A NONCOMMUTATIVE TRANSPORT METRIC AND SYMMETRIC QUANTUM MARKOV SEMIGROUPS AS GRADIENT FLOWS OF THE ENTROPY

MELCHIOR WIRTH

Abstract. We study quantum Dirichlet forms and the associated symmetric quantum Markov semigroups on noncommutative $L^2$ spaces. It is known from the work of Cipriani and Sauvageot that these semigroups induce a first order differential calculus, and we use this differential calculus to define a noncommutative transport metric on the set of density matrices. This construction generalizes both the $L^2$-Wasserstein distance on a large class of metric spaces as well as the discrete transport distance introduced by Maas, Mielke, and Chow–Huang–Li–Zhou. Assuming a Bakry–Émery-type gradient estimate, we show that the quantum Markov semigroup can be viewed as a metric gradient flow of the entropy with respect to this transport metric. Under the same assumption we also establish that the set of density matrices with finite entropy endowed with the noncommutative transport metric is a geodesic space and that the entropy is semi-convex along these geodesics.

Contents

Introduction 1
1. Quantum Dirichlet Forms 5
2. Carré du champ 11
3. Operator means and the algebra $A_\theta$ 14
4. The noncommutative transport metric $W$ 22
5. Von Neumann entropy and Fisher information 32
6. Gradient flow of the entropy 42
7. Geodesic convexity 56
References 63

Introduction

Since the seminal work of Jordan–Kinderlehrer–Otto [JKO98] and Otto [Ott01] it is known that the space of probability measures on $\mathbb{R}^d$ endowed with the $L^2$-Wasserstein metric can, at least formally, be viewed as an infinite-dimensional Riemannian manifold and the heat flow as gradient flow of the Boltzmann entropy.

This insight spawned a lot of subsequent activity, extending the gradient flow characterization to various other geometric settings (see [AGS14a, AS18, Erb10, GKO13, Jui14, OS09]) as well as to other evolution equations (see [Erb16, Ott01]).

Not the least, the characterization of the heat flow as gradient flow of the entropy played a crucial role in the work of Ambrosio, Gigli and Savaré [AGS14a, AGS14b, AGS15] that provided an understanding of the connection between synthetic lower 2020 Mathematics Subject Classification. 81S22 (primary), 35K20, 35R20, 46L57, 47D06 (secondary).

The author was financially supported by the German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes) and the German Research Foundation (DFG) via RTG 1523.
bounded Ricci curvature bounds in the sense of Lott–Sturm–Villani [LV09, Stu06a, Stu06b] and Bakry–Émery [BE85].

In contrast, for discrete spaces respectively evolution equations with nonlocal generators, the Monge–Kantorovich formulation of transport distances has turned out not to be useful in this direction: While one can define the Wasserstein distances $W_p$ for an arbitrary metric $d$, absolutely continuous curves in the Wasserstein space are constant when $d$ is discrete and thus there are no non-trivial gradient flows.

However, Maas [Maa11], Mielke [Mie11], and Chow–Huang–Li–Zhou [CHLZ12] independently defined a discrete transport metric $W$ on the set of probability densities over a finite graph such that the heat flow for the graph Laplacian coincides with the gradient flow of the entropy with respect to $W$. Instead of the Monge–Kantorovich optimal transport problem, their approach is based on a discrete version of the Benamou–Brenier formula [BB00], which gives an equivalent description of the $L^2$-Wasserstein metric on probability measures on Euclidean space.

This new metric has already proven to be very fertile. On the one hand, the gradient flow characterization has been generalized to the heat equation for generators of jump processes [Erb14] as well as a variety of other evolution equations on graphs [CLZ18, EM14, EFLS16, LM13]. On the other, (variants of) the metric $W$ has been used (among other things) to define lower Ricci curvature bounds for graphs [EM12] and to study a new discrete version of the nonlinear Schrödinger equation [CLZ19].

Moreover, in recent years have seen new activity in the study of matrix-valued optimal transport with several groups studying a version of the metric $W$ for matrix algebras (see [CM14, CM17, CGGT17, CGT18, MM17]); and, independently, Brenier [Bre17, Bre18] discovered a surprising connection between matrix-valued optimal transport and fluid dynamics. Notably, Carlen and Maas [CM14, CM17] showed that the metric $W$ allows to view the flow determined by a finite-dimensional Markovian quantum master equation as gradient flow of the von Neumann entropy.

Both in the case of graphs and matrix algebras, all work so far has been limited to a finite-dimensional setting and the problem of extending it to the infinite-dimensional case has been raised in several of the aforementioned articles. This problem is solved in the present article. More precisely, we give a definition of $W$ and a characterization of the flow defined by a Markovian quantum master equation as gradient flow of the entropy for quantum Dirichlet forms – a setting that generalizes many of the ones above – based on the first order differential calculus developed by Cipriani and Sauvageot [CS03].

In particular, this article gives the first unified approach to the results in the local case (for example the heat equation on Euclidean space, manifolds, infinitesimally Riemannian metric measure spaces) on the one hand and non-local case (e.g. heat equation on graphs, for fractional powers of the Laplacian) on the other hand, which could only be treated by analogy until now.

On the noncommutative side, this setting does not only treat infinite-dimensional quantum systems, but also some classical examples of noncommutative geometry such as the noncommutative heat semigroup on the noncommutative torus. This could open the door to a theory of Ricci curvature for noncommutative spaces, a concept that has been notoriously elusive in noncommutative geometry until now.

Let us shortly comment on the differences to prior work. In contrast to the case of metric measure spaces, many powerful tools of optimal transport are not available here, and in the Benamou–Brenier formulation, the continuity equation depends linearly on the measure density in the local case, while in our setting, it is in general a nonlinear equation in the density.

These problems have already been tackled successfully in the non-local case of graphs and jump processes, however, the necessary analysis of monotonicity and
convexity properties turns out more difficult in the noncommutative setting as operator monotonicity and operator convexity are decidedly more rigid notions than their commutative counterparts.

Compared to previous work on matrix-valued optimal transport, we deal not only with quantum Markov semigroups on an infinite-dimensional space, but in general with semigroups that are not uniformly bounded. This means that the Lindblad characterization of generators of quantum Markov semigroups, which is a central tool for the construction of \( \mathcal{W} \) for matrix algebras, is no longer applicable. Moreover, the density matrices in our setting are not only operators on an infinite-dimensional space (as opposed to matrices), but in general unbounded ones. For that reason our analysis requires a careful adaptation of classical tools for operator monotonicity and convexity, which are usually only developed for bounded operators. Furthermore, it is only in the infinite-dimensional case that the full power of the theory of gradient flows in metric spaces is needed, whereas in finite dimensions the elementary theory of gradient flows on Riemannian manifolds suffices.

Let us therefore stress that the infinite-dimensional setting does not only include new classes of examples, but that it is also necessary for unified treatment of the local and non-local case is possible since locality is a purely infinite-dimensional phenomenon (incidentally, it did not appear in the seminal work of Beurling–Deny [BD58] on Dirichlet forms, as they only treated the finite-dimensional case).

Among other possible applications, we hope to lay the ground for a systematic study of displacement convexity of the entropy for infinite-dimensional quantum systems, a topic which has already proven useful for convergence results in the finite-dimensional case [CM17].

Moreover, the theory developed here could provide a framework for approximation results of smooth spaces or infinite-dimensional systems by discrete spaces or finite-dimensional systems, which so far have only been treated in some particular cases [GM13, Gar17].

Let us summarize the content of this article in some more detail. In Section 1 we recall some basic facts about noncommutative integration and quantum Dirichlet forms, including the first order differential calculus of Cipriani and Sauvageot. One classical example of a (commutative) Dirichlet form is the Dirichlet energy on \( \mathbb{R}^n \), that is,

\[
\mathcal{E}(u) = - \int u \Delta u \, dx.
\]

By partial integration, \( \mathcal{E} \) can equivalently be expressed as

\[
\mathcal{E}(u) = \int_{\mathbb{R}^n} |\nabla u|^2,
\]

and \( \nabla \) is a derivation in the sense that it satisfies the product rule \( \nabla(uv) = u\nabla v + v\nabla u \).

Now, if \( \mathcal{E} \) is a Dirichlet form on the noncommutative \( L^2 \) space \( L^2(\mathcal{M}, \tau) \) with generator \( \mathcal{L} \), the first order differential calculus of Cipriani and Sauvageot (Theorem 1.9) asserts that it can be represented in the same way, that is, there exists a Hilbert bimodule \( \mathcal{H} \) and an operator \( \partial \) with values in \( \mathcal{H} \) such that

\[
\mathcal{E}(a) = \| \partial a \|^2_{\mathcal{H}}
\]

and \( \partial \) satisfies the product rule \( \partial(ab) = a\partial b + (\partial a)b \). Notice that unlike in the case of the Dirichlet energy on \( \mathbb{R}^n \), the left and right multiplication on \( \mathcal{H} \) may be different.

A priori, the left and right multiplication on \( \mathcal{H} \) are only defined for elements in \( D(\mathcal{E}) \cap \mathcal{M} \). In Section 2 we study when they can be extended to all of \( \mathcal{M} \). It turns
out that this question is closely related to the carré du champ
\[ \Gamma(a)(x) = \langle x\partial a, \partial a \rangle_H. \]

The carré du champ is \( \sigma \)-weakly continuous (in the commutative case, this holds if and only if the energy measure is absolutely continuous with respect to the reference measure) for all \( a \in D(E) \) if and only if the left and right multiplication have a \( \sigma \)-weakly continuous extension to all of \( M \) (Theorem 2.4). For the rest of the article we work under the standing assumption that this property holds.

In the classical Otto calculus on Euclidean space, the tangent space of \( \mathcal{P}(\mathbb{R}^n) \) at \( \rho \, dx \) is identified with \( H^1(\mathbb{R}^n, \rho \, dx) \). In particular, the norm of the tangent vector \( \psi \) is given by \( \int |\nabla \psi|^2 \rho \, dx \). In our framework, the latter could be replaced by \( \langle \rho \partial a, \partial a \rangle_H \) or \( \langle (\partial a)\rho, \partial a \rangle_H \) or any “mean” of these two extreme cases (these two examples correspond to the left and right trivial mean).

We study a class of such means in Section 3. For a suitable mean \( \theta \) we define \( \hat{\rho} = \theta(L(\rho), R(\rho)) \). A key technical role in this study plays the Lipschitz algebra \( A_\theta \) of bounded elements \( a \) of \( D(E) \) such that
\[ \|\partial a\|_\rho^2 = \langle \hat{\rho} \partial a, \partial a \rangle_H \]
is bounded on the space of density matrices. In this case, \( \|\partial a\|_\rho^2 \) depends upper semicontinuously on \( \rho \) (Theorem 3.15). The elements of \( A_\theta \) can be tested uniformly against \( \hat{\rho} \) for all density matrices \( \rho \), which makes them a suitable choice of test “function” in the continuity equation discussed below.

While the discussion of Section 3 is rather general, we will later focus on the logarithmic mean
\[ \theta(s, t) = \frac{s - t}{\log s - \log t}, \]
which was already identified by Maas and Mielke as the correct mean to study gradient flows of the entropy \( \text{Ent}(\rho) = \tau(\rho \log \rho) \).

In Section 4 we introduce the noncommutative transport metric \( \mathcal{W} \) via an analogue of the Benamou–Brenier formula (Definition 4.12). It is given as the length metric associated with the action functional
\[ (\rho_t) \mapsto \int \langle \hat{\rho}_t D\rho_t, D\rho_t \rangle_H \, dt \]
defined on a class of admissible curves, where the velocity vector field \( (D\rho_t) \) is determined by the abstract continuity equation
\[ \dot{\rho}_t = \partial^\ast(\hat{\rho}_t D\rho_t) \]
in a suitable weak sense. Further we establish some basic properties such as the convexity of \( \mathcal{W} \) (Lemma 4.24) and lower semicontinuity of the action functional with respect to pointwise weak convergence in \( L^1 \) (Theorem 4.30).

Section 5, which is quite technical in nature, deals with the entropy and the Fisher information. The latter is formally given by \( \mathcal{I}(\rho) = \mathcal{E}(\rho, \log \rho) \), but this expression suffers from several regularity issues (not all density matrices are in the domain of \( \mathcal{E} \), the logarithm is not a Lipschitz function), so we spend much of this section giving a rigorous definition via approximation and showing that several different approximations yield the same result. Then we go on to show that solutions of the quantum master equation are admissible curves in the definition of \( \mathcal{W} \) (Proposition 5.22, Corollary 5.24) and the entropy dissipation rate along these curves is given by the Fisher information (Proposition 5.23).

In Section 6 we come to the identification of the flow defined by the Markovian quantum master equation with the gradient flow of the entropy. First, we introduce the gradient estimate \( \mathcal{G}(K, \infty) \) in Definition 6.11. If \( \mathcal{E} \) is the Dirichlet energy on
a Riemannian manifold, then GE($K, \infty$) reduces to the well-known Bakry–Émery gradient estimate
\[ \Gamma(P_t f) \leq e^{-2Kt} P_t \Gamma(f), \]
which is equivalent to Ric \( \geq K \).

After a technical subsection singling out a suitable class of regular curves, we can then show in Theorem 6.26 that GE($K, \infty$) implies that the flow defined by the Markovian quantum Master equation
\[ \dot{\rho}_t = -\mathcal{L} \rho_t \]
is an EVI$_K$ gradient flow of the entropy, that is,
\[ \frac{1}{2} \frac{d}{dt} W(P_t \rho, \sigma)^2 + \frac{K}{2} W(P_t \rho, \sigma)^2 + \text{Ent}(P_t \rho) \leq \text{Ent}(\sigma). \]
In Section 7 we study consequences of the gradient flow characterization with regard to semi-convexity of the entropy along geodesics. The EVI$_K$ implies that the distance $W$ between two density matrices with finite entropy can be realized by a sequence of curves with uniformly bounded entropy. Combining this fact with the weak $L^1$-compactness of sublevel sets of the entropy, we conclude that density matrices with finite entropy are joined by a minimizing geodesic (Theorem 7.7). Finally, by an abstract result on gradient flows, the entropy is $K$-convex along geodesics, that is,
\[ \text{Ent}(\rho_t) \leq (1 - t) \text{Ent}(\rho_0) + t \text{Ent}(\rho_1) - \frac{K}{2} t(1 - t) W(\rho_0, \rho_1)^2. \]
The relation between GE($K, \infty$), EVI$_K$ and geodesic $K$-convexity are summarized in Theorem 7.12.

The content of this article was included in the author’s PhD thesis at the University of Jena.

Acknowledgments. The author would like to express his gratitude to Daniel Lenz for the support and helpful advice as his supervisor. He wants to thank Matthias Erbar and Jan Maas for fruitful discussions on the topic of this article, Aljosha Sukeylo for the translation of the article [Tik87] and Simon Puchert for the proof of Lemma 5.11.

After the project had grown, we learned that ideas for a related, but slightly different definition of the metric $W$ had also been developed independently by David Hornshaw [Hor18]. The author is grateful for the exchange of draft versions and discussions.

1. Quantum Dirichlet Forms

In this section we give a short overview over the theory of noncommutative integration and noncommutative Dirichlet forms, and show how some examples encountered later fit into that framework. In particular, we review the first order differential calculus developed by Cipriani and Sauvageot, which will be a central tool later on.

To begin, let us recall some basics of noncommutative integration theory as developed in [Seg53a, Seg53b]; a good overview is given in [PX03].

An algebra $\mathcal{M}$ of bounded operators on a Hilbert space $H$ is called von Neumann algebra if it is closed under taking adjoints, contains 1 and is closed in the weak operator topology. The set $\mathcal{M}' = \{ x \in B(H) \mid xy = yx \text{ for all } y \in \mathcal{M} \}$ is called the commutant of $\mathcal{M}$. By the bicommutant theorem, a unital $*$-subalgebra of $B(H)$ is a von Neumann algebra if and only if $\mathcal{M} = \mathcal{M}'$. The set of all positive operators in $\mathcal{M}$ is denoted by $\mathcal{M}_+$. 
Let $\mathcal{M}$ be a von Neumann algebra. A map $\tau: \mathcal{M}_+ \to [0, \infty]$ is called a weight if $\tau(\lambda x) = \lambda \tau(x)$ and $\tau(x + y) = \tau(x) + \tau(y)$ for all $\lambda \geq 0$, $x, y \in \mathcal{M}_+$ (with the convention $0 \cdot \infty = 0$).

The weight $\tau$ is called

- **normal** if $\tau(\text{sup} \{x_i\}) = \text{sup} \{\tau(x_i)\}$ for every increasing net $(x_i)$ in $\mathcal{M}_+$,
- **semi-finite** if $\{x \in \mathcal{M}_+ \mid \tau(x) < \infty\}$ generates $\mathcal{M}$ as von Neumann algebra,
- **faithful** if $\tau(x^* x) = 0$ implies $x = 0$,
- **tracial** or a trace if $\tau(x^* x) = \tau(xx^*)$ for all $x \in \mathcal{M}$.

We say that $\tau$ is an n.s.f. trace if it is a normal, semi-finite, faithful, tracial weight, and call the pair $(\mathcal{M}, \tau)$ a tracial von Neumann algebra. Every n.s.f. trace $\tau$ induces a faithful normal representation $\pi_\tau$ on a Hilbert space $H_\tau$. We will routinely identify $\mathcal{M}$ with $\pi_\tau(\mathcal{M})$.

A closed, densely defined operator $x$ is said to be affiliated with $\mathcal{M}$ if $xu = ux$ for every unitary $u \in \mathcal{M}'$. A self-adjoint operator $x$ is affiliated with $\mathcal{M}$ if and only if $\varphi(x) \in \mathcal{M}$ for every bounded Borel function $\varphi: \mathbb{R} \to \mathbb{R}$. An affiliated operator $x$ is called $\tau$-measurable if $\tau(1_{(\lambda, \infty)}(|x|)) < \infty$ for some $\lambda \geq 0$. The set of all $\tau$-measurable operators is denoted by $L^0(\mathcal{M}, \tau)$.

The trace $\tau$ can be extended to the set of positive $\tau$-measurable operators via

$$\tau(x) = \int_{[0, \infty)} \lambda d(\tau \circ e)(\lambda),$$

where $e$ denotes the spectral measure of $x$. Equivalently, $\tau(x) = \sup_{n \in \mathbb{N}} \tau(x \wedge n)$.

The noncommutative $L^p$ spaces are defined as

$$L^p(\mathcal{M}, \tau) = \{x \in L^0(\mathcal{M}, \tau) \mid \tau(|x|^p) < \infty\}$$

for $p \in [1, \infty)$ and endowed with the norm $\|\|_p = \tau(|\cdot|^p)^{1/p}$. For $p = \infty$ one sets $L^\infty(\mathcal{M}, \tau) = \mathcal{M}$.

The space $L^p(\mathcal{M}, \tau)$ coincides with the completion of $\{x \in \mathcal{M} \mid \tau(|x|^p) < \infty\}$ with respect to $\|\|_p$. In particular, it is a Banach space. Moreover, $L^2(\mathcal{M}, \tau) \cong H_\tau$ with the action of $\mathcal{M}$ on $L^2(\mathcal{M}, \tau)$ given by left multiplication. In the following, we will always identify $\mathcal{M}$ with its image in $B(L^2(\mathcal{M}, \tau))$.

**Example 1.1.** If $(X, \mathcal{B}, m)$ is a localizable (for example $\sigma$-finite) measure space, then $L^\infty(X, m)$ is a von Neumann algebra (acting on $L^2(X, m)$ by multiplication) and the functional

$$\tau_m: L^\infty_+(X, m) \to [0, \infty], \quad \tau_m(f) = \int_X f \, dm$$

is an n.s.f. trace. Moreover, every commutative tracial von Neumann algebra arises in this way.

The space $L^p(L^\infty(X, m), \tau_m)$ is isometrically isomorphic to $L^p(X, m)$, and the isomorphism can be chosen in such a way that it is consistent for all $p \in [1, \infty]$.

**Remark 1.2.** If $A$ is a $C^*$-algebra and $\tau: A_+ \to [0, \infty]$ a lower semicontinuous, semi-finite, faithful trace, one also gets a GNS representation $\pi_\tau$ on a Hilbert space $H_\tau$. In this case, $L^\infty(A, \tau) = \pi_\tau(A)'$ is a von Neumann algebra and $\tau$ extends to an n.s.f. weight on $L^\infty(A, \tau)$. One sets $L^p(A, \tau) := L^p(L^\infty(A, \tau), \tau)$.

From a theoretical point of view one can therefore always assume to be given a tracial von Neumann algebra $(\mathcal{M}, \tau)$, but in the examples it will sometimes be more convenient to give a description in terms of $(A, \tau)$.

