Journal of Convex Analysis
Volume 28 (2021), No. 2, 599–612

From Linear to Convex

Jürgen Jost
Max-Planck-Institut für Mathematik, Leipzig, Germany
jjost@mis.mpg.de

Dedicated to Umberto Mosco on the occasion of his 80th birthday.

Received: January 4, 2020
Accepted: April 13, 2020

Umberto Mosco has developed a powerful theory of variational convergence for convex functionals and sets, or with a different terminology, for Dirichlet forms. Such forms are defined on Hilbert spaces. In this contribution, I shall describe how his theory can be extended from Hilbert spaces to metric spaces that themselves satisfy suitable convexity properties.

Keywords: Mosco convergence, Gamma convergence, Dirichlet form, convex functional, generalized harmonic mapping, nonpositive curvature, weak convergence, regularity theory.

2010 Mathematics Subject Classification: 58E20, 49J45, 54A20, 31C25, 46E36, 54E35.

1. Introduction

Dirichlet forms were introduced by Beurling and Deny [4, 5] as an abstract version of classical potential theory. The basic example is the Dirichlet integral

$$I(u) = \int_\Omega \|\nabla u(x)\|^2 dx \tag{1}$$

for an L^2-function on a domain $\Omega \subset \mathbb{R}^n$. Since the Sobolev space $W^{1,2}(\Omega)$, on which I is finite, is dense in $L^2(\Omega)$. I is densely defined in L^2, and one puts $D(u) = \infty$ if u is not in $W^{1,2}(\Omega)$. The Dirichlet integral I has some nice properties. It is convex and lower semicontinuous on the Hilbert space $H = L^2(\Omega)$. It can be turned into a bilinear quadratic form $I(u,v) = \int_\Omega \langle \nabla u(x), \nabla v(x)\rangle dx$. This form is symmetric and positive definite. If we denote the Hilbert space product by (u,v), we can introduce the Laplace operator Δ (with the usual sign convention that makes it a nonpositive operator) via

$$I(u,v) = (u, -\Delta v) = (\sqrt{-\Delta}u, \sqrt{-\Delta}v) \tag{2}$$

on a dense subspace of H, its domain of definition $D(\Delta)$. Δ generates the heat semigroup, which in turn describes random walks on Ω. The theory of Dirichlet forms provides an abstract version of these relations. One considers some Hilbert space H, often assumed separable, in order to facilitate convergence notions, and one considers a nonnegative definite bilinear form A that is defined on some linear subspace $D(A)$ of H. Usually, A is assumed to be symmetric [10], but the theory can also be developed without that assumption [26].

For our purpose, it is important that the nonnegativity makes $A(u,u)$ convex. In fact, since $A(u-v, u-v) \geq 0$, the bilinearity of A implies the Cauchy-Schwarz inquality

$$A(u,v) + A(v,u) \leq A(u,u) + A(v,v), \tag{3}$$

ISSN 0944-6532 / $ 2.50

and this in turn yields, for $0 \le t \le 1$,

$$A(tu+(1-t)v, tu+(1-t)v) \le (t^2+t(1-t))A(u,u)+((1-t)^2+t(1-t))A(v,v) \\ = tA(u,u) + (1-t)A(v,v) \tag{4}$$

When $u \notin D(A)$, one puts $A(u,u) = \infty$. Some further assumptions are needed in the theory of Dirichlet forms. The form should be closable, i.e. admit a closed extension, that is, an extension that is lower semicontinuous w.r.t. convergence in H. Even if the form is not closable, it admits a relaxation $\underline{A}$, the largest lower semicontinuous form with

$$\underline{A}(u,u) \le A(u,u) \text{ for all } u \in H. \tag{5}$$

For a closable form, $\underline{A}$ is the smallest closed extension, called its closure. A closed form A is called Markovian if whenever $u \in D(A)$, then also

$$v = (0 \vee u) \wedge 1 (= \inf(\sup(u,0),1) \in D(A) \text{ and } A(v,v) \le A(u,u). \tag{6}$$

That is, we can truncate u from above and below, or equivalently, project its values onto the interval $[0,1]$ without increasing the value of A. Note that this interval is a convex subset of the real line $\mathbb{R}$, the range of $A(u,u)$. Of course, a property analogous to (6) then holds for any other interval of the form $[0,a]$ for $a > 0$ or $[b,0]$ for $b < 0$. This cut-off property is important for showing the regularity of minimizers.

A Dirichlet form then is defined as a closed Markovian bilinear form, and as mentioned, mostly assumed symmetric. A Dirichlet form then generates a structure analogous to that described for the classical Dirichlet integral above. We shall henceforth also assume that the Dirichlet form A is symmetric, Such a Dirichlet form then defines a non-positive definite self-adjoint operator L on H that satisfies on its domain of definition $D(L)$

$$A(u,v) = -(Lu,v) = (\sqrt{-L}u, \sqrt{-L}v), \tag{7}$$

where the last relation holds for all $u, v \in D(E)$.

Umberto Mosco has applied the theory of Dirichlet forms with great success to composite media [28] and fractals [29]. The point is that such structures are often difficult to describe explicitly, and while they may obey variational properties, the corresponding variational principle can then not be written down explicitly, but only described asymptotically. Therefore, a theory is needed that can treat variational integrals abstractly, without depending on their explicit form, and at the same time allows for suitable notions of convergence in order to recover the properties of the structure from an asymptotic analysis. That theory, as Umberto Mosco argued with great success, is the theory of Dirichlet forms, and he has fashioned it into the appropriate tool to understand such structures in much detail. In particular, he has developed appropriate notions of convergence.

