A VARIATIONAL FORMULATION OF THE BDF2 METHOD FOR METRIC GRADIENT FLOWS

Daniel Matthes[1,*] and Simon Plazotta[1]

Abstract. We propose a variational form of the BDF2 method as an alternative to the commonly used minimizing movement scheme for the time-discrete approximation of gradient flows in abstract metric spaces. Assuming uniform semi-convexity – but no smoothness – of the augmented energy functional, we prove well-posedness of the method and convergence of the discrete approximations to a curve of steepest descent. In a smooth Hilbertian setting, classical theory would predict a convergence order of two in time, we prove convergence order of one-half in the general metric setting and under our weak hypotheses. Further, we illustrate these results with numerical experiments for gradient flows on a compact Riemannian manifold, in a Hilbert space, and in the L^2-Wasserstein metric.

1. Introduction

1.1. The main idea in short

This article is concerned with a particular temporal discretization for gradient flows in metric spaces. That is, we study the approximation of curves of steepest descent in the energy landscape of a functional $\mathcal{E} : \boldsymbol{X} \to \mathbb{R} \cup \{\infty\}$ with respect to a metric $\boldsymbol{d}$ on $\boldsymbol{X}$. Before we elaborate on our motivation and results, we briefly outline the concept in the simplest setting, namely when $\boldsymbol{X} = \mathbb{R}^d$, $\boldsymbol{d}$ is the Euclidean metric, and $\mathcal{E} \in C^\infty(\mathbb{R}^d)$, in which case the problem amounts to approximate solutions to

$$\dot{u} = -\nabla\mathcal{E}(u). \tag{1.1}$$

Assuming in addition that $\mathcal{E}$ is uniformly semi-convex, *i.e.*, $\nabla^2\mathcal{E} \geq \lambda\mathbf{1}_d$ for some $\lambda \in \mathbb{R}$, then it follows that the implicit Euler method with any sufficiently small uniform time step $\tau > 0$,

$$\frac{u_\tau^k - u_\tau^{k-1}}{\tau} = -\nabla\mathcal{E}(u_\tau^k), \tag{1.2}$$

is well-defined, *i.e.*, the initial condition u_τ^0 determines the entire sequence $(u_\tau^k)_{k\in\mathbb{N}}$ uniquely. It is further well-known that this is a first order approximation of the true solution u to (1.1), *i.e.*, $u_\tau^k = u(k\tau) + O(\tau)$ as $\tau \to 0$.

Our point is that under the same semi-convexity hypothesis, also the second order Backward Differentiation Formula (BDF2) method,

$$\frac{3u_\tau^k - 4u_\tau^{k-1} + u_\tau^{k-2}}{2\tau} = -\nabla\mathcal{E}(u_\tau^k), \tag{1.3}$$

is well-defined and convergent for well-prepared initial data (u_τ^0, u_τ^1). The strength of the BDF2 method in comparison to the implicit Euler scheme is that the former – at least in the smooth setting at hand – converges to second order in τ.

Studies on the BDF2 scheme in the above mentioned ODE setting have been an active topic in the 1950's and 1960's [11,12]. Subsequently, BDF2-based problem adapted methods for the integration of flows on (finite-dimensional) submanifolds $\boldsymbol{X}$, *e.g.*, for ODEs with constraints, have been proposed and analyzed, see *e.g.* [10,21]. More recently, the method in the Hilbertian setup, where $\boldsymbol{X}$ is a Hilbert space and $\boldsymbol{d}$ is induced by the norm, has attracted a lot of attention. There is now a rich literature on convergence results, particularly for very general nonlinear right-hand sides, see *e.g.* [1,2,4,8,16–18,27,33,36,51]. The analysis appears to be more or less complete now, at least under reasonable conditions on the nonlinearity.

We are apparently the first to analyze (a variational formulation of) the BDF2 method *for approximation of gradient flows in abstract metric spaces*, and to prove its convergence just under the hypothesis of semi-convexity. Our proof is different from the one in ODE text books [15,20,26], also from the ones typically given in the Hilbertian setting, like in [16]. The key difference is that due to the possible "roughness" of the metric space $\boldsymbol{X}$, there is no appropriate notion of smooth solution for the gradient flow (in general, there does not even exist a good definition for the differentiability of a curve). Hence, we cannot invoke error estimates that rely on Taylor expansions around the limiting solution.

These difficulties are already present in the convergence analysis of the implicit Euler method in metric spaces, and have been overcome in [3] by formulating all essential estimates in a robust way that requires no smoothness. Naturally, the strategy for our own convergence analysis of the variational BDF2 method is inspired by that from [3], and there are various similarities also on the technical level. For instance, being unable to estimate the error between the genuine and the approximating discrete solutions directly, we resort to a Cauchy-type argument that compares discrete solutions with different time steps as in Section 4.1 of [3]. Further, the basis for the control of the local error is a convexity inequality for the variational functional, which estimates the change of distance to a fixed "observer point" during one iteration: this is Corollary 4.1.3 of [3] for the implicit Euler method, and (4.10) for the BDF2 method. While the accumulation of the global error is relatively easy to control for the one-step Euler discretization, see Section 4.4 of [3], this is an extremely tedious piece of work for our two-step method.

Our approach yields a control on the global approximation error, which is not of order two, but only of order one-half. We also provide an example to show that indeed, even for specific, seemingly harmless choices of $(\boldsymbol{X}, \boldsymbol{d})$ and $\mathcal{E}$, convergence takes place at first order only. In view of the results in Section 4.4 of [3] on the implicit Euler method, it seems likely that our variational BDF2 converges to first order in general. Currently, we are not able to close the apparent gap between order one-half and order one. There is little hope to adapt the methods leading to improved convergence in Section 4.4 of [3], since there, profound properties of the Yosida regularization play a pivotal role in estimating the local error. No comparable estimates are known for our BDF2 functional, and it seems unlikely that an appropriate surrogate exists *e.g.* for the duality formula for the slope ([3], Lem. 3.1.5). And according to our general philosophy, that we describe below, any further investigations in the direction of improving the rate beyond one-half appear rather pointless.

We emphasize that the proven slow convergence order one-half does not contradict our initial intention of providing a method of faster convergence than the implicit Euler one. Indeed, *if* the approximated solution is smooth enough (which, in specific situations, can often be verified a posteriori by considering it in a different setting or in an ambient space), *then* the classical convergence proofs from text books apply and yield the desired rate of order two. That philosophy is justified by a series of numerical experiments that all show second order convergence. *Our contribution is that – regardless of the regularity of the limiting solution under consideration –*

convergence of the method is guaranteed, even with an explicit rate. And our proof utilizes solely the variational structure of the scheme (1.3) *and the semi-convexity hypothesis on* $\mathcal{E}$.

We remark that other, conceptually different approaches to construct time discretizations with formally higher order have been investigated, for instance, variational formulations of Runge–Kutta methods [34, 37]. However, there is no analytically proven rate of convergence available so far, not even order one-half.

1.2. Metric gradient flows

For definiteness, we are working inside the abstract framework developed in the first part of the book [3]; only a few basic notations from that comprehensive theory will be relevant to us, and these are summarized in Section 2 below. Although our considerations are very general, we have three specific settings in mind. The first is that of gradient flows for smooth functions on a finite dimensional compact manifold, the second concerns uniformly semi-convex functionals on Hilbert spaces, and in the third, we consider flows for uniformly displacement semi-convex functionals on the space $\boldsymbol{X} = \mathcal{P}_2(\Omega)$ of probability measures with respect to the L^2-Wasserstein distance $\boldsymbol{d} = \mathbf{W}_2$.

Particularly the third setting has gained a lot of popularity in the past two decades, see [50, 52, 53] for a general introduction to the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ and see [43] for the concept of displacement convexity. It has been shown that a variety of important nonlinear dissipative evolution equations can be cast in the form of a Wasserstein gradient flow, from linear, nonlinear, and non-local Fokker–Planck equations [9, 28, 46] over fourth order fluid and quantum models [22, 23, 41] to chemotaxis systems [6, 7, 54]. The L^2-Wasserstein framework has been extended in many different directions, that allow to consider reaction diffusion equations [24, 44], Poisson–Nernst–Planck equations [32], multi-component fluid systems [35], and Cahn–Hilliard equations [38], as well as Markov chains [40, 45] as gradient flows in suitable metrics.

We shall not discuss the manifold advantages of such a variational formulation of the evolution, but focus only on the fact that under certain convexity properties of $\mathcal{E}$, a gradient flow can be rather easily obtained as the limit of a convergent time-discrete approximation, for instance by the variational forms of the implicit Euler and BDF2 methods discussed below. While the order of convergence of the time discretization is usually irrelevant for existence proofs (and the simpler Euler scheme might be advantageous over the more complicated BDF2 method), it is a limiting factor for efficiency of numerical discretizations that are based on these approximations. Indeed, a variety of structure preserving full (spatio-temporal) discretizations for numerical solution of Wasserstein gradient flows that have been developed recently, see *e.g.* [5, 29, 47], and essentially all of these methods discretize by implicit Euler in time direction. It seems that the variational character of the Euler scheme is essential to inherit desirable properties – like monotonicity of the energy – from the original flow to its discrete approximation, and is even the key element for convergence proofs [42]. *One of the conclusions from our work is that the variational BDF2 method introduced below might give a better approximation in time, without losing convergence or the flow's variational character.*

1.3. Minimizing movement scheme

Before we introduce our own temporal discretization, we recall some properties of the celebrated minimizing movement scheme, also known as implicit Euler method or JKO stepping, depending on the context. Let the metric space $(\boldsymbol{X}, \boldsymbol{d})$ and the functional $\mathcal{E}$ be fixed; we wish to construct a curve of steepest descent emerging from the initial data $u_0 \in \boldsymbol{X}$. For metric gradient flows in the sense of [3, 14], that scheme is defined as follows.

For each sufficiently small time step $\tau > 0$, let an initial condition u_τ^0 be given that approximates u^0. Then define inductively a discrete solution $(u_\tau^k)_{k \in \mathbb{N}}$ such that each u_τ^k with $k = 1, 2, \ldots$ is a minimizer of the Yosida-penalized energy functional

$$w \mapsto \Phi(\tau, u_\tau^{k-1}; w) := \frac{1}{2\tau} \boldsymbol{d}^2(u_\tau^{k-1}, w) + \mathcal{E}(w).$$

There are various "soft" conditions that guarantee well-definedness of this scheme, *i.e.*, the inductive solvability of the minimization problems; one is formulated as Condition 1.1 below. It is easy to verify that, in the trivial setting discussed in Section 1.1, the Euler–Lagrange equation for minimization of $\Phi(\tau, u_\tau^{k-1}; \cdot)$ yields precisely the induction formula (1.2) for the implicit Euler scheme.

One of the remarkable strengths of this method in the abstract setting is its stability: by an easy induction argument, one proves that for arbitrary $0 \le m \le n$,

$$\mathcal{E}(u_\tau^n) + \frac{\tau}{2} \sum_{k=m+1}^{n} \left(\frac{\boldsymbol{d}(u_\tau^{k-1}, u_\tau^k)}{\tau} \right)^2 \le \mathcal{E}(u_\tau^m), \tag{1.4}$$

which immediately implies monotonicity of the energy, $\mathcal{E}(u_\tau^n) \le \mathcal{E}(u_\tau^m)$, as well as a τ-uniform bound on the "integrated kinetic energy". These bounds are usually sufficient to conclude the convergence of the discrete solution to a continuous curve u_* in the limit $\tau \searrow 0$. Some additional work is needed to prove that u_* is indeed a curve of steepest descent. A famous hypothesis that makes this last step work is the convexity condition ([3], Assumption 4.0.4):

Condition 1.1. *There exists a $\lambda \in \mathbb{R}$ such that for each sufficiently small $\tau > 0$, for each reference point $u \in \mathcal{D}(\mathcal{E})$, and for each pair of points $\gamma_0, \gamma_1 \in \boldsymbol{X}$, there is some curve $(\gamma_s)_{s \in [0,1]}$ connecting γ_0 to γ_1 along which $s \mapsto \Phi(\tau, u; \gamma_s)$ is uniformly convex of modulus $\frac{1}{\tau} + \lambda$.*

We remark that recent adaptations of the minimizing movement scheme, see *e.g.* [19, 48, 49], allow to treat also non-autonomous gradient flows along the same lines, as long as the time-dependent energy functional has some "convexity in the average".

1.4. BDF2 method

We propose the following construction of a discrete approximation $(u_\tau^k)_{k \in \mathbb{N}}$:

For each sufficiently small time step $\tau > 0$, let a pair of initial conditions (u_τ^0, u_τ^1) be given that approximate u^0. Then define inductively a discrete solution $(u_\tau^k)_{k \in \mathbb{N}}$ such that each u_τ^{k+1} with $k \in \mathbb{N}$ is a minimizer of the following functional,

$$w \mapsto \Psi(\tau, u_\tau^{k-1}, u_\tau^k; w) := \frac{1}{\tau} \boldsymbol{d}^2(w, u_\tau^k) - \frac{1}{4\tau} \boldsymbol{d}^2(w, u_\tau^{k-1}) + \mathcal{E}(w).$$

In the euclidean setting of Section 1.1, the minimizer u_τ^k satisfies the BDF2 recursion (1.3). The BDF2 method is known to converge to second order under suitable smoothness hypotheses, see *e.g.* [15]. A key feature of this approach is that a simple induction argument produces an intrinsic stability estimate that is similar to – although slightly weaker than – (1.4) above:

$$\mathcal{E}(u_\tau^n) + \frac{\tau}{4} \sum_{k=m+1}^{n} \left(\frac{\boldsymbol{d}(u_\tau^{k-1}, u_\tau^k)}{\tau} \right)^2 \le \mathcal{E}(u_\tau^m) + \frac{1}{4\tau} \boldsymbol{d}^2(u_\tau^{m-1}, u_\tau^m). \tag{1.5}$$

From (1.5) for $n = m + 1$, it is clear that the energy value $\mathcal{E}(u_\tau^k)$ is not necessarily diminished in each time step, but the potential loss of monotonicity is well controlled. Moreover, one still has the crucial bound on the "integrated kinetic energy". The validity of estimate (1.5) is related to the intrinsic *A-stability* of the BDF2 scheme (1.3), see [11, 12, 15, 26]. Since the implicit Euler (BDF1) and the BDF2 methods are *the only* A-stable backward differentiation formulae, it is not to be expected that one can invent a variant of these schemes with a consistency order higher than two, which also satisfies a stability estimate like (1.5). In this sense, our approach is optimal.

1.5. Main result

Our main result, stated in Theorem 5.1, can be informally rephrased as follows. We assume that $\mathcal{E}$ is lower semi-continuous, admits a sort of lower bound, and satisfies the following adaptation of Condition 1.1:

Condition 1.2. *There exists a $\lambda \in \mathbb{R}$ such that for each sufficiently small $\tau > 0$, for each pair of reference points $u, v \in \mathcal{D}(\mathcal{E})$, and for each pair of points $\gamma_0, \gamma_1 \in \boldsymbol{X}$, there is some curve $(\gamma_s)_{s \in [0,1]}$ connecting γ_0 to γ_1 along which $s \mapsto \Psi(\tau, u, v; \gamma_s)$ is uniformly convex of modulus $\frac{3}{2\tau} + \lambda$.*

Then, firstly, the induction described above is well-defined, *i.e.*, from each initial pair (u_τ^{-1}, u_τ^0) one obtains a unique discrete solution $(u_\tau^k)_{k \in \mathbb{N}}$. Secondly, these discrete solutions converge to a curve u_* of steepest descent. Thirdly, the convergence can be made quantitative: roughly, the distance of u_τ^k to the associated limit value $u_*(k\tau)$ is at most of order $\sqrt{\tau}$.

1.6. Outline of the paper

After recalling some basic notions from the theory of gradient flows in metric spaces in Section 2, we specify our hypotheses and discuss examples in Section 3, Section 4 is devoted to the well-posedness of the BDF2 scheme, and to the derivation of various *a priori* estimates. Section 5 is the heart of the paper, with the statement and the proof of our main result on convergence, see Theorem 5.1. In Section 6, we show numerical convergence of the methods on different examples.

2. Notations for gradient flows in abstract metric spaces

We briefly recall a few basic notations from the theory of gradient flows in abstract metric spaces, following [3]. Here and below, $(\boldsymbol{X}, \boldsymbol{d})$ is a separable and complete metric space.

