

A NEW MINIMIZING-MOVEMENTS SCHEME FOR CURVES OF MAXIMAL SLOPE

ULISSE STEFANELLI^{1,2,3,*}

Abstract. Curves of maximal slope are a reference gradient-evolution notion in metric spaces and arise as variational formulation of a vast class of nonlinear diffusion equations. Existence theories for curves of maximal slope are often based on minimizing-movements schemes, most notably on the Euler scheme. We present here an alternative minimizing-movements approach, yielding more regular discretizations, serving as a-posteriori convergence estimator, and allowing for a simple convergence proof.

Mathematics Subject Classification. 35K55.

Received March 1, 2021. Accepted April 7, 2022.

1. INTRODUCTION

Gradient-flow evolution in metric spaces has been the subject of intense research in the last years. Starting from the pioneering remarks in [28], the theory has been boosted by the monograph by Ambrosio, Gigli, & Savaré [6] and now encompasses existence and approximation results, as well as long-time behavior, decay to equilibrium, and regularity [48].

The applicative interest in evolution equations in metric spaces has been revived by the seminal observations in [32] and the work by Otto [41] that a remarkably large class of diffusion equations can be variationally reinterpreted as gradient flows in Wasserstein spaces. More precisely, consider the nonlinear diffusion equation

$$\partial_t \rho - \operatorname{div}(\rho \nabla(V + F'(\rho) + W * \rho)) = 0 \quad \text{in } \mathbb{R}^d \times (0, T). \quad (1.1)$$

Here, $\rho = \rho(x, t) \geq 0$ is a time-dependent density with fixed total mass $\int_{\mathbb{R}^d} \rho(x, t) dx = 1$ and finite second moment $\int_{\mathbb{R}^d} |x|^2 \rho(x, t) dx < \infty$. Finally, $V : \mathbb{R}^d \rightarrow \mathbb{R}$ is a *confinement potential*, $F : [0, \infty) \rightarrow \mathbb{R}$ is an *internal-energy density*, $W : \mathbb{R}^d \rightarrow \mathbb{R}$ is an *interaction potential*, and $*$ stands for the standard convolution in \mathbb{R}^d .

Equation (1.1) can be variationally reformulated in terms of the gradient flow in the metric space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ of probability measures with finite second moment, endowed with the 2-Wasserstein distance W_2 , of the

Keywords and phrases: Curves of maximal slope, minimizing movements, generalized geodesic convexity, nonlinear diffusion, Wasserstein spaces.

¹ Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria.

² Vienna Research Platform on Accelerating Photoreaction Discovery, University of Vienna, Währingerstraße 17, 1090 Wien, Austria.

³ Istituto di Matematica Applicata e Tecnologie Informatiche E. Magenes, via Ferrata 1, 27100 Pavia, Italy.

* Corresponding author: ulisse.stefanelli@univie.ac.at

functional ϕ defined as

$$\phi(u) = \int_{\mathbb{R}^d} V(x) \, du(x) + \int_{\mathbb{R}^d} F(\rho(x)) \, dx + \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y) \, d(u \otimes u)(x, y) \quad (1.2)$$

if $u = \rho \mathcal{L}^d$ and $\phi(u) = \infty$ if u is not absolutely continuous with respect to the Lebesgue measure \mathcal{L}^d in \mathbb{R}^d , see [6] and Section 8.

The reference notion of solution to gradient flows in metric spaces is that of *curves of maximal slope* [28], see Definition 2.1 below. This is based on a specific reformulation of (1.1) in the form of a single scalar relation, featuring specific scalar quantities playing the role of the norm of time derivative of the trajectory and of the gradient of the energy, in the spirit of (1.4) below. Existence and decay to equilibrium of curves of maximal slope for ϕ in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ are available, see [6, 21, 22], for instance.

In this paper, we focus on a novel time-discretization scheme for gradient flows in metric spaces, falling within the class of *Minimizing Movements* in the sense of De Giorgi [5, 27]. Our theory is framed in abstract metric spaces, see Sections 2–5, and applied in linear and Wasserstein spaces in Sections 7 and 8, respectively. To keep this introductory discussion as simple as possible, we present here the idea in the case of the doubly nonlinear ODE system driven by a smooth potential ϕ on \mathbb{R}^d , namely

$$|u'|^{p-2} u' + \nabla \phi(u) = 0 \quad \text{in } (0, T) \quad (1.3)$$

for $p > 1$, where the prime denotes time differentiation. This equation can be equivalently rewritten as

$$\phi(u(t)) + \frac{1}{p} \int_0^t |u'(r)|^p \, dr + \frac{1}{q} \int_0^t |\nabla \phi(u(r))|^q \, dr - \phi(u(0)) = 0 \quad \forall t \in (0, T) \quad (1.4)$$

where now $q = p/(p-1)$ is conjugate to p . Note that the left-hand side above is always nonnegative, so that (1.4) corresponds indeed to a so-called *null-minimization* principle: the left-hand side is minimized and one checks that the minimum value is 0. This approach has been lately referred to as *De Giorgi's Energy-Dissipation* principle and has already been applied in a variety of different contexts, including generalized gradient flows [13, 45], rate-independent [37, 43] and GENERIC systems [29, 34], and optimal control [42].

We complement equation (1.3) by specifying the initial condition $u(0) = u^0$ for some $u^0 \in \mathbb{R}^d$. By introducing a time partition of $(0, T)$ with uniform steps $\tau = T/N > 0$, $N \in \mathbb{N}$ (note however that we consider nonuniform partitions below), and letting $u_0 = u^0$, the new minimizing-movements scheme reads

$$u_i \in \arg \min_u \left(\phi(u) + \frac{\tau^{1-p}}{p} |u - u_{i-1}|^p + \frac{\tau}{q} |\nabla \phi(u)|^q - \phi(u_{i-1}) \right) \quad (1.5)$$

for $i = 1, \dots, N$. With respect to the classical implicit Euler method, scheme (1.5) includes an extra term featuring the norm of the gradient. This modification with respect to Euler makes the function to be minimized in (1.5) a discrete and localized version of the left-hand side in (1.4). As such, scheme (1.5) is nothing by the canonical *variational integrator* scheme [31] associated with the De Giorgi's Energy-Dissipation principle.

Compared to Euler, the new minimizing-movements scheme (1.5) shows some distinguishing features. First of all, the direct occurrence of the gradient in (1.5) entails additional regularity of discrete solutions, see (3.11). As a matter of illustration, in the case of the *linear* heat equation ($p = 2$) with homogeneous Dirichlet boundary conditions scheme (1.5), corresponds to solving the problem

$$\frac{u_i - u_{i-1}}{\tau} - \Delta u_i + \tau \Delta^2 u_i = 0,$$

which is reminiscent of a *singular perturbation* of the Euler scheme, see Section 3.3.

Secondly, the exact correspondence of (1.5) to the left-hand side of (1.4) allows to check convergence of discrete solutions without the need of introducing the so-called *De Giorgi's variational interpolation* function ([6], Def. 3.2.1).

Thirdly, in using a time discretization to detect a minimum point of ϕ by iterating on the time steps, the new scheme shows enhanced performance with respect to Euler for large time steps, see [34] and (3.20) below.

Finally, the functional under minimization in (1.5) may serve as an a-posteriori estimator for the convergence of *any* discrete solution, regardless of the specific method used to obtain it. In particular, one can resort to approximate minimizers instead of true minimizers.

The minimizing-movements scheme (1.5) was already analyzed in [34] in the case of gradient flows in Hilbert spaces. In particular, convergence of the scheme for ϕ being a $C^{1,\alpha}$ perturbation of a convex function and sharp, order-one error estimates in finite dimensions can be found there. The case of curves of maximal slope in metric spaces is also mentioned in [34], where nevertheless the analysis is limited to $p = 2$ and geodesically convex potentials.

In this note, we extend the analysis of [34] to the case $p > 1$ and to potentials ϕ being (λ, p) -generalized-geodesically convex for $\lambda \in \mathbb{R}$. More precisely, the combination of our main results, Theorems 3.1-3.2, entails that solutions to the new minimizing-movements scheme (1.5) in metric spaces, see (3.2), converge to curves of maximal slope for all $p > 1$, if $\lambda \geq 0$, and for $p > 2$, if $\lambda < 0$.

In addition, in Theorem 3.3 we are able to provide a convergence result for not geodesically convex functionals, provided that some weak differentiability of its slope in form of a generalized one-sided Taylor expansion condition holds, see (3.13).

Before closing this introduction let us mention that alternative time-discrete scheme with respect to Euler are available, also in the nonlinear setting of metric spaces [25, 35, 36, 50]. Time-discrete schemes for gradient flows bear interest in connection with optimization as well [14, 16]. Since some notation and preliminary material is needed, we postpone a discussion on the literature and on the connection with optimization to Section 3.4.

This is the plan of the paper. We introduce some notation and preliminaries in Section 2 and present our main convergence results in Section 3. In particular, assumptions are collected in Section 3.1 and statements are given in Section 3.2. Some illustration of the theory on two linear equations, both in finite and infinite dimensions, is in Section 3.3. The convergence results are then proved in Sections 4–6. Eventually, we comment on the application of the abstract theory in linear spaces in Section 7 and in Wasserstein spaces in Section 8.

2. PRELIMINARIES

We briefly collect here some classical notation and preliminaries on evolution in metric spaces, for completeness. The reader familiar with the classical reference [6] may consider moving directly to Section 3.

In all of the following, (U, d) denotes a complete metric space and $\phi : U \rightarrow (-\infty, \infty]$ is a proper functional, *i.e.*, the *effective domain* $D(\phi) := \{u \in U : \phi(u) < \infty\}$ is assumed to be nonempty.

Let $p, q > 1$ be given with $1/p + 1/q = 1$. A curve $u : [0, T] \rightarrow U$ is said to belong to $AC^p([0, T]; U)$ if there exists $m \in L^p(0, T)$ with

$$d(u(s), u(t)) \leq \int_s^t m(r) dr \quad \text{for all } 0 \leq s \leq t < T. \quad (2.1)$$

If $u \in AC^p([0, T]; U)$, the limit

$$|u'| (t) := \lim_{s \rightarrow t} \frac{d(u(s), u(t))}{|t - s|}$$

exists for almost all $t \in (0, T)$, see ([6], Thm. 1.1.2), and is referred to as *metric derivative* of u at t . Moreover, the map $t \mapsto |u'| (t)$ is in $L^p(0, T)$ and is minimal within the class of functions $m \in L^p(0, T)$ fulfilling (2.1).

The *local slope* [6, 23, 28] of ϕ at $u \in D(\phi)$ is defined *via*

$$|\partial\phi|(u) := \limsup_{v \rightarrow u} \frac{(\phi(u) - \phi(v))^+}{d(u, v)}.$$

If U is a Banach space and ϕ is Fréchet differentiable, we have that $|\partial\phi|(u) = \|\mathbf{D}\phi(u)\|_*$ (dual norm).

In the following, we will make use of the notion of *geodesic convexity* for ϕ . More precisely, we call (*constant-speed*) *geodesic* any curve $\gamma : [0, 1] \rightarrow U$ such that $d(\gamma(t), \gamma(s)) = (t - s)d(\gamma(0), \gamma(1))$ for all $0 \leq s \leq t \leq 1$ and we say that ϕ is (κ, p) -*geodesically convex* for $\kappa \in \mathbb{R}$ if for all $v_0, v_1 \in D(\phi)$ there exists a geodesic with $\gamma(0) = v_0$ and $\gamma(1) = v_1$ such that

$$\phi(\gamma(\theta)) \leq \theta\phi(v_1) + (1 - \theta)\phi(v_0) - \frac{\kappa}{p}\theta(1 - \theta)d^p(v_0, v_1) \quad \forall \theta \in [0, 1] \quad (2.2)$$

The definition is classical for $p = 2$. For this p -extension see ([6], Rem. 2.4.7) or [1]. Note that geodesic convexity in particular implies that U is a *geodesic space*, for each pair v_0, v_1 is connected by a geodesic. More generally, we say that ϕ is (κ, p) -*generalized-geodesically convex* if (2.2) holds for some curve γ connecting v_0 and v_1 , not necessarily being a geodesic. In this case, U is implicitly assumed to be path-connected.

From Proposition 2.7 of [46] we have that if ϕ is (κ, p) -geodesically convex and d -lower semicontinuous, the local slope $|\partial\phi|$ is d -lower semicontinuous as well. In addition, $|\partial\phi|$ admits the representation

$$|\partial\phi|(u) = \sup_{v \neq u} \left(\frac{\phi(u) - \phi(v)}{d(u, v)} + \frac{\kappa}{p}d^{p-1}(u, v) \right)^+ \quad \forall u \in D(\phi). \quad (2.3)$$

We denote by $D(|\partial\phi|)$ the effective domain of $|\partial\phi|$, namely, $D(|\partial\phi|) = \{u \in D(\phi) : |\partial\phi|(u) < \infty\}$. Under the above-mentioned geodesic convexity assumption, the local slope $|\partial\phi|$ is a *strong upper gradient* ([6], Def. 1.3.2). Namely, for all $u \in AC^p([0, T]; U)$, the map $r \mapsto |\partial\phi|(r)$ is Borel and

$$|\phi(u(t)) - \phi(u(s))| \leq \int_s^t |\partial\phi|(u(r)) |u'(r)| dr \quad \forall 0 \leq s \leq t \leq T.$$

Note that, if $r \mapsto |\partial\phi|(u(r)) |u'(r)| \in L^1(0, T)$ the latter entails that $\phi \circ u \in W^{1,1}(0, T)$ and $(\phi \circ u)' = |\partial\phi|(u) |u'|$ almost everywhere in $(0, T)$.

Along with the above provisions, we specify the notion of gradient-driven evolution as follows.

Definition 2.1 (Curve of maximal slope). The trajectory $u \in AC^p([0, T]; U)$ is said to be a *curve of maximal slope* if $\phi \circ u \in W^{1,1}(0, T)$ and

$$\phi(u(t)) + \frac{1}{p} \int_0^t |u'|^p(r) dr + \frac{1}{q} \int_0^t |\partial\phi|^q(u(r)) dr = \phi(u(0)) \quad \forall t \in [0, T]. \quad (2.4)$$

3. MAIN RESULTS

With each time partition $0 = t_0 < t_1 < \dots < t_N = T$ we associate the time steps $\tau_i = t_i - t_{i-1}$ and the diameter $\tau = \max \tau_i$. Given the vector $\{u_i\}_{i=0}^N \in U^{N+1}$ we define its backward piecewise constant interpolant $\bar{u} : [0, T] \rightarrow U$ on the time partition to be

$$\bar{u}(0) = u_0 \quad \text{and} \quad \bar{u}(t) = u_i \quad \forall t \in (t_{i-1}, t_i], \quad i = 1, \dots, N.$$

Moreover, we define the piecewise constant function $|\widehat{u}'| : [0, T] \setminus \{t_0, \dots, t_N\} \rightarrow [0, \infty)$ as

$$|\widehat{u}'|(t) := \frac{d(u_{i-1}, u_i)}{\tau_i} \quad \forall t \in (t_{i-1}, t_i), \quad i = 1, \dots, N.$$

The notation $|\widehat{u}'|(t)$ alludes to the fact that in the Hilbert-space case the latter is nothing but the norm of the time derivative of the piecewise affine interpolant of the values $\{u_i\}_{i=0}^N$ on the time partition.