Now the present author in his work on generalized harmonic mappings into metric spaces has encountered a somewhat analogous problem in a nonlinear setting. When the metric space (Y,d), the target of a generalized harmonic map u, does not have a differential structure, we cannot directly define a derivative ∇u, and therefore, we cannot write down the nonlinear version of the Dirichlet integral. Instead of derivatives, we only have distances $d(u(x), u(y))$, and we can define approximate Dirichlet integrals through such differences, when the points x and y in the domain get closer and closer. It seemed therefore natural to try to extend the theory of Dirichlet forms to such a nonlinear setting, and this is what will be recalled here.

For our purposes, the crucial properties of Dirichlet forms then are lower semicontinuity, convexity, and the fact that we can project onto certain convex sets in the range. The main argument of the present contribution then is that convexity leads to a natural generalization of the linear structure in the classical Dirichlet form theory. Convexity will allow us to extend many of the basic constructions of that theory to the nonlinear setting of metric spaces of nonpositive curvature. This will include resolvents and semigroups as well variational aspects and the regularity of minimizers. In particular, fundamental constructions developed by Umberto Mosco for Dirichlet forms can naturally be extended to spaces of nonpositive curvature. This will be our first theme, see Section 3. The second theme is concerned with the regularity of minimizers. Again, Umberto Mosco has fashioned the appropriate tools in the context of Dirichlet forms, in his work with Marco Biroli. In particular, they obtained an abstract Harnack principle. In the same way that the Dirichlet integral (1) can be extended for maps $u : \Omega \to N$ into a Riemannian manifold N (see for instance [21]), Dirichlet forms can be extended from a Hilbert space of the form $L^2(\Omega)$ to mapping spaces $L^2(\Omega, (Y, d))$ into metric space (Y, d) (and of course for domains that are more general than subsets of $\mathbb{R}^n$). We shall describe that theory in Section 5 and then use the Harnack principle of [6] to show the regularity of local minimizers.

I thank the referee for a very careful reading of this paper.

2. Spaces of nonpositive curvature

For our purpose, we first need to explain the geometry of nonpositive curvature. We shall summarize the theory developed in [14]. Let (Y, d) be a metric space in which any two points can be connected by a shortest arc. Shortest arcs are geodesics, and we always parametrize them proportionally to arclength. Such a Y is connected, and we also assume that it is simply connected.

The key assumption for this paper is that (Y, d) be nonpositively curved in the sense of Alexandrov [1] (for systematic expositions of the theory, see [3, 18]). This means that the distance function is at least as convex as the Euclidean distance $\|.-.\|$ in $\mathbb{R}^2$. Formally, this can be expressed as follows. Whenever

$$\gamma : [0, b] \to Y \quad \text{and} \quad \tilde{\gamma} : [0, b] \to \mathbb{R}^2$$

are geodesic arcs (to repeat, parametrized by arclength) ($\tilde{\gamma}$ as a Euclidean geodesic is simply a straight line) and if $p \in Y$ and $\tilde{p} \in \mathbb{R}^2$ and satisfy

$$d(p, \gamma(0)) = \|\tilde{p} - \tilde{\gamma}(0)\| \quad \text{and} \quad d(p, \gamma(b)) = \|\tilde{p} - \tilde{\gamma}(b)\|, \tag{8}$$

then for all $t \in [0, b]$
$$d(p, \gamma(t)) \leq \|\tilde{p} - \tilde{\gamma}(t)\|. \tag{9}$$

In particular, Euclidean spaces have nonpositive curvature, and since we do not require our spaces to be locally compact, also Hilbert spaces are nonpositively curved.

In the rest of this section, we shall assume, even where not explicitly mentioned, that (Y, d) is such a space on nonpositive curvature. An easy consequence of nonpositive curvature is

Lemma 2.1. *If (Y, d) has nonpositive curvature in the above sense, then the geodesic arc between any $p_1, p_h 2 \in Y$ is unique and depends continuously on p_1 and p_2.*

Another consequence is Busemann's condition:

Lemma 2.2. *Let $\gamma_i : [0, b_i] \to Y, i = 1, 2$ be geodesic arcs with $\gamma_1(0) = \gamma_2(0)$. Then*

$$d(\gamma_1(b_1), \gamma_2(b_2)) \geq 2d(\gamma_1(b_1/2), \gamma_2(b_2/2), \tag{10}$$

that is, geodesics are diverging from each other at least as fast as in the Euclidean case.

The notion of nonpositive curvature enjoys the following *iteration principle* [15, 16].

Theorem 2.3. *Let (Y, d) be a metric space of nonpositive curvature. Let (X, μ) be a space equipped with a measure μ. Then the $L^2(X, \mu; Y, d)$ of maps into the metric space (Y, d), where the L^2-metric is defined by*

$$d^2(f, g)_{L^2} = \int_X d^2(f(x), g(x)) d\mu(x) \tag{11}$$

is again a metric space of nonpositive curvature.

Here, we may want to choose some base map $g_o : X \to Y$ and take the space of all $f : X \to Y$ with $d^2(f, g_0) < \infty$. Of course, when Y is a linear space, we may naturally choose $g_0 = 0$.