Definition 2.1 (AC curves). A curve $u : [0, \infty) \to \boldsymbol{X}$ is said to be L^2-*absolutely continuous*, written as $u \in \mathrm{AC}^2([0, \infty), \boldsymbol{X})$, if there exists a function $m \in L^2_{\mathrm{loc}}([0, \infty))$ such that

$$\boldsymbol{d}(u(t), u(s)) \leq \int_s^t m(r) \, \mathrm{d}r \qquad \text{for all} \ \ 0 \leq s \leq t.$$

See Definition 1.1.1 of [3] for further details. It can be shown ([3], Thm. 1.1.2) that among all possible choices for m, there is a minimal one, called the *metric derivative* $|u'| \in L^2_{\mathrm{loc}}([0, \infty))$, given by

$$|u'|(t) := \lim_{s \to t} \frac{\boldsymbol{d}(u(s), u(t))}{|s - t|} \qquad \text{for a.e. } t.$$

The main definition is that of a *gradient flow* in the energy landscape of a functional $\mathcal{E} : \boldsymbol{X} \to \mathbb{R} \cup \{\infty\}$ with respect to the metric $\boldsymbol{d}$. Here we adopt the strong but rather restrictive notion of the *evolution variational inequality* (EVI). This is particularly well adapted to dealing with gradient flows in the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$, which will be one of our main examples.

Definition 2.2 (EVI). Let a proper and $\boldsymbol{d}$-lower-semicontinuous functional $\mathcal{E} : \boldsymbol{X} \to \mathbb{R} \cup \{\infty\}$ be given. We say that $\mathcal{E}$ generates a λ-contractive gradient flow on $(\boldsymbol{X}, \boldsymbol{d})$ with some $\lambda \in \mathbb{R}$ if the following is true:

For each initial condition $u_0 \in \mathcal{D}(\mathcal{E})$, there is a corresponding solution curve $u \in \mathrm{AC}^2([0, \infty), \boldsymbol{X})$ with $u(0) = u_0$, that satisfies the evolution variational inequality associated to $\mathcal{E}$:

$$\frac{1}{2}\boldsymbol{d}^2(u(t), w) - \frac{1}{2}\boldsymbol{d}^2(u(s), w) \leq \int_s^t \left[\mathcal{E}(w) - \mathcal{E}(u(r)) - \frac{\lambda}{2}\boldsymbol{d}^2(u(r), w) \right] \mathrm{d}r \tag{2.1}$$

holds for arbitrary $0 \leq s \leq t$ and for every $w \in \mathcal{D}(\mathcal{E})$.

Remark 2.3. The inequalities (2.1) are actually the integrated versions of the differential EVI,

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{d}^2(u(t),w) + \frac{\lambda}{2}\boldsymbol{d}^2(u(t),w) + \mathcal{E}(u(t)) \leq \mathcal{E}(w), \tag{2.2}$$

which is supposed to hold for a.e. $t > 0$. The integrated form (2.1) is more convenient for our purposes, and it is easy to show that validity of (2.1) for arbitrary $0 \leq s \leq t$ implies validity of (2.2) for a.e. $t > 0$.

For a detailed discussion of the role of the EVI in gradient flow theory, we refer to Chapter 4.0 of [3]. The EVI is a more restrictive characterization of gradient flows than, *e.g.*, the energy-dissipation-equality. Most notably, validity of the EVI implies that the gradient flow is λ-contractive on $(\boldsymbol{X}, \boldsymbol{d})$, so in particular, solutions are uniquely determined by their initial datum. Moreover, if the metric space $(\boldsymbol{X}, \boldsymbol{d})$ is "almost Euclidean" – for instance, $\boldsymbol{X}$ is a Hilbert space, or $\boldsymbol{X}$ is the space $\mathcal{P}_2(\Omega)$ of probability measures endowed with the Wasserstein metric $\mathbf{W}_2$ – then if $\mathcal{E}$ generates a λ-contractive gradient flow then $\mathcal{E}$ is uniformly semi-convex see [13]. Thus, (2.1) is not available for gradient flows of non-semi-convex functionals $\mathcal{E}$.

3. Setup, assumptions and examples

3.1. Definition of the method

Define the *BDF2 penalization* $\Psi : (0, \tau_*) \times \boldsymbol{X} \times \boldsymbol{X} \times \boldsymbol{X} \to \mathbb{R} \cup \{\infty\}$ of $\mathcal{E}$ by

$$\Psi(\tau, u, v; \cdot) : \boldsymbol{X} \to \mathbb{R} \cup \{\infty\}; \quad \Psi(\tau, u, v; w) := \frac{1}{\tau}\boldsymbol{d}^2(v, w) - \frac{1}{4\tau}\boldsymbol{d}^2(u, w) + \mathcal{E}(w).$$

The *discrete solution* (for $\mathcal{E}$ on $(\boldsymbol{X}, \boldsymbol{d})$) corresponding to a time step size $\tau \in (0, \tau_*)$ and pair of initial data $(u_\tau^0, u_\tau^1) \in \boldsymbol{X} \times \boldsymbol{X}$ is the sequence $(u_\tau^k)_{k \in \mathbb{N}}$, which is inductively obtained via

$$u_\tau^{k+1} \in \underset{w \in \boldsymbol{X}}{\operatorname{argmin}} \ \Psi(\tau, u_\tau^{k-1}, u_\tau^k; w) \tag{3.1}$$

for $k \in \mathbb{N}$. Finally, the *interpolated solution* $\bar{u}_\tau : [0, \infty) \to \boldsymbol{X}$ is obtained by piecewise constant interpolation of the values u_τ^k in time,

$$\bar{u}_\tau(0) = u_\tau^0, \quad \bar{u}_\tau(t) = u_\tau^k \quad \text{for } t \in ((k-1)\tau, k\tau] \quad \text{and } k \in \mathbb{N}. \tag{3.2}$$

3.2. Main assumptions

From now on, we assume that an energy functional $\mathcal{E} : \boldsymbol{X} \to \mathbb{R} \cup \{\infty\}$ with nonempty domain $\mathcal{D}(\mathcal{E})$ is given, which satisfies the following hypotheses.

(E1) **Semi-continuity:** $\mathcal{E}$ is sequentially lower semi-continuous on $(\boldsymbol{X}, \boldsymbol{d})$:

$$u_n \xrightarrow{\boldsymbol{d}} u \quad \Longrightarrow \quad \mathcal{E}(u) \leq \liminf_{n \to \infty} \mathcal{E}(u_n).$$

(E2) **Coercivity:** There exist $\tau_* > 0$ and $u_* \in \boldsymbol{X}$, such that

$$c_* := \inf_{v \in \boldsymbol{X}} \frac{1}{2\tau_*}\boldsymbol{d}^2(u_*, v) + \mathcal{E}(v) > -\infty.$$

(E3) **Semi-convexity:** There exists a constant λ such that for every $u, v, \gamma_0, \gamma_1 \in \mathcal{D}(\mathcal{E})$ and every $\tau \in [0, \tau_*)$, there exists a continuous curve $\gamma_{(\cdot)} : [0, 1] \to \boldsymbol{X}$ joining the given end points γ_0 and γ_1, along which the penalized energy Ψ satisfies

$$\Psi(\tau, u, v; \gamma_s) \leq (1-s)\Psi(\tau, u, v; \gamma_0) + s\Psi(\tau, u, v; \gamma_1) - \frac{1}{2}\left(\frac{3}{2\tau} + \lambda\right)s(1-s)\boldsymbol{d}^2(\gamma_0, \gamma_1). \tag{3.3}$$

Moreover, without loss of generality, we assume that

$$\lambda \leq 0 \quad \text{and} \quad (-\lambda)\tau_* \leq \frac{1}{2} \quad \text{and} \quad \tau_* \leq 1. \tag{3.4}$$

Note that for $0 < \tau < \tau_*$, the last term on the right hand side of (3.3) is positive for $\gamma_0 \neq \gamma_1$ and $0 < s < 1$, implying that $s \mapsto \Psi(\tau, u, v; \gamma_s)$ is strictly convex.

Remark 3.1. Assumptions (E1) & (E2) are standard minimal hypotheses on the energy in the context of metric gradient flows. (E3) is a reformulation of Condition 1.2 from the introduction. As mentioned there, it plays an analogous role for the BDF2 discretization as Condition 1.1 plays for the minimizing movement scheme.

3.3. Examples

In this section we discuss three general situations in which the convexity assumption (E3) is satisfied, namely that of uniformly semi-convex functionals $\mathcal{E}$ on a Hilbert space H, that of semi-convex C^1-functions $\mathcal{E}$ on Riemannian manifolds of non-negative cross-curvature, and that of functionals $\mathcal{E}$ on the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$ that are uniformly displacement semi-convex.

3.3.1. Hilbert space

Uniformly semi-convex functionals on Hilbert spaces provide a class of fairly easy examples for the validity of assumption (E3), thanks to the linear structure of the space.

Theorem 3.2. *Assume that the metric space $(\boldsymbol{X}, \boldsymbol{d})$ is a Hilbert space $\boldsymbol{X} = H$, with the distance $\boldsymbol{d}$ induced by the norm $\|\cdot\|$. Assume further that $\mathcal{E}$ is uniformly semi-convex with modulus λ. Then (E3) is satisfied, with γ_s the straight line between γ_0, γ_1 and with the same λ.*

Proof. Let $\gamma_0, \gamma_1 \in \mathcal{D}(\mathcal{E})$ as well as $u, v \in \mathcal{D}(\mathcal{E})$ and $\tau > 0$ be given. We verify (3.3) for the particular curve $\gamma_s := (1-s)\gamma_0 + s\gamma_1$. On the one hand, by the convexity hypothesis on $\mathcal{E}$, we know that

$$\mathcal{E}(\gamma_s) \leq (1-s)\mathcal{E}(\gamma_0) + (1-s)\mathcal{E}(\gamma_1) - \frac{\lambda}{2}s(1-s)\|\gamma_0 - \gamma_1\|^2. \tag{3.5}$$

On the other hand, a direct calculation using the property of the scalar product yields

$$\|\gamma_s - v\|^2 - \frac{1}{4}\|\gamma_s - u\|^2 = (1-s)\big(\|\gamma_0 - v\|^2 - \frac{1}{4}\|\gamma_0 - u\|^2\big)$$
$$+ s\big(\|\gamma_1 - v\|^2 - \frac{1}{4}\|\gamma_1 - u\|^2\big) - \frac{3}{4}s(1-s)\|\gamma_0 - \gamma_1\|^2. \tag{3.6}$$

Adding $\frac{1}{\tau}$ times (3.6) to (3.5) yields (3.3). $\qquad\square$

3.3.2. Riemannian manifolds

Another situation of interest is that of the gradient flow on a compact smooth Riemannian manifold $(\mathcal{M}, \mathbf{g})$, which is induced by a semi-convex function $\mathcal{E} \in C^1(\mathcal{M})$. Here, our very general approach is clearly not optimal: in that finite dimensional setting, gradient flows can be characterized in a direct way instead of using the EVI (2.1). Further, there are explicit and local variants of the BDF2 method (avoiding the global minimization of Ψ in each time step), see *e.g.* [25], which are simpler to implement, and whose convergence is expected under more easily varifiable hypotheses than (E3). Still, for the sake of completeness, we shall detail a sufficient criterion for applicability of our results in that situation.

To indicate why the verification of (E3) indeed poses a (surprisingly hard) problem, observe that it is in general *not* possible to use the geodesic $\tilde{\gamma}_{(\cdot)}$ for the curve connecting γ_0 to γ_1 in (3.3). Indeed, for $s \mapsto \Psi(\tau, u, v; \tilde{\gamma}_s)$ to be uniformly convex of modulus $\frac{3}{2\tau} + \lambda$, independently of u and v, one would essentially need that both $s \mapsto \boldsymbol{d}^2(v, \gamma_s)$ and $s \mapsto -\boldsymbol{d}^2(u, \gamma_s)$ are uniformly convex of modulus $\boldsymbol{d}^2(\gamma_0, \gamma_1)$. By Toponogov's theorem, the

first condition would imply that $\mathcal{M}$ has non-negative sectional curvature, and the later would imply that $\mathcal{M}$ has non-positive sectional curvature; hence, $\mathcal{M}$ would need to be flat.

A more appropriate class of connecting curves are *segments*, which are defined with the help of the exponential map exp as follows. Fix $v \in \mathcal{M}$, and let $\gamma_0, \gamma_1 \in \mathcal{M}$ lie outside of v's cut locus $\mathrm{cut}(v)$. Then, there are unique ξ_0, ξ_1 in the injectivity domain $I(v) \subset T_v\mathcal{M}$ of the exponential map $\exp_v : T_v\mathcal{M} \to \mathcal{M}$ such that $\exp_v(\xi_0) = \gamma_0$ and $\exp_v(\xi_1) = \gamma_1$. Further, assume that the straight line from ξ_0 to ξ_1 lies in $I(v)$. The segment $[\gamma_0, \gamma_1; v]_{(\cdot)} : [0, 1] \to \mathcal{M}$ with base v connecting γ_0 to γ_1 is then defined by $[\gamma_0, \gamma_1; v]_s = \exp_v((1 - s)\xi_0 + s\xi_1)$.

Kim and McCann ([31], Cor. 2.11) have established a sufficient criterion for the convexity of

$$[0, 1] \ni s \mapsto \boldsymbol{d}^2\big(v, [\gamma_0, \gamma_1; v](s)\big) - \boldsymbol{d}^2\big(u, [\gamma_0, \gamma_1; v]_s\big), \tag{3.7}$$

independently of $u \in \mathcal{M}$. Their hypotheses are as follows.

(KM0) The squared metric $\boldsymbol{d}^2(\cdot, \cdot)$, induced on $\mathcal{M}$ via $\mathbf{g}$, is C^4-regular outside of the cut locus.
(KM1) For each $v \in \mathcal{M}$, its injectivity domain $I(v)$ is convex, so segments $[\gamma_0, \gamma_1; v]$ can be defined for arbitrary $\gamma_0, \gamma_1 \notin \mathrm{cut}(v)$.
(KM2) For each segment $[\gamma_0, \gamma_1; v]$, there is a dense subset $U \subset \mathcal{M}$, such that there is no $u \in U$ and no $s \in [0, 1]$ with $u \in \mathrm{cut}([\gamma_0, \gamma_1; v]_s)$.
(KM3) $(\mathcal{M}, \mathbf{g})$ has non-negative cross curvature; the definition is given in Appendix A.

Apart from (KM0), each of these conditions is rather demanding. A class of examples satisfying (KM0)–(KM3) are the round spheres $\mathbb{S}^d$. For these, (KM0)–(KM2) are easily verified since $\mathrm{cut}(v) = \{-v\}$ only contains the antipodal point, and $I(v)$ is the open d-dimensional ball of radius π, for each $v \in \mathbb{S}^d$. In contrast, the proof of (KM3) has been a challenge even for spheres, that has been mastered in Theorem 6.2 of [31]. It seems that – apart from products and quotients of spheres – no further explicit examples satisfying (KM0)–(KM3) are currently known.

Theorem 3.3. *Assume that $(\boldsymbol{X}, \boldsymbol{d})$ is a compact Riemannian manifold $(\mathcal{M}, \mathbf{g})$ that satisfies (KM0)–(KM3) above. Assume further that $\mathcal{E} \in C^1(\mathcal{M})$ is semi-convex. Then (E3) is satisfied with some $\lambda \in \mathbb{R}$, with $\gamma_{(\cdot)} := [\gamma_0, \gamma_1; v]_{(\cdot)}$.*

Proof. For given $u, v \in \mathcal{M}$ and $\gamma_0, \gamma_1 \in \mathcal{M} \setminus \mathrm{cut}(v)$, let $\gamma_{(\cdot)} := [\gamma_0, \gamma_1; v]_{(\cdot)}$; the result for general $\gamma_0, \gamma_1 \in \mathcal{M}$ follows by continuity a forteriori. Further, we shall assume that $\mathcal{E} \in C^2(\mathcal{M})$ during the computations. Since $\mathcal{E}$ is semi-convex, and $\mathcal{M}$ is compact, there is a global modulus $\lambda' \le 0$ of convexity, *i.e.*, $\mathrm{Hess}\,\mathcal{E}(v) \ge \lambda'$ as a quadratic form on each $T_v\mathcal{M}$. The final estimate (3.8) depends only on λ', so (E3) follows for general semi-convex $\mathcal{E} \in C^1(\mathcal{M})$ by approximation.

