Nonlinear Dirichlet Forms

Jürgen Jost

October 10, 2002

Introduction

The theory of Dirichlet forms started with the work of Beurling-Deny [BD1, BD2]. It was conceived as an abstract version of the theory of the Dirichlet integral, i.e. of the variational theory of harmonic functions. It included also the discrete harmonic functions that had received their mathematical treatment already by Courant-Friedrichs-Lewy [CFL]. Courant-Friedrichs-Lewy also pointed out a connection with discretized Brownian motion, and more recently, the stochastic aspects of Dirichlet forms have found an important place in the subject. The theory of Dirichlet forms is well documented in the monographs of Fukushima-Oshima-Takeda [FOT], Ma-Roeckner [MR], and Bouleau-Hirsch [BH]. As explained in other contributions in this volume, Dirichlet forms also constitute the proper axiomatic framework for many aspects of the theory of homogenization, self-similar fractals, or infinite particle systems, to name but a few.

The theory of Dirichlet forms is essentially linear as it is essentially concerned with scalar valued functions which are square integrable. In Riemannian geometry, however, a nonlinear generalization of harmonic functions is quite important. For that purpose, one generalized the classical Dirichlet integral to a so-called energy integral for maps with values in a Riemannian manifold. Minimizers (or other critical points) of this energy integral are called harmonic maps. Harmonic maps have found many applications in Riemannian geometry, but for other applications (as. e.g. by Gromov-Schoen [GS] to p-adic superrigidity) it was found necessary to consider target spaces that are more general than Riemannian manifolds. In that context, existence results were established by Jost [J2] and Korevaar-Schoen [KS]. In order to gain a deeper understanding and at the same time establish the most general existence results, the present author found it desirable to develop a more abstract and axiomatic version of the theory ([J3], [J4], [J5]). In particular, the approach taken in [J5] was to consider a theory of nonlinear or generalized Dirichlet forms as the appropriate setting for these generalized harmonic maps. In a certain sense, this brings together the two quite different generalizations of harmonic functions alluded to above - Dirichlet forms and harmonic maps - and it is hoped that this will lead to new insights in both fields. This theory is the topic of the present lecture notes.

In order to have a good theory of harmonic maps, one needed to impose the condition that the target manifold has nonpositive curvature. This condition can be interpreted as a convexity condition on the energy functional. Convexity conditions will also be fundamental for the abstract theory developed here. In fact, it turns out that convexity is the appropriate replacement for the linear structure in the classical Dirichlet form theory. Convexity will allow to extend many of the basic constructions of that theory to our setting, from resolvents and semigroups to variational aspects and the regularity of minimizers. This will be a basic theme in these notes. An important technical point will be to avoid assumptions of local compactness on the image. We shall concentrate on the analytic and axiomatic aspects. The geometric aspects have been treated already in detail in the author's companion lecture notes [J6]. Therefore, we shall be brief about geometric concepts and do not mention geometric applications. The geometric definitions and results needed for understanding these notes are summarized in an appendix. The reader can consult this appendix whenever he/she encounters an unfamiliar geometric concept in the main body of the text. In particular, the notion of nonpositive curvature in the sense of Alexandrov for a metric space will be important. For further details, we refer to [J6] as already mentioned. For the basic notions of Riemannian geometry and the classical theory of harmonic maps, the reader can also consult the author's textbook [J1].

We shall also not treat the very interesting stochastic aspects of harmonic maps and refer the reader to Kendall's contribution to this volume instead.

Acknowledgement: The research underlying these lecture notes was generously supported by the DFG through the Leibniz program.

1 Definition and properties of generalized Dirichlet forms

Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space. A key notion for us is the space

$$L^2(X,\mu) := \left\{ f : X \to \mathbb{R} \cup \{\pm\infty\}, \mu \text{ - measurable}, \int_X f^2(x)\mu(dx) < \infty \right\}$$

of square integrable functions. (Here and in the sequel, functions that differ only on a set of vanishing μ - measure will be identified.) On $L^2(X,\mu)$, one has the inner product

$$(f,g) := \int_X f(x)g(x)\,\mu(dx)\,,$$

and this product makes $L^2(X,\mu)$ into a Hilbert Space.

The main aim of these lectures is to extend the theory based on this Hilbert space to a more general setting where instead of functions, we consider mappings $f : X \to Y$ with values in some complete metric space (Y,d). To make this generalization meaningful, we need some elementary preparations.

First of all, since L^2 - functions need to be defined only μ - almost everywhere (abbreviated as μ - a.e., or simply a.e. in the sequel), and since L^2 - functions can assume the values $\pm\infty$ at most on a set of μ measure 0, in the definition of $L^2(X,\mu)$ it obviously suffices to consider functions $f : X \to \mathbb{R}$, as we may simply redefine f on the set where it assumes the values $\pm\infty$. Next, we need to explain what a μ - measurable map with values in a complete metric space is. The starting point here is a result that a real - valued function is measurable if it is the pointwise limit of simple functions. In general, we call a function

$$\varphi : X \to Y$$

simple if there exists pairwise disjoint, μ measurable sets $\mathcal{B}_1, \ldots, \mathcal{B}_n \subset X$ whose union is X such that φ is constant on each $\mathcal{B}_j$.

$$f : X \to Y$$

now is called μ - measurable if it is the pointwise limit of simple functions. As in the case of real-valued functions, we then have the following important composition property:
If $f : X \to Y$ is measurable, $\eta : Y \to \mathbb{R}$ continuous, then $\eta \circ f : X \to Y$ is a measurable function in the usual sense. In particular, if $f, g : X \to Y$ are measurable, then

$$d(f(x), g(x))$$

is a measurable function on X.

This observation will now enable us to define L^2-spaces of mappings. Namely, for measurable maps $f, g : X \to Y$, we may define the L^2 - distances as

$$d(f,g) := d_{L^2}(f,g) := \left(\int d^2(f(x), g(x))\, \mu(dx) \right)^{1/2}.$$

Of course, this distance may possibly be infinite for certain maps f, g.

We now make a **digression** that can be omitted at a first reading, in particular by readers who are more interested in the analytic concepts involved than in the geometric applications. For certain applications this notion needs to be redefined slightly. For example, if Y happens to be a strong geodesic length space, one may wish to prescribe a homotopy class for the maps f, g. If f and g are in the same homotopy class, i.e. if there exists a map

$$F : X \times [0,1] \to Y$$

with $F(x,0) = f(x)$, $F(x,1) = g(x)$ and for which $F(x,\cdot)$ is continuous for all $x \in X$, we may consider the homotopy distance between f and g that is defined as follows:
We put

$$d_h(f(x), g(x)) := \inf \{ L(\gamma)\colon \gamma\colon [0,1] \to Y \text{ curve with } \gamma(0) = f(x), \gamma(1) = g(x), \gamma(\cdot) \text{ homotopic to } F(x,\cdot) \text{ with fixed endpoints} \}$$

and

$$d_h(f,g) := \left(\int_X d_h^2\,(f(x), g(x))\, \mu(dx) \right)^{1/2}.$$

Obviously, this homotopy distance does not depend on the particular choice of the homotopy F. If X and Y are manifolds, or more generally if they are amenable to covering space theory, this homotopy distance may be expressed in the following equivalent form. If $\pi_X : \tilde{X} \to X$, $\pi_Y : \tilde{Y} \to Y$ are the universal coverings of X and Y, respectively, we lift $f : X \to Y$ to $\tilde{f} : \tilde{X} \to \tilde{Y}$. Thus we have

$$\pi_Y \circ \tilde{f} = f \circ \pi_X.$$

Given homotopic maps $f, g : X \to Y$, we choose compatible lifts $\tilde{f}, \tilde{g} : X \to Y$, i.e. we choose a lift $\tilde{F}$ of the homotopy F between f and g, (same notations as above) and put $\tilde{f} = \tilde{F}(\cdot, 0)$, $\tilde{g} = \tilde{F}(\cdot, 1)$. Then

$$d_h(f,g) = \left(\int_X d^2(\tilde{f}(x), \tilde{g}(x))\, \mu(dx) \right)^{1/2}$$

where the integration now is over a fundamental domain for X in $\tilde{X}$, and where we have lifted the measure μ to $\tilde{X}$ in such a manner that it is invariant under all covering transformations. More generally, we may assume that X admits a cover in the following sense: There exists a measure space $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$ and a group Γ of measure preserving homomorphisms of $\tilde{X}$ such that $X = \tilde{X}/\Gamma$ and that $\tilde{\mathcal{B}}, \tilde{\mu}$ induce $\mathcal{B}, \mu$ on X under the projection $\pi_X : \tilde{X} \to X = \tilde{X}/\Gamma$. We also require that Y admits a cover $\tilde{Y}$ for which every map $f : X \to Y$ lifts to a map $\tilde{f} : \tilde{X} \to \tilde{Y}$ as above.

We then suppose that we are given a homomorphism

$$\rho : \Gamma \to I(\tilde{Y})$$

from Γ into the isometry group of $\tilde{Y}$ (Note that the metric d of Y lifts to a metric on $\tilde{Y}$ that will again be denoted by d.). $f : \tilde{X} \to \tilde{Y}$ is called ρ-equivariant if for all $x \in \tilde{X}$, $\gamma \in \Gamma$

$$f(\gamma x) = \rho(\gamma) f(x).$$

If $f, g : \tilde{X} \to \tilde{Y}$ are ρ-equivariant, we put

$$d(f, g) := \left(\int_X d^2(f(x), g(x)) \tilde{\mu}(dx) \right)^{1/2}.$$

Again, the integration is over a fundamental domain F_X for X in $\tilde{X}$, i.e. over a maximal set of points in $\tilde{X}$ that are mutually Γ-inequivalent, i.e. given $x, y \in F_X$, there is no $\gamma \in \Gamma$ with $\gamma(x) = y$.

This homotopy distance will be used subsequently without further explicit mentioning.
This **ends** the geometric **digression**.

In order to define an L^2-space for maps from X into a metric space (Y, d), we need to select a μ-measurable base map $f_0 : X \to Y$. In case (Y, d) is $\mathbb{R}$ or $\mathbb{R}^n$ with its standard Euclidean metric, the canonical base map is $f_0 \equiv 0$. In the case of a general Y, one might then take a constant map $f_0 \equiv p$ for some $p \in Y$, but one should take the following two comments into consideration:

1) If X is noncompact, the subsequent construction may depend on the choice of p, as happens already for $X = Y = \mathbb{R}$, and in general there will be no canonical choice for p.

2) In view of the above digression, one may wish to fix a homotopy class of maps from X to Y, and it will then be natural to require that f_0 also belongs to the chosen homotopy class.

In any case, having selected f_0, we put

$$\begin{aligned} L^2(X,Y) &:= L^2(X,\mu;Y,d) := L^2(X,\mu;Y,d;f_0) \\ &:= \{f : X \to Y \ \mu - \text{measurable } : d(f,f_0) < \infty\}. \end{aligned}$$

In the sequel, we shall identify maps that differ only on a set of measure 0, as is customary in the theory of L^p-spaces. Since (Y,d) is a complete metric space, $L^2(X,\mu;Y,d)$ then also becomes a complete metric space. This means that any Cauchy sequence $(f_n)_{n\in\mathbb{N}} \subset L^2(X,Y)$, i.e. any sequence of maps $f_n : X \to Y$ satisfying

$$\begin{gathered} \forall \varepsilon > 0 \; \exists N \in \mathbb{N} \; \forall n,m \geq \varepsilon \\ \int_X d^2(f_n(x), f_m(x))\, \mu(dx) < \varepsilon \end{gathered}$$

converges to some $f \in L^2(X,Y)$, i.e.

$$\lim_{n\to\infty} \int_X d^2(f_n(x), f(x))\, \mu(dx) = 0.$$

Let $(X,\mathcal{B},\mu)$ be given as before. The basic object of study are functionals E defined on

$$\{f \in L^2(X,\mu;Y,d;f_0) \quad \text{for some metric space } (Y,d) \text{ and some base map } f_0 : X \to Y\}$$

and taking values in $\mathbb{R}^+ \cup \{\infty\}$. The letter E here stands for energy, and the original motivation for the theory was to find an axiomatic version for the energy functional maps between Riemannian manifolds that leads to the theory of harmonic mappings.