Proof. A geodesic between $f_0, f_1 \in L^2(X, \mu; Y, d)$ is given by the family of maps

$$f_t(x) = \gamma_x(t) \tag{12}$$

where $\gamma_x : [0, 1] \to Y$ is the geodesic in Y from $f_0(x)$ to $f_1(x)$. The nonpositive curvature (9) then extends to the L^2-space by integration. $\square$

This is a reason why we did not require that (Y, d) be locally compact. In fact, even it were, then $L^2(X, \mu; Y, d)$ in general is no longer locally compact (unless X is finite).

Nonpositive curvature in this sense has many further nice consequences. In particular, it implies the uniqueness of mean values (we shall call them *Karcher means*, as Karcher developed them as a systematic tool in Riemannian geometry).

Lemma 2.4. *Let $p_1, ..., p_m \in Y$. Then there exists a unique $p = m(p_1, ..., p_m) \in Y$ with*

$$p = \operatorname{argmin}_{q \in Y} \sum_{i=1}^{m} d^2(q, p_i). \tag{13}$$

If $\gamma : [0, b] \to Y$ is a geodesic arc, then $m(\gamma(0), \gamma(b) = \gamma(b/2)$.

Definition 2.5. A function $F : Y \to \mathbb{R} \cup \infty$ is called *convex* if for any $p, q \in Y$

$$F(m(p, q)) \leq \frac{1}{2}(F(p) + F(q)). \tag{14}$$

It was already observed in [14] that on spaces of nonpositive curvature, we have a version of Jensen's inequality which in this simple situation can be formulated as

Lemma 2.6. *Let Y have nonpositive curvature, $F : Y \to \mathbb{R} \cup \infty$ convex. Then, for $p_1, \dots, p_m \in Y$, we have*

$$F(m(p_1, \dots, p_m)) \leq \frac{1}{m} \sum_{i=1}^{m} F(p_i). \tag{15}$$

Jensen inequalities on more general metric spaces were obtained in [23, 24, 32].

Definition 2.7. The set $C \subset Y$ is *convex* if whenever $p_1, p_2 \in C$, the geodesic arc between p_1 and p_2 is also contained in C. For any $B \subset Y$ its *convex hull* $C(B)$ is the smallest convex set $C \subset Y$ with $B \subset C$.

The convex hull is unique because the intersection of two convex sets is again convex. In a space of nonpositive curvature, points outside a convex set can be uniquely projected onto its closure (which is also convex).

Lemma 2.8. *Let $C \subset Y$ be a convex set. Then for any $p_0 \in Y$ there exists a unique $p = \pi(p_0, C)$ in the closure $\bar{C}$ with*

$$d(p_0, p) \leq d(p_0, q) \text{ for all } q \in \bar{C}. \tag{16}$$

This point $p = \pi(p_0, C)$ is called the projection of p_0 onto $\bar{C}$.

Corollary 2.9. *If $C \subset Y$ is convex and $p_1, ..., p_m \in C$, then also $m(p_1, ..., p_m) \in C$.*

In fact, if $m(p_1, \dots, p_m)$ were outside C, we could project it onto C by Lemma 2.8 without increasing (13), contradicting either its minimizing property or its uniqueness.

Lemma 2.10. *The convex hull of B can be iteratively constructed by starting with $C_0 := B$ and defining C_{n+1} as the union of all geodesic arcs between points in C_n. Then*

$$C(B) = \bigcup_n C_n. \tag{17}$$

Corollary 2.11. *The convex hull of a separable set is separable.*

We can now introduce one of the fundamental concepts of [14].

Definition 2.12. $p_0 \in Y$ is called the *weak limit* of a sequence $(p_n)_{n \in \mathbb{N}} \subset Y$ if for every geodesic arc γ with $\gamma(0) = p_0$, the projections $\pi(p_n, \gamma)$ converge to p_0.

Of course, in a Hilbert space, this reduces to the standard concept of weak convergence. The point we want to make here is that the notions of Hilbert space theory around weak convergence naturally generalize to spaces of nonpositive curvature. In that sense, it was proved in [14] that nonpositively curved spaces are weakly compact, that is

Theorem 2.13. *Every bounded sequence $(p_n) \subset Y$ contains a weakly convergent subsequence.*

In order to strongly distinguish ordinary convergence, that is, convergence w.r.t. the metric d, from weak convergence, we shall also call the former *strong convergence*.

We also have an analogue of the Banach-Saks lemma.

Theorem 2.14. *For a bounded sequence $(p_n) \subset Y$, after selection of a subsequence, the mean values $m(p_1, \dots, p_n)$ converge to some $p_0 \in Y$ for $n \to \infty$.*

We recall the

Proof. Let p_0 be the weak of a subsequence, again denoted by (p_n) for simplicity. When the sequence itself converges (in which case of course the limit is p_0), then also the mean values converge. Thus, we have to consider the case where the sequence does not converge. Then, after further selection of a subsequence, there is some $\eta > 0$ with

$$d^2(p_0, p_n) \geq \eta \text{ for all } n, \tag{18}$$

and hence we may also assume $\lim d^2(p_0, q_\nu) = \eta > 0$. Since (p_n) does not converge, it cannot be a Cauchy sequence, and so, we may also assume

$$d^2(p_m, p_n) \geq \eta \text{ for all } m \neq n. \tag{19}$$

We now claim that we can find a subsequence (as always without changing the notation) with

$$d^2(p_0, \pi(p_m, \gamma_n)) + d^2(p_0, \pi(p_n, \gamma_m)) < \frac{1}{n} \text{ for } m < n \tag{20}$$

where γ_k is the geodesic arc from p_0 to p_k. In fact, that $d^2(p_0, \pi(p_n, \gamma_m))$ can be made small for $m < n$ is a consequence of the weak convergence, and that $d^2(p_0, \pi(q_\mu, \gamma_\nu)$ can be made small then follows from this and (19). The latter can be seen by contradiction. Otherwise $d^2(p_0, \pi(p_n, \gamma_m)) \to 0$ for some fixed m and $n \to infty$, while $d^2(p_0, \pi(p_m, \gamma_n)) > \sigma$ for some $\sigma > 0$. But the convexity of the ball $B(p_m, d(p_m, \pi(p_m, \gamma_n)))$ yields a contradiction.