Denote by $L^p_h(\mathcal{M}, \tau)$ the set of self-adjoint (as operators on $H_\tau$) elements of $L^p(\mathcal{M}, \tau)$. One advantage of the definition via affiliated operators over that as abstract completion is the fact that we can use functional calculus for elements of $L^p_h(\mathcal{M}, \tau)$. 


For the next lemma recall that for a nonempty, closed, convex subset $C$ of a Hilbert space $H$ and $x \in H$ there is a unique element $y \in C$ with $\|x - y\| = \inf_{z \in C}\|x - z\|$. The map $P_C: x \mapsto y$ is called (metric) projection onto $C$. The element $P_C(x)$ can alternatively be characterized as the unique $y \in C$ such that
\[ \text{Re}(x - y, z - y) \leq 0 \]
for all $z \in C$.

We write $\alpha \wedge \beta = \min\{\alpha, \beta\}$ and $\alpha \lor \beta = \max\{\alpha, \beta\}$ for $\alpha, \beta \in \mathbb{R}$. If $x$ is a self-adjoint operator, $x \wedge \alpha$ stands for the application of the function $\min\{\cdot, \alpha\}$ to $x$, which is the infimum of $x$ and $\alpha 1$ in the (commutative) unital $C^*$-algebra generated by $x$.

**Lemma 1.3.** Let $(\mathcal{M}, \tau)$ be a tracial von Neumann algebra and let $C$ be the closure of $\{x \in L^2_b(\mathcal{M}, \tau) \cap \mathcal{M} \mid x \leq 1\}$ in $L^2(\mathcal{M}, \tau)$. Then $C$ is convex and the projection $P_C$ onto $C$ is given by $P_C(a) = a \wedge 1$ for all $a \in L^2_b(\mathcal{M}, \tau)$.

**Proof.** It is easy to see that $C$ is convex. For $a \in L^2_b(\mathcal{M}, \tau)$ let $a_n = (a \wedge 1) \lor (-n)$. Then $a_n \in L^2_b(\mathcal{M}, \tau) \cap \mathcal{M}$, $a_n \leq 1$ and $a_n \rightarrow a \wedge 1$ in $L^2(\mathcal{M}, \tau)$, hence $a \wedge 1 \in C$. If $b \in \mathcal{M} \cap L^2_b(\mathcal{M}, \tau)$ with $b \leq 1$, then
\[ \tau((a - a \wedge 1)(b - a \wedge 1)) = \tau((a - 1)^{1/2}(b - a \wedge 1)(a - 1)^{1/2}) \]
\[ \leq \tau((a - 1)^{1/2}(1 - a \wedge 1)(a - 1)^{1/2}) \]
\[ = \tau((a - 1)_+(a - 1)_-) \]
\[ = 0. \]

For arbitrary $b \in C$, the inequality above follows by continuity. Thus $P_C(a) = a \wedge 1$. \(\square\)

Now we can turn to the theory of Dirichlet forms and Markovian semigroups in the noncommutative setting. For some basic references see [AH77, DL92], for the first order differential calculus described below see [CS03] and the expository article [Cip08].

A quadratic form $\mathcal{E}: L^2(\mathcal{M}, \tau) \rightarrow [0, \infty]$ is real if $\mathcal{E}(a^*) = \mathcal{E}(a)$ for all $a \in L^2(\mathcal{M}, \tau)$ and Markovian if $\mathcal{E}(a \wedge 1) \leq \mathcal{E}(a)$ for all $a \in L^2_b(\mathcal{M}, \tau)$. The lemma above shows that the cut-off $a \wedge 1$ can be understood either as an application of functional calculus or as projection in $L^2(\mathcal{M}, \tau)$.

By the next lemma (see [DL92, Proposition 2.12] and [CS03, Theorem 10.2]), Markovian forms automatically satisfy a stronger contraction property with respect to Lipschitz functional calculus.

**Lemma 1.4.** A closed real quadratic form $\mathcal{E}: L^2(\mathcal{M}, \tau) \rightarrow [0, \infty]$ is Markovian if and only if $\mathcal{E}(f(a)) \leq \mathcal{E}(a)$ for all $a \in L^2_b(\mathcal{M}, \tau)$ and all 1-Lipschitz functions $f: \mathbb{R} \rightarrow \mathbb{R}$ with $f(0) = 0$.

For $n \in \mathbb{N}$ denote by $\text{tr}_n$ the normalized trace on $M_n(\mathbb{C})$ and let $\tau_n = \tau \otimes \text{tr}_n$ on $(\mathcal{M} \otimes M_n(\mathbb{C}))+ \cong M_n(\mathcal{M})+$, that is,
\[ \tau_n: M_n(\mathcal{M})+ \rightarrow [0, \infty], \quad \tau_n((a_{ij})) = \frac{1}{n} \sum_{i=1}^n \tau(a_{ii}). \]

The quadratic form $\mathcal{E}$ can be extended to $L^2(M_n(\mathcal{M}), \tau_n)$ via
\[ \mathcal{E}_n: L^2(M_n(\mathcal{M}), \tau_n) \rightarrow [0, \infty], \quad \mathcal{E}_n((a_{ij})) = \sum_{i,j=1}^n \mathcal{E}(a_{ij}). \]

We say that $\mathcal{E}$ is completely Markovian if $\mathcal{E}_n$ is Markovian for all $n \in \mathbb{N}$. 
A lower semicontinuous, densely defined, real, completely Markovian quadratic form \( \mathcal{E} \) on \( L^2(\mathcal{M}, \tau) \) is called \textit{completely Dirichlet form} on \( (\mathcal{M}, \tau) \).

\textbf{Remark 1.5.} For every quadratic form \( q \) on a Hilbert space \( H \) there is an associated sesquilinear form \( \tilde{q} \) defined as

\[
\tilde{q}: D(q) \times D(q) \to \mathbb{C}, \quad \tilde{q}(u, v) = \frac{1}{4} \sum_{k=0}^{3} i^{k} q(u + i^{k} v),
\]

where \( D(q) = \{ u \in H \mid q(u) < \infty \} \). We will use these two points of view interchangeably and write \( q \) for both of these maps.

\textbf{Remark 1.6.} In [CS03], an additional condition called \textit{regularity} is imposed in most results. This property depends not only on the form \( \mathcal{E} \), but also on the choice of some \( C^* \)-subalgebra of \( \mathcal{M} \). Every completely Dirichlet form \( \mathcal{E} \) is regular with respect to the norm closure of \( D(\mathcal{E}) \cap \mathcal{M} \).

\textbf{Example 1.7.} Let \( (X, \mathcal{B}, m) \) be a localizable measure space. Every Markovian form on \( L^\infty(X, m) \) is completely Markovian so that Dirichlet forms on \( L^2(X, m) \) in the sense of Beurling–Deny [BD58, BD59] can be identified with completely Dirichlet forms on \( L^2(L^\infty(X, m), \tau_m) \).

There is a bijective correspondence between quantum Dirichlet forms and quantum sub-Markov semigroups on \( (\mathcal{M}, \tau) \) analogous to the commutative case: The semigroup \( (P_t) \) generated by a positive self-adjoint operator \( \mathcal{L} \) on \( L^2(\mathcal{M}, \tau) \) is sub-Markovian, that is, \( 0 \leq P_t(a) \leq 1 \) for \( 0 \leq a \leq 1 \), if and only if the quadratic form generated by \( \mathcal{L} \) is a Markovian form (see [AH77] Theorems 2.7, 2.8] in the finite case and [DL92] Theorems 2.13, 3.3] in the semi-finite case).

Moreover, \( (P_t) \) extends uniquely to strongly continuous semigroups on \( L^p(\mathcal{M}, \tau) \) for \( p \in [1, \infty) \) and to a quantum sub-Markov semigroup on \( \mathcal{M} \). We will usually denote these extensions by the same symbol, occasionally also writing \( (P_t^{(p)}) \) when the space on which the semigroup acts is important. Similarly, \( L^{(p)} \) denotes the generator of \( (P_t) \) on \( L^p(\mathcal{M}, \tau) \).

The curve \( (P_t^{(p)} a)_{t \geq 0} \) is the unique (mild) solution of the initial value problem for the Markovian quantum master equation

\[
\begin{cases}
\dot{x}_t = -\mathcal{L}^{(p)} x_t, \\
x_0 = a
\end{cases}
\]

in \( L^p(\mathcal{M}, \tau) \).

The semigroup \( (P_t)_{t \geq 0} \) is called \textit{conservative} if \( P_1 1 = 1 \). As we want to study the evolution on density matrices, conservativeness is a natural assumption, so we reserve a special name for the associated Dirichlet forms (motivated by the term quantum Markov semigroup for the corresponding semigroup on \( \mathcal{M} \)).

\textbf{Definition 1.8 (Quantum Dirichlet form).} A completely Dirichlet form \( \mathcal{E} \) is called \textit{quantum Dirichlet form} if the associated semigroup \( (P_t) \) is conservative.

The following representation theorem for completely Dirichlet forms by Cipriani and Sauvageot (see [CS03] Theorems 4.7, 8.2, 8.3]) is central to our investigations.

\textbf{Theorem 1.9 (First order differential calculus).} Let \( \mathcal{E} \) be a quantum Dirichlet form on the tracial von Neumann algebra \( (\mathcal{M}, \tau) \) and \( \mathcal{C} = D(\mathcal{E}) \cap \mathcal{M} \).

Then \( \mathcal{C} \) is a \( * \)-algebra and there exist a Hilbert space \( \mathcal{H} \), commuting non-degenerate \( * \)-representations \( L \) of \( \mathcal{C} \) and \( R \) of \( \mathcal{C}^* \) on \( \mathcal{H} \), an anti-linear isometric involution \( J: \mathcal{H} \to \mathcal{H} \), and a closed operator \( \partial: D(\mathcal{E}) \to \mathcal{H} \) such that

\begin{itemize}
\item \( \{ L(a) \partial b \mid a, b \in \mathcal{C} \} \) is dense in \( \mathcal{H} \),
\end{itemize}
Remark

The first order differential calculus associated with $\mathcal{E}$. The $\ast$-representations $L, R$ are to be understood as left and right multiplication of $\mathcal{C}$ on $\mathcal{H}$. Accordingly, we will write $a \cdot \xi$ and $\xi \cdot b$ for $L(a)\xi$ and $R(b)\xi$, respectively.

**Remark 1.10.** In the theorem, $\mathcal{C}^\circ$ denotes the opposite algebra of $\mathcal{C}$, that is, the $\ast$-algebra with same underlying vector space and involution, but with multiplication given by $a \circ b = ba$ for $a, b \in \mathcal{C}$.

**Remark 1.11.** Instead of conservativeness, it suffices to assume that the killing term of $\mathcal{E}$ vanishes in the sense of [CS03, Theorem 8.1].

An important consequence of the product rule for the first order differential calculus is a (two-variable) chain rule. For that purpose, let

$$\tilde{f}: I \times I \rightarrow \mathbb{R}, \quad \tilde{f}(s, t) = \begin{cases} \frac{f(s) - f(t)}{s - t} & \text{if } s \neq t, \\ f'(s) & \text{if } s = t \end{cases}$$

for $f \in C^1(I)$. The function $\tilde{f}$ is sometimes called the quantum derivative of $f$. With this notation, the chain rule reads as follows ([CS03, Lemma 7.2]).

**Lemma 1.12** (Chain rule). If $f \in C^1(\mathbb{R})$ has bounded derivative and $f(0) = 0$, then

$$\partial f(a) = \tilde{f}(L(a), R(a))\partial a.$$  

for all $a \in D(\mathcal{E})$.

If $\mathcal{E}$ is strongly local, then $L = R$ and one recovers the usual chain rule $\partial f(a) = f'(a)\partial a$.

**Example 1.13** (Weighted graphs). Let $X$ be a countable set, $m: X \rightarrow (0, \infty)$ and $b: X \times X \rightarrow [0, \infty)$ such that

- $b(x, x) = 0$ for all $x \in X$,
- $b(x, y) = b(y, x)$ for all $x, y \in X$,
- $\sum_y b(x, y) < \infty$ for all $x \in X$.

The triple $(X, b, m)$ is called a weighted graph (compare [KL10, KL12]). Often one allows for an additional killing weight $c: X \rightarrow [0, \infty)$, but the associated Dirichlet form will never be conservative if $c \neq 0$, so we drop it from the beginning.

The associated Dirichlet form with Neumann boundary conditions is

$$\mathcal{E}^{(N)}: \ell^2(X, m) \rightarrow [0, \infty], \quad \mathcal{E}^{(N)}(u) = \frac{1}{2} \sum_{x,y} b(x, y)|u(x) - u(y)|^2.$$  

The associated Dirichlet form with Dirichlet boundary conditions $\mathcal{E}^{(D)}$ is the closure of the restriction of $\mathcal{E}^{(D)}$ to $C_c(X)$.

The first order differential calculus associated with $\mathcal{E}^{(N)}$ is given by $\mathcal{H} = \ell^2(X \times X, \frac{1}{2}b)$, $(u \cdot \xi)(x, y) = u(x)\xi(x, y)$, $(\xi \cdot v)(x, y) = \xi(x, y)v(y)$, $\partial u(x, y) = u(x) - u(y)$ and $(J\xi)(x, y) = -\xi(y, x)$. 
The first order differential calculus associated with $\mathcal{E}^{(D)}$ is obtained by suitable restriction.

**Example 1.14** (Riemannian manifolds). Let $(M, g)$ be a complete Riemannian manifold and $\mathcal{E}$ the standard Dirichlet integral

$$\mathcal{E}: L^2(M) \longrightarrow [0, \infty], \quad \mathcal{E}(u) = \begin{cases} \int_M |\nabla u|^2 \, d\text{vol}_g & \text{if } \nabla u \in L^2(M), \\ \infty & \text{otherwise.} \end{cases}$$

The first order differential calculus for $\mathcal{E}$ is given by $\mathcal{H} = L^2(M; TM)$, $(u\xi)(x) = (\xi u)(x) = u(x)\xi(x)$, $\partial = \nabla$ and $J\xi = \overline{\xi}$.

**Example 1.15** (Metric measure spaces). If $(X, d, m)$ is an infinitesimally Hilbertian metric measure space (see [AGS14b]) and $\mathcal{E}$ the associated Dirichlet form (twice the Cheeger energy), then the first order differential calculus described above coincides with first order differential calculus developed in [Gig14].

Notice that the crucial difference between Example 1.13 on the one hand and Examples 1.14, 1.15 on the other hand is that left and right multiplication on $\mathcal{H}$ coincide for the Dirichlet forms on Riemannian manifolds and metric measure spaces while they differ for graphs. More generally, left and right multiplication coincide in the commutative setting whenever $\mathcal{E}$ is a strongly local regular Dirichlet form (see [IRT12, Theorem 2.7]).

**Example 1.16** (Noncommutative torus). Let $\theta \in (0,1)$ be irrational and let $U, V \in B(H)$ be unitaries with $VU = e^{2\pi i \theta} UV$. The unital $C^*$-algebra $A_\theta$ generated by $U, V$ is called *noncommutative torus* (and, up to $*$-isomorphism, it is indeed independent of the choice of $U, V$). Let $A_\theta$ be the linear hull of $\{U^mV^n \mid m, n \in \mathbb{Z}\}$, which is clearly a dense $*$-subalgebra of $A_\theta$.

The map

$$\tau: A_\theta \longrightarrow \mathbb{C}, \quad \tau(U^mV^n) = \delta_{m,0}\delta_{n,0}$$

extends to a tracial state on $A_\theta$. Furthermore, the map

$$P_t: A_\theta \longrightarrow A_\theta, \quad P_t(U^mV^n) = e^{-t(m^2+n^2)}U^mV^n$$

extends to a bounded linear operator on $L^2(A_\theta, \tau)$ and $(P_t)_{t \geq 0}$ is a $\tau$-symmetric quantum Markov semigroup, called *noncommutative heat semigroup*. The associated Dirichlet form $\mathcal{E}$ acts on $A_\theta$ as

$$\mathcal{E}\left( \sum_{m,n \in \mathbb{Z}} \alpha_{m,n}U^mV^n \right) = \sum_{m,n \in \mathbb{Z}} (m^2 + n^2)|\alpha_{m,n}|^2.$$

Let $\partial_1, \partial_2: A_\theta \longrightarrow A_\theta$ be defined by $\partial_1(U^mV^n) = imU^mV^n$ and $\partial_2(U^mV^n) = inU^mV^n$, and let $\mathcal{H}$ be the the closed linear hull of $\{(a\partial_1(b), a\partial_2(b)) \mid a, b \in A_\theta\}$ in $L^2(A_\theta, \tau) \oplus L^2(A_\theta, \tau)$.

The first order differential calculus associated with $\mathcal{E}$ is given by $\partial = \partial_1 \oplus \partial_2$, $L(a)(u,v) = (au, av)$, $R(b)(u,v) = (ub, vb)$, $J(u,v) = -(v^*, u^*)$.

**Example 1.17** (Fermionic Clifford algebra). Let $H$ be an infinite-dimensional, separable real Hilbert space and $\mathcal{C}l(H)$ the Clifford $C^*$-algebra over $H$ (see [SS64]). It is well-known that $\mathcal{C}l(H)$ is a simple $C^*$-algebra with a unique tracial state $\tau$. The von Neumann algebra $L^\infty(\mathcal{C}l(H), \tau)$ is the hyperfinite type $II_1$ factor.

Let $(e_i)_{i \in \mathbb{N}}$ be an orthonormal basis of $H$. The linear hull of all products of the form $e_{i_1}\cdots e_{i_k}$ with $i_1 < \cdots < i_k$ and $k \in \{0,1,\ldots\}$ is a dense $*$-subalgebra of $\mathcal{C}l(H)$, which we denote by $\mathcal{A}$. 


Let $\mathcal{F}_-(H)$ be the fermionic Fock space over $H$, that is, $\mathcal{F}_-(H) = \bigoplus_{k \geq 0} \wedge^k H$.

The map

$$A \rightarrow \mathcal{F}_-(H), \quad \sum_{i_1 < \cdots < i_k} \alpha_{i_1 \cdots i_k} e_{i_1} \cdots e_{i_k} \mapsto \sum_{i_1 < \cdots < i_k} \alpha_{i_1 \cdots i_k} e_{i_1} \wedge \cdots \wedge e_{i_k}$$

extends to an isometric isomorphism $\Phi: L^2(\mathcal{C}^\ell(H), \tau) \rightarrow \mathcal{F}_-(H)$, the Chevalley–Segal isomorphism.

The number operator on $\mathcal{F}_-(H)$ is defined by

$$D(N) = \{ (\psi_k) \in \mathcal{F}_-(H) \mid \sum_{k \geq 0} k^2 \| \psi_k \|^2_{\Lambda^k H} < \infty \}, \quad N(\psi_k) = (k\psi_k),$$

and $\Phi^{-1} N \Phi$ generates a conservative quantum Dirichlet form $\mathcal{E}_N$ on $L^2(\mathcal{C}^\ell(H), \tau)$.

Let $a_i$ be the annihilation operator on $\mathcal{F}_-(H)$ characterized by

$$a_i(e_{j_1} \wedge \cdots \wedge e_{j_k}) = \frac{1}{\sqrt{k}} \sum_{l=1}^k (-1)^l e_{j_1} \wedge \cdots \wedge e_{j_l} \wedge \cdots e_{j_k}$$

and by $\gamma: L^\infty(\mathcal{C}^\ell(H), \tau) \rightarrow L^\infty(\mathcal{C}^\ell(H), \tau)$ the grading operator.

The first order differential calculus for $\mathcal{E}_N$ is given by $\mathcal{H} = \sum_{i \geq 0} L^2(\mathcal{C}^\ell(H), \tau), L(x)(\xi_i) = (x\xi_i), R(x)(\xi_i) = (\gamma(x)\xi_i), J(\xi_i) = - (\xi_i^*)$ and $\partial = \bigoplus_{i \geq 0} \Phi^{-1} a_i \Phi$.

2. Carré du champ

In this section we study the question of when the first order differential calculus introduced in the last section can be extended to $\mathcal{M}$. It turns out that this question is closely related to the so-called carré du champ operator defined below. More precisely we show in [2.2] that the carré du champ $\Gamma(a)$ has a density with respect to $\tau$ if and only if the left and right action of $D(\mathcal{E}) \cap \mathcal{M}$ have normal extensions to $\mathcal{M}$. This provides a characterization of the noncommutative analogue of energy dominant measures.

Throughout the section let $(\mathcal{M}, \tau)$ be a tracial von Neumann algebra, $\mathcal{E}$ a quantum Dirichlet form on $L^2(\mathcal{M}, \tau)$, $\mathcal{C} = D(\mathcal{E}) \cap \mathcal{M}$, and $(\partial, \mathcal{H}, L, R, J)$ the associated first order differential calculus.