Given such a functional E, we define $D(E)$ as its domain of definition, i.e. as the set of those f with

$$E(f) < \infty.$$

We point out that while the measure space (X,μ) is fixed, the image or target space (Y,d) is kept variable. For certain purposes, however, one may wish to restrict the class of permitted target spaces (Y,d). Now comes the fundamental

Definition 1.1 *A functional E as defined above with domain of definition $D(E)$ is called a generalized Dirichlet form or energy form if the following three conditions hold:*

1) *Quadratic contraction property:*
Let $f : X \to Y$ *be in* $D(E)$, $\varphi : (f(X), d) \to (Z, d')$ *for some metric space* (Z, d') *(here* $f(X)$ *carries the metric induced from the metric* d *on* Y*) an L-Lipschitz function, i.e.*

$$d'(\varphi(x), \varphi(y)) \leq L\, d(x, y) \text{ for all } x, y \in f(X).$$

Then also $\varphi \circ f \in D(E)$, *and*

$$E(\varphi \circ f) \leq L^2\, E(f).$$

2) *Density:*
$D(E) \cap L^2(X, \mu; \mathbb{R})$ *is dense in* $L^2(X, \mu; \mathbb{R}) = L^2(X, \mu)$. *Furthermore, if for some* $f \in L^2(X, Y)$,

$$\varphi \circ f \in D(E)$$

for every 1-Lipschitz map $\varphi : (Y, d) \to \mathbb{R}$, *then also* $f \in D(E)$.

3) *Closedness:*
E *is lower semicontinuous w.r.t. convergence in* $L^2(X, \mu; Y, d)$.

Remark: If $f \in L^2(X, \mu; Y, d; f_0), \varphi : (f(X), d) \to (Z, d')$ L-Lipschitz, then $\varphi \circ f \in L^2(X, \mu; Z, d'; \varphi \circ f_0)$. Thus, it is useful to allow the base map to vary.

Let E be an energy form. In order to understand how our definition is related to the standard definition of a Dirichlet form as given e.g. in [FOT], we restrict our E to $L^2(X, \mu; \mathbb{R})$, i.e. to $Y = \mathbb{R}$. We start with the following simple observation.

Lemma 1.1 *If* $f : X \to \mathbb{R}$ *is in* $D(E)$, $a, \lambda \in \mathbb{R}$, *then*

$$E(\lambda f) = \lambda^2\, E(f)$$
$$E(f + a) = E(f)$$

Proof: A simple consequence of property 1), considering the Lipschitz functions

$$\varphi_\lambda : \mathbb{R} \to \mathbb{R}, \quad \varphi_\lambda(x) : \lambda x, \quad \varphi_\lambda^{-1}(x) = \varphi_1(x) = \frac{1}{\lambda} x$$
$$\varphi_a : \mathbb{R} \to \mathbb{R}, \quad \varphi_a(x) : x + a, \quad \varphi_a^{-1}(x) = \varphi_{-a}(x) = x - a.$$

□

By polarization

$$E(f, g) := \frac{1}{4}(E(f + g) - E(f - g))$$

we obtain a Dirichlet form in the usual sense. The closedness and density conditions reduce to the standard ones, but the quadratic contraction property is stronger than the one usually required for Dirichlet forms as can be seen from Lemma 1. Namely, there one requires that

$$E(g,g) \le E(f,f)$$

whenever $|g(x)-g(y)| \le |f(x)-f(y)|$ and $|g(x)| \le |f(x)|$ for all $x,y \in X$. A Dirichlet form $E(\cdot,\cdot)$ is called regular if $D(E) \cap C_0^0(X)$ (where $C_0^0(x)$ is the space of continuous functions with compact support) is dense in $C_0^0(X)$ w.r.t. the C^0-norm as well as dense in $D(E)$ w.r.t. the norm $\|\cdot\|_1$ induced by the product

$$(f,g)_1 := (f,g) + E(f,g)$$

(where $(\cdot,\cdot)$ denotes the L^2-product).

$E(\cdot,\cdot)$ is called strongly local if

$$E(f,g) = 0$$

whenever f is constant on some neighborhood of the support of g (or vice versa).

With these definitions, we may state the **representation theorem of Beurling-Deny** (see e.g. [FOT]): *Suppose that X is a locally compact, separable metric space and μ is a positive Radon measure with supp(μ) $= X$. (This means that μ is a non-negative Borel measure on X that is finite on compact sets and strictly positive on non-empty open sets.) Then a regular Dirichlet form E on X admits the following decompositions: For all $f,g \in D(X) \cap C_0^0(X)$*

$$\begin{aligned} E(f,g) &= E^c(f,g) + \int_{X\times X\setminus diagonal} (f(x)-f(y))\,(g(x)-g(y))\,J(dx,dy) \\ &\quad + \int_X f(x)g(x)\,k(dx) \end{aligned}$$

with a strongly local Dirichlet form E^c, a symmetric positive Radon measure J on $X \times X \setminus$ diagonal and a positive Radon measure k on X, the so-called killing measure.

It is clear from the next lemma that our quadratic contraction property implies that for a Dirichlet form that comes from an energy form in our sense, the killing measure has to vanish.

Lemma 1.2 *Let E be an energy form.*

1) If $\varphi : (Y,d) \to (Y,d)$ is an isometry, then for every $f \in D(E)$, $f : X \to Y$

$$E(\varphi \circ f) = E(f).$$

2) Every constant map $f_0 : X \to Y$ *satisfies*

$$E(f_0) = 0.$$

Proof:

1) If φ is an isometry, so is φ^{-1}, and φ and φ^{-1} both are 1-Lipschitz. Thus, by property 1)

$$E(\varphi \circ f) \leq E(f) = E(\varphi^{-1} \circ \varphi \circ f) \leq E(\varphi \circ f),$$

and equality has to hold.

2) It is clear from property 1) that $g_0 \equiv 0$ satisfies

$$E(g_0) = 0.$$

Next, by 2) of Lemma 1.1,

$$E(h_0) = 0 \text{ for any constant function } h_0 : X \to \mathbb{R}$$

Finally, using properties 1) and 2) gives the result for constant maps f_0.

Remark: If $\mu(X) = \infty$, constant functions $\neq 0$ are not in $L^2(X, \mu; \mathbb{R})$. Thus, in that case, the reader needs to interprete the preceding statements appropriately.

2 Resolvents, semigroups, and variational aspects

In the classical theory of Dirichlet forms, one has a correspondance between such forms, semigroups, resolvents and selfadjoint operators satisfying suitable conditions that we now summarize. For that purpose, one may take any (real) Hilbert space H, not necessarily $L^2(X, \mu)$. The norm and scalar product on H will be denoted by $\|\cdot\|$, $(\cdot,\cdot)$. A reference for the sequel is [FOT].

Definition 2.1 *A family* $(T_t)_{t>0}$ *of symmetric linear operators defined on all of* H *is called a semigroup if*

(i) $T_t \cdot T_s = T_{t+s}$ *for all* $t, s > 0$

(ii) $\|T_t u\| \leq \|u\| \quad u \in H, t > 0.$

Such a semigroup is called strongly continuous if *(iii)* $\lim_{t\searrow 0} T_t u = u$ *for all* $u \in H$

Definition 2.2 *A family* $(G_\alpha)_{\alpha>0}$ *of symmetric linear operators defined on all of* H *is called a resolvent if*

(i) $G_\alpha - G_\beta + (\alpha-\beta)G_\alpha G_\beta = 0$ *for all* $\alpha, \beta > 0$

(ii) $\|\alpha G_\alpha u\| \le \|u\|$ *for all* $u \in H, \alpha > 0$.

Such a resolvent is called strongly continuous if in addition

(iii) $\lim_{\alpha\to\infty} \alpha G_\alpha u = u$ *for all* $u \in H$.

The following is a basic result in the theory of Dirichlet forms, see [BH],[FOT], or [MR].

Theorem *The following classes of objects correspond to each other:*

(i) *Closed symmetric forms on* H.

(ii) *Non-positive definite selfadjoint operators on* H.

(iii) *Strongly continuous semigroups.*

(iv) *Strongly continuous resolvents.*

This correspondance goes as follows.

A closed symmetric form E corresponds to a non-positive definite self-adjoint operator L via

$$\begin{aligned} D(E) &= D(\sqrt{-L}) \\ E(u,v) &= (\sqrt{-L}u, \sqrt{-L}v) \text{ for all } u, v \in D(E). \end{aligned}$$

If $u \in D(L), v \in H$, this becomes

$$E(u,v) = -(Lu, v).$$

A strongly continuous semigroup $(T_t)_{t>0}$ generates a non-positive selfadjoint operator L as its generator

$$Lu = \lim_{t\searrow 0} \frac{1}{t}(T_t u - u)$$

with $D(L) = \{u \in H$ for which this limit exists$\}$.

Conversely, given such an L, a strongly continuous semigroup is generated as

$$T_t = \exp tL = \lim_{n\to\infty} (Id - \frac{t}{n}L)^{-n}.$$

(The limit exists because for all $\lambda > 0$, $\|(Id - \lambda L)^{-1}\| \le 1$ as L is non-positive.) A strongly continuous resolvent generates a non-positive selfadjoint operator L via

$$\frac{1}{\alpha}Lu = u - \frac{1}{\alpha}G_\alpha^{-1}u \qquad \text{(this does not depend on } \alpha > 0\text{)}$$
$$D(L) = G_\alpha(H)$$

and the converse relation is

$$\alpha G_\alpha = (Id - \frac{1}{\alpha}L)^{-1} \text{ for } \alpha > 0.$$

The resolvent can also be characterized by the fact that

$$J_\lambda u := \frac{1}{\lambda}G_{1/\lambda}u = (Id - \lambda L)^{-1}$$

is the unique minimizer of $\lambda E(v) + \|u - v\|^2$, i.e.

$$\lambda E(J_\lambda u) + \|u - J_\lambda u\|^2 = \inf_{v \in H}(\lambda E(v) + \|u - v\|^2).$$

The semigroup may be obtained from the family $(J_\lambda)_{\lambda > 0}$ via

$$T_t u = \lim_{n \to \infty} J_{\frac{t}{n}}^n x.$$

as follows from the preceding formulae.

The result that a closed densely defined linear operator L on a Banach space B that is non-positive, or, more generally, satisfies

$$\|(Id - \lambda L)^{-1}\| \le 1 \text{ for all } \lambda > 0$$

defines a semigroup via

$$T_t u = \lim_{n \to \infty}(Id - \frac{t}{n}L)^{-n}u$$

and furthermore that $u(t) := T_t u$ solves the Cauchy problem

$$\frac{du}{dt} = Lu, \; u(0) = v \qquad \text{for } v \in D(L)$$

is the Hille-Yosida theorem. Crandall-Liggett [CL] then showed that this result extends to the case where L is a nonlinear operator on the Banach space B, if the other assumptions continue to hold.

In the case of $H = L^2(X, \mu)$, a Dirichlet form was a symmetric closed form that satisfies a contraction property. In terms of the corresponding semigroup or resolvent, this is characterized by the so-called Markovian property that the T_t or G_α have to satisfy. A linear operator S defined on all of $L^2(X, \mu)$ is called Markovian if

$$0 \leq u \leq 1 \quad \mu - \text{ a.e. } \Rightarrow 0 \leq Su \leq 1 \quad \mu - \text{ a.e for all } u \in L^2(X, \mu).$$

It will turn out that a substantial part of this correspondance extends to the case of energy forms at least if we impose an additional condition on the target space (Y, D), namely to be non-positively curved in the sense of Alexandrov.

Remark: For many applications, it is important to realize that there is a still more general condition than Alexandrov's, namely Busemann's nonpositive curvature condition under which the theory still works. For example, the L^p spaces are nonpositively curved in the sense of Busemann for $1 < p < \infty$, but not in the sense of Alexandrov unless $p = 2$. The resulting theory is developed in [J6]. Since Busemann's condition is not as easy to work with as Alexandrov's, in the present notes we restrict our attention to the latter.

We start with some general definitions. Let Z be simply connected complete geodesic length space. In our applications below, we shall use

$$Z = L^2(X, \mu; Y, d) \tag{2.1}$$

equipped with the L^2-metric that we shall also denote by d. Here, we assume that (Y, d) itself is a simply connected complete geodesic length space. Points in Z will be denoted by x, y, z, also in the context of (2.1), where these would correspond to L^2 - maps that should rather be called f, g in accordance with the terminology of §1.