The proof is then completed by comparison arguments. We consider any geodesic startting at the weak limit p_0. As a consequence of nonpositive curvature (see for instance (11)), the distances from the p_n then are not smaller than the comparison Euclidean distances. Therefore for any point on that line different from the weak limit p_0, the sum of the squared distances eventually becomes larger than that from p_0. Therefore p_0 asymptotically becomes the mean value. □

A Banach-Saks lemma on more general metric spaces was obtained in [32].

As a consequence, we have

Corollary 2.15. *Let $F : Y \to \mathbb{R} \cup \infty$ be a convex function on a metric space (Y, d) of nonpositive curvature. If F is lower semicontinuous, that is, whenever $p_n \to p$ in Y, then*

$$F(p) \leq \liminf_{n \to \infty} F(p_n), \tag{21}$$

then it is also weakly lower semicontinuous, that is, (21) also holds when the p_n converge to p weakly.

Proof. Let p_n converge to p weakly. After selection of a subsequence, we may assume that $\lim F(p_n)$ exists. Then by Theorem 2.14, after selection of a subsequence, the family of mean values $q_n = m(p_1, \dots, p_n)$ converges (strongly) to p.

By Jensen's inequality (Lemma 2.6), we obtain $F(p) \leq \liminf F(q_n) \leq \lim F(p_n)$, if F is lower semicontinous. □

3. Nonlinear Dirichlet forms

Let Z be a topological space that for simplicity (so that we can work with sequences instead of filters) satisfies the first axiom of countability. For a functional

$$F\colon Z \to \mathbb{R} \cup \{\pm\infty\}, \quad \text{we let} \quad D(F) = \{z \in Z : F(z) \in \mathbb{R}\}$$

be its domain of definition. Let $F_n\colon Z \to \mathbb{R} \cup \{\pm\infty\}$ be functionals.

Since the aim of the theory is to find minimizers of such functionals and their possible limits, we shall assume that the functionals are bounded from below. In order to stress the analogy with Dirichlet forms, we assume that the infimum of each F_n (although, strictly speaking, we would need that only for their possible limits, under the notions of convergence to be introduced shortly) satisfies

$$\inf_{z\in Z} F_n(z) = 0. \tag{22}$$

Moreover, we shall be mostly interested in convex functionals. In order to have an appropriate notion of convexity, we therefore have the case in mind where Z is a metric space of nonpositive curvature.

Below, we shall utilize a technical notion, that of a Rellich condition. It abstracts Rellich's theorem that a bounded sequence in the Hilbert space $L^2(\Omega)$ for an open domain $\Omega \subset \mathbb{R}^d$ whose Dirichlet integrals are bounded contains a convergent subsequence.

Definition 3.1. $F\colon Z \to \mathbb{R}_+ \cup \{\infty\}$ satisfies the *Rellich property* if every bounded sequence (u_n) in Z with $\liminf\limits_{n\to\infty} F(u_n) < \infty$ contains a convergent subsequence.

The sequence $F_n\colon Z \to \mathbb{R}_+ \cup \{\infty\}$ satisfies the *asymptotic Rellich property* if every bounded sequence (u_n) with $\liminf_{n\to\infty} F_n(u_n) < \infty$ contains a convergent subsequence.

For variational purposes, one needs notions of convergence under which the minimizers of a sequence of approximating functionals converge to minimizers of the limit functional. Inverting the historical order, let us first discuss de Giorgi's general notion of Γ-convergence [9, 8] (see [7] as a textbook reference and [22] for a short introduction). Here, we do not yet need the assumption (22).

Definition 3.2. A sequence $(F_n)_{n\in\mathbb{N}}$ of functionals Γ-*converges* to the functional $F\colon Z \to \mathbb{R} \cup \{\pm\infty\}$, or shortly, $F = \Gamma - \lim_{n\to\infty} F_n$, if

(i) if $(z_n)_{n\in\mathbb{N}} \subset Z$ converges to $z \in Z$, then $F(z) \le \liminf_{n\to\infty} F_n(_n)$, and

(ii) for every $z \in Z$, there exists some sequence $(\zeta_n)_{n\in\mathbb{N}} \subset Z$ converging to z that satisfies $F(z) \ge \limsup_{n\to\infty} F_n(\zeta_n)$.

The first condition here is the important one; the second condition only serves the purpose of normalization. Γ-limits have some useful properties

(1) Γ-limits are lower semicontinuous.

(2) If all the F_n are convex, then Γ-lim F_n is also convex.

(3) If z_n is a minimizer of F_n, and if $z_n \to z$, then z is a minimizer of F.

(4) If Z is second countable, then every sequence $(F_n)_{n\in\mathbb{N}}$ of functionals possesses a Γ-convergent subsequence.

The notion of Γ-convergence naturally fits with variational constructions like that of a *relaxation*, to be recalled below.