We split

$$\Phi(\tau, u, v; \gamma_s) = h_1(s) + h_2(s) + h_3(s),$$

with $h_1, h_2, h_3 : [0, 1] \to \mathbb{R}$ given by

$$h_1(s) = \frac{3}{4\tau}\boldsymbol{d}^2(v, \gamma_s), \quad h_2(s) = \frac{1}{4\tau}\big(\boldsymbol{d}^2(v, \gamma_s) - \boldsymbol{d}^2(u, \gamma_s)\big), \quad h_3(s) = \mathcal{E}(\gamma_s).$$

First, by definition of the segment $\gamma_{(\cdot)}$ via the exponential map, $s \mapsto \boldsymbol{d}^2(v, \gamma_s)$ is twice differentiable with

$$\frac{3}{4\tau}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\boldsymbol{d}^2(v, \gamma_s) \equiv \frac{3}{2\tau}\|\xi_1 - \xi_0\|_v^2,$$

where $\|\xi\|_v^2 = \mathbf{g}_v(\xi, \xi)$. Second, by the hypotheses (KM0)–(KM3), the result from Corollary 2.11 of [31] applies, so h_2 is convex. Finally, concerning h_3: in the normal coordinates induced by $\exp_v : I(v) \to \mathcal{M}$, the segment

$\gamma_{(\cdot)}$ is the straight line connecting ξ_0 to ξ_1, hence (recalling the definition of the Hessian, and that $\exp_v$ is a 1-Lipschitz map):

$$
\begin{aligned}
h_3''(s) = \frac{\mathrm{d}^2}{\mathrm{d}s^2}\mathcal{E}(\gamma_s) &= \operatorname{Hess}\mathcal{E}(\gamma_s)[\dot{\gamma}_s] + d\mathcal{E}(\gamma_s)\big[\nabla_{\dot{\gamma}_s}\dot{\gamma}_s\big] \\
&\geq \lambda'\|\dot{\gamma}_s\|_{\gamma_s}^2 - \|\mathcal{E}\|_{C^1}\left\|\sum_{i,j,k}\Gamma_{i,j}^k(\xi_1^i-\xi_0^i)(\xi_1^j-\xi_0^j)\frac{\partial}{\partial x_k}\right\|_{\gamma_s} \geq \big(\lambda'-K\|\mathcal{E}\|_{C^1}\big)\|\xi_1-\xi_0\|_v^2.
\end{aligned}
$$

Here K is a bound on the Christoffel symbols Γ_{ij}^k on the smooth and compact manifold $(\mathcal{M},\mathbf{g})$, and for the estimate $\|\dot{\gamma}_s\| \leq \|\xi_1-\xi_0\|_v$, we have used that (KM3) implies that $(\mathcal{M},\mathbf{g})$ is of non-negative sectional curvature.

In summary, we have shown that $s \mapsto \Phi(\tau,u,v;\gamma_s)$ is uniformly convex of modulus

$$
\left(\frac{3}{4\tau}+\lambda\right)\|\xi_1-\xi_0\|_v^2 \quad \text{with} \quad \lambda := \lambda' - K\|\mathcal{E}\|_{C^1}. \tag{3.8}
$$

Recalling that (KM3) implies non-negative sectional curvature on $(\mathcal{M},\mathbf{g})$, we conclude that $\boldsymbol{d}^2(\gamma_0,\gamma_1) \leq \|\xi_1-\xi_0\|_v^2$, so the claim (E3) follows. $\qquad\square$

3.3.3. Wasserstein space

In our last example, we consider the L^2-Wasserstein space $(\mathcal{P}_2(\Omega),\mathbf{W}_2)$ of the probability measures of finite second moment over a convex, possibly unbounded domain $\Omega \subset \mathbb{R}^d$. And we assume that $\mathcal{E}$ is uniformly displacement semi-convex; the definition is recalled below. We remark that the class of gradient flows generated in this setting encompasses nonlinear drift-diffusion-aggregation equations of the form

$$
\partial_t u = \Delta(u^m) + \nabla\cdot(u\nabla V) + \nabla\cdot(u*\nabla W\, u),
$$

under the restrictions that $m \geq (d-1)/d$, and that $V,W \in C^2(\Omega)$ are uniformly semi-convex.

$(\mathcal{P}_2(\Omega),\mathbf{W}_2)$ is a complete geodesic space, which has non-negative curvature in the sense of Alexandrov. Similarly as in the case of (non-negatively cross-curved) Riemannian manifolds discussed above, one cannot expect that hypothesis (E3) is satisfied for the geodesic $\tilde{\gamma}_{(\cdot)}$ connecting the two given measures $\gamma_0,\gamma_1 \in \mathcal{P}_2(\Omega)$. Indeed, $s \mapsto \mathbf{W}_2^2(\tilde{\gamma}_s,u)$ is typically *not* uniformly convex of modulus $\mathbf{W}_2^2(\gamma_0,\gamma_1)$, see Example 7.3.3 of [3]. Again, segments with prescribed base point are more appropriate.

We need to recall some basic notations from the theory of optimal mass transport. $\mathcal{P}_2(\Omega_j\times\Omega_k)$ is the space of probability measures with finite second moment on the cross product $\Omega\times\Omega$, and the indices j and k indicate that we use coordinates $x_j \in \Omega$ and $x_k \in \Omega$ on the components, *i.e.*, we write $x = (x_j,x_k) \in \Omega\times\Omega$. We introduce the canonical projections $\pi_j : (x_j,x_k) \mapsto x_j$, and for $s \in [0,1]$, we define $\pi_s : \Omega_0\times\Omega_1 \to \Omega$ by $\pi_s := (1-s)\pi_0 + s\pi_1$ for brevity. We write $\pi_j\#\boldsymbol{\mu}$ for the j-marginal of $\boldsymbol{\mu} \in \mathcal{P}_2(\Omega_j\times\Omega_k)$, and analogously, for $\boldsymbol{\mu} \in \mathcal{P}_2(\Omega_0\times\Omega_1)$ and $s \in [0,1]$, the interpolating measure $\pi_s\#\boldsymbol{\mu} \in \mathcal{P}_2(\Omega)$ is characterized by

$$
\int\varphi(y)\,\mathrm{d}\pi_s\#\boldsymbol{\mu}(y) = \int\varphi\big((1-s)x_0+sx_1\big)\,\mathrm{d}\boldsymbol{\mu}(x), \quad \text{for all } \varphi \in C_b^0(\Omega).
$$

A transport plan from $\mu_0 \in \mathcal{P}_2(\Omega_0)$ to $\mu_1 \in \mathcal{P}_2(\Omega_1)$ is any $\boldsymbol{\mu} \in \mathcal{P}_2(\Omega_0\times\Omega_1)$ satisfying the marginal constraints $\pi_0\#\boldsymbol{\mu} = \mu_0$ and $\pi_1\#\boldsymbol{\mu} = \mu_1$. Such a plan $\boldsymbol{\mu}$ is called optimal if it is a minimizer in the Kantorovich problem

$$
\boldsymbol{\mu} \mapsto \int|x_0-x_1|^2\,\mathrm{d}\boldsymbol{\mu}(x). \tag{3.9}
$$

For any given $\mu_0,\mu_1 \in \mathcal{P}_2(\Omega)$, there exists at least one optimal plan. The corresponding minimal value of the integral in (3.9) defines the squared Wasserstein distance $\mathbf{W}_2^2(\mu_0,\mu_1)$.

We are going to use the following two facts, which are essentially Lemma 5.3.2 of [3] and Proposition 7.3.1 of [3]:

(1) *Glueing lemma*: Given $\alpha \in \mathcal{P}_2(\Omega_0 \times \Omega_2)$ and $\beta \in \mathcal{P}_2(\Omega_1 \times \Omega_2)$ with $\pi_2\#\alpha = \pi_2\#\beta$, there exists a $\mu \in \mathcal{P}_2(\Omega_0 \times \Omega_1 \times \Omega_2)$ such that $(\pi_0, \pi_2)\#\mu = \alpha$ and $(\pi_1, \pi_2)\#\mu = \beta$.
(2) *Curve lemma*: Given $\alpha \in \mathcal{P}_2(\Omega_0 \times \Omega_1)$, $\beta \in \mathcal{P}_2(\Omega_3)$ and $t \in [0,1]$, there exists a $\mu \in \mathcal{P}_2(\Omega_0 \times \Omega_1 \times \Omega_3)$ such that $(\pi_0, \pi_1)\#\mu = \alpha$, and $(\pi_t, \pi_3)\#\mu$ is an optimal transport plan from $\pi_t\#\alpha$ to β.

Segments are defined as follows. Let $\boldsymbol{\mu}_{02} \in \mathcal{P}_2(\Omega_0 \times \Omega_2)$ and $\boldsymbol{\mu}_{12} \in \mathcal{P}_2(\Omega_1 \times \Omega_2)$ be optimal plans for the transport of γ_0 and γ_1, respectively, to $v \in \mathcal{P}_2(\Omega_2)$. By the glueing lemma, there exists a $\boldsymbol{\mu}_{012}$ such that $(\pi_0, \pi_2)\#\boldsymbol{\mu}_{012} = \boldsymbol{\mu}_{01}$ and $(\pi_1, \pi_2)\#\boldsymbol{\mu}_{012} = \boldsymbol{\mu}_{12}$. Then $[\gamma_0, \gamma_1; v]_s := \pi_s\#\boldsymbol{\mu}_{012}$. Finally, we recall that $\mathcal{E}$ being uniformly displacement semi-convex of modulus λ means that

$$\mathcal{E}\big([\gamma_0, \gamma_1; v]_s\big) \leq (1-s)\mathcal{E}(\gamma_0) + s\mathcal{E}(\gamma_1) - \frac{\lambda}{2} \int |x_0 - x_1|^2 \, d\boldsymbol{\mu}_{012}(x).$$

In the language of [3], this property is referred to as λ-convexity along generalized geodesics.

Theorem 3.4. *Assume that the metric space $(\boldsymbol{X}, \boldsymbol{d})$ is the L^2-Wasserstein space $(\mathcal{P}_2(\Omega), \mathbf{W}_2)$, with $\Omega \subset \mathbb{R}^d$. Assume further that $\mathcal{E}$ is uniformly displacement semi-convex of modulus λ. Then (E3) is satisfied, with the same λ, for $\gamma_{(\cdot)} = [\gamma_0, \gamma_1; v]_{(\cdot)}$.*

Proof. Let $u, v, \gamma_0, \gamma_1 \in \mathcal{P}_2(\Omega)$ be given, and let $\boldsymbol{\mu}_{012}$ be as above. We are going to prove the inequality (3.3) directly for a fixed value $s \in (0,1)$.

Since $(\pi_s, \pi_2)\#\boldsymbol{\mu}_{012}$ is *some* transport from γ_s to v, and $(\pi_0, \pi_2)\#\boldsymbol{\mu}_{012}$, $(\pi_1, \pi_2)\#\boldsymbol{\mu}_{012}$ are both optimal,

$$\begin{aligned}
\mathbf{W}_2^2(\gamma_s, v) &\leq \int |(1-s)x_0 + sx_1 - x_2|^2 \, d\boldsymbol{\mu}_{012}(x) \\
&= \int \big[(1-s)|x_0 - x_2|^2 + s|x_1 - x_2|^2 - s(1-s)|x_0 - x_1|^2\big] \, d\boldsymbol{\mu}_{012}(x) \\
&= (1-s)\mathbf{W}_2^2(\gamma_0, v) + s\mathbf{W}_2^2(\gamma_1, v) - s(1-s)\int |x_0 - x_1|^2 \, d\boldsymbol{\mu}_{012}(x).
\end{aligned}$$

By the curve extension lemma, there exists a $\boldsymbol{\mu}_{013} \in \mathcal{P}_2(\Omega_0 \times \Omega_1 \times \Omega_3)$, such that $(\pi_0, \pi_1)\#\boldsymbol{\mu}_{013} = (\pi_0, \pi_1)\#\boldsymbol{\mu}_{012}$, and $(\pi_s, \pi_3)\#\boldsymbol{\mu}_{013}$ is an optimal plan from γ_s to u. It follows that $(\pi_0, \pi_3)\#\boldsymbol{\mu}_{013}$ and $(\pi_1, \pi_3)\#\boldsymbol{\mu}_{013}$ are some transport plans from γ_0 and γ_1, respectively, to u, and so

$$\begin{aligned}
\mathbf{W}_2^2(\gamma_s, u) &= \int |(1-s)x_0 + sx_1 - x_3|^2 \, d\boldsymbol{\mu}_{013}(x) \\
&= \int \big[(1-s)|x_0 - x_3|^2 + s|x_1 - x_3|^2 - s(1-s)|x_0 - x_1|^2\big] \, d\boldsymbol{\mu}_{013}(x) \\
&\geq (1-s)\mathbf{W}_2^2(\gamma_0, u) + s\mathbf{W}_2^2(\gamma_1, u) - s(1-s)\int |x_0 - x_1|^2 \, d\boldsymbol{\mu}_{012}(x).
\end{aligned}$$

In combination with the definition of λ-uniform displacement convexity of $\mathcal{E}$, we arrive at

$$\Psi(\tau, u, v; \gamma_s) \leq (1-s)\Psi(\tau, u, v; \gamma_0) + s\Psi(\tau, u, v; \gamma_1) - \frac{1}{2}\left(\frac{3}{2\tau} + \lambda\right) s(1-s) \int |x_1 - x_0|^2 \, d\boldsymbol{\mu}_{012}(x).$$

Clearly, the integral above is larger or equal to $\mathbf{W}_2^2(\gamma_0, \gamma_1)$, hence (3.3) for any $\tau > 0$ so small that $\frac{3}{2\tau} + \lambda > 0$. $\square$

4. Well-posedness of the scheme and classical estimates

In this section, we study basic properties of the BDF2 scheme. First, we prove well-posedness in the sense that for all sufficiently small $\tau > 0$, and arbitrary data $u, v \in \mathcal{D}(\mathcal{E})$, the functional $\Psi(\tau, u, v; \cdot)$ possesses a unique minimizer in $\mathcal{D}(\mathcal{E})$. Consequently, for an arbitrary pair (u_τ^0, u_τ^1) of initial conditions, one obtains inductively a unique global discrete solution $(u_\tau^k)_{k \in \mathbb{N}}$ by solving the corresponding sequence of minimization problems in (3.1). Subsequently, we derive some fundamental estimates that are needed for the convergence proof later.

Recall that Assumptions (E1)–(E3) are supposed to hold, with (3.4).

Theorem 4.1 (Existence of a minimizer). *For all $\tau \in (0, \tau_*)$ and for all $u, v \in \boldsymbol{X}$, there exists a unique minimizer $w_* \in \mathcal{D}(\mathcal{E})$ of $w \mapsto \Psi(\tau, u, v; w)$.*

Proof. Let $\tau \in (0, \tau_*)$ and $u, v \in \boldsymbol{X}$ be fixed. For brevity, we write $\psi(w) := \Psi(\tau, u, v; w)$.

First, we show that ψ is bounded from below. By the triangle inequality and the binomial formula, we have that

$$\boldsymbol{d}^2(u, w) \leq 2\boldsymbol{d}^2(u, v) + 2\boldsymbol{d}^2(v, w), \quad \boldsymbol{d}^2(u_*, w) \leq \frac{2\tau_*}{\tau_* + \tau}\boldsymbol{d}^2(w, v) + \frac{2\tau_*}{\tau_* - \tau}\boldsymbol{d}^2(v, u_*).$$

Substituting these estimates into the definition of $\psi(w) = \Psi(\tau, u, v; w)$ and using Assumption (E2), we obtain for each $w \in \mathcal{D}(\mathcal{E})$:

$$
\begin{aligned}
\psi(w) &\geq \frac{1}{\tau_* + \tau}\boldsymbol{d}^2(v, w) + \frac{1}{2\tau}\boldsymbol{d}^2(v, w) - \frac{1}{4\tau}\boldsymbol{d}^2(u, w) + \mathcal{E}(w) \\
&\geq -\frac{1}{\tau_* - \tau}\boldsymbol{d}^2(v, u_*) + \frac{1}{2\tau_*}\boldsymbol{d}^2(u_*, w) - \frac{1}{2\tau}\boldsymbol{d}^2(v, u) + \mathcal{E}(w) \\
&\geq -\frac{1}{\tau_* - \tau}\boldsymbol{d}^2(v, u_*) - \frac{1}{2\tau}\boldsymbol{d}^2(v, u) + c_*.
\end{aligned}
$$

The last expression, which only depends on the given quantities u and v, constitutes the sought for lower bound on ψ. Consequently,

$$\underline{\psi} := \inf_{w \in \mathcal{D}(\mathcal{E})} \psi(w) > -\infty.$$

Now, choose a minimizing sequence $(w_n)_{n \in \mathbb{N}}$ in $\mathcal{D}(\mathcal{E})$, *i.e.*,

$$\psi(w_n) \searrow \underline{\psi}. \tag{4.1}$$

We are going to prove that this is a Cauchy sequence. Towards that goal, we invoke Assumption (E3): specifically, for given indices m and n, we choose $\gamma_0 = w_m$, $\gamma_1 = w_n$, and we define $w_{m,n} := \gamma_{\frac{1}{2}}$, the midpoint of the respective curve joining w_m to w_n. Then, by (3.3),

$$\psi(w_{m,n}) \leq \frac{1}{2}\psi(w_m) + \frac{1}{2}\psi(w_n) - \frac{1}{8}\left(\frac{3}{2\tau} + \lambda\right)\boldsymbol{d}^2(w_m, w_n).$$

Since $\tau < \tau_*$ by hypothesis, and $3 + 2\lambda\tau_* \geq 2$ thanks to (3.4), this yields an estimate on the distance from w_m to w_n:

$$\boldsymbol{d}^2(w_m, w_n) \leq \frac{8\tau}{3 + 2\tau\lambda}\left(\psi(w_m) + \psi(w_n) - 2\psi(w_{m,n})\right) \leq \frac{8\tau}{3 + 2\tau\lambda}\left(\psi(w_m) + \psi(w_n) - 2\underline{\psi}\right).$$

In view of (4.1), this verifies the Cauchy property of $(w_n)_{n \in \mathbb{N}}$. Consequently, and by completeness of $(\boldsymbol{X}, \boldsymbol{d})$, that sequence converges to a limit $w_* \in \boldsymbol{X}$.