Let $D(F) \subset Z$, and let $F : D(F) \to \mathbb{R}$ be a functional. We say that F is densely defined if $D(F)$ is dense in Z. We say that F is convex if whenever $\gamma : [0, 1] \to Z$ is geodesic (parametrized proportionally to arclength, as always), and if $\gamma(0), \gamma(1) \in D(F)$, then also $\gamma(t) \in D(F)$ and

$$F(\gamma(t)) \leq t\, F(\gamma(0)) + (1 - t) F(\gamma(1)) \tag{2.2}$$

for all $0 \leq t \leq 1$.

We extend any convex functional $F : D(F) \to \mathbb{R}$ to a functional

$$F : Z \to \mathbb{R} \cup \{\infty\}$$

by putting

$$F(x) = \infty \quad \text{ if } x \in Z \setminus D(F).$$

This extension still satisfies (2.2) as the left hand side can only assume the value ∞ if the right hand side does. Consequently, any functional $F : Z \to \mathbb{R} \cup \{\infty\}$ satisfying (2.2) is called convex.

Definition 2.3 *Let $F : Z \to \mathbb{R} \cup \{\infty\}$. The Moreau-Yosida approximation F^λ of F is*

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

From now on, we shall make the following general
Assumption: Z is a complete global NPC space.

Lemma 2.1 *Let $F : Z \to \mathbb{R}_+ \cup \{\infty\}$ be convex, $\not\equiv \infty$, and lower semicontinuous. Then for every $x \in Z$, and every $\lambda > 0$, there exists a unique $y_\lambda = J_\lambda(x)$ with*

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

Proof: Let $(y_n)_{n\in\mathbb{N}}$ be a minimizing sequence for F^λ, i.e.

$$\lambda F(y_n) + d^2(x, y_n) \to \inf_{y \in Z}(\lambda F(y) + d^2(x, y)) =: \kappa_\lambda.$$

For $m, n \in \mathbb{N}$, we let $y_{m,n}$ be the midpoint of y_m and y_n. The convexity of F implies

$$F(y_{m,n}) \le \frac{1}{2}F(y_m) + \frac{1}{2}F(y_n)$$

and the NPC inequality gives

$$d^2(x, y_{m,n}) \le \frac{1}{2}(d^2(x, y_m) + d^2(x, y_n)) - \frac{1}{4}d^2(y_m, y_n).$$

Together, we get

$$\begin{aligned} \lambda F(y_{m,n}) + d^2(x, y_{m,n}) &\le \frac{1}{2}(\lambda F(y_m) + d^2(x, y_m)) \\ &+ \frac{1}{2}(\lambda F(y_n) + d^2(x, y_n)) \\ &- \frac{1}{4}d^2(y_m, y_n). \end{aligned}$$

From the definition of κ_λ, the left hand side cannot be smaller then κ_λ, while from the definition of (y_n), the first two terms on the right hand side both tend to $\frac{1}{2}\kappa_\lambda$. It follows that $(y_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. It thus has a uniquely determined limit point y_λ which must realize the above infimum as $d^2(x, \cdot)$ is continuous and F is lower semicontinuous by assumption. □

Lemma 2.2 *Assumptions as in the previous lemma. For* $x \in Z$*, let* $y_\lambda = J_\lambda(x)$ *as before. If* $x \in \overline{D(F)}$*, then*

$$x = \lim_{\lambda \to 0} J_\lambda(x).$$

In particular, if F *is densely defined, i.e.* $\overline{D(F)} = Z$*, this holds for all* $x \in Z$*.*

Proof: Since x is assumed to be in the closure of $D(F)$, for every $\rho > 0$, we may find $x_\rho \in B(x, \rho)$ with

$$F(x_\rho) < \infty.$$

Thus

$$\lim_{\lambda \to \infty} (\lambda F(x_\rho) + d^2(x, x_\rho)) \leq \rho^2,$$

and consequently (recall $\kappa_\lambda := \inf_{y \in Z}(\lambda F(y) + d^2(x, y))$)

$$\limsup_{\lambda \to 0} \kappa_\lambda \leq 0. \tag{2.3}$$

We shall prove the result by deriving a contradiction from the assumption that for some sequence $(\lambda_n)_{n \in \lambda}$ with $\lambda_n \to 0$ as $n \to \infty$

$$d^2(x_n, y_{\lambda_n}) \geq \beta > 0 \quad \text{for all } n.$$

Then, because of

$$\limsup_{n \to \infty} (\lambda_n F(y_{\lambda_n}) + d^2(x, y_{\lambda_n})) \leq 0 \quad \text{by (2.3)},$$

$$F(y_{\lambda_n}) \to -\infty \quad \text{for } n \to \infty$$

contradicting our assumption that F is nonnegative. □

We may now derive the main existence result of [J3].

Theorem 2.1 *As always in this §, we assume that* Z *is a complete NPC space, and that* $F : Z \to \mathbb{R}_+ \cup \{\infty\}$ *is convex,* $\not\equiv \infty$*, and lower semicontinuous. For* $x \in Z$ *and* $\lambda > 0$*, let* $y_\lambda = J_\lambda(x)$ *be as in Lemma 2.1. If* $(y_{\lambda_n})_{n \in \mathbb{N}}$ *is bounded for some sequence* $\lambda_n \to \infty$*, then* $(y_\lambda)_{\lambda > 0}$ *converges to a minimizer of* F *for* $\lambda \to \infty$*.*

Proof: The proof will be divided into six simple steps.

1) y_{λ_n} minimizes $F(y) + \frac{1}{\lambda_n} d^2(x, y)$. Since (y_{λ_n}) is bounded, it therefore is a minimizing sequence for F.

2) Let $0 < \mu_1 < \mu_2$. The minimizing property of y_{μ_1} implies

$$F(y_{\mu_1}) + \frac{1}{\mu_1} d^2(x, y_{\mu_1}) \leq F(y_{\mu_2}) + \frac{1}{\mu_1} d^2(x, y_{\mu_2}),$$

hence

$$\begin{aligned} F(y_{\mu_1}) \quad &+ \quad \frac{1}{\mu_2} d^2(x, y_{\mu_2}) + (\frac{1}{\mu_1} - \frac{1}{\mu_2})(d^2(x, y_{\mu_1}) - d^2(x, y_{\mu_2})) \\ &\leq \quad F(y_{\mu_2}) + \frac{1}{\mu_2} d^2(x, y_{\mu_2}). \end{aligned}$$

The minimizing property of y_{μ_2} therefore implies

$$d^2(x, y_{\mu_1}) \leq d^2(x, y_{\mu_2}).$$

Thus, $d^2(x, y_\lambda)$ is a monotonically increasing function of λ.

3) Since $d^2(x, y_\lambda)$ is monotonically increasing by 2) and bounded on the sequence (y_{λ_n}) by assumption, it follows that $d^2(x, y_\lambda)$ is bounded as $\lambda \to \infty$.

4) From the definition of y_λ,

$$F(y_\lambda) = \inf\{F(y) : d^2(x, y) \leq d^2(x, y_\lambda)\}.$$

Since $d^2(x, y_\lambda)$ is nondecreasing by 2), it follows that $F(y_\lambda)$ is nonincreasing in λ, and

$$\lim_{\lambda \to \infty} F(y_\lambda) = \inf_{y \in Z} F(y) \quad \text{by 1)}.$$

5) Let $\varepsilon > 0$. We choose Λ so large that for $\lambda \geq \mu \geq \Lambda$

$$d^2(x, y_\lambda) - d^2(x, y_\mu) < \varepsilon/2 \tag{2.4}$$

which is possible by 2), 3).
By 4)

$$F(y_\lambda) \leq F(y_\mu). \tag{2.5}$$

We let $y_{\lambda,\mu}$ be the mean value of y_λ and y_μ. Then the convexity of F, (2.5), and the NPC condition imply

$$\begin{aligned} & F(y_{\lambda,\mu}) + \frac{1}{\mu} d^2(x, y_{\lambda,\mu}) \\ \leq \quad & F(y_\mu) + \frac{1}{\mu}(\frac{1}{2} d^2(x, y_\mu) + \frac{1}{2} d^2(x, y_\lambda) - \frac{1}{4} d^2(y_\lambda, y_\mu)) \\ < \quad & F(y_\mu) + \frac{1}{\mu}(d^2(x, y_\mu) + \varepsilon/4 - \frac{1}{4} d^2(y_\lambda, y_\mu)) \quad \text{by (2.4)}. \end{aligned}$$

The minimizing property of y_μ then implies

$$d^2(y_\lambda, y_\mu) < \varepsilon.$$

Thus, $(y_\lambda)_{\lambda>0}$ satisfies the Cauchy property for $\lambda \to \infty$.

6) Since Z is complete, (y_λ) therefore converges to some limit y_∞. 4) and the lower semicontinuity of F imply that y_∞ minimizes F. □

In the classical case, where $Z = L^2(x, \mu)$ and F is a Dirichlet form, $J_\lambda(x)$ equals $\lambda G_\lambda x$ where $(G_\lambda)_{\lambda>0}$ is the resolvent assiciated with F. In this light, one should consider the following **resolvent identity**.

Lemma 2.3 *Let $F : Z \to \mathbb{R}_+ \cup \{\infty\}$ be a function. Then the resolvent equation*

$$\frac{1}{\mu}(\frac{1}{\lambda}F^\lambda)^\mu = \frac{1}{\lambda+\mu}F^{\lambda+\mu}$$

holds.

Proof:

$$\begin{aligned}\frac{1}{\mu}(\frac{1}{\lambda}F^\lambda)^\mu(x) &= \inf_{y\in Z}(\frac{1}{\lambda}F^\lambda(y) + \frac{1}{\mu}d^2(x,y)) \\ &= \inf_{y\in Z}(\inf_{z\in Z}(F(z) + \frac{1}{\lambda}d^2(y,z) + \frac{1}{\mu}d^2(x,y)).\end{aligned}$$

For each $z \in Z$,

$$\inf_{y\in Z}(\frac{1}{\lambda}d^2(y,z) + \frac{1}{\mu}d^2(x,y))$$

is realized by a unique point y_0, namely the point on the geodesic arc from x to z with

$$d(x, y_0) = \frac{\mu}{\lambda+\mu}d(x,z),\ d(z,y_0) = \frac{\lambda}{\lambda+\mu}d(x,z).$$

y_0 thus satisfies

$$\frac{1}{\lambda}d^2(y_0,z) - \frac{1}{\mu}d^2(x,y_0) = \frac{1}{\lambda+\mu}d^2(x,z)$$

and

$$\frac{1}{\mu}(\frac{1}{\mu}F^\lambda)^\mu(x) = \inf_{z\in Z}(F(z) + \frac{1}{\lambda+\mu}d^2(x,z)) = \frac{1}{\lambda+\mu}F^{\lambda+\mu}(x).$$

□

Another version of the **resolvent identity** is

Corollary 2.1 *Let $F : Z \to \mathbb{R}_+ \cup \{\infty\}$ be convex, $\not\equiv \infty$, lower semicontinuous. Let $x \in Z, \lambda > 0, y_\lambda = J_\lambda(x)$ as in Lemma 2.1. Let $\gamma : [0,1] \to Z$ be the geodesic arc (as always, parametrized proportionally to arclength) with $\gamma(0) = x, \gamma(1) = y_2$, and put*

$$y_{\lambda,s} := \gamma(s).$$

Then

$$J_{(1-s)\lambda}(y_{\lambda,s}) = y_\lambda.$$

Proof: This follows from the proof of Lemma 2.3 if one uses the uniqueness result of Lemma 2.1. □

We continue to derive some further properties of F^λ and J_λ.