In [27, 28], with convex sets and functionals in mind, Mosco had defined a different notion of convergence. When viewed from the perspective of Γ-convergence, it can be seen as a stronger notion that amounts to replacing (i) by

(i') if $(z_n)_{n\in\mathbb{N}} \subset Z$ *weakly* converges to $z \in Z$, then $F(z) \leq \liminf_{n\to\infty} F_n(z_n)$.

We then have the analogue of Lemma 2.3 in [28].

Lemma 3.3. *If the family (F_n) satisfies an asymptotic Rellich condition, then it Γ-converges to a functional F if and only if it converges to F in the sense of Mosco.*

Of course, convergence in the sense of Mosco implies Γ-convergence. Therefore, it needs only be shown that Γ-convergence implies convergence in Mosco's sense. The crucial point in Mosco's *proof* is that since a weakly convergent sequence is bounded, by the asymptotic Rellich property, we can extract a (strongly) convergent subsequence.

Mosco's approach, however, allows for further insight as we shall now describe. For every functional F as above, there exists a largest one among the lower semicontinuous functionals G with $G(u) \leq F(u)$ for all u. This functional is called the *relaxation* $\underline{F}$ of F, given by

$$\underline{F}(u) = \min\{\liminf_{n\to\infty} F(u_n) : u_n \to u \text{ in } Z \text{ for } n \to \infty\}$$

Its domain of definition satisfies $D(F) \subset D(\underline{F})$.

To now make closer contact with Mosco's theory of variational convergence [28], we introduce the notion of the *Moreau-Yosida approximation* (called Deny-Yosida approximation by Mosco, but we follow here the terminology of [2]) of a convex functional F

$$F^\lambda(x) = \inf_{y\in Z}\{\lambda F(y) + d^2(x,y)\}. \tag{23}$$

Importantly, here we assume that Z is a metric space, and in fact, we want it to be a space of nonpositive curvature, in order to have d^2 be strictly convex in a quantitative sense. The purpose of the Moreau-Yosida approximation is precisely to regularize a given functional by adding a strictly convex penalty term. By tuning the parameter λ, we can start at any given pont x for $\lambda = 0$ and try to move towards a minimizer of F for $\lambda \to \infty$. – Of course, we could develop a corresponding theory on other metric spaces that support such a family of strictly convex functions.

We observe the following inequality

$$F^\lambda(x) \leq \lambda F(x) \text{ for all } x, \lambda. \tag{24}$$

We also need the *resolvents* $J_\lambda x$ that satisfy.

$$F^\lambda(x) = \lambda F(J_\lambda x) + d^2(x, J_\lambda x),$$

assuming that the infimum in (23) is achieved, which is the case, for instance, if F is convex, $\not\equiv \infty$, and lower semicontinuous.

Of course, the resolvents cannot distinguish between constants, that is, the functionals F and $F + c$ with $c \in \mathbb{R}$ have the same resolvents. This problem is avoided by the normalization (22).

We can then extend Mosco's theory [28].

Definition 3.4. The convex functionals $F_n : Z \to \mathbb{R}_+ \cup \{\infty\}$ *Mosco-converge* to the functional F if for every $\lambda > 0$, the Moreau-Yosida approximations $(\underline{F}_n)^\lambda$ of the relaxations of the F_n converge to F^λ pointwise in Z, or equivalently (assuming the normalization (22)), if the resolvents $J_{n,\lambda}$ of the relaxed forms $\underline{F}_n$ converge to the resolvents J_λ.

A result of [19] generalizes Mosco's characterization of his notion of convergence to the nonlinear setting of the present contribution.

Theorem 3.5. *If $(F_n)_{n\in\mathbb{N}}$ Mosco-converges to F, then (F_n) also Γ-converges to F. The converse holds when the $F_n : Z \to \mathbb{R}_+ \cup \{\infty\}$ are convex and satisfy an asymptotic Rellich property.*

The advantage of the concept of Mosco-convergence then lies in the fact that it provides a concrete criterion to verify Γ-convergence.

4. Regularity theory

In [6], Biroli and Mosco considered certain families of Dirichlet forms that describe the variational behaviour of possibly highly nonhomogeneous and nonisotropic materials. They proved a Harnack inequality for their local solutions. From the perspective adopted in this contribution, the important achievement of [6] is an abstract and structural version of Moser's Harnack principle [30, 31] which constitutes one of the most fundamental contributions to the regularity of partial differential equations (see [20] for a textbook presentation). The Harnack inequality can be seen as a quantitative version of the maximum principle for elliptic and parabolic differential equations and therefore allows for quantitative estimates for their solutions.

In [6], Biroli and Mosco derived such a Harnack inequality for minimizers of Dirichlet forms. To get such an inequality, some further conditions need to be imposed on the Dirichlet form. We shall now recall them. Let (X, μ) be a locally compact, separable metric space with a positive Radon measure μ with $\mathrm{supp}(\mu) = X$. Let $E(\cdot,\cdot)$ be a Dirichlet form on $L^2(X, \mu)$. We assume that $E(\cdot,\cdot)$ is strongly local in the sense that

$$E(u, v) = 0$$

whenever v is constant on some neighborhood of $\mathrm{supp}(u)$. We also assume that E can be represented as

$$E(u, v) = \int \eta(u, v)(dx) \tag{25}$$

where $\eta(\cdot,\cdot)$ is a Radon-measure valued, nonnegative, bilinear form on $D(E)$. By the representation theorem of Beurling-Deny, this is possible under rather general circumstances, see [10]. We consider the bilinear form

$$(u, v)_1 := (u, v) + E(u, v)$$

on $D(E)$ and let $\| \cdot \|_1$ be the norm induced by that bilinear form.