Our new minimizing-movements scheme is specified by means of the *incremental functional* $G : (0, \infty) \times D(\phi) \times D(\phi) \times D(|\partial\phi|)$ given by

$$G(\tau, v, u) := \phi(u) + \frac{\tau^{1-p}}{p} d^p(v, u) + \frac{\tau}{q} |\partial\phi|^q(u) - \phi(v). \quad (3.1)$$

In the setting of the assumptions specified later in Section 3.1, for all $(\tau, v) \in (0, \infty) \times D(\phi)$ the functional $u \in D(|\partial\phi|) \mapsto G(\tau, v, u)$ admits a minimizer, possibly being not unique. We indicate the set of such minimizers by $M_G(\tau, v)$ and the minimum value of $G(\tau, v, \cdot)$ by $\widehat{G}(\tau, v)$, namely,

$$M_G(\tau, v) := \arg \min_{u \in D(|\partial\phi|)} G(\tau, v, u), \quad \widehat{G}(\tau, v) := \min_{u \in D(|\partial\phi|)} G(\tau, v, u).$$

With this notation, the new minimizing-movements scheme reads

$$u_0 = u^0 \quad \text{and} \quad u_i \in M_G(\tau_i, u_{i-1}) \quad \text{for } i = 1, \dots, N, \quad (3.2)$$

for some given initial datum $u^0 \in D(\phi)$.

For later purposes, we introduce also the incremental functional $E : (0, \infty) \times D(\phi) \times D(\phi)$ associated with the classical backward Euler method

$$E(\tau, v, u) := \phi(u) + \frac{\tau^{1-p}}{p} d^p(v, u) - \phi(v), \quad (3.3)$$

as well as the corresponding notation

$$M_E(\tau, v) := \arg \min_{u \in D(\phi)} E(\tau, v, u), \quad \widehat{E}(\tau, v) := \min_{u \in D(\phi)} E(\tau, v, u).$$

In particular, the Euler method corresponds to the incremental problem

$$u_0 = u^0 \quad \text{and} \quad u_i \in M_E(\tau_i, u_{i-1}) \quad \text{for } i = 1, \dots, N. \quad (3.4)$$

In the context of Wasserstein spaces, see Section 8, the latter is often referred to as Jordan-Kinderlehrer-Otto scheme [32].

3.1. Assumptions

In this subsection, we fix our assumptions and collect some comments. We start by asking that

$$(U, d) \text{ is a complete metric space.} \quad (3.5)$$

In addition to the metric topology, (U, d) is assumed to be endowed with

$$\text{a Hausdorff topology } \sigma, \text{ compatible with the metric } d. \quad (3.6)$$

The latter compatibility is intended in the following sense

$$u_n \xrightarrow{\sigma} u, v_n \xrightarrow{\sigma} v \Rightarrow d(u, v) \leq \liminf_{n \rightarrow \infty} d(u_n, v_n) \quad (3.7)$$

and, in essence, means that σ is weaker than the topology induced by d . An early example for σ complying with (3.6) is the topology induced by d . In applications it may however be useful to keep the two topologies separate. In particular, if U is a Banach space, σ is often chosen to be some weak topology whereas d usually corresponds to the strong one.

The initial datum is assumed to satisfy

$$u^0 \in D(\phi). \quad (3.8)$$

We assume the proper potential $\phi : U \rightarrow (-\infty, \infty]$ to be such that

$$\text{the sublevels of } \phi \text{ are sequentially } \sigma\text{-compact}. \quad (3.9)$$

The latter in particular entails that ϕ is sequentially σ -lower semicontinuous and bounded from below. In the following, we hence assume with no loss of generality that ϕ is nonnegative. Note however that assumption (3.9) could be weakened by asking compactness on d -bounded sublevels of ϕ only.

In addition, we ask that

$$\begin{aligned} |\partial\phi| \text{ is a strong upper gradient for } \phi \text{ and it is sequentially} \\ \sigma\text{-lower semicontinuous on } d\text{-bounded sublevels of } \phi. \end{aligned} \quad (3.10)$$

The latter assumption could be weakened by developing the theory for some relaxation of $|\partial\phi|$. Still, Proposition 2.7 of [46] ensures that (3.10) hold, as soon as ϕ is (λ, p) -geodesically convex and σ is the metric topology induced by d .

In the setting of assumptions (3.5)–(3.10), the solvability of the incremental minimization problem (3.2) follows from the Direct Method. Indeed, for all $\tau > 0$ and $v \in D(\phi)$ the incremental functional $u \in D(|\partial\phi|) \mapsto G(\tau, v, u)$ is coercive and lower semicontinuous by (3.9)–(3.10). We will later check in (6.7) that indeed

$$u \in M_G(\tau, v) \Rightarrow |\partial(\phi + \tau|\partial\phi|^q/q)|(u) < \infty. \quad (3.11)$$

In particular, minimizers of $G(\tau, v, \cdot)$ enjoy additional regularity. This extra regularity may be not preserved by the time-continuous limit.

Under the sole (3.9) the incremental Euler minimization problem (3.4) is solvable as well. In particular, for all $\tau > 0$ and $v \in D(\phi)$ the functional $u \in D(\phi) \mapsto E(\tau, v, u)$ admits a minimizer.

Along the analysis, we will make reference to specific generalized geodesically convex cases. In particular, we may ask for

$$\begin{aligned} \exists \tau_* > 0, \lambda \in \mathbb{R} \text{ such that } \forall \tau \in (0, \tau_*), \forall v \in D(\phi) \\ u \mapsto E(\tau, v, u) \text{ is } (\kappa, p)\text{-generalized-geodesically convex} \\ \text{with } \kappa = (p-1)\tau^{1-p} + \lambda. \end{aligned} \quad (3.12)$$

Note that (3.12) holds if ϕ is (λ, p) -geodesically convex and the p -power of the distance is $(p-1, p)$ -geodesically convex. In case $p = 2$, the $(1, 2)$ -geodesic convexity of $u \mapsto d^2(u, v)/2$ qualifies nonpositively curved spaces in the Alexandrov sense [3, 33]. In particular, Euclidean and Hilbert spaces, as well as Riemannian manifolds of nonpositive sectional curvature ([6], Rem. 4.0.2), fall into this class.

Condition (3.12) is more demanding for $p \neq 2$. In fact, by letting $\tau \rightarrow 0$ it implies that the p -power of the distance is $(p-1, p)$ -geodesically convex. This is actually not the case in linear spaces, as one can check already in \mathbb{R} , but see also ([2], Lem. 3.1). Indeed, let $\theta = 1/2$ and $v_0 = -1, v_1 = 1, \theta = 1/2$ for $p > 2$ and $v_0 = 0, v_1 = 1$ for $p < 2$ in order to get

$$\frac{1}{p}|\theta v_1 + (1-\theta)v_0|^p > \frac{\theta}{p}|v_1|^p + \frac{1-\theta}{p}|v_0|^p - \theta(1-\theta)\frac{p-1}{p}|v_1 - v_0|^p$$

contradicting $(p-1, p)$ -geodesic convexity. See ([33], Exm. 1, p. 55) for some similar argument, proving the failure of $(1, 2)$ -geodesic convexity of $(x_1, x_1) \in \mathbb{R}^2 \mapsto (x_1^p + x_2^p)^{1/p}$. This indicates that condition (3.12) for $p \neq 2$ is actually meaningful only in spaces of qualified negative curvature. Such class of metric spaces unfortunately does not contain the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, which is actually of positively curved, see Section 8. This forces us to use condition (3.12) exclusively for $p = 2$ in Sections 7 and 8, where we deal with applications in linear and Wasserstein spaces.

In case of not geodesically convex potentials, we are still in the position of providing a convergence result under the following generalized one-sided Taylor-expansion condition on $|\partial\phi|$

$$\begin{aligned} & \exists \tau_* > 0, \forall C > 0, \exists g : (0, \tau_*) \rightarrow [0, \infty] \text{ with } \frac{1}{\tau} \int_0^\tau g(r) dr \searrow 0 \text{ as } \tau \rightarrow 0 \text{ such that} \\ & \forall \tau \in (0, \tau_*), \forall v \in D(|\partial\phi|) \text{ with } \max\{\phi(v), \tau|\partial\phi|^q(v)\} \leq C, \forall u \in M_G(\tau, v) \\ & \text{we have that } |\partial\phi|^q(u) - |\partial(\phi + \tau|\partial\phi|^q/q)|^q(u) \leq g(\tau). \end{aligned} \quad (3.13)$$

Notice that the last inequality makes sense for $u \in M_G(\tau, v)$, for we have the additional regularity (3.11). We discuss some applications fulfilling condition (3.13) in Sections 7 and 8.

A caveat on notation: In the following we use the same symbol C in order to indicate a generic positive constant, possibly depending on data and changing from line to line. Where needed, dependencies are indicated by subscripts.

3.2. Convergence results

We are now ready to state our main results.

Theorem 3.1 (Conditional convergence). *Under (3.5)–(3.10) let $\{0 = t_0^n < t_1^n < \dots < t_{N^n}^n = T\}$ be a sequence of partitions with $\tau^n := \max(t_i^n - t_{i-1}^n) \rightarrow 0$ as $n \rightarrow \infty$. Moreover, let $\{u_i^n\}_{i=0}^{N^n}$ be such that u_0^n are d -bounded, $u_0^n \xrightarrow{\sigma} u^0$, $\phi(u_0^n) \rightarrow \phi(u^0)$, and*

$$\sum_{i=1}^{N^n} (G(\tau_i^n, u_{i-1}^n, u_i^n))^+ \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (3.14)$$

Then, up to a not relabeled subsequence, we have that $\bar{u}^n(t) \xrightarrow{\sigma} u(t)$, where u is a curve of maximal slope with $u(0) = u^0$.

Note that the statement of Theorem 3.1 does not require that $u_i^n \in M_G(\tau_i^n, u_{i-1}^n)$, namely that $\{u_i^n\}_{i=0}^{N^n}$ is a solution of the new minimizing-movements scheme (3.2). In particular, Theorem 3.1 can serve as an a-posteriori tool to check the convergence of time-discrete approximations, regardless of the method used to generate them.

In particular, the above conditional convergence result directly applies to *approximate* minimizers, namely solutions of

$$u_0^n = u^0 \quad \text{and} \quad G(\tau_i, u_{i-1}^n, u_i^n) \leq \inf G(\tau_i^n, u_{i-1}^n, \cdot) + g_i^n \quad \text{for } i = 1, \dots, N^n$$

(compare with (3.2)) as long as $\sum_{i=1}^{N^n} g_i^n \rightarrow 0$ as $n \rightarrow \infty$. See [30] for a result on approximate minimizers of $E(\tau_i^n, u_{i-1}^n, \cdot)$ instead.

The conditional convergence result of Theorem 3.1 thus relies on the possibility of solving the inequality $G(\tau_i^n, u_{i-1}^n, u_i^n) \leq 0$ up to a small, controllable error, and establishing some a priori bounds on the discrete solution. The validity of condition (3.14) is to be checked on the specific problem at hand. In the specific case of (λ, p) -generalized-geodesically convex functionals ϕ on a properly nonpositively curved space, condition (3.14) actually holds for solutions of the new minimizing-movements scheme (3.2). This is the content of our second main result.

Theorem 3.2 (Convergence in the geodesically convex case). *Under assumptions (3.5)–(3.10) and (3.12), let $\{0 = t_0^n < t_1^n < \dots < t_{N^n}^n = T\}$ be a sequence of partitions with $\tau^n := \max(t_i^n - t_{i-1}^n) < \tau_*$ and $\tau^n \rightarrow 0$ as $n \rightarrow \infty$. Moreover, assume that either $\lambda \geq 0$ or $p > 2$ in (3.12). Then, solutions $\{u_i^n\}_{i=0}^{N^n}$ of (3.2) fulfill condition (3.14). Hence, \bar{u}^n converges pointwise to a curve of maximal slope up to subsequences.*

We now turn to a convergence result in the not geodesically convex case. Here, some stronger topological assumption, an approximation of the initial datum, and the generalized one-sided Taylor-expansion assumption (3.13) for $|\partial\phi|$ are necessary.

Theorem 3.3 (Convergence without geodesic convexity). *Under assumptions (3.5)–(3.10), let σ be the metric topology induced by d , U be separable, and ϕ fulfill (3.13). Moreover, let $\{0 = t_0^n < t_1^n < \dots < t_{N^n}^n = T\}$ be a sequence of partitions with $\tau^n := \max(t_i^n - t_{i-1}^n) < \tau_*$, $(\tau_i^n - \tau_{i-1}^n)^+ / \tau_{i-1}^n \leq C\tau^n$ for $i = 2, \dots, N^n$, and $\tau^n \rightarrow 0$ as $n \rightarrow \infty$. Choose a sequence $\{u^{0n}\}$ with $u^{0n} \in M_E(\tau^n, u^0)$. Then, solutions $\{u_i^n\}_{i=0}^{N^n}$ of (3.2) with $u_0^n = u^{0n}$ fulfill condition (3.14). Hence, \bar{u}^n converges pointwise to a curve of maximal slope up to subsequences.*

Note that the one-sided nondegeneracy condition $(\tau_i^n - \tau_{i-1}^n)^+ / \tau_{i-1}^n \leq C\tau^n$ in the statement is fulfilled if $i \mapsto \tau_i^n$ is nonincreasing. In particular, it holds for uniform partitions. In case $u^0 \in D(|\partial\phi|)$ no approximation of the initial datum as in Theorem 3.3 is actually needed.

Theorems 3.1, 3.2, and 3.3 are proved in Sections 4, 5, and 6, respectively.

3.3. An illustration on linear equations

The focus of our theory is on nonlinear problems. Still, as a way of illustrating the results, we present here two linear ODE and PDE examples. Nonlinear applications are then discussed in Sections 7 and 8.