According to Assumption (E1), $\mathcal{E}$ is lower $\boldsymbol{d}$-semi-continuous. Since the distance to a given point is clearly a continuous function, also ψ is lower $\boldsymbol{d}$-semi-continuous. By the usual argument

$$\underline{\psi} \leq \psi(w_*) \leq \liminf_{n\to\infty} \psi(w_n) = \underline{\psi},$$

we conclude that ψ attains its infimum $\underline{\psi}$ at w_*, *i.e.*, w_* is a minimizer.

Uniqueness of the minimizer follows by Assumption (E3) as well: by the remarks following (3.4), ψ is *strictly convex* along some curve that connects two potentially different minimizers. But that would mean that ψ attains a value lower than that at the minimizers, a contradiction. $\qquad\square$

In the following, we assume that discrete initial data (u_τ^0, u_τ^1) are given for each $\tau \in (0, \tau_*)$, and we consider the – according to Theorem 4.1 above – well-defined family of discrete solutions $(u_\tau^k)_{k\in\mathbb{N}}$. We recall that one of the key features of the implicit Euler method is that the energy values $\mathcal{E}(u_\tau^{k+1})$ are monotonically decreasing with k. This is not quite the case for the BDF2 scheme at hand, but we can prove a slightly weaker property.

Lemma 4.2 (Almost energy diminishing). *Each discrete solution $(u_\tau^k)_{k\in\mathbb{N}}$ satisfies*

$$\mathcal{E}(u_\tau^{k+1}) + \frac{1}{2\tau}\boldsymbol{d}^2(u_\tau^k, u_\tau^{k+1}) \leq \mathcal{E}(u_\tau^k) + \frac{1}{4\tau}\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^k) \tag{4.2}$$

at each step $k = 1, 2, \ldots$.

Proof. Since u_τ^{k+1} is a minimizer of $w \mapsto \Psi(\tau, u_\tau^{k-1}, u_\tau^k; w)$, it satisfies

$$\Psi(\tau, u_\tau^{k-1}, u_\tau^k; u_\tau^{k+1}) \leq \Psi(\tau, u_\tau^{k-1}, u_\tau^k; w)$$

for all $w \in \boldsymbol{X}$. For the choice $w = u_\tau^k$, we obtain

$$\frac{1}{\tau}\boldsymbol{d}^2(u_\tau^k, u_\tau^{k+1}) - \frac{1}{4\tau}\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^{k+1}) + \mathcal{E}(u_\tau^{k+1}) \leq -\frac{1}{4\tau}\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^k) + \mathcal{E}(u_\tau^k). \tag{4.3}$$

By the triangle inequality and the binomial formula,

$$\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^{k+1}) \leq 2\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^k) + 2\boldsymbol{d}^2(u_\tau^k, u_\tau^{k+1}). \tag{4.4}$$

Substitute this in the left-hand side of (4.3). This yields (4.2) $\qquad\square$

Next, we derive the classical estimates on energy and distance. These require

Assumption 4.3. *There are constants K_1 and K_2, such that the discrete initial data (u_τ^0, u_τ^1) satisfy for all $\tau \in (0, \tau_*)$:*

$$\mathcal{E}(u_\tau^0) \leq K_1, \quad \mathcal{E}(u_\tau^1) \leq K_1, \quad and \quad \boldsymbol{d}(u_\tau^0, u_\tau^1) \leq K_2\tau.$$

Theorem 4.4 (Classical estimates). *Fix a time horizon $T > 0$, and let Assumption 4.3 hold. Then there is a constant C, depending only on K_1, K_2 and T, such that the corresponding discrete solutions $(u_\tau^k)_{k\in\mathbb{N}}$ satisfy*

$$\sum_{k=1}^N \frac{1}{2\tau}\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^k) \leq C, \tag{4.5}$$

$$|\mathcal{E}(u_\tau^N)| \leq C, \tag{4.6}$$

$$\boldsymbol{d}^2(u_*, u_\tau^N) \leq C, \tag{4.7}$$

for all $N \in \mathbb{N}$ with $N\tau \leq T$.

Proof. The main estimate is easy to obtain: sum up the inequalities (4.2) for $k = 1$ to $k = N - 1$. After cancellation of corresponding terms on both sides, we remain with

$$\mathcal{E}(u_\tau^N) + \frac{1}{4\tau} \sum_{k=1}^{N} \boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^k) \leq \mathcal{E}(u_\tau^1) + \frac{1}{4\tau} \boldsymbol{d}^2(u_\tau^0, u_\tau^1) \leq K_1 + \frac{1}{4} K_2^2 \tau_*, \tag{4.8}$$

where Assumption 4.3 has been used in the last inequality. Clearly, if $\mathcal{E}$ would be bounded below, then (4.5)–(4.7) would follow immediately.

Since we only assume the weaker lower bound (E2), more work is required. First, we show that

$$\boldsymbol{d}^2(u_*, u_\tau^k) - \boldsymbol{d}^2(u_*, u_\tau^{k-1}) \leq 2\boldsymbol{d}(u_\tau^{k-1}, u_\tau^k) \boldsymbol{d}(u_*, u_\tau^k). \tag{4.9}$$

We only need to consider the case that $\boldsymbol{d}(u_*, u_\tau^k) \geq \boldsymbol{d}(u_*, u_\tau^{k-1})$, since otherwise the inequality is trivially true. But then an application of the triangle inequality yields:

$$
\begin{aligned}
\boldsymbol{d}^2(u_*, u_\tau^k) - \boldsymbol{d}^2(u_*, u_\tau^{k-1}) &= \big(\boldsymbol{d}(u_*, u_\tau^k) + \boldsymbol{d}(u_*, u_\tau^{k-1})\big)\big(\boldsymbol{d}(u_*, u_\tau^k) - \boldsymbol{d}(u_*, u_\tau^{k-1})\big) \\
&\leq \big(\boldsymbol{d}(u_*, u_\tau^k) + \boldsymbol{d}(u_*, u_\tau^k)\big)\big(\boldsymbol{d}(u_*, u_\tau^{k-1}) + \boldsymbol{d}(u_\tau^{k-1}, u_\tau^k) - \boldsymbol{d}(u_*, u_\tau^{k-1})\big) \\
&= 2\boldsymbol{d}(u_\tau^{k-1}, u_\tau^k) \boldsymbol{d}(u_*, u_\tau^k),
\end{aligned}
$$

which is (4.9). We use (4.9) and the binomial formula to estimate

$$
\begin{aligned}
\frac{1}{2} \boldsymbol{d}^2(u_*, u_\tau^N) - \frac{1}{2} \boldsymbol{d}^2(u_*, u_\tau^1) &= \frac{1}{2} \sum_{k=2}^{N} \big[\boldsymbol{d}^2(u_*, u_\tau^k) - \boldsymbol{d}^2(u_*, u_\tau^{k-1})\big] \\
&\leq \sum_{k=2}^{N} \boldsymbol{d}(u_\tau^{k-1}, u_\tau^k) \boldsymbol{d}(u_*, u_\tau^k) \leq \sum_{k=2}^{N} \frac{\tau_*}{8\tau} \boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^k) + \sum_{k=2}^{N} \frac{2\tau}{\tau_*} \boldsymbol{d}^2(u_*, u_\tau^k).
\end{aligned}
$$

At this point, we substitute estimate (4.9) and use Assumption (E2) to obtain

$$
\begin{aligned}
\frac{1}{2} \boldsymbol{d}^2(u_*, u_\tau^N) - \frac{1}{2} \boldsymbol{d}^2(u_*, u_\tau^1) &\leq \frac{\tau_*}{2} \left(\mathcal{E}(u_\tau^1) - \mathcal{E}(u_\tau^N) + \frac{1}{4\tau} \boldsymbol{d}^2(u_\tau^0, u_\tau^1) \right) + \frac{2\tau}{\tau_*} \sum_{k=2}^{N} \boldsymbol{d}^2(u_*, u_\tau^k) \\
&\leq \frac{\tau_*}{2} \left(\mathcal{E}(u_\tau^1) - c_* + \frac{1}{2\tau_*} \boldsymbol{d}^2(u_*, u_\tau^N) + \frac{1}{4\tau} \boldsymbol{d}^2(u_\tau^0, u_\tau^1) \right) + \frac{2\tau}{\tau_*} \sum_{k=2}^{N} \boldsymbol{d}^2(u_*, u_\tau^k).
\end{aligned}
$$

We rearrange terms and use Assumption 4.3 to arrive at the following time-discrete Gronwall inequality:

$$\boldsymbol{d}^2(u_*, u_\tau^N) \leq 2K_0^2 + 2\tau_* (K_1 - c_*) + \frac{\tau_*}{2} K_2^2 + \frac{8\tau}{\tau_*} \sum_{k=2}^{N} \boldsymbol{d}^2(u_*, u_\tau^k).$$

One verifies by induction on N that

$$\boldsymbol{d}^2(u_*, u_\tau^N) \leq \left[2K_0^2 + 2\tau_* (K_1 - c_*) + \frac{\tau_*}{2} K_2^2 \right] \left(1 + \frac{8\tau}{\tau_*} \right)^N \leq \hat{C} \exp\left(\frac{8N\tau}{\tau_*} \right) \leq \hat{C} \exp\left(\frac{8T}{\tau_*} \right),$$

proving (4.7).

From here, we conclude (4.6): the bound on $\mathcal{E}(u_\tau^N)$ from above follows immediately from (4.8), for the bound from below, we combine (4.7) with Assumption (E2). Having (4.6) at hand, the bound (4.5) follows again from (4.8). □

As a final preparation for the convergence proof, we derive a time-discrete version of the differential EVI (2.2). That estimate does not require any further assumptions on the discrete initial data.

Lemma 4.5 (Discrete EVI). *The discrete solution* $\left(u_\tau^k\right)_{k\in\mathbb{N}}$ *satisfies*

$$\left(\frac{3}{4\tau}+\frac{\lambda}{2}\right)\boldsymbol{d}^2(u_\tau^{k+1},w)-\frac{1}{\tau}\boldsymbol{d}^2(u_\tau^k,w)+\frac{1}{4\tau}\boldsymbol{d}^2(u_\tau^{k-1},w)\le\mathcal{E}(w)-\mathcal{E}(u_\tau^{k+1})-\frac{1}{\tau}\boldsymbol{d}^2(u_\tau^k,u_\tau^{k+1})+\frac{1}{4\tau}\boldsymbol{d}^2(u_\tau^{k-1},u_\tau^{k+1})$$
(4.10)

for all $k\in\mathbb{N}$, *and for all* $w\in\mathcal{D}(\mathcal{E})$.

Proof. This follows from Assumption (E3). Choose $\gamma_0=u_\tau^{k+1}$ and $\gamma_1=w$, and let $(\gamma_s)_{s\in[0,1]}$ be the corresponding connecting curve such that (3.3) holds. Combine (3.3) with the fact that u_τ^{k+1} minimizes $\Psi(\tau,u_\tau^{k-1},u_\tau^k,\cdot)$ to obtain, for each $s\in(0,1)$,

$$0\le\Psi(\tau,u_\tau^{k-1},u_\tau^k;\gamma_s)-\Psi(\tau,u_\tau^{k-1},u_\tau^k;u_\tau^{k+1})$$
$$\le s\Psi(\tau,u_\tau^{k-1},u_\tau^k;w)-s\Psi(\tau,u_\tau^{k-1},u_\tau^k;u_\tau^{k+1})-\frac{1}{2}\left(\frac{3}{2\tau}+\lambda\right)s(1-s)\boldsymbol{d}^2(u_\tau^{k+1},w).$$

Divide by $s\in(0,1)$ and pass to the limit $s\searrow0$. This yields

$$0\le\Psi(\tau,u_\tau^{k-1},u_\tau^k;w)-\Psi(\tau,u_\tau^{k-1},u_\tau^k;u_\tau^{k+1})-\frac{1}{2}\left(\frac{3}{2\tau}+\lambda\right)\boldsymbol{d}^2(u_\tau^{k+1},w),$$

which, by definition of Ψ, is the desired inequality (4.10). $\square$

5. CONVERGENCE

5.1. Statement of the main result

Once again, we recall that $(\boldsymbol{X},\boldsymbol{d})$ is a separable and complete metric space, on which a functional $\mathcal{E}:\boldsymbol{X}\to\mathbb{R}\cup\{\infty\}$ is given, that satisfies Assumptions (E1)–(E3), with (3.4). Our main result is the following.

Theorem 5.1 (Convergence result). *Consider a vanishing sequence* $(\tau_n)_{n\in\mathbb{N}}$ *of step sizes* $\tau_n\in(0,\tau_*)$ *that is strictly decreasing, and which is such that the quotients* τ_n/τ_{n+1} *are natural numbers. Let further initial data* $(u_{\tau_n}^0,u_{\tau_n}^1)$ *be given that satisfy Assumption 4.3, and that such that* $u_{\tau_n}^0\to u^0$.

Then the associated discrete solution $(u_{\tau_n}^k)_{k\in\mathbb{N}}$ *is well-defined for each* n, *and the sequence of piecewise constant interpolations* $(\bar{u}_{\tau_n})_{n\in\mathbb{N}}$ *converges locally uniformly with respect to time to a solution* $u_*\in\mathrm{AC}^2\left([0,\infty),\boldsymbol{X}\right)$ *of the gradient flow for* $\mathcal{E}$, *i.e., the limit* u_* *satisfies* (2.1). *More precisely, for every time horizon* $T>0$, *there is a constant* C *that can be expressed in terms of* K_1, K_2 *and* T *alone, such that*

$$\boldsymbol{d}\left(\bar{u}_{\tau_n}(t),u_*(t)\right)\le C\left(\boldsymbol{d}(u_{\tau_n}^0,u^0)+\sqrt{\tau_n}\right)$$
(5.1)

for all $t\in[0,T]$.

Remark 5.2. The hypothesis that consecutive τ_n's have an integer quotient has been made in order not to make the already quite technical proof even more technical. Under that hypothesis, the time grid associated to some τ_n is always a refinement of the grid for τ_m if $n>m$. That simplifies our calculations considerably.