We assume that E admits a μ-separating core Γ, that is, a subspace of $D(E)\cap C_0^0(X)$ that is dense in $C_0^0(X)$ w.r.t. the uniform norm and dense in $D(E)$ w.r.t. $\|\cdot\|_1$ and that has the following separation property: For any $x, y \in X, x \neq y$, there is some $\phi \in \Gamma$ with

$$\eta(\phi,\phi) \leq \mu \text{ on } X, \text{ and } \phi(x) \neq \phi(y).$$

E then induces the metric

$$d_E(x,y) := \sup\{\phi(x) - \phi(y) : \varphi \in \Gamma, \eta(\phi,\phi) \leq \mu \text{ on } X\}$$

on X. We assume that the topology on X induced by the metric d_E coincides with the original topology of X. Biroli-Mosco [6] then imposed the following crucial conditions

(A1) (Ball doubling property): The measure μ has the ball doubling property w.r.t. d_E. That means that there exist constants $c_0 < \infty$ and $r_0 > 0$ such that the balls $B(x,r) := \{y \in X : d_E(x,y) < r\}$ for $x \in X, 0 < r \leq r_0$ satisfy

$$\mu(B(x,2r)) \leq c_0 \mu(B(x,r)).$$

(A2) (Poincaré inequality): For any $X_0 \Subset X$, there exist constants $0 < \lambda \leq 1$, $c_1 < \infty$, s.t. for every $x_0 \in X_0$, $r > 0$ with $B(x_0,r) \subset X_0$, and every $u \in D(E)$,

$$\int\limits_{B(x_0,\lambda r)} \|u(x) - \bar{u}_{B(x_0,\lambda r)}\|^2 \mu(dx) \leq c_1 r^2 \int\limits_{B(x_0,r)} \eta(u,u)dx,$$

where $\bar{u}_\Omega$ denotes the mean value of u on a set Ω, that is,

$$\bar{u}_\Omega = \operatorname{argmin}_q \int\limits_\Omega \|u(x) - q\|^2 \mu(dx).$$

Under these assumptions, Biroli and Mosco could derive a Harnack inequality. In [17], we have used their inequality to obtain the following version. We shall formulate it with the following notation for $v \in L^\infty(X)$

$$v_{+,R} := \sup_{B(x_0,R)} v\,, \quad v_R := \fint\limits_{B(x_0,R)} v\mu(dx) \quad \text{ on the ball } B(x_0,R) \subset X.$$

Theorem 4.1. *There exists a constant δ_0, such that for all weak subsolutions $v \in D(E) \cap L^\infty(X,\mu)$ of $(Lv \geq 0)$ and all $R > 0$ with $B(x_0,4R) \Subset X$ we have*

$$v_{+,R} \leq (1-\delta_0)v_{+,4R} + \delta_0 v_R. \tag{26}$$

This tells us that the quantitative control of a subsolution gets better and sharper if we move further inside the domain on which it is defined.

From their Harnack inequality, Biroli-Mosco [6] can also derive estimates for Green functions of such operators L that generalize the classical estimates that are collected in [13]. We shall also use those estimates below.

5. Generalized Dirichlet forms

In Section 3, we have seen that Mosco's convergence theory for Dirichlet forms can be extended to convex functionals on spaces of nonpositive curvature. There, we have considered such a functional as a nonlinear version of a Dirichlet form. In other words, we have replaced the Hilbert space underlying a Dirichlet form by a metric space of nonpositive curvature. Of course, Hilbert spaces fall into that class (they have zero curvature), and so, this is a natural generalization. Now, the prototype of such a Hilbert space is $L^2(X,\mu)$, where X is a topological space equipped with a measure μ, for instance $L^2(\Omega)$ for an open $\Omega \subset \mathbb{R}^n$ with the Lebesgue measure, as explained in Section 1. It is therefore also natural to replace that Hilbert space by the L^2-space $L^2(X,\mu;Y,d)$ of maps into a metric space (Y,d), where as in (12) the L^2-metric is defined by

$$d^2(f,g)_{L^2} = \int_X d^2(f(x),g(x))d\mu(x). \tag{27}$$

In the sequel, we shall assume that (X,μ) is a locally compact, separable metric space with a positive Radon measure μ with $\mathrm{supp}(\mu) = X$ satisfying the properties of Biroli-Mosco [6] as presented in Section 4.

When we extend a Dirichlet form on the Hilbert space $L^2(X,\mu) = L^2(X,\mu;\mathbb{R})$ to such L^2-mappings, the bilinearity is replaced by the *Quadratic contraction property*: If $f\colon X \to (Y,d)$ is in $D(E)$ and $g\colon (f(X),d)(\subset (Y,d)) \to (Z,d')$ for some metric space (Z,d') is a Lipschitz mapping, i.e.

$$d'(g(p),g(q)) \le \lambda d(p,q) \quad \text{for all } p,q \in f(X)$$

for some constant λ, then $g \circ f \in D(E))$ and $E(g \circ f) \le \lambda^2 E(f)$.

We note that this quadratic contraction property implies the Markov condition (6).

