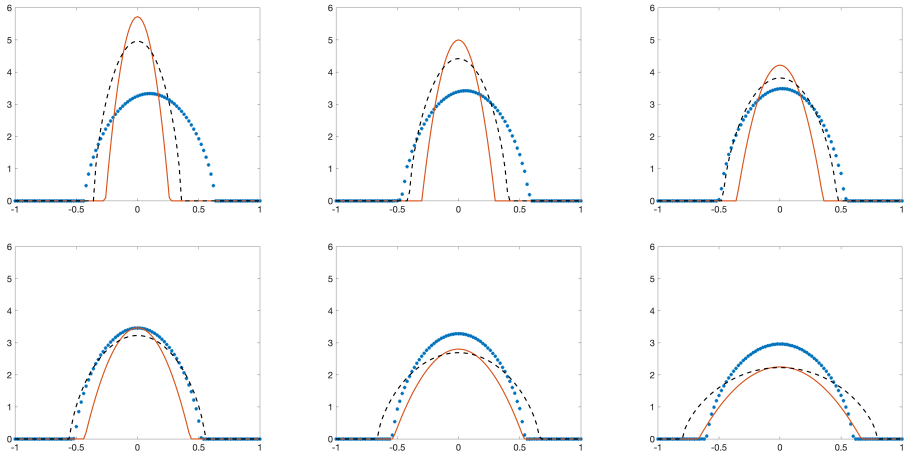

FIG. 5.5. *Barenblatt solution for $m^\dagger = 2$ (solid red), perturbed $m = 2.5$ (dashed black), and its daJKO approximation (dotted blue) using the perturbed $m = 2.5$, a shifted initial condition (5.3) with $\bar{x} = 0.1$ and $\bar{t} = 6\tau$, and expectation $B_h^1(u_h)$ and variance data $B_h^2(u_h)$ for the data term ($\vartheta = 1$). The results are shown for $k = 2^i$ for $i = 0, \dots, 5$ from top left to bottom right.*

In order to improve the data fitting, we repeat the experiment for $\bar{x} = 0.1$ and $\bar{t} = 6\tau$, using additional variance data of the ground truth, i.e., $\vartheta = 1$ in the definition of B_h after (5.5). Given that the value of m is known to us, we can determine the Barenblatt profile by its expected value and variance. Therefore, we expect that the daJKO scheme can match the true profile better, when the expected value and the variance are used as measurements. This expectation is confirmed by the results shown in Figure 5.4.

5.1.2. Unknown m. In addition to perturbed initial data, we consider model perturbations as a next experiment. We assume that $m^\dagger = 2$ describes the true dynamics, while the computational model employs the perturbed value $m = 2.5$. Except for $\theta = 1/400$, we keep the parameters as in the previous section. In particular, we use a shift $\bar{x} = 0.1$ and $\bar{t} = 6\tau$ for the initial condition.

In our first experiment, we use again the expected value and the variance of the Barenblatt profile (5.2) with the ground-truth values $m^\dagger$, $\bar{x} = 0$, and $\bar{t} = 0$. The results of the daJKO scheme are displayed in Figure 5.5, together with the ground-truth Barenblatt profile and a Barenblatt profile for the perturbed parameter $m = 2.5$. We observe that the daJKO scheme is able to correctly centre the numerical solution. Furthermore, by inspecting the supports of the three functions, we observe that using available data can correct for the speed of propagation. Comparing to the results of the previous section, however, the numerical approximation does not match the ground-truth solution very well. This can be explained by the fact that the Barenblatt profiles for $m = 2.5$ and $m^\dagger = 2$ have different masses.

Therefore, let us assume that we have access to the mass of the ground-truth Barenblatt profile. Incorporating the mass constraint via an observation operator similar to (5.4) counteracts enforcing (4.11). Despite this obvious mismatch, a computational consequence is that the underlying optimization, Algorithm 1, does not converge any more within 2×10^5 iterations. Therefore, we initialize the daJKO scheme differently, by scaling the perturbed initial condition computed from (5.2) and (5.3) with $\bar{x} = 0.1$, $\bar{t} = 6\tau$, and $m = 2.5$ such that its mass matches the mass of the ground truth. The results of this approach are displayed in Figure 5.6. In this case, the daJKO scheme is able to match the ground truth much better.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