Remark 5.3. We give a simple example showing that under the given assumptions, in general one cannot expect second order convergence of the BDF2 method, *i.e.*, τ_n^2 in place of $\sqrt{\tau_n}$ in (5.1). Our example is placed on the (very regular) metric space $\boldsymbol{X} = \mathbb{R}$ with the usual distance, with the convex but not globally differentiable potential

$$\mathcal{E}(u) = \begin{cases} u & \text{for } u \geq 0, \\ 0 & \text{for } u < 0. \end{cases}$$

The associated gradient flow with initial condition $u_0 = 1$ is the continuous curve

$$u_*(t) = \begin{cases} 1 - t & \text{for } 0 \leq t \leq 1, \\ 0 & \text{for } t > 1, \end{cases}$$

that fails to be differentiable at $t = 1$. The solution u_τ^{k+1} to the kth minimization problem in (3.1) is elementary to compute – making a case distinction whether the minimizer is positive, negative or zero – and explicitly given by

$$u_\tau^{k+1} = \begin{cases} \frac{4}{3} u_\tau^k - \frac{1}{3} u_\tau^{k-1} - \frac{2}{3}\tau & \text{if that expression is positive,} \\ \frac{4}{3} u_\tau^k - \frac{1}{3} u_\tau^k & \text{if that expression is negative,} \\ 0 & \text{otherwise, } i.e., \text{ if } 0 \leq \frac{4}{3} u_\tau^k - \frac{1}{3} u_\tau^{k-1} \leq \frac{2}{3}\tau. \end{cases} \tag{5.2}$$

One easily concludes that for the initial conditions $u_\tau^0 = 1$ and $u_\tau^1 = 1 - \tau$, the kth approximation equals $u_\tau^{k+1} = 1 - (k+1)\tau$ as long as that expression is positive. Indeed, one has

$$\frac{4}{3} u_\tau^k - \frac{1}{3} u_\tau^{k-1} - \frac{2}{3}\tau = \frac{4}{3}\big(1 - k\tau\big) - \frac{1}{3}\big(1 - (k-1)\tau\big) - \frac{2}{3}\tau = 1 - (k+1)\tau > 0,$$

so the first case in the recursion (5.2) applies. Accordingly, let N_τ be the smallest index $k \geq 1$ for which $k\tau \geq 1$. For simplicity, we assume that $u_\tau^{N_\tau} = 0$, *i.e.*, that the third case in (5.2) applies:

$$1 - N_\tau \tau = \frac{4}{3} u_\tau^{N_\tau - 1} - \frac{1}{3} u_\tau^{N_\tau - 2} - \frac{2}{3}\tau \in \left[-\frac{2}{3}\tau, 0\right]. \tag{5.3}$$

The other case, in which $-\tau < 1 - N_\tau \tau < -\frac{2}{3}\tau$, leads to a similar result, but with more complicated formulae. Recalling that the two-step recursion $a_{k+1} = \frac{4}{3} a_k - \frac{1}{3} a_{k-1}$ has the general solution $a_k = p + 3^{-k} q$ with real parameters p and q, one easily deduces from (5.2) in combination with $u_\tau^{N_\tau} = 0$ and $u_\tau^{N_\tau - 1} = 1 - (N_\tau - 1)\tau \in [\frac{1}{3}\tau, \tau]$ because of (5.3) that

$$u_\tau^k = \frac{4}{3} u_\tau^{k-1} - \frac{1}{3} u_\tau^{k-2} = -\frac{1}{2}\big(1 - 3^{-(k-N_\tau)}\big) u_\tau^{N_\tau - 1} \leq -\frac{1}{6}\big(1 - 3^{-(k-N_\tau)}\big)\tau < 0$$

for each index $k > N_\tau$. In conclusion, we have exact approximation for $t < 1$, *i.e.*,

$$u_\tau^k = u_*(k\tau) \quad \text{for every } k \text{ with } k\tau < 1,$$

but a residual of order τ at every point $t > 1$: with indices $k_\tau(t)$ chosen such that $k_\tau(t)\tau \to t > 1$ as $\tau \to 0$, it follows that

$$\lim_{\tau \to 0} \frac{u_*(t) - u_\tau^{k_\tau(t)}}{\tau} \geq \frac{1}{6} \lim_{\tau \to 0} \big(1 - 3^{-(k_\tau(t) - N_\tau)}\big) = \frac{1}{6}.$$

This clearly excludes the possibility of second order convergence.

5.2. Comparison principle

The main ingredient of the proof of Theorem 5.1 is the following comparison principle, which estimates the rate at which two discrete solutions with almost identical initial data may diverge from each other.

Theorem 5.4 (Comparison principle)**.** *In the setting of Theorem 5.1, fix two time step sizes $\tau := \tau_n$ and $\eta := \tau_m$ with $m > n$. Then, there is a constant C, expressible in terms of K_1, K_2 and T alone, such that*

$$\boldsymbol{d}^2\big(\bar{u}_\tau(t), \bar{v}_\eta(t)\big) \le C\left(\boldsymbol{d}^2(u_\tau^0, v_\eta^0) + \tau\right) \tag{5.4}$$

for all $t \in [0, T]$.

Proof. By hypothesis, $R := \tau/\eta \in \mathbb{N}$. The basic idea is to derive bounds on the distance between the discrete solutions $(u_\tau^k)_{k\in\mathbb{N}}$ and $(v_\eta^\ell)_{\ell\in\mathbb{N}}$ at comparable times, *i.e.*, for $(k-1)R \le \ell \le Rk$, by using the time-discrete EVI (4.10) for each of the two solutions and substituting the respective other solution for the "observer point" w.

More specifically, multiplication of (4.10) for u_τ^{k+1} by $(4\tau)/(3 + 2\lambda\tau)$ yields

$$\boldsymbol{d}^2(u_\tau^{k+1}, w) - \frac{4}{3 + 2\lambda\tau}\boldsymbol{d}^2(u_\tau^k, w) + \frac{1}{3 + 2\lambda\tau}\boldsymbol{d}^2(u_\tau^{k-1}, w)$$
$$\le \frac{4\tau}{3 + 2\lambda\tau}\left(\mathcal{E}(w) - \mathcal{E}(u_\tau^{k+1}) - \frac{1}{\tau}\boldsymbol{d}^2(u_\tau^k, u_\tau^{k+1}) + \frac{1}{4\tau}\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^{k+1})\right). \tag{5.5}$$

For brevity, we introduce

$$g_\tau := \frac{1}{2 + \sqrt{1 - 2\lambda\tau}} = \frac{1}{3} + O(\tau), \quad h_\tau := 2 - \sqrt{1 - 2\lambda\tau} = 1 + O(\tau), \quad \lambda_\tau := \frac{\log(h_\tau)}{\tau} = \lambda + O(\tau),$$

where the Landau symbol $O(\tau)$ is understood for the limit $\tau \to 0$, and further

$$a_\tau^k(\bar{u}_\tau; w) := h_\tau \boldsymbol{d}^2(u_\tau^{k+1}, w) - \boldsymbol{d}^2(u_\tau^k, w),$$
$$b_\tau^k(\bar{u}_\tau; w) := 4g_\tau\left(\mathcal{E}(w) - \mathcal{E}(u_\tau^{k+1}) - \frac{1}{\tau}\boldsymbol{d}^2(u_\tau^k, u_\tau^{k+1}) + \frac{1}{4\tau}\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^{k+1})\right).$$

With these notations, the variational inequality (5.5) attains the following convenient form:

$$a_\tau^k(\bar{u}_\tau; w) \le g_\tau\, a_\tau^{k-1}(\bar{u}_\tau; w) + \tau b_\tau^k(\bar{u}_\tau; w).$$

An iteration of this inequality yields

$$a_\tau^k(\bar{u}_\tau; w) \le g_\tau^k a_\tau^0(\bar{u}_\tau; w) + \tau \sum_{n=1}^k g_\tau^{k-n}\, b_\tau^n(\bar{u}_\tau; w). \tag{5.6}$$

Analogously, define $g_\eta, h_\eta, \lambda_\eta$, as well as $a_\eta^\ell(\bar{v}_\eta; w), b_\eta^\ell(\bar{v}_\eta; w)$, replacing $u_\tau^{k-1}, u_\tau^k, u_\tau^{k+1}$ and τ with $v_\eta^{\ell-1}, v_\eta^\ell, v_\eta^{\ell+1}$ and η, respectively. By the same argument as above, one obtains a corresponding estimate for $a_\eta^\ell(\bar{v}_\eta; w)$.

Now fix a time $t \in [0, T]$, and define the three quantities

$$N := \max\{n \mid n\tau \le t\}, \quad M := \max\{m \mid m\eta \le t\}, \quad L := M - \mathrm{RN}.$$

To simplify notations in the next calculations, introduce further

$$q^{k,\ell} := h_\tau^k h_\eta^\ell \boldsymbol{d}^2(u_\tau^k, v_\eta^\ell) = e^{\lambda_\tau k\tau + \lambda_\eta \ell\eta}\boldsymbol{d}^2\big(\bar{u}_\tau(k\tau), \bar{v}_\eta(\ell\eta)\big).$$

The goal is to derive an estimate on the difference

$$q^{N,M} - q^{0,0} = h_\tau^N h_\eta^M \boldsymbol{d}^2(u_\tau^N, v_\eta^M) - \boldsymbol{d}^2(u_\tau^0, v_\eta^0).$$

We expand this difference into telescopic sums:

$$q^{N,M} - q^{0,0} = \left(q^{N,M} - q^{N,\mathrm{RN}}\right) + \sum_{k=0}^{N-1} \left(q^{k+1,R(k+1)} - q^{k,Rk}\right)$$

$$= \sum_{\ell=\mathrm{RN}}^{M-1} (q^{N,\ell+1} - q^{N,\ell}) + \sum_{k=0}^{N-1} \left((q^{k+1,Rk} - q^{k,Rk}) + \sum_{\ell=Rk}^{R(k+1)-1} (q^{k+1,\ell+1} - q^{k+1,\ell}) \right).$$

By definition of a and b, the differences inside the sums satisfy

$$q^{k+1,\ell} - q^{k,\ell} = h_\tau^k h_\eta^\ell a_\tau^k(\bar{u}_\tau; v_\eta^\ell), \qquad q^{k,\ell+1} - q^{k,\ell} = h_\tau^k h_\eta^\ell a_\eta^\ell(\bar{v}_\eta; u_\tau^k).$$

Insert this above and use the estimates (5.6) to obtain

$$h_\tau^N h_\eta^M \boldsymbol{d}^2(u_\tau^N, v_\eta^M) - \boldsymbol{d}^2(u_0, v_0) \leq I_{\tau,\eta}^{N,M}(\bar{u}_\tau, \bar{v}_\eta)$$

$$:= \sum_{\ell=\mathrm{RN}}^{M-1} h_\tau^N h_\eta^\ell \left[g_\eta^\ell a_\eta^0(\bar{v}_\eta; u_\tau^N) + \eta \sum_{n=1}^{\ell} g_\eta^{\ell-n} b_\eta^n(\bar{v}_\eta; u_\tau^N) \right] \tag{5.7}$$

$$+ \sum_{k=0}^{N-1} h_\tau^k h_\eta^{Rk} \left[g_\tau^k a_\tau^0(\bar{u}_\tau; v_\eta^{Rk}) + \tau \sum_{n=1}^{k} g_\tau^{k-n} b_\tau^n(\bar{u}_\tau; v_\eta^{Rk}) \right] \tag{5.8}$$

$$+ \sum_{k=0}^{N-1} \sum_{\ell=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell \left[g_\eta^\ell a_\eta^0(\bar{v}_\eta; u_\tau^{k+1}) + \eta \sum_{n=1}^{\ell} g_\eta^{\ell-n} b_\eta^n(\bar{v}_\eta; u_\tau^{k+1}) \right]. \tag{5.9}$$

The core part of the proof of Theorem 5.4 is to show that under the given hypotheses,

$$I_{\tau,\eta}^{N,M}(\bar{u}_\tau, \bar{v}_\eta) \leq C'\tau. \tag{5.10}$$

The proof of (5.10) can be found at the end of this section. In conclusion, we have

$$e^{\lambda_\tau t + \lambda_\eta t} \boldsymbol{d}^2(\bar{u}_\tau(t), \bar{v}_\eta(t)) \leq h_\tau^N h_\eta^M \boldsymbol{d}^2(u_\tau^N, v_\eta^M) \leq C'\tau + \boldsymbol{d}^2(u_0, v_0),$$

which implies the inequality (5.4) with $C = e^{-2\lambda T}(1 + C')$. □

5.3. Proof of the main theorem

With Theorem 5.4 at hand, we finish the proof of Theorem 5.1.

Proof of Theorem 5.1. From Theorem 5.4 it follows that $(\bar{u}_{\tau_n})_{n\in\mathbb{N}}$ is a Cauchy family with respect to uniform convergence on each interval $[0, T]$. Indeed, from estimate (5.4), it follows that the values $(\bar{u}_{\tau_n}(t))_{n\in\mathbb{N}}$ converge in the complete metric space $(\boldsymbol{X}, \boldsymbol{d})$ to a limit $u_*(t)$, uniformly for $t \in [0, T]$, and that the estimate (5.1) holds. Since this argument holds for arbitrary $T > 0$, the limit $u_*(t)$ is defined for all $t \geq 0$.

To prove absolute continuity of the limit curve u_*, we argue as usual: we assign time-discrete derivatives $|u'_{\tau_n}|$ to the interpolated solutions $\bar{u}_{\tau_n}(t)$ by

$$|u'_{\tau_n}|(t) := \frac{\boldsymbol{d}(\bar{u}_{\tau_n}(t - \tau_n), \bar{u}_{\tau_n}(t))}{\tau_n} = \frac{\boldsymbol{d}(u_{\tau_n}^{k-1}, u_{\tau_n}^k)}{\tau_n} \qquad \text{for } t \in ((k-1)\tau_n, k\tau_n].$$

Thanks to the classical estimate (4.5), $|u'_{\tau_n}|$ is uniformly bounded in $L^2(0,T)$. Hence, $|u'_{\tau_n}|$ possesses a $L^2(0,T)$-weakly convergent subsequence (not relabelled) with limit $A \in L^2(0,T)$. Choose arbitrary s,t with $0 \le s \le t \le T$, and define $k_n(r) := \max\{k|k\tau_n \le r\}$, then

$$\boldsymbol{d}(\bar{u}_{\tau_n}(s), \bar{u}_{\tau_n}(t)) \le \sum_{k=k_n(s)+1}^{k_n(t)} \boldsymbol{d}(u_{\tau_n}^{k-1}, u_{\tau_n}^k) = \int_{k_n(s)\tau_n}^{k_n(t)\tau_n} |u'_{\tau_n}|(r)\mathrm{d}r.$$

In the limit $n \to \infty$, this yields

$$\boldsymbol{d}(u_*(s), u_*(t)) = \lim_{n\to\infty} \boldsymbol{d}(\bar{u}_{\tau_n}(s), \bar{u}_{\tau_n}(t)) \le \lim_{n\to\infty} \int_{k_n(s)\tau_n}^{k_n(t)\tau_n} |u'_{\tau_n}| = \int_s^t A(r)\mathrm{d}r.$$

Hence $u_* \in \mathrm{AC}^2([0,\infty), \boldsymbol{X})$.

It remains to prove that the limit curve u_* satisfies the integrated form (2.1) of the EVI. Again, let $0 \le s \le t \le T$, and define $k_n(r)$ be as above. Multiply the time-discrete EVI (4.10) for $(u_{\tau_n}^k)_{k\in\mathbb{N}}$ by τ_n, and sum from $k = k_n(s)$ to $k = k_n(t) - 1$. On the left-hand side, we obtain after elementary manipulations:

$$J_n^{(1)}(s,t) := \tau_n \sum_{k=k_n(s)}^{k_n(t)-1} \left[\left(\frac{3}{4\tau_n} + \frac{\lambda}{2}\right) \boldsymbol{d}^2(u_{\tau_n}^{k+1}, w) - \frac{1}{\tau_n}\boldsymbol{d}^2(u_{\tau_n}^k, w) + \frac{1}{4\tau_n}\boldsymbol{d}^2(u_{\tau_n}^{k-1}, w) \right]$$

$$= \frac{\lambda}{2} \int_{k_n(s)\tau_n}^{k_n(t)\tau_n} \boldsymbol{d}^2(\bar{u}_{\tau_n}(r), w)\,\mathrm{d}r$$

$$+ \frac{1}{4}\left[\left(3\boldsymbol{d}^2(u_{\tau_n}^{k_n(t)}, w) - \boldsymbol{d}^2(u_{\tau_n}^{k_n(t)-1}, w)\right) - \left(3\boldsymbol{d}^2(u_{\tau_n}^{k_n(s)}, w) - \boldsymbol{d}^2(u_{\tau_n}^{k_n(s)-1}, w)\right) \right].$$

Thanks to the r-uniform convergence of $\bar{u}_{\tau_n}(r)$ to $u_*(r)$, and since u_* is continuous, we obtain in the limit

$$\lim_{n\to\infty} J_n^{(1)}(s,t) = \frac{\lambda}{2}\int_s^t \boldsymbol{d}^2(u_*(r), w)\,\mathrm{d}r + \frac{1}{2}\boldsymbol{d}^2(u_*(t), w) - \frac{1}{2}\boldsymbol{d}^2(u_*(s), w).$$

On the other hand, after summation of the right-hand side of (4.10), we estimate once again with the help of the elementary inequality (4.4) and thus obtain

$$J_n^{(2)}(s,t) := \tau_n \sum_{k=k_n(s)}^{k_n(t)-1} \left[\mathcal{E}(w) - \mathcal{E}(u_{\tau_n}^{k+1}) - \frac{1}{\tau_n}\boldsymbol{d}^2(u_{\tau_n}^k, u_{\tau_n}^{k+1}) + \frac{1}{4\tau_n}\boldsymbol{d}^2(u_{\tau_n}^{k-1}, u_{\tau_n}^{k+1}) \right]$$

$$\le \int_{k_n(s)\tau_n}^{k_n(t)\tau_n} \left[\mathcal{E}(w) - \mathcal{E}(\bar{u}_{\tau_n}(r))\right]\mathrm{d}r - \frac{1}{2}\boldsymbol{d}^2(u_{\tau_n}^{k_n(t)-1}, u_{\tau_n}^{k_n(t)}) + \frac{1}{2}\boldsymbol{d}^2(u_{\tau_n}^{k_n(s)-1}, u_{\tau_n}^{k_n(s)}).$$

Again, thanks to local uniform convergence of $\bar{u}_{\tau_n}$ to the continuous limit u_*, and since $\mathcal{E}$ is lower semi-continuous thanks to Assumption (E1), Fatou's lemma yields that

$$\lim_{n\to\infty} J_n^{(2)}(s,t) \le \int_s^t \left[\mathcal{E}(w) - \mathcal{E}(u_*(r))\right]\mathrm{d}r.$$

Since $J_n^{(1)}(s,t) \le J_n^{(2)}(s,t)$ for all n by (4.10), the respective inequality follows for the limits, that is

$$\frac{\lambda}{2}\int_s^t \boldsymbol{d}^2(u_*(r), w)\,\mathrm{d}r + \frac{1}{2}\boldsymbol{d}^2(u_*(t), w) - \frac{1}{2}\boldsymbol{d}^2(u_*(s), w) \le \int_s^t \left[\mathcal{E}(w) - \mathcal{E}(u_*(r))\right]\mathrm{d}r.$$

This implies the integrated EVI (2.1). $\square$

5.4. Proof of the estimate (5.10)

This is the most technical part of the convergence proof, that uses only elementary inequalities and the classical estimates (4.5)–(4.7) (which, in turn, are valid thanks to Assumption 4.3). Throughout this section, we adopt the convenient notation that C is a generic constant, which is in principle expressible in terms of K_1 and K_2 from Assumption 4.3 and the terminal time T alone, and whose value may change from one line to the next.