Lemma 2.4 *Let F be as in Theorem 2.1, $x \in D(F)$, $0 < \mu \leq \lambda$. Then*

$$F(J_\mu x) \leq \left(1 - \frac{d(x, J_\mu x)}{d(x, J_\lambda x)}\right) F(x) + \frac{d(x, J_\mu x)}{d(x, J_\lambda x)} = F(J_\lambda x).$$

Proof: By 2) in the proof of Theorem 2.1,

$$d^2(x, J_\mu x) \leq d^2(x, J_\lambda x).$$

Let x_μ be the point on the geodesic form x to $J_\lambda x$ with $d(x, x_\mu) = d(x, J_\mu x)$. Then by the minimuing property of $J_\mu x$,

$$\begin{aligned} F(J_\lambda x) &\leq F(x_\mu) \\ &\leq (1 - \frac{d(x, x_\mu)}{d(x, J_\lambda x)})F(x) + \frac{d(x, x_\mu)}{d(x, J_\lambda x)}F(J_\lambda x) \end{aligned}$$

by the convexity of F. □

Lemma 2.5 *Let F be as in Theorem 2.1. For any x_1, $x_2 \in Z$, $\lambda > 0$,*

$$d(J_\lambda x_1, J_\lambda x_2) \leq d(x_1, x_2).$$

Thus J_λ is a 1-Lipschitz map.

Proof: Put $y_i := J_\lambda x_i, i = 1, 2$, and let $\gamma : [0,1] \to Z$ be the shortest geodesic with $\gamma(0) = y_1, \gamma(1) = y_2$. Since $F(y_1)$ and $F(y_2)$ are finite and F is convex, F is finite on γ, and we have

$$F(\gamma(t)) + F(\gamma(1-t)) \leq F(y_1) + F(y_2) \text{ for } 0 \leq t \leq 1. \tag{2.6}$$

By (G.2) of the appendix,

$$\begin{array}{rcl} d^2(\gamma(t),x_1)+d^2(\gamma(1-t),x_2) & \leq & d^2(x_1,y_1)+d^2(x_2,y_2) \\ & + & td^2(x_1,x_2)-td^2(y_1,y_2) \\ & + & 2t^2d^2(y_1,y_2) \\ & - & t(d(x_1,x_2)-d(y_1,y_2))^2. \end{array}$$

If we had

$$d(y_1,y_2) > d(x_1,x_2), \tag{2.7}$$

then for small $t > 0$,

$$d^2(\gamma(t),x_1)+d^2(\gamma(1-t),x_2) < d^2(x_1,y_1)+d^2(x_2,y_2), \tag{2.8}$$

and together with (2.6) either

$$\lambda F(\gamma(t))+d^2(\gamma(t),x_1) < \lambda F(y_1)+d^2(y_1,x_2) \text{ or}$$
$$\lambda F(\gamma(1-t))+d^2(\gamma(1-t),x_2) < \lambda F(y_2)+d^2(y_2,x_2).$$

This, however, would contradict the minimizing property of y_1 or y_2. Thus, (2.7) cannot hold. □

We now wish to construct a strongly continuous semigroup $(T_t)_{t>0}$ with the help of the resolvents. More precisely, we shall extend the Crandall - Liggett theorem [CL] to the present situation.

Let $x \in R(J_\lambda)$ for some $\lambda > 0$, i.e.

$$x = J_\lambda z \text{ for some } z \in Z \tag{2.9}$$

Then also $x \in R(J_\mu)$ whenever $0 < \mu \leq \lambda$, namely

$$x = J_\mu(m_{\frac{\mu}{\lambda}}(z,x)) \quad \text{by Cor. 2.1.} \tag{2.10}$$

Also

$$\frac{d(x,z)}{\lambda} = \frac{d(x,m_{\frac{\mu}{\lambda}}(z,x))}{\mu}$$

and we denote this value by

$$l_F(x).$$

Note that $l_F(x)$ may depend on the choice of z with $x = J_\lambda z$ above. If $\lambda \geq \mu > 0$, however, we see from (2.10) that F having fixed λ and z, $l_F(x)$ does not depend on μ anymore. In order to eliminate the dependence on z, we may take the infimum over all z and $\lambda > 0$ satisfying (2.9). For $x \in \overline{D(F)}$, we may then define

$$l_F(x) = \lim_{n\to\infty} l_F(J_{\frac{1}{n}}x)$$

because $J_{\frac{1}{n}}x \in R(J_{\frac{1}{n}})$ and $\lim_{n\to\infty} J_{\frac{1}{n}}x = x$ for $x \in \overline{D(F)}$ by Lemma 2.2. Note that $l_F(x)$ may be infinite for some $x \in \overline{D(F)}$.

Lemma 2.6 *Let $x \in R(J_\lambda)$ for some $\lambda > 0$. Then*

$$d(x, J_\lambda x) \leq \lambda l_F(x).$$

Proof: Let $x = J_\lambda z$ as above. Then

$$\begin{aligned} d(x, J_\lambda x) = d(J_\lambda z, J_\lambda(J_\lambda z)) &\leq d(z, J_\lambda z) \text{ by Lemma 2.5} \\ &= d(x,z) = \lambda l_F(x). \end{aligned}$$

□

Lemma 2.7 *Let $\lambda \geq \mu > 0, m, n \in \mathbb{Z}$. Then for $x \in D(F) \cap R(J_\lambda)$*

(2.11) $$d(J_\mu^n x, J_\lambda^m x) \leq (\sqrt{(m\lambda - n\mu)^2 + m\lambda^2} + \sqrt{(m\lambda - n\mu)^2 + \lambda\mu n})l_F(x).$$

Proof: By Cor. 2.1

$$\begin{aligned} a_{n,m} := d(J_\mu^n x, J_\lambda^m x) &= d(J_\mu^n x, J_\mu(m_{\frac{\mu}{\lambda}}(J_\lambda^{m-1}x, J_\lambda^m x)) \\ &\leq d(J_\mu^{n-1}x, m_{\frac{\mu}{\lambda}}(J_\lambda^{m-1}x, J_\lambda^m x)) \text{ by Lemma 2.5} \\ &\leq \frac{\mu}{\lambda} d(J_\mu^{n-1}x, J_\lambda^{m-1}x) + \frac{\lambda-\mu}{\lambda} d(J_\mu^{n-1}x, J_\lambda^m x) \\ &\qquad \text{by nonpositive curvature (see (G.1)).} \end{aligned}$$

With $\alpha = \frac{\mu}{\lambda}, \beta = \frac{\lambda-\mu}{\lambda}$, we thus have

(2.12) $$a_{n,m} \leq \alpha a_{n-1,m-1} + \beta a_{n-1,m}$$

Using (2.12), we now want to prove (2.11) by induction. First

$$\begin{aligned} a_{0,m} = d(x, J_\lambda^m x) &\leq \sum_{i=0}^{m-1} d(J_\lambda^i x, J_\lambda^{i+1}x) \text{ by the triangle inequality} \\ &\leq m d(x, J_\lambda x) \text{ by Lemma 2.5} \end{aligned}$$

(2.13) $$\leq m\lambda l_F(x) \text{ by Lemma 2.6.}$$

Likewise

$$a_{n,0} \leq n\mu l_F(x).$$

We shall show that if (2.12) holds for the pairs (n, m) and $(n, m-1)$ then it also holds for $(n+1, m)$. Induction then yields the claim. From (2.12)

$$
\begin{aligned}
a_{n+1,m} &\leq \alpha a_{n,m-1} + \beta a_{n,m} \\
&\leq \Big\{\alpha\left(\sqrt{((m-1)\lambda - n\mu)^2 + (m-1)\lambda^2} + \sqrt{((m-1)\lambda - n\mu)^2 + \lambda\mu n}\right) \\
&\quad + \beta\left(\sqrt{(m\lambda - n\mu)^2 + m\lambda^2} + \sqrt{(m\lambda - n\mu)^2 + \lambda\mu m}\right)\Big\}\, l_F(x) \\
&\quad \text{by our inductive hypothesis} \\
&\leq \Big(\sqrt{\alpha\{((m-1)\lambda - m\mu)^2 + (m-1)\lambda^2\} + \beta\{(m\lambda - n\mu)^2 + m\lambda^2\}} \\
&\quad + \sqrt{\alpha\{((m-1)\lambda - n\mu)^2 + \lambda\mu n\} + \beta\{(m\lambda - n\mu)^2 + \lambda\mu n\}}\Big)\, l_F(x) \\
&\quad \text{by the Schwarz inequality, noting that } \alpha + \beta = 1 \\
&\leq \left(\sqrt{(m\lambda - (n+1)\mu)^2 + m\lambda^2} + \sqrt{(m\lambda - (n+1)\mu)^2 + (n+1)\lambda\mu}\right) l_F(x) \\
&\quad \text{noting that } \alpha = \frac{\mu}{\lambda} \text{ and } \alpha + \beta = 1.
\end{aligned}
$$

This is (2.12) for the pair $(n+1, m)$. □

Theorem 2.2 *Let $F : Z \to \mathbb{R}_+ \cup \{\infty\}$ be a convex lower semicontinuous functional, $\not\equiv \infty$. Then*

$$
T_t x := \lim_{n\to\infty} J^n_{\frac{t}{n}} x \quad t \geq 0, x \in \overline{D(F)} \tag{2.14}
$$

defines a strongly continuous semigroup of contractions on $\overline{D(F)}$. The convergence here is uniform on bounded subintervals of $[0,\infty)$. If $x \in D(F) \cap R(J_\lambda)$ for some $\lambda > 0$, then for $s, t \geq 0$

$$
d(T_t x, T_s x) \leq 2 l_F(x)|t - s|. \tag{2.15}
$$

Proof: For $x \in D(F) \cap R(J_\lambda)$, we may apply (2.11). We choose $\mu = \frac{t}{n}, \lambda = \frac{t}{m}$, $(n \geq m)$ in (2.11). Then

$$
d(J^n_{\frac{t}{n}} x, J^m_{\frac{t}{m}} x) \leq \frac{2t}{\sqrt{m}} l_F(x) \tag{2.16}
$$

If $x \in \overline{D(F)}$ and $m, n \to \infty$, then $x_n = J_{\frac{t}{n}} x$ constitutes a sequence in $D(F) \cap R(J_{\frac{t}{n}})$ converging to x (see Lemma 2.2). Then

$$
\begin{aligned}
& d(J^n_{\frac{t}{n}} x, J^m_{\frac{t}{m}} x) \\
\leq\ & d(J^n_{\frac{t}{n}} x, J^n_{\frac{t}{n}} x_n) + d(J^m_{\frac{t}{m}} x, J^m_{\frac{t}{m}} x_n) + d(J^m_{\frac{t}{n}} x_n, J^m_{\frac{t}{m}} x_n) \\
\leq\ & 2d(x, x_n) + \frac{2t}{\sqrt{m}} l_F(x_n) \quad \text{by Lemma 2.5, (2.11).}
\end{aligned}
$$

Therefore, for $T > 0$, we obtain

$$\limsup_{m,n\to\infty} \big(\sup_{0\le t\le T} d(J^n_{\frac{t}{n}}x, J^m_{\frac{t}{m}}x) \big) = 0 \tag{2.17}$$

Therefore, as $n \to \infty$, $J^n_{\frac{t}{n}}x$ converges uniformly for $0 \le t \le T$ to some point in X, denoted by $T_t x$.

For $t = 0$, we have

$$T_0 = Id.$$

If we choose $\mu = \frac{s}{n}$, $\lambda = \frac{t}{n}$ $(s \le t)$ and $m = n$ in (2.11), we get

$$d(J^n_{\frac{s}{n}}x, J^n_{\frac{t}{n}}x) \le \left(\sqrt{(s-t)^2 + \frac{t^2}{n}} + \sqrt{(s-t)^2 + \frac{st}{n}} \right) l_F(x),$$

hence for $n \to \infty$

$$d(T_s x, T_t x) \le 2(t-s) l_F(x) \tag{2.18}$$

which implies the continuity of $T_t x$ w.r.t. t. This also holds for $s = 0$; that latter fact alternatively follows from (2.13), i.e.

$$d(J^n_{\frac{t}{n}}x, x) \le t l_F(x)$$

and passing to $n \to \infty$.