Let us start from the finite-dimensional example of the gradient flow in $(\mathbb{R}^d, |\cdot|)$ of $\phi(u) = \lambda|u|^2/2$ with $\lambda \in \mathbb{R}$ and take $p = 2$. In this case, the incremental functional G reads

$$G(\tau, v, u) = \frac{\lambda}{2}|u|^2 + \frac{1}{2\tau}|u - v|^2 + \frac{\tau\lambda^2}{2}|u|^2 - \frac{\lambda}{2}|v|^2.$$

For all $v \in \mathbb{R}^d$ given, the latter can be readily minimized, giving the only minimum point $u = v/(1 + \lambda\tau + \lambda^2\tau^2)$. Correspondingly, the minimal value $\widehat{G}(t, v)$ can be checked to be

$$\widehat{G}(t, v) = -\frac{|v|^2\lambda^3\tau^2}{2(1 + \lambda\tau + \lambda^2\tau^2)}. \quad (3.15)$$

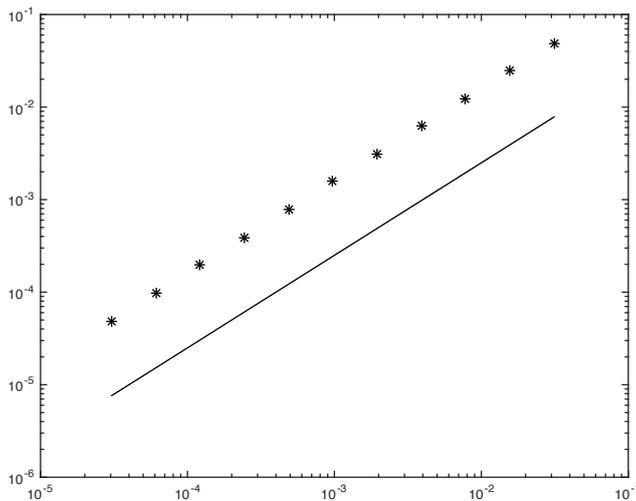


FIGURE 1. Values r^n from (3.16) against τ^n (stars) with respect to order 1 (solid) in log-log scale.

If $\lambda \geq 0$ the minimal value is nonpositive and condition (3.14) trivially holds. If $\lambda < 0$, the minimal value scales as τ^2 and condition (3.14) still holds. Indeed, by letting

$$r^n := \sum_{i=1}^{N^n} (G(\tau_i^n, u_{i-1}^n, u_i^n))^+ \quad (3.16)$$

we have that

$$r^n = \sum_{i=1}^{N^n} \frac{|u_{i-1}^n|^2 (\lambda^-)^3 (\tau_i^n)^2}{2(1 + \lambda \tau_i^n + \lambda^2 (\tau_i^n)^2)} \leq C \max_i |u_i^n|^2 \tau^n \quad (3.17)$$

where we tacitly assumed that $\lambda^- \tau^n \leq \lambda^- \tau_* < 1$ and we used the standard notation for the negative part $\lambda^- = \max\{0, -\lambda\}$. Condition (3.14) hence follows as soon as $\max_i |u_i^n|$ stays bounded with respect to n , which happens to be the case as the evolution takes place in the finite time interval $[0, T]$.

In fact, the order of convergence in (3.17) is sharp, as illustrated in Figure 1 for the choice $d = 1$, $\lambda = -1$, $u^0 = 1$, $T = 1$. Here, r_n is computed for the uniform partition $\tau_i^n = \tau^n = 2^{-n}$, $n = 1, \dots, 12$ or, equivalently, for $N^n = 2^n$.

On a uniform partition of time step $\tau > 0$, the solution of the new minimizing movement scheme $\{u_i\}$ and the solution $\{u_i^e\}$ of the Euler scheme read

$$u_i = \frac{u_0}{(1 + \lambda\tau + \lambda^2\tau^2)^i} \quad \text{and} \quad u_i^e = \frac{u_0}{(1 + \lambda\tau)^i}, \quad (3.18)$$

respectively. It is hence a standard matter to compute

$$|u_i - u_i^e| = |u_0| \left| \frac{(1 + \lambda\tau)^i - (1 + \lambda\tau + \lambda^2\tau^2)^i}{(1 + \lambda\tau + \lambda^2\tau^2)^i (1 + \lambda\tau)^i} \right| \quad (3.19)$$

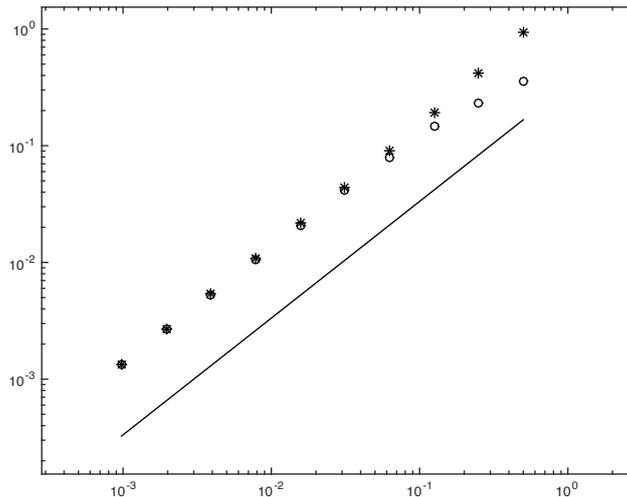


FIGURE 2. L^∞ error with respect to τ^n for the new minimizing-movements scheme (stars) and the Euler scheme (dots) in log–log scale. The solid line represents order 1.

which scales like τ^2 as $\tau \rightarrow 0$. As the Euler scheme is of first order, the same holds true for the new minimizing-movements scheme, see Figure 2 for $d = 1$, $\lambda = -1$, $u_0 = 1$. Indeed, Figure 2 shows that this order is sharp. Note in fact that the new minimizing-movements scheme is proved in Proposition 4.3 of [34] to be of first order for all nonnegative potentials ϕ in C^2 in finite dimensions.

Assume now to be interested in computing the minimum of ϕ by following the discrete scheme for a fixed number m of iterations, a classical strategy in optimization [15, 44]. In the specific case of our ODE example we compute from (3.18)

$$\phi(u_m) = \frac{\lambda}{2(1 + \lambda\tau + \lambda^2\tau^2)^{2m}} \quad \text{and} \quad \phi(u_m^e) = \frac{\lambda}{2(1 + \lambda\tau)^{2m}}. \quad (3.20)$$

Due to the presence of the extra term $\lambda^2\tau^2$ in the denominator, the new scheme is advantageous with respect to Euler as for reduction of the potential after a fixed number of iterations. Note that this effect is enhanced by choosing *large* time steps.

Let us move to an infinite-dimensional example by considering the standard heat equation on the space time cylinder $\Omega \times (0, T)$ where $\Omega \subset \mathbb{R}^d$ is a smooth, open, and bounded set and homogeneous Dirichlet conditions are imposed (other choices being of course possible). We classically reformulate this as the gradient flow in $(L^2(\Omega), \|\cdot\|)$, of the *Dirichlet* energy

$$\phi(u) = \begin{cases} \frac{1}{2} \int_{\Omega} |\nabla u(x)|^2 dx & \text{for } u \in H_0^1(\Omega) \\ \infty & \text{elsewhere in } L^2(\Omega). \end{cases}$$

where $\|\cdot\|$ is the norm corresponding to the natural L^2 scalar product (\cdot, \cdot) . In this case, we have that $\partial\phi(u) = -\Delta u$ with $D(\partial\phi) = H^2(\Omega) \cap H_0^1(\Omega)$. The symbol ∂ indicates the *subdifferential* in the sense of convex analysis [19]. In particular, $\partial\phi$ is single-valued and $|\partial\phi|(u) = \|\Delta u\|$ for all $u \in D(\partial\phi)$. The incremental functional

$G : (0, \infty) \times H_0^1(\Omega) \times H^2(\Omega) \cap H_0^1(\Omega)$ hence reads

$$G(\tau, v, u) = \int_{\Omega} \left(\frac{1}{2} |\nabla u|^2 + \frac{1}{2\tau} |u - v|^2 + \frac{\tau}{2} |\Delta u|^2 - \frac{1}{2} |\nabla v|^2 \right) dx.$$

For all $v \in H_0^1(\Omega)$ given, the latter can be readily minimized in $H^2(\Omega) \cap H_0^1(\Omega)$. Relying on linearity one can easily identify the subgradient of $u \mapsto G(\tau, v, u)$ as

$$(\partial G(\tau, v, \cdot))(u) = -\Delta u + \frac{u - v}{\tau} + \tau \Delta^2 u$$

and $D(\partial G(\tau, v, \cdot)) = \{u \in H^4(\Omega) \cap H_0^1(\Omega) : \Delta u = 0 \text{ on } \partial\Omega\}$. Hence, the minimizer u of $G(\tau, v, \cdot)$ solves

$$u - \tau \Delta u + \tau^2 \Delta^2 u = v \text{ a.e. in } \Omega, \quad u = \Delta u = 0 \text{ on } \partial\Omega. \quad (3.21)$$

The latter is reminiscent of a singular perturbation of

$$u^e - \tau \Delta u^e = v \text{ a.e. in } \Omega, \quad u^e = 0 \text{ on } \partial\Omega, \quad (3.22)$$

corresponding instead to the incremental step of the Euler scheme.

Let now $\{w^k\}$ be a complete orthonormal basis of L^2 of eigenfunctions of $-\Delta$ with homogeneous Dirichlet boundary conditions, namely, $w^k \in H^2(\Omega) \cap H_0^1(\Omega)$ with $w^k \neq 0$ and $-\Delta w^k = \lambda^k w^k$ for some $\lambda^k > 0$. By inserting in (3.21)–(3.22) $u = \sum_k u^k w^k$, $u^e = \sum_k (u^e)^k w^k$, and $v = \sum_k v^k w^k$ for $u^k := (u, w^k)$, $(u^e)^k := (u^e, w^k)$, and $v^k := (v, w^k)$, respectively, we get that

$$u^k = \frac{v^k}{1 + \tau \lambda^k + (\tau \lambda^k)^2} \quad \text{and} \quad (u^e)^k = \frac{v^k}{1 + \tau \lambda^k}.$$

In particular, by arguing as in (3.15) one readily checks that

$$\widehat{G}(\tau, v) = - \sum_k \frac{|v^k|^2 (\lambda^k)^3 \tau^2}{2(1 + \tau \lambda^k + (\tau \lambda^k)^2)} \leq 0$$

and condition (3.14) holds. By iterating on the time steps, the solution $\{u_i\}$ of the new minimizing movement scheme and that $\{u_i^e\}$ of the Euler scheme read $u_i = \sum_k u_i^k w^k$ and $u_i^e = \sum_k (u_i^e)^k w^k$ where

$$u_i^k = \frac{(u^0)^k}{(1 + \tau \lambda^k + (\tau \lambda^k)^2)^i} \quad \text{and} \quad (u_i^e)^k = \frac{(u^0)^k}{(1 + \tau \lambda^k)^i}$$

and $(u^0)^k := (u^0, w^k)$. Proceeding as in (3.20) one computes

$$\begin{aligned} \phi(u_m) &= \frac{1}{2} \sum_k \lambda^k (u_m^k)^2 = \frac{1}{2} \sum_k \frac{\lambda^k ((u^0)^k)^2}{(1 + \tau \lambda^k + (\tau \lambda^k)^2)^{2m}}, \\ \phi(u_m) &= \frac{1}{2} \sum_k \lambda^k ((u_m^e)^k)^2 = \frac{1}{2} \sum_k \frac{\lambda^k ((u^0)^k)^2}{(1 + \tau \lambda^k)^{2m}} \end{aligned}$$

and the same observations as in the ODE case on the effectiveness of the reduction of the potential for a fixed number of iterations apply.

3.4. Literature

Before moving on, let us record here some other alternatives to the Euler scheme, specifically focusing on the case $p = 2$.

Legendre and Turinici advance in [35] the *midpoint* scheme

$$u_i \in \arg \min_u \left(\inf \left(2\phi(w) + \frac{1}{2\tau} d(u, u_{i-1}) : w \in \Gamma(u, u_{i-1}) \right) \right)$$

where

$$\Gamma(u, u_{i-1}) = \{\gamma(1/2) : \gamma : [0, 1] \rightarrow U \text{ geodesic with } \gamma(0) = u_{i-1} \text{ and } \gamma(1) = u\}.$$

By assuming (3.9)–(3.10), as well as some additional closure property relating to the specific structure of the set Γ , they prove that this midpoint scheme is solvable and convergent.

A variant of this scheme is also proposed in [35] in the specific case of nonbranching geodesic spaces, namely, spaces where any two points are connected by a unique geodesic. In these spaces, for all w and u_{i-1} there exists a unique u such that $w \in \Gamma(u, u_{i-1})$. An *extrapolated* version of the Euler scheme is hence defined by the relations

$$u_{1/2}^e \in \Gamma(u_i, u_{i-1}) \quad \text{where} \quad u_{1/2}^e \in M_E(\tau/2, u_{i-1}).$$

Albeit not purely variational, this scheme is based on the solution of the Euler scheme with halved time step.

Matthes and Plazotta [36] address a variational version of the Backward Differentiation Formula (BDF2) method, namely,

$$u_i \in \arg \min_{u \in D(\phi)} \left(\frac{1}{\tau} d^2(u, u_{i-1}) - \frac{1}{4\tau} d^2(u, u_{i-2}) + \phi(u) \right) \quad \text{for } i = 2, \dots, N$$

where now both u_0 and u_1 are given. Under some lower semicontinuity and convexity conditions, it is proved in [36] that the scheme admits a solution, whose piecewise-in-time interpolant converges to a curve of maximal slope with rate $\tau^{1/2}$. It also shown that under natural regularity assumptions on the limiting time-continuous curve of maximal slope, the convergence rate can be τ at best.

Perturbations of the Euler method of the form

$$u_i \in \arg \min_{u \in D(\phi)} \left(\frac{a_i^\tau}{2\tau} d^2(u, u_{i-1}) + \phi(u) \right) \quad \text{for } i = 2, \dots, N,$$

are considered by Tribuzio in [50]. Here, one is given the sequence of positive weights defined as $a_i^\tau = a^\tau(i\tau)$ for some functions $a^\tau : (0, \infty) \rightarrow (0, \infty)$. This generalization with respect to the classical Euler scheme yields a modification of the metric as time evolves. By asking $1/a^\tau$ to be locally equiintegrable with respect to τ , one can prove that minimizers converge to curves of maximal slope according to a specific time-dependent limiting metric. Under some more general assumptions on a^τ , discontinuous evolutions can also be obtained. These can be proved to be capable of exploring the different wells of a multiwell potential ϕ .