The carré du champ $\Gamma$ of $\mathcal{E}$ is defined as

$$\Gamma: \mathcal{C} \times \mathcal{C} \rightarrow \mathcal{C}^*, \quad \Gamma(a, b)(x) = \langle x \partial a, \partial b \rangle_{\mathcal{H}}.$$

We write $\Gamma(a)$ for $\Gamma(a, a)$. It is easy to see that $\Gamma$ is sesquilinear and $\|\Gamma(a, b)\|_{\mathcal{C}^*} \leq \mathcal{E}(a)^{1/2} \mathcal{E}(b)^{1/2}$ for all $a, b \in \mathcal{C}$.

Remark 2.1. In terms of $\mathcal{E}$, the carré du champ can be expressed as

$$\Gamma(a)(x) = \frac{1}{2}(\mathcal{E}(a, ax^*) + \mathcal{E}(ax, a) - \mathcal{E}(a^*a, x^*))$$

for all $a, x \in \mathcal{C}$.

For $\xi = \sum_i a_i \partial b_i$ we define (compare [HRT13] in the commutative case)

$$\Gamma_H(\xi) = \sum_{i,k} a_i \Gamma(b_i, b_k) a_k^*.$$
are equivalent: 
richlet form on the tracial von Neumann algebra 

adjoints are $(i)$ 

Proof. $D$ 

rem II.4.8] asserts that 

Then $L$ 

on $M$ 

Remark 2.3 

dense in $M$ 

Proof. Let $(P_t)_{t \geq 0}$ be the quantum Markov semigroup associated with $E$. If $a \in L^2(M, \tau) \cap M$, then $P_t(a) \in D(E) \cap M$ for all $t > 0$ and $P_t(a) \to a$ weakly as $t \searrow 0$. Now the assertions follows from the fact that $L^2(M, \tau) \cap M$ is weakly dense in $M$. 

Remark 2.3. Since $C$ is a $*$-algebra, the Kaplansky density theorem ([Tak02, Theorem II.4.8]) asserts that $D(E) \cap M_1$ is even strongly dense in $M_1$, where $M_1$ denotes the unit ball in $M$. 

Theorem 2.4 (Characterization energy dominant trace). Let $E$ be a quantum Dirichlet form on the tracial von Neumann algebra $(M, \tau)$. The following assertions are equivalent:

(i) $L$ is weakly continuous 
(ii) $R$ is weakly continuous 
(iii) $\Gamma(a)$ is weakly continuous for all $a \in C$ 
(iv) $\Gamma_H(\xi)$ is weakly continuous for all $\xi \in H$ 

Proof. (i) $\iff$ (ii): Since multiplication by a fixed bounded operator and taking adjoints are weakly continuous, the equivalence of (i) and (ii) follows from $L(\cdot) = JR(\cdot) J$. 

(iii) $\implies$ (iv): It is easy to see that $\Gamma_H(\xi)$ is weakly continuous for $\xi \in \text{lin}\{a\partial b \mid a, b \in C\}$. Combined with the fact that the norm limit of weakly continuous functionals is weakly continuous, (iv) follows.
Hence the tracial von Neumann algebra \((\xi_n)_{n=1}^\infty\) is uniformly continuous with respect to the \(\sigma\)-\(\xi\)-weak topology (see [Rud91, Theorem 1.17]). Thus, by [Bou89, Theorem II.2], for every \(\xi \in \mathcal{H}\) there is a unique \(\sigma\)-\(\xi\)-weakly continuous extension of \(\Gamma_{\mathcal{H}}(\xi)\) to \(\mathcal{M}\) with the same norm. We continue to write \(\Gamma_{\mathcal{H}}(\xi)\) for this extension.

For \(L\) to be \(\sigma\)-\(\xi\)-weakly continuous it suffices to show that \(\varphi \circ L\) is \(\sigma\)-\(\xi\)-weakly continuous for all \(\varphi \in B(\mathcal{H})_+\). Every \(\varphi \in B(\mathcal{H})_+\) is of the form \(\varphi = \sum_n \langle \xi_n, \eta_n \rangle_{\mathcal{H}}\) for sequences \((\xi_n), (\eta_n)\) in \(\mathcal{H}\) such that \(\sum_n (\|\xi_n\|_H^2 + \|\eta_n\|_H^2) < \infty\). Then

\[
\sum_{n=1}^\infty \|\Gamma_{\mathcal{H}}(\xi_n, \eta_n)\|_{\mathcal{M}^*} \leq \sum_{n=1}^\infty \|\xi_n\|_H \|\eta_n\|_H \leq \frac{1}{2} \sum_{n=1}^\infty \left(\|\xi_n\|_H^2 + \|\eta_n\|_H^2\right).
\]

Hence \(\sum_n \Gamma_{\mathcal{H}}(\xi_n, \eta_n)\) converges absolutely with respect to \(\|\cdot\|_{\mathcal{M}^*}\) to some \(\omega \in \mathcal{M}^*\). Since the space \(\mathcal{M}_*\) of \(\sigma\)-\(\xi\)-weakly continuous linear functionals is closed in \(\mathcal{M}^*\), we have \(\omega \in \mathcal{M}_*\).

Now let \((a_i)\) be a sequence in \(\mathcal{M}\) such that \(a_i \to 0\) \(\sigma\)-\(\xi\)-weakly. Then

\[
\sum_n (a_i \xi_n, \eta_n)_{\mathcal{H}} = \sum_n \Gamma_{\mathcal{H}}(\xi_n, \eta_n)(a_i) = \omega(a_i) \to 0.
\]

Hence \(L\) is \(\sigma\)-\(\xi\)-weakly continuous. \(\square\)

**Definition 2.5** (Energy dominant trace). Let \(\mathcal{E}\) be a quantum Dirichlet form on the tracial von Neumann algebra \((\mathcal{M}, \tau)\). We say that \(\tau\) is energy dominant if one of the equivalent assertions of Theorem 2.4 holds.

As already seen in the proof of Theorem 2.4 if the trace \(\tau\) is energy dominant, the functional \(\Gamma_{\mathcal{H}}(\xi)\) has a unique \(\sigma\)-\(\xi\)-weakly continuous extension to \(\mathcal{M}\) for all \(\xi \in \mathcal{H}\) of the same norm. Since \(L\) is non-degenerate, this extension is still a positive functional.

We denote by \(\Gamma_{\mathcal{H}}(\xi)\) the preimage of \(\Gamma_{\mathcal{H}}(\xi)\) under the isomorphism \(L^1(\mathcal{M}, \tau) \to \mathcal{M}_*\), \(x \mapsto \tau(x \cdot)\), that is, \(\Gamma_{\mathcal{H}}(\xi)\) is the unique element in \(L^1(\mathcal{M}, \tau)\) such that \(\Gamma_{\mathcal{H}}(\xi)(x) = \tau(x \Gamma_{\mathcal{H}}(\xi))\) for all \(x \in \mathcal{C}\). Similarly we define \(\Gamma(a) \in L^1(\mathcal{M}, \tau)\) for \(a \in D(\mathcal{E})\).

On the other hand, if \(\tau\) is energy dominant, also the left and right action \(L\) and \(R\) have unique \(\sigma\)-\(\xi\)-weakly continuous extensions \(\tilde{L}\) and \(\tilde{R}\) to \(\mathcal{M}\) and \(\mathcal{M}^*\), respectively. These extensions are characterized by

\[
\langle \tilde{L}(a)\xi, \eta \rangle_{\mathcal{H}} = \tau(a \Gamma_{\mathcal{H}}(\xi, \eta))
\]

for \(a \in \mathcal{M}\), \(\xi, \eta \in \mathcal{H}\), and similarly for \(\tilde{R}\).

Since the vector space operations as well as the multiplication and the involution on \(\mathcal{M}\) are all (separately) \(\sigma\)-\(\xi\)-weakly continuous, the extensions \(\tilde{L}\) and \(\tilde{R}\) are again *-homomorphisms. From now on we will denote these extensions simply by \(L\), \(R\).

**Remark 2.6.** If \(\mathcal{E}\) is a Dirichlet form on \(L^2(X, m)\), then \(\Gamma\) is twice the (linear functional induced by the) energy measure as defined in [POT94, Section 3.2]. In this case, the measure \(m\) is energy dominant if and only if \(\Gamma(u)\) is absolutely continuous with respect to \(m\) for all \(u \in D(\mathcal{E})\). This concept was introduced by Kusuoka [Kus89, Kus93] in the study of Dirichlet forms on fractals.

**Remark 2.7.** In the noncommutative setting, energy dominant traces were studied for example in [JZ15], where the corresponding semigroups are called noncommutative diffusion semigroups. We do not adopt this terminology as it conflicts with the well-established definition of diffusion semigroups in the commutative case.
Remark 2.8. For an irreducible local Dirichlet form $E$ it is always possible to construct an energy dominant measure $\mu$ such that $E$ is closable in $L^2(X, \mu)$, see [HRT13, Theorem 5.1]. In the noncommutative setting, it is not clear why an analogously constructed weight should be tracial.

3. Operator means and the algebra $A_\theta$

In this section we study means of the left and right action on $\mathcal{H}$, which will later appear both in the action functional and the constraint in the definition of the metric $W$. As an important tool we introduce the space $A_\theta$, which will take on the role of a space of test “functions”. We prove several continuity properties of these means, which are important technical tools for the remainder of the thesis, especially the semicontinuity property established in Theorem 3.15.

In the later sections we will focus on the logarithmic mean as it gives the connection to gradient flows of the entropy. In this section however we keep the discussion more general since means other than the logarithmic one have also proven useful in the commutative case (see for example [CLLZ17] for an application to evolutionary games).

Throughout this section let $(\mathcal{M}, \tau)$ be a tracial von Neumann algebra and $E$ a quantum Dirichlet form on $L^2(\mathcal{M}, \tau)$ with associated first-order differential calculus $(\partial, \mathcal{H}, L, R, J)$. We further assume that $\tau$ is energy dominant.

Since we assume $\tau$ to be energy dominant, the left and right action $L$ and $R$ extend to $\mathcal{M}$ by Theorem 2.4. Using the spectral theorem, we can even extend them to operators affiliated with $\mathcal{M}$ in the following way.

For self-adjoint $a \in \mathcal{M}$ let

$$a = \int_{\mathbb{R}} \lambda \, d(e(\lambda))$$

be the spectral decomposition. Since $L, R$ are normal $*$-homomorphisms, the maps $L \circ e$ and $R \circ e$ are spectral measures on $\mathcal{H}$ and

$$L(a) = \int_{\mathbb{R}} \lambda \, d(L \circ e)(\lambda),$$

and analogously for $R(a)$. This formula obviously extends to self-adjoint operators affiliated with $\mathcal{M}$. We continue to denote also these extensions by $L$ and $R$. For arbitrary $a$ affiliated with $\mathcal{M}$ with polar decomposition $a = u |a|$ we define $L(a) = L(u)L(|a|)$ and $R(a) = R(|a|)R(u)$. Again this definition is clearly consistent for $a \in \mathcal{M}$, which justifies the use of the same symbol both for the maps on $\mathcal{M}$ and their extensions to operators affiliated with $\mathcal{M}$.

It is easy to see that for self-adjoint $a, b$ affiliated with $\mathcal{M}$ the operators $L(a)$ and $R(b)$ commute strongly, that is, the spectral measures of $L(a)$ and $R(b)$ commute. Hence we can make sense of expressions of the form $\theta(L(\rho), R(\rho))$ via functional calculus (see [Sch12, Section 5.5]) for positive self-adjoint $\rho$ affiliated with $\mathcal{M}$.

Definition 3.1. Let $\rho$ be a positive self-adjoint operator affiliated with $\mathcal{M}$ and let $\theta : [0, \infty)^2 \rightarrow [0, \infty)$ be measurable. Let $e$ denote the joint spectral measure of $L(\rho)$ and $R(\rho)$. The multiplication operator $\hat{\rho}$ is defined by

$$D(\hat{\rho}) = \left\{ \xi \in \mathcal{H} \left| \int_{[0, \infty)^2} \theta(s, t)^2 \, d\langle e(s, t)\xi, \xi \rangle_{\mathcal{H}} < \infty \right. \right\},$$

$$\langle \hat{\rho} \xi, \eta \rangle_{\mathcal{H}} = \int_{[0, \infty)^2} \theta(s, t) \, d\langle e(s, t)\xi, \eta \rangle_{\mathcal{H}}.$$
Remark 3.2. If $\mathcal{M}$ is commutative, one could alternatively define $\hat{\rho}$ separately for the strongly local and the jump part of $\mathcal{E}$ (recall that one can always regularize $\mathcal{E}$, even if at the cost of a huge state space). Indeed, in the light of the discussion in \cite[Section 10.1]{CS03} it is not hard to see that

$$\|\hat{\rho}|^{1/2}u\|_{\mathcal{H}}^2 = \int \theta(\rho(x), \rho(x))|u(x)|^2 d\Gamma^{(c)}(v)(x)$$

$$+ \frac{1}{2} \int \theta(\rho(x), \rho(y))|u(x)|^2 |v(x) - v(y)|^2 dJ(x,y).$$

However, such a definition would be against the spirit of the present thesis to give the strongly local and the jump part of $\mathcal{E}$.

Assume that $\rho \xi \in (\mathcal{H}, \mathcal{N})$ it is not hard to see that even if at the cost of a huge state space). Indeed, in the light of the discussion in \cite[Section 10.1]{CS03} it is not hard to see that

$$\left\|\hat{\rho}^{1/2}\xi\right\|_{\mathcal{H}}^2 = \sup_{n \in \mathbb{N}} \langle \hat{\rho}_n \xi, \xi \rangle_{\mathcal{H}}.$$

**Proof.** Let $\xi \in \mathcal{H}$ and let $e$ be a joint spectral measure for $L(\rho)$ and $R(\rho)$. Then

$$\langle \hat{\rho}_n \xi, \xi \rangle_{\mathcal{H}} = \int_{[0,\infty)^2} \theta(s \wedge n, t \wedge n) d\langle e(s,t)\xi, \xi \rangle.$$

By assumption, $\theta(s \wedge n, t \wedge n) \nearrow \theta(s,t)$ for all $s,t \geq 0$. The monotone convergence theorem gives

$$\int_{[0,\infty)^2} \theta(s,t) d\langle e(s,t)\xi, \xi \rangle = \sup_{n \in \mathbb{N}} \langle \hat{\rho}_n \xi, \xi \rangle_{\mathcal{H}}.$$

Thus $\xi \in D(\hat{\rho}^{1/2}) = D(\theta(L(\rho), R(\rho))^{1/2})$ if and only if $\sup_{n \in \mathbb{N}} \langle \hat{\rho}_n \xi, \xi \rangle_{\mathcal{H}} < \infty$, and in this case $\|\hat{\rho}^{1/2}\xi\|_{\mathcal{H}}^2 = \sup_{n \in \mathbb{N}} \langle \hat{\rho}_n \xi, \xi \rangle_{\mathcal{H}}.$

**Definition 3.4.** For a positive self-adjoint operator $\rho$ affiliated with $\mathcal{M}$ and a measurable function $\theta : [0, \infty)^2 \to [0, \infty)$ we define

$$\|\cdot\|_\rho : \mathcal{H} \to [0, \infty], \|\xi\|_\rho = \begin{cases} \|\hat{\rho}^{1/2}\xi\|_{\mathcal{H}} & \text{if } \xi \in D(\hat{\rho}^{1/2}), \\ \infty & \text{otherwise.} \end{cases}$$

In other words, $\|\cdot\|_\rho^2$ is the quadratic form generated by $\rho$. Note that this definition implicitly depends on the choice of $\theta$. Lemma 3.3 shows that if $\theta$ is increasing in both arguments, this norm can alternatively be computed as $\|\xi\|_\rho^2 = \sup_{n \in \mathbb{N}} \langle \hat{\rho}_n \xi, \xi \rangle_{\mathcal{H}}$ for $\xi \in \mathcal{H}$.

**Lemma 3.5.** Assume that $\theta : [0, \infty)^2 \to [0, \infty)$ is continuous, increasing in both arguments and $\theta(s,t) > 0$ for $s,t > 0$. If $\rho$ is an invertible positive self-adjoint operator affiliated with $\mathcal{M}$, then the map

$$L^2(\mathcal{M}, \tau) \to [0, \infty], \quad a \mapsto \|\hat{\rho}^{1/2}a\|_{\mathcal{H}}$$

is lower semicontinuous.

**Proof.** First assume that $\rho$ is bounded. Since $\rho$ is invertible and $\theta(s,t) > 0$ for $s,t > 0$, the operator $\hat{\rho}$ is also invertible. Thus $\hat{\rho}^{1/2} \partial a$ is closed and the lower semicontinuity follows from a standard Hilbert space argument. If $\rho$ is not necessarily bounded, the lower semicontinuity follows from Lemma 3.3 and the first part. \hfill $\square$
Definition 3.6. Let \( \theta : [0, \infty)^2 \rightarrow [0, \infty) \) be measurable. For \( a \in D(\mathcal{E}) \) let
\[
\|a\|_\theta^2 = \sup_{\rho \in L_1^0(\mathcal{M}, \tau)} \frac{\|\partial a\|_\rho^2}{\|\rho\|_1}.
\]
The test space \( \mathcal{A}_\theta \) is the set of all \( a \in D(\mathcal{E}) \cap \mathcal{M} \) with \( \|a\|_\theta < \infty \).

Example 3.7. If \( \mathcal{E} \) is a strongly local commutative Dirichlet form on \( L^2(X, m) \), then
\[
\|\partial f\|_\rho^2 = \int_X \theta(\rho(x), \rho(x)) \Gamma(f) \, dm.
\]
In particular, if \( \theta(s, s) = s \), then \( \|f\|_\rho^2 = \|\Gamma(f)\|_\infty \) and
\[
\mathcal{A}_\theta = \{ f \in D(\mathcal{E}) \cap \mathcal{M} | \, \Gamma(f) \in L^\infty(X, m) \}.
\]
This is the space of test functions used in [AEST16].

Example 3.8. If \( \theta = \text{AM} \), the arithmetic mean, then
\[
\|\partial a\|_\rho^2 = \frac{1}{2} \tau((\Gamma(a) + \Gamma(a^*))\rho)
\]
and
\[
\mathcal{A}_{\text{AM}} = \{ a \in D(\mathcal{E}) \cap \mathcal{M} | \, \Gamma(a), \Gamma(a^*) \in \mathcal{M} \}.
\]

A variant of this algebra (without the assumption \( \Gamma(a^*) \in \mathcal{M} \)) was introduced in [Cip16] Definition 10.7 under the name Lipschitz algebra. By [Cip16] Proposition 10.6 the boundedness of \( \Gamma(a) \) is equivalent to the boundedness of the commutator \([D, a^*]\), where
\[
D = \begin{pmatrix}
0 & \partial^* \\
\partial & 0
\end{pmatrix}
\]
is the Dirac operator acting on \( L^2(\mathcal{M}, \tau) \oplus \mathcal{H} \). Hence the space \( \mathcal{A}_{\text{AM}} \) is closely related to spectral triples in Connes’ noncommutative geometry [Con94] (compare also Remark 4.21).

Remark 3.9. In general, it does not seem feasible to give a more explicit description of \( \mathcal{A}_\theta \). Note however that if \( \theta \) is concave, there exist \( \alpha, \beta > 0 \) such that \( \theta \leq \alpha \text{AM} + \beta \) and thus \( \mathcal{A}_{\text{AM}} \subset \mathcal{A}_\theta \).

Lemma 3.10. If \( \theta : [0, \infty)^2 \rightarrow [0, \infty) \) is a symmetric measurable function, then
\[
\|\partial a^*\|_\rho^2 = \|\partial a\|_\rho^2
\]
for all \( a \in D(\mathcal{E}) \cap \mathcal{M} \) and positive self-adjoint operators \( \rho \) affiliated with \( \mathcal{M} \). In particular, \( \mathcal{A}_\theta \) is self-adjoint.

Proof. It follows from the properties of the first-order differential calculus that
\[
J1_A(L(\rho))1_B(R(\rho)) = 1_B(L(\rho))1_A(R(\rho))J
\]
for all Borel sets \( A, B \subset [0, \infty) \). Thus, if \( e \) denotes the joint spectral measure of \( L(\rho) \) and \( R(\rho) \), then
\[
\|\partial a^*\|_\rho^2 = \int_{[0, \infty)^2} \theta(s, t) \, d\langle e(s, t)J\partial a, J\partial a\rangle_{\mathcal{H}}
\]
\[
= \int_{[0, \infty)^2} \theta(s, t) \, d\langle e(t, s)\partial a, \partial a\rangle
\]
\[
= \|\partial a\|_\rho^2,
\]
since \( \theta \) is symmetric. \( \square \)

Lemma 3.11. If \( \tau \) is a state and \( \theta : [0, \infty)^2 \rightarrow [0, \infty) \) is continuous, increasing in both arguments and \( \theta(s, t) > 0 \) for \( s, t > 0 \), then \( \|\cdot\|_\theta \) is lower semicontinuous on \( L^2(\mathcal{M}, \tau) \).
Proof. By Lemma 3.3 it suffices to show that the supremum in the definition of $\|\cdot\|_\theta$ can be taken over all invertible $\rho \in L^1_+(\mathcal{M}, \tau)$.