Since all J_λ are contractions by Lemma 2.5, so is T_t. In particular, $T_t x$ is Lipschitz continuous in x. Let us now verify the semigroup property. For $k \in \mathbb{N}$, $t > 0$, we have

$$\begin{aligned} T_{kt} &= \lim_{n\to\infty} J^n_{\frac{kt}{n}} = \lim_{m\to\infty} J^{km}_{\frac{kt}{km}} = \lim_{m\to\infty} (J^m_{\frac{t}{m}})^k \\ &= (\lim_{m\to\infty} J^m_{\frac{t}{m}})^k \text{ since the } J^m_{\frac{t}{m}} \text{ are Lipschitz continuous} \\ &\quad \text{with uniform Lipschitz constant 1 (see Lemma 2.5)} \\ &= T^k_t. \end{aligned}$$

For $p, q, r, s \in \mathbb{N}$, one obtains

$$T_{\frac{p}{q}+\frac{r}{s}} = T_{\frac{ps+qr}{qs}} = T^{ps}_{\frac{1}{qs}} \circ T^{rq}_{\frac{1}{qs}} = T_{\frac{p}{q}} \circ T_{\frac{r}{s}}.$$

Consequently, the semigroup property

$$T_{t+s} = T_t \circ T_s$$

holds for rational $t, s \geq 0$, and then also for real $t, s \geq 0$ because of the continuity in t and the Lipschitz continuity in x of $T_t x$. $\square$

Theorem 2.2 is the main result of [Ma]. However, this last result and the proof given here were already known to the present author in September '93. It essentially follows the original proof of the Crandall-Liggett theorem [CL] and also uses the presentation of Miyadera [Mi]. One may ask whether under the assumption of Theorem 2.1, i.e. if we suppose that $T_{t_n} x$ stays bounded for some sequence $t_n \to \infty$, $T_t(x)_{t>0}$ converges to a minimizer of F as $t \to \infty$. This was verified by Mayer [Ma] under various additional assumptions, for example if F is uniformly convex, but the general case is unknown.

Applications of nonlinear resolvents to other geometric variational or evolution problems have been suggested for example by de Giorgi [dG] and Jost [J7].

3 Convergence properties

We consider a family of functionals

$$F_n : Z \to \mathbb{R} \cup \{\pm\infty\}$$

defined on some topological space Z. In view of our subsequent applications, we shall assume

$$Z = L^2(X, \mu; Y, d)$$

for some global NPC space (Y, d), but we invite the reader to search for the most general spaces permitted in our subsequent considerations. We also assume that Z satisfies the first axiom of countability; this is just for simplicity, so that we can use sequences instead of filters. We first recall de Giorgi's notion of Γ-convergence, using [dM] as our basic reference.

Definition 3.1 *$F : Z \to \mathbb{R} \cup \{\pm\infty\}$ is the Γ-limit of the sequence $(F_n)_{n\in\mathbb{N}}$, $F = \Gamma - \lim_{n\to\infty} F_n$ if*

(i) whenever $(x_n)_{n\in\mathbb{N}} \subset Z$ converges to $x \in Z$, then

$$F(x) \leq \liminf_{n\to\infty} F_n(x_n)$$

(ii) for every $x \in Z$, there exists some sequence $(y_n)_{n\in\mathbb{N}} \subset Z$ that converges to x and satisfies

$$F(x) \geq \limsup_{n\to\infty} F_n(y_n)$$

Let us list some properties of Γ-limits:

(1) Γ-limits are lower semicontinuous.

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

(3) If all the F_n satisfy the quadratic contraction property of Def. 2.1, so does Γ-lim F_n.

(4) If x_n is a minimizer for F_n, and if $\lim_{n\to\infty} x_n = x$, then x minimizes F.

(5) The following important compactness result holds in case Z is second countable: Every sequence $(F_n)_{n\in\mathbb{N}}$ contains a Γ-convergent subsequence.

Given a functional F as above, but not assumed to be lower semicontinuous, there exists a greatest lower semicontinuous functional $\underline{F}$ with $\underline{F}(u) \leq F(u)$ for all u. $\underline{F}$ is called the **relaxation** of F. Since $\underline{F}(u)$ is finite whenever $F(u)$ is finite, we have

$$D(F) \subset D(\underline{F}).$$

$\underline{F}$ is given by the formula

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

We recall the Moreau-Yosida approximations of a convex functional F

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

and the "resolvents" $J_\lambda x$ characterized by

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

We now extend Mosco's theory [M] to our situation.

Definition 3.2 *A family of convex functionals $F_n : Z \to \mathbb{R}_+ \cup \{\infty\}$ converges to a functional F in the sense of Mosco 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, if the resolvents $J_{n,\lambda}$ associated with the relaxed forms $\underline{F}_n$ converge to the J_λ.*

Lemma 3.1 *If $(F_n)_{n\in\mathbb{N}}$ converges to F in the sense of Mosco, then (F_n) also Γ-converges to F.*

Proof:

(i) Suppose that $(u_n)_{n\in\mathbb{N}}$ converges to u in Z. Then

$$\begin{aligned} F_n(u_n) \geq \frac{1}{\lambda} F_n^\lambda(u_n) &\geq \frac{1}{\lambda} \underline{F}_n^\lambda(u_n) \\ &\geq \frac{1}{\lambda} \underline{F}_n^\lambda(u) - 2d(u,u_n)d(u,J_\lambda u). \end{aligned}$$

Since $d(u,u_n) \to 0$ for $n \to \infty$, and $\underline{F}_n^\lambda(u)$ converges to $F^\lambda(u)$ by assumption, we obtain

$$\liminf_{n\to\infty} F_n(u_n) \geq \frac{1}{\lambda} F^\lambda(u) \quad \text{for every } \lambda > 0,$$

hence also by letting $\lambda \to 0$

$$\liminf_{n\to\infty} F_n(u_n) \geq F(u).$$

(ii) Let $u \in Z$. We may find a decreasing sequence $\lambda_n \to 0$ for $n \to \infty$ with

$$F(u) \geq \lim_{\lambda\to 0} \lim_{n\to\infty} \frac{1}{\lambda} F_n^\lambda(u) \geq \lim_{n\to\infty} \frac{1}{\lambda_n} F_n^{\lambda_n}(u).$$

We put $u_n := J_{n,\lambda_n} u$, hence

$$\frac{1}{\lambda_n} F_n^{\lambda_n}(u) = F_n(u_n) + \frac{1}{\lambda_n} d^2(u,u_n).$$

Inserting this into the preceding inequality yields

$$F(u) \geq \limsup_{n\to\infty} F_n(u_n).$$

We may assume $u \in D(F)$ in which case $u_n \to u$ in Z by Lemma 2.2.

Thus, we have verified the two properties required for Γ-convergence. □

Definition 3.3 *We say that a functional $F : Z \to \mathbb{R}_+ \cup \{\infty\}$ satisfies the Rellich property if every bounded sequence*

$$(u_n)_{n\in\mathbb{N}} \subset Z$$

with

$$\liminf_{n\to\infty} F(u_n) < \infty$$

contains a subsequence that converges in Z. A sequence $F_n : Z \to \mathbb{R}_+ \cup \{\infty\}$ satisfies an asymptotic Rellich condition if every bounded sequence $(u_n)_{n\in\mathbb{N}} \subset Z$ with

$$\liminf_{n\to\infty} F_n(u_n) < \infty$$

contains a converging subsequence.

Remark: A Rellich condition is typically implied by a Poincaré inequality (as exhibited for example in §4).

Theorem 3.1 *A sequence of convex $F_n : Z \to \mathbb{R}_+ \cup \{\infty\}$ satisfying an asymptotic Rellich property Γ-converges to some F if and only if it converges to F in the sense of Mosco.*

Proof: Mosco convergence always implies Γ-convergence by Lemma 3.1. For the reverse direction, we recall that a Γ-limit is automatically lower semicontinuous, and so the F_n Γ-converge to F if and only if the relaxations $\underline{F}_n$ Γ-converge to F. Thus, we may assume w.l.o.g. that the F_n themselves are lower semicontinuous. Let $y_\lambda = J_\lambda u$, $y_{n,\lambda} = J_{n,\lambda} u$ for some $u \in Z$. Then

$$\begin{aligned} F_n(y_{n,\lambda}) + \frac{1}{\lambda} d^2(y_{n,\lambda}, u) &= \frac{1}{\lambda} F_n^\lambda(u) \\ &\le F_n(u_n) + \frac{1}{\lambda} d^2(u, u_n) \end{aligned}$$

for any u_n. We then choose a sequence $(u_n)_{n\in\mathbb{N}}$ that converges to u and satisfies

$$\limsup_{n\to\infty} F_n(u_n) \le F(u)$$

according to (ii) of Γ-convergence. Thus, $(y_{n,\lambda})_{n\in\mathbb{N}}$ is bounded in Z, and $F_n(y_{n,\lambda})$ is bounded as well. By the asymptotic Rellich property, after selection of a subsequence, $(y_{n,\lambda})$ converges to some z_λ as $n \to \infty$. We want to show that $z_\lambda = y_\lambda$. For any $v \in Z$, by (ii) of Γ-convergence, we find a sequence $(v_n)_{n\in\mathbb{N}}$ converging to v with

$$\limsup_{n\to\infty} F_n(v_n) \le F(v).$$

For every n

$$F_n(y_{n,\lambda}) + \frac{1}{\lambda} d^2(u, y_{n,\lambda}) \le F_n(v_n) + \frac{1}{\lambda} d^2(u, v_n).$$

Using (i) of Γ-convergence for $(y_{n,\lambda})_{n\in\mathbb{N}}$, we then get

$$F(z_\lambda) + \frac{1}{\lambda} d^2(u, z_\lambda) \le F(v) + \frac{1}{\lambda} d^2(u, v).$$

Since this holds for any v, we must have $z_\lambda = J_\lambda u = y_\lambda$. Therefore, $J_{n,\lambda} u = y_{n,\lambda}$ converges to $J_\lambda u$ as $n \to \infty$. We thus have verified Mosco convergence. □

4 Generalized harmonic maps

Although, as mentioned in the introduction, we shall be brief about the geometric aspects in these notes and refer to [J6] for a detailed treatment, it might be useful to introduce the original example that motivated the theory presented here, harmonic maps between compact Riemannian manifolds M and N. For a - sufficiently regular - map $f : M \to N$, we consider the energy functional

$$E(f) = \int_M \|df(x)\|^2 \, d\, vol_m(x) \tag{4.1}$$

where $d\, vol_M$ is the measure given by the Riemannian metric of M and $\|df(x)\|$ is the norm of the differential $df(x)$ considered as a linear map from the tangent space $T_x M$ to $T_{f(x)}N$, where the metrics on these spaces are given by the corresponding Riemannian metrics. If $x^\alpha, \alpha = 1, ..., m$, and $f^i, i = 1, ..., n$ are local coordinates in M and N, resp., and if we denote the metric tensors in these coordinates as $(\gamma_{\alpha,\beta})_{\alpha,\beta=1,...,m}$ and $(g_{ij})_{i,j=1,...,n}$ resp., and if $(\gamma^{\alpha\beta}) := (\gamma_{\alpha\beta})^{-1}$, then in those coordinates

$$\|df(x)\|^2 = \sum_{\substack{\alpha,\ \beta \\ i,\ j}} \gamma^{\alpha\beta}(x) g_{ij}(f(x)) \frac{\partial f^i}{\partial x^\alpha} \frac{\partial f^j}{\partial x^\beta}. \tag{4.2}$$

Thus, E is a quadratic variational integral, but since the coefficients g_{ij} depend on the value of the unknown map f, the corresponding Euler-Lagrange equations are nonlinear; they are of the form

$$\Delta_M f + \Gamma(f)(df, df) = 0 \tag{4.3}$$

where Δ_M is the Laplace-Beltrami operator of M (a linear second order elliptic differential operator, the Riemannian version of the usual Laplace operator), while the nonlinear term which is quadratic in the first derivatives of f stems from the Riemannian geometry of N. The coeffiecients $\Gamma(f)$ are given by the so-called Christoffel symbols of N which in turn are expressions involving first derivatives of the metric of N. The details are not so important here. Solutions of (4.3) are called harmonic maps. (4.3) is a semilinear elliptic system, and in general weak solutions need not be regular, or from a more geometric point of view, solutions need not exist in given homotopy classes. It was however discovered in the work of Al'ber [A1, A2] and Eells-Sampson [ES] that these problems disappear if one assumes that N has nonpositive sectional curvature. That condition resulted in a differential inequality

$$\Delta_M \|df(x)\|^2 \geq -\text{const } \|df(x)\|^2 \tag{4.4}$$

for a solution f that implied estimates on solutions by standard PDE techniques. We shall not enter into those details here, as our approach will be conceptually different. Namely, we shall translate the curvature condition on N into a convexity condition for the functional E that makes the results of §2 applicable. Also, motivated both by geometric applications and by the intrinsic desire to isolate the essential features of the subject, we shall study a more general situation of L^2-maps between a measure space (X, μ) that will take the role of the Riemannian manifold M (or its universal cover) with the Riemannian volume form and a metric space (Y, d) that will generalize the target N (or its universal cover) with the distance function derived from the Riemannian metric.

