

THE BRØNSTED–ROCKAFELLAR THEOREM IN GEODESIC SPACES

ALBERTO DOMÍNGUEZ CORELLA^{1,2}  AND ALEJANDRO VILLEGAS-ACUÑA^{3,*} 

Abstract. We present a constructive version of the Brøndsted–Rockafellar theorem in general geodesic metric spaces. Applications include a constructive form of the Caristi theorem and quantitative relations between metric slope error bounds and the global growth of functionals.

Mathematics Subject Classification. 47H10, 49J52, 49K27, 49Q22, 51F99.

Received November 11, 2025. Accepted May 29, 2026.

1. INTRODUCTION, MAIN RESULTS AND STATE OF THE ART

1.1. Motivation, the classical Brøndsted–Rockafellar theorem and summary of results

Any measurement comes with finite precision, and the *true state* of a system can only be observed up to a tolerance. For this reason, approximate solutions are often as good as, and frequently indistinguishable, from true ones. Since mathematical laws governing physical systems are typically formulated in variational form, it is natural to consider *almost critical points*.

The Brøndsted–Rockafellar theorem occupies a central place in variational analysis. It establishes that every approximate minimizer of a convex functional can be *lifted* to a nearby point admitting a subgradient of small norm—that is, an almost critical point.

Theorem (Brøndsted–Rockafellar [1] Lemma on page 608). *Let X be a real Banach space and $\Phi : X \rightarrow \mathbb{R} \cup \{+\infty\}$ a proper convex lower semicontinuous functional bounded from below. Let $u \in X$ and $\varepsilon > 0$ be such that*

$$\Phi(u) \leq \inf_{v \in X} \Phi(v) + \varepsilon.$$

Then, for any $\lambda > 0$, there exist $\hat{u} \in X$ and $\xi \in \partial\Phi(\hat{u})$ satisfying

$$(a) \quad \|u - \hat{u}\| \leq \lambda \quad (b) \quad \Phi(\hat{u}) \leq \Phi(u) \quad (c) \quad \|\xi\| \leq \frac{\varepsilon}{\lambda}.$$

Keywords and phrases: Variational principle, Wasserstein space, geodesic metric space, Brøndsted–Rockafellar, Carlier inequality.

¹ Institut für Mathematik und Wissenschaftliches Rechnen, Universität Graz, Heinrichstraße 36, 8010 Graz, Austria.

² Institut für Stochastik und Wirtschaftsmathematik, Technische Universität Wien, Wiedner Hauptstraße 8, VADOR E105-04, 1040 Wien, Austria.

³ Departamento de Matemáticas, Universidad de Sonora, Hermosillo, Sonora 83000, Mexico.

* Corresponding author: francisco.villegas@unison.mx

One of the main drawbacks of this result is that it is not constructive; its original proof relied on an order argument *via* Zorn's lemma, adapted from the proof of the celebrated Bishop–Phelps lemma [2], Lemma 1 and Theorem 2. Modern proofs use Ekeland's variational principle [3], Theorem 1.1, which does not require the axiom of choice.

Recently, in [4], Theorem 4.1, a constructive version of the theorem was established in Hilbert spaces. The result provides the desired point by means of the resolvent of the subdifferential. The proof relies on what is now referred to as *Carlier's inequality* [5–7], an estimate relating the Fenchel–Young gap to the squared resolvent residual. The construction in [4], Theorem 4.1 yields that the choices

$$\hat{u} = \left(\text{id} + \frac{\lambda^2}{\varepsilon} \partial\Phi \right)^{-1} (u) \quad \text{and} \quad \xi = \frac{\varepsilon}{\lambda^2} (u - \hat{u}) \quad (1.1)$$

satisfy items (a)–(c) in the theorem stated above. An advantage of this explicit construction is that it enables more refined arguments in applications of the theorem. It also fits naturally within the frameworks of constructive mathematics and computable analysis.

The Brøndsted–Rockafellar theorem has been extended in many directions, including operator-theoretic [8–11] and convex-analytic [12–17]. But so far only [4], Theorem 4.1 has given an explicit statement where a construction is provided; a similar construction was used in the proof of [16], Proposition 2.3.

This paper is devoted to a constructive extension of the Brøndsted–Rockafellar theorem to geodesic metric spaces. Such spaces include, for instance, normed linear spaces endowed with the distance induced by the norm, the hyperbolic space equipped with its Riemannian metric (and, more generally, CAT(0) spaces), and the space of probability measures with finite second moment on the Euclidean space endowed with the Wasserstein distance. The extension to this latter space is particularly significant, since a notion of subdifferential is available; see [18], Section 10 and [19], Section 4.

The contributions of this paper can be summarized as follows.

- (1) **Extension to geodesic spaces.** We provide two versions of the Brøndsted–Rockafellar theorem. The first one comes in the form of Ekeland's variational principle (Thm. 1.1 below), while the second one (Thm. 1.7 below) is more faithful to the original theorem, replacing the bound on the subdifferential by a bound on the metric slope. The resolvent construction in (1.1) is replaced by proximal points (in Hilbert spaces there is no distinction between resolvent and proximal points). We circumvent Carlier's inequality, which is intrinsically related to the linear structure due to the use of the Legendre–Fenchel transform, through purely metric arguments. We give some pertinent examples illustrating the applicability of the results.
- (2) **Caristi fixed point theorem.** As an application of Theorem 1.1, we recover a metric version of the Caristi fixed point theorem for geodesically convex functionals. The result (stated in Prop. 1.10 below) is constructive and relies on the existence of proximal points.
- (3) **A result on global error bounds.** We show (in Prop. 1.11 below) that if a geodesically convex functional satisfies an error bound controlling the distance to the minimizer set by a power of the metric slope, then it necessarily satisfies a quantitative growth estimate. This extends to geodesic metric spaces the classical result, known for convex functionals in Banach spaces, that the metric sub-regularity of the subdifferential implies a polynomial growth condition at the set of minimizers.
- (4) **A construction in the Wasserstein space.** In Section 3, we provide more refined results than the general theorems in this section for the particular case of the space of probability measures with the Wasserstein distance. In Theorem 3.1, we obtain an analogue of the construction in (1.1), formulated in terms of optimal transport maps and the minimizing movement scheme. Furthermore, in Theorem 3.3, we establish the equivalence between the metric sub-regularity of the subdifferential and the so-called entropy–transportation inequality (also referred to as the Talagrand inequality).

Below in this section we present the main results; their proofs are given in the next section, and the results specific to the Wasserstein space are collected in Section 3.

1.2. The theorem in geodesic spaces and some remarks

Throughout this subsection, (X, d) denotes a geodesic space; that is, a metric space such that for every $u, v \in X$ there exists a constant-speed¹ geodesic $\gamma : [0, 1] \rightarrow X$ such that $\gamma(0) = u$ and $\gamma(1) = v$. Additionally, we consider a proper functional $\Phi : X \rightarrow \mathbb{R} \cup \{+\infty\}$ bounded from below such that

$$\Phi(\gamma(t)) \leq (1-t)\Phi(\gamma(0)) + t\Phi(\gamma(1)) \quad \forall t \in [0, 1] \quad (1.2)$$

for every constant-speed geodesic $\gamma : [0, 1] \rightarrow X$. Condition (1.2) is referred to as geodesic convexity of Φ . We fix numbers $p, q \in (1, +\infty)$ such that $p^{-1} + q^{-1} = 1$.

1.2.1. The theorem in variational form

Consider the set-valued mapping $\text{prox}_\Phi : X \rightrightarrows X$ given by

$$\text{prox}_\Phi(u) := \arg \min_{v \in X} \left\{ \Phi(v) + \frac{1}{p} d(v, u)^p \right\}.$$

This operator is usually referred to as the *proximal mapping*, and in some contexts it can be interpreted as a resolvent operator. The main result of this subsection establishes that the proximal mapping naturally yields points satisfying the conditions of Ekeland's variational principle.