As before, we require lower semicontinuity, as well as a *density property* that extends the requirement that $D(E) \cap L^2(X,\mu;\mathbb{R})$ be dense in $L^2(X,\mu;\mathbb{R})$. If for every Lipschitz map $g : (Y,d) \to \mathbb{R}$, $g \circ f \in D(E)$, then also $f \in D(E)$.

In a variational setting, one then seeks local minimizers of such Dirichlet forms. Here, $u : X \to Y$ is a local minimizer if for every $x \in X$, there is some radius $R_x > 0$ such that for all $0 < R < R_x$

$$E(u_{B(x,R)}) \le E(v_{B(x,R)})$$

for every $v \in L^2(B(x,R_x),\mu;Y,d)$ with $u = v$ on $B(x,R_x)\backslash B(x,R)$. (Here, u_B denotes the restriction of u to the set $B \subset X$. Our generalized Dirichlet form naturally restricts from X to B.

In [17], a regularity theorem for local minimizers was obtained.

Theorem 5.1. *Suppose E is a generalized Dirichlet form satisfying the assumptions that we have stated. Let (Y,d) be a metric space of nonpositive curvature. Then every local minimizer of E is Hölder continuous.*

Such a Hölder regularity was independently obtained by F. H. Lin in a somewhat different context. He used an elegant compactness argument that needs the assumption that the target (Y,d) be locally compact.

Under additional assumptions on the domain X, local minimizers are even Lipschitz continuous, as was recently shown by H. C. Zhang and X. P. Zhu [33].

To conclude this paper, we present the crucial step in the proof of Theorem 5.1 from [17], a telescoping lemma (a strategy originally developed in [11, 12]) that yields a self improving property when restricting the local minimizer to smaller balls.

Lemma 5.2. *Let (Y, d) be a metric space of nonpositive curvature, and $u\colon X \to Y$ be a local minimizer for E. Let $p \in Y$. Suppose $B(x_0, 4R) \Subset X$.*

Then,
$$v(x) := d^2(u(x), p) \tag{28}$$

is a subsolution for the selfadjoint operator L associated with the Dirichlet form E, and with $|B(x_0, R)| := \mu(B(x_0, R))$, we have

$$\frac{R^2}{|B(x_0, R)|} \int\limits_{B(x_0,R)} \eta(u, u)\mu(dx) \le c_5(v_{+,4R} - v_{+,R})\,. \tag{29}$$

Proof. We have
$$Lv \ge 2\eta(u, u) \tag{30}$$

by the nonpositive curvature of Y (see Lemma 5 in [17]), that is, a quantitative version of the subsolution property of v. This comes from the strict convexity of d^2 on a space of nonpositive curvature, see Section 2.

We then use the mollified Green function $G^R(x, y)$ on $B(x_0, R)$ relative to $B(x_0, 2R)$, with $B(x_0, 4R) \Subset X$. We have

$$G^R(x_0, \cdot) \in D_0(E, B(x_0, 2R)) \quad (:= \text{closure of } D(E) \cap C_0^0(B(x_0, 2R)),$$

and it is characterized by the relation

$$\int\limits_{B(x_0,2R)} \eta(\varphi(x), G^R(x_0, x))dx = \fint\limits_{B(x_0,R)} \varphi(x)\mu(dx) \tag{31}$$

for all $\varphi \in D(E)$ with supp $\varphi \Subset B(x_0, 2R)$. (For $R \to 0$, we obtain the usual Green function; with the mollification, we avoid the latter's singularity.)

We then define the test function $w^R(x) := \dfrac{|B(x_0, 4R)|}{R^2} G^R(x_0, x)$.

It satisfies
$$\int\limits_{B(x_0,2R)} \eta(\varphi(x), w^R(x))dx = \frac{1}{R^2} \int\limits_{B(x_0,R)} \varphi(x)\mu(dx) \tag{32}$$

for all $\varphi \in D(E)$ with supp $\varphi \Subset B(x_0, 2R)$. We can then invoke the estimates for G^R of [6],

$$0 \le w^R \le \gamma_1 \quad \text{in } B(x_0, 2R) \tag{33}$$

$$w^R \ge \gamma_2 > 0 \quad \text{in } B(x_0, R) \tag{34}$$

for constants γ_1, γ_2 that do not depend on R.

$z := v - v_{+,4R}$ then satisfies

$$\begin{aligned}
\lambda \int\limits_{B(x_0,2R)} \eta(u,u)(w^R)^2 &\leq \int\limits_{B(x_0,2R)} (w^R)^2 Lz \\
&= - \int\limits_{B(x_0,2R)} \eta((w^R)^2, z) \qquad \text{since } w^R \in D_0(E, B(x_0,2R)) \\
&= -2 \int \eta(w^R, w^R z) + 2 \int z\eta(w^R, w^R) \leq -2 \int \eta(w^R, w^R z) \qquad \text{since } z \leq 0.
\end{aligned}$$

Using (32), (33), (34), we can then apply the Harnack inequality of Theorem 4.1 to obtain the desired estimate,

$$\begin{aligned}
\int\limits_{B(x_0,R)} \eta(u,u) \leq \frac{c_4}{R^2} \int\limits_{B(x_0,R)} (v_{+,4R} - v) &\leq c_4 \frac{|B(x_0,R)|}{R^2}(v_{+,4R} - v_R) \\
&\leq c_5 \frac{|B(x_0,R)|}{R^2}(v_{+,4R} - v_{+,R}). \qquad \square
\end{aligned}$$

References

[1] A. D. Alexandrov: *Über eine Verallgemeinerung der Riemannschen Geometrie*, Schriftenreihe d. Forschungsinstituts f. Mathematik b. d. Deutschen Akademie der Wissenschaften zu Berlin 1 (1957) 33–84.