To begin with, observe that since we assumed $\lambda \leq 0$, we have that $g_\tau \leq \frac{1}{3}$ and $g_\eta \leq \frac{1}{3}$, and therefore

$$\sum_{k=0}^{\infty} g_\tau^k \leq \frac{3}{2}, \quad \sum_{\ell=0}^{\infty} g_\eta^\ell \leq \frac{3}{2}. \tag{5.11}$$

Further, we have that $h_\tau \leq 1$ and $h_\eta \leq 1$, which means that

$$h_\tau^k h_\eta^\ell \leq 1 \tag{5.12}$$

for arbitrary $k, \ell \geq 0$. On the other hand, since $h_\tau \leq 1$ and due to (4.9), it follows that

$$a_\tau^0(\bar{u}_\tau; w) = h_\tau \boldsymbol{d}^2(u_\tau^1, w) - \boldsymbol{d}^2(u_\tau^0, w) \leq \boldsymbol{d}^2(u_\tau^1, w) - \boldsymbol{d}^2(u_\tau^0, w) \leq 2\boldsymbol{d}(u_\tau^0, u_\tau^1)\boldsymbol{d}(u_\tau^1, w).$$

Substituting $w := v_\eta^\ell$, we obtain by the triangle inequality, and thanks to estimate (4.7), that

$$a_\tau^0(\bar{u}_\tau; v_\eta^\ell) \leq 2\boldsymbol{d}(u_\tau^0, u_\tau^1)\big[\boldsymbol{d}(u_*, u_\tau^1) + \boldsymbol{d}(u_*, v_\eta^\ell)\big] \leq C\boldsymbol{d}(u_\tau^0, u_\tau^1) \leq CK_2\tau. \tag{5.13}$$

Analogously, one derives

$$a_\tau^0(\bar{v}_\eta; u_\tau^k) \leq C\eta. \tag{5.14}$$

With (5.11)–(5.14) at hand, it is now straight-forward to estimate the terms inside $I_{\tau,\eta}^{N,M}(\bar{u}_\tau, \bar{v}_\eta)$ involving a_τ^0 or a_η^0. For the expression in (5.7),

$$\sum_{\ell=\mathrm{RN}}^{M-1} h_\eta^N h_\eta^\ell g_\eta^\ell a_\eta^0(\bar{v}_\eta; u_\tau^N) \leq \sum_{\ell=\mathrm{RN}}^{M-1} g_\eta^\ell C\eta \leq \frac{3}{2}C\eta.$$

For (5.8),

$$\sum_{k=0}^{N-1} h_\tau^k h_\eta^{Rk} g_\tau^k a_\tau^0(\bar{u}_\tau; v_\eta^{Rk}) \leq \sum_{k=0}^{N-1} g_\tau^k C\tau \leq \frac{3}{2}C\tau.$$

And finally, for (5.9),

$$\sum_{k=0}^{N-1} \sum_{l=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell g_\eta^\ell a_\eta^0(\bar{v}_\eta; u_\tau^{k+1}) \leq \sum_{\ell=0}^{\mathrm{RN}} g_\eta^\ell C\eta \leq \frac{3}{2}C\eta.$$

We turn to estimate the terms involving b_τ^k and b_η^ℓ. First, by (4.4),

$$b_\tau^k(\bar{u}_\tau; w) \leq 4g_\tau \left(\mathcal{E}(w) - \mathcal{E}(u_\tau^{k+1}) + \frac{1}{2\tau}\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^k) \right) \leq 4g_\tau \left(\mathcal{E}(w) - \mathcal{E}(u_\tau^{k+1}) \right) + \frac{1}{\tau}\boldsymbol{d}^2(u_\tau^{k-1}, u_\tau^k),$$

where we have used that $g_\tau \leq 1/2$. An analogous estimate holds for b_η^ℓ. Substitute this into the expression for $I_{\tau,\eta}^{N,M}(\bar{u}_\tau, \bar{v}_\eta)$ in (5.7)–(5.9). This gives rise to two groups of terms: one related to the metric, the other related to $\mathcal{E}$.

We begin by estimating the terms related to the metric. This is done using the classical estimate (4.5): for the expression in (5.7),

$$\eta \sum_{\ell=\mathrm{RN}}^{M-1} h_\tau^N h_\eta^\ell \sum_{n=1}^\ell g_\eta^{\ell-n} \frac{\boldsymbol{d}^2(v_\eta^{n-1}, v_\eta^n)}{\eta} \le \eta \sum_{\ell=\mathrm{RN}}^{M-1} \sum_{n=1}^\ell \frac{\boldsymbol{d}^2(v_\eta^{n-1}, v_\eta^n)}{\eta} \le R\eta \sum_{\ell=1}^{M-1} \frac{\boldsymbol{d}^2(v_\eta^{\ell-1}, v_\eta^\ell)}{\eta} \le C\tau.$$

Here we have used that $M - NR \le R$ and that $R\eta = \tau$. For (5.8),

$$\tau \sum_{k=0}^{N-1} h_\tau^k h_\eta^{Rk} \sum_{n=1}^k g_\tau^{k-n} \frac{\boldsymbol{d}^2(u_\tau^{n-1}, u_\tau^n)}{\tau} \le \tau \sum_{n=1}^{N-1} \left[\left(\sum_{k=n}^{N-1} g_\tau^{k-n} \right) \frac{\boldsymbol{d}^2(u_\tau^{n-1}, u_\tau^n)}{\tau} \right] \le \frac{3}{2} C\tau.$$

Finally, for (5.9),

$$\eta \sum_{k=0}^{N-1} \sum_{\ell=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell \sum_{n=1}^\ell g_\eta^{\ell-n} \frac{\boldsymbol{d}^2(v_\eta^{n-1}, v_\eta^n)}{\eta} \le \eta \sum_{\ell=1}^{\mathrm{RN}} \sum_{n=1}^\ell g_\eta^{\ell-n} \frac{\boldsymbol{d}^2(v_\eta^{n-1}, v_\eta^n)}{\eta}$$
$$\le \eta \sum_{n=1}^{\mathrm{RN}} \left[\left(\sum_{\ell=n}^{\mathrm{RN}} g_\eta^{\ell-n} \right) \frac{\boldsymbol{d}^2(v_\eta^{n-1}, v_\eta^n)}{\eta} \right] \le \frac{3}{2} C\eta.$$

The estimates on the expressions involving the differences of the energy values are a bit more involved. To simplify calculations, we use that the b's only contain the difference between two values of $\mathcal{E}$; hence adding a constant to $\mathcal{E}$ does not change the b values. Consequently, since $\mathcal{E}(u_\tau^k)$ and $\mathcal{E}(v_\eta^\ell)$ are bounded from below thanks to (4.6), we may assume without loss of generality that all $\mathcal{E}(u_\tau^k)$ and $\mathcal{E}(v_\eta^\ell)$ are non-negative.

The contribution of the $\mathcal{E}$ terms to (5.7) is immediately controlled, recalling (5.11), (5.12), and that $M < R(N+1)$:

$$4\eta \sum_{\ell=\mathrm{RN}}^{M-1} h_\tau^N h_\eta^\ell \sum_{n=1}^\ell g_\eta^{\ell+1-n} \big[\mathcal{E}(u_\tau^N) - \mathcal{E}(v_\eta^n) \big] \le 4R\eta \mathcal{E}(u_\tau^N) 4 \sum_{\ell=1}^M g_\eta^l \le \frac{3}{2} C\tau.$$

Collecting all terms containing evaluations of $\mathcal{E}$ in (5.8) and (5.9) yields the following:

$$4\tau \sum_{k=0}^{N-1} h_\tau^k h_\eta^{Rk} \sum_{n=1}^k g_\tau^{k-n+1} \big[\mathcal{E}(v_\eta^{Rk}) - \mathcal{E}(u_\tau^{n+1}) \big] + 4\eta \sum_{k=0}^{N-1} \sum_{\ell=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell \sum_{n=1}^\ell g_\eta^{\ell-n+1} \big[\mathcal{E}(u_\tau^{k+1}) - \mathcal{E}(v_\eta^{n+1}) \big]$$
$$= 4I_1 + 4I_2,$$

where I_1 and I_2 collect terms with $\mathcal{E}(u_\tau^k)$ and $\mathcal{E}(v_\eta^\ell)$, respectively,

$$I_1 = \tau \sum_{k=0}^{N-1} \mathcal{E}(u_\tau^{k+1}) \left[\frac{1}{R} \sum_{\ell=Rk}^{R(k+1)-1} h_\tau^{k+1} h_\eta^\ell \sum_{n=1}^\ell g_\eta^{\ell-n+1} - \sum_{n=k}^{N-1} h_\tau^n h_\eta^{Rn} g_\tau^{n-k+1} \right] + \tau \mathcal{E}(u_\tau^1) \sum_{n=0}^{N-1} h_\tau^n h_\eta^{Rn} g_\tau^{n+1},$$
$$I_2 = \tau \sum_{k=0}^{N-1} \left[\mathcal{E}(v_\eta^{Rk}) h_\tau^k h_\eta^{Rk} \sum_{n=1}^k g_\tau^{k-n+1} \right] - \eta \sum_{k=0}^{N-1} \sum_{\ell=Rk}^{R(k+1)-1} \sum_{n=1}^\ell h_\tau^{k+1} h_\eta^\ell g_\eta^{\ell-n+1} \mathcal{E}(v_\eta^{n+1}).$$

For I_1, we obtain

$$I_1 \leq \tau \sum_{k=0}^{N-1} \mathcal{E}(u_\tau^{k+1}) \, h_\tau^k h_\eta^{Rk} \left[\frac{1}{R} h_\tau \sum_{\ell=0}^{R-1} \left(h_\eta^\ell \sum_{n=1}^{Rk+l} g_\eta^n \right) - g_\tau \sum_{n=0}^{N-k-1} (h_\tau h_\eta^R g_\tau)^n \right] + \tau \mathcal{E}(u_\tau^1) \sum_{n=0}^{N-1} g_\tau^n$$

$$\leq \tau \sum_{k=0}^{N-1} \mathcal{E}(u_\tau^{k+1}) \, h_\tau^k h_\eta^{Rk} \left[\frac{1}{R} \frac{1}{3} \left(\sum_{l=0}^{R-1} \frac{3}{2} \right) - g_\tau \frac{1 - (h_\tau h_\eta^R g_\tau)^{N-k}}{1 - h_\tau h_\eta^R g_\tau} \right] + C\tau$$

$$\leq \tau \sum_{k=0}^{N} \mathcal{E}(u_\tau^{k+1}) \, h_\tau^k h_\eta^{Rk} \left[\frac{1}{2} - \frac{g_\tau}{1 - h_\tau h_\eta^R g_\tau} + \frac{g_\tau^{N-k+1}}{1 - h_\tau h_\eta^R g_\tau} \right] + C\tau.$$

Recalling that $-1 \leq \lambda\tau \leq 0$, and observing that both g_τ and h_τ are convex functions of τ, a Taylor expansion yields that

$$g_\tau \geq \frac{1}{3} \left(1 + \frac{1}{3} \lambda\tau \right) \geq 0, \quad h_\tau \geq 1 + \lambda\tau \geq 0, \tag{5.15}$$

and similarly for g_η and h_η. Therefore, in combination with Bernoulli's inequality,

$$h_\tau h_\eta^R g_\tau \geq (1 + \lambda\tau)(1 + R\lambda\eta) \frac{1}{3} \left(1 + \frac{1}{2} \lambda\tau \right) \geq \frac{1}{3} (1 + \lambda\tau)^3.$$

With another application of (5.15),

$$\frac{1}{2} - \frac{g_\tau}{1 - h_\tau h_\eta^R g_\tau} \leq \frac{1}{2} - \frac{\frac{1}{3}(1 + \lambda\tau)}{1 - \frac{1}{3}(1 + \lambda\tau)^3} \leq \frac{1}{2} - \frac{(1 + \lambda\tau)^3}{3 - (1 + \lambda\tau)^3} = \frac{3}{2} \cdot \frac{1 - (1 + \lambda\tau)^3}{3 - (1 + \lambda\tau)^3} \leq \frac{3}{2} \cdot \frac{(-3)\lambda\tau}{2} = -\frac{9}{4} \lambda\tau. \tag{5.16}$$

This yields, in combination with the bound (4.6) on $\mathcal{E}$,

$$I_1 \leq \tau \sum_{k=0}^{N-1} \mathcal{E}(u_\tau^{k+1}) \, h_\tau^k h_\eta^{Rk} \left[\frac{g_\tau^{N-k+1}}{1 - h_\tau h_\eta^R g_\tau} - \frac{9}{4} \lambda\tau \right] + C\tau \leq \tau C \sum_{k=1}^{N} \left[\frac{3}{2} g_\tau^{N-k+1} - \frac{9}{4} \lambda\tau \right] + C\tau \leq C[1 + (-\lambda)T] \, \tau.$$

We turn to the estimate of I_2. In order to merge the difference of the two sums into a single sum – similar to what we did for I_1 above – we are going to apply a shift of no more than R to the indices inside $\mathcal{E}(v_\eta^{Rk})$. To control the error introduced by that shift, observe that an iteration of the energy estimate (4.2) yields

$$\mathcal{E}(v_\eta^{Rk}) \leq \mathcal{E}(v_\eta^\ell) + \frac{1}{4\eta} \boldsymbol{d}^2(v_\eta^{\ell-1}, v_\eta^\ell)$$

as soon as $0 \leq \ell \leq Rk$. Further, for such k and ℓ, we have $h_\tau^k \leq h_\tau^{\ell/R}$ since $h_\tau \leq 1$. This allows us to estimate the first sum in I_2 as follows:

$$\tau \sum_{k=1}^{N-1} \left[\mathcal{E}(v_\eta^{Rk}) h_\tau^k h_\eta^{Rk} g_\tau \sum_{n=1}^{k} g_\tau^{k-n} \right] \leq \frac{\tau}{2} \sum_{k=1}^{N-1} h_\tau^k h_\eta^{Rk} \mathcal{E}(v_\eta^{Rk})$$

$$\leq \frac{\tau}{2R} \sum_{k=1}^{N} \sum_{\ell=R(k-1)+1}^{Rk} h_\tau^{\ell/R} h_\eta^\ell \left[\mathcal{E}(v_\eta^\ell) + \frac{1}{4\eta} \boldsymbol{d}^2(v_\eta^{\ell-1}, v_\eta^\ell) \right]$$

$$\leq \frac{\eta}{2} \sum_{\ell=1}^{R(N-1)} h_\tau^{\ell/R} h_\eta^\ell \mathcal{E}(v_\eta^\ell) + \frac{\eta}{4} \sum_{\ell=1}^{R(N-1)} \frac{\boldsymbol{d}^2(v_\eta^{\ell-1}, v_\eta^\ell)}{2\eta}$$

$$\leq \frac{\eta}{2} \sum_{\ell=0}^{RN-1} h_\tau^{\ell/R} h_\eta^\ell \mathcal{E}(v_\eta^{\ell+1}) + C\eta,$$

where we have used the classical estimate (4.5) and a lower bound for h_τ in the last inequality. The second sum in I_2 is estimated as follows, using that $h_\tau^k \geq h_\tau^{\ell/R}$ for $Rk \leq \ell$:

$$\eta \sum_{k=0}^{N-1} \sum_{\ell=Rk}^{R(k+1)-1} \sum_{n=1}^{l} h_\tau^{k+1} h_\eta^\ell g_\eta^{\ell-n+1} \mathcal{E}(v_\eta^{n+1}) \geq \eta h_\tau g_\eta \sum_{\ell=0}^{RN-1} \sum_{n=1}^{\ell} h_\tau^{\ell/R} h_\eta^\ell g_\eta^{\ell-n} \mathcal{E}(v_\eta^{n+1})$$

$$= \eta g_\eta h_\tau \sum_{n=0}^{RN-1} \mathcal{E}(v_\eta^{n+1}) h_\tau^{n/R} h_\eta^n \sum_{\ell=0}^{RN-n-1} \left(h_\tau^{1/R} h_\eta g_\eta \right)^n - \eta h_\tau g_\eta \mathcal{E}(v_\eta^1) \sum_{\ell=0}^{RN-1} h_\tau^{\ell/R} h_\eta^\ell g_\eta^\ell$$

$$\geq \eta g_\eta h_\tau \sum_{n=0}^{RN-1} \mathcal{E}(v_\eta^{n+1}) h_\tau^{n/R} h_\eta^n \frac{1-(h_\tau^{1/R} h_\eta g_\eta)^{RN-n}}{1-h_\tau^{1/R} h_\eta g_\eta} - C\eta.$$