Theorem 1.1. *Let $u \in X$ and $\varepsilon > 0$ be such that*

$$\Phi(u) \leq \inf_{v \in X} \Phi(v) + \varepsilon.$$

Then, for any $\lambda > 0$ and $\hat{u} \in \text{prox}_{\frac{\lambda^p}{\varepsilon}\Phi}(u)$, the following estimates are satisfied.

$$(a) \ d(u, \hat{u}) \leq \lambda \quad (b) \ \Phi(\hat{u}) \leq \Phi(u) \quad (c) \ \Phi(\hat{u}) \leq \Phi(v) + \frac{\varepsilon}{\lambda} d(v, \hat{u}) \quad \forall v \in X.$$

We now make some remarks.

Remark 1.2. While Theorem 1.1 does not require completeness of the underlying space, it does rely on the existence of minimizers of a regularized problem. An advantage of it is that it gives an explicit and computable representation of the approximate minimizer, making the result more constructive rather than purely existential. The assumption that proximal points exist is common in the literature, as it simplifies the analysis and avoids technical complications concerning the well-posedness of minimizing movement schemes; *e.g.*, [18], Assumption (10.1.1b).

Remark 1.3. We observe that Theorem 1.1 does not require the functional to be lower semicontinuous at every point. However, whenever a proximal point exists, the functional is automatically lower semicontinuous at that point.

Remark 1.4. The point $\hat{u}$ in Theorem 1.1 can be viewed as a step of the so-called p -generalized minimizing movement scheme in the calculus of variations framework; see, *e.g.*, [18], Definition 2.0.6.

¹A function $\gamma : [0, 1] \rightarrow X$ is said to be a *constant-speed geodesic* if

$$d(\gamma(s), \gamma(t)) = |s - t| d(\gamma(0), \gamma(1)) \quad \forall s, t \in [0, 1].$$

Remark 1.5. In normed spaces, the inequality in Theorem 1.1-(c) is equivalent to the existence of an element $\xi \in X^*$ satisfying $\xi \in \partial\Phi(\hat{u})$ and $\|\xi\| \leq \varepsilon/\lambda$. A refinement of Theorem 1.1 in the particular case of normed spaces is given in Appendix A.

Remark 1.6. Theorem 1.1 is stated for exponents strictly between one and infinity. The remaining cases are discussed in the appendix. The limiting regime in which the proximal regularization becomes a hard constraint is treated in Appendix B. Appendix C presents a more general theorem based on arbitrary convex proximal kernels, which includes, in particular, the case of exponent one.

1.2.2. The theorem with the metric slope

The *metric slope* of Φ at a point $u \in X$ is given by

$$|\partial\Phi|(u) := \max \left\{ \limsup_{v \rightarrow u} \frac{\Phi(u) - \Phi(v)}{d(u, v)}, 0 \right\}. \quad (1.3)$$

Quantity (1.3) is always nonnegative, and vanishes exactly at the minimizers of Φ . If X is a real normed space, then the metric slope coincides with the distance of the convex subdifferential to zero, *i.e.*, $|\partial\Phi|(u) = \text{dist}(0, \partial\Phi(u))$ for all $u \in X$, where $\partial\Phi : X \rightrightarrows X^*$ denotes the convex differential of Φ ; see [20], Proposition 2.1 (vii). This identity is one of the reasons for the term *metric slope*.

We now give an alternative version of Theorem 1.1.

Theorem 1.7. *Let $u \in X$ and $\varepsilon > 0$ be such that*

$$\Phi(u) \leq \inf_{v \in X} \Phi(v) + \varepsilon.$$

Then, for any $\lambda > 0$ and $\hat{u} \in \text{prox}_{\frac{\lambda^p}{\varepsilon}\Phi}(u)$, the following estimates are satisfied.

$$(a) \ d(u, \hat{u}) \leq \lambda \quad (b) \ \Phi(\hat{u}) + \frac{\lambda^q}{p\varepsilon^{\frac{q}{p}}} |\partial\Phi|(\hat{u})^q \leq \Phi(u) \quad (c) \ |\partial\Phi|(\hat{u}) \leq \frac{\varepsilon}{\lambda}.$$

We now make a couple of remarks.

Remark 1.8. Interpreting Φ as a potential energy and $(u, |\partial\Phi|(u))$ as a position–speed pair, the inequality in Theorem 1.7-(b) takes the form of an energy balance. The functional

$$X \ni v \mapsto \Phi(v) + \frac{\lambda^q}{p\varepsilon^{\frac{q}{p}}} |\partial\Phi|(v)^q \in \mathbb{R} \cup \{+\infty\}$$

can be viewed as a total energy (the sum of a potential and a kinetic term). The inequality then states that moving from the initial rest position u to the new one $\hat{u}$ does not increase the total energy.

Remark 1.9. It follows from Theorem 1.7 that if for every $u \in \text{dom } \Phi$ there exists a sequence $(t_n)_{n \in \mathbb{N}} \subseteq (0, +\infty)$ converging to zero such that $\text{prox}_{t_n \Phi}(u) \neq \emptyset$, then $\text{dom } |\partial\Phi|$ is dense in $\text{dom } \Phi$.

1.3. A couple of consequences

We now give a couple of applications of Theorem 1.1. As in the previous subsection, we consider a geodesic space (X, d) and a proper geodesically convex functional $\Phi : X \rightarrow \mathbb{R} \cup \{+\infty\}$ bounded from below.

1.3.1. The fixed-point theorem of Caristi

A classical consequence of Ekeland’s variational principle is the so-called Caristi fixed-point theorem [21], Theorem (2.1)’. A constructive variant of this result can be recovered for geodesically convex functionals by means of Theorem 1.1.

Proposition 1.10 (Caristi fixed-point theorem). *Let $T : X \rightarrow X$ be a mapping. Assume that*

$$\Phi(v) \geq \Phi(T(v)) + d(v, T(v)) \quad \forall v \in X. \quad (1.4)$$

Suppose further that there exist $u, \hat{u} \in X$ and $t > 0$ such that

$$\Phi(u) < \inf_{v \in X} \Phi(v) + t^{\frac{q}{p}} \quad \text{and} \quad \hat{u} \in \text{prox}_{t\Phi}(u).$$

Then, $\hat{u}$ is a fixed point of T , i.e., $T(\hat{u}) = \hat{u}$.

This formulation shows that the existence of minimizers of regularized problems, combined with geodesic convexity and condition (1.4), is sufficient to guarantee fixed-point properties, even without completeness or lower semicontinuity.

1.3.2. A necessary condition for metric sub-regularity

Metric sub-regularity is a fundamental stability property of set-valued mappings; it relates the distance to the solution set with a suitable residual measure; see [22–25] for detailed references. This property for convex subdifferentials was first characterized in [26], Theorem 3.3 through growth conditions *via* Ekeland’s variational principle. Using the metric slope, the global sub-regularity of the subdifferential may be interpreted as a quantitative relation between the slope of a functional and its deviation from the minimum value; see [27], Remark 6.2.

The following result provides a necessary condition for the sub-regularity property in terms of an error bound over the domain of the proximal mapping.

Proposition 1.11. *Suppose that Φ has at least one minimizer. If there exists $\kappa > 0$ such that*

$$\text{dist}(u, \arg \min_{v \in X} \Phi(v)) \leq \kappa |\partial \Phi|(u)^{q/p} \quad \forall u \in X, \quad (1.5)$$

then, for any $t > 0$, there holds

$$\Phi(u) \geq \inf_{v \in X} \Phi(v) + \frac{t^{q/p}}{(t^{q/p} + \kappa)^p} \text{dist}(u, \arg \min_{v \in X} \Phi(v))^p \quad \forall u \in \text{dom prox}_{t\Phi}. \quad (1.6)$$

The growth condition (1.6) is a well-known error bound; see [20, 27, 28]. In Wasserstein spaces, it is common to refer to conditions such as (1.6) as entropy–transport inequalities or Talagrand inequalities; see [29], Section 3.8 and [30], Section 3.1.