This will involve a setting where no differentiable structure is assumed, and so the derivatives occuring in the definition E need to be replaced by expressions involving distances only. The guiding idea for those constructions will be that for a smooth map $f : M \to N$ with M and N having Riemannian distances functions d_M, d_N, resp., we have approximations of the norm of df of the type

$$\|df(x)\|^2 = c(\dim M) \lim_{r\to 0} \int_{B(x,r)} \frac{d_N^2(f(x), f(y))}{d_M^2(x,y)}\, \mu(dy) \tag{4.5}$$

with a constant c depending only on the dimension of M, where $B(x, r) := \{y \in M : d_M(x, y) < r\}$, and μ again is the Riemannian volume measure. Alternatively

$$\|df(x)\|^2 = c'(\dim M) \lim_{N\to 0} \frac{\int_{B(x,r)} d_N^2(f(x), f(y))\, \mu(dy)}{\int_{B(x,r)} d_M^2(x,y) \mu(dy)} \tag{4.6}$$

which has the advantage of not involving a singular kernel. Of course, there exist many other possibilities for approximating $\|df(x)\|^2$, for example by using Gaussian integral kernels with variance tending to 0.

The point of such approximations is that for example the right hand sides of (4.5) and (4.6) remain meaningful without assuming differentiable structures and hence can be used to define energy integrals in the more general setting envisioned here.

Let us consider a general situation:
We let $h : X \times X \to \mathbb{R}$ be a nonnegative (but not $\not\equiv 0$) and symmetric (μ-measurable) function, i.e. $h(x, y) \geq 0$ and $h(x, y) = h(y, x)$ for all $x, y \in X$. For geometric applications, we may also wish to impose some invariance condition $h(\gamma x, \gamma y) = h(x, y)$ for all elements γ of some group Γ that operates on X and also leaves the measure μ on X invariant. For a map $f : X \to Y$ into a metric space (Y, d), we then put

$$E_h(f) := \int\int h(x,y) d^2(f(x), f(y))\, \mu(dy)\, \mu(dx).$$

Minimizers of E_h can be characterized by a local mean value property.

Lemma 4.1 *$f : X \to Y$ minimizes E_h if and only if for μ-almost all $x \in X$, $f(x)$ is the $h(x, \cdot)$ weighted mean value of f, i.e. if $f(x)$ minimizes*

$$\Phi(p) := \int h(x, y)\, d^2(p, f(y))\, \mu(dy).$$

For the easy proof, see e.g. [J6]. To put this result into a proper perspective, we observe that for a global NPC space (Y, d), the function Φ has a unique minimizer for every x and h as above, because in that case $d^2(\cdot, q)$ is strictly convex for every q.

In a certain sense, harmonic maps between Riemannian manifolds can be considered as satisfying an infinitesimal mean value property, and in this sense, we are trying to approximate such maps by ones that satisfy local mean value properties. (It should be pointed out, however, that the present mean value property is weaker than the one of harmonic functions on domains in spaces of constant curvature which satisfy a mean value property on every distance ball contained in their domain of definition.) We should also point out that Kendall [K] had arrived at the importance of mean value properties in Riemannian geometry from a stochastic point of view.

Lemma 4.1 suggests the following iterative procedure for constructing minimizers of E_h (see [J2]): We start with $f_0 : X \to Y$ in the desired class and define

$$f : X \times \mathbb{N} \to Y$$

via

$$f(x, 0) := f_0(x)$$

and

$$f(x, n+1) := h(x, \cdot) \text{weighted mean value of} f(\cdot, n) \quad \text{for} x \in X, \sim \in \mathbb{N},$$

in order to get the minimizer as

$$\lim_{n \to \infty} f(x, n).$$

As explained above, one may then consider an appropriate family of kernels h in order to also obtain a minimizer of the original functional E. The reader should compare this with the very interesting stochastic mean value constructions in Kendall's contribution.

Remark: As explained in the geometric digression in §1, in order to make the theory applicable in geometry, one should consider ρ-invariant maps and integrate over some fundamental domain for X in a cover $\tilde{X}$, in order to have a simply connected target space. We shall ignore this point here, but we assure the reader that this can easily be taken into account. See e.g. [J6].

Let us consider some

Examples:

1) X is a discrete space, for example a grid as used in numerical analysis, a discrete approximation of a Riemannian manifold, or a discrete group, equipped with a symmetric neighborhood relation, and put

$$h(x,y) = \begin{cases} 1 & \text{if } x \text{ and } y \text{ are neighbors} \\ 0 & \text{otherwise} \end{cases}$$

2) X carries a metric d_X; we put

$$h_\varepsilon(x,y) = \begin{cases} c(\varepsilon) & \text{if } d_X(x,y) < \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

where $c(\varepsilon) > 0$ is so chosen that for example

$$\lim_{\varepsilon \to 0} E_{h_\varepsilon}$$

is nontrivial (i.e. not identically 0 or ∞) (We shall specify the nature of this limit below.). Thus, $c(\varepsilon)$ is just a suitable normalization constant. Taking this limit of the E_{h_ε} for $\varepsilon \to 0$ leads to the energy functional for maps between Riemannian manifolds. In fact, with the appropriate choice of $c(\varepsilon)$, the functionals E_{h_ε} converge to the energy functional E in a monotically increasing manner up to some error term that is controlled by a lower bound on the Ricci curvature of the domain X, see [J1] or [KS]. For other purpose, it might also be useful and interesting to consider

$$\lim_{R \to \infty} E_{h_R}$$

for example if X is a global NPC space, in order to capture the asymptotic geometry of X.

3) We let $h(x,y)$ be some kind of Gaussian kernel, e.g. of the form

$$\frac{1}{t^{n/2}} \exp\left(-\frac{d_X(x,y)}{t^2}\right)$$

if X is a Riemannian manifold of dimension n with distance function d_X, and consider the limit as t goes to 0 or ∞ of the resulting functionals E_h. More generally, one may consider transition density functions $P_t(x, y)$ for the Markov process associated with some Dirichlet form E. In this context, Sturm [St3, St4] investigated the question raised in [J5] under which conditions the resulting E_h monotonically converge for $t \to 0$. Essentially, his result is that one gets monotone convergence provided the target satisfies a suitable lower curvature bound.

Lemma 4.2 *Suppose Y is a global NPC space. Then the functionals E_h as defined above are convex on $Z = L^2(X, \mu; Y, d)$.*

Proof: This follows from the fact that the squared distance functions on Y are convex. In the notations of 4) of the Appendix, we have

$$\begin{aligned} d^2(\gamma(x,t), \gamma(y,t)) \;\le\; & (1-t)d^2(f(x), f(y)) \\ & + td^2(g(x), g(y)) \end{aligned}$$

for the geodesic $\gamma(\cdot, t)$ from f to g. Integrating this inequality implies the convexity of E_h. □

As already indicated, we wish to consider energy functionals E obtained as limits of suitable E_{h_ε} for $\varepsilon \to 0$ as in example 2). It was the author's idea in [J2] to use Γ-limits in this context. By property (5) of Γ-limits, we may find suitable sequences $(\varepsilon_n)_{n\in\mathbb{N}}$ that converge to 0 for $n \to \infty$ for which

$$E = \Gamma - \lim_{n\to\infty} E_{h_{\varepsilon_n}}$$

exists.

By general properties of Γ-limits as listed in §3, E is lower semicontinuous and satisfies the quadratic contraction property. Furthermore, if Y is a global NPC space, the convexity of the E_h as verified in Lemma 4.2 carries over to the Γ-limit E.

There is still some arbitrariness involved in the choice of the h_ε, and one may wish to exploit this flexibilty to get some additional control over E. For example, it will be convenient to have a representation of E by so-called energy measures $\eta(u, u)$, i.e. (nonnegative) Radon measures satisfying

$$E(u) = \int_X \eta(u, u)\, dx.$$

This has been achieved by Korevaar-Schoen [KS]. In particular, this allows to restrict E to μ-measurable subsets Ω of X by using the characteristic function χ_Ω of Ω and considering

$$E(u;\Omega) := \int_X \chi_\Omega(x)\eta(u,u)\,dx.$$

When we restrict E to scalar valued functions, the existence of the energy measure $\eta(u,u)$ is quite simple, see e.g. [FOT, p.110 f, p. 159 f]. In that case, we have in terms of the polarized form $E(\cdot,\cdot)$ and for $u \in D(E) \cap L^\infty(X;\mathbb{R}); \varphi \in D(E) \cap C_0^0(X,\mathbb{R})$

$$\begin{aligned} 2E(\varphi u, u) - E(u^2, \varphi) &= \int_X \varphi(x)\eta(u,u)\,dx \\ &= \lim_{t \searrow 0} \frac{1}{2t} \int_{X\times X\setminus diagonal} \varphi(x)(u(x)-u(y))^2\, p_t(x,dy)\,\mu(dx), \end{aligned} \tag{4.7}$$

where $(p_t)_{t\geq 0}$ is the family of transition functions for the Markov process associated with E, or equivalently for $u, v \in D(E)$

$$\begin{aligned} E(u,v) &= \int_X \eta(u,v)\,dx \\ &= \lim_{t \searrow 0} \frac{1}{2t} \int_{X\times X\setminus diagonal} (u(x)-u(y))(v(x)-v(y))p_t(x,dy)\mu(dx). \end{aligned} \tag{4.8}$$

(However, as found by Sturm [St3], the corresponding expressions in the general case, namely

$$\frac{1}{2t} \int_{X\times X\setminus diagonal} d^2(u(x),u(y))p_t(x,dy)\mu(dx)$$

need not always converge in the same monotonic manner for $t \searrow 0$ as they do in the scalar case.)

One may also wish that E be local in the sense that

$$E(u;\Omega) = 0$$

whenever u is constant on the open set Ω. Sturm [St2] has an argument to show the strong locality of ordinary Dirichlet forms (i.e. restrictions to $L^2(X,\mu;\mathbb{R})$ of generalized Dirichlet forms) obtained as Γ-limits of functionals as above for suitable families (h_ε) similar as the ones considered above, and using the quadratic contraction property, one may then deduce locality in our sense. In any case, since by Lemma 4.2 the E_h and hence also the Γ-limit E are convex, we get

Theorem 4.1 *The existence result for minimizers of Theorem 2.2 applies to the functionals E_h and E.*

Here we do not want to present the precise existence results because in order to ensure the boundedness of the resolvents J_λ in Theorem 2.1, one needs an additional necessary assumption that can be phrased as a so-called reductivity condition. We refer to [J6] for details. Theorem 4.1, as presented in detail in that reference, includes the previous existence results of Gromov-Schoen [GS], Jost [J2], and Korevaar-Schoen [KS]. As a Corollary, we just quote the original existence result of Al'ber [A2] and Eells-Sampson [ES].

Corollary 4.1 *Let M and N be compact Riemannian manifolds, with N of nonpositive sectional curvature. Then every continuous $g : M \to N$ is homotopic to a harmonic map.*

To prove this result, one looks at the universal covers X of M and Y of N. Y then is a global NPC space. The group of covering translations Γ for M and the homotopy class of g determine a homomorphism

$$\rho : \Gamma \to I(Y),$$

and a map $\tilde{f} : X \to Y$ passes to a map $f : M \to N$ that is homotopic to ρ if and only if for all $x \in X$, $\gamma \in \Gamma$

$$\tilde{f}(\gamma x) = \rho(\gamma)\tilde{f}(x).$$

This is the equivariance condition on $\tilde{f}$. Theorem 4.1 then implies that one may minimize E or E_h among all such ρ-equivariant maps in order to produce the desired harmonic map. □

Of course, so far the map has been produced only in the space $L^2(X, Y) \cap D(E)$ and so we need to discuss regularity results for minimizers of energy functionals. If domain and target both are Riemannian manifolds and the latter has nonpositive sectional curvature, regularity results were already known to Al'ber [A1] and Eells-Sampson [ES]. More precise results were obtained by Hamilton [Ha] and Hildebrandt-Kaul-Widman [HKW]. Recently also a very interesting stochastic approach to regularity was put forward e.g. in [K].