Let us also mention the approach à la Crandall-Liggett by Clément and Desch [24, 25], see also [26], who recursively define $u_0^n = u^0$ and $u_i^n = J(u_{i-1}^n)$ for $i = 1, \dots, N = T/\tau$, where $J(u_{i-1}^n)$ is the set of points $u \in D(\phi)$ fulfilling the inequality

$$\frac{1}{2\tau} d^2(u, w) - \frac{1}{2\tau} d^2(u_{i-1}^n, w) + \frac{1}{2\tau} d^2(u, u_{i-1}^n) + \phi(u) \leq \phi(w) \quad \forall w \in D(\phi).$$

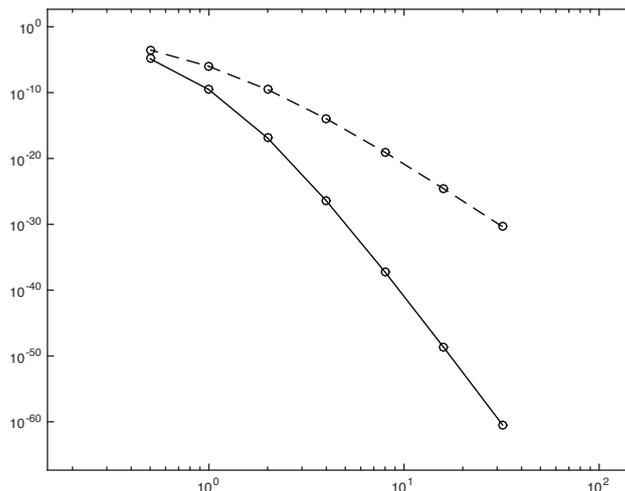


FIGURE 3. Values of the potential $\phi(u) = \lambda u^2/2$ ($\lambda > 0, d = 1$) after 20 iterations from $u_0 = 1$ with respect to $\lambda\tau_i$ (log–log scale): the new minimizing-movements scheme (1.5) for $p = 2$ (solid) and the proximal algorithm (dashed).

Such points exist for ϕ geodesically convex and the corresponding interpolants \bar{u}^n converge to *evolutionary variational inequality* solutions [38], a specific class of curves of maximal slope.

As mentioned in the Introduction, time-discrete schemes for gradient flows play a distinguished role in optimization, as the minimization of the potential ϕ is often tackled by iterative approximations. In case of a convex potential ϕ , a first example in this direction is the *proximal algorithm* $u_i = \text{prox}_{\tau_i\phi}(u_{i-1})$ where $\text{prox}_{\tau_i\phi} := (\text{id} + \tau_i\nabla\phi)^{-1} : \mathbb{R}^d \rightarrow \mathbb{R}^d$. (Here and in the rest of this section we assume the potential $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ to be smooth and $p = 2$, for the sake of notational definiteness.) The proximal algorithm corresponds to iterations of the implicit Euler method, starting from an initial guess u_0 .

A variety of modifications of the proximal algorithm has been advanced in order to solve optimization problems with specific structure with higher efficiency [14, 16]. In this connection, the new minimizing-movements scheme (1.5) corresponds to the proximal algorithm for the *augmented* potential $u \mapsto \phi(u) + \tau_i|\nabla\phi(u)|^2/2$, having the same critical points as ϕ , as soon as $D^2\phi$ is bounded below and τ_i is small enough. Note however that such augmented functional is generally not convex even if ϕ convex.

For ϕ convex, the new minimizing-movements scheme (1.5) corresponds to the following modification of the proximal algorithm

$$u_i = \text{prox}_{\tau_i\phi} \left(u_{i-1} - \tau_i D^2\phi(u_i) \nabla\phi(u_i) \right),$$

which is however implicit, due to the occurrence of the correction $\tau_i D^2\phi(u_i) \nabla\phi(u_i)$. Such correction may be however beneficial in terms of accelerating convergence along iterations. In Figure 3 we compare the reduction of the potential $\phi(u) = u^2/2$ ($d = 1$) after 20 iterations from $u_0 = 1$ for the proximal algorithm and the new minimizing-movement scheme (1.5). The reader is referred to [49] for discussion on the relation between discrete and continuous gradient-flow dynamics in presence of higher-order corrections and to [12] for an example of a second-order a stochastic-gradient descent method of Runge-Kutta type.

The literature on dynamic-systems methods for optimization is indeed broad. Without claiming completeness, we would like to point out the relevance of the discretization of inertial gradient systems of the form

$$u''(t) + \alpha(t)u'(t) + \nabla\phi(u(t)) = 0 \tag{3.23}$$

where $t \mapsto \alpha(t)$ is a suitably chosen damping coefficient, vanishing for $t \rightarrow \infty$. Starting from Nesterov's result [39], optimization based on discretization of inertial gradient flows has attracted attention, see [9, 10, 17, 20] among many others. Recent developments feature hessian-damping terms [4] of the form

$$u''(t) + \alpha u'(t) + \beta D^2\phi(u)u' + \nabla\phi(u(t)) = 0,$$

for $\alpha, \beta > 0$. Compared with the approach *via* (3.23), the hessian damping generally shows less oscillations along iterations [11, 18]. The new minimizing-movements scheme (1.5) could in principle be extended in the above mentioned directions.

Before concluding, let us remark that, in order to make the connection with optimization algorithms complete, the study of the convergence of the discrete sequence u_i as $i \rightarrow \infty$ would be required. This would call for studying the long-time discrete dynamics, possibly in relation to its corresponding continuous counterpart. Although some of our results may indeed be stated for the whole time semiline $t > 0$, the study of such asymptotic behavior is presently out of the scope of our analysis.

4. CONDITIONAL CONVERGENCE

This section is devoted to the proof of Theorem 3.1. The ingredients of the argument are quite classical. Still, as already mentioned, the current minimizing-movement setting of (3.2) expedites the proof, for there is no need to resort to the De Giorgi variational interpolant ([6], Def. 3.2.1).

Let $\{u_0^n\}$ be d -bounded with $u_0^n \xrightarrow{\sigma} u^0$ and $\phi(u_0^n) \rightarrow \phi(u^0)$. We have that

$$\begin{aligned} & \phi(\bar{u}^n(t_m^n)) + \frac{1}{p} \int_0^{t_m^n} |(\widehat{u}^n)'|^p(r) \, dr + \frac{1}{q} \int_0^{t_m^n} |\partial\phi|^q(\bar{u}^n(r)) \, dr \\ &= \phi(u_m^n) + \frac{1}{p} \sum_{i=1}^m (\tau_i^n)^{1-p} d^p(u_{i-1}^n, u_i^n) + \frac{1}{q} \sum_{i=1}^m \tau_i^n |\partial\phi|^q(u_i^n) \\ &= \sum_{i=1}^m G(\tau_i^n, u_{i-1}^n, u_i^n) + \phi(u_0^n). \end{aligned} \tag{4.1}$$

Condition (3.14) ensures that the above right-hand is bounded independently of $m = 1, \dots, N^n$ and n . A first consequence of estimate (4.1) is that $\{u_m^n\}$ is d -bounded independently of $m = 1, \dots, N^n$ and n . Indeed, one has that

$$\begin{aligned} d^p(u_0^n, u_m^n) &\leq 2^{p-1} \sum_{i=1}^m d^p(u_{i-1}^n, u_i^n) \leq 2^{p-1} (\tau^n)^{p-1} \sum_{i=1}^m (\tau_i^n)^{1-p} d^p(u_{i-1}^n, u_i^n) \\ &\leq 2^{p-1} (\tau^n)^{p-1} p \left(\sum_{i=1}^m G(\tau_i^n, u_{i-1}^n, u_i^n) + \phi(u_0^n) \right). \end{aligned}$$

The right-hand side is bounded independently of $m = 1, \dots, N^n$ and n . Since $\{u_0^n\}$ are d -bounded, the d -boundedness of $\{u_m^n\}$ follows.

As the sublevels of ϕ are sequentially σ -compact, one can apply the extended Ascoli-Arzelà Theorem from Proposition 3.3.1 of [6] and find a not relabeled subsequence $\{\bar{u}^n\}$ such that $\bar{u}^n \xrightarrow{\sigma} u$ pointwise, where $u : [0, T] \rightarrow U$, and $|(\widehat{u}^n)'| \rightarrow m$ weakly in $L^p(0, T)$. In particular, we have that $u(0) = \lim_{n \rightarrow \infty} u^n(0) = \lim_{n \rightarrow \infty} u_0^n = u^0$. For all $0 < s \leq t < T$, define $s^n = \max\{t_i^n : t_i^n < s\}$ and $t^n = \min\{t_i^n : t < t_i^n\}$. Then,

$$d(u(s), u(t)) \stackrel{(3.7)}{\leq} \liminf_{n \rightarrow \infty} d(\bar{u}^n(s), \bar{u}^n(t)) \leq \liminf_{n \rightarrow \infty} \int_{s^n}^{t^n} |(\widehat{u}^n)'|(r) \, dr = \int_s^t m(r) \, dr.$$

This entails that $u \in AC^p([0, T]; U)$ since we just checked that the function $m \in L^p(0, T)$ fulfills (2.1). As $|u'|$ is the minimal function in $L^p(0, T)$ fulfilling (2.1), we also have that $|u'| \leq m$ almost everywhere and

$$\int_0^t |u'|^p(r) dr \leq \int_0^t m^p(r) dr \leq \liminf_{\tau \rightarrow 0} \int_0^t |(\widehat{u}^n)'|^p(r) dr \quad \forall t > 0.$$

For all fixed $t \in (0, T]$, choose $t_m^n = t^n$ in (4.1) in order to get that

$$\phi(\bar{u}^n(t)) + \frac{1}{p} \int_0^{\bar{t}^n(t)} |(\widehat{u}^n)'|^p(r) dr + \frac{1}{q} \int_0^{\bar{t}^n(t)} |\partial\phi|^q(\bar{u}^n(r)) dr \stackrel{(4.1)}{\leq} \sum_{i=1}^{N^n} (G(\tau_i^n, u_{i-1}^n, u_i^n))^+ + \phi(u_0^n).$$

Owing to the sequential σ -lower semicontinuity of ϕ and $|\partial\phi|$, see (3.9)–(3.10), we can pass to the lim inf in the latter inequality and, using again condition (3.14) and the fact that $\phi(u_0^n) \rightarrow \phi(u^0)$, we obtain

$$\phi(u(t)) + \frac{1}{p} \int_0^t |u'|^p(r) dr + \frac{1}{q} \int_0^t |\partial\phi|^q(u(r)) dr \leq \phi(u(0)) \quad \forall t \in [0, T]. \quad (4.2)$$

As $|\partial\phi|$ is a strong upper gradient for ϕ by (3.10), we have that

$$\phi(u(0)) \leq \phi(u(t)) + \int_0^t |\partial\phi|(u(r)) |u'| (r) dr \leq \phi(u(t)) + \frac{1}{p} \int_0^t |u'|^p(r) dr + \frac{1}{q} \int_0^t |\partial\phi|^q(u(r)) dr$$

so that (4.2) is actually an equality and u is a curve of maximal slope in the sense of Definition 2.1.

5. CONVERGENCE IN THE GEODESICALLY CONVEX CASE

We now turn to the proof of Theorem 3.2.

Recall that for all $\tau_i^n > 0$ and $v \in D(\phi)$ the functional $u \in D(\phi) \mapsto E(\tau_i^n, v, u)$ admits a minimizer. We first prove a p -variant for $p > 1$ of the slope estimate ([6], Lem. 3.1.3, p. 61), which was originally proved for $p = 2$. In particular, we aim at the following

$$|\partial\phi|(u) \leq (\tau_i^n)^{1-p} d^{p-1}(v, u) \quad \forall u \in M_E(\tau_i^n, v). \quad (5.1)$$

Note that this estimate is already mentioned in Remark 3.1.7 of [6] without proof. We give an argument here. Let $w \in D(\phi)$ be given. From the minimality $E(\tau_i^n, v, u) \leq E(\tau_i^n, v, w)$ we deduce that

$$\begin{aligned} \phi(u) - \phi(w) &\leq \frac{(\tau_i^n)^{1-p}}{p} \left(d^p(v, w) - d^p(v, u) \right) \leq \frac{(\tau_i^n)^{1-p}}{p} \left((d(u, w) + d(v, u))^p - d^p(v, u) \right) \\ &= \frac{(\tau_i^n)^{1-p}}{p} \left(\sum_{k=0}^{\infty} \binom{p}{k} d^k(u, w) d^{p-k}(v, u) - d^p(v, u) \right) = d(u, w) \frac{(\tau_i^n)^{1-p}}{p} \sum_{k=1}^{\infty} \binom{p}{k} d^{k-1}(u, w) d^{p-k}(v, u) \end{aligned}$$

where we have made use of the generalized binomial formula and the generalized binomial coefficients

$$\binom{p}{k} = \frac{p(p-1)\dots(p-k+1)}{k!}.$$

Assume now that $w \neq u$, divide by $d(u, w)$, and compute the lim sup as $w \rightarrow u$ in order to get

$$\begin{aligned} |\partial\phi|(u) &= \limsup_{w \rightarrow u} \frac{(\phi(u) - \phi(w))^+}{d(u, w)} \leq \limsup_{w \rightarrow u} \frac{(\tau_i^n)^{1-p}}{p} \sum_{k=1}^{\infty} \binom{p}{k} d^{k-1}(u, w) d^{p-k}(v, u) \\ &= \frac{(\tau_i^n)^{1-p}}{p} \binom{p}{1} d^{p-1}(v, u) = (\tau_i^n)^{1-p} d^{p-1}(v, u) \end{aligned}$$

so that (5.1) holds. Above, we have used the fact that

$$\begin{aligned} 0 &\leq \lim_{w \rightarrow u} \sum_{k=2}^{\infty} \binom{p}{k} d^{k-1}(u, w) d^{p-k}(v, u) \leq \lim_{w \rightarrow u} d(u, w) \sum_{k=2}^{\infty} \binom{p}{k} d^{p-k}(v, u) \\ &= \lim_{w \rightarrow u} d(u, w) \left((1 + d(v, u))^p - \binom{p}{1} d^{p-1}(v, u) - \binom{p}{0} d^p(v, u) \right) = 0. \end{aligned}$$

Let now $u^e \in D(\phi)$ be a minimizer of $u \mapsto E(\tau_i^n, u_{i-1}^n, u)$. Taking into account the convexity assumption (3.12), let $\gamma : [0, 1] \rightarrow U$ be a curve with $\gamma(0) = u_{i-1}^n$ and $\gamma(1) = u^e$, so that

$$\begin{aligned} E(\tau_i^n, u_{i-1}^n, u^e) &\leq E(\tau_i^n, u_{i-1}^n, \gamma(\theta)) \\ &\stackrel{(3.12)}{\leq} \theta E(\tau_i^n, u_{i-1}^n, u^e) + (1 - \theta) E(\tau_i^n, u_{i-1}^n, u_{i-1}^n) \\ &\quad - \theta(1 - \theta) \frac{(p-1)(\tau_i^n)^{1-p} + \lambda}{p} d^p(u_{i-1}^n, u^e) \end{aligned}$$

where in the first inequality we have again used minimality. Let $\theta \in [0, 1)$, divide by $1 - \theta$, and take $\theta \rightarrow 1$ in order to get

$$E(\tau_i^n, u_{i-1}^n, u^e) + \frac{(\tau_i^n)^{1-p}}{q} d^p(u_{i-1}^n, u^e) \leq E(\tau_i^n, u_{i-1}^n, u_{i-1}^n) - \frac{\lambda}{p} d^p(u_{i-1}^n, u^e). \quad (5.2)$$