1.4. Some pertinent examples

We present some relevant examples concerning the applicability of Theorem 1.1 in situations where the assumptions of the classical Brøndsted–Rockafellar theorem are not satisfied.

1.4.1. Lack of lower semicontinuity

Let H be a Hilbert space and $C \subseteq H$ a nonempty open bounded convex set. Fix $\xi \in H$ and consider the functional $\Phi : H \rightarrow \mathbb{R} \cup \{+\infty\}$ given by

$$\Phi(v) = -\langle \xi, v \rangle + \iota_C(v),$$

where $\iota_C : H \rightarrow \mathbb{R} \cup \{+\infty\}$ denotes the indicator function of C . It is clear from its definition that Φ is proper, convex, and bounded from below. Furthermore, it can be seen that

- (i) Φ is not lower semicontinuous;
- (ii) For any $u \in H$ and $t > 0$,

$$\text{prox}_{t\Phi}(u) = \begin{cases} \{u + t^{q-1}\|\xi\|^{q-2}\xi\} & \text{if } u \in C - t^{q-1}\|\xi\|^{q-2}\xi, \\ \emptyset & \text{otherwise.} \end{cases}$$

While the functional is not lower semicontinuous, and the classical Brøndsted–Rockafellar theorem cannot be applied, one can still apply Theorem 1.1 for specific choices of parameters.

1.4.2. Lack of completeness

Let $\Omega \subseteq \mathbb{R}^d$ be a convex bounded open set. Consider the set

$$\mathcal{P}_2(\Omega) := \{\mu \in \mathcal{P}_2(\mathbb{R}^d) : \mu(\Omega) = 1\}$$

Preliminaries on the Wasserstein space are given in Subsection 3.1. The tuple $(\mathcal{P}_2(\Omega), W_2)$, where W_2 is the 2-Wasserstein distance, forms a non-complete geodesic metric space. Consider the entropy functional $\Phi : \mathcal{P}_2(\Omega) \rightarrow \mathbb{R} \cup \{+\infty\}$ given by

$$\Phi(\nu) := \begin{cases} \int_{\Omega} \rho(x) \log \rho(x) \, dx & \text{if } \nu = \rho \, dx \\ +\infty & \text{otherwise.} \end{cases}$$

It is clear from definition that Φ is proper, geodesically convex, and bounded from below. It can further be seen from standard arguments, *e.g.*, [31], Lemma 3.3.2, that

- (i) Φ is lower semicontinuous;
- (ii) For any $\mu \in \mathcal{P}_2(\Omega)$ and $t > 0$, $\text{prox}_{t\Phi}(\mu) \neq \emptyset$.

From this example we see that Theorem 1.1 can replace the usual variational principle on a non-complete geodesic space for certain functionals.

1.4.3. Non-existence of proximal points

Let $(c_{00}, \|\cdot\|_{\ell^2})$ be the space of finitely supported real sequences equipped with the ℓ^2 -norm. Consider the functional $\Phi : c_{00} \rightarrow \mathbb{R} \cup \{+\infty\}$ given by

$$\Phi(u) = \sum_{n=1}^{\infty} \frac{n^2}{2} (u_n - n^{-2})^2.$$

It is clear that Φ is proper, convex, and bounded from below. It was proved in [32], Theorem 2.1-(d) that it is nowhere sub-differentiable. Using this result, one can verify that

- (i) Φ is lower semicontinuous;
- (ii) For every $u \in c_{00}$ and $t > 0$, $\text{prox}_t \Phi(u) = \emptyset$.

Item (i) follows from [32], Theorem 2.1-(d), while item (ii) follows from the emptiness of the subdifferential at all points and the sum rule for subdifferentials [33], Theorem 9.5.4.

This example shows that certain pathologies of non-complete spaces cannot be avoided; here Theorem 1.1 becomes vacuous as there are no proximal points.

2. PROOFS OF SUBSECTIONS 1.2 AND 1.3

2.1. Proof of Theorem 1.1

Let $\hat{u} \in X$ and $\lambda > 0$ be such that

$$\hat{u} \in \arg \min_{v \in X} \left\{ \Phi(v) + \frac{\varepsilon}{p\lambda^p} d(v, u)^p \right\}.$$

Let us begin proving item (a); we proceed by contradiction. Suppose that $d(u, \hat{u}) > \lambda$. Since X is a geodesic metric space, there exists a curve $\gamma: [0, 1] \rightarrow X$ with $\gamma(0) = u$ and $\gamma(1) = \hat{u}$ such that $d(\gamma(s), u) = s d(u, \hat{u})$ for all $s \in [0, 1]$. Define

$$t := \frac{\lambda}{d(u, \hat{u})}.$$

Since we assumed $d(u, \hat{u}) > \lambda$, it must be that $t \in (0, 1)$. The geodesic convexity of Φ gives

$$\Phi(\gamma(t)) \leq (1 - t)\Phi(u) + t\Phi(\hat{u}). \quad (2.1)$$

By optimality of $\hat{u}$, we have

$$\Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} d(\hat{u}, u)^p \leq \Phi(\gamma(t)) + \frac{\varepsilon}{p\lambda^p} d(\gamma(t), u)^p.$$

Combining this with (2.1) gives

$$\Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} d(\hat{u}, u)^p \leq (1 - t)\Phi(u) + t\Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} d(\gamma(t), u)^p. \quad (2.2)$$

As γ is a constant-speed geodesic, we have that $d(\gamma(t), u) = td(\hat{u}, u) = \lambda$, and hence

$$\Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} d(\hat{u}, u)^p \leq (1 - t)\Phi(u) + t\Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} t^p d(u, \hat{u})^p$$

Using that $\Phi(u) \leq \Phi(\hat{u}) + \varepsilon$ and simplifying, we get

$$\begin{aligned} \Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} d(\hat{u}, u)^p &\leq (1 - t)(\Phi(\hat{u}) + \varepsilon) + t\Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} t^p d(u, \hat{u})^p \\ &= \Phi(\hat{u}) + \varepsilon \left(1 - \frac{\lambda}{d(u, \hat{u})} \right) + \frac{\varepsilon}{p}. \end{aligned}$$

Rearranging and simplifying, we obtain

$$\frac{1}{p\lambda^p}d(\hat{u}, u)^p \leq \frac{d(u, \hat{u}) - \lambda}{d(u, \hat{u})} + \frac{1}{p}.$$

Multiplying both sides by λ^p and rearranging,

$$\frac{\lambda - d(u, \hat{u})}{d(u, \hat{u})}\lambda^p \leq \frac{\lambda^p}{p} - \frac{1}{p}d(\hat{u}, u)^p \quad (2.3)$$

Define $\omega : [0, +\infty) \rightarrow [0, +\infty)$ as $\omega(s) = s^p/p$. By convexity of ω , it follows from (2.3) that

$$\frac{\lambda - d(u, \hat{u})}{d(u, \hat{u})}\lambda^p \leq \omega(\lambda) - \omega(d(u, \hat{u})) \leq \omega'(\lambda)(\lambda - d(u, \hat{u})) = \lambda^{p-1}(\lambda - d(u, \hat{u})).$$

Dividing both sides by $\lambda^{p-1}(\lambda - d(u, \hat{u})) < 0$, we obtain $\lambda/d(u, \hat{u}) \geq 1$. A contradiction.