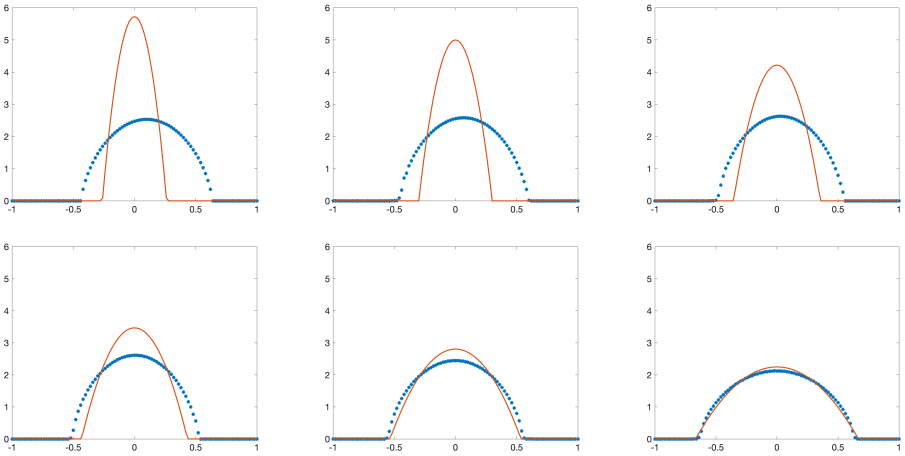

FIG. 5.6. *Barenblatt solution for $m^\dagger = 2$ (solid red) and its daJKO approximation (dotted blue) using the perturbed $m = 2.5$, a shifted and scaled initial condition (5.3) with $\bar{x} = 0.1$ and $\bar{t} = 6\tau$, and expectation $B_h^1(u_h)$ and variance data $B_h^2(u_h)$ for the data term ($\vartheta = 1$). The results are shown for $k = 2^i$ for $i = 0, \dots, 5$ from top left to bottom right.*

5.2. Chemotaxis with logarithmic kernel. Our second example is the chemotaxis model presented in [7]. Chemotaxis refers to the directed motion of bacteria in response to the gradient of a chemical signal. Usually the bacteria move along the negative gradient of the chemical, which, for example, drives them towards a food source. Denoting by $u = u(x,t)$ the density of bacteria and by $c = c(x,t)$ that of the chemical potential, the model reads as

$$\begin{aligned} \partial_t u &= \Delta u - \chi \nabla \cdot [u \nabla c], && t > 0,\ x \in \mathbb{R}^d, \\ c(t,x) &= -\frac{1}{d\pi} \int_{\mathbb{R}^d} \log|x-y| u(t,y)\,\mathrm{d}y, && t > 0,\ x \in \mathbb{R}^d, \\ u(0,x) &= u_0 \geq 0, && x \in \mathbb{R}^d. \end{aligned} \tag{5.6}$$

Again, this model constitutes a 2-Wasserstein gradient flow with respect to the energy

$$F(u) = \int_{\mathbb{R}}^d u(x) \log(u(x))\,\mathrm{d}x + \frac{\chi}{2d\pi} \int_{\mathbb{R}^d \times \mathbb{R}^d} \log|x-y| u(x) u(y)\,\mathrm{d}x\,\mathrm{d}y.$$

There has been great interest in the mathematical analysis of chemotaxis models, since many of them exhibit an interesting dichotomy: If their initial mass is below a given threshold, solutions exist for all time; if, on the other hand, the mass is large enough, a finite-time blow-up occurs [8]. For (5.6), the dichotomy also depends on the value of the sensitivity parameter χ. Denoting by M the mass on the initial datum, for $M\chi > 2d^2\pi$, finite-time blow-up will occur, while for $M\chi < 2d^2\pi$ solutions exist globally in time [12].