Substituting these estimates into the expression for I_2 yields a single sum,

$$I_2 \leq C\eta + \eta \sum_{\ell=0}^{RN-1} h_\tau^{\ell/R} h_\eta^\ell \mathcal{E}(v_\eta^{\ell+1}) \left[\frac{1}{2} - \frac{g_\eta h_\tau}{1-h_\tau^{1/R} h_\eta g_\eta} + \frac{g_\eta^{RN-\ell+1}}{1-h_\tau^{1/R} h_\eta g_\eta} \right].$$

Arguing similarly as in the derivation of (5.16), we estimate

$$h_\tau^{1/R} h_\eta g_\eta \geq (1+\lambda\tau)^{1/R} (1+\lambda\eta) \frac{1}{3} \left(1 + \frac{1}{3}\lambda\eta \right) \geq \frac{1}{3}(1+\lambda\tau)^3.$$

and consequently, just as in (5.16),

$$\frac{1}{2} - \frac{g_\eta h_\tau}{1-h_\tau^{1/R} h_\eta g_\eta} \leq \frac{1}{2} - \frac{\frac{1}{3}(1+\lambda\tau)^2}{1-\frac{1}{3}(1+\lambda\tau)^3} \leq -\frac{9}{4}\lambda\tau.$$

In conclusion, we obtain with the help of (4.6) that

$$I_2 \leq C\eta + \eta \sum_{\ell=0}^{RN-1} \mathcal{E}(v_\eta^{\ell+1}) \left[\frac{3}{2} g_\eta^{RN+1-\ell} + \frac{9}{4}\lambda\tau \right] C[1+(-\lambda)T] \, \tau.$$

Collecting all terms, we finally obtain the desired estimate (5.10).

6. ILLUSTRATION BY NUMERICAL EXPERIMENTS

In this section, we illustrate the convergence of our variational BDF2 method in comparison to the implicit Euler scheme in several numerical experiments. As examples, we have chosen a flow on the two-dimensional sphere $\mathbb{S}^2$, a reaction-diffusion equation as flow on the Hilbert space $L^2([0,1])$, and an aggregation-diffusion equation as flow on the space $\mathcal{P}([-1,1])$ of probability measures on $[-1,1]$, equipped with the the L^2-Wasserstein distance $\mathbf{W}_2$. We observe that the order of convergence is indeed very close to two in each of our simulations. This underlines our philosophy that one reaches the optimal order in "typical" problems, despite the fact that our main Theorem 5.1 only provides order one-half, and that there are specific counter-examples with sub-optimal converge rates, like in Remark 5.3.

In each of the examples below, we compare the numerical results for the implicit Euler scheme and for the BDF2 method at various moderately small time steps $\tau > 0$ to a reference solution that is obtained by the BDF2 method with a very small time step τ_{ref}. The approximation with the implicit Euler method of step size $\tau > 0$ – see Section 1.3 for details – is denoted by $\bar{u}_\tau^{(1)}$, and the approximation with BDF2 by $\bar{u}_\tau^{(2)}$, respectively. For the time-discrete initial data, we choose the original datum u^0 for both schemes at $t = 0$, and for the second initial

datum (at $t = \tau$) of the BDF2 method, we use the result of the first step of the implicit Euler scheme. This choice ensures in the ODE setting the enhanced convergence rate of order two, since the startup calculation with one step of the implicit Euler scheme is of order two ([15], Thm. 7.23).

The numerical rate of convergence is then computed as follows. In addition to the very small reference time step τ_{ref}, we choose a moderately large time step τ_{coarse} that is an integer multiple of τ_{ref}. Then, we calculate $\bar{u}_\tau^{(1)}$ and $\bar{u}_\tau^{(2)}$ for several intermediate time steps $\tau \in (\tau_{\mathrm{ref}}, \tau_{\mathrm{coarse}})$ that are chosen such that τ is an integer multiple of τ_{ref}, and τ_{coarse} is an integer multiple of τ. For each such choice of τ, the respective solutions $\bar{u}_\tau^{(1)}$ and $\bar{u}_\tau^{(2)}$ are compared to the reference solution $\bar{u}_{\mathrm{ref}}^{(2)}$: specifically, we calculate a mean numerical error by taking the average of the distances $\boldsymbol{d}(\bar{u}_\tau^{(i)}(t_k), u_{\mathrm{ref}}(t_k))$ at times $t_k = k\tau_{\mathrm{coarse}} \in [0, T]$ on the coarsest grid.

All simulations have been performed with `MATLAB`. Both variational schemes are implemented by solving the sequence of variational problems using the built-in function `fmincon`.

6.1. Gradient flow on the sphere $\mathbb{S}^2$

The first test problem is placed on the unit 2-sphere $\mathbb{S}^2$ equipped with the intrinsic (great-circle) distance $\boldsymbol{d}_{\mathbb{S}^2}$,

$$\boldsymbol{X} := \mathbb{S}^2 = \{u \in \mathbb{R}^3 \,|\, u_1^2 + u_2^2 + u_3^2 = 1\} \subset \mathbb{R}^3, \qquad \boldsymbol{d}_{\mathbb{S}^2}(u, v) = \arccos(u_1 v_1 + u_2 v_2 + u_3 v_3).$$

For the potential $\mathcal{E} : \mathbb{S}^2 \to \mathbb{R}$, we choose the restriction of

$$\tilde{\mathcal{E}}(u) = \sum_{i=1}^3 \left(u_i - \frac{1}{2}\right)\left(u_i + \frac{1}{2}\right)^2.$$

The corresponding gradient flow satisfies the ODE

$$\dot{u} = -\nabla_{\mathbb{S}^2}\mathcal{E}(u) = \Pi_u\left[-\nabla\tilde{\mathcal{E}}(u)\right],$$

where $\Pi_u[v] = v - u^T v$ is the projection of a vector v to the tangent space of $\mathbb{S}^2$ at u. Its flow lines are sketched in Figure 1 (left). The example falls into the class of gradient flows on Riemannian manifolds that is covered by Theorem 3.3.

A series of simulations has been performed for the initial datum

$$u^0 = \frac{1}{\sqrt{30}}(1, 2, 5)$$

and the reference step size $\tau_{\mathrm{ref}} = 10^{-5}$. The observed numerical convergence rates are 1.00 for the implicit Euler method, and 2.06 for BDF2, see Figure 1 (*right*). Further experiments with different initial data and other potentials yield very similar results. In this smooth, finite dimensional setting, second order convergence of the BDF2 method was naturally expected, as the solution curve and $\nabla_{\mathbb{S}^2}\mathcal{E}$ are smooth in the ambient space.

6.2. Reaction-diffusion equation with obstacle

Next, we consider the constrained reaction-diffusion equation

$$\partial_t u = \Delta u + 60u^3 \quad \text{subject to} \quad |u| \le 1$$

on $\Omega = [0, 1]$, subject to homogeneous Neumann boundary conditions. This PDE constitutes a gradient flow on the Hilbert space $L^2([0, 1])$ for the energy

$$\mathcal{E}(u) = \begin{cases} \frac{1}{2}\int_0^1 \left(\partial_x u(x)\right)^2 \mathrm{d}x - 15\int_0^1 u(x)^4 \mathrm{d}x, & \text{for } u \in H^1([0, 1]), \ |u| \le 1, \\ +\infty, & \text{otherwise.} \end{cases}$$
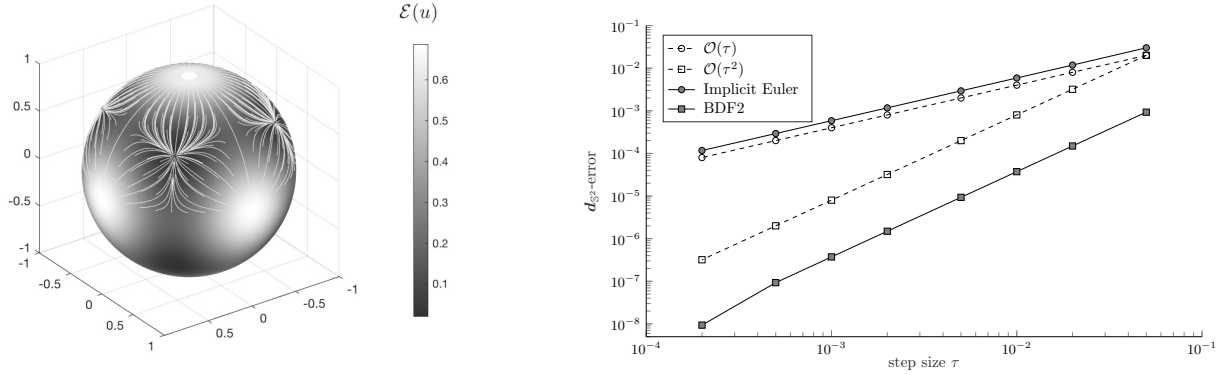
FIGURE 1. *Gradient flow on the Sphere* $\mathbb{S}^2$. *Left*: the values of $\mathcal{E}$ are color-coded by gray scale. The white lines are sample trajectories of the gradient flow generated by $\mathcal{E}$. *Right*: the $\boldsymbol{d}_{\mathbb{S}^2}$-error plot of $\bar{u}_\tau^{(i)}$ compared with $\bar{u}_{\tau_{\mathrm{ref}}}$.

The second variation of $\mathcal{E}$ amounts to

$$\mathrm{D}^2\mathcal{E}(u)[\varphi]^2 = \int_0^1 \left(\partial_x\varphi(x)\right)^2\mathrm{d}x - 180\int_0^1 u(x)^2\varphi(x)^2\mathrm{d}x \geq -180\|\varphi\|_{L^2}^2,$$

since $0 \leq u(x)^2 \leq 1$. Hence $\mathcal{E}$ is uniformly semi-convex of modulus $\lambda = -90$.

For numerical approximation, we first perform the implicit Euler or BDF2 method for discretization in time, then we apply a spatial discretization of the PDE, using central finite differences. The qualitative behavior of the approximate solution for the initial condition

$$u_0(x) = \frac{1}{2}\sin(2\pi x) + \frac{1}{4}$$

has been plotted in Figure 2 (*left*). Notice that the upper barrier is hit after a short transient time. The reference step size is $\tau_{\mathrm{ref}} = 10^{-6}$. Since we are interested in the convergence rate of the temporal discretization for the PDE, we need to estimate the influence of the additional spatial discretization on the numerical error. For that reason, the experiment is carried out with different choices of the spatial resolution, using $K = 50, 100, 250, 500, 1000$ grid points.

Our results on the numerical error are given in Figure 2 (*right*). The error curve for the implicit Euler scheme is proportional to τ, as expected. For time steps $\tau > 10^{-5}$, the error curve for the BDF2 scheme is almost perfectly proportional to τ^2, and there is no significant dependence on the spatial discretization. For very small steps $\tau \leq 10^{-5}$, there is apparently an additional contribution to the numerical error due to the spatial discretization, however as K is increased, the error curve extends its approximate proportionality to τ^2 also into that regime. This is a strong indication that for a purely temporal discretization by BDF2, the order of convergence is indeed quadratic in τ. We performed further experiments with different initial data, and with variants of the energy functional. The results remain approximately the same.

6.3. Aggregation-diffusion equations

In our last example, we study discretizations of the following aggregation-diffusion equation,

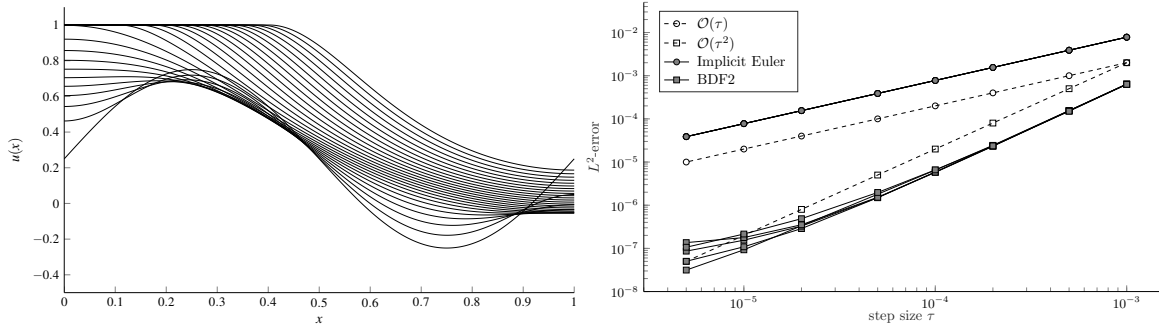$$\partial_t u = \Delta u + \partial_x(uW' * u). \tag{6.1}$$

FIGURE 2. *Reaction-diffusion equation with obstacle.* Evolution of the reference solution $\bar{u}_{\tau_{\mathrm{ref}}}$ (*left*). The L^2-error plot of $\bar{u}_\tau^{(i)}$ compared with $\bar{u}_{\tau_{\mathrm{ref}}}$ for different K (*right*).

on $\Omega = [-1, 1]$, with no-flux boundary condition, *i.e.*, $\partial_x u + W' * u = 0$ at $x = \pm 1$. For the interaction kernel, we use $W(x) = 2x^4 - x^2$. Weak solutions to (6.1) conserve mass and positivity, so we restrict attention to solutions u that are probability densities. Under this restriction, solutions to (6.1) correspond to the gradient flow on the space $\boldsymbol{X} = \mathcal{P}([-1, 1])$ of probability measures μ with respect to the L^2-Wasserstein distance $\boldsymbol{d} = \mathbf{W}_2$ for the energy functional

$$\mathcal{E}(\mu) := \begin{cases} \int_\Omega u(x) \log(u(x)) \mathrm{d}x + \frac{1}{2} \iint_{\Omega \times \Omega} u(x) W(x - y) u(y) \mathrm{d}y \mathrm{d}x, & \text{if } \mu(x) = u(x)\mathrm{d}x \text{ with density } u \in L^1(\Omega), \\ +\infty, & \text{otherwise.} \end{cases}$$

For numerical simulation, we employ the isometry of the Wasserstein space $(\mathcal{P}(\Omega), \mathbf{W}_2)$ and the space $\tilde{\boldsymbol{X}}$ of non-decreasing càdlàg functions $X : [0, 1] \to \Omega$, equipped with the $L^2([0, 1])$-norm. This isometry is realized by assigning to each μ its inverse distribution function $X_\mu : [0, 1] \to \Omega$, *i.e.*, $\mu\big([-1, X_\mu(\xi)]\big) = \xi$ for all $\xi \in [0, 1]$.

Accordingly, the Wasserstein gradient flow transforms into an $L^2([0, 1])$-gradient flow on $\tilde{\boldsymbol{X}}$ with the energy functional

$$\tilde{\mathcal{E}}(X) := -\int_{[0,1]} \log(\partial_\xi X(\xi)) \mathrm{d}\xi + \frac{1}{2} \iint_{[0,1] \times [0,1]} W(X(\xi) - X(\eta)) \mathrm{d}\xi \mathrm{d}\eta.$$

Remark 6.1. Note that the change of coordinates has transformed the original gradient flow, that had been posed on the metric Wasserstein space, into a gradient flow on (a closed subset of) the Hilbert space $L^2([0, 1])$. This example is thus rather another illustration for flows on Hilbert spaces. The combination of the BDF2 time discretization with one of the spatial discretizations for Wasserstein gradient flows in multiple space dimensions – where a transformation into a Hilbert space flow is not possible anymore – is currently under investigation.