Let us now proceed with item (b). By optimality of $\hat{u}$,

$$\Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p}d(\hat{u}, u)^p \leq \Phi(v) + \frac{\varepsilon}{p\lambda^p}d(v, u)^p \quad \forall v \in X. \quad (2.4)$$

In particular, taking $v = u$ gives

$$\Phi(\hat{u}) \leq \Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p}d(\hat{u}, u)^p \leq \Phi(u).$$

Finally, we prove item (c). Fix $v \in X \setminus \{\hat{u}\}$ and let $\gamma : [0, 1] \rightarrow X$ be a constant-speed geodesic with $\gamma(0) = \hat{u}$ and $\gamma(1) = v$. For each $t \in [0, 1]$, set $v_t := \gamma(t)$. Then,

$$\Phi(v_t) \leq (1-t)\Phi(\hat{u}) + t\Phi(v) \quad \text{and} \quad d(v_t, u) \leq d(\hat{u}, u) + td(v, \hat{u}) \quad \forall t \in (0, 1]. \quad (2.5)$$

Plugging v_t into (2.4), using (2.5), and simplifying yields

$$t(\Phi(\hat{u}) - \Phi(v)) \leq \frac{\varepsilon}{\lambda^p} \left(\frac{d(v_t, u)^p}{p} - \frac{d(\hat{u}, u)^p}{p} \right).$$

As in the proof of item (a), let $\omega : [0, +\infty) \rightarrow [0, +\infty)$ be given by $\omega(s) = s^p/p$. By convexity of ω , we see that, for all $t \in (0, 1]$,

$$\begin{aligned} t(\Phi(\hat{u}) - \Phi(v)) &\leq \frac{\varepsilon}{\lambda^p} (\omega(d(v_t, u)) - \omega(d(\hat{u}, u))) \\ &\leq \frac{\varepsilon}{\lambda^p} \omega'(d(v_t, u))(d(v_t, u) - d(\hat{u}, u)) \\ &= \frac{\varepsilon}{\lambda^p} d(v_t, u)^{p-1} (d(v_t, u) - d(\hat{u}, u)). \end{aligned}$$

Using (2.5), we get that for all $t \in (0, 1]$,

$$t(\Phi(\hat{u}) - \Phi(v)) \leq \frac{\varepsilon}{\lambda^p} d(v_t, u)^{p-1} td(v, \hat{u}).$$

This leads to $\Phi(\hat{u}) - \Phi(v) \leq \varepsilon/\lambda^p d(v_t, u)^{p-1} d(v, \hat{u})$ for all $t \in (0, 1]$. Letting $t \rightarrow 0^+$,

$$\Phi(\hat{u}) - \Phi(v) \leq \frac{\varepsilon}{\lambda^p} d(\hat{u}, u)^{p-1} d(v, \hat{u}). \quad (2.6)$$

Combining this with the estimate in item (a) yields

$$\Phi(\hat{u}) \leq \Phi(v) + \frac{\varepsilon}{\lambda^p} d(\hat{u}, u)^{p-1} d(v, \hat{u}) \leq \Phi(v) + \frac{\varepsilon}{\lambda} d(\hat{u}, v).$$

This finishes the proof, as $v \in X \setminus \{\hat{u}\}$ was arbitrary. $\square$

2.2. Proof of Theorem 1.7

From Theorem 1.1 we obtain items (a) and (c) directly. Proceeding as in the proof of Theorem 1.1, one arrives at inequality (2.6) for any $v \in X \setminus \{\hat{u}\}$. From this, the definition of metric slope and (2.6), we obtain

$$|\partial\Phi|(\hat{u}) \leq \frac{\varepsilon}{\lambda^p} d(\hat{u}, u)^{p-1}, \quad (2.7)$$

then $\frac{\lambda^p}{\varepsilon} |\partial\Phi|(\hat{u}) \leq d(\hat{u}, u)^{p-1}$. Raising both sides to the power $q = \frac{p}{p-1}$ yields

$$\left(\frac{\lambda^p}{\varepsilon} |\partial\Phi|(\hat{u}) \right)^q \leq d(\hat{u}, u)^p. \quad (2.8)$$

Using inequality (2.8) into

$$\Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} d(\hat{u}, u)^p \leq \Phi(u),$$

it follows

$$\Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} \left(\frac{\lambda^p}{\varepsilon} |\partial\Phi|(\hat{u}) \right)^q \leq \Phi(\hat{u}) + \frac{\varepsilon}{p\lambda^p} d(\hat{u}, u)^p \leq \Phi(u).$$

Thus, we obtain

$$\Phi(\hat{u}) + \frac{\lambda^q}{p\varepsilon^{\frac{q}{p}}} |\partial\Phi|(\hat{u})^q \leq \Phi(u).$$

This completes the proof. $\square$

2.3. Proof of Proposition 1.10

Let $u \in \text{dom } \Phi$ and $t > 0$ such that $\Phi(u) < \inf_{v \in X} \Phi(v) + t^{\frac{q}{p}}$. Let

$$\hat{u} \in \arg \min_{v \in X} \left\{ \Phi(v) + \frac{1}{pt} d(v, u)^p \right\}.$$

If $\Phi(u) = \inf_{v \in X} \Phi(v)$, then $\hat{u} \in \arg \min_{v \in X} \Phi(v)$, and hence by (1.4), $d(\hat{u}, T(\hat{u})) \leq \Phi(\hat{u}) - \Phi(T(\hat{u})) \leq 0$; so $\hat{u}$ is a fixed point. Suppose now that $\Phi(u) - \inf_{v \in X} \Phi(v) > 0$. Define $\varepsilon := \Phi(u) - \inf_{v \in X} \Phi(v)$. Set $\lambda = \sqrt[p]{t\varepsilon}$ and

observe that $(pt)^{-1} = \varepsilon(p\lambda^p)^{-1}$. We see that the conditions in Theorem 1.1 are satisfied. By Theorem 1.1, we have

$$\Phi(\hat{u}) \leq \Phi(v) + \frac{\varepsilon}{\lambda} d(v, \hat{u}) \quad \forall v \in X.$$

Combining this with (1.4), we conclude that

$$d(\hat{u}, T(\hat{u})) + \Phi(T(\hat{u})) \leq \Phi(\hat{u}) \leq \Phi(T(\hat{u})) + \frac{\varepsilon}{\lambda} d(T(\hat{u}), \hat{u}).$$

This can be rewritten as

$$(1 - \frac{\varepsilon}{\lambda}) d(\hat{u}, T(\hat{u})) \leq 0.$$

Since $\varepsilon/\lambda = \varepsilon^{1-\frac{1}{p}}/t^{\frac{1}{p}} = (\varepsilon^{\frac{p}{q}}/t)^{\frac{1}{p}} < 1$, the result follows.

2.4. Proof of Proposition 1.11

Set $\Lambda := \arg \min_{v \in X} \Phi(v)$, assume that (1.5) holds and fix $t > 0$. We will prove that for any $\delta \in (0, 1)$ there holds

$$\Phi(u) - \inf_{v \in X} \Phi(v) \geq \frac{t^{q/p}(1-\delta)^p}{(t^{q/p} + \kappa)^p} d(u, \Lambda)^p \quad \forall u \in \text{dom prox}_{t\Phi}. \quad (2.9)$$

The result will follow by letting $\delta \rightarrow 0^+$. We proceed by contradiction; suppose there exist $\delta > 0$ and $u \in \text{dom prox}_{t\Phi}$ such that (2.9) does not hold. Then,

$$\Phi(u) - \inf_{v \in X} \Phi(v) < \frac{t^{q/p}(1-\delta)^p}{(t^{q/p} + \kappa)^p} d(u, \Lambda)^p.$$

Since $u \in \text{dom prox}_{t\Phi}$, there exists $\hat{u} \in X$ such that

$$\hat{u} \in \arg \min_{v \in X} \left\{ \Phi(v) + \frac{1}{pt} d(v, u)^p \right\}.$$