For $\rho \in L^1_+(\mathcal{M}, \tau)$ and $\varepsilon > 0$ let $\rho^\varepsilon = \rho + \varepsilon$. Evidently, $\rho^\varepsilon \in L^1_+(\mathcal{M}, \tau)$ is invertible and $\|\rho^\varepsilon\|_1 = \|\rho\|_1 + \varepsilon$. Since $\theta$ is increasing in both arguments, one sees as in the proof of Lemma 3.3 that

$$\|\partial a\|_{\rho^\varepsilon}^2 = \|\partial a\|_{\rho + \varepsilon}^2 \geq \|\partial a\|_{\rho}^2.$$

Thus

$$\sup_{\varepsilon > 0} \frac{\|\partial a\|_{\rho^\varepsilon}^2}{\|\rho\|_1^2} \geq \sup_{\varepsilon > 0} \frac{\|\partial a\|_{\rho + \varepsilon}^2}{\|\rho\|_1^2} = \frac{\|\partial a\|_{\rho}^2}{\|\rho\|_1^2}. \quad \square$$

Corollary 3.12. If $\tau$ is a state and $\theta$ is continuous, increasing in both arguments and $\theta(s,t) > 0$ for $s,t > 0$, then $A_0$ is complete in the norm $\|\cdot\|_M + \|\cdot\|_\theta$.

The following technical lemma is certainly known to experts, but because we could not find a reference we include its proof for the convenience of the reader.

Lemma 3.13. If $(x_n)$ is a sequence in $L^p_h(\mathcal{M}, \tau)$ and $x \in L^p_h(\mathcal{M}, \tau)$ such that $x_n \to x$ in $L^p$, then $(x_n)$ converges in the strong resolvent sense to $x$, that is, $(x_n + z)^{-1} \to (x + z)^{-1}$ strongly for every $z \in \mathbb{C} \setminus \mathbb{R}$.

Proof. Since $((x_n + z)^{-1})_n$ is bounded in $B(L^2(\mathcal{M}, \tau))$, it is enough to prove that

$$\tau(a((x_n + z)^{-1} - (x + z)^{-1})b) \to 0$$

for $a, b \in L^2(\mathcal{M}, \tau) \cap L^{2q}(\mathcal{M}, \tau)$, where $q$ is the dual exponent of $p$.

Using the resolvent formula, we see

$$|\tau(a((x_n + z)^{-1} - (x + z)^{-1})b)| = |\tau(a(x + z)^{-1}(x - x_n)(x_n + z)^{-1}b)|$$

$$\leq \|x_n - x\|_p \|a(x + z)^{-1}\|_p \|\|a \|^{-1}b\|_q \|x_n + i\|^{-1}b\|_q$$

$$\leq \frac{1}{(\text{Im } z)^2} \|x - x_n\|_p \|a\|_2 \|b\|_q \|b\|_q \to 0. \quad \square$$

Next we study continuity properties of the map $\rho \mapsto \|\hat{\rho}^{1/2} \partial a\|_\mathcal{H}^2$. We start with an auxiliary result for bounded $\theta$.

Lemma 3.14. Denote by $C_h(\mathcal{H})$ the set of all self-adjoint operators on $\mathcal{H}$. If $\theta: [0, \infty)^2 \to [0, \infty)$ is continuous, then the map

$$L^1_+(\mathcal{M}, \tau) \to C_h(\mathcal{H}), \rho \mapsto \theta(L(\rho), R(\rho))$$

is continuous with respect to the norm topology on $L^1$ and the strong resolvent topology on $C_h(\mathcal{H})$.

If $\theta$ is additionally bounded, then the map

$$L^1_+(\mathcal{M}, \tau) \to [0, \infty), \rho \mapsto \|\xi\|_\rho^2$$

is continuous for all $\xi \in \mathcal{H}$.

Proof. Let $(\rho_n)$ be a sequence in $L^1_+(\mathcal{M}, \tau)$ and $\rho \in L^1_+(\mathcal{M}, \tau)$ such that $\rho_n \to \rho$ in the strong $L^1$ topology. By Lemma 3.13 the sequence $(\rho_n)$ also converges to $\rho$ in the strong resolvent sense. Since $L$ and $R$ are normal $*$-homomorphisms, $L(\rho_n) \to L(\rho)$ and $R(\rho_n) \to R(\rho)$ in the strong resolvent sense as well.

Now $\theta(L(\rho_n), R(\rho_n)) \to \theta(L(\rho), R(\rho))$ follows from [RSt78, Theorem VIII.20]. As the strong resolvent topology coincides with the strong topology on norm bounded subsets of $B(\mathcal{H})$, the last part is clear. \qed
Theorem 3.15. If \( \theta : [0, \infty)^2 \rightarrow [0, \infty) \) is continuous, then
\[
\Lambda : L^+_1(\mathcal{M}, \tau) \rightarrow [0, \infty), \; \rho \mapsto \|\xi\|^2_\rho
\]
is lower semicontinuous with respect to \( \|\cdot\|_1 \) for all \( \xi \in \mathcal{H} \). If \( \theta \) is additionally concave, then \( \Lambda \) is continuous for \( \xi = \partial a \) with \( a \in \mathcal{A}_\mathcal{M} \).

In particular, if \( \Lambda \) is concave, then it is weakly upper semicontinuous for \( \xi = \partial a \) with \( a \in \mathcal{A}_\mathcal{M} \).

Proof. For \( k \in \mathbb{N} \) let \( \theta_k = \theta \wedge k \). By Lemma 3.14 the map
\[
\Lambda_k : L^+_1(\mathcal{M}, \tau) \rightarrow [0, \infty), \; \rho \mapsto \langle \theta_k(L(\rho), R(\rho))\xi, \xi \rangle_\mathcal{H}
\]
is continuous with respect to \( \|\cdot\|_1 \) for all \( \xi \in \mathcal{H} \). Since \( \Lambda_k \not\rightarrow \Lambda \) by functional calculus, the map \( \Lambda \) is lower semicontinuous as supremum of continuous maps.

To prove the continuity when \( \xi = \partial a \) with \( a \in \mathcal{A}_\mathcal{M} \), it only remains to show upper semicontinuity. Since \( \theta \) is concave, there exist \( \alpha, \beta \geq 0 \) such that \( \theta \leq \alpha \Lambda + \beta \).

Since \( a \in \mathcal{A}_\mathcal{M} \), we have
\[
\|(\alpha \Lambda(L(\rho), R(\rho)) + \beta)^{1/2} \partial a\|^2_\mathcal{H} = \frac{\alpha}{2} \tau((\Gamma(a) + \Gamma(a^*))\rho) + \beta \mathcal{E}(a),
\]
which clearly depends continuously on \( \rho \).

Moreover, the map
\[
L^+_1(\mathcal{M}, \tau) \rightarrow [0, \infty), \; \rho \mapsto \|(\alpha \Lambda + \beta - \theta)^{1/2}(L(\rho), R(\rho))\partial a\|^2_\mathcal{H}
\]
is lower semicontinuous by the first part. Thus \( -\Lambda \) is lower semicontinuous as the sum of two lower semicontinuous maps in this case.

Finally, the weak upper semicontinuity for concave \( \Lambda \) follows from the Hahn-Banach theorem. \( \square \)

Remark 3.16. In general, concavity of \( \theta \) is not sufficient for concavity of \( \Lambda \). However, in the following we will study a class of functions \( \theta \) for which \( \Lambda \) is concave.

Remark 3.17. Note that for the upper semicontinuity part we need \( a \in \mathcal{A}_\mathcal{M} \) instead of \( \mathcal{A}_\phi \). Whether upper semicontinuity still holds for \( a \) in the bigger space \( \mathcal{A}_\phi \) is unclear.

In the first part of this section we saw that a crucial property of the multiplication operator is the concavity of the map \( \rho \mapsto \hat{\rho} \). An important class of functions \( \theta \) for which this property holds are those that can be represented as an operator mean in the sense of Kubo–Ando [KA80]. We will now review the definition and a representation theorem for operator means before we turn to the application to the multiplication operator.

Definition 3.18. Let \( H \) be an infinite-dimensional Hilbert space. An operator mean is a map \( \# : B(H)_+ \times B(H)_+ \rightarrow B(H)_+ \) such that
\[
\begin{align*}
x_1 &\leq x_2 \text{ and } y_1 \leq y_2 \text{ imply } x_1 \# y_1 \leq x_2 \# y_2, \\
z(x \# y)z &\leq (zxx) \# (yz) \text{ for } x,y,z \in B(H)_+, \\
x_n \searrow x \text{ and } y_n \searrow y \text{ imply } x_n \# y_n \searrow x \# y, \\
1 \# 1 &= 1.
\end{align*}
\]
If \( H \) is finite-dimensional, a map \( \# : B(H)_+ \times B(H)_+ \rightarrow B(H)_+ \) is called operator mean if \( H \) embeds into an infinite-dimensional Hilbert space \( K \) such that \( \# \) extends to an operator mean on \( K \).

An operator mean \( \# \) is called symmetric if \( x \# y = y \# x \) for all \( x,y \in B(H)_+ \).

We say that a continuous function \( \theta : [0, \infty)^2 \rightarrow [0, \infty) \) can be represented by a (symmetric) operator mean if there exists a (symmetric) operator mean such that \( \theta(x,y) = x \# y \) for all commuting \( x,y \in B(H)_+ \).
Example 3.19. Examples of symmetric operator means include

- the arithmetic operator mean \( (x, y) \mapsto \frac{1}{2}(x + y) \),
- the logarithmic operator mean, given by the generating function \( f(t) = (t - 1)/\log t \) (see Proposition 3.20),
- the harmonic operator mean \( (x, y) \mapsto 2(x^{-1} + y^{-1})^{-1} \),
- the geometric operator mean \( (x, y) \mapsto x^{1/2}(x^{-1/2}y-x^{-1/2})^{1/2}x^{1/2} \).

Two examples of non-symmetric operator means are

- the left trivial mean \( (x, y) \mapsto x \), and
- the right trivial mean \( (x, y) \mapsto y \).

There is a close relation between operator means and operator monotone functions. A continuous function \( f : I \to \mathbb{R} \) is called operator monotone if and only if there exists a Borel probability measure \( \mu \) on \( [0, 1] \) such that

\[ x \# y = x^{1/2}f(x^{-1/2}y-x^{-1/2})x^{1/2} \]

for all invertible \( x, y \in B_+(H) \), and every operator mean arises this way.

In the situation of the proposition above, the operator monotone function \( f \) is called the generating function of \( \# \). An important result of Löwner’s seminal work on operator monotone functions (see [Lö34]) is that every operator monotone function admits an integral representation. A variant of this theorem reads as follows.

Proposition 3.20 ([KA80, Theorem 3.2]). Let \( H \) be a Hilbert space. For every operator monotone function \( f : (0, \infty) \to (0, \infty) \) with \( f(1) = 1 \) there exists a unique operator mean \( \# \) such that

\[ x \# y = x^{1/2}f(x^{-1/2}y-x^{-1/2})x^{1/2} \]

for all invertible \( x, y \in B_+(H) \), and every operator mean arises this way.

In the situation of the proposition above, the operator monotone function \( f \) is called the generating function of \( \# \). An important result of Löwner’s seminal work on operator monotone functions (see [Lö34]) is that every operator monotone function admits an integral representation. A variant of this theorem reads as follows.

Proposition 3.21 ([Han80, Theorem 4.9]). A function \( f : (0, \infty) \to (0, \infty) \) is operator monotone if and only if there exists a finite Borel measure \( \mu \) on \( [0, 1] \) such that

\[ f(t) = \int_0^1 \frac{t}{\lambda + (1-\lambda)t} \, d\mu(t) \]

for \( t > 0 \).

Corollary 3.22. A function \( \theta : [0, \infty)^2 \to [0, \infty) \) can be represented by an operator mean if and only if there exists a Borel probability measure \( \mu \) on \( [0, 1] \) such that

\[ \theta(s, t) = \int_0^1 \frac{st}{\lambda s + (1-\lambda)t} \, d\mu(\lambda) \]

for \( s, t > 0 \).

In this case \( \theta \) is increasing in both arguments, positively homogeneous and satisfies \( \theta(s, s) = s \) for \( s \geq 0 \). The resulting mean is symmetric if and only if \( \mu(A) = \mu(1-A) \) for all Borel sets \( A \subset [0, 1] \).

Conversely, if \( \# \) is an operator mean with generating function \( f \) and one defines \( \theta \) by \( \theta(s, t) = sf(s/t) \) for \( s, t > 0 \), then \( x \# y = \theta(x, y) \) for all commuting \( x, y \in B(H)_+ \).

After we introduced the multiplication operator \( \hat{\rho} = \theta(L(\rho), R(\rho)) \) and operator means, we will now discuss some additional properties of the multiplication operator in the case when \( \theta \) can be represented by an operator mean.

Proposition 3.23. If \( \theta \) can be represented by a symmetric operator mean, then \( A_\theta \) is a *-algebra.

Proof. We have already proven that \( A_\theta \) is self-adjoint in Lemma 3.10. It remains to show that \( A_\theta \) is an algebra.
Let \(a, b \in A_\theta\). By Lemma 3.3 it suffices to show that there exists a constant \(C > 0\) such that \(\|\partial(ab)\|^2 \leq C\|\rho\|_1\) for all \(\rho \in L^1_+(\mathcal{M}, \tau) \cap \mathcal{M}\). We can assume without loss of generality \(\|a\|_\mathcal{M}, \|b\|_\mathcal{M} \leq 1\).

By the product rule we have
\[
\|\partial(ab)\|^2 = \|a^{1/2}(L(a)\partial b + R(b)\partial a)\|^2_H
\leq 2\langle L(a^*)\hat{\rho}L(a)\partial b, \partial b \rangle_H^2 + 2\langle R(b^*)\hat{\rho}R(b)\partial a, \partial a \rangle_H.
\]
(3.1)

Let \(f\) be the generating function of the operator mean \# that is represented by \(\theta\). If \(x \in \mathcal{M}_+\) is invertible, then
\[
L(a^*)(L(x)\# R(x))L(a) = R(x)^{1/2}L(a^*)f(R(x)^{-1/2}L(x)R(x)^{-1/2})L(a)R(x)^{1/2}
\leq R(x)^{1/2}f(R(x)^{-1/2}L(a^* xa)R(x)^{-1/2})R(x)^{1/2}
= L(a^* xa)\# R(x),
\]
where the inequality in the second line follows from the operator monotonicity of \(f\) (see [Han80]). If \(x\) is not necessarily invertible, the same inequality still holds by the continuity property of \#. Thus
\[
L(a^*)\hat{\rho}L(a) \leq L(a^*\rho a)\# R(\rho)
\]
(3.2)
\[
\leq L(a^*\rho a)\# R(\rho) + L(\rho)\# R(a^* \rho a)
\leq L(a^*\rho a + \rho)\# R(\rho + a^* \rho a),
\]
where we used the concavity of operator means ([KA80 Theorem 3.5]) for the last inequality.

Since \# is assumed to be symmetric, the inequality
\[
R(b^*)\hat{\rho}R(b) \leq L(b^* \rho b + \rho)\# R(\rho + b^* \rho b)
\]
(3.3)
follows analogously.

If we combine (3.1), (3.2) and (3.3), we obtain
\[
\|\partial(ab)\|^2 \leq 2\|\partial a\|_{\rho + b^* \rho b}^2 + 2\|\partial b\|_{\rho + a^* \rho a}^2
\leq 2\|a\|^2_2\|\rho + b^* \rho b\|_1 + 2\|b\|^2_2\|\rho + a^* \rho a\|_1
\leq 4(\|a\|^2_2 + \|b\|^2_2)\|\rho\|_1.
\]
Hence \(ab \in A_\theta\). \qed

**Lemma 3.24.** If \(\theta\) can be represented by a symmetric operator mean, then
\[
\|\partial a\|^2_\rho \leq \frac{1}{2}\tau((\Gamma(a) + (\Gamma(a^*)))\rho)
\]
for \(a \in D(\mathcal{E})\) and \(\rho \in L^1_+(\mathcal{M}, \tau)\).

**Proof.** By [KA80 Theorem 4.5] we have \(\theta \leq \text{AM}\). Now it suffices to notice that
\[
\langle L(\rho)\partial a, \partial a \rangle_H = \tau(\Gamma(a)\rho)
\]
and
\[
\langle R(\rho)\partial a, \partial a \rangle_H = \tau(\Gamma(a^*)\rho).
\]
\qed

**Lemma 3.25.** Assume that \(\theta\) can be represented by an operator mean. If \(\rho_0, \rho_1 \in L^1_+(\mathcal{M}, \tau)\) with \(\rho_0 \leq \rho_1\), then \(\|\xi\|_{\rho_0} \leq \|\xi\|_{\rho_1}\) for all \(\xi \in \mathcal{H}\).

**Proof.** If \(\rho_0, \rho_1\) are bounded, then the claim is immediate from the definition of operator means. In the general case let \(f_\varepsilon(\rho) = (1 + \varepsilon \rho)^{-1}\). This function is operator monotone, hence \(f_\varepsilon(\rho_0) \leq f_\varepsilon(\rho_1)\). Moreover, \(f_\varepsilon(\rho) \nearrow \rho\) as \(\varepsilon \to 0\) implies
\[ \|\xi\|_{f_\epsilon(\rho)} \to \|\xi\|_\rho \text{ as } \epsilon \to 0 \text{ for } i \in \{0,1\}. \] Combining this convergence with the monotonicity in the bounded case, we obtain
\[ \|\xi\|_{\rho_0} = \lim_{\epsilon \to 0} \|\xi\|_{f_\epsilon(\rho_0)} \leq \lim_{\epsilon \to 0} \|\xi\|_{f_\epsilon(\rho_1)} = \|\xi\|_{\rho_1}. \]

**Corollary 3.26.** Assume that \( \theta \) can be represented by an operator mean. If \( \rho_n \to \rho \) in \( L_+^1(M, \tau) \) and \( \rho_n \leq \rho \), then \( \|\xi\|_{\rho_n} \to \|\xi\|_\rho \) for \( \xi \in \mathcal{H} \). If moreover \( \xi \in D(\hat{\rho}^{1/2}) \), then \( \xi \in D(\hat{\rho}_n^{1/2}) \) for all \( n \in \mathbb{N} \) and \( \hat{\rho}_n^{1/2} \xi \to \hat{\rho}^{1/2} \xi \) in \( \mathcal{H} \).

**Proof.** The first part is an immediate consequence of Theorem 3.15 and Lemma 3.25. For the second part first note that \( \hat{\rho}_n \wedge N^{1/2} \to \hat{\rho} \wedge N^{1/2} \) strongly as \( n \to \infty \) by Lemma 3.14.

Let \( e_n \) denote the joint spectral measure of \( L(\rho_n) \) and \( R(\rho_n) \). Then
\[ \|(\hat{\rho}_n^{1/2} - \hat{\rho}^{1/2})\xi\|_\mathcal{H}^2 = \int_{[0,\infty)^2} (\theta(s, t) - 1^{1/2} - \theta(s \wedge N, t \wedge N)) \cdot d\langle e_n(s, t)\xi, \xi \rangle_{\mathcal{H}} \]
\[ \leq \int_{[0,\infty)^2} (\theta(s, t) - \theta(s \wedge N, t \wedge N)) \cdot d\langle e_n(s, t)\xi, \xi \rangle_{\mathcal{H}} \]
\[ = \|\xi\|_{\rho_n}^2 - \|\xi\|_{\rho_n \wedge N}^2. \]
The same holds for \( \rho_n \) replaced by \( \rho \). Thus
\[ \|(\hat{\rho}_n^{1/2} - \hat{\rho}^{1/2})\xi\|_\mathcal{H} \leq \|(\hat{\rho}_n^{1/2} - \hat{\rho}_n \wedge N^{1/2})\xi\|_\mathcal{H} + \|(\hat{\rho}_n \wedge N^{1/2} - \hat{\rho} \wedge N^{1/2})\xi\|_\mathcal{H} \]
\[ + \|(\hat{\rho} \wedge N^{1/2} - \hat{\rho}^{1/2})\xi\|_\mathcal{H} \]
\[ \leq (\|\xi\|_{\rho_n}^2 - \|\xi\|_{\rho_n \wedge N}^2)^{1/2} + \|(\hat{\rho}_n \wedge N^{1/2} - \hat{\rho} \wedge N^{1/2})\xi\|_\mathcal{H} \]
\[ + (\|\xi\|_{\rho}^2 - \|\xi\|_{\rho \wedge N}^2)^{1/2}. \]
Hence
\[ \limsup_{n \to \infty} \|(\hat{\rho}_n^{1/2} - \hat{\rho}^{1/2})\xi\|_\mathcal{H} \leq 2(\|\xi\|_{\rho}^2 - \|\xi\|_{\rho \wedge N}^2)^{1/2}, \]
which goes to zero as \( N \to \infty \). \( \square \)

**Lemma 3.27.** If \( \theta \) can be represented by an operator mean, then
\[ L_+^1(M, \tau) \to [0, \infty), \rho \to \|\xi\|_\rho^2 \]
is concave for all \( \xi \in \mathcal{H} \).