If X is a Riemannian manifold and (Y, d) is a locally finite Euclidean Bruhat - Tits building (and thus satisfies an NPC condition), very precise regularity results were obtained in Gromov-Schoen [GS]. If X is still a Riemannian manifold and (Y, d) is an NPC space, Korevaar - Schoen [KS] showed Lipschitz continuity of minimizers. We shall now describe a Hölder continuity result for more general domains.

We need some additional assumptions, in order to put the Harnack inequality of Biroli - Mosco [BM] at our disposal. Let X be a locally compact, separable topological space with a positive Radon measure μ with $\operatorname{supp} \mu = X$. Let E be a generalized Dirichlet form as above with energy measure η and let $E(\cdot, \cdot)$ be the Dirichlet form on $L^2(X, \mu; \mathbb{R})$ obtained by polarization. We assume that $E(\cdot, \cdot)$ is strongly local, i.e. $E(u, v) = 0$ whenever u, v have compact support and v is constant on a neighbarhood of $\operatorname{supp} u$. We assume that $E(\cdot, \cdot)$ admits a μ-separating core C. This is a subspace C of $D(E) \cap C_0^0(X)$ (where $D(E)$ now means the domain of definition of the Dirichlet form on $L^2(X, \mu; \mathbb{R})$ that is dense in $C_0^0(X)$ w.r.t the norm $\| \cdot \|_{L^2}^2 + E(\cdot)$ and which satisfies the condition that for any $x \neq y \in X$, there exists $\psi \in C$ with $\psi(x) \neq \psi(y)$ and

$$\eta(\psi, \psi) \leq \mu$$

on X.

In that case,

$$d_E(x, y) := \sup\{\psi(x) - \psi(y) : \psi \in C, \eta(\psi, \psi) \leq \mu\}$$

yields a metric on X. (If X is a Riemannian manifold and E is the classical Dirichlet integral, then d_E is the original Riemannian metric.) We assume that the topology induced by E is the original topology of X. We also assume the so-called **ball doubling property**.

(A) For $x \in X, r > 0$, we put $B(x, r) := \{y \in X : d_E(x, y) < r\}$, and require that there exist constants $c_1 < \infty, R > 0$ such that

$$\mu(B(x, 2r)) \leq c_1 \mu(B(x, r))$$

whenever $x \in X, 0 < r \leq R$.

This ball doubling property should be considered as some kind of generalization of a lower bound for the Ricci curvature in the case of a Riemannian manifold, because such a bound implies a control on the growth of the volume of balls by the Bishop-Gromov volume comparison theorem. It also seems that such a ball doubling property automatically implies that X is locally compact.

Let now a metric space (Y, d) be given. Secondly, we need a **Poincaré inequality**.

(B) For any $X_0 \subset\subset X$, there exist constants $0 < \alpha \leq 1$, $c_2 < \infty$ s.t. for any ball $B(x_0, r) \subset X_0$ and every $f : X \to Y, f \in D(E)$

$$\int_{B(x_0, \alpha r)} d^2(f(x), \bar{f}_{B(x_0, \alpha r)}) \mu(dx) \leq c_2 r^2 \int_{B(x_0, r)} \eta(f, f)\, dx.$$

Here $\bar{f}_A$ is the mean value of f on a set A, i.e. the minimum of the function

$$\Phi(q) := \int_A d^2(f(x), q)\, \mu(dx).$$

Another version of the Poincaré inequality that implies the preceding one is

$$\int_{B(x_0,\alpha r)} \int_{B(x_0,\alpha r)} d^2(f(x), f(y))\, \mu(dx)\, \mu(dy) \leq c_3 r^2 \int_{B(x_0,r)} \eta(f,f)\, dx$$

for some constant c_3.

These Poincaré inequalities are generalizations of the corresponding ones for scalar valued functions, and in some (and perhaps all) cases they may be deduced from the latter:

1) If X is a domain in some $\mathbb{R}^d$, Poincaré inequalities for maps $f : X \to Y$ can be derived in the same way as the ones for functions.

2) If (Y, d) is a compact Riemannian manifold, Nash's theorem yields an isometric embedding i of (Y, d) into some Euclidean space, and one applies the Poincaré inequality for $i \circ f$ to get one for f.

Biroli - Mosco [BM] showed that the preceding assumptions, i.e. the ball doubling property and the Poincaré inequality imply a Sobolev inequality which in term implies a Harnack inequality for supersolutions of the nonpositive self-adjoint operator A corresponding to the Dirichlet form $E(\cdot, \cdot)$. This work was extended in [St1]. (Note however that the reverse implications do not always hold, i.e. in the presence of a ball doubling property, a Harnack inequality need not imply a Sobolev inequality, and the latter need not imply a Poincaré inequality.) The Harnack inequality is the following

Lemma 4.3 *Let $v \in D(E)$ be a nonnegative, bounded supersolution*

$$Av \leq 0 \text{ in } B(x_0, 4R) \subset\subset X$$

in the weak sense, i.e.

$$\int_{X \cap B(x_0,4R)} \eta(v, \varphi)\, dx \leq 0 \qquad \text{for all } \varphi \geq 0, \varphi \in D(E),$$
$$\text{with supp } \varphi \subset\subset X \cap B(x_0, 4R).$$

Then for some $p \geq 1$ and some constant c_4

$$\left(\int_{B(x_0,2R)} |v|^p \mu(dx) \right)^{1/p} \leq c_4 \inf_{B(x_0,R)} v.$$

A consequence of the Harnack inequality is the following Hölder continuity result for minimizers of E with values in a NPC space.

Theorem 4.2 *Let the energy form E satisfy the preceding assumptions, let (Y, d) be a complete NPC space, and let $f : X \to Y$ be a local minimizer for E in the sense that*

$$E(f_{|B(x_0,R))} \leq E(g_{|B(x_0,R)})$$

for all $x_0 \in X, 0 < R < R_0$ for some fixed R_0, and all

$$g \in L^2(B(x_0, R_0), \mu; Y, d) \text{ with } g = f \text{ on } B(x_0, R_0) \setminus B(x_0, R).$$

Then f is locally Hölder continuous on X.

This result was obtained in [J5]. Another proof of Hölder continuity based on a Harnack inequality was given by F. H. Lin [L] (according to the introduction of [L] this was known to him considerably earlier than [L] appeared). The argument of [L] is shorter than the one of [J5], but it assumes the local compactness of the target Y which [J5] does not need.

An interesting question is to find out the appropriate assumptions on X that are needed to improve Hölder continuity to Lipschitz continuity (cf. the result of Korevaar - Schoen [KS] that works if X is a Riemannian manifold).

In order to employ the Harnack inequality for the proof of Theorem 4.2, of course one has to produce supersolutions of A. In fact, by changing signs it is as good to construct subsolutions. This is achieved in

Lemma 4.4 *Let $f : X \to Y$ be a local minimizer for E where Y is a global NPC space. Then for any $p \in Y$*

$$Ad^2(f(\cdot), p) \geq 2\eta(f, f) \text{ weakly.}$$

In particular $d^2(f(x), p)$ is a subsolution of A in the weak sense.

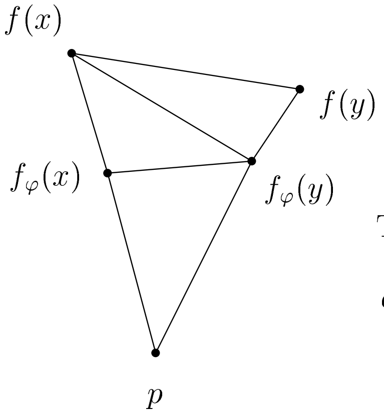
Proof: We consider a function $\varphi \in D(E)$ with compact support and $0 \leq \varphi \leq 1$. For $x \in X$, we consider the geodesic

$$\gamma_x : [0, 1] \to Y$$

with $\gamma_x(0) = f(x), \gamma_x(1) = p$. We put

$$f_\varphi(x) := \gamma_x(\varphi(x)).$$

The triangle comparison theorem yields



$$\begin{aligned} d^2(f(x), f_\varphi(y)) &\leq (1-\varphi(y))d^2(f(x), f(y)) \\ &\quad +\varphi(y)d^2(f(x), p) \\ &\quad -\varphi(y)(1-\varphi(y))d^2(f(y), p). \end{aligned}$$

Triangle comparison also yields

$$\begin{aligned} d^2(f_\varphi(x), f_\varphi(y)) &\leq (1-\varphi(x))d^2(f(x), f_\varphi(y)) \\ &\quad +\varphi(x)d^2(f_\varphi(y), p) \\ &\quad -\varphi(x)(1-\varphi(x))d^2(f(x), p). \end{aligned}$$

Inserting the first inequality into the second one and using

$$d(f_\varphi(y), p) = (1-\varphi(y))d(f(y), p)$$

yields

$$\begin{aligned} d^2(f_\varphi(x), f_\varphi(y)) &\leq (1-\varphi(x))(1-\varphi(y))d^2(f(x), f(y)) \\ &\quad -(\,\varphi(x)-\varphi(y)\,)(\,1-\varphi(x)\,)\,(\,d^2(f(x),p)-(\,1-\varphi(y)\,)d^2(f(y),p)\,). \end{aligned}$$

Since f is minimizing, we have

$$\int \eta(f, f)(dx) \leq \int \eta(f_\varphi, f_\varphi)(dx).$$

Combining this inequality with the previous one expressed in terms of $\eta(\cdot,\cdot)$, we get

$$\begin{aligned} 0 &\leq -\int 2\varphi(y)\eta(f, f)(dx) \\ &\quad -\int \eta(\varphi(x), d^2(f(x), p))(dx) \\ &\quad + \text{terms that are at least quadratic in } \varphi. \end{aligned}$$

Replacing φ by $t\varphi$ with $t > 0$, dividing the resulting inequality by t and letting t tend to zero eliminates the quadratic and higher terms in φ, and we obtain

$$-\int \eta(\varphi(x), d^2(f(x), p))(dx) \geq 2\int \varphi(x)\eta(f, f)(dx)$$

which is equivalent to the claim. □

Appendix: Geometric notions

1) Let (Y, d) be a metric space.
A curve in Y is a continuous map

$$\gamma : [0,1] \to Y.$$

The length of a curve γ is defined as

$$L(\gamma) := \sup\{\sum_{i=0}^{n} d(\gamma(t_{i-1}), \gamma(t_i)) \quad 0 = t_0 < t_1 < \ldots < t_n = 1, n \in \mathbb{N}\},$$

γ is called rectifiable if $L(\gamma) < \infty$. (Y, d) is called a length space if for any $x, y \in Y$

$$d(x, y) = \inf\{L(\gamma) : \gamma : [0,1] \to Y \text{ curve with } \gamma(0) = x, \gamma(1) = y\},$$

i.e. if the distance between any two points equals the infimum of the lengths of curves connecting them.
(Y, d) is called a geodesic length space if this infimum is achieved, i.e. if for any $x, y \in Y$, we may find a curve $\gamma : [0,1] \to Y$ with $\gamma(0) = x, \gamma(1) = y$, and

$$L(\gamma) = d(x, y).$$

Any such minimizing curve γ is geodesic. A shortest geodesic is characterized by the property that $d(\gamma(t_1), \gamma(t_2)) = |t_2 - t_1| \ d(\gamma(0), \gamma(1))$ $(= |t_2 - t_1| L(\gamma))$ whenever $0 \leq t_1, t_2 \leq 1$. More generally, a curve $\gamma : [0,1] \to N$ is called geodesic if there exists $\varepsilon > 0$ with

$$L(\gamma_{|[t_1,t_2]}) = d(\gamma(t_1), \gamma(t_2)) \text{ whenever } 0 \leq t_2 - t_1 < \varepsilon (0 \leq t_1, t_2 \leq 1).$$

Obviously, the geodesic property can also be defined for curves defined on an interval different from $[0,1]$, and if a curve is geodesic, its restriction to any subinterval of its domain of definition remains geodesic.
Two curves $\gamma_0, \gamma_1 : [0,1] \to Y$ with the same end points, i.e $\gamma_0(0) = \gamma_1(0) =: x_0$, $\gamma_0(1) = \gamma_1(1) =: x_1$, are called homotopic if there exists a continuous

$$\Gamma : [0,1] \times [0,1] \to Y$$

with

$$\Gamma(t, 0) = \gamma_0(t), \Gamma(t, 1) = \gamma_1(t) \text{ for all } t \in [0,1]$$
$$\Gamma(0, s) = x_0, \Gamma(1, s) = x_1 \text{ for all } s \in [0,1].$$

We point out that our definition of homotopy between curves with the same end points requires that these end points remain fixed during the homotopy.