Consider the positive numbers

$$\lambda := \frac{1-\delta}{1 + \kappa t^{-q/p}} d(u, \Lambda) \quad \text{and} \quad \varepsilon := \lambda^p/t.$$

Observe that by construction,

$$\frac{\varepsilon}{p\lambda^p} = \frac{1}{pt} \quad \text{and} \quad \frac{\varepsilon}{\lambda} = \frac{\lambda^{p-1}}{t}.$$

Since $\Phi(u) \leq \inf_{v \in X} \Phi(v) + \varepsilon$, Theorem 1.7 yields that $d(u, \hat{u}) \leq \lambda$ and

$$|\partial\Phi|(\hat{u}) \leq \frac{\lambda^{p-1}}{t}.$$

This and condition (1.5) yield

$$d(\hat{u}, \Lambda) \leq \kappa |\partial\Phi|(\hat{u})^{q/p} \leq \kappa \left(\frac{\lambda^{p-1}}{t} \right)^{q/p} = \lambda \kappa t^{-q/p}.$$

Combining the previous inequality with the estimate $d(u, \hat{u}) \leq \lambda$, we obtain

$$\begin{aligned} d(u, \Lambda) &\leq d(u, \hat{u}) + d(\hat{u}, \Lambda) \leq \lambda + \lambda \kappa t^{-q/p} \\ &= \lambda(1 + \kappa t^{-q/p}) = \frac{1 - \delta}{1 + \kappa t^{-q/p}} d(u, \Lambda)(1 + \kappa t^{-q/p}) \\ &= (1 - \delta) d(u, \Lambda) < d(u, \Lambda). \end{aligned}$$

This is the sought contradiction. □

3. THE THEOREM IN THE SPACE OF PROBABILITY MEASURES

3.1. Preliminaries and notation

In this subsection, we recall basic notions from the theory of optimal transport and the structural properties of the Wasserstein space, following [18], Chapters 5 and 7.

3.1.1. The space of probability measures

The space of Borel probability measures $\mu : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ is denoted by $\mathcal{P}(\mathbb{R}^d)$. For a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ and a Borel mapping $\mathfrak{t} : \mathbb{R}^d \rightarrow \mathbb{R}^d$, the *pushforward* of μ through $\mathfrak{t}$ is the probability measure $\mathfrak{t}_\# \mu \in \mathcal{P}(\mathbb{R}^d)$ defined by

$$\mathfrak{t}_\# \mu(B) := \mu(\mathfrak{t}^{-1}(B)) \quad \forall B \in \mathcal{B}(\mathbb{R}^d).$$

A probability measure $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is said to be a *coupling* between $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\nu \in \mathcal{P}(\mathbb{R}^d)$ if

$$\text{proj}_1 \# \gamma = \mu \quad \text{and} \quad \text{proj}_2 \# \gamma = \nu,$$

where $\text{proj}_1, \text{proj}_2 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ denote the canonical projections. The set of all couplings between μ and ν is denoted by $\Gamma(\mu, \nu)$. A sequence $(\mu_n) \subset \mathcal{P}(\mathbb{R}^d)$ is said to converge *narrowly* to $\mu \in \mathcal{P}(\mathbb{R}^d)$ if

$$\int_{\mathbb{R}^d} \varphi \, d\mu_n \longrightarrow \int_{\mathbb{R}^d} \varphi \, d\mu \quad \forall \varphi \in C_b(\mathbb{R}^d),$$

where $C_b(\mathbb{R}^d)$ denotes the space of bounded continuous functions $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$.

3.1.2. The Wasserstein space distance

Given $\mu \in \mathcal{P}(\mathbb{R}^d)$, $L^2(\mu)$ denotes the usual space of square-integrable Borel functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$. We denote

$$\mathcal{P}_2(\mathbb{R}^d) := \{\mu \in \mathcal{P}(\mathbb{R}^d) : \text{id} \in L^2(\mu)^d\}.$$

This is precisely the space of Borel probability measures with finite second moment. The *Wasserstein distance* between $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ is defined by

$$W_2(\mu, \nu) := \inf_{\gamma \in \Gamma(\mu, \nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y) \right)^{1/2}.$$

The tuple $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ forms a complete and separable metric space.

3.1.3. Geodesics in the Wasserstein space

Given $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, we consider the set

$$\Gamma_o(\mu, \nu) := \left\{ \gamma \in \Gamma(\mu, \nu) : W_2(\mu, \nu)^2 = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y) \right\}.$$

Given $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, the set $\Gamma_o(\mu, \nu)$ is always nonempty. We say that a Borel mapping $\mathfrak{t} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is an *optimal transport* between $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ if $(\text{id}, \mathfrak{t})_{\#}\mu$ belongs to $\Gamma_o(\mu, \nu)$.

A curve $\gamma : [0, 1] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ is said to be a *Wasserstein geodesic* if there exists $\pi \in \Gamma_o(\mu, \nu)$ such that

$$\gamma(t) = ((1 - t) \text{proj}_1 + t \text{proj}_2)_{\#} \pi \quad \forall t \in [0, 1].$$

It can be seen that Wasserstein geodesics are precisely the constant-speed geodesics in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, and that this space is geodesic.

3.1.4. The regular subdifferential

Let $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper geodesically convex functional such that

$$\text{dom } \Phi \subseteq \{ \mu \in \mathcal{P}_2(\mathbb{R}^d) : \mu = \rho dx \text{ for some } \rho \in L^1(\mathbb{R}^d) \}. \quad (3.1)$$

Given $\mu, \nu \in \text{dom } \Phi$ we will denote by $\mathfrak{t}_{\mu}^{\nu} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ the unique optimal transport between μ and ν ; the existence and uniqueness is guaranteed by the Brenier–McCann theorem. The subdifferential $\partial\Phi(\mu)$ of Φ at a probability measure $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ consists of all $\xi \in L^2(\mu)^d$ such that

$$\Phi(\nu) \geq \Phi(\mu) + \int_{\mathbb{R}^d} \langle \xi(x), \mathfrak{t}_{\mu}^{\nu}(x) - x \rangle d\mu(x) \quad \forall \nu \in \mathcal{P}_2(\mathbb{R}^d).$$

3.2. The Brøndsted–Rockafellar theorem in the Wasserstein space

Let $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper geodesically convex functional bounded from below such that (3.1) holds. Given $\tau > 0$, we consider the set-valued mapping $J_{\tau}^{\Phi} : \mathcal{P}_2(\mathbb{R}^d) \rightrightarrows \mathcal{P}_2(\mathbb{R}^d)$ given by

$$J_{\tau}^{\Phi}(\mu) := \arg \min_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \Phi(\nu) + \frac{1}{2\tau} W_2(\nu, \mu)^2 \right\}.$$

This is precisely the proximal mapping (for $p = 2$) introduced in Subsection 1.2; it also corresponds to a single implicit–Euler step of the minimizing movement scheme for gradient flows, often referred to as the Jordan–Kinderlehrer–Otto (JKO) scheme (see [34], Prop. 4.1 and [18], Def. 2.0.6).

We now formulate the metric analogue of the Brøndsted–Rockafellar theorem in the Wasserstein framework, and give a couple of remarks before the proof.