**Proof.** Since operator means are jointly concave by [KA80, Theorem 3.5], the map \( \rho \to \|\xi\|_\rho^2 \) is concave on \( L_+^1(M, \tau) \cap M \). Hence, if \( \rho_0, \rho_1 \in L_+^1(M, \tau) \) and \( \lambda > 0 \), then
\[ \|\xi\|^2_{(1-\lambda)\rho_0 \wedge \lambda \rho_1 \wedge \lambda \rho_1} \geq (1 - \lambda)\|\xi\|_{\rho_0 \wedge \lambda \rho_1}^2 + \lambda \|\xi\|_{\rho_1 \wedge \lambda \rho_1}^2. \]
By Lemma 3.3 the right-hand side converges to \((1 - \lambda)\|\xi\|^2_{\rho_0} + \lambda \|\xi\|^2_{\rho_1}\) as \( n \to \infty \). On the other hand, Lemma 3.25 gives
\[ \|\xi\|^2_{(1-\lambda)\rho_0 + \lambda \rho_1} \geq \|\xi\|^2_{(1-\lambda)\rho_0 \wedge \lambda \rho_1}. \]
Thus \( \rho \to \|\xi\|_{\rho}^2 \) is concave. \( \square \)

As mentioned before, we will later focus on the case when \( \theta \) is the logarithmic mean
\[ \text{LM}: [0, \infty)^2 \to [0, \infty), (s, t) \mapsto \begin{cases} \frac{s - t}{\log s - \log t} & \text{if } s \neq t, \\ s & \text{otherwise}. \end{cases} \]
Alternatively, it can be represented as

\[ \text{LM}(s, t) = \int_0^1 s^\alpha t^{1-\alpha} \, d\alpha. \]

A direct calculation shows that LM can be represented by a symmetric operator mean, namely the logarithmic operator mean from Example 3.19. Thus, all the results from this section are applicable in this case.

It is the following identity that sets the logarithmic mean apart from other possible choices of operator means in our context:

\[ \text{LM}(L(a), R(a)) \partial \log(a) = \text{LM}(L(a), R(a)) \tilde{\log}(L(a), R(a)) \partial a = \partial a. \]

This cancellation effect relies only on the chain rule for the first-order differential calculus. It would therefore be natural to consider more general functions \( \theta \) of the form

\[ \theta(s, t) = \frac{s - t}{\psi(s) - \psi(t)}. \]

However, if we additionally require that \( \theta \) can be represented by an operator mean, then it is not hard to see that \( \psi \) is already forced to be the logarithm (up to an additive constant). Thus the choice of the logarithmic mean (and the von Neumann entropy later) is not arbitrary, but a consequence of these two simple structural assumptions.

4. THE NONCOMMUTATIVE TRANSPORT METRIC \( W \)

In this section we define a transport metric on the space of density operators that generalizes both the discrete transport metric \( W \) from \[Maa11, Mie11, CHLZ12\] and the Wasserstein metric \( W_2 \) on Riemannian manifolds.

The study of the optimal transport problem

\[ \int_{X \times X} d(x, y)^2 \, d\pi(x, y) \to \min \]

\[ (\text{pr}_1)_\# \pi = \mu, (\text{pr}_2)_\# \pi = \nu \]

defining the \( L^2 \)-Wasserstein metric goes back to the work of Monge [Mon81] and Kantorovich [Kan42, Kan04], who formulated the relaxed problem in the modern form. Especially for the quadratic case metric, the name “Wasserstein metric” is misleading, and some authors prefer to call it Monge-Kantorovich metric or some variations of that. More information on the history of the Wasserstein metric as well as optimal transport in general can be found in the bibliographical notes in Villani’s book [Vil09].

The Benamou–Brenier formula

\[ W_2(\mu, \nu)^2 = \inf \left\{ \int_0^1 \int_{\mathbb{R}^n} |v_t|^2 \, d\mu_t \, dt \left| \begin{array}{l} \dot{\mu}_t + \nabla \cdot (\mu_t v_t) = 0, \mu_0 = \mu, \mu_1 = \nu \end{array} \right. \right\} \]

gives an equivalent definition of the Wasserstein metric on Borel probability measures over \( \mathbb{R}^n \) as dynamical optimization problem. It was found by Benamou and Brenier [BB00] in relation to numerical algorithms for the Wasserstein distance and later generalized to considerably more general settings (see for example [AES16]).

Our definition of the transport metric \( W \) relies on a modification of the Benamou–Brenier formula. As already observed in the articles mentioned above in the case of finite graphs and matrix algebras, the crucial step is to not only replace the action functional in the classical Benamou–Brenier formula, but also the constraint by a suitable noncommutative version of the continuity equation.
While the form of this continuity equation is easily adapted from the previous work on the finite-dimensional case, finding a good weak formulation is still challenging. As it turns out, especially in view of the results in Section 6, the algebra $\mathcal{A}_{AM}$ introduced in the last section is a good choice of test “functions”. Among several other useful properties of the metric $\mathcal{W}$, we will use the continuity properties from the last section to prove lower semicontinuity of the energy functional defining $\mathcal{W}$ (Theorem 1.3.30).

As usual, $(\mathcal{M}, \tau)$ is a tracial von Neumann algebra, $\mathcal{E}$ a quantum Dirichlet form on $L^2(\mathcal{M}, \tau)$ such that $\tau$ is energy dominant, and $(\partial, \mathcal{H}, L, R, J)$ the associated first-order differential calculus. We further assume that $\theta: [0, \infty)^2 \to [0, \infty)$ is a continuous function that can be represented by a symmetric operator mean. In particular, all results from Section 5 are applicable. All expressions like $\hat{\rho}$, $\|\cdot\|_{\rho}$ etc. are to be understood with respect to this particular choice of $\theta$.

**Definition 4.1** (Density operator). A *density operator* is an element $\rho$ of $L^1_+(\mathcal{M}, \tau)$ with $\tau(\rho) = 1$. The space of all density operators over $(\mathcal{M}, \tau)$ is denoted by $\mathcal{D}(\mathcal{M}, \tau)$.

Under the map $\rho \mapsto \tau(\cdot, \rho)$, the density matrices correspond exactly to the normal states on $\mathcal{M}$. Of course, if $\mathcal{M}$ is commutative, the density operators over $(\mathcal{M}, \tau)$ are just the classical probability densities.

**Definition 4.2** (Hilbert space $\mathcal{H}_\rho$). For $\rho \in \mathcal{D}(\mathcal{M}, \tau)$ let $\mathcal{H}_\rho$ be the Hilbert space obtained from $D(\hat{\rho}^{1/2})$ after separation and completion with respect to $\|\cdot\|_{\rho}$. Let $\mathcal{H}_\rho$ be the closure of $\partial\mathcal{A}_{AM}$ in $\mathcal{H}_\rho$.

If $(\rho_t)_{t \in I}$ is a curve in $\mathcal{D}(\mathcal{M}, \tau)$, we say that a curve $(\xi_t)_{t \in I}$ with $\xi_t \in \mathcal{H}_{\rho_t}$ is measurable if $t \mapsto \|\xi_t\|_{\rho_t}$ is measurable and $t \mapsto (\xi_t, \partial a)_{\rho_t}$ is measurable for all $a \in \mathcal{A}_{AM}$. The space of all a.e.-equivalence classes of measurable curves $(\xi_t)$ such that $\int_I \|\xi_t\|_{\rho_t}^2 dt < \infty$ is denoted by $L^2(I; (\mathcal{H}_{\rho_t})_{t \in I})$. The space $L^2_{\text{loc}}(I; (\mathcal{H}_{\rho_t})_{t \in I})$ is defined accordingly.

**Remark 4.3.** If there exists a countable subset $E$ of $\mathcal{A}$ such that $\partial E$ is dense in $\mathcal{H}_{\rho_t}$ for all $t \in I$, then $(\mathcal{H}_{\rho_t})_{t \in I}$ is a measurable field of Hilbert spaces in the sense of [Tak02, Definition 8.9] and $L^2(I; (\mathcal{H}_{\rho_t})_{t \in I})$ is just a different notation for the direct integral $\int_I \mathcal{H}_{\rho_t} dt$.

**Definition 4.4** (Admissible curves). A curve $(\rho_t)_{t \in I}$ in $\mathcal{D}(\mathcal{M}, \tau)$ is *admissible* if $t \mapsto \tau(\rho_t)$ is locally absolutely continuous for all $a \in \mathcal{A}_{AM}$ and there exists $\xi \in L^2_{\text{loc}}(I; (\mathcal{H}_{\rho_t})_{t \in I})$ such that for all $a \in \mathcal{A}_{AM}$ the continuity equation

\[ \frac{d}{dt} \tau(a \rho_t) = \langle \partial a, \xi_t \rangle_{\rho_t} \]

holds for a.e. $t \in I$.

It if exists, such an element $\xi \in L^2_{\text{loc}}(I; (\mathcal{H}_{\rho_t})_{t \in I})$ is necessarily unique since $\partial \mathcal{A}_{AM}$ is dense in $\mathcal{H}_{\rho_t}$ for all $t \in I$, and we write $D \rho = \xi$ in this case.

A couple of remarks are in order. First, the definition of absolutely continuous functions allows for an integral characterization of admissible curves that will be useful later on.

**Remark 4.5.** Let $(\rho_t)_{t \in I}$ be a curve in $\mathcal{D}(\mathcal{M}, \tau)$. It is easy to see that $(\rho_t)$ is admissible if and only if there exists a $c \in L^2_{\text{loc}}(I)$ such that

\[ |\tau(a \rho_t) - \tau(a \rho_s)| \leq \int_s^t c(r) \|\partial a\|_{\rho_r} dr \]

for all $s, t \in I$ and $a \in \mathcal{A}_{AM}$, and in this case, $r \mapsto \|D \rho_r\|_{\rho_r}$ is the minimal function $c$ with this property.
Remark 4.6. First rudiments of a solution theory of equations of similar type based on the noncommutative differential calculus have been developed in [Zae16].

Remark 4.7. If $E$ is the standard Dirichlet energy on a complete Riemannian manifold $(M, g)$, then [CE] reduces to the classical continuity equation
\[ \dot{\rho}_t + \text{div}(\rho_t \xi_t) = 0 \]
(weakly in duality with the bounded Lipschitz functions).

Accordingly, if $(M, g)$ has lower bounded Ricci curvature, then a curve $(\rho_t)_{t \in I}$ of probability densities is admissible if and only if it is in $\text{AC}^2_{\text{loc}}(I; (P_2(M), W_2))$ by [Erb10 Proposition 2.5]. Compare also Example 4.11 and Proposition 7.5.

Finally, let us also discuss two possible variants of the definition of admissible curves.

Remark 4.8. Instead of restricting to $\xi_t \in H_{\rho_t}$ in [CE], one might want to take $\xi_t \in \mathcal{H}_{\rho_t}$. This is no longer unique, but if it exists, the orthogonal projection $\eta_t$ of $\xi_t$ onto $H_{\rho_t}$ still satisfies [CE] and $\|\eta_t\|_{\rho_t} \leq \|\xi_t\|_{\rho_t}$. Instead of minimizing over all admissible curves $(\rho_t)$ with unique “velocity vector field” $(D\rho_t)$ in the definition $W$ below, one can therefore equivalently minimize over all pairs of curves $(\rho_t, \xi_t)$ satisfying [CE], where we only assume $\xi_t \in \mathcal{H}_{\rho_t}$.

Remark 4.9. Since the definition of the multiplication operator $\hat{\rho}$ uses the mean $\theta$, it might appear more natural to replace $\mathcal{A}_{AM}$ by the bigger space $\mathcal{A}_\theta$ both in the definition of $\mathcal{H}_\rho$ and the weak continuity equation [CE]. The crucial point is that the upper semicontinuity property from Theorem 3.15 is only guaranteed for $\mathcal{A}_{AM}$.

However, under suitable conditions on the Dirichlet form $E$ we introduce in Chapter 6.1 the closure of $\partial \mathcal{A}_\theta$ in $\mathcal{H}_\rho$ coincides with $\mathcal{H}_\rho$ and the duality in the continuity equation can be extended to $a \in \mathcal{A}_\theta$ so that both of these possible definitions finally yield the same result.

Under strong conditions on the curve $(\rho_t)$, the duality in (CE) can be extended beyond to $D(E)$.

Lemma 4.10. Assume that $\mathcal{A}_{AM} \subset D(E)$ is dense. If $(\rho_t)_{t \in I}$ is an admissible curve in $D(M, \tau)$ such that
\[ \sup_{J \subset I} \|\rho_t\|_M < \infty \]
for all compact $J \subset I$, then $t \mapsto \tau(\rho_t a)$ is locally absolutely continuous for all $a \in D(E)$ and
\[ \frac{d}{dt} \tau(\rho_t a) = \langle \partial a, D\rho_t \rangle_{\rho_t} \]
for a.e. $t \in I$.

Proof. Let $(a_k)$ be a sequence in $\mathcal{A}_{AM}$ such that $a_k \to a$ w.r.t. $\|\cdot\|_E$. Since $\rho_t \in D(M, \tau) \cap M \subset L^2(M, \tau)$, we have $\tau(a_k \rho_t) \to \tau(a \rho_t)$ as $k \to \infty$. On the other hand, since $\dot{\rho}_t$ is bounded and $\partial a_k \to \partial a$, we also have $\langle \partial a_k, D\rho_t \rangle_{\rho_t} \to \langle \partial a, D\rho_t \rangle_{\rho_t}$ as $k \to \infty$. Moreover,
\[ |\langle \partial a_k, D\rho_t \rangle_{\rho_t}| \leq \|\rho_t\|_M^{1/2} \mathcal{E}(a_k)^{1/2} \|D\rho_t\|_{\rho_t}. \]
Since $(\mathcal{E}(a_k))_k$ is bounded and $t \mapsto \|\rho_t\|_M$ is bounded on compact intervals, we can apply the dominated convergence theorem to get
\[ \tau(a(\rho_t - \rho_s)) = \lim_{k \to \infty} \tau(a_k(\rho_t - \rho_s)) = \lim_{k \to \infty} \int_s^t \langle \partial a_k, D\rho_r \rangle_{\rho_r} dr = \int_s^t \langle \partial a, D\rho_r \rangle_{\rho_r} dr. \]
From this equality, both the claimed absolute continuity and the identity for the derivative follow easily. \qed
We are now in the position to introduce the transport metric $W$ as a length metric with a length functional defined on the class of admissible curves. Strictly speaking, the map $W$ will not be a metric since it might be degenerate and take the value infinity. Let us therefore recall the following extended concept of metrics.

**Definition 4.11 (Extended pseudometric).** Let $X$ be a set. An extended pseudometric on $X$ is a map $d: X \times X \rightarrow [0, \infty]$ such that

- $d(x, x) = 0$ for $x \in X$,
- $d(x, y) = d(y, x)$ for $x, y \in X$,
- $d(x, y) \leq d(x, z) + d(z, y)$ for $x, y, z \in X$.

An extended pseudometric $d$ is an extended metric if $d(x, y) = 0$ implies $x = y$.

**Definition 4.12 (Transport metric $W$).** The extended pseudometric $W$ on $\mathcal{D}(\mathcal{M}, \tau)$ is defined by

$$W(\bar{\rho}_0, \bar{\rho}_1) = \inf \left\{ \int_0^1 \|D\rho_t\|_{\rho_t} dt \left| (\rho_t) \text{ admissible}, \rho_0 = \bar{\rho}_0, \rho_1 = \bar{\rho}_1 \right. \right\}.$$

**Remark 4.13.** If we endow $\mathcal{D}(\mathcal{M}, \tau)$ with the topology induced by the seminorms $\tau(a \cdot)$ for $a \in A_\mathcal{M}$, then the class of admissible curves together with the map that sends an admissible curve $(\rho_t)_{t \in I}$ to $\int_I \|D\rho_t\|_{\rho_t} dt$ is a length structure in the sense of [BBI01, Chapter 2] and $W$ is the associated length metric. The topological condition from their definition is verified in Proposition 4.20.

**Remark 4.14.** Contrary to the Wasserstein metric, but also the metric $W$ defined for certain jump processes in [Erb14], we define $W$ only on densities. This is enough to study gradient flows of the entropy, which is only finite on measures with density anyway, but it would be interesting to see if there is an extension of $W$ to a larger class of states.

**Remark 4.15.** A different approach to noncommutative analogues of the Wasserstein distances, which relies on approximation by commutative subalgebras, has been studied in [Zae15]. Contrary to our construction, if the algebra $\mathcal{M}$ is commutative, the metric $W_2$ defined by Zaev is the usual $L^2$-Wasserstein distance. In particular, in some examples it coincides and in some examples it is different from the metric constructed here. It is not clear if there is any deeper connection between these two approaches in the noncommutative case.

**Example 4.16.** Let $(X, b, m)$ be a weighted graph and $\mathcal{E}(N)$ as in Example 1.13. Then

$$\|\xi\|_{\rho}^2 = \frac{1}{2} \sum_{x,y} b(x, y) \theta(\rho(x), \rho(y)) |\xi(x, y)|^2$$

for $\rho \in \mathcal{P}(X, m)$ and $\xi \in \ell^2(X \times X, \frac{1}{2}b)$. In particular, if $X$ is finite, this norm coincides with the one defined in [Maa11]. Consequently, our metric $W$ coincides with the metric $W$ defined in [Maa11] for finite graphs.

**Example 4.17.** If $\mathcal{E}$ is the standard Dirichlet energy on $\mathbb{R}^n$, then

$$\|\partial u\|_{\rho}^2 = \int_{\mathbb{R}^n} |\nabla u|^2 \rho \, dx$$

and the definition of $W$ coincides with the Benamou–Brenier formulation [BB00] of the $L^2$-Wasserstein distance.
Example 4.18. More generally, let $\mathcal{E}$ be a strongly local regular Dirichlet form on $L^2(X, m)$ and assume that $m$ is energy dominant. Then

$$\|\partial u\|^2_\rho = \int_X \Gamma(u)\rho \, dm.$$  

In this case $\mathcal{W}$ coincides with the metric $\mathcal{W}_\mathcal{E}$ defined in [AES16] Definition 10.4]. This in turn was shown in [AES16] Theorem 12.5] to coincide with the $L^2$-Wasserstein distance $\mathcal{W}_2$ if $(X, d, m)$ is an RCD($K, \infty$) space and $\mathcal{E}$ is twice the Cheeger energy (see also [AGS14b] for the relevant definitions).

Note that in the last two examples the transport metric $\mathcal{W}$ does not depend on the choice of the mean $\theta$. That is because if $\mathcal{E}$ is strongly local, only the values of $\theta$ on the diagonal matter, and these are already determined by the assumption that $\theta$ can be represented by an operator mean.

If one is only interested in the commutative case, one might want to relax the condition of operator concavity of $\theta$ to mere concavity. In this case, $\theta$ does not necessarily reduce to the identity on the diagonal. Metrics of this type (in the strongly local case) were studied in [DNS09, CLSS10].

Example 4.19. Let $\mathcal{M}$ be a finite-dimensional von Neumann algebra, $\text{tr}$ the normalized trace on $\mathcal{M}$, and $(P_t)$ a quantum Markov semigroup on $\mathcal{M}$. Under the assumption that $(P_t)$ satisfies the quantum detailed balance condition, Carlen and Maas [CM17] defined a Riemannian metric on the space $D_+(\mathcal{M}, \text{tr})$ of strictly positive density matrices. Let us shortly summarize their construction.

Given $\omega = (\omega_1, \ldots, \omega_n) \in \mathbb{R}^n$ and $c = (c_1, \ldots, c_n) \in \mathbb{R}^n$ (which are canonically associated with $(P_t)$) and a density matrix $\rho$, they define

$$[\rho]_\omega : L^2(\mathcal{M}, \text{tr}) \rightarrow L^2(\mathcal{M}, \text{tr}), \quad [\rho]_\omega \omega_j = \int_0^1 (e^{-\omega_j/2}L_\rho)^s(e^{x_j/2}R_\rho)^{1-s} \, ds,$$

where $L_\rho$ and $R_\rho$ are the left and right multiplication with $\rho$ on $\mathcal{M}$. Further, $[\rho]_\omega = [\rho]_{\omega_1} \oplus \cdots \oplus [\rho]_{\omega_n}$.