In this example, we demonstrate that the daJKO scheme can prevent blow-up. Therefore, we assume that the true value of $\chi = 2$ and the initial condition is given by

$$\rho_0(x) = G\left(x - \frac{1}{3}\right) + G\left(x + \frac{1}{3}\right) \quad \text{with } G(x) = \frac{1}{2\pi\eta^2} e^{-|x|^2/\eta^2}$$

and $\eta = 1/5$. We have that $\int_\Omega \rho_0\,\mathrm{d}x \approx 2.8$, and hence no blow-up occurs. If, however, a wrong value of $\chi = 10$ is used, the solution will blow up; see Figure 5.7. As in the previous

ETNA
Kent State University and
Johann Radon Institute (RICAM)

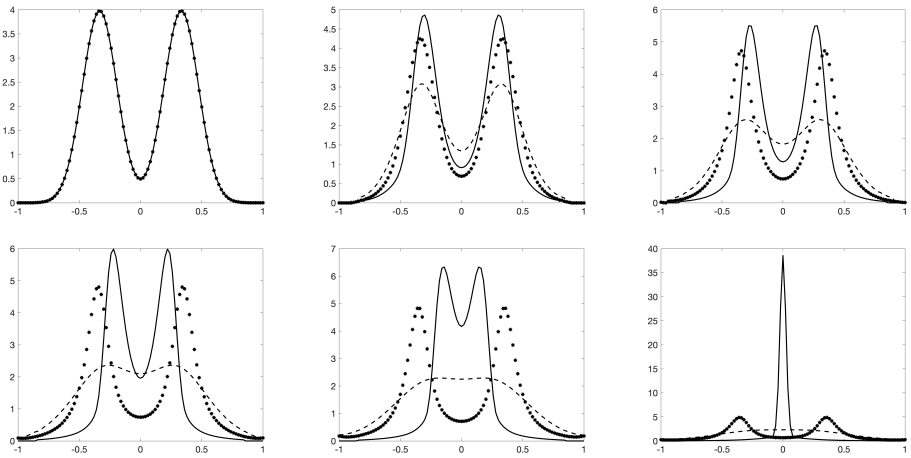

FIG. 5.7. *Solution to the chemotaxis equations (5.6) for $\chi = 2$ (dashed) and for $\chi = 10$ (solid) for $t = k\tau$ with $k = 10^{i-1}$, $i = 0, \dots, 5$ from top left to bottom right. We use different scalings of the axis. The solution of the daJKO scheme (dotted) remains bounded.*

section, we employ the expected value and the variance of the unperturbed solution ($\chi = 2$) to prevent blow-up in the perturbed model ($\chi = 10$); cf. (5.4) and (5.5) for the definition of the observations. We change τ to 10^{-3}, and keep the other parameters as in the previous section. Similar to [15], we regularize the logarithmic convolution kernel by approximating it at zero by an integral average over two grid cells. As can be seen from Figure 5.7, the daJKO approximation does not blow up in time. However, the daJKO approximation does not fit the ground truth as well as in the previous section, which might be explained by the rather large perturbation in χ and the lack of sufficient data. This is in line with the analytical results in [12], which show that, if the mass is above the critical mass, the second moment of the solution will eventually become negative, which is a contradiction to the non-negativity of the solution. Thus, it makes sense that steering the second moment via measurements is able to prevent the blow-up.

REMARK 5.1. Model (5.6) is a special case of the more general class of aggregation–diffusion equations of the form

$$\partial_t u = \nabla \cdot (\nabla u + u \nabla (K * u)),$$

where $K : \mathbb{R}^d \to \mathbb{R}$ denotes a fixed convolution kernel. They are also Wasserstein gradient flows with respect to the energy

$$F(u) = \int_\Omega u \log(u) \, \mathrm{d}x + \frac{1}{2} \iint_{\Omega \times \Omega} K(x-y) u(x) u(y) \, \mathrm{d}x \, \mathrm{d}y.$$

6. Outlook. We have presented a variational data assimilation approach based on a modified minimizing movement scheme together with a numerical scheme. We believe that the following extensions are interesting future research questions.

- *Non-linear mobilities and geometry identification.* Currently, we are able to treat equations of the form (1.4) which have a linear mobility $m(u) = u$. In many application, the mobility might also be non-linear, which yields

$$\partial_t u = \nabla \cdot \left(m(u) \nabla \frac{\delta F}{\delta u}(u) \right), \quad \Omega \times (0, T).$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Here, m is a concave function, e.g., $m(u) = u(1-u)$, which appears in models with volume filling [50]; see also [21] for model discovery in this context. For such equations, a modified Wasserstein distance is needed which includes the mobility. They can be obtained by replacing the term m^2/ρ in (2.2) by

$$\frac{J^2}{m(\rho)};$$

see [20, 35] for details. This clearly requires a modification of the numerical scheme as the proximal operator changes. Independent of the context of this work, an interesting question is whether it is possible to uniquely determine m from measurements of geodesics with respect to the modified distance.