Theorem 3.1. *Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\varepsilon > 0$ such that*

$$\Phi(\mu) \leq \inf_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Phi(\nu) + \varepsilon.$$

Let $\lambda > 0$ and suppose there exist $\hat{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$ and $\xi \in L^2(\hat{\mu})^d$ such that

$$\hat{\mu} \in J_{\frac{\lambda^2}{\varepsilon}}^{\Phi}(\mu) \quad \text{and} \quad \mathbf{t}_{\hat{\mu}}^{\mu} = \text{id} + \frac{\lambda^2}{\varepsilon} \xi.$$

Then the following statements are satisfied.

$$(a) \ W_2(\mu, \hat{\mu}) \leq \lambda \quad (b) \ \Phi(\hat{\mu}) + \frac{\lambda^2}{2\varepsilon} \|\xi\|_{L^2(\hat{\mu})^d}^2 \leq \Phi(\mu) \quad (c) \ \xi \in \partial\Phi(\hat{\mu}) \quad (d) \ \|\xi\|_{L^2(\hat{\mu})^d} \leq \frac{\varepsilon}{\lambda}.$$

Proof. Item (a) follows directly from Theorem 1.1-(a). Item (c) follows from the first-order optimality condition for the proximal problem

$$\frac{\varepsilon}{\lambda^2} (\mathbf{t}_{\hat{\mu}}^{\mu} - \text{id}) \in \partial\Phi(\hat{\mu}).$$

Using that $\mathbf{t}_{\hat{\mu}}^{\mu} = \text{id} + \frac{\lambda^2}{\varepsilon} \xi$, we obtain $\xi \in \partial\Phi(\hat{\mu})$. Observe that

$$\|\xi\|_{L^2(\hat{\mu})^d} = \frac{\varepsilon}{\lambda^2} \|\mathbf{t}_{\hat{\mu}}^{\mu} - \text{id}\|_{L^2(\hat{\mu})^d} = \frac{\varepsilon}{\lambda^2} W_2(\hat{\mu}, \mu). \quad (3.2)$$

Since $W_2(\hat{\mu}, \mu) \leq \lambda$, equality (3.2) yields immediately item (d). Finally, item (b) can be obtained by repeating the argument in the proof of Theorem 1.7 replacing (2.7) with (3.2). $\square$

Remark 3.2. Interpreting Φ as a potential energy functional on the Wasserstein space and $\xi \in L^2(\nu)^d$ as the velocity field acting on the particles of a probability measure $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, inequality in Theorem 3.1-(b) expresses an energy balance. The quantity

$$\mathcal{E}(\nu) := \Phi(\nu) + \frac{\lambda^2}{2\varepsilon} \|\xi\|_{L^2(\nu)^d}^2$$

can be viewed as a total energy, consisting of a potential part and a kinetic term associated with the transport cost. Inequality in Theorem 3.1-(b) therefore states that a single JKO step does not increase the total energy, reflecting the dissipation principle of gradient flows.

3.3. Entropy-transportation and metric sub-regularity

Let $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper geodesically convex functional such that (3.1) holds.

Given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $C \subseteq \mathcal{P}_2(\mathbb{R}^d)$, we denote

$$W_2(\mu, C) := \inf_{\nu \in C} W_2(\mu, \nu) \quad \text{and} \quad d_{L^2}(0, \partial\Phi(\mu)) := \inf_{\xi \in \partial\Phi(\mu)} \|\xi\|_{L^2(\mu)^d}.$$

The first quantity can be interpreted as a gauge of how close a given probability measure is to belonging to a specified set, while the second—being the minimal L^2 -norm among all elements of the subdifferential of Φ at μ —serves as a gauge of how close μ is to having zero as a subgradient, and hence to being a minimizer of Φ .

As a consequence of Theorem 3.1, we obtain an equivalence between the metric sub-regularity of the subdifferential and a quantitative growth condition of entropy–transport type.

Theorem 3.3. *Assume that $\arg \min_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Phi(\nu) \neq \emptyset$ and that there exists $\tau > 0$ such that $J_\tau^\Phi(\mu) \neq \emptyset$ for every $\mu \in \text{dom } \Phi$. Then, the following statements are equivalent.*

(i) *There exists $\kappa > 0$ such that*

$$W_2(\mu, \arg \min_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Phi(\nu)) \leq \kappa d_{L^2}(0, \partial\Phi(\mu)) \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d). \quad (3.3)$$

(ii) *There exists $c > 0$ such that*

$$\Phi(\mu) \geq \inf_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Phi(\nu) + c W_2(\mu, \arg \min_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Phi(\nu))^2 \quad \forall \mu \in \mathcal{P}_2(\mathbb{R}^d). \quad (3.4)$$

Proof. Set $\Lambda := \arg \min_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Phi(\nu)$.

(ii) $\implies$ (i). Fix $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. If $\partial\Phi(\mu) = \emptyset$, then $d_{L^2}(0, \partial\Phi(\mu)) = +\infty$ and the inequality is trivial. Assume henceforth $\partial\Phi(\mu) \neq \emptyset$. Fix $\xi \in \partial\Phi(\mu)$. By the definition of the subdifferential we have

$$\Phi(\nu) \geq \Phi(\mu) + \int_{\mathbb{R}^d} \langle \xi(x), \mathbf{t}_\mu^\nu(x) - x \rangle d\mu(x) \quad \forall \nu \in \mathcal{P}_2(\mathbb{R}^d).$$

From this,

$$\Phi(\mu) - \Phi(\nu) \leq \int_{\mathbb{R}^d} \langle \xi(x), x - \mathbf{t}_\mu^\nu(x) \rangle d\mu(x) \leq \|\xi\|_{L^2(\mu)^d} W_2(\mu, \nu).$$

Taking the infimum over $\nu \in \Lambda$, we obtain

$$\Phi(\mu) - \inf_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Phi(\nu) \leq \|\xi\|_{L^2(\mu)^d} W_2(\mu, \Lambda).$$

Combining this with (3.4) gives $c W_2(\mu, \Lambda) \leq \|\xi\|_{L^2(\mu)^d}$. Taking the infimum over $\xi \in \partial\Phi(\mu)$ gives the desired inequality with $\kappa = 1/c$.

(i) $\implies$ (ii). Fix $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\mu^\tau \in J_\tau^\Phi(\mu)$. By Theorem 3.1, there exists $\xi \in \partial\Phi(\mu^\tau)$ such that

$$\mathbf{t}_{\mu^\tau}^\mu = \text{id} + \tau \xi, \quad \|\xi\|_{L^2(\mu^\tau)^d} = \frac{1}{\tau} W_2(\mu, \mu^\tau), \quad \Phi(\mu) - \Phi(\mu^\tau) \geq \frac{1}{2\tau} W_2(\mu, \mu^\tau)^2. \quad (3.5)$$

Using (3.3) and (3.5), we get

$$W_2(\mu^\tau, \Lambda) \leq \kappa d_{L^2}(0, \partial\Phi(\mu^\tau)) \leq \kappa \|\xi\|_{L^2(\mu^\tau)^d} = \frac{\kappa}{\tau} W_2(\mu, \mu^\tau).$$

By the triangle inequality,

$$W_2(\mu, \Lambda) \leq W_2(\mu, \mu^\tau) + W_2(\mu^\tau, \Lambda) \leq \left(1 + \frac{\kappa}{\tau}\right) W_2(\mu, \mu^\tau).$$

We deduce from (3.5) that

$$\Phi(\mu) - \inf_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Phi(\nu) \geq \frac{1}{2\tau} W_2(\mu, \mu^\tau)^2 \geq \frac{\tau}{2(\tau + \kappa)^2} W_2(\mu, \Lambda)^2.$$

Hence (3.4) holds with $c = 2^{-1} \tau (\tau + \kappa)^{-2}$. $\square$

Remark 3.4. In Theorem 3.3 we assume uniform existence of minimizers for the JKO steps. This assumption is sometimes adopted in the literature to simplify the analysis and to avoid technical complications concerning the well-posedness of the minimizing movement scheme; see, for instance, [18], Assumption (10.1.1b).

Remark 3.5. There are several examples of functionals on Wasserstein spaces satisfying condition (3.4) (and consequently (3.3)), for instance, those given by the sum of internal and potential energies for specific parameter choices; we refer to [30], Section 3 and [29], Section 4.2 for a comprehensive account.

FUNDING

The first author was supported by the Austrian Science Foundation (FWF) under grants P 36344-N and F 100800, and the second author by the SECIHTI under Grant No. 921835.