The norm of a tangent vector $\dot{\rho}_0$ is defined by

$$g(\dot{\rho}_0, \dot{\rho}_0) = \inf_V \sum_{j=1}^n c_j \langle V_j, [\rho_0]_{\omega_j} V_j \rangle_{L^2(\mathcal{M}, \text{tr})},$$

where the infimum is taken over all $V$ satisfying a continuity equation of the form

$$\dot{\rho}_0 = \text{div}([\rho_0]_{\omega} V).$$

In the case when $(P_t)$ is tracially symmetric and $\theta = \text{LM}$, one has $c_j = 1$, $\omega_j = 0$ for all $j \in \{1, \ldots, n\}$, and it is easily checked that $L$, $R$, $\text{div}$ etc. coincide with the operations obtained from the first-order differential calculus. Therefore, $[\dot{\rho}]_0 = \dot{\rho}$ and the distance function induced by $g$ coincides with $\mathcal{W}$.

However, it should be stressed that the class of quantum Markov semigroups satisfying the detailed balance condition is larger than the class of tracially symmetric ones, so we do not fully recover the construction from [CM17]. It is an interesting open question how one can generalize the construction of the metric $\mathcal{W}$ to the case of infinite-dimensional quantum Markov semigroups satisfying the detailed balance condition for a non-tracial state or weight.

Next we collect some basic properties of $\mathcal{W}$. The first one is a sufficient condition to make $\mathcal{W}$ non-degenerate.

Proposition 4.20. If $\rho_0, \rho_1 \in D(\mathcal{M}, \tau)$ and $a \in \mathcal{A}_{\mathcal{AM}}$, then

$$|\tau(a(\rho_0 - \rho_1))|^2 \leq \|a\|_{\mathcal{AM}}^2 \mathcal{W}(\rho_0, \rho_1)^2 \leq \|a\|_{\mathcal{AM}}^2 \mathcal{W}_\sigma(\rho_0, \rho_1).$$

In particular, if $\mathcal{A}_{\mathcal{AM}}$ is $\sigma$-weakly dense in $\mathcal{M}$, then $\mathcal{W}$ is non-degenerate.
Proof. We can assume that \( W(\rho_0, \rho_1) < \infty \), otherwise there is nothing to prove. Let \((\rho_t)_{t \in [0, 1]}\) be an admissible curve connecting \(\rho_0\) and \(\rho_1\). By definition \(\|\frac{d}{dt}a\|^2_{\rho_t} \leq \|a\|^2_{\theta}\) for all \(t \in [0, 1]\). Thus

\[
|\tau(a(\rho_1 - \rho_0))| \leq \int_0^1 |\langle \frac{d}{dt}a, D\rho_t \rangle_{\rho_t}| \, dt \leq \|[a\|_{\theta}\int_0^1 \|D\rho_t\| \, dt.
\]

Taking the infimum over all admissible curves connecting \(\rho_0\) and \(\rho_1\) yields the first inequality.

The second inequality follows directly from the first an Lemma 4.24. Finally, the last claim is an immediate consequence of the first inequality. \(\square\)

Remark 4.21. According to [Cip16 Proposition 10.6], the seminorm

\[
\|\cdot\|_{AM}: A_{AM} \to [0, \infty), \quad a \mapsto \left(\frac{1}{2}(\|\Gamma(a) + \Gamma(a^*)\|_{AM})\right)^{1/2}
\]

is a Lipschitz seminorm in the spirit of [Con89, Rie99]. The induced metric \(W_\Gamma\) on \(D(M, \tau)\) given by

\[
W_\Gamma(\rho, \sigma) = \sup\{|	au(a(\rho - \sigma))|: a \in A_{AM}, \|a\|_{AM} \leq 1\}
\]

is a noncommutative analog of the \(L^1\)-Wasserstein distance (depending on the context, it is also called Connes distance or spectral distance).

The following lemma is standard.

Lemma 4.22. If \((\rho_s)_{s \in [0, 1]}\) is an admissible curve with \(D\rho_s \neq 0\) for a.e. \(s \in [0, 1]\), then \((\rho_s)\) can be reparametrized so that the resulting curve \((\sigma_t)_{t \in I}\) has constant speed and

\[
\int_0^1 \|D\sigma_t\|_{\sigma_t}^2 \, dt \leq \int_0^1 \|D\rho_s\|^2_{\rho_s} \, ds.
\]

Proof. By assumption, the map

\[
[0, 1] \to [0, 1], \quad s \mapsto \frac{\int_0^s \|D\rho_r\|_{\rho_r} \, dr}{\int_0^1 \|D\rho_r\|_{\rho_r} \, dr}
\]

is continuous and strictly increasing, hence a homeomorphism. Denote its inverse by \(\theta\) and let \(\sigma_t = \rho_{\theta(t)}\). It is immediate from the definition that \(\sigma\) is admissible and \(D\sigma_t = \dot{\theta}(t)D\rho_{\theta(t)}\) for a.e. \(t \in [0, 1]\). Note that

\[
\dot{\theta}(t) = \frac{\int_0^1 \|D\rho_r\|_{\rho_r} \, dr}{\|D\rho_{\theta(t)}\|_{\rho_{\theta(t)}}}.
\]

Thus \((\sigma_t)\) has constant speed and

\[
\int_0^1 \|D\sigma_t\|_{\sigma_t}^2 \, dt = \left(\int_0^1 \|D\rho_r\|_{\rho_r} \, dr\right)^2 \leq \int_0^1 \|D\rho_r\|_{\rho_r}^2 \, dr.
\]

Corollary 4.23. The pseudometric \(W\) can alternatively be calculated as

\[
W(\bar{\rho}_0, \bar{\rho}_1)^2 = \inf \left\{ \int_0^1 \|D\rho_t\|_{\rho_t}^2 \, dt \mid (\rho_t) \text{ admissible, } \rho_0 = \bar{\rho}_0, \rho_1 = \bar{\rho}_1 \right\}.
\]

Lemma 4.24 (Convexity of the squared distance). For \(\rho^j_i \in D(M, \tau), i, j \in \{0, 1\}\), let \(\rho^i_t = (1 - t)\rho^0_i + t\rho^1_i\) for \(i \in \{0, 1\}, t \in [0, 1]\). Then

\[
W^2(\rho^0_i, \rho^1_i) \leq (1 - t)W^2(\rho^0_0, \rho^1_0) + tW^2(\rho^0_1, \rho^1_1)
\]

for all \(t \in [0, 1]\).
Proof. We can assume that \( W(\rho^0_0, \rho^1_0), W(\rho^0_1, \rho^1_1) < \infty \). For \( j \in \{0, 1\} \) let \( (\rho^*_{t})_{t \in [0, 1]} \) be admissible curves connecting \( \rho^0_j \) and \( \rho^1_j \) and let \( \xi^*_{t} = D_s \rho^* \). Define \( \rho^* = (1-t)\rho^0 + t\rho^1 \) for \( s, t \in [0, 1] \). Obviously, \( s \mapsto \tau(ap^*_t) \) is locally absolutely continuous for all \( a \in A_{AM} \) and \( t \in [0, 1] \).

We will show that the map \( \partial A_{AM} \rightarrow \mathbb{C}, \partial a \mapsto \frac{d}{ds} \tau(ap^*_t) \) is well-defined and continuous with respect to \( \|\cdot\|_{\rho^*_t} \). Indeed,

\[
\left| \frac{d}{ds} \tau(ap^*_t) \right|^2 \leq \frac{\|D_s \rho^*_t\|_{\rho^*_t}^2}{\|\partial a\|_{\rho^*_t}^2} \leq \frac{\|D_s \rho^*_t\|_{\rho^*_t}^2}{\|\partial a\|_{\rho^*_t}^2} \leq (1-t) \left< \frac{\|D_s \rho^*_t\|_{\rho^*_t}^2}{\|\partial a\|_{\rho^*_t}^2} \right> + t \left\| \frac{\|D_s \rho^*_t\|_{\rho^*_t}^2}{\|\partial a\|_{\rho^*_t}^2} \right\|^2.
\]

For the first inequality we used Lemma 3.27, while the second inequality follows from the convexity of the function \( (x,y) \mapsto \frac{x}{y} \).

Thus, \( (\rho^*_t)_{t \in [0,1]} \) is admissible for every \( t \in [0, 1] \) and \( (D_s \rho^*_t)_{t \in [0,1]} \) satisfies

\[
\|D_s \rho^*_t\|_{\rho^*_t}^2 = \sup_{a \in A_{AM}} \left( \frac{\|D_s \rho^*_t\|_{\rho^*_t}^2}{\|\partial a\|_{\rho^*_t}^2} \right) \leq (1-t) \|D_s \rho^0_t\|_{\rho^0_t}^2 + t \|D_s \rho^1_t\|_{\rho^1_t}^2.
\]

Therefore

\[
W^2(\rho^0_t, \rho^1_t) \leq (1-t) \int_0^1 \|D_s \rho^0_t\|_{\rho^0_t}^2 ds + t \int_0^1 \|D_s \rho^1_t\|_{\rho^1_t}^2 ds.
\]

Taking the infimum over all admissible curves \( (\rho^*_t)_{t \in [0,1]} \) connecting \( \rho^0_j \) and \( \rho^1_j \) yields the assertion. \( \square \)

Definition 4.25. Let \( (X,d) \) be an extended metric space. A curve \( \gamma : I \rightarrow X \) is called \( p \)-locally absolutely continuous if there exists a positive function \( g \in L^p_{\text{loc}}(I) \) such that

\[
\left( AC_p \right) \quad d(\gamma_s, \gamma_t) \leq \int_s^t g(r) \, dr
\]

for all \( s, t \in I \). We write \( AC^p_{\text{loc}}(I; (X,d)) \) for the space of all \( p \)-locally absolutely continuous curves in \( (X,d) \). If \( \gamma \in AC^p_{\text{loc}}(I; (X,d)) \), then the metric speed

\[
|\dot{\gamma}|_d := \lim_{h \to 0} \frac{d(\gamma_{t+h}, \gamma_t)}{|h|}
\]

exists for a.e. \( t \in I \) and \( |\dot{\gamma}|_d \) is the minimal \( g \in L^p_{\text{loc}}(I) \) such that \( \left( AC_p \right) \) holds.

It is immediate from the definition that every admissible curve \( (\rho_t)_{t \in I} \) belongs to \( AC^p_{\text{loc}}(I; (\mathcal{D}(\mathcal{M}, \tau, W))) \) and \( |\dot{\rho}_t|_W \leq |\|D\rho_t\|_{\rho_t} \) for a.e. \( t \in I \).

Corollary 4.26 (Convexity squared metric speed). Let \( (\rho^*_t)_{t \in I}, i \in \{0, 1\} \), be locally absolutely continuous curves in \( (\mathcal{D}(\mathcal{M}, \tau, W)) \) and \( \rho^* = (1-s)\rho^0 + s\rho^1 \) for \( s \in [0,1] \). Then \( \rho^* \) is locally absolutely continuous and

\[
|\rho^*|^2_W \leq (1-s)|\rho^0|^2_W + s|\rho^1|^2_W
\]
for all $s \in [0, 1]$.  

At the present stage we cannot say much about when the distance $W$ between two density matrices is finite. However, if $E$ satisfies some functional inequalities, we get estimates on $W$.

**Proposition 4.27.** Assume that $\tau$ is a state. If $E$ satisfies the Poincaré inequality with constant $c_P > 0$, that is,

$$\|a - \tau(a)\|^2_2 \leq c_P E(a)$$

for all $a \in D(E)$, then

$$W(\rho_0, \rho_1) \leq \frac{c_p}{\lambda} \|\rho_1 - \rho_0\|_2$$

for all $\rho_0, \rho_1 \in D(M, \tau) \cap L^2(M, \tau)$ with $\rho_0, \rho_1 \geq \lambda^2 > 0$

*Proof.* Let $\rho_t = (1 - t)\rho_0 + t\rho_1$ and notice that $\rho_t \geq \lambda^2$ implies $\dot{\rho}_t \geq \lambda^2$. For all $a \in A_{AM}$ we have

$$|\tau(a(\rho_t - \rho_s))| = |t - s||\tau((a - \tau(a))(\rho_1 - \rho_0))|$$

$$\leq c_P |t - s|\|\rho_1 - \rho_0\|_2\|\partial a\|_{H}$$

$$\leq \frac{c_p}{\lambda} \|\rho_1 - \rho_0\|_2 \int_s^t \|\partial a\|_{\rho_r} \, dr.$$  

Hence $(\rho_t)_{t \in [0,1]}$ is admissible with $\|D \rho_t\|_{\rho_r} \leq \frac{c_p}{\lambda} \|\rho_1 - \rho_0\|_2$. \hfill \square

**Remark 4.28.** Qualitatively, this result can be rephrased as follows: The form $E$ satisfies a Poincaré inequality if and only if ker $\partial$ is spanned by 1 and $\partial^* \eta$ has closed range. In this case, if $(\rho_t)$ is the linear interpolation between two density matrices in $L^2$, then

$$\dot{\rho}_t = \partial^* \eta_t$$

has a solution $\eta_t \in \mathcal{H}$. If $\rho_t$ is additionally bounded away from zero, then there is a solution $\xi_t$ to

$$\dot{\rho}_t \xi_t = \eta_t,$$

and $(\rho_t)$ satisfies the continuity equation for the vector field $(\xi_t)$.

**Remark 4.29.** One typical problem for length spaces we have not touched upon yet is the existence of geodesics, that is, length-minimizing curves for a given start and endpoint. A length space is called geodesic if any two points with finite distance are joined by a geodesic. This property is one of the advantages of the metric $W$ in the finite-dimensional case compared to the Wasserstein distance, with the geometry of the geodesics an object of recent attention (see [GLM17, EMW19]).

Unfortunately, $(D(M, \tau), W)$ can fail to be geodesic even in the commutative case, as was pointed out to the author by Erbar. However, we will see in Chapter 6 that (under suitable conditions) the subset of all density matrices with finite entropy is indeed geodesic, and this is enough for the study of geodesic convexity of the entropy.

Next we prove that the action functional appearing in the definition of $W$ is lower semicontinuous with respect to pointwise weak convergence in $L^1$ and show some first consequences. This property will later be important for several approximation arguments.
**Theorem 4.30 (Lower semicontinuity of the action).** If $L^1(\mathcal{M}, \tau)$ is separable, then the action functional

$$E : \mathcal{D}(\mathcal{M}, \tau)^{[0,1]} \longrightarrow [0, \infty], \ (\rho_t) \mapsto \int_0^1 \|D\rho_t\|_{\rho_t}^2 \ dt \quad \text{if} \ (\rho_t) \ \text{is admissible},$$

$$\infty \quad \text{otherwise}$$

is lower semicontinuous with respect to pointwise weak convergence in $L^1(\mathcal{M}, \tau)$.

**Proof.** Let $(\rho^n)$ be a sequence in $\mathcal{D}(\mathcal{M}, \tau)^{[0,1]}$ and $\rho : [0,1] \longrightarrow \mathcal{D}(\mathcal{M}, \tau)$ such that $\rho^n_t \rightarrow \rho_t$ weakly in $L^1$ for all $t \in [0,1]$. Otherwise passing to subsequence, we may assume that $(E(\rho^n))_n$ is convergent. Moreover, if the limit is infinite, there is nothing to prove, so we assume additionally that $\sup_n E(\rho^n) < \infty$. In particular, the curve $(\rho^n_t)_t$ is admissible for all $n \in \mathbb{N}$.

Fix $a \in \mathcal{A}_{AM}$. Since $\rho^n_t \rightarrow \rho_t$ weakly for all $t \in [0,1]$, we have

$$|\tau(a(\rho_t - \rho_s))| = \lim_{n \to \infty} |\tau(a(\rho^n_t - \rho^n_s))| \leq \liminf_{n \to \infty} \int_s^t \|D\rho^n_t\|_{\rho^n_t} \|\partial a\|_{\rho^n_t} \ dr.$$

Let $c^n : [0,1] \longrightarrow [0, \infty)$, $c^n(r) = \|D\rho^n_r\|_{\rho^n_r}$. By assumption, $(c^n)$ is bounded in $L^2([0,1])$, hence we may assume that $c^n \rightarrow c$ weakly in $L^2([0,1])$.

Note that the separability of $L^1(\mathcal{M}, \tau)$ implies that the weak $L^1$-topology restricted to $\mathcal{D}(\mathcal{M}, \tau)$ is metrizable (see [DS88, Theorem V.5.1]). Since $\rho \mapsto \|\partial a\|_{\rho}$ is upper semicontinuous (Theorem 3.15), there is a decreasing sequence $(G_k)$ of weakly continuous functions on $\mathcal{D}(\mathcal{M}, \tau)$ such that

$$\|\partial a\|_{\rho} = \inf_{k \in \mathbb{N}} G_k(\rho)$$

for all $\rho \in \mathcal{D}(\mathcal{M}, \tau)$. Moreover, we can assume that $G_k \leq \|a\|_\theta$ for all $k \in \mathbb{N}$.

Let $g^n(r) = G_k(\rho^n_r)$ and $g(r) = G_k(\rho_r)$. The dominated convergence theorem gives $g^n \rightarrow g$ strongly in $L^2([0,1])$.

Hence

$$|\tau(a(\rho_t - \rho_s))| \leq \lim_{n \to \infty} \int_s^t c^n(r)g^n(r) \ dr = \int_s^t c(r)g(r) \ dr = \int_s^t c(r)G_k(\rho_r) \ dr$$

for all $k \in \mathbb{N}$.

Finally, another application of the dominated convergence theorem yields

$$|\tau(a(\rho_t - \rho_s))| \leq \lim_{k \to \infty} \int_s^t c(r)G_k(\rho_r) \ dr = \int_s^t c(r)\|\partial a\|_{\rho_r} \ dr.$$

This inequality implies that $(\rho_t)$ is admissible and $\int_0^1 \|D\rho_t\|_{\rho_t}^2 \ dt \leq \|c\|_{L^2([0,1])}^2$. Thus

$$\int_0^1 \|D\rho_t\|_{\rho_t}^2 \ dt \leq \|c\|_{L^2([0,1])}^2 \leq \liminf_{n \to \infty} \int_0^1 \|D\rho^n_t\|_{\rho^n_t}^2 \ dt. \quad \square$$

**Remark 4.31.** The separability of $L^1(\mathcal{M}, \tau)$ is equivalent to each of the following properties:

(i) The $\sigma$-weak topology on the unit ball of $\mathcal{M}$ is metrizable.

(ii) $\mathcal{M}$ has a faithful normal representation on a separable Hilbert space.

A von Neumann algebra with one of these properties is often called separable or separably acting.

**Lemma 4.32.** If $\mathcal{A}_{AM}$ is $\sigma$-weakly dense in $\mathcal{M}$ and $L^1(\mathcal{M}, \tau)$ is separable, then every admissible curve is measurable in $L^1(\mathcal{M}, \tau)$.

**Proof.** Let $A$ be the uniform closure of $\mathcal{A}_{AM}$. By Kaplansky’s density theorem, $A \cap \mathcal{M}_1$ is $\sigma$-weakly dense in $\mathcal{M}_1$. Since $L^1(\mathcal{M}, \tau)$ is separable, the $\sigma$-weak topology is metrizable on $\mathcal{M}_1$. Thus, for every $a \in \mathcal{M}$ there exists a sequence $(a_k)$ in $A$ such that $a_k \rightarrow a$ $\sigma$-weakly.
If \((\rho_t)\) is an admissible curve, then \(t \mapsto \tau(\rho_t a_k)\) is continuous for all \(k \in \mathbb{N}\). Therefore \(t \mapsto \tau(\rho_t a)\) is measurable as pointwise limit of a sequence of continuous functions. Using once more the separability of \(L^1(\mathcal{M}, \tau)\), we conclude that \((\rho_t)\) is measurable in \(L^1(\mathcal{M}, \tau)\) due to Pettis' measurability theorem (see [DU77, Theorem II.2]).

**Lemma 4.33.** Assume that \(\mathcal{A}_{AM}\) is \(\sigma\)-weakly dense in \(\mathcal{M}\) and \(L^1(\mathcal{M}, \tau)\) is separable. If \((\rho_t)_{t \in [0,1]}\) is an admissible curve in \(\mathcal{D}(\mathcal{M}, \tau)\), then there exists a family of admissible curves \((\rho_t^\varepsilon) \in C^\infty([0,1]; L^1(\mathcal{M}, \tau))\) such that \(\rho_0^\varepsilon = \rho_0\), \(\rho_1^\varepsilon = \rho_1\) for \(\varepsilon > 0\), and

\[
\limsup_{\varepsilon \to 0} \int_0^1 \|D\rho_t^\varepsilon\|^2_{\rho_t^\varepsilon} dt \leq \int_0^1 \|D\rho_t\|^2_{\rho_t} dt,
\]

\[
\limsup_{\varepsilon \to 0} \text{ess sup}_{t \in [0,1]} \|D\rho_t^\varepsilon\|^2_{\rho_t^\varepsilon} \leq \text{ess sup}_{t \in [0,1]} \|D\rho_t\|^2_{\rho_t},
\]

In particular, the infimum in the definition of \(W\) can alternatively be taken over \(L^1\)-admissible curves.