By taking the q -power of the slope estimate (5.1) with $v = u_{i-1}^n$ we get

$$|\partial\phi|^q(u^e) \leq (\tau_i^n)^{-p} d^p(u_{i-1}^n, u^e).$$

We use this to estimate from below the second term on the left-hand side of (5.2) obtaining

$$E(\tau_i^n, u_{i-1}^n, u^e) + \frac{\tau_i^n}{q} |\partial\phi|^q(u^e) \leq E(\tau_i^n, u_{i-1}^n, u_{i-1}^n) - \frac{\lambda}{p} d^p(u_{i-1}^n, u^e).$$

As $E(\tau_i^n, u_{i-1}^n, u_{i-1}^n) = 0$, given any $u_i^n \in M_G(\tau_i^n, u_{i-1}^n)$ the latter entails that

$$G(\tau_i^n, u_{i-1}^n, u_i^n) \leq G(\tau_i^n, u_{i-1}^n, u^e) = E(\tau_i^n, u_{i-1}^n, u^e) + \frac{\tau_i^n}{q} |\partial\phi|^q(u^e) \leq -\frac{\lambda}{p} d^p(u_{i-1}^n, u^e). \quad (5.3)$$

Recall now that the minimality $u^e \in M_E(\tau_i^n, u_{i-1}^n)$ and the nonnegativity of ϕ ensure that

$$\frac{(\tau_i^n)^{1-p}}{p} d^p(u_{i-1}^n, u^e) \leq \phi(u_{i-1}^n).$$

Hence, inequality (5.3) yields

$$G(\tau_i^n, u_{i-1}^n, u_i^n) \leq \lambda^- (\tau_i^n)^{p-1} \phi(u_{i-1}^n). \quad (5.4)$$

Taking the sum on $i = 1, \dots, m$ for $m \leq N^n$ we get

$$\begin{aligned} \phi(u_m^n) + \frac{1}{p} \sum_{i=1}^m (\tau_i^n)^{1-p} d^p(u_{i-1}^n, u_i^n) + \frac{1}{q} \sum_{i=1}^m \tau_i^n |\partial\phi|^q(u_i^n) - \phi(u^0) \\ = \sum_{i=1}^m G(\tau_i^n, u_{i-1}^n, u_i^n) \leq \lambda^- (\tau^n)^{p-2} \sum_{i=0}^{m-1} \tau_i^n \phi(u_i^n). \end{aligned}$$

We can hence use the discrete Gronwall Lemma and deduce that

$$\phi(u_m^n) + \frac{1}{p} \sum_{i=1}^m (\tau_i^n)^{1-p} d^p(u_{i-1}^n, u_i^n) + \frac{1}{q} \sum_{i=1}^m \tau_i^n |\partial\phi|^q(u_i^n) \leq \phi(u^0) \exp(\lambda^- (\tau^n)^{p-2} t_m^n).$$

Going back to (5.4), this entails that

$$(G(\tau_i^n, u_{i-1}^n, u_i^n))^+ \leq \lambda^- (\tau_i^n)^{p-1} \phi(u^0) \exp(\lambda^- (\tau^n)^{p-2} T).$$

Adding up for $i = 1, \dots, N^n$ we get

$$\sum_{i=1}^{N^n} (G(\tau_i^n, u_{i-1}^n, u_i^n))^+ \leq \lambda^- (\tau^n)^{p-2} T \phi(u^0) \exp(\lambda^- (\tau^n)^{p-2} T) =: R^n.$$

If $\lambda \geq 0$, we have that $R^n = 0$ and condition (3.14) trivially holds. If $\lambda < 0$ and $p > 2$, one can readily check that $R^n \rightarrow 0$ as $n \rightarrow \infty$ and (3.14) again holds.

6. CONVERGENCE WITHOUT GEODESIC CONVEXITY

We now turn to the proof of Theorem 3.3, where the convexity assumption is replaced by the generalized one-sided Taylor-expansion assumption (3.13). The argument follows the general strategy of Chapter 3 in [6], by revisiting the theory and adapting it to the incremental functional G and to the case $p > 1$. In particular, it is fairly different with respect to that of Section 5 and does not rely on the existence of solutions of the Euler scheme. We prepare some preliminary arguments in Sections 6.1–6.4, deduce an a priori estimate in Section 6.5 and eventually present the proof of Theorem 3.3 in Section 6.6.

6.1. A measurable selection in $\tau \mapsto M_G(\tau, v)$

Let us recall that for all $\tau \in (0, \tau_*]$ and $v \in D(\phi)$ the set of minimizers $M_G(\tau, v)$ is not empty. By additionally defining $M_G(0, v) = \{v\}$, the set-valued function $\tau \in [0, \tau_*] \mapsto M_G(\tau, v)$ has nonempty values. The aim of this section is to check that it admits a measurable selection, namely,

$$\exists \tau \in [0, \tau_*] \mapsto u_\tau \in M_G(\tau, v) \text{ measurable.} \quad (6.1)$$

To this aim, we firstly check that $M_G(\tau, v)$ is closed for all $\tau \in [0, \tau_*]$. Indeed, assume $\tau > 0$ (the case $\tau = 0$ being trivial) and let $u_k \in M_G(\tau, v)$ with $u_k \rightarrow u_\infty$. In particular, we have that

$$\phi(u_k) + \frac{\tau^{1-p}}{p} d^p(v, u_k) + \frac{\tau}{q} |\partial\phi|^q(u_k) - \phi(v) = G(\tau, v, u_k) \leq G(\tau, v, w)$$

for any $w \in D(|\partial\phi|)$. Owing to the lower semicontinuity (3.9)–(3.10) we can pass to the lower limit and check that $G(\tau, v, u_\infty) \leq G(\tau, v, w)$, so that $u_\infty \in M_G(\tau, v)$ as well.

Secondly, we check that $\tau \mapsto M_G(\tau, v)$ is measurable in the sense of set-valued functions [51]. In particular, we have to check that, for all $C \subset U$ closed, the set

$$A = \{\tau \in [0, \tau_*] : M_G(\tau, v) \cap C \neq \emptyset\}$$

is measurable. Indeed, one can prove that A is closed: Take $\tau_k \in A$ such that $\tau_k \rightarrow \tau_\infty$ and let $u_k \in M_G(\tau_k, v) \cap C$. We have that

$$\begin{aligned} \phi(u_k) + \frac{\tau_k^{1-p}}{p} d^p(v, u_k) + \frac{\tau_k}{q} |\partial\phi|^q(u_k) - \phi(v) &= G(\tau_k, v, u_k) \\ &\leq G(\tau_k, v, v) = \frac{\tau_k}{q} |\partial\phi|^q(v) < \infty. \end{aligned} \quad (6.2)$$

One can hence deduce uniform estimates for u_k and from compactness (3.9) one extracts a not relabeled subsequence such that $u_k \rightarrow u_\infty$. If $\tau_\infty > 0$, by passing to the liminf in the minimality condition for u_k one gets

$$G(\tau_\infty, v, u_\infty) \leq \liminf_{k \rightarrow \infty} G(\tau_k, v, u_k) \leq \liminf_{k \rightarrow \infty} G(\tau_k, v, w) = G(\tau_\infty, v, w)$$

for any $w \in D(|\partial\phi|)$. This implies that $u_\infty \in M_G(\tau_\infty, v)$. On the other hand, if $\tau_\infty = 0$ we obtain from (6.2) that

$$d^p(v, u_k) \leq p\tau_k^{p-1} \phi(v) + \frac{p\tau_k^p}{q} |\partial\phi|^q(v) \rightarrow 0,$$

so that $u_\infty = v \in M_G(0, v)$. Since C is closed, $u_\infty \in C$ as well and we have proved that $M_G(\tau_\infty, v) \cap C$ is not empty. In particular, $\tau_\infty \in A$ which is hence closed.

As the metric space (U, d) is complete and separable and $\tau \mapsto M_G(\tau, v)$ has nonempty and closed values, the Ryll-Nardzewski Theorem [47] applies and (6.1) holds.

6.2. Continuity of $\tau \mapsto \widehat{G}(\tau, v)$

We now turn our attention to the real map $\tau \in [0, \tau_*] \mapsto \widehat{G}(\tau, v)$, where we recall that

$$\begin{aligned} \widehat{G}(\tau, v) &= \min_{u \in D(|\partial\phi|)} G(\tau, v, u) \quad \text{with} \\ G(\tau, v, u) &= \phi(u) + \frac{\tau^{1-p}}{p} d^p(v, u) + \frac{\tau}{q} |\partial\phi|^q(u) - \phi(v) \end{aligned}$$

and define $\widehat{G}(0, v) = 0$. In order to check that $\tau \mapsto \widehat{G}(\tau, v)$ is continuous on $[0, \tau_*]$, let $\tau_k \in [0, \tau_*] \rightarrow \tau_\infty$ and take $u_k \in M_G(\tau_k, v)$. As in Section 6.1, we can extract a not relabeled subsequence such that $u_k \rightarrow u_\infty \in M_G(\tau_\infty, v)$.

If $\tau_\infty > 0$ the lower semicontinuity (3.9)–(3.10) implies that

$$\begin{aligned} G(\tau_\infty, v, u_\infty) &\leq \liminf_{k \rightarrow \infty} G(\tau_k, v, u_k) \leq \limsup_{k \rightarrow \infty} G(\tau_k, v, u_k) \\ &\leq \limsup_{k \rightarrow \infty} G(\tau_k, v, u_\infty) = G(\tau_\infty, v, u_\infty). \end{aligned}$$

The case $\tau_\infty = 0$ is even simpler as $u_\infty = v$ and we can compute

$$\begin{aligned} 0 = \widehat{G}(0, v) &= \phi(u_\infty) - \phi(v) \leq \liminf_{k \rightarrow \infty} \phi(u_k) - \phi(v) \leq \liminf_{k \rightarrow \infty} G(\tau_k, v, u_k) \\ &\leq \limsup_{k \rightarrow \infty} G(\tau_k, v, u_k) \leq \limsup_{k \rightarrow \infty} G(\tau_k, v, v) = \lim_{k \rightarrow \infty} \frac{\tau_k}{q} |\partial\phi|^q(v) = 0. \end{aligned} \quad (6.3)$$

In both cases, we have proved that $\widehat{G}(\tau_k, v) \rightarrow \widehat{G}(\tau_\infty, v)$.

6.3. Differentiability of $\tau \mapsto \widehat{G}(\tau, v)$

The aim of the subsection is to show that $\tau \mapsto \widehat{G}(\tau, v)$ is even locally Lipschitz continuous and to compute its almost-everywhere derivative, see equation (6.6) below.

Take $0 < \tau_0 < \tau_1 < \tau_*$, $u_0 \in M_G(\tau_0, v)$, and $u_1 \in M_G(\tau_1, v)$ where $v \in D(\phi)$ is fixed. From minimality we deduce

$$\widehat{G}(\tau_1, v) \leq G(\tau_1, v, u_0) = \widehat{G}(\tau_0, v) + \frac{\tau_1^{1-p} - \tau_0^{1-p}}{p} d^p(v, u_0) + \frac{\tau_1 - \tau_0}{q} |\partial\phi|^q(u_0)$$

so that one has

$$\widehat{G}(\tau_1, v) - \widehat{G}(\tau_0, v) \leq \frac{\tau_1^{1-p} - \tau_0^{1-p}}{p} d^p(v, u_0) + \frac{\tau_1 - \tau_0}{q} |\partial\phi|^q(u_0).$$

By exchanging the roles of τ_0 and τ_1 we also get

$$\widehat{G}(\tau_0, v) - \widehat{G}(\tau_1, v) \leq \frac{\tau_0^{1-p} - \tau_1^{1-p}}{p} d^p(v, u_1) + \frac{\tau_0 - \tau_1}{q} |\partial\phi|^q(u_1).$$

By dividing by $\tau_1 - \tau_0$ we hence obtain

$$\begin{aligned} \frac{\tau_1^{1-p} - \tau_0^{1-p}}{p(\tau_1 - \tau_0)} d^p(v, u_1) &\leq \frac{\tau_1^{1-p} - \tau_0^{1-p}}{p(\tau_1 - \tau_0)} d^p(v, u_1) + \frac{1}{q} |\partial\phi|^q(u_1) \leq \frac{\widehat{G}(\tau_1, v) - \widehat{G}(\tau_0, v)}{\tau_1 - \tau_0} \\ &\leq \frac{\tau_1^{1-p} - \tau_0^{1-p}}{p(\tau_1 - \tau_0)} d^p(v, u_0) + \frac{1}{q} |\partial\phi|^q(u_0) \leq \frac{1}{q} |\partial\phi|^q(u_0). \end{aligned} \quad (6.4)$$

The latter implies that $\tau \mapsto \widehat{G}(\tau, v)$ is locally Lipschitz continuous on $(0, \tau_*]$. Indeed, take $0 < \underline{\tau} < \tau_*$ and $\tau \in [\underline{\tau}, \tau_*]$. Given $u_\tau \in M_G(\tau, v)$, we readily deduce that

$$\begin{aligned} d^p(v, u_\tau) &\leq p\tau_*^{p-1} \phi(v) + \frac{p\tau_*^p}{q} |\partial\phi|^q(v), \\ \frac{1}{q} |\partial\phi|^q(u_\tau) &\leq \frac{1}{\underline{\tau}} \phi(v) + \frac{1}{q} |\partial\phi|^q(v), \end{aligned}$$

$$-\frac{\tau_1^{1-p} - \tau_0^{1-p}}{p(\tau_1 - \tau_0)} \leq \frac{1}{q\tau_0^p} \leq \frac{1}{q\underline{\tau}^p}.$$

In particular, moving from (6.4), for all $\underline{\tau} \in (0, \tau_*]$ we find $C_{\underline{\tau}}$ depending on $\underline{\tau}$, $\phi(v)$, and $|\partial\phi|(v)$ such that

$$\left| \frac{\widehat{G}(\tau_1, v) - \widehat{G}(\tau_0, v)}{\tau_1 - \tau_0} \right| \leq C_{\underline{\tau}} \quad \forall \underline{\tau} < \tau_0 < \tau_1 < \tau_*.$$

Hence, $\tau \in (0, \tau_*] \mapsto \widehat{G}(\tau, v)$ is locally Lipschitz continuous and therefore almost everywhere differentiable in $(0, \tau_*)$.