REFERENCES

- [1] A. Brøndsted and R.T. Rockafellar, On the subdifferentiability of convex functions. *Proc. Am. Math. Soc.* **16** (1965) 605–611.
- [2] E. Bishop and R.R. Phelps, The support functionals of a convex set, in *Proc. Sympos. Pure Math.*, Vol. VII. American Mathematical Society, Providence, RI (1963) 27–35.
- [3] I. Ekeland, On the variational principle. *J. Math. Anal. Appl.* **47** (1974) 324–353.
- [4] G. Carlier, Fenchel-Young inequality with a remainder and applications to convex duality and optimal transport. *SIAM J. Optim.* **33** (2023) 1463–1472.
- [5] S. Armeniakos and A. Daniilidis, Characterizing maximal monotone operators with unique representation. [arXiv:2510.09368](https://arxiv.org/abs/2510.09368) (2025).
- [6] H.H. Bauschke, S. Singh and X. Wang, On Carlier’s inequality. *J. Convex Anal.* **30** (2023) 499–514.
- [7] R.S. Burachik and J.E. Martínez-Legaz, A note on carlier’s inequality. *Optimization* (2025) 1–6.
- [8] R.I. Boş and E.R. Csetnek, A Brøndsted–Rockafellar theorem for diagonal subdifferential operators, in *Computational and analytical mathematics*, vol. 50 of *Springer Proc. Math. Stat.*. Springer, New York (2013) 105–112.
- [9] A.N. Iusem and B.F. Svaiter, On diagonal subdifferential operators in nonreflexive Banach spaces. *Set-Valued Var. Anal.* **20** (2012) 1–14.
- [10] M. Marques Alves and B.F. Svaiter, Brøndsted–Rockafellar property and maximality of monotone operators representable by convex functions in non-reflexive Banach spaces. *J. Convex Anal.* **15** (2008) 693–706.
- [11] S. Simons, *From Hahn–Banach to monotonicity*, vol. 1693 of *Lecture Notes in Mathematics*, 2nd edn. Springer, New York (2008).
- [12] T. Amahroq and A. Oussarhan, An extension of Brøndsted–Rockafellar’s theorem with applications. *J. Math. Anal. Appl.* **531** (2024) Paper No. 127810, 10.
- [13] Z. Chbani and H. Riahi, Variational principles for monotone and maximal bifunctions. *Serdica Math. J.* **29** (2003) 159–166.
- [14] R. Correa, A. Hantoute and P. Pérez-Aros, On Brøndsted–Rockafellar’s Theorem for convex lower semicontinuous epi-pointed functions in locally convex spaces. *Math. Program.* **168** (2018) 631–643.
- [15] J. Falcó and D. Isert, Group invariant variational principles. *Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat.* **118** (2024) Paper No. 91, 20.
- [16] M. Lassonde, Brøndsted–Rockafellar property of subdifferentials of prox-bounded functions. *J. Convex Anal.* **22** (2015) 485–492.
- [17] X.Y. Zheng, Corrigendum: convex optimization problems on differentiable sets. *SIAM J. Optim.* **33** (2023) 2484–2488.
- [18] L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich, 2nd edn. Birkhäuser Verlag, Basel (2008).
- [19] L. Ambrosio and G. Savaré, Gradient flows of probability measures, in *Handbook of differential equations: evolutionary equations*, vol. III of *Handbook of Differential Equations*. Elsevier/North-Holland, Amsterdam (2007) 1–136.
- [20] A.Y. Kruger, Error bounds and metric subregularity. *Optimization* **64** (2015) 49–79.

- [21] J. Caristi, Fixed point theorems for mappings satisfying inwardness conditions. *Trans. Am. Math. Soc.* **215** (1976) 241–251.
- [22] R. Cibulka, A.L. Dontchev and A.Y. Kruger, Strong metric subregularity of mappings in variational analysis and optimization. *J. Math. Anal. Appl.* **457** (2018) 1247–1282.
- [23] A.L. Dontchev, H. Gfrerer, A.Y. Kruger and J.V. Outrata, The radius of metric subregularity. *Set-Valued Var. Anal.* **28** (2020) 451–473.
- [24] A.L. Dontchev and R.T. Rockafellar, *Implicit functions and solution mappings*. Springer Series in Operations Research and Financial Engineering, 2nd edn. Springer, New York (2014). A view from variational analysis.
- [25] N.A. Jork, N.P. Osmolovskii and V.M. Veliov, Strong metric (sub)regularity in optimal control. *J. Convex Anal.* **32** (2025) 375–398.
- [26] F.J. Aragón Artacho and M.H. Geoffroy, Characterization of metric regularity of subdifferentials. *J. Convex Anal.* **15** (2008) 365–380.
- [27] D. Azé and J.-N. Corvellec, Nonlinear local error bounds via a change of metric. *J. Fixed Point Theory Appl.* **16** (2014) 351–372.
- [28] D. Azé and J.-N. Corvellec, Nonlinear error bounds via a change of function. *J. Optim. Theory Appl.* **172** (2017) 9–32.
- [29] D. Hauer and J.M. Mazón, Kurdyka–Łojasiewicz–Simon inequality for gradient flows in metric spaces. *Trans. Am. Math. Soc.* **372** (2019) 4917–4976.
- [30] A. Blanchet and J. Bolte, A family of functional inequalities: Łojasiewicz inequalities and displacement convex functions. *J. Funct. Anal.* **275** (2018) 1650–1673.
- [31] A. Figalli and F. Glaudo, *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows*. EMS Textbooks in Mathematics. EMS Press, Berlin (2021).
- [32] G. Wachsmuth, A convex, finite and lower semicontinuous function with empty subdifferential. *J. Convex Anal.* **32** (2025) 877–882.
- [33] H. Attouch, G. Buttazzo and G. Michaille, *Variational analysis in Sobolev and BV spaces*, volume 17 of *MOS-SIAM Series on Optimization*, 2nd edn. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA (2014). Applications to PDEs and optimization.
- [34] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **29** (1998) 1–17.
- [35] M. Bačák and U. Kohlenbach, On proximal mappings with Young functions in uniformly convex Banach spaces. *J. Convex Anal.* **25** (2018) 1291–1318.



Please help to maintain this journal in open access!

This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

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

More information, including a list of supporters and financial transparency reports, is available at <https://edpsciences.org/en/subscribe-to-open-s2o>.

APPENDIX A. THE PARTICULAR CASE OF NORMED SPACES

In this subsection of the appendix, we specialize the metric constructions of Section 1.2 to the classical setting of a real normed space. In this framework, the optimality condition for a proximal step admits a concrete expression in dual variables, and it becomes transparent how the choice of a proximal point produces an approximate minimizer together with a subgradient of controlled size.

The following theorem can be regarded as an extension of [4], Theorem 4.1 from Hilbert spaces to general normed spaces. Its proof is a direct consequence of Theorem 1.1.