[2] H. Attouch: *Variational Convergence for Functions and Operators*, Pitman, London (1984).

[3] V. N. Berestovskij, I. G. Nikolaev: *Multidimensional Generalized Riemannian Spaces*, Encyclopedia of the Mathematical Sciences 70, Springer, Berlin (1993).

[4] A. Beurling, J. Deny: *Espaces de Dirichlet. I: Le cas élémentaire*, Acta Math. 99 (1958) 203–224.

[5] A. Beurling, J. Deny: *Dirichlet spaces*, Proc. Nat. Acad. Sci. U.S.A. 45 (1959) 208–215.

[6] M. Biroli, U. Mosco: *A Saint-Venant type principle for Dirichlet forms on discontinuous media*, Ann. Mat. Pura Appl. 169 (1995) 125–181.

[7] G. Dal Maso: *An Introduction to Γ–Convergence*, Birkhäuser, Basel (1993).

[8] E. De Giorgi: *Convergence problems for functionals and operators*, in: *Ennio De Giorgi. Selected Papers*, Springer, Berlin (2006) 487–516.

[9] E. De Giorgi, T. Franzoni: *Su un tipo di convergenza variazionale*, Atti Accad. Naz. Lincei, Cl. Sci. Fis. Mat. Nat., Rendiconti 58/6 (1975) 842–850.

[10] M. Fukushima, Y. Oshima, M. Takeda: *Dirichlet Forms and Symmetric Markov Processes*, De Gruyter, Berlin (1994).

[11] M. Giaquinta, E. Giusti: *On the regularity of the minima of variational integrals*, Acta Math. 148 (1982) 31–46.

[12] M. Giaquinta, S. Hildebrandt: *A priori estimates for harmonic mappings*, J. Reine Angew. Math. 336 (1982) 124–164.

[13] M. Grüter, K.-O. Widman: *The Green function for uniformly elliptic equations*, Manuscripta Math. 37 (1982) 303–342.

[14] J. Jost: *Equilibrium maps between metric spaces*, Calc. Var. 2 (1994) 173–204.

[15] J. Jost: *Convex functionals and generalized harmonic maps into spaces of non-positive curvature*, Comment. Math. Helv. 70 (1995) 659–673.

[16] J. Jost: *Generalized harmonic maps between metric spaces*, in: *Geometric Analysis and the Calculus of Variations for Stefan Hildebrandt*, J. Jost (ed.), International Press, Boston (1996) 143–174.

[17] J. Jost: *Generalized Dirichlet forms and harmonic maps*, Calc. Var. 5 (1997) 1–19.

[18] J. Jost: *Nonpositive Curvature: Geometric and Analytic Aspects*, Birkhäuser, Basel (1997).

[19] J. Jost: *Nonlinear Dirichlet forms*, in: *New Directions in Dirichlet Forms*, J. Jost, W. S. Kendall, U. Mosco, M. Röckner, K.-T. Sturm (eds.), AMS/IP Studies in Advanced Mathematics, American Mathemtical Society, Providence (1998) 1–47.

[20] J.Jost: *Partial Differential Equations*, 3rd ed., Springer, Berlin (2013).

[21] J. Jost: *Riemannian Geometry and Geometric Analysis*, 7th ed., Springer, Berlin (2017).

[22] J. Jost, X. Li-Jost: *Calculus of Variations*, Cambridge University Press, Cambridge (1998).

[23] M. Kell: *Uniformly convex metric spaces*, Anal. Geom. Metr. Spaces 2 (2014) 359–380.

[24] K. Kuwae: *Jensen's inequality on convex spaces*, Calc. Var. Partial Diff. Equations 49 (2014) 1359–1378.

[25] F. H. Lin: *Analysis on singular spaces*, in: *Geometry, Analysis and Mathematical Physics, in Honor of Prof. Chaohao Gu*, Ta-Tsien Li (ed.), World Scientific, Singapore (1997) 114–126.

[26] Z. M. Ma, M. Roeckner: *An Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*, Springer, Berlin (1992).

[27] U. Mosco: *Convergence of convex sets and of solutions of variational inequalities*, Adv. Math. 3/4 (1969) 510–585.

[28] U. Mosco: *Composite media and asymptotic Dirichlet forms*, J. Funct. Analysis 123 (1994) 368–421.

[29] U. Mosco: *Dirichlet forms and self-similarity*, in: *New Directions in Dirichlet Forms*, J. Jost, W. S. Kendall, U. Mosco, M. Röckner, K.-T. Sturm (eds.), AMS/IP Studies in Advanced Mathematics, American Mathemtical Society, Providence (1998) 117–155.

[30] J. Moser: *On Harnack's inequality for elliptic differential equations*, Comm. Pure Appl. Math. 14 (1961) 377–591.

[31] J. Moser: *A Harnack inequality for parabolic differential equations*, Comm. Pure Appl. Math. 17/1 (1964) 101–134.

[32] T. Yokota: *Convex functions and barycenter on CAT(1)-spaces of small radii*, J. Math. Soc. Japan 68/3 (2016) 1297–1323.

[33] H. C. Zhang, X. P. Zhu: *Lipschitz continuity of harmonic maps between Alexandrov spaces*, Inventiones Mathematicae 211/3 (2018) 863–934.