**Proof.** Let \((\rho_t)_{t \in [0,1]}\) be an admissible curve. Extend it to a curve \((\rho_t)_{t \in \mathbb{R}}\) by setting \(\rho_t = \rho_0\) for \(t < 0\) and \(\rho_t = \rho_1\) for \(t > 1\). Note that the extended curve is still admissible with \(D\rho_t = 0\) for \(t \in (-\infty, 0) \cup (1, \infty)\).

Let \((\eta_t)_{t > 0}\) be a mollifying kernel on \(\mathbb{R}\) with \(\text{supp } \eta_t \subset (-\varepsilon, \varepsilon)\) and set

\[
\rho_t^\varepsilon = \int_{\mathbb{R}} \eta_t(s) \rho_{t-s} ds,
\]

where the integral is to be understood as Pettis integral in \(L^1(\mathcal{M}, \tau)\). The measurability of \((\rho_t)\) is guaranteed by Lemma [4.32].

The curve \((\rho_t^\varepsilon)_{t \in \mathbb{R}}\) is in \(C^\infty(\mathbb{R}; L^1(\mathcal{M}, \tau))\) and satisfies \(\rho_t^\varepsilon = \rho_0\) for \(t \leq -\varepsilon\) and \(\rho_t^\varepsilon = \rho_1\) for \(t \geq 1 + \varepsilon\). Moreover, if \(a \in \mathcal{A}_{AM}\), then

\[
|\tau(a(\rho_t^\varepsilon - \rho_s^\varepsilon))| \leq \int_{\mathbb{R}} \eta_t(r) |\tau(a(\rho_{t-r} - \rho_{s-r}))| dr
\]

\[
\leq \int_{\mathbb{R}} \eta_t(r) \int_{s-r}^t \|\partial a\|_{\rho_u} \|D\rho_u\|_{\rho_u} du dr
\]

\[
= \int_s^t \int_{\mathbb{R}} \eta_t(r) \|\partial a\|_{\rho_u} \|D\rho_{u-r}\|_{\rho_{u-r}} du dr
\]

\[
\leq \int_s^t \left( \int_{\mathbb{R}} \eta_t(r) \|\partial a\|^2_{\rho_{u-r}} dr \right)^{\frac{1}{2}} \left( \int_{\mathbb{R}} \eta_t(r) \|D\rho_{u-r}\|^2_{\rho_{u-r}} dr \right)^{\frac{1}{2}} du
\]

Since \(\rho \mapsto \|\partial a\|^2\rho\) is upper semicontinuous and concave by Lemmas [3.15, 3.27], we can apply the vector-valued version of Jensen’s inequality (see [Per74, Theorem 3.10]) to get

\[
\int_{\mathbb{R}} \eta_t(r) \|\partial a\|^2_{\rho_{u-r}} dr \leq \|\partial a\|^2_{\rho_u}.
\]

Thus \((\rho_t^\varepsilon)_{t \in \mathbb{R}}\) is admissible and

\[
\|D\rho_t^\varepsilon\|^2_{\rho_t^\varepsilon} \leq \int_{\mathbb{R}} \eta_t(r) \|D\rho_{t-r}\|^2_{\rho_{t-r}} dr.
\]

for a.e. \(t \in [0,1]\).

Finally one can reparametrize \((\rho_t^\varepsilon)\) in such a way that \(\rho_0^\varepsilon = \rho_0\), \(\rho_1^\varepsilon = \rho_1\) and the claimed inequalities are retained. 

\(\square\)
Remark 4.34. Mollifying in the time variable to restrict minimization problems to smooth curves is a standard argument, but the nonlinearity of $\|\xi\|_\rho$ in $\rho$ requires some additional care in our case. In particular, the upper semicontinuity of $\rho \mapsto \|\partial a\|_\rho^2$ is crucial here, because the vector-valued version of Jensen’s inequality may fail otherwise (see [Per74]).

5. Von Neumann entropy and Fisher information

In this section we introduce two important quantities for the gradient flow characterization of the heat flow, namely the (von Neumann) entropy and the Fisher information or entropy production, and carefully analyze convexity and continuity properties of these two as well as some related functionals.

The Fisher information appears in quantum information theory as the derivative of the entropy along heat flow curves, hence the name entropy production, but we will see that it also occurs in the metric speed of heat flow trajectories with respect to $\mathcal{W}$. This foreshadows already the close relation between entropy, heat flow and the metric $\mathcal{W}$, which we will exploit for the gradient flow characterization in the next section.

Throughout this section let $(\mathcal{M}, \tau)$ be a tracial von Neumann algebra, $\mathcal{E}$ a quantum Dirichlet form on $L^2(\mathcal{M}, \tau)$, $(\partial, \mathcal{H}, L, R, J)$ the associated first order differential calculus and assume that $\tau$ is energy dominant. We further assume that $\theta$ can be represented by a symmetric operator mean. Denote by $(P_t)_{t \geq 0}$ the quantum Markov semigroup associated with $\mathcal{E}$ and by $\mathcal{L} = \partial^\ast \partial$ its generator.

The von Neumann entropy is defined as

$$\text{Ent}: \mathcal{D}(\mathcal{M}, \tau) \longrightarrow [-\infty, \infty], \quad \text{Ent}(\rho) = \begin{cases} \tau(\rho \log \rho) & \text{if } (\rho \log \rho)_+ \in L^1(\mathcal{M}, \tau), \\ \infty & \text{otherwise}. \end{cases}$$

Its domain of definition is $\mathcal{D}(\text{Ent}) = \{\rho \in \mathcal{D}(\mathcal{M}, \tau) \mid \text{Ent}(\rho) \in \mathbb{R}\}$. Here and in the following, the expression $\rho \log \rho$ is to be understood as $f(\rho)$ for the (continuous) function

$$f: [0, \infty) \longrightarrow \mathbb{R}, \quad x \mapsto \begin{cases} x \log x & \text{if } x > 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Remark 5.1. If $\tau(1) = 1$, an application of Jensen’s inequality shows that $\text{Ent} \geq 0$. If $\tau(1) = \infty$, the entropy can be rather ill-behaved already in the commutative case (see e.g. [Stu06a, Example 4.4]). For that reason, we will from now on concentrate on the finite case. However, we believe that this assumption is not essential and that similar modifications as in [AGS14a] should also work in our setting.

Remark 5.2. Some authors, especially in the physics community, define the entropy with the opposite sign. We choose the sign in such a way that the entropy is positive if $\tau$ is a state and the semigroup $(P_t)$ is a gradient flow of $\text{Ent}$ instead of $-\text{Ent}$.

An important property of the entropy is its lower semicontinuity with respect to suitable topologies on $\mathcal{D}(\mathcal{M}, \tau)$, in our case the topology induced by $\mathcal{W}$. To prove it, we first establish a variational formulation of the entropy. In the noncommutative case, it is (along with the idea of proof presented here) originally due to Petz [Pet88]. We just adapt it to be applicable in duality with $\mathcal{A}_{\text{LM}}$ instead of $\mathcal{M}$.

Proposition 5.3 (Variational formula for the entropy). Assume $\tau(1) = 1$. If $A$ is a $\sigma$-weakly dense $C^*$-subalgebra of $\mathcal{M}$, then

$$\text{Ent}(\rho) = \sup \{\tau(\rho a) - \log \tau(e^a) \mid a \in A_+\}$$

for all $\rho \in \mathcal{D}(\mathcal{M}, \tau)$. 

Proof. Step 1: The equality
\[ \operatorname{Ent}(\rho) = \max \{ \tau(a\rho) - \log \tau(e^a) \mid a \in \mathcal{M}_+ \} \]
holds for invertible \( \rho \in \mathcal{D}(\mathcal{M}, \tau) \cap \mathcal{M} \), and the maximum is attained at \( a = \log M\rho \) for all sufficiently large \( M \).

By assumption, there are constants \( c, C > 0 \) such that \( c \leq \rho \leq C \). Let \( a = \log M\rho \in \mathcal{M}_+ \) for \( M \geq e^{-1} \). Then
\[ \tau(a\rho) - \log \tau(e^a) = \tau(\rho \log \rho) + \log M - \log \tau(M\rho) = \tau(\rho \log \rho). \]

For the converse inequality let \( x = \log \rho, \ y = a - \log \tau(e^a) \). By Klein’s inequality (Lemma 5.9) we have
\[ 0 \leq \tau(e^y - e^x - e^x(y - x)) = \tau\left(\rho - \frac{e^a}{\tau(e^a)} - \rho(a - \log \tau(e^a) - \log \rho)\right) = \tau(\rho \log \rho) - (\tau(a\rho) - \log \tau(e^a)). \]

Step 2: The equality
\[ \operatorname{Ent}(\rho) = \sup \{ \tau(a\rho) - \log \tau(e^a) \mid a \in \mathcal{M}_+ \} \]
holds for all \( \rho \in \mathcal{D}(\mathcal{M}, \tau) \):

Let \( \rho^{(n)} = (\rho \wedge n) \vee \frac{1}{n} \), denote by \( e \) the spectral measure of \( \rho \) and let \( \mu = \tau \circ e \).

Then
\[ \operatorname{Ent}(\rho^{(n)}) = \int_{[0, e^{-1}]} \left( \lambda \vee \frac{1}{n} \right) \log \left( \lambda \vee \frac{1}{n} \right) d\mu(\lambda) + \int_{(e^{-1}, \infty)} (\lambda \wedge n) \log(\lambda \wedge n) d\mu(\lambda) \]
\[ \to \int \lambda \log \lambda d\mu(\lambda), \ n \to \infty \]
by monotone convergence.

Moreover,
\[ \tau(\log \rho^{(n)}(\rho^{(n)} - \rho)) = \int_{[0, \infty)} \log \lambda^{(n)}(\lambda^{(n)} - \lambda) d\mu(\lambda) \leq 0, \]
as can be seen by a decomposition of \([0, \infty)\) into \([0, 1/n), [1/n, n]\) and \((n, \infty)\).

Let \( a_n = \log n\rho^{(n)} \). Using the first step, we get
\[ \operatorname{Ent}^{(n)}(\rho) = \lim_{n \to \infty} \operatorname{Ent}(\rho^{(n)}) = \lim_{n \to \infty} (\tau(\rho^{(n)} a_n) - \log \tau(e^{a_n})) \leq \sup_{n \in \mathbb{N}} (\tau(\rho a_n) - \log \tau(e^{a_n})). \]

For the converse inequality, first observe that
\[ \tau(|\rho - \rho^{(n)}|) = \int_{(0, 1/n)} \left( \frac{1}{n} - \lambda \right) d\mu(\lambda) + \int_{(n, \infty)} (\lambda - n) d\mu(\lambda) \to 0, \ n \to \infty \]
by dominated convergence.

Now let \( a \in \mathcal{M}_+ \). Then
\[ \operatorname{Ent}(\rho) = \lim_{n \to \infty} \operatorname{Ent}(\rho^{(n)}) \geq \lim_{n \to \infty} \tau(a\rho^{(n)}) - \log \tau(e^a) = \tau(a\rho) - \log \tau(e^a). \]

Step 3: The equality
\[ \operatorname{Ent}(\rho) = \sup \{ \tau(a\rho) - \log \tau(e^a) \mid a \in A_+ \} \]
holds for all \( \rho \in \mathcal{D}(\mathcal{M}, \tau) \):

Let \( \varepsilon > 0 \). By the second step, there is an \( a \in \mathcal{M}_+ \) such that
\[ \operatorname{Ent}(\rho) \leq \tau(a\rho) - \log \tau(e^a) + \frac{\varepsilon}{3}. \]

Since \( A \subset \mathcal{M} \) is \( \sigma \)-weakly dense, the unit ball of \( A \) is strongly dense in the unit ball of \( \mathcal{M} \) by Kaplansky’s density theorem. Thus there is a net \( (a_i) \) in \( A_+ \) with
Proof. Let $i \in I$ such that $\tau(a) \leq \tau(a_i) + \varepsilon/3$, $-\log \tau(e^a) \leq -\log \tau(e^{a_i}) + \varepsilon/3$. Then

$$\text{Ent}(\rho) \leq \tau(a_i) - \log \tau(e^{a_i}) + \varepsilon.$$ 

Thus $\text{Ent}(\rho) \leq \sup \{\tau(a) - \log \tau(e^a) \mid a \in A_+\}$. The converse inequality is clear from Step 2. \qed

For the next corollary recall that a convex function is called proper if it is not identically $\infty$.

**Corollary 5.4.** If $\tau(1) = 1$ and $A_{LM} \subset \mathcal{M}$ is $\sigma$-weakly dense, then $\text{Ent}$ is a proper lower semicontinuous convex functional on $(\mathcal{D}(\mathcal{M}, \tau), W)$.

**Proof.** Denote by $A_{LM}$ the uniform closure of $A_{LM}$. Let $(\rho_n)$ be a sequence in $\mathcal{D}(\mathcal{M}, \tau)$ such that $W(\rho_n, \rho) \to 0$. It follows from Proposition 4.20 that $\tau(\rho_n) \to \tau(\rho)$ for all $\rho \in A_0$, and, since $(\rho_n)$ is bounded in $L^1(\mathcal{M}, \tau)$, indeed for all $\rho \in A_0$.

Combined with Proposition 5.5 we infer that $\text{Ent}$ is the supremum of affine, continuous functions on $(\mathcal{D}(\mathcal{M}, \tau), W)$, hence lower semicontinuous and convex. Since $\text{Ent}(1) = 0$, the entropy is also proper. \qed

**Remark 5.5.** As we already used in the proof of Theorem 3.15, functionals of the form $\rho \mapsto \tau(f(\rho))$ are convex for all convex functions $f$.

For approximating functionals we will also need the following (weaker) lower semi-continuity property, which holds for convex trace functionals.

**Proposition 5.6** (Lower semicontinuity of convex trace functionals). Assume that $\tau(1) < \infty$. For an interval $I \subset \mathbb{R}$ let $L^1(\mathcal{M}, \tau)_I = \{a \in L^1(\mathcal{M}, \tau) \mid a = a^*, \sigma(a) \subset I\}$. If $f : I \to \mathbb{R}$ is a proper lower semicontinuous convex function, then

$$F : L^1(\mathcal{M}, \tau)_I \to (-\infty, \infty], \ a \mapsto \tau(f(a))$$

is lower semicontinuous.

**Proof.** Since $f$ is proper, lower semicontinuous and convex, by [Roc97, Theorem 12.1] there is a sequence $(f_i)$ of affine functions such that $f = \sup_i f_i$. In particular, $F$ is well-defined.

Fix $n \in \mathbb{N}$. For $j \in \{1, \ldots, n\}$ let

$$E_j = \{t \in I \mid \max_{1 \leq i < j} f_i(t) < f_j(t), \max_{j < i \leq n} f_i(t) \leq f_j(t)\}.$$

In particular, $f_j(t) = \max_{1 \leq i \leq n} f_i(t)$ for $t \in E_j$, and $I = \bigcup_j E_j$.

Let $(a_k)$ be a sequence in $L^1(\mathcal{M}, \tau)_I$ converging to $a \in L^1(\mathcal{M}, \tau)_I$. Then

$$\liminf_{k \to \infty} \tau(f(a_k)) = \liminf_{k \to \infty} \sum_{j=1}^n \tau(\mathbb{1}_{E_j}(a) f(a_k))$$

$$\geq \liminf_{k \to \infty} \sum_{j=1}^n \tau(\mathbb{1}_{E_j}(a) f_j(a_k))$$

$$= \sum_{j=1}^n \tau(\mathbb{1}_{E_j}(a) f_j(a))$$

$$= \sum_{j=1}^n \tau \left( \mathbb{1}_{E_j}(a) \max_{1 \leq i \leq n} f_i(a) \right)$$

$$= \tau \left( \max_{1 \leq i \leq n} f_i(a) \right).$$
Since \( f_1(a) \) is integrable, \( \tau (\max_{1 \leq i \leq n} f_i(a)) \) \( \geq \tau (f(a)) \). This proves the claim. \( \square \)

**Remark 5.7.** By the Hahn-Banach theorem, lower semicontinuity and weak lower semicontinuity are equivalent for convex functionals on convex subsets of Banach spaces.

**Lemma 5.8.** Assume that \( \tau \) is finite. Let \( (\rho_t)_{t \in I} \) be an \( L^1 \)-differentiable curve in \( D(\mathcal{M}, \tau) \). Assume that there exists \( M > 0 \) such that \( \| \rho_t \|_\mathcal{M} \leq M \) for all \( t \in I \). If \( f \in C^1(\mathbb{R}) \), then the map

\[
F : I \longrightarrow \mathbb{R}, \quad t \mapsto \tau(f(\rho_t))
\]

is differentiable with derivative \( F'(t) = \tau(f'(\rho_t)\hat{\rho}_t) \).

**Proof.** By approximation in \( C^1 \) we can assume that \( f \) is a polynomial, and by linearity, even that \( f \) is a monomial. Observe that

\[
\tau(a^{k+1} - b^{k+1}) = \tau \left( \sum_{j=0}^{k} a^{k-j} b^j (a-b) \right)
\]

for all \( a, b \in \mathcal{M} \) and \( k \geq 0 \) (of course it is crucial here that \( \tau \) is a trace).

Thus

\[
\frac{1}{h} \tau(\rho_{t+h}^{k+1} - \rho_t^{k+1}) = \sum_{j=0}^{k} \tau \left( \rho_{t+h}^{k-j} \rho_t^j \left( \frac{\rho_{t+h} - \rho_t}{h} \right) \right)
\]

Since \( (\rho_t) \) is uniformly bounded and \( L^1 \)-continuous, we have \( \rho_{t+h} \rightarrow \rho_t \) as \( h \rightarrow 0 \) in \( L^1 \) and, by a standard approximation argument, also \( \sigma \)-weakly for all \( j \geq 0 \). Hence

\[
\left| \tau \left( \rho_{t+h}^{k-j} \rho_t^j \left( \frac{\rho_{t+h} - \rho_t}{h} \right) - \rho_t^k \hat{\rho}_t \right) \right| \leq \tau \left( \rho_{t+h}^{k-j} \rho_t^j \left( \frac{\rho_{t+h} - \rho_t}{h} - \hat{\rho}_t \right) \right)
\]

\[
+ \tau \left( \rho_{t+h}^{k-j} \rho_t^j \hat{\rho}_t \right) \]

\[
\leq M^k \tau \left( \left| \rho_{t+h} - \rho_t \right| \right)
\]

\[
(\rho_t \rightarrow \rho_t^j \hat{\rho}_t) \rightarrow 0, \quad h \rightarrow 0.
\]

All put together, we have proven that

\[
\frac{1}{h}(F(t+h) - F(t)) = \frac{1}{h} \tau(\rho_{t+h}^{k+1} - \rho_t^{k+1}) \rightarrow \tau((k+1)\rho_t^k \hat{\rho}_t)
\]

as \( h \rightarrow 0 \). \( \square \)

**Lemma 5.9** (Klein’s inequality). Assume that \( \tau \) is finite. Let \( f : \mathbb{R} \longrightarrow \mathbb{R} \) be a convex \( C^1 \) Lipschitz function. If \( \rho_0, \rho_1 \in L^1_0(\mathcal{M}, \tau) \), then \( f(\rho_0), f(\rho_1) \in L^1(\mathcal{M}, \tau) \) and

\[
\tau(f(\rho_1) - f(\rho_0)) \leq \tau(f'(\rho_1)(\rho_1 - \rho_0)).
\]

**Proof.** If \( \rho_0, \rho_1 \in L^1_0(\mathcal{M}, \tau) \cap \mathcal{M} \), then one can use Lemma 5.8 to adapt the proof of the finite-dimensional case (see [Car10, Theorem 2.11]).

Now let \( \rho_0, \rho_1 \in L^1_0(\mathcal{M}, \tau) \) be arbitrary. We already know that the inequality holds for \( \rho_0, \rho_1 \) replaced by \( \rho_0 \wedge n, \rho_1 \wedge n \). It remains to show that both sides converge to the correct limit as \( n \rightarrow \infty \).
Since \( f' \) is bounded and increasing, \( \|f'(\rho_1 \land n)\|_{L^1} \leq \|f'\|_{\infty} \) for all \( n \in \mathbb{N} \) and \( f'(\rho_1 \land n) \to f'(\rho_1) \) \( \sigma \)-weakly as \( n \to \infty \). It is elementary that \( \rho_0 \land n \to \rho_0 \) and \( \rho_1 \land n \to \rho_1 \) in \( L^1 \). Put together we get
\[
\tau(f'(\rho_1 \land n)(\rho_1 \land n - \rho_0 \land n)) \to \tau(f'(\rho_1)(\rho_1 - \rho_0)).
\]
For the convergence of the left-hand side one can use that \( \mathbb{R} \) can be decomposed into at most two intervals such that \( f \) is decreasing on the first one and decreasing on the second. Using monotone convergence on both parts one gets \( \tau(f(\rho_1 \land n)) \to \tau(f(\rho_1)) \) and the same for \( \rho_0 \).