- *Reaction terms and boundary conditions*. Another possible extension to a larger class of PDEs is to include reaction terms. This immediately results in the total mass being no longer conserved. Thus it is not feasible to work in the space of probability measures but one must use non-negative measures instead. In this case, the Wasserstein distance has to be replaced by the Fisher–Rao type distance, which allows for a change of mass [17]. In the case of non-homogeneous boundary conditions, different modifications are necessary; see [28] for a formal discussion.
- *Random effects*. So far, we have remained in a purely deterministic setting, different from the classical approach in data assimilation. It would be interesting to include both random measurement as well as model errors.
- *General geometries*. The present scheme is based on a finite difference approximation, which makes it difficult to handle more complicated geometries when going to higher dimensions. Here, finite volume [13] or finite element discretizations [1] for the divergence constraint could be used as long as the application of the proximal operators is computationally efficient.

Acknowledgments. The authors thank G. Heinze (TU Chemnitz) for useful discussions.

Appendix A. In order to compute the largest real root ρ^* required in the evaluation of the proximal operator of the action functional (4.13), we use a similar approach as in [15] and compute ρ^* explicitly using Cardano's formula. For the convenience of the reader, we provide some detail.

We start by introducing new variables $z = x - \rho$, $p = \rho + \lambda$, and $c = \lambda|m|^2/2$. Hence, we look for the largest real root z^* of the cubic equation

$$z(z+p)^2 = c.$$

Clearly, $z^* \geq \max\{0, -p\}$, and z^* is the only root in that range. In particular, $z^* + p \geq 0$, whence z^* is equivalently characterized by the solution of

$$z + p = \sqrt{c/z}, \quad z > \max\{0, -p\}.$$

Multiplying this equation by $\sqrt{z}$ and introducing $y = \sqrt{z}$ and $q = -\sqrt{c}$, we obtain the depressed cubic

$$y^3 + pt + q = 0.$$

Defining the discriminant $\Delta = (p/3)^3 + (q/2)^2$ and $C = (-q/2 + \sqrt{D})^{1/3}$, the three roots are given by

$$y_1 = C - p/(3C), \quad y_2 = \xi_1 C - p/(3\xi_1 C), \quad y_3 = \xi_2 C - p/(3\xi_2 C),$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

where $\xi_1 = (-1 + i\sqrt{3})/2$ and $\xi_2 = (-1 - i\sqrt{3})/2$ are two cube roots of 1. Note that $C \neq 0$, since $\Re(\sqrt{D}) \geq 0$. The largest real root ρ^* is then obtained by back-transformation and a simple post-processing.

REFERENCES

[1] G. ALBI, M. BURGER, J. HASKOVEC, P. MARKOWICH, AND M. SCHLOTTBOM, *Continuum modeling of biological network formation*, in Active Particles. Vol. 1. Advances in Theory, Models, and Applications, N. Bellomo, P. Degond, and E. Tadmor, eds., Model. Simul. Sci. Eng. Technol., Birkhäuser/Springer, Cham, 2017, pp. 1–48.

[2] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Lectures in Mathematics, ETH Zürich, Birkhäuser, Basel, 2008.

[3] R. A. ANTHES, *Data assimilation and initialization of hurricane prediction models*, J. Atmospheric Sci., 31 (1974), pp. 702–719.

[4] J.-D. BENAMOU AND Y. BRENIER, *A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem*, Numer. Math., 84 (2000), pp. 375–393.

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

[6] J. D. BENAMOU, G. CARLIER, AND M. LABORDE, *An augmented Lagrangian approach to Wasserstein gradient flows and applications*, in Gradient Flows: From Theory to Application, B. Düring, C.-B. Schönlieb, and M.-T. Wolfram, eds., ESAIM: Proceedings and Surveys, 54, EDP Sciences, Les Ulis, 2016, pp. 1–17.

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

[8] A. BLANCHET, J. DOLBEAULT, AND B. PERTHAME, *Two-dimensional Keller–Segel model: optimal critical mass and qualitative properties of the solutions*, Electron. J. Differential Equations, 2006 (2006), Art. No. 44, 32 pages.

[9] A. BRAIDES, *Local Minimization, Variational Evolution and* Γ*-Convergence*, Springer, Cham, 2014.