We say that (Y, d) is a strong geodesic length space if the infimum of the length of curves in a given homotopy class is always achieved. This means that given

$$\gamma_0 : [0,1] \to Y$$

with $\gamma_0(0) = x, \gamma_0(1) = y$, there exists a curve $\gamma : [0,1] \to Y$ with $\gamma(0) = x, \gamma(1) = y$, homotopic to γ_0, whose length equals the infimum of the length of all such curves homotopic to γ_0.

We point out, however, that we do not require that geodesics are unique in a given homotopy class. However, such uniqueness holds if one makes a suitable assumption about nonpositive curvature as will be done below.

Any rectifiable curve, in particular any geodesic in Y may be parametrized proportionally to arclength. This means that after composition with a suitable homomorphism $\sigma : [0,1] \to [0,1]$ the resulting curve $\gamma : [0,1] \to Y$ satisfies

$$L(\gamma_{|[0,t]}) = tL(\gamma) \text{ for all } 0 \le t \le 1.$$

In these notes, we shall automatically assume that all geodesics are parametrized proportionally to arclength.

A function $F : Y \to \mathbb{R}$ is called convex if its restriction to any geodesic $\gamma : [0,1] \to Y$ is a convex function of the arclength parameter t, i.e. $F(\gamma(t)) : [0,1] \to \mathbb{R}$ is a convex function in the usual sense.

2) If (Y, d) is a metric space, we denote by

$$I(Y)$$

the group of isometries of Y, i.e. the group of all homeomorphisms

$$h : Y \to Y$$

with

$$d(h(x), h(y)) = d(x, y) \text{ for any } x, y \in Y.$$

3) A geodesic length space(Y,d) is said to have nonpositive curvature in the sense of Alexandrov or as we shall say for abbreviation, to be an NPC space if for every $p \in Y$, there exists $\rho_p > 0$ such that for any $x, y, z \in B(p, \rho_p)$ and any shortest geodesic $\gamma : [0,1] \to Y$ with $\gamma(0) = x, \gamma(1) = z$ and all $0 \le t \le 1$

(G.1) $d^2(\gamma(t), y) \le (1-t)d^2(\gamma(0), y) + td^2(\gamma(1), y) - t(1-t)L(\gamma)^2.$

(G.1) becomes an equality for certain triangles in $\mathbb{R}^2$ with the Euclidean metric. The geometric content of (G.1) is that the distance function in (Y,d) is at least as convex as the Euclidean distance function. Thus (G.1) is a quantitative convexity statement. In an NPC space, we have the following useful quadrilateral comparison theorem of Reshetnyak [Re]. (Proofs can also be found in [KS] and [J6].)

The statement is that for any geodesics $\gamma_1, \gamma_2 : [0,1] \to Y$ with images continued in some ball $B(p, \rho_p)$ as above and $(0 \le s, t \le 1)$

$$\text{(G.2)} \quad \begin{aligned} & d^2(\,\gamma_1(0), \gamma_2(t)\,) + d^2(\,\gamma_1(1), \gamma_2(1-t)\,) \\ \le\ & d^2(\,\gamma_1(0), \gamma_2(0)\,) + d^2(\,\gamma_1(1), \gamma_2(1)\,) + 2\,t^2\, d^2(\,\gamma_2(0), \gamma_2(1)\,) \\ & + t\ (\,d^2(\,\gamma_1(0), \gamma_1(1)\,) - d^2(\,\gamma_2(0), \gamma_2(1)\,)\,) \\ & - s\,t\ (\,d(\,\gamma_1(0), \gamma_2(0)\,) - d(\,\gamma_1(1), \gamma_2(1)\,))^2 \\ & -(1-s)\,t\ (\,d(\,\gamma_1(0), \gamma_1(1)\,) - d(\,\gamma_2(0), \gamma_2(1)\,)\,)^2 . \end{aligned}$$

Again, equality in (G.2) holds for certain quadrilaterals in the Euclidean plane.

(Y,d) is called a global NPC space if we may put $\rho_p = \infty$ for some (and hence for all) p in the above condition.

The global NPC condition implies in particular that the geodesic curve between any two points $x, y \in Y$ is unique. If $\gamma_{x,y} : [0,1] \to Y$ is this geodesic, with $\gamma_{x,y}(0) = x$, $\gamma_{x,y}(1) = y$, we put for $t \in [0,1]$

$$m_t(x,y) = \gamma_{x,y}(t).$$

In particular, we shall call

$$m_{\frac{1}{2}}(x,y)$$

the midpoint of x and y. $m_{\frac{1}{2}}(x,y)$ is uniquely determined by the relations

$$\text{(G.3)} \qquad d(m_{\frac{1}{2}}(x,y), x) = d(m_{\frac{1}{2}}(x,y), y) = \frac{1}{2} d(x,y).$$

The uniqeness of geodesics also implies that a global NPC space is simply connected. It is shown in [AB] that the universal cover of an NPC space is a global NPC space. Since in the context of the present notes, one may always lift to universal covers, we shall usually assume that our spaces are global NPC spaces.

4) We now observe that geometric properties of (Y, d) carry over to $Z = L^2(X, \mu; Y, d)$. If (Y, d) is complete, so is Z by definition of L^2-spaces. If (Y, d) is a geodesic length space, so is Z. Namely given $f, g \in Z$, for each $x \in X$, we let $\gamma_x : [0,1] \to Y$ be a shortest geodesic with $\gamma(0) = f(x), \gamma(1) = g(x)$. We put $\gamma(x, y) := \gamma_x(t)$.

$$\Gamma : [0,1] \to Z, \quad \Gamma(t) = \gamma(\cdot, t)$$

is a shortest geodesic in Z between f and g. In fact,

$$\begin{aligned} d(\Gamma(t_1), \Gamma(t_2)) &= \left(\int d^2(\gamma(x, t_1), \gamma(x, t_2))\, \mu(dx) \right)^{1/2} \\ &= |t_1 - t_2| \left(\int d^2(\, f(x), g(x)\,)\, \mu(dx) \right)^{1/2} \end{aligned}$$

by our above characterization, since $\gamma(x, \cdot)$ is a shortest geodesic for every x, and that characterization then conversely shows that $\Gamma(\cdot)$ is geodesic. Thus, geodesics in $Z = L^2(X, \mu; Y, d)$ are just obtained as families of geodesics in Y. One may also perform this construction in fixed homotopy classes. This means that if $F : X \times [0,1] \to Y$ is a homotopy between f and g, one can choose the geodesics from $f(x)$ to $g(x)$ homotopic to $F(x, \cdot)$ to obtain the geodesic family $\Gamma(t) = \gamma(x, t)$ homotopic to $F(x, t)$. Also, curvature properties of Y extend to Z. Namely, if Y is e.g. a global NPC space, the NPC inequality continuous to hold for Z, simply by integrating it over the relevant families of geodesics that constitute the geodesics in Z.

References

[A1] S. I. Al'ber, On n-dimensional problems in the calculus of variations in the large, Sov. Math. Dokl. **5** (1964), 700-704.

[A2] S. I. Al'ber, Spaces of mappings into a manifold of negative curvature, Soc. Math. Dokl. **9** (1967), 6-9.

[AB] S. Alexander, R. Bishop, The Hadamard-Cartan theorem in locally convex metric spaces, L'Ens. Math. **36** (1990), 309-320.

[BD2] A. Beurling, J. Deny, Dirichlet spaces, Proc. Nat. Acad. Sci. USA **45** (1959), 208-215.

[BD1] A. Beurling, J. Deny, Espaces de Dirichlet I, le cas élémentaire, Acta Math. **99** (1958), 203-224.

[BH] N. Bouleau, F. Hirsch, Dirichlet forms and analysis on Wiener space, de Gruyter, 1991.

[BM] M. Biroli, U. Mosco, A Saint-Venant type principle for Dirichlet forms on discontinuous media, Annal. Mat. Pure Appl. **169** (1995), 125-181.

[CFL] R. Courant, K. O. Friedrichs, H. Lewy, Über die partiellen Differentialgleichungen der mathematischen Physik, Math. Ann. **100** (1928).

[CL] M. Crandall, T. Liggett, Generation of semigroups of nonlinear transformations on general Banach spaces, Amer. J. Math. **93** (1971), 265-298.

[dG] E. de Giorgi, New problems on minimizing movements, in: Boundary value problems for partial differential equations and applications (J. L. Lions, C. Baiochi, eds.), RMA **29**, Masson, Paris, 1993.

[dM] G. dal Maso, An introduction to Γ-convergence, Birkhäuser, 1993.

[ES] J. Eells, J. Sampson, Harmonic mappings of Riemannian manifolds, Amer. J. Math. **85** (1964), 109-160.

[FOT] M. Fukushima, Y. Oshima, M. Takeda, Dirichlet forms and symmetric Markov processes, de Gruyter (1994).

[GS] M. Gromov, R. Schoen, Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one, Publ. Math. IHES **76** (1992), 165-246.

[Ha] R. Hamilton, Harmonic maps of Riemannian manifolds with boundary, Springer LNM **471**, 1975.

[HKW] S. Hildebrandt, H. Kaul, K. O. Widman, An existence theorem for harmonic mappings of Riemannian manifolds, Acta Math. **138** (1997), 1-16.

[J1] J. Jost, Riemannian geometry and geometric analysis, Springer, 1995.

[J2] J. Jost, Equilibrium maps between metric spaces, Calc. Var. **2** (1994), 173-204.

[J3] J. Jost, Convex functionals and generalized harmonic maps into spaces of nonpositive curvature, Comment. Math. Helv. **70** (1995), 659-673.

[J4] J. Jost, Generalized harmonic maps between metric spaces. in: Geometric analysis and the calculus of variations for Stefan Hildebrandt (J. Jost, ed.), International Press, 1996, pp. 143-174.

[J5] J. Jost, Generalized Dirichlet forms and harmonic maps, Calc. Var. **5** (1997), 1-19.

[J6] J. Jost, Nonpositive curvature: Geometric and analytic aspects, Birkhäuser, 1997.

[J7] J. Jost, A. weak notion of mean curvature and a generalized mean curvature flow for singular sets, in: Advances in Geometric Analysis and Continuum Mechanics (P. Concus, K. Lancester, eds.), International Press, pp. 138-144.

[K] W. Kendall, Brownian motion and partial differential equations. From the heat equation to harmonic maps, Proc. 151, 49^{th} session, Firenze, 1993, pp. 85-101.

[KS] N. Korevaar, R. Schoen, Sobolev spaces and harmonic maps for metric space targets, Comm. Anal. Geom. **1** (1993), 561-659.

[L] F. H. Lin, Analysis on singular spaces, Preprint, 1997.

[M] U. Mosco, Composite media and asymptotic Dirichlet forms, J. Funct. Anal. **123** (1994), 368-421.

[Ma] U. Mayer, Gradient flow on nonpositively curved metric spaces, Thesis, Univ. Utah, Salt Lake City, 1994.

[Mi] Miyadera, Nonlinear Semigroups, Transl. Math. Monographs **109**, AMS, 1991.

[MR] Z. M. Ma, M. Roeckner, An introduction to the theory of (non-symmetric) Dirichlet forms, Springer, 1992.

[Re] Y. G. Reshetnyak, Inextensible mappings in a space of curvature no greater than K, Sib. Math. J. **9** (1968), 683-689.

[St1] K. Sturm, Analysis on local Dirichlet spaces II: Upper Gaussian estimates for the fundamental solutions of parabolic equations, Osaka J. Math. **32** (1995), 275-312.

[St2] K. Sturm, Diffusion processes and heat kernels on length spaces.

[St3] K. Sturm, Monotone approximation of energy functionals for mappings into metric spaces, I, J. reine angew. Math. **486** (1997), 129-151.

[St4] K. Sturm, Monotone approximation of energy functionals for mappings into metric spaces, II, Preprint, 1997.