Define now

$$\begin{aligned} \bar{f}(\tau_0, \tau_1) &= \sup \left\{ \frac{\tau_1^{1-p} - \tau_0^{1-p}}{p(\tau_1 - \tau_0)} d^p(v, u_1) + \frac{1}{q} |\partial\phi|^q(u_1) : u_1 \in M_G(\tau_1, v) \right\}, \\ \underline{f}(\tau_0, \tau_1) &= \inf \left\{ \frac{\tau_1^{1-p} - \tau_0^{1-p}}{p(\tau_1 - \tau_0)} d^p(v, u_0) + \frac{1}{q} |\partial\phi|^q(u_0) : u_0 \in M_G(\tau_0, v) \right\}. \end{aligned}$$

By using again relation (6.4) one has

$$\bar{f}(\tau_0, \tau_1) \leq \frac{\widehat{G}(\tau_1, v) - \widehat{G}(\tau_0, v)}{\tau_1 - \tau_0} \leq \underline{f}(\tau_0, \tau_1). \quad (6.5)$$

Let $\tau \in (0, \tau_*)$ be such that $\tau \mapsto \widehat{G}(\tau, v)$ is differentiable at τ , take $h \in (0, \tau_* - \tau)$ and any $u_{\tau+h} \in M_G(\tau+h, v)$. By arguing as in Section 6.1, one can extract a not relabeled subsequence $u_{\tau+h} \rightarrow u_\tau$ as $h \rightarrow 0$ and check that $u_\tau \in M_G(\tau, v)$. Moreover, going back to (6.5) and choosing $\tau_0 = \tau$ and $\tau_1 = \tau + h$ we deduce that

$$\begin{aligned} -\frac{\tau^{-p}}{q} d^p(v, u_\tau) + \frac{1}{q} |\partial\phi|^q(u_\tau) &\leq \liminf_{h \rightarrow 0} \bar{f}(\tau, \tau + h) \\ &= \frac{d}{d\tau} \widehat{G}(\tau, v) \leq \liminf_{h \rightarrow 0} \underline{f}(\tau, \tau + h) \leq -\frac{\tau^{-p}}{q} d^p(v, \tilde{u}_\tau) + \frac{1}{q} |\partial\phi|^q(\tilde{u}_\tau) \end{aligned}$$

where \tilde{u}_τ is any element of $M_G(\tau, v)$. Passing to the infimum in $M_G(\tau, v)$ left and right we get

$$\frac{d}{d\tau} \widehat{G}(\tau, v) = \inf \left\{ -\frac{\tau^{-p}}{q} d^p(v, u_\tau) + \frac{1}{q} |\partial\phi|^q(u_\tau) : u_\tau \in M_G(\tau, v) \right\} \quad (6.6)$$

almost everywhere in $(0, \tau_*)$.

6.4. Slope estimate

Let us prepare a version of the slope estimate (5.1) adapted to our setting, namely for points in $u \in M_G(\tau, v)$ for $v \in D(\phi)$ instead of $M_E(\tau, v)$. Let $w \in D(|\partial\phi|)$ be given. From minimality we deduce that

$$\begin{aligned} \phi(u) - \phi(w) + \frac{\tau}{q} |\partial\phi|^q(u) - \frac{\tau}{q} |\partial\phi|^q(w) &\leq \frac{\tau^{1-p}}{p} (d^p(v, w) - d^p(v, u)) \\ &\leq d(u, w) \frac{\tau^{1-p}}{p} \sum_{k=1}^{\infty} \binom{p}{k} d^{k-1}(u, w) d^{p-k}(v, u). \end{aligned}$$

By assuming that $w \neq u$, dividing by $d(u, w)$, and taking $w \rightarrow u$ we get

$$|\partial(\phi + \tau|\partial\phi|^q/q)|(u) \leq \tau^{1-p}d^{p-1}(v, u) \quad \forall u \in M_G(\tau, v). \quad (6.7)$$

This proves in particular the additional regularity

$$M_G(\tau, v) \subset D(\partial(\phi + \tau|\partial\phi|^q/q))$$

for minimizers of G .

6.5. A priori estimate

Let now $\{u_i^n\}_{i=0}^{N^n}$ solve the incremental minimization problem (3.2) with u^0 replaced by the approximating $u^{0n} \in M_E(\tau^n, u^0)$. From minimality we obtain that

$$\phi(u_i^n) + \frac{\tau_i^n}{q}|\partial\phi|^q(u_i^n) + \frac{1}{p}(\tau_i^n)^{1-p}d^p(u_{i-1}^n, u_i^n) \leq \phi(u_{i-1}^n) + \frac{\tau_i^n}{q}|\partial\phi|^q(u_{i-1}^n). \quad (6.8)$$

Taking into account the one-sided nondegeneracy of the time partition

$$(\tau_i^n - \tau_{i-1}^n)^+ / \tau_{i-1}^n \leq C\tau^n$$

we can control the above right-hand side of (6.8) as follows

$$\begin{aligned} \phi(u_{i-1}^n) + \frac{\tau_i^n}{q}|\partial\phi|^q(u_{i-1}^n) &= \phi(u_{i-1}^n) + \frac{\tau_{i-1}^n}{q}|\partial\phi|^q(u_{i-1}^n) + \frac{\tau_i^n - \tau_{i-1}^n}{q}|\partial\phi|^q(u_{i-1}^n) \\ &\leq \phi(u_{i-1}^n) + \frac{\tau_{i-1}^n}{q}|\partial\phi|^q(u_{i-1}^n) + C\tau^n \frac{\tau_{i-1}^n}{q}|\partial\phi|^q(u_{i-1}^n). \end{aligned}$$

Owing to this bound, we can take the sum in (6.8) for $i = 2, \dots, m$ and get

$$\begin{aligned} \phi(u_m^n) + \frac{\tau_m^n}{q}|\partial\phi|^q(u_m^n) + \frac{1}{p} \sum_{i=1}^m (\tau_i^n)^{1-p}d^p(u_{i-1}^n, u_i^n) &\leq \phi(u_1^n) + \frac{\tau_1^n}{q}|\partial\phi|^q(u_1^n) + C \sum_{i=2}^m \tau^n \frac{\tau_{i-1}^n}{q}|\partial\phi|^q(u_{i-1}^n) \\ &\stackrel{(6.8)}{\leq} \phi(u^{0n}) + \frac{\tau^n}{q}|\partial\phi|^q(u^{0n}) + C \sum_{j=1}^{m-1} \tau^n \frac{\tau_j^n}{q}|\partial\phi|^q(u_j^n). \end{aligned}$$

By applying the discrete Gronwall Lemma we hence obtain

$$\phi(u_m^n) + \frac{\tau_m^n}{q}|\partial\phi|^q(u_m^n) + \frac{1}{p} \sum_{i=1}^m (\tau_i^n)^{1-p}d^p(u_{i-1}^n, u_i^n) \leq C \left(\phi(u^{0n}) + \frac{\tau^n}{q}|\partial\phi|^q(u^{0n}) \right). \quad (6.9)$$

Recall now that $u^{0n} \in M_E(\tau^n, u^0)$ and use the slope estimate (5.1) to get that

$$\frac{\tau^n}{q}|\partial\phi|^q(u^{0n}) \leq \frac{1}{q}(\tau^n)^{1-p}d^p(u^0, u^{0n}) \leq \frac{p}{q}\phi(u^0).$$

Hence, $\{u^{0n}\}$ are in particular d -bounded and the bound (6.9) entails the estimate

$$\phi(u_m^n) + \frac{\tau_m^n}{q} |\partial\phi|^q(u_m^n) + \frac{1}{p} \sum_{i=1}^m \tau^{1-p} d^p(u_{i-1}^n, u_i^n) \leq C \left(1 + \frac{p}{q}\right) \phi(u^0) \quad \forall m = 1, \dots, N^n, \quad \forall n. \quad (6.10)$$

6.6. Conclusion of the proof

For all $i = 1, \dots, N^n$ and $\tau_0 \in (0, \tau_i^n]$ we use the Lipschitz continuity of $\tau \in (0, \tau_i^n] \mapsto \widehat{G}(\tau, u_{i-1}^n)$ and write

$$\widehat{G}(\tau_i^n, u_{i-1}^n) = \widehat{G}(\tau_0, u_{i-1}^n) + \int_{\tau_0}^{\tau_i^n} \frac{d}{d\tau} \widehat{G}(\tau, u_{i-1}^n) d\tau. \quad (6.11)$$

Let now $\tau \in [\tau_0, \tau_i^n] \mapsto u_\tau$ be a measurable selection in $M_G(\tau, u_{i-1}^n)$. The existence of such a selection is ascertained in Section 6.1. Take $\tau_0 \rightarrow 0$ in (6.11), use $\widehat{G}(\tau_0, u_{i-1}^n) \rightarrow 0$ from (6.3) and (6.6) to get

$$G(\tau_i^n, u_{i-1}^n, u_i^n) \leq \int_0^{\tau_i^n} \left(-\frac{\tau^{-p}}{q} d^p(u_{i-1}^n, u_\tau) + \frac{1}{q} |\partial\phi|^q(u_\tau) \right) d\tau. \quad (6.12)$$

In order to conclude the proof of Theorem 3.2, one has to check that condition (3.14) holds, so that Theorem 3.1 applies. This calls for controlling the right-hand side of (6.12). By means of the slope estimate (6.7) for $v = u_{i-1}^n$ and $u = u_\tau$ we can control the right-hand of (6.12) as

$$G(\tau_i^n, u_{i-1}^n, u_i^n) \leq \int_0^{\tau_i^n} \left(\frac{1}{q} |\partial\phi|^q(u_\tau) - \frac{1}{q} |\partial(\phi + \tau |\partial\phi|^q/q)|^q(u_\tau) \right) d\tau.$$

As estimate (6.10) provides uniform bounds on $\phi(u_m^n)$ and $\tau_m^n |\partial\phi|^q(u_m^n)$, the generalized one-sided Taylor expansion condition (3.13) entails that

$$\sum_{i=1}^{N^n} (G(\tau_i^n, u_{i-1}^n, u_i^n))^+ \leq \frac{1}{q} \sum_{i=1}^{N^n} \int_0^{\tau_i^n} g(\tau) d\tau = \frac{1}{q} \sum_{i=1}^{N^n} \tau_i^n \left(\frac{1}{\tau_i^n} \int_0^{\tau_i^n} g(\tau) d\tau \right) \leq \frac{T}{q} \frac{1}{\tau^n} \int_0^{\tau^n} g(\tau) d\tau.$$

As $(1/\tau^n) \int_0^{\tau^n} g(\tau) d\tau \searrow 0$ as $\tau^n \rightarrow 0$ condition (3.14) holds. The statement hence follows from Theorem 3.1.

7. APPLICATIONS IN LINEAR SPACES

We collect in this section some comments on the application of the abstract convergence results of Theorem 3.1-3.3 in linear finite and infinite-dimensional spaces.

Let us start from the convex case of Theorem 3.2. We hence restrict to $p = 2$, for assumption (3.12) cannot hold for $p \neq 2$ in linear spaces, as commented in Section 3.1. Correspondingly, the potential ϕ is required to be convex ($\lambda \geq 0$).

In the finite-dimensional ODE case, let the proper, convex potential $\phi : \mathbb{R}^d \rightarrow [0, \infty]$ and the initial datum $u^0 \in D(\phi)$ be given. In this case, we have that $|\partial\phi|(u) = |(\partial\phi(u))^\circ|$, where $(\partial\phi(u))^\circ$ is the element of minimal norm in the convex and closed set $\partial\phi(u)$. In particular, $|\partial\phi|(u)$ is lower semicontinuous. As such, the new minimizing-movements scheme (3.2) has a solution $\{u_i^n\}$ for any partition and the corresponding interpolants converge to a solution of $u' + \partial\phi(u) \ni 0$, up to subsequences.

In order to give an application of Theorem 3.2 in infinite dimensions, we consider

$$\partial_t u - \nabla \cdot \beta(\nabla u) + \alpha(u) \ni 0 \quad \text{in } \Omega \times (0, T). \quad (7.1)$$

Here, $\Omega \subset \mathbb{R}^d$ is open, bounded, and smooth, $u : \Omega \times (0, T) \rightarrow \mathbb{R}$ is scalar-valued, and ∂_t and ∇ indicate partial derivatives in time and space, respectively. We assume that $\beta = \partial \widehat{\beta}$ and $\alpha = \partial \widehat{\alpha}$ where the potentials $\widehat{\beta} : \mathbb{R}^d \rightarrow [0, \infty]$ and $\widehat{\alpha} : \mathbb{R} \rightarrow [0, \infty]$ are proper and convex. In addition, we assume $\widehat{\beta}$ to be coercive in the following sense

$$\exists c_\beta > 0, m > \frac{2d}{d+2} : \quad \widehat{\beta}(\xi) \geq c_\beta |\xi|^m - \frac{1}{c_\beta} \quad \forall \xi \in \mathbb{R}^d. \quad (7.2)$$

Equation (7.1) is intended to be complemented with homogeneous Dirichlet boundary conditions (other choices being of course possible) hence corresponding to the gradient flow in $U = L^2(\Omega)$ of the functional

$$\phi(u) = \begin{cases} \int_{\Omega} (\widehat{\beta}(\nabla u) + \widehat{\alpha}(u)) \, dx & \text{for } u \in W_0^{1,m}(\Omega), \\ \infty & \text{with } \widehat{\beta}(\nabla u) + \widehat{\alpha}(u) \in L^1(\Omega) \\ \infty & \text{elsewhere in } L^2(\Omega). \end{cases} \quad (7.3)$$

As $\phi : U \rightarrow [0, \infty]$ is convex, proper, and lower semicontinuous, we have that $u \mapsto \partial\phi(u)$ is strongly-weakly closed and $|\partial\phi|(u) = \|(\partial\phi(u))^\circ\|$ (norm in $L^2(\Omega)$) is lower semicontinuous. Moreover, $\partial\phi$ fulfills the chain rule ([19], Lem. 3.3), so that $|\partial\phi|$ is a strong upper gradient. Note that the sublevels of ϕ are bounded in $W_0^{1,m}(\Omega)$, which embeds compactly into $L^2(\Omega)$. We can hence apply Theorem 3.2. In particular, the new minimizing-movements scheme (3.2) has a solution, which converges to a solution of (7.1), up to subsequences.