In the numerical experiments, we prescribe an initial datum u_0 via its inverse distribution function

$$X_{u_0}(\xi) := 2\xi - 1 + \frac{1}{8\pi} \sin(8\pi\xi) \cdot \left(10(\xi(\xi - 0.5)(x - 1)) + 1\right).$$

Concerning the discretization in space, we proceed as in the previous example, using central finite differences with $K = 50, 100, 250, 500, 1000$ spatial grid points. The qualitative behavior of the reference solution (in original variables with $\tau_{\mathrm{ref}} = 10^{-6}$, and $K = 1000$) is sketched in Figure 3 (*left*).

Our results on the numerical error are given in Figure 3 (*right*). The error curves for the implicit Euler and the BDF2 schemes, respectively, are almost perfectly proportional to τ and τ^2. The results are comparable to (and even better than in) the previous example; we do not observe any significant effect of the spatial discretization, even for very small time steps. This indicates that the purely temporal discretization of the original PDE with BDF2 leads an approximation error τ^2.
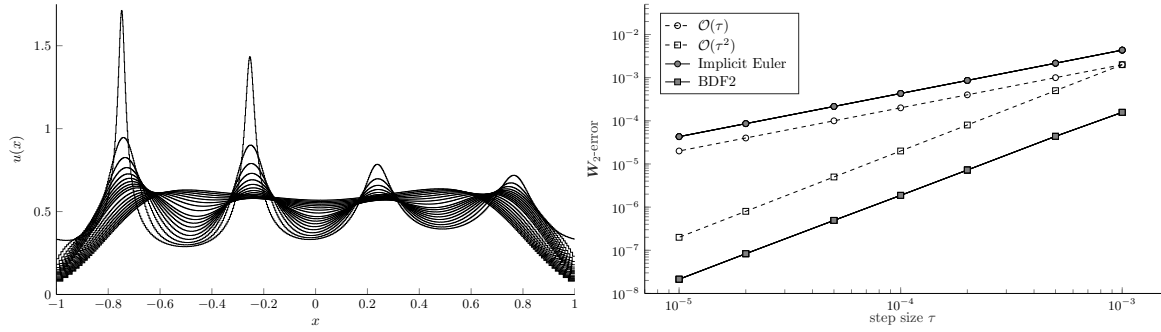
FIGURE 3. *Aggregation-diffusion equation.* Evolution of the reference solution $\bar{u}_{\tau_{\text{ref}}}$ (*left*) and the $\mathbf{W}_2$-error plot of $\bar{u}_\tau^{(i)}$ compared with $\bar{u}_{\tau_{\text{ref}}}$ for different K (*right*).

APPENDIX A. CROSS CURVATURE

In this appendix, we define the term *cross curvature* that we have used in hypothesis (KM3) in Section 3.3.2, and we briefly discuss its significance in the proof of Corollary 2.11 from [31], which we have been used for showing Theorem 3.3. Starting from the Riemannian manifold $(\mathcal{M}, \mathbf{g})$, one defines a further manifold $\mathcal{N}$ with indefinite metric tensor $\mathbf{h}$ as follows:

$$\mathcal{N} = \{(x,y) \mid x,y \in \mathcal{M}, \ x \notin \text{cut}(y), \ y \notin \text{cut}(x)\}, \quad \mathbf{h}_{(x,y)}[(\xi,\eta)]^2 = -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0}\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \boldsymbol{d}^2\big(\exp_x(s\xi), \exp_y(t\eta)\big).$$

The cross-curvature of $(\mathcal{M}, \mathbf{g})$ at the point pair $(x,y) \in \mathcal{N}$ is defined as the tensor $\mathfrak{S}_{(x,y)}^{(\mathcal{N},\mathbf{h})}$ of sectional curvatures in $(\mathcal{N}, \mathbf{h})$ at (x,y). More explicitly, for $\hat{z} = (\hat{v}, \hat{w}), \check{z} = (\check{v}, \check{w}) \in T_{(x,y)}\mathcal{N}$,

$$\mathfrak{S}_{(x,y)}^{(\mathcal{N},\mathbf{h})}[\hat{z}, \check{z}]^2 = \sum_{i,j,k,l} R_{i,j,k,\ell}\hat{z}^i \check{z}^j \hat{z}^k \check{z}^\ell,$$

with R the Riemann curvature tensor of $(\mathcal{N}, \mathbf{h})$ at (x,y). One says that $(\mathcal{M}, \mathbf{g})$ has non-negative cross curvature if, for each $(x,y) \in \mathcal{N}$ and $v \in T_x\mathcal{M}$, $w \in T_y\mathcal{M}$,

$$\mathfrak{S}_{(x,y)}^{(\mathcal{N},\mathbf{h})}[(v,0),(0,w)]^2 \geq 0. \tag{A.1}$$

The role of $\mathfrak{S}^{(\mathcal{N},\mathbf{h})}$ is probably best understood from the following alternative representation that has been derived in Lemma 4.5 of [30]: given a curve $y : (-\epsilon, \epsilon) \to \mathcal{M}$ and points $x_0, x_1 \notin \text{cut}(y(0))$, then

$$-\frac{\mathrm{d}^2}{\mathrm{d}t^2}\Big|_{t=0}\frac{\mathrm{d}^2}{\mathrm{d}s^2}\Big|_{s=0} \boldsymbol{d}^2\big([x_0, x_1; y_0]_t, y(s)\big) = \mathfrak{S}_{(x_0,y_0)}^{(\mathcal{N},\mathbf{h})}[(v,0),(0,w)]^2, \tag{A.2}$$

where $y_0 := y(0)$, $w := \dot{y}(0) \in T_{y_0}\mathcal{M}$, and $v \in T_{x_0}\mathcal{M}$ the derivative of $t \mapsto [x_0, x_1; y_0]_t$ at $t = 0$. The crucial point here is that the fourth order mixed derivative on the left-hande side only depends on v and w, without any requirement on y's second derivative, like y being a segment.

The representation (A.2) is also the key ingredient to the proof of Corollary 2.11 from [31], which we have built upon in Section 3.3.2. The cited corollary states the following: under the hypotheses (KM0)–(KM3) given in Section 3.3.2 – including non-negative cross-curvature – the function defined in (3.7) is convex, for any choice of $u, v, \gamma_0, \gamma_1 \in \mathcal{M}$. To prove this, Kim and McCann consider for each $s \in [0,1]$ the segment $z_{(\cdot)}^{(s)} := [u, v; \gamma_s]_{(\cdot)}$ connecting u to v with base γ_s, and define the extended function $F : [0,1] \times [0,1] \to \mathbb{R}$ with

$$F(s,t) = \boldsymbol{d}^2(u, \gamma_s) - \boldsymbol{d}^2(z_t^{(s)}, \gamma_s). \tag{A.3}$$

In these notations, the claim is that $\partial_s^2 F(s, 1) \geq 0$ for all $s \in [0, 1]$. By definition of F, it is clear that $\partial_s^2 F(s, 0) = 0$. A direct calculation in local coordinates shows that $\partial_t \partial_s^2 F(s, 0) = 0$ as well. Finally, thanks to non-negative cross-curvature, it follows from (A.2) that $\partial_t^2 \partial_s^2 F(s, t) \geq 0$, for each $s, t \in [0, 1]$. Notice that γ plays the role of the "general" curve y, whereas z is of the required exponential form. This concludes the proof.

Remark A.1. It turns out that $\mathfrak{S}^{(\mathcal{N}, \mathbf{h})}$ is identical to the celebrated *MTW tensor*, which is usually defined via the expression on the left-hand side of (A.2). The weak regularity condition of Ma, Trudinger and Wang [39] is (A.1), but only for those $v, w \in T_{x(0)}\mathcal{M}$ with $\mathbf{g}_{x(0)}[v, w] = 0$. It constitutes a necessary condition for the regularity of optimal transport maps on $\mathcal{M}$ with cost function $c(x, y) = \frac{1}{2}\boldsymbol{d}^2(x, y)$.

REFERENCES

[1] G. Akrivis and M. Crouzeix, Linearly implicit methods for nonlinear parabolic equations. *Math. Comput.* **73** (2004) 613–635.

[2] G. Akrivis and C. Lubich, Fully implicit, linearly implicit and implicit–explicit backward difference formulae for quasi-linear parabolic equations. *Numer. Math.* **131** (2015) 713–735.

[3] 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, 2nd edn. (2008).

[4] C. Baiocchi and M. Crouzeix, On the equivalence of *A*-stability and *G*-stability. Recent theoretical results in numerical ordinary differential equations. *Appl. Numer. Math.* **5** (1989) 19–22.

[5] J.-D. Benamou, G. Carlier, Q. Mérigot and E. Oudet, Discretization of functionals involving the Monge–Ampère operator. *Numer. Math.* **134** (2016) 611–636.

[6] 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) 691–721.

[7] A. Blanchet and P. Laurençot, The parabolic-parabolic Keller-Segel system with critical diffusion as a gradient flow in $\mathbb{R}^d$, $d \geq 3$. *Comm. Part. Differ. Equ.* **38** (2013) 658–686.

[8] J.H. Bramble, J.E. Pasciak, P.H. Sammon and V. Thomée, Incomplete iterations in multistep backward difference methods for parabolic problems with smooth and nonsmooth data. *Math. Comput.* **52** (1989) 339–367.

[9] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent and D. Slepčev, Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Math. J.* **156** (2011) 229–271.

[10] P.E. Crouch and R. Grossman, Numerical integration of ordinary differential equations on manifolds. *J. Nonlinear Sci.* **3** (1993) 1–33.

[11] G. Dahlquist, Convergence and stability in the numerical integration of ordinary differential equations. *Math. Scand.* **4** (1956) 33–53.

[12] G.G. Dahlquist, A special stability problem for linear multistep methods. *BIT Numer. Math.* **3** (1963) 27–43.

[13] S. Daneri and G. Savaré, Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.* **40** (2008) 1104–1122.

[14] E. De Giorgi, New problems on minimizing movements. *Ennio de Giorgi: Sel. Pap.* **29** (1993) 699–713.

[15] P. Deuflhard and F. Bornemann, Scientific Computing with Ordinary Differential Equations. Springer Science & Business Media, Berlin, Heidelberg (2012).

[16] E. Emmrich, Stability and error of the variable two-step BDF for semilinear parabolic problems. *J. Appl. Math. Comput.* **19** (2005) 33–55.

[17] E. Emmrich, Convergence of the variable two-step BDF time discretisation of nonlinear evolution problems governed by a monotone potential operator. *BIT Numer. Math.* **49** (2009) 297–323.

[18] E. Emmrich, Two-step BDF time discretisation of nonlinear evolution problems governed by monotone operators with strongly continuous perturbations. *Comput. Methods Appl. Math.* **9** (2009) 37–62.

[19] L.C. Ferreira and J.C. Valencia-Guevara, Gradient flows of time-dependent functionals in metric spaces and applications to pdes. *Monatsh. Math.* **185** (2018) 231–268.

[20] C.W. Gear, Numerical Initial Value Problems in Ordinary Differential Equations. Prentice Hall PTR, Upper Saddle River, NJ, USA (1971).

[21] C.W. Gear and L.R. Petzold, ODE methods for the solution of differential/algebraic systems. *SIAM J. Numer. Anal.* **21** (1984) 716–728.

[22] L. Giacomelli and F. Otto, Variatonal formulation for the lubrication approximation of the Hele-Shaw flow. *Calc. Var. Part. Differ. Equ.* **13** (2001) 377–403.

[23] U. Gianazza, G. Savaré and G. Toscani, The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation. *Arch. Ration. Mech. Anal.* **194** (2009) 133–220.

[24] A. Glitzky and A. Mielke, A gradient structure for systems coupling reaction–diffusion effects in bulk and interfaces. *Z. Angew. Math. Phys.* **64** (2013) 29–52.

[25] E. Hairer and C. Lubich, Energy-diminishing integration of gradient systems. *IMA J. Numer. Anal.* **34** (2014) 452–461.

[26] E. Hairer and G. Wanner, Solving Ordinary Differential Equations II – Stiff and Differential - Algebraic Problems. Springer Science & Business Media, Berlin, Heidelberg (2013).

[27] E. Hansen, Convergence of multistep time discretizations of nonlinear dissipative evolution equations. *SIAM J. Numer. Anal.* **44** (2006) 55–65.

[28] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** (1998) 1–17.

[29] O. Junge, D. Matthes and H. Osberger, A fully discrete variational scheme for solving nonlinear Fokker–Planck equations in multiple space dimensions. *SIAM J. Numer. Anal.* **55** (2017) 419–443.

[30] Y.-H. Kim and R.J. McCann, Continuity, curvature, and the general covariance of optimal transportation. *J. Eur. Math. Soc. (JEMS)* **12** (2010) 1009–1040.

[31] Y.-H. Kim and R.J. McCann, Towards the smoothness of optimal maps on Riemannian submersions and Riemannian products (of round spheres in particular). *J. Peine Angew. Math. (Crelles J.)* **2012** (2012) 1–27.

[32] D. Kinderlehrer, L. Monsaingeon and X. Xu, A Wasserstein gradient flow approach to Poisson-Nernst-Planck equations. *ESAIM: COCV* **23** (2017) 137–164.

[33] H. Kreth, Time-discretisations for nonlinear evolution equations. In: Numerical Treatment of Differential Equations in Applications. In Vol. 679. Springer (1978) 57–63.

[34] L. Laguzet, High order variational numerical schemes with application to Nash-MFG vaccination games. *Ricerche Mat.* **67** (2018) 247–269.

[35] P. Laurençot and B.-V. Matioc, A gradient flow approach to a thin film approximation of the Muskat problem. *Calc. Var. Part. Differ. Equ.* **47** (2013) 319–341.

[36] M.-N. Le Roux, Semidiscretization in time for parabolic problems. *Math. Comput.* **33** (1979) 919–931.

[37] G. Legendre and G. Turinici, Second-order in time schemes for gradient flows in Wasserstein and geodesic metric spaces. *C. R. Math.* **355** (2017) 345–353.

[38] S. Lisini, D. Matthes and G. Savaré. Cahn-Hilliard and thin film equations with nonlinear mobility as gradient flows in weighted-Wasserstein metrics. *J. Differ. Equ.* **253** (2012) 814–850.

[39] X.-N. Ma, N.S. Trudinger and X.-J. Wang, Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.* **177** (2005) 151–183.

[40] J. Maas, Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.* **261** (2011) 2250–2292.

[41] D. Matthes, R.J. McCann and G. Savaré, A family of nonlinear fourth order equations of gradient flow type. *Comm. Part. Differ. Equ.* **34** (2009) 1352–1397.

[42] D. Matthes and H. Osberger, Convergence of a variational lagrangian scheme for a nonlinear drift diffusion equation. *ESAIM: M2AN* **48** (2014) 697–726.

[43] R.J. McCann, A convexity principle for interacting gases. *Adv. Math.* **128** (1997) 153–179.

[44] A. Mielke, A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity* **24** (2011) 1329.

[45] A. Mielke, Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Part. Differ. Equ.* **48** (2013) 1–31.

[46] F. Otto, The geometry of dissipative evolution equations: the porous medium equation. *Comm. Part. Differ. Equ.* **26** (2001) 101–174.

[47] G. Peyré, Entropic approximation of Wasserstein gradient flows. *SIAM J. Imaging Sci.* **8** (2015) 2323–2351.

[48] S. Plazotta and J. Zinsl, High-frequency limit of non-autonomous gradient flows of functionals with time-periodic forcing. *J. Differ. Equ.* **261** (2016) 6806–6855.

[49] R. Rossi, A. Mielke and G. Savaré, A metric approach to a class of doubly nonlinear evolution equations and applications. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **7** (2008) 97–169.

[50] F. Santambrogio, Optimal Transport for Applied Mathematicians. Birkäuser, NY (2015).

[51] V. Thomée, Galerkin Finite Element Methods for Parabolic Problems. Springer, Berlin, Heidelberg, 2 edn., Vol. 25 (2006).

[52] C. Villani. Topics in optimal transportation. Vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence (2003).

[53] C. Villani, Optimal Transport: Old and New., Vol. 338. Springer Science & Business Media, Berlin, Heidelberg (2008).

[54] J. Zinsl and D. Matthes, Exponential convergence to equilibrium in a coupled gradient flow system modeling chemotaxis. *Anal. Part. Differ. Equ.* **8** (2015) 425–466.