[10] A. BRAIDES, M. COLOMBO, M. GOBBINO, AND M. SOLCI, *Minimizing movements along a sequence of functionals and curves of maximal slope*, C. R. Math. Acad. Sci. Paris, 354 (2016), pp. 685–689.

[11] M. BURGER, I. HUMPERT, AND J.-F. PIETSCHMANN, *On Fokker–Planck equations with in- and outflow of mass*, Kinet. Relat. Models, 13 (2020), pp. 249–277.

[12] V. CALVEZ, B. PERTHAME, AND M. SHARIFI TABAR, *Modified Keller–Segel system and critical mass for the log interaction kernel*, in Stochastic Analysis and Partial Differential Equations, G.-Q. Chen, E. Hsu, and M. Pinsky, eds., vol. 429 of Contemp. Math., Amer. Math. Soc., Providence, 2007, pp. 45–62.

[13] C. CANCÈS, T. O. GALLOUËT, AND G. TODESCHI, *A variational finite volume scheme for Wasserstein gradient flows*, Numer. Math., 146 (2020), pp. 437–480.

[14] C. CANCÈS, T. GALLOUËT, M. LABORDE, AND L. MONSAINGEON, *Simulation of multiphase porous media flows with minimising movement and finite volume schemes*, European J. Appl. Math., 30 (2019), pp. 1123–1152.

[15] J. A. CARRILLO, K. CRAIG, L. WANG, AND C. WEI, *Primal dual methods for Wasserstein gradient flows*, Found. Comput. Math., 22 (2022), pp. 389–443.

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

[17] L. CHIZAT, G. PEYRÉ, B. SCHMITZER, AND F.-X. VIALARD, *An interpolating distance between optimal transport and Fisher–Rao metrics*, Found. Comput. Math., 18 (2018), pp. 1–44.

[18] S.-N. CHOW, W. HUANG, Y. LI, AND H. ZHOU, *Fokker–Planck equations for a free energy functional or Markov process on a graph*, Arch. Ration. Mech. Anal., 203 (2012), pp. 969–1008.

[19] M. CUTURI, *Sinkhorn distances: lightspeed computation of optimal transport*, in Advances in Neural Information Processing Systems Vol. 26, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, eds., Curran Associates, Red Hook, 2013.

[20] J. DOLBEAULT, B. NAZARET, AND G. SAVARÉ, *A new class of transport distances between measures*, Calc. Var. Partial Differential Equations, 34 (2009), pp. 193–231.

[21] H. EGGER, J.-F. PIETSCHMANN, AND M. SCHLOTTBOM, *Identification of chemotaxis models with volume-filling*, SIAM J. Appl. Math., 75 (2015), pp. 275–288.

[22] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, vol. 375 of Mathematics and its Applications, Kluwer, Dordrecht, 1996.

[23] M. ERBAR, *A gradient flow approach to the Boltzmann equation*, Preprint on arXiv, 2016. https://arxiv.org/abs/1603.00540

ETNA
Kent State University and
Johann Radon Institute (RICAM)