Let us now turn to some application of Theorem 3.3 to nonconvex problems. In the finite-dimensional case, assume ϕ to be twice differentiable and coercive with $\nabla\phi$ and $D^2\phi$ locally bounded. Then, one computes

$$\begin{aligned} & |\nabla\phi(u)|^q - |\nabla(\phi(u) + \tau|\nabla\phi(u)|^q/q)|^q \\ &= |\nabla\phi(u)|^q - |\nabla\phi(u) + \tau|\nabla\phi(u)|^{q-2}D^2\phi(u)\nabla\phi(u)|^q \\ &\leq (|\nabla\phi(u) + \tau|\nabla\phi(u)|^{q-2}D^2\phi(u)\nabla\phi(u)| + |\tau|\nabla\phi(u)|^{q-2}D^2\phi(u)\nabla\phi(u)|)^q \\ &\quad - |\nabla\phi(u) + \tau|\nabla\phi(u)|^{q-2}D^2\phi(u)\nabla\phi(u)|^q \\ &\leq \tau \sum_{k=1}^{\infty} \binom{q}{k} |\nabla\phi(u) + \tau|\nabla\phi(u)|^{q-2}D^2\phi(u)\nabla\phi(u)|^{q-k} |\nabla\phi(u)|^{q-2}D^2\phi(u)\nabla\phi(u)|^k \\ &\leq \tau (|\nabla\phi(u) + \tau|\nabla\phi(u)|^{q-2}D^2\phi(u)\nabla\phi(u)| + |\nabla\phi(u)|^{q-2}D^2\phi(u)\nabla\phi(u)|)^q. \end{aligned}$$

Hence, the one-sided Taylor-expansion condition (3.13) holds for the choice

$$g(\tau) = \tau \sup_{\phi(v) \leq C} (|\nabla\phi(v)| + 2|\nabla\phi(v)|^{q-1}|D^2\phi(v)|)^q.$$

Note that the above computation simplifies in case $p = 2$, for we have

$$\begin{aligned} & |\nabla\phi(u)|^2 - |\nabla(\phi(u) + \tau|\nabla\phi(u)|^2/2)|^2 = |\nabla\phi(u)|^2 - |\nabla\phi(u) + \tau D^2\phi(u)\nabla\phi(u)|^2 \\ & \quad - \tau^2 |D^2\phi(u)\nabla\phi(u)|^2 - 2\tau \nabla\phi(u) \cdot (D^2\phi(u)\nabla\phi(u)). \end{aligned}$$

In particular, if ϕ is convex condition (3.13) holds with the trivial choice $g(\tau) = 0$. In all cases, if $D^2\phi(u)$ is bounded below on sublevels of ϕ in the following sense

$$\forall C > 0, \exists c > 0, \forall v, \xi \in \mathbb{R}^d \text{ with } \phi(v) \leq C : \quad \xi \cdot D^2\phi(v)\xi \geq -\frac{c}{2}|\xi|^2 \quad (7.4)$$

and $\nabla\phi(u)$ is bounded on the sublevels of ϕ , namely, $|\nabla\phi(u)| \leq \ell(\phi(u))$ for some ℓ increasing, we can choose $g(\tau) = 2\tau c(\ell(C))^2$ in order to get again condition (3.13). This in particular applies to $\phi \in C^2$ and coercive. In all cases, we can apply Theorem 3.3 and deduce that the solution of the new minimizing-movements scheme converges up to subsequences to a solution of (1.3).

Let us now turn to the infinite-dimensional case. To simplify notation, let again $p = 2$ and $U = L^2(\Omega)$ (the case $p \neq 2$ and $U = L^p(\Omega)$ can also be treated) and define ϕ as in (7.3) by dropping the convexity requirement on $\hat{\alpha}$. More precisely, we ask $\beta = D\hat{\beta} \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ and $\alpha = \hat{\alpha}' \in C^2(\mathbb{R})$ and $\hat{\beta}$ fulfill the coercivity (7.2). In this case, we have that

$$\begin{aligned} \partial\phi(u) &= -\nabla \cdot \beta(\nabla u) + \alpha(u), \\ \text{with } D(\partial\phi) &= \{u \in L^2(\Omega) : -\nabla \cdot \beta(\nabla u) + \alpha(u) \in L^2(\Omega)\}. \end{aligned}$$

Recall that the Fréchet subdifferential [44] of $\psi : U \rightarrow [0, \infty]$ at $u \in D(\psi)$ is the set

$$\partial\psi(u) = \left\{ \xi \in U : \liminf_{v \rightarrow u} \frac{\psi(v) - \psi(u) - (\xi, v - u)}{\|v - u\|} \geq 0 \right\}$$

and $D(\partial\psi) = \{u \in D(\psi) : \partial\psi(u) \neq \emptyset\}$. In case of $\psi(u) = \|\partial\phi(u)\|^2/2$ we obtain that the Fréchet subdifferential is single-valued and

$$\partial \frac{1}{2} \|\partial\phi(u)\|^2 = \nabla \cdot (D\beta(\nabla u) \nabla (\nabla \cdot \beta(\nabla u) - \alpha(u))) - (\nabla \cdot \beta(\nabla u) - \alpha(u)) \alpha'(u)$$

with domain given by

$$\begin{aligned} D \left(\partial \frac{1}{2} \|\partial\phi(u)\|^2 \right) &= \left\{ u \in D(\partial\phi) : \partial \|\partial\phi(u)\|^2 \in L^2(\Omega), \right. \\ &\quad \left. \text{with } (\nabla \cdot \beta(\nabla u) - \alpha(u)) D\beta(\nabla u) \nu = 0 \text{ on } \partial\Omega \right\}. \end{aligned} \quad (7.5)$$

In particular, an extra natural boundary condition arises, where ν denotes the outer normal vector to $\partial\Omega$. In the linear case of $\beta(\xi) = \xi$ (see Sect. 3.3), we have that $D\beta = I$ (identity matrix) and we deduce again

$$\begin{aligned} \partial\phi(u) &= -\Delta u, \quad D(\partial\phi) = \{u \in H_0^1(\Omega) : -\Delta u \in L^2(\Omega)\} = H^2(\Omega) \cap H_0^1(\Omega), \\ \partial \frac{1}{2} \|\Delta u\|^2 &= \Delta^2 u, \\ D \left(\partial \frac{1}{2} \|\Delta u\|^2 \right) &= \{u \in H^2(\Omega) \cap H_0^1(\Omega) : \Delta^2 u \in L^2(\Omega) \text{ and } \Delta u = 0 \text{ on } \partial\Omega\} \\ &= \{u \in H^4(\Omega) \cap H_0^1(\Omega) : \Delta u \in H^2(\Omega) \cap H_0^1(\Omega)\}. \end{aligned}$$

In order to assess the one-sided Taylor-expansion condition 3.13 we argue as follows

$$\begin{aligned} &|\partial\phi|^2(u) - |\partial(\phi + \tau\partial|\partial\phi|^2/2)|^2(u) \\ &= \|\partial\phi(u)\|^2 - \|\partial\phi(u) + \tau\partial\|\partial\phi(u)\|^2/2\|^2 \\ &= -\tau^2 \|\nabla \cdot (D\beta(\nabla u) \nabla (\nabla \cdot \beta(\nabla u) - \alpha(u))) - (\nabla \cdot \beta(\nabla u) - \alpha(u)) \alpha'(u)\|^2 \\ &\quad + 2\tau \int_{\Omega} \left(\nabla \cdot (D\beta(\nabla u) \nabla (\nabla \cdot \beta(\nabla u) - \alpha(u))) - (\nabla \cdot \beta(\nabla u) - \alpha(u)) \alpha'(u) \right) \cdot (\nabla \cdot \beta(\nabla u) - \alpha(u)) \, dx \end{aligned}$$

$$\begin{aligned}
&\leq -2\tau \int_{\Omega} \nabla(\nabla \cdot \beta(\nabla u) - \alpha(u)) \cdot D\beta(\nabla u) \nabla(\nabla \cdot \beta(\nabla u) - \alpha(u)) \, dx \\
&\quad - 2\tau \int_{\Omega} \alpha'(u) (\nabla \cdot \beta(\nabla u) - \alpha(u))^2 \, dx
\end{aligned} \tag{7.6}$$

where we have used also the additional natural condition from (7.5) in the last inequality. The one-sided Taylor-expansion condition (3.13) then holds if $\widehat{\beta}$ and $\widehat{\alpha}$ are convex.

In addition, some nonconvex $\widehat{\alpha}$ can be considered as well. Assume $m > d$. Due to the coercivity of β , one has that the sublevels of ϕ are bounded in $W^{1,m}$ hence in L^∞ . In particular, $\phi(u) \leq c \Rightarrow \|u\|_{L^\infty} \leq \ell(c)$ for some $\ell : (0, \infty) \rightarrow (0, \infty)$ increasing. Assume u^0 to be given and use (6.10) to bound $\phi(u)$. Owing to the above discussion we hence have that $\|u\|_{L^\infty} \leq \ell(2C\phi(u^0))$ along the discrete evolution, where C is the constant in (6.10). Let now $C_P > 0$ be the Poincaré constant giving $\|w\|_{L^2}^2 \leq C_P \|\nabla w\|_{L^2}^2$ for all $w \in H_0^1(\Omega)$. Assume $\widehat{\alpha}$ to be such that α' is locally bounded from below. Under the following smallness assumption

$$\inf \{ \alpha'(r) : |r| \leq \ell(2C\phi(u^0)) \} \geq -\frac{c_\beta}{C_P}$$

one has that the right hand side of (7.6) can be controlled from above as

$$-2\tau c_\beta \|\nabla(\nabla \cdot \beta(\nabla u) - \alpha(u))\|_{L^2}^2 + 2\tau c_\beta |\Omega| + 2\tau \frac{c_\beta}{C_P} \|\nabla \cdot \beta(\nabla u) - \alpha(u)\|_{L^2}^2$$

and the one-sided Taylor-expansion condition (3.13) follows with $g(\tau) = 2\tau c_\beta |\Omega|$, at least on the relevant energy sublevel. In this case, Theorem 3.3 again ensures that the solution of the new minimizing-movement scheme converges to a solution of (7.1), up to subsequences.

8. APPLICATIONS IN WASSERSTEIN SPACES

Let us now give some detail in the direction of the application of the above theory to the case of the nonlinear diffusion equation (1.1). To start with, let us specify the space of probability measures of finite p -moment as

$$U = \mathcal{P}_p(\mathbb{R}^d) = \left\{ u \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^p du(x) < +\infty \right\}$$

where $\mathcal{P}(\mathbb{R}^d)$ denotes probability measures on \mathbb{R}^d , and endow it with the p -Wasserstein distance

$$W_p^p(u_1, u_2) = \inf \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\mu(x, y) : \mu \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_{\#}^1 \mu = u_1, \pi_{\#}^2 \mu = u_2 \right\}$$

where $u_1, u_2 \in \mathcal{P}_p(\mathbb{R}^d)$ and $\pi_{\#}^i$ denotes the push-forward of the projection π^i on the i -th component. Let σ indicate the *narrow* topology, namely, $u_n \xrightarrow{\sigma} u$ iff

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f(x) du_n(x) \rightarrow \int_{\mathbb{R}^d} f(x) du(x) \quad \forall f : \mathbb{R}^d \rightarrow \mathbb{R} \text{ continuous and bounded.}$$

Note that $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is a complete metric space ([6], Prop. 7.1.5) and that σ is compatible with W_p ([6], Lem. 7.1.4), namely, assumptions (3.5)–(3.7) hold.

Let us now fix some assumptions on potentials V , F , and W . We follow the setting of Section 10.4.7 of [6], also referring to Section 7 in [46] for some additional discussion. In particular, we assume

$$V : \mathbb{R}^d \rightarrow [0, \infty) \quad (\lambda, 2)\text{-convex with } \limsup_{|x| \rightarrow \infty} \frac{V(x)}{|x|^2} = \infty, \quad (8.1)$$

$$\begin{aligned} F : [0, \infty) \rightarrow \mathbb{R} \quad &\text{convex, differentiable, superlinear for } |x| \rightarrow \infty, F(0) = 0, \text{ and} \\ \exists C_F > 0 : \quad &F(x+y) \leq C_F(1 + F(x) + F(y)) \quad \forall x, y \in \mathbb{R}^d, \\ r \in (0, \infty) \mapsto r^d F(r^{-d}) \quad &\text{is convex and nonincreasing,} \end{aligned} \quad (8.2)$$

$$\begin{aligned} W : \mathbb{R}^d \rightarrow [0, \infty) \quad &\text{convex, differentiable, even, such that} \\ \exists C_W > 0 : \quad &W(x+y) \leq C_W(1 + W(x) + W(y)) \quad \forall x, y \in \mathbb{R}^d. \end{aligned} \quad (8.3)$$

Note that the assumptions on F cover the classical cases $F(r) = r \ln r$ and $F(r) = r^m$ for $m > 1$, respectively related to Fokker-Planck and porous media equations.

Under assumptions (8.1)–(8.3) we have that the potential ϕ from (1.2) is $(\lambda, 2)$ -geodesically convex. Combining this with the $(1, 2)$ -generalized-geodesic convexity of $u \mapsto W_2^2(v, u)$ ([6], Lem. 9.2.1) one has that condition (3.12) holds with $p = 2$. Note that resorting to generalized-geodesic convexity is here crucial, for the Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is positively curved ([40], Prop. 3.1), namely, $u \mapsto W_2^2(v, u)$ is actually $(1, 2)$ -geodesically *concave*. In addition, ϕ has σ -sequentially compact sublevels and its local slope $|\partial\phi|$ is a strong upper gradient and is σ -sequentially lower semicontinuous ([6], Prop. 10.4.14). In particular, (3.9)–(3.10) hold and we have the following.