Theorem A.1. *Let $(X, \|\cdot\|)$ be a real normed space, and $\Phi : X \rightarrow \mathbb{R} \cup \{+\infty\}$ a proper convex functional. Fix $p, q \in (1, +\infty)$ such that $p^{-1} + q^{-1} = 1$. Let $u \in X$ and $\varepsilon > 0$ be such that*

$$\Phi(u) \leq \inf_{v \in X} \Phi(v) + \varepsilon.$$

Let $\lambda > 0$ be given. Suppose that $\hat{u} \in X$ and $\xi \in \partial\Phi(\hat{u})$ satisfy

$$\langle \xi, u - \hat{u} \rangle = \|\xi\| \|u - \hat{u}\| \quad \text{and} \quad \|\xi\| = \frac{\varepsilon}{\lambda^p} \|u - \hat{u}\|^{p-1} \quad (\text{A.1})$$

Then, the following estimates are satisfied.

$$(a) \quad \|u - \hat{u}\| \leq \lambda \quad (b) \quad \Phi(\hat{u}) + \frac{\lambda^q}{p\varepsilon^{\frac{q}{p}}} \|\xi\|^q \leq \Phi(u) \quad (c) \quad \|\xi\| \leq \frac{\varepsilon}{\lambda}.$$

Proof. Let us begin with item (a). Condition (A.1) is equivalent to

$$0 \in \partial\Phi(\hat{u}) + \frac{\varepsilon}{\lambda^p} J_p(\hat{u} - u),$$

where $J_p : X \rightrightarrows X^*$ denotes the duality mapping. This in turn implies $\hat{u} \in \text{prox}_{\frac{\lambda^p}{\varepsilon}\Phi}(u)$. Item (a) then follows from Theorem 1.1. Using the identities $\|\xi\| = \frac{\varepsilon}{\lambda^p} \|u - \hat{u}\|^{p-1}$ and $(p-1)q = p$, we obtain

$$\frac{\varepsilon}{p\lambda^p} \|u - \hat{u}\|^p = \frac{\lambda^q}{p\varepsilon^{\frac{q}{p}}} \|\xi\|^q.$$

Combining this with the fact that $\hat{u} \in \text{prox}_{\frac{\lambda^p}{\varepsilon}\Phi}(u)$ yields then item (b). Finally,

$$\|\xi\| = \frac{\varepsilon}{\lambda^p} \|u - \hat{u}\|^{p-1} \leq \frac{\varepsilon}{\lambda^p} \lambda^{p-1} = \frac{\varepsilon}{\lambda},$$

whence item (c) follows. $\square$

Remark A.2. In the Hilbertian case with the choice $p = 2$, the duality mapping coincides with the identity under the canonical identification of X with its dual. In this situation, condition (A.1) reduces to the construction in [4], Theorem 4.1.

Remark A.3. Item (b) strengthens the classical estimate $\Phi(\hat{u}) \leq \Phi(u)$ by giving an explicit lower bound on the decrease $\Phi(u) - \Phi(\hat{u})$ in terms of the size of a subgradient at $\hat{u}$. In particular, if $\hat{u}$ is not a critical point, then the proximal construction yields a strict decrease, that is, $\Phi(\hat{u}) < \Phi(u)$.

APPENDIX B. THE LIMITING CASE $p = +\infty$

This part of the appendix records the limiting regime in which the proximal regularization becomes a hard constraint. As the exponent in the powered-distance kernel tends to infinity, the penalized problems enforce confinement to a fixed ball; in the sense of Γ -convergence the kernels converge to the indicator function of a ball. Accordingly, the approximate minimizer produced by the proximal step is replaced by a minimizer under a ball constraint, and the same Ekeland-type estimate is recovered.

Theorem B.1. *Let (X, d) be a geodesic space, and $\Phi : X \rightarrow \mathbb{R} \cup \{+\infty\}$ a proper geodesically convex functional. Let $u \in X$ and $\varepsilon > 0$ be such that*

$$\Phi(u) \leq \inf_{v \in X} \Phi(v) + \varepsilon.$$

Assume that there exist $\hat{u} \in X$ and $\lambda > 0$ such that

$$\hat{u} \in \arg \min_{v \in X} \{ \Phi(v) + \iota_{\mathbb{B}(u, \lambda)}(v) \}.$$

Then the following estimates are satisfied.

$$(a) \ d(u, \hat{u}) \leq \lambda \quad (b) \ \Phi(\hat{u}) \leq \Phi(u) \quad (c) \ \Phi(\hat{u}) \leq \Phi(v) + \frac{\varepsilon}{\lambda} d(v, \hat{u}) \quad \forall v \in X.$$

Proof. Items (a) and (b) follow trivially. Let $v \in X$; the computation is divided in two cases. If $d(u, v) \leq \lambda$, then $v \in \mathbb{B}(u, \lambda)$ and by minimality of $\hat{u}$ on $\mathbb{B}(u, \lambda)$ we have $\Phi(\hat{u}) \leq \Phi(v)$, hence

$$\Phi(\hat{u}) \leq \Phi(v) \leq \Phi(v) + \frac{\varepsilon}{\lambda} d(v, \hat{u}).$$

If $d(u, v) > \lambda$, let $\gamma : [0, 1] \rightarrow X$ be a constant-speed geodesic from u to v , and set $t := \lambda/d(u, v)$. Then $\gamma(t) \in \mathbb{B}(u, \lambda)$, so $\Phi(\hat{u}) \leq \Phi(\gamma(t))$. By geodesic convexity,

$$\Phi(\hat{u}) \leq \Phi(\gamma(t)) \leq (1-t)\Phi(u) + t\Phi(v) \leq (1-t)(\Phi(v) + \varepsilon) + t\Phi(v) = \Phi(v) + (1-t)\varepsilon.$$

Since $1-t = (d(u, v) - \lambda)/d(u, v) \leq (d(u, v) - \lambda)/\lambda$, it follows that

$$\Phi(\hat{u}) \leq \Phi(v) + \frac{\varepsilon}{\lambda} (d(u, v) - \lambda).$$

Finally, $d(u, v) \leq d(u, \hat{u}) + d(\hat{u}, v) \leq \lambda + d(\hat{u}, v)$ by (a), hence $d(u, v) - \lambda \leq d(v, \hat{u})$, and therefore

$$\Phi(\hat{u}) \leq \Phi(v) + \frac{\varepsilon}{\lambda} d(v, \hat{u}).$$

This proves item (c) and completes the proof. $\square$

APPENDIX C. THE CASE OF GENERAL PROXIMAL KERNELS

In this subsection of the appendix, we present a more general version of Theorem 1.1, in which powered-distance perturbations are replaced by more general ones.

Let $\omega : [0, +\infty) \rightarrow [0, +\infty)$ be a convex function satisfying

$$\omega(0) = 0 \quad \text{and} \quad \omega(s) > 0 \quad \forall s > 0.$$

These properties imply the existence of the right-derivative $\omega'_+ : [0, +\infty) \rightarrow [0, +\infty)$ and furthermore that $\omega'_+(s) > 0$ for all $s > 0$. We are now ready to state the general theorem.

Theorem C.1. *Let (X, d) be a geodesic space, and $\Phi : X \rightarrow \mathbb{R} \cup \{+\infty\}$ a proper geodesically convex functional. Let $u \in X$ and $\varepsilon > 0$ be such that*

$$\Phi(u) \leq \inf_{v \in X} \Phi(v) + \varepsilon.$$

Assume that there exist $\hat{u} \in X$ and $\lambda > 0$ such that

$$\hat{u} \in \arg \min_{v \in X} \left\{ \Phi(v) + \frac{\varepsilon}{\lambda \omega'_+(\lambda)} \omega(d(v, u)) \right\}.$$

Then the following estimates are satisfied.

$$(a) \ d(u, \hat{u}) \leq \lambda \quad (b) \ \Phi(\hat{u}) \leq \Phi(u) \quad (c) \ \Phi(\hat{u}) \leq \Phi(v) + \frac{\varepsilon}{\lambda} d(v, \hat{u}) \quad \forall v \in X.$$

Proof. The proof is analogous to that of Theorem 1.1. □

Remark C.2. Examples of functions $\omega : [0, +\infty) \rightarrow [0, +\infty)$ satisfying the hypotheses above include power-type (*e.g.*, $\omega(s) = s^p/p$ for $p \geq 1$), exponential-type (*e.g.*, $\omega(s) = e^s - 1$), logarithmic-type (*e.g.*, $\omega(s) = s \log(1 + s)$), and hyperbolic-type (*e.g.*, $\omega(s) = \log \cosh s$). We refer to [35] for the study of proximal operators with Young moduli of continuity in uniformly convex Banach spaces.