[24] A. ESPOSITO, R. S. GVALANI, A. SCHLICHTING, AND M. SCHMIDTCHEN, *On a novel gradient flow structure for the aggregation equation*, Preprint on arXiv, 2021. https://arxiv.org/abs/2112.08317
[25] A. ESPOSITO, F. S. PATACCHINI, A. SCHLICHTING, AND D. SLEPCEV, *Nonlocal-interaction equation on graphs: gradient flow structure and continuum limit*, Arch. Ration. Mech. Anal., 240 (2021), pp. 699–760.
[26] G. HEINZE, J.-F. PIETSCHMANN, AND M. SCHMIDTCHEN, *Nonlocal cross-interaction systems on graphs: nonquadratic Finslerian structure and nonlinear mobilities*, Preprint on arXiv, 2021. https://arxiv.org/abs/2107.11289
[27] ———, *Nonlocal cross-interaction systems on graphs: energy landscape and dynamics*, Preprint on arXiv, 2022. https://arxiv.org/abs/2107.11289
[28] I. HUMPERT, *Mathematical Models of Transport Phenomena with In- and Outflow*, PhD. Thesis, WWU Münster, Münster, 2012.
[29] K. V. HUYNH AND B. KALTENBACHER, *Some application examples of minimization based formulations of inverse problems and their regularization*, Inverse Probl. Imaging, 15 (2021), pp. 415–443.
[30] R. JORDAN, D. KINDERLEHRER, AND F. OTTO, *The variational formulation of the Fokker–Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17.
[31] B. KALTENBACHER, *Minimization based formulations of inverse problems and their regularization*, SIAM J. Optim., 28 (2018), pp. 620–645.
[32] J. KITAGAWA, Q. MÉRIGOT, AND B. THIBERT, *Convergence of a Newton algorithm for semi-discrete optimal transport*, J. Eur. Math. Soc. (JEMS), 21 (2019), pp. 2603–2651.
[33] K. LAW, A. STUART, AND K. ZYGALAKIS, *Data Assimilation*, vol. 62 of Texts in Applied Mathematics, Springer, Cham, 2015.
[34] G. LEGENDRE AND G. TURINICI, *Second-order in time schemes for gradient flows in Wasserstein and geodesic metric spaces*, C. R. Math. Acad. Sci. Paris, 355 (2017), pp. 345–353.
[35] S. LISINI AND A. MARIGONDA, *On a class of modified Wasserstein distances induced by concave mobility functions defined on bounded intervals*, Manuscripta Math., 133 (2010), pp. 197–224.
[36] J. MAAS, *Gradient flows of the entropy for finite Markov chains*, J. Funct. Anal., 261 (2011), pp. 2250–2292.
[37] P. A. MARKOWICH AND C. VILLANI, *On the trend to equilibrium for the Fokker–Planck equation: an interplay between physics and functional analysis*, in VI Workshop on Partial Differential Equations. Part II. (Rio de Janeiro, 1999), H. Frid, J. Hounie, R. Iório, D. Marchesin, A. Nachbin, and J. P. Zubelli, eds., Mat. Contemp. 19 (2000), Sociedade Brasileira de Matemática, Rio de Janeiro, 2000, pp. 1–29.
[38] D. MATTHES AND S. PLAZOTTA, *A variational formulation of the BDF2 method for metric gradient flows*, ESAIM Math. Model. Numer. Anal., 53 (2019), pp. 145–172.
[39] A. MIELKE, *A gradient structure for reaction–diffusion systems and for energy-drift–diffusion systems*, Nonlinearity, 24 (2011), pp. 1329–1346.
[40] J.-J. MOREAU, *Proximité et dualité dans un espace Hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
[41] F. OTTO, *Dynamics of labyrinthine pattern formation in magnetic fluids: a mean-field theory*, Arch. Rational Mech. Anal., 141 (1998), pp. 63–103.
[42] ———, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.
[43] N. PAPADAKIS, G. PEYRÉ, AND E. OUDET, *Optimal transport with proximal splitting*, SIAM J. Imaging Sci., 7 (2014), pp. 212–238.
[44] G. PEYRÉ AND M. CUTURI, *Computational optimal transport*, Found. and Trends Machine Learning, 11 (2019), pp. 355–607.
[45] S. REICH, *A nonparametric ensemble transform method for Bayesian inference*, SIAM J. Sci. Comput., 35 (2013), pp. A2013–A2024.
[46] F. SANTAMBROGIO, *Euclidean, metric, and Wasserstein gradient flows: an overview*, Bull. Math. Sci., 7 (2017), pp. 87–154.
[47] S. K. TAMANG, A. EBTEHAJ, P. J. VAN LEEUWEN, D. ZOU, AND G. LERMAN, *Ensemble Riemannian data assimilation over the Wasserstein space*, Nonlin. Processes Geophys., 28 (2021), pp. 295–309.
[48] S. K. TAMANG, A. EBTEHAJ, D. ZOU, AND G. LERMAN, *Regularized variational data assimilation for bias treatment using the Wasserstein metric*, Quart. J. Roy. Meteorolog. Soc., 146 (2020), pp. 2332–2346.
[49] J. L. VÁZQUEZ, *The Porous Medium Equation*, Oxford University Press, Oxford, 2007.
[50] D. WRZOSEK, *Volume filling effect in modelling chemotaxis*, Math. Model. Nat. Phenom., 5 (2010), pp. 123–147.
[51] M. YAN, *A new primal-dual algorithm for minimizing the sum of three functions with a linear operator*, J. Sci. Comput., 76 (2018), pp. 1698–1717.