Proposition 8.1. *Assume (8.1)–(8.3) and $u^0 \in \mathcal{P}_2(\mathbb{R}^d)$ with $\phi(u^0) < \infty$. Let $\{0 = t_0^n < t_1^n < \dots < t_{N^n}^n = T\}$ be a sequence of partitions with $\tau^n := \max(t_i^n - t_{i-1}^n) \rightarrow 0$ as $n \rightarrow \infty$. Moreover, let $u_i^n \in M_G(\tau_i^n, u_{i-1}^n)$ for $i = 1, \dots, N^n$. Then, up to a not relabeled subsequence, we have that $\bar{u}^n(t) \xrightarrow{\sigma} u(t)$, where $u \in AC^2([0, T]; \mathcal{P}_2(\mathbb{R}^d))$ and there exists a density $\rho : t \in [0, T] \rightarrow L^1(\mathbb{R}^d)$ such that $u(t) = \rho(t)\mathcal{L}^d$, $\int_{\mathbb{R}^d} \rho(x, t) d\mathcal{L}^d(x) = 1$, and $\int_{\mathbb{R}^d} |x|^2 \rho(x, t) d\mathcal{L}^d(x) < \infty$ for all $t \in [0, T]$, satisfying $u^0 = \rho(\cdot, 0)\mathcal{L}^d$ and the nonlinear diffusion equation*

$$\partial_t \rho - \operatorname{div}(\rho \nabla(V + F'(\rho) + W * \rho)) = 0 \quad \text{in } \mathcal{D}'(\mathbb{R}^d \times (0, T)).$$

The latter proposition is specifically targeting the case $p = 2$. Let us now turn to an application of the one-sided Taylor-expansion condition (3.13), which would then allow for general p . In the metric situation of (1.2), one can use condition (3.13) in the purely transport case $F = 0$ and $W = 0$. By assuming periodic boundary conditions, we formulate the problem on the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. Let $V \in C^3(\mathbb{T}^d)$ and define

$$\phi(u) = \int_{\mathbb{T}^d} V(x) du(x) \quad \forall u \in \mathcal{P}(\mathbb{T}^d). \quad (8.4)$$

From Proposition 10.4.2 of [6] we have that

$$|\partial\phi|^q(u) = \int_{\mathbb{T}^d} |\nabla V(x)|^q du(x).$$

Again in case $p = 2$ we obtain

$$\phi(u) + \frac{\tau}{2} |\partial\phi|^2(u) = \int_{\mathbb{T}^d} \left(V(x) + \frac{\tau}{2} |\nabla V(x)|^2 \right) du(x) =: \int_{\mathbb{T}^d} \tilde{V}(x) du(x).$$

One readily checks that $D^2\tilde{V} = D^2V + \tau D^3V \nabla V + \tau D^2V D^2V$ is bounded below. We can hence apply ([6], Prop. 10.4.2) once more and deduce that

$$\begin{aligned} |\partial\phi|^2(u) - |\partial(\phi + \tau|\partial\phi|^2/2)|^2(u) &= \int_{\mathbb{T}^d} \left(|\nabla V(x)|^2 - |\nabla\tilde{V}(x)|^2 \right) du(x) \\ &= \int_{\mathbb{T}^d} \left(|\nabla V(x)|^2 - |\nabla V(x) + \tau D^2V(x)\nabla V(x)|^2 \right) du(x) \\ &= - \int_{\mathbb{T}^d} \tau^2 |D^2V(x)\nabla V(x)|^2 du(x) - 2 \int_{\mathbb{T}^d} \tau \nabla V(x) \cdot D^2V(x)\nabla V(x) du(x) \\ &\leq 2\tau\lambda^- \int_{\mathbb{T}^d} |\nabla V(x)|^2 du(x) \leq 2\tau\lambda^- \|\nabla V\|_{L^\infty(\mathbb{T}^d)}^2 \end{aligned}$$

where we have defined $\lambda = \min\{\xi \cdot D^2V(x)\xi : x \in \mathbb{T}^d, \xi \in \mathbb{R}^d, |\xi| = 1\}$. Hence, condition (3.13) holds with $g(\tau) := 2\tau\lambda^- \|\nabla V\|_{L^\infty(\mathbb{T}^d)}^2$ (and, in particular, $g(\tau) = 0$ if V is convex).

In fact, the above computation can be adapted to the case $p \neq 2$ by letting $\tilde{V} = V + \tau|\nabla V|^q/q$. Let us shorten notation by denoting by $\xi(x) = \nabla V(x)$ and by $A(x) = D^2V(x)$. Then, $\nabla\tilde{V}(x) = \xi(x) + \tau|\xi(x)|^{q-2}A(x)\xi(x)$. We compute

$$\begin{aligned} |\partial\phi|^q(u) - |\partial(\phi + \tau|\partial\phi|^q/q)|^q(u) &= \int_{\mathbb{T}^d} \left(|\nabla V|^q - |\nabla\tilde{V}|^q \right) du = \int_{\mathbb{T}^d} \left(|\xi|^q - |\xi + \tau|\xi|^{q-2}A\xi|^q \right) du \\ &\leq \int_{\mathbb{T}^d} \left(\left(|\xi + \tau|\xi|^{q-2}A\xi| + |\tau|\xi|^{q-2}A\xi| \right)^q - |\xi + \tau|\xi|^{q-2}A\xi|^q \right) du \\ &= \int_{\mathbb{T}^d} \sum_{k=1}^{\infty} \binom{q}{k} |\xi + \tau|\xi|^{q-2}A\xi|^{q-k} |\tau|\xi|^{q-2}A\xi|^k du \leq \tau \sum_{k=1}^{\infty} \binom{q}{k} \|\xi + \tau|\xi|^{q-2}A\xi\|_{L^\infty(\mathbb{T}^d)}^{q-k} \|\tau|\xi|^{q-2}A\xi\|_{L^\infty(\mathbb{T}^d)}^k \\ &\leq \tau \left(\|\xi + \tau|\xi|^{q-2}A\xi\|_{L^\infty(\mathbb{T}^d)} + \|\tau|\xi|^{q-2}A\xi\|_{L^\infty(\mathbb{T}^d)} \right)^q. \end{aligned}$$

The one-sided Taylor-expansion condition (3.13) hence follows with the choice

$$g(\tau) = \tau \left(\|\nabla V\|_{L^\infty(\mathbb{T}^d)} + 2\|\nabla V\|_{L^\infty(\mathbb{T}^d)}^{q-1} \|D^2V\|_{L^\infty(\mathbb{T}^d)} \right)^q.$$

By applying Theorem 3.3 we obtain the following.

Proposition 8.2. *Assume $V \in C^3(\mathbb{T}^d)$ and $u^0 \in \mathcal{P}(\mathbb{T}^d)$. Let $\{0 = t_0^n < t_1^n < \dots < t_{N^n}^n = T\}$ be a sequence of partitions with $\tau^n := \max(t_i^n - t_{i-1}^n) \rightarrow 0$ as $n \rightarrow \infty$ and $(\tau_i^n - \tau_{i-1}^n)^+ / \tau_{i-1}^n \leq \hat{C}\tau^n$ for $i = 1, \dots, N^n$. Moreover, let $u_i^n \in M_G(\tau_i^n, u_{i-1}^n)$ for $i = 1, \dots, N^n$ and ϕ defined in (8.4). Then, up to a not relabeled subsequence, we have that $\bar{u}^n(t) \xrightarrow{\sigma} u(t)$, where $u \in AC^p([0, T]; \mathcal{P}(\mathbb{T}^d))$ satisfies $u(0) = u_0$ and the transport equation*

$$\partial_t u - \operatorname{div} (u|\nabla V|^{q-2}\nabla V) = 0 \quad \text{in } \mathcal{D}'(\mathbb{T}^d \times (0, T)).$$

Acknowledgements. This research is supported by the Austrian Science Fund (FWF) projects F 65, W 1245, I 4354, and P 32788 and by the OeAD-WTZ project CZ 01/2021.

REFERENCES

- [1] M. Agueh, Asymptotic behavior for doubly degenerate parabolic equations. *C. R. Math. Acad. Sci. Paris* **337** (2003) 331–336.
- [2] M. Agueh, Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory. *Adv. Differ. Equ.* **10** (2005) 309–360.

- [3] A.D. Alexandrov, A theorem on triangles in a metric space and some applications. *Trudy Math. Inst. Steklov* **38** (1951) 5–23.
- [4] F. Alvarez, H. Attouch, J. Bolte and P. Redont, A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. *J. Math. Pures Appl. (9)* **81** (2002) 747–779.
- [5] L. Ambrosio, Minimizing movements. *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5)*, **19** (1995) 191–246.
- [6] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures, second ed., Birkhäuser Verlag, Basel (2008).
- [7] E. Asplund, Averaged norms. *Israel J. Math.* **5** (1967) 227–233.
- [8] E. Asplund, Topics in the theory of convex functions. In *Theory and Applications of Monotone Operators (Proc. NATO Advanced Study Inst., Venice, 1968)*. Oderisi, Gubbio (1969), pp. 1–33.
- [9] H. Attouch and A. Cabot, Asymptotic stabilization of inertial gradient dynamics with time dependent viscosity. *J. Differ. Equ.* **263** (2017) 5412–5458.
- [10] H. Attouch, Z. Chbani, J. Peyrouquet and P. Redont, Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program. Ser. B* **168** (2018) 123–175.
- [11] H. Attouch, J. Peyrouquet and P. Redont, Fast convex optimization via inertial dynamics with Hessian driven damping. *J. Differ. Equ.* **261** (2016) 5734–5783.
- [12] I. Ayadi and G. Turinici, Stochastic Runge-Kutta methods and adaptive SGD-G2 stochastic gradient descent. Preprint [arXiv:2002.09304](https://arxiv.org/abs/2002.09304) (2020).
- [13] A. Bacho, E. Emmrich and A. Mielke, An existence result and evolutionary Γ -convergence for perturbed gradient systems. *J. Evol. Equ.* **19** (2019) 479–522.
- [14] H.H. Bauschke and P.L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York (2011).
- [15] H.H. Bauschke and P.L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces. Second edition. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham (2017).
- [16] D.P. Bertsekas, Convex optimization algorithms. Athena Scientific, Belmont, MA (2015).
- [17] R.I. Bot, E.R. Csetnek and S.C. László, Approaching nonsmooth nonconvex minimization through second order proximal-gradient dynamical systems. *J. Evol. Equ.* **18** (2018) 1291–1318.
- [18] R.I. Bot and E.R. Csetnek, A second-order dynamical system with Hessian-driven damping and penalty term associated to variational inequalities. *Optimization* **68** (2019) 1265–1277.
- [19] H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. Number 5 in North Holland Math. Studies. North-Holland, Amsterdam (1973).
- [20] A. Cabot, H. Engler and S. Gadat, On the long time behavior of second order differential equations with asymptotically small dissipation. *Trans. Am. Math. Soc.* **361** (2009) 5983–6017.
- [21] J.A. Carrillo, R.J. McCann and C. Villani, Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana* **19** (2003) 971–1018.
- [22] J.A. Carrillo, R.J. McCann and C. Villani, Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.* **179** (2006) 217–263.
- [23] J. Cheeger, Differentiability of Lipschitz functions on metric measure spaces. *Geom. Funct. Anal.* **9** (1999) 428–517.
- [24] P. Clément and W. Desch, Some remarks on the equivalence between metric formulations of gradient flows. *Boll. Unione Mat. Ital. (9)* **3** (2010) 583–588.
- [25] P. Clément and W. Desch, A Crandall-Liggett approach to gradient flows in metric spaces. *J. Abstr. Differ. Equ. Appl.* **1** (2010) 46–60.
- [26] P. Clément and J. Maas, A Trotter product formula for gradient flows in metric spaces. *J. Evol. Equ.* **11** (2011) 405–427.
- [27] E. De Giorgi, New problems on minimizing movements. In: *Boundary value problems for PDE and applications*, edited by C. Baiocchi and J.-L. Lions. Masson, Paris (1993), pp. 81–98.
- [28] E. De Giorgi, A. Marino and M. Tosques, Problems of evolution in metric spaces and maximal decreasing curve. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)* **68** (1980) 180–187.
- [29] M.H. Duong, M.A. Peletier and J. Zimmer, GENERIC formalism of a Vlasov–Fokker–Planck equation and connection to large-deviation principles. *Nonlinearity* **26** (2013) 2951–2971.
- [30] F. Fleißner, Γ -convergence and relaxations for gradient flows in metric spaces: a minimizing movement approach. *ESAIM: COCV* **25** (2019) Paper No. 28, 29 pp.
- [31] E. Hairer, C. Lubich and G. Wanner, Geometric Numerical Integration, second edition. Springer, Berlin (2006).
- [32] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** (1998) 1–17.
- [33] J. Jost, *Nonpositive curvature: geometric and analytic aspects*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (1997).
- [34] A. Jüngel, U. Stefanelli and L. Trussardi, Two structure-preserving time discretizations for gradient flows. *Appl. Math. Optim.* **80** (2020) 733–764.
- [35] G. Legendre and G. Turinici, Second-order in time schemes for gradient flows in Wasserstein and geodesic metric spaces. *C. R. Math. Acad. Sci. Paris* **355** (2017) 345–353.
- [36] D. Matthes and S. Plazotta, A variational formulation of the BDF2 method for metric gradient flows. *ESAIM: Math. Model. Numer. Anal.* **53** (2019) 145–172.

- [37] A. Mielke, R. Rossi and G. Savaré, BV solutions and viscosity approximations of rate-independent systems. *ESAIM: COCV* **18** (2012) 36–80.
- [38] M. Muratori and G. Savaré, Gradient flows and evolution variational inequalities in metric spaces. I: Structural properties. *J. Funct. Anal.* **278** (2020) 108347, 67 pp.
- [39] Y.E. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR* **269** (1983) 543–547.
- [40] S. Ohta, Gradient flows on Wasserstein spaces over compact Alexandrov spaces. *Am. J. Math.* **131** (2009) 475–516.
- [41] F. Otto, The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differ. Equ.* **26** (2001) 101–174.
- [42] L. Portinale and U. Stefanelli, Penalization via global functionals of optimal-control problems for dissipative evolution. *Adv. Math. Sci. Appl.* **28** (2019) 425–447.
- [43] T. Roche, R. Rossi and U. Stefanelli, Stability results for doubly nonlinear differential inclusions by variational convergence. *SIAM J. Control Optim.* **52** (2014) 1071–1107.
- [44] R.T. Rockafellar, Convex analysis. Princeton University Press (1970).
- [45] 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.
- [46] R. Rossi, A. Segatti and U. Stefanelli, Global attractors for gradient flows in metric spaces. *J. Math. Pures Appl.* **95** (2011) 204–244.
- [47] C. Ryll-Nardzewski, On Borel measurability of orbits. *Fund. Math.* **56** (1964) 129–130.
- [48] F. Santambrogio, {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bull. Math. Sci.* **7** (2017) 87–154.
- [49] B. Shi, S.S. Du, M.I. Jordan and W.J. Su, Understanding the acceleration phenomenon via high-resolution differential equations. *Math. Program.* (2021). <https://doi.org/10.1007/s10107-021-01681-8>.
- [50] A. Tribuzio, Perturbations of minimizing movements and curves of maximal slope. *Netw. Heterog. Media* **13** (2018) 423–448.
- [51] D.H. Wagner, Survey of measurable selection theorems. *SIAM J. Control Optim.* **15** (1977) 859–903.

Subscribe to Open (S2O)

A fair and sustainable open access model



This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

Please help to maintain this journal in open access!

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting subscribers@edpsciences.org

More information, including a list of sponsors and a financial transparency report, available at: <https://www.edpsciences.org/en/math-s2o-programme>