**Corollary 5.10.** Assume that \( \tau \) is finite. If \( f: \mathbb{R} \to \mathbb{R} \) is a convex \( C^1 \) Lipschitz function and \( (\rho_t)_{t \in I} \) is an \( L^1 \)-differentiable curve in \( L^1_h(M,\tau) \), then \( t \mapsto \tau(f(\rho_t)) \) is locally absolutely continuous and
\[
\frac{d}{dt}\tau(f(\rho_t)) = \tau(f'(\rho_t)\dot{\rho}_t)
\]
for a.e. \( t \in I \).

**Proof.** By Klein’s inequality we have
\[
\tau(f'(\rho_t)(\rho_t - \rho_s)) \leq \tau(f(\rho_t) - f(\rho_s)) \leq \tau(f'(\rho_t)(\rho_t - \rho_s))
\]
for all \( s, t \in I \). Thus
\[
|\tau(f(\rho_t) - f(\rho_s))| \leq \|f'\|_{\infty}\|\rho_t - \rho_s\|_1.
\]
Since \( (\rho_t) \) is \( L^1 \)-differentiable, it follows that \( t \mapsto \tau(f(\rho_t)) \) is locally absolutely continuous.

Now assume that \( t \in I \) is a point of differentiability. By Klein’s inequality,
\[
\frac{d}{dt}\tau(f(\rho_t)) = \lim_{s \nearrow t} \frac{1}{t-s}\tau(f(\rho_t) - f(\rho_s)) \leq \lim_{s \nearrow t} \tau \left( \frac{f'(\rho_t)\rho_t - \rho_s}{t-s} \right) = \tau(f'(\rho_t)\dot{\rho}_t),
\]
and
\[
\frac{d}{dt}\tau(f(\rho_t)) = \lim_{s \nearrow t} \frac{1}{t-s}\tau(f(\rho_t) - f(\rho_s)) \geq \lim_{s \nearrow t} \tau \left( \frac{f'(\rho_t)\rho_t - \rho_s}{t-s} \right) = \tau(f'(\rho_t)\dot{\rho}_t).
\]
This settles the claim. \( \square \)

**Lemma 5.11.** If \( a \in D(\mathcal{E})_h \) and \( C_1, C_2: \mathbb{R} \to \mathbb{R} \) are increasing Lipschitz functions with \( C_1(0) = C_2(0) = 0 \), then
\[
\mathcal{E}(C_1(a), C_2(a)) \geq 0.
\]

**Proof.** Since \( C_1(s) - C_1(t) \) and \( C_2(s) - C_2(t) \) have the same sign,
\[
|C_1(s) - C_1(t) + C_2(t) - C_2(s)| \leq |C_1(s) + C_1(t) - C_2(s) - C_2(t)|
\]
for all \( s, t \in \mathbb{R} \). Hence there exists a \( 1 \)-Lipschitz function \( C: \mathbb{R} \to \mathbb{R} \) with \( C(0) = 0 \) such that \( C \circ (C_1 + C_2) = C_1 - C_2 \). Thus
\[
\mathcal{E}(C_1(a), C_2(a)) = \frac{1}{4}(\mathcal{E}(C_1(a) + C_2(a)) - \mathcal{E}(C_1(a) - C_2(a))) \geq 0. \quad \square
\]

**Remark 5.12.** If \( C_1, C_2 \) in the previous lemma are continuously differentiable, we can also use the chain rule to get
\[
\mathcal{E}(C_1(a), C_2(a)) = \langle (\tilde{C}_1\tilde{C}_2)(L(a), R(a))\partial a, \partial a \rangle \geq 0.
\]

**Remark 5.13.** The short proof of Lemma 5.11 was shown to the author by S. Puchert, replacing a quite involved proof arguing by approximation of \( \mathcal{E} \).
Corollary 5.14. Let $C, C_1, C_2 : \mathbb{R} \to \mathbb{R}$ be increasing Lipschitz functions with $C(0) = C_1(0) = C_2(0) = 0$ and $|C_1(s) - C_1(t)| \leq |C_2(s) - C_2(t)|$ for all $s, t \in \mathbb{R}$. Then

$$\mathcal{E}(C(a), C_1(a)) \leq \mathcal{E}(C(a), C_2(a))$$

for all $a \in D(\mathcal{E})$.

Lemma 5.15. Let $\mathcal{E}$ be a quantum Dirichlet form on $L^2(\mathcal{M}, \tau)$, denote by $\mathcal{L}^{(p)}$ the generator of the associated semigroup on $L^p(\mathcal{M}, \tau)$ for $p \in [1, \infty)$, and let $a \in D(\mathcal{L}^{(p)})$.

If $C_1, C_2 : \mathbb{R} \to \mathbb{R}$ are increasing 1-Lipschitz functions and there exists a constant $\alpha > 0$ such that $|C_i(t)| \leq \alpha|t|^{p-1}$ for $t \in \mathbb{R}$, $i \in \{1, 2\}$, then $C_1(a), C_2(a) \in D(\mathcal{E})$ and

$$\mathcal{E}(C_1(a), C_2(a)) \leq \tau(C_1(a)\mathcal{L}^{(p)}a).$$

Proof. First note that $C_i(a) \in L^p(\mathcal{M}, \tau) \cap L^q(\mathcal{M}, \tau) \subset L^2(\mathcal{M}, \tau)$, where $q$ is the dual exponent of $p$. To prove $C_i(a) \in D(\mathcal{E})$, it suffices to show

$$\frac{1}{t}\tau(C_i(a)(C_i(a) - P_tC_i(a)) \leq \frac{1}{t}\tau(C_i(a)(a - P_ta))$$

for $t > 0$, since the right-hand side converges to $\tau(C_i(a)\mathcal{L}^{(p)}a)$ as $t \to 0$.

Since the approximating form

$$L^2(\mathcal{M}, \tau) \twoheadrightarrow [0, \infty), x \mapsto \frac{1}{t}\tau(x(x - P_tx))$$

is a quantum Dirichlet form, Equation (5.2) holds for $a \in L^2(\mathcal{M}, \tau)$ by Corollary 5.14.

In the general case let $(a_k)$ be a sequence in $L^2(\mathcal{M}, \tau) \cap L^p(\mathcal{M}, \tau)$. By [Tik87, Theorem 3.2] we have $C_i(a_k) \to C_i(a)$ with respect to $||\cdot||_p$ and $||\cdot||_q$ (resp. $\sigma$-weakly in the case $p = 1$). Using the continuity of $P_t$ with respect to $||\cdot||_p$ and $||\cdot||_q$ (and additionally the bound $||C_i(a_k)||_M \leq ||C_i||_\infty$ in the case $p = 1$), we see that (5.2) continues to hold for arbitrary $a \in L^p(\mathcal{M}, \tau)$.

Finally, by Corollary 5.14 and the same approximation argument as above, we obtain

$$\frac{1}{t}\tau(C_1(a)(C_2(a) - P_tC_2(a))) \leq \frac{1}{t}\tau(C_1(a)(a - P_ta)),$$

which gives

$$\mathcal{E}(C_1(a), C_2(a)) \leq \tau(C_1(a)\mathcal{L}^{(p)}a)$$

in the limit $t \to 0$. \qed

Lemma 5.16. Let $(C_n)$ be a sequence of continuously differentiable, increasing normal contractions on $[0, \infty)$ with

$$\tilde{C}_n(s, t) \nearrow \log(s, t)$$

for all $s, t \geq 0$ (with the convention that the right-hand side equals $\infty$ whenever $s = 0$ or $t = 0$).

Then $\lim_{n \to \infty} \mathcal{E}(a, C_n(a))$ exists in $[0, \infty]$ for all $a \in D(\mathcal{E})_+$ and is independent from the choice of the sequence $(C_n)$.

Proof. We use the chain rule to get

$$\mathcal{E}(a, C_n(a)) = \langle \tilde{C}_n(L(a), R(a))\partial a, \partial a \rangle_H.$$
Denote by \( e \) the joint spectral measure of \( L(a) \) and \( R(a) \). Then
\[
\langle \tilde{C}_n(L(a), R(a)) \partial a, \partial a \rangle_H = \int_{[0, \infty)^2} \tilde{C}_n(s, t) d\langle e(s, t) \partial a, \partial a \rangle_H
\]
\[
\not\rightarrow \int_{[0, \infty)^2} \log(s, t) d\langle e(s, t) \partial a, \partial a \rangle_H,
\]
where the integrand is interpreted as \( \infty \) whenever \( s = 0 \) or \( t = 0 \).

**Definition 5.17** (Fisher information). The *Fisher information* of \( a \in D(E)_+ \) is defined as
\[
\mathcal{I}(a) = \lim_{n \to \infty} \mathcal{E}(a, C_n(a)) \in [0, \infty]
\]
for some (any) sequence \( (C_n) \) of continuously differentiable, increasing normal contractions with \( \tilde{C}_n \not\rightarrow \log \) pointwise.

An example of a sequence \( (C_n) \) that is admissible in the definition of the Fisher information is given by
\[
C_n: [0, \infty) \to [0, \infty), \quad C_n(t) = \log(t + e^{-n}) + n.
\]

We will also need the following two different approximation results of the Fisher information:

**Lemma 5.18.** If \( a \in D(E)_+ \) with \( \mathcal{I}(a) < \infty \), then
\[
\mathcal{I}(a) = \lim_{n \to \infty} \mathcal{E}(a, ((\log \wedge n) \vee (-n) + n)(a)).
\]

**Proof.** Let \( f_n = (\log \wedge n) \vee (-n) + n \). By Corollary 5.14, the sequence \( \mathcal{E}(a, f_n(a)) \) is increasing, hence it suffices to show convergence along a subsequence. Let
\[
C_k: [0, \infty) \to [0, \infty), \quad C_k(t) = \log(t + e^{-k}) + k
\]
and \( n_k \in \mathbb{N} \) such that \( \mathcal{E}((C_k(a) \wedge n_k) \vee (-n_k)) \geq \mathcal{E}(C_k(a)) - \frac{1}{k} \). Then
\[
\mathcal{E}(a, (C_k(a) \wedge n_k) \vee (-n_k)) \to \mathcal{I}(a), \quad k \to \infty,
\]
and
\[
\left| (C_k(s) \wedge n_k) \vee (-n_k) - (C_k(t) \wedge n_k) \vee (-n_k) \right| \leq |f_{n_k}(s) - f_{n_k}(t)| \leq |\log s - \log t|.
\]
An application of Corollary 5.14 yields
\[
\mathcal{E}(a, f_{n_k}(a)) \to \mathcal{I}(a), \quad k \to \infty.
\]

**Lemma 5.19.** Let \( a \in D(E)_+ \). Then \( \mathcal{I}(a) = \sup_n \mathcal{I}(a \wedge n) \).

**Proof.** By Corollary 5.14, the sequence \( (\mathcal{I}(a \wedge n))_n \) is increasing and bounded from above by \( \mathcal{I}(a) \). For the converse inequality let \( (C_m) \) be a sequence as in the definition of \( \mathcal{I} \).

Since \( C_m(a) = \lim_{n \to \infty} C_m(a \wedge n) \) in \( L^2(M, \tau) \) and \( \mathcal{E}(C_m(a \wedge n)) \leq \mathcal{E}(C_m(a)) \) by Corollary 5.14, we have \( \mathcal{E}(C_m(a \wedge n)) \to \mathcal{E}(C_m(a)) \) as \( n \to \infty \) by the lower semicontinuity of \( \mathcal{E} \). The same argument holds for \( \mathcal{E}(a \wedge n) \) so that we get
\[
\mathcal{I}(a) \geq \mathcal{I}(a \wedge n) \geq \mathcal{E}(a \wedge n, C_m(a \wedge n)) \to \mathcal{E}(a, C_m(a)).
\]
Taking the supremum over \( m \in \mathbb{N} \), the assertion follows.

With the aid of the previous lemma, we can extend the Fisher information to \( L^1_+(M, \tau) \) via \( \mathcal{I}(\rho) = \sup_n \mathcal{I}(\rho \wedge n) \) if \( \rho \wedge n \in D(E)_+ \) for all \( n \in \mathbb{N} \) and \( \mathcal{I}(\rho) = \infty \) otherwise.

**Lemma 5.20.** The Fisher information is lower semicontinuous on \( L^1_+(M, \tau) \).
Proof. By monotone approximation it suffices to show that maps of the form

$$L^1_+ ( \mathcal{M}, \tau ) \rightarrow [0, \infty), \rho \mapsto \begin{cases} \mathcal{E}(\rho \wedge n, C(\rho \wedge n)) & \text{if } \rho \wedge n \in D(\mathcal{E}), \\ \infty & \text{otherwise} \end{cases}$$

are continuous for $n \in \mathbb{N}$ and $C \in C^1(\mathbb{R})$ with $C' > 0$.

Let $(\rho_k)$ be a sequence in $L^1_+ ( \mathcal{M}, \tau )$ such that $\rho_k \rightarrow \rho$ in $L^1_+ ( \mathcal{M}, \tau )$ and $\sup_k \mathcal{E}(\rho_k \wedge n, C(\rho_k \wedge n)) < \infty$. By [Tik87, Theorem 3.2] we have $\rho_k \wedge n \rightarrow \rho \wedge n$ in $L^1_+ ( \mathcal{M}, \tau )$ and $L^2(\mathcal{M}, \tau)$. Moreover, since $C' > 0$, the sequence $(\mathcal{E}(\rho_k \wedge n))_k$ is bounded. Thus $\rho \wedge n \in D(\mathcal{E})$ and $\rho_k \wedge n \rightarrow \rho \wedge n$ weakly in $D(\mathcal{E})$.

By Lemma 3.14 we have $\tilde{\mathcal{C}}(\rho_k \wedge n), R(\rho_k \wedge n)) \rightarrow \tilde{\mathcal{C}}(\rho \wedge n, R(\rho \wedge n))$ strongly as $k \rightarrow \infty$. If we combine these convergences, we obtain

$$\mathcal{E}(\rho_k \wedge n, C(\rho_k \wedge n)) \rightarrow (\partial(\rho \wedge n), \tilde{\mathcal{C}}(\rho \wedge n), R(\rho \wedge n)\partial(\rho \wedge n))_{\mathcal{H}}$$

by the lower semicontinuity of $\mathcal{E}$, and

$$\mathcal{I}((1 - \lambda)\rho_0 + \lambda \rho_1) \leq \liminf_{n \rightarrow \infty} \mathcal{I}((1 - \lambda)(\rho_0 \wedge n) + \lambda(\rho_1 \wedge n))$$

by the lower semicontinuity of $\mathcal{I}$, and

$$(1 - \lambda)\mathcal{I}(\rho_0 \wedge n) + \lambda \mathcal{I}(\rho_1 \wedge n) \leq (1 - \lambda)\mathcal{I}(\rho_0) + \lambda \mathcal{I}(\rho_1)$$

by definition.

For that purpose let

$$C_k : [0, \infty) \rightarrow [0, \infty), t \mapsto \log(t + e^{-n}) + n$$

and let $\mathcal{E}_\varepsilon$ be the quadratic form generated by $\mathcal{L}(1 + \varepsilon \mathcal{L})^{-1}$. Since

$$\mathcal{I}(a) = \lim_{k \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \mathcal{E}_\varepsilon(a, C_k(a))$$

for $a \in D(\mathcal{E})_+ \cap \mathcal{M}$, it is sufficient to prove that $a \mapsto \mathcal{E}_\varepsilon(a, C_k(a))$ is convex for all $\varepsilon > 0$ and $k \in \mathbb{N}$.

Let $(\mathcal{H}_\varepsilon, \partial_\varepsilon, L_\varepsilon, R_\varepsilon, J_\varepsilon)$ be the first-order differential calculus associated with $\mathcal{E}_\varepsilon$. By [CS03, Section 10.3] there exists $\eta_\varepsilon \in \mathcal{H}_\varepsilon$ such that $\partial_\varepsilon a = (L_\varepsilon(a) - R_\varepsilon(a))\eta_\varepsilon$ for all $a \in D(\mathcal{E})_+ \cap \mathcal{M}$. Thus

$$\mathcal{E}_\varepsilon(a, C_k(a)) = \langle (L_\varepsilon(a) - R_\varepsilon(a))\eta_\varepsilon, (C_k(L_\varepsilon(a)) - C_k(R_\varepsilon(a)))\eta_\varepsilon \rangle_{\mathcal{H}_\varepsilon}.$$ 

For $\alpha > 0$ let

$$\Phi_\alpha : (0, \infty)^2 \rightarrow \mathbb{R}, (s, t) \mapsto \frac{s^{\alpha+1}t^{\alpha} + t - s^{\alpha}t^{1-\alpha} - t}{\alpha}.$$ 

Note that $\Phi_\alpha(s, t) \rightarrow (s - t)(\log s - \log t)$ as $\alpha \searrow 0$. By dominated convergence we have

$$\Phi_\alpha(L_\varepsilon(a) + e^{-k}, R_\varepsilon(a) + e^{-k}) \rightarrow (L_\varepsilon(a) - R_\varepsilon(a))(C_k(L_\varepsilon(a)) - C_k(R_\varepsilon(a))).$$

as $\alpha \searrow 0$ in the weak operator topology. Hence it suffices to show that

$$D(\mathcal{E})_+ \cap \mathcal{M} \rightarrow B(\mathcal{H}), a \mapsto \Phi_\alpha(L_\varepsilon(a) + e^{-k}, R_\varepsilon(a) + e^{-k})$$

is convex for all $k \in \mathbb{N}$ and $\alpha, \varepsilon > 0$. 

\[ \square \]

**Proposition 5.21.** The Fisher information is convex and weakly lower semicontinuous on $L^1_+ ( \mathcal{M}, \tau ).$
Let \( l(a) = L_e(a) + e^{-k} \) and \( r(a) = R_e(a) + e^{-k} \). Since the images of \( L_e \) and \( R_e \) commute, we have

\[
\Phi_\alpha(l(a), r(a)) = \frac{1}{\alpha}(l(a)(l(a)\#\alpha r(a))^{-1}l(a) + r(a) - l(a)\#\alpha r(a) - l(a)),
\]
where \( \#\alpha \) is the operator mean with generating function \( t \mapsto t^\alpha \). As mentioned before, the mean \( \#\alpha \) is jointly operator concave. Together with the joint operator convexity of the map \( (x, y) \mapsto xy^{-1} \) this implies the convexity of \( a \mapsto l(a)(l(a)\#\alpha r(a))^{-1}l(a) \). Hence \( a \mapsto \Phi_\alpha(l(a), r(a)) \) is convex as the sum of convex maps.

Finally, the weak lower semicontinuity follows from the convexity and the strong lower semicontinuity by the Hahn-Banach theorem.

\[ \square \]

**Proposition 5.22.** Assume that \( \theta \) is the logarithmic mean. If \( \rho \in D(M, \tau) \cap L^2(M, \tau) \) and

\[
\int_s^t \mathcal{I}(P_t \rho) \, dr < \infty
\]

for all \( s, t > 0 \), then the curve \( (P_t \rho)_{t>0} \) is admissible and

\[
\|D(P_t \rho)\|^2_{P_t \rho} \leq \mathcal{I}(P_t \rho).
\]

**Proof.** Let

\[
C_n: (e^{-n}, \infty) \to [0, \infty), \ t \mapsto \log(t + e^{-n}) + n.
\]

Let \( \rho_t = P_t \rho \) and \( \xi_t^n = -\partial(C_n(\rho_t)) \). Since \( \rho \in L^2(M, \tau) \), we have \( \rho_t \in D(L) \). As \( C_n \) is continuously differentiable on a neighborhood of \( \sigma(\rho_t) \), we have \( \partial(C_n(\rho_t)) = \dot{C}_n(L(\rho_t), R(\rho_t)) \partial \rho_t \).

Denote by \( e \) the joint spectral measure of \( L(\rho_t) \) and \( R(\rho_t) \). An application of \( 0 \leq \dot{C}_n \cdot LM \leq 1 \) gives

\[
\|\xi_t^n\|^2_{\rho_t} = \|LM^{1/2} \dot{C}_n)(L(\rho_t), R(\rho_t)) \partial \rho_t\|^2_H
\]

\[
= \int_{[0, \infty)^2} LM(s, t) \dot{C}_n(s, t)^2 \, d\langle e(s, t) \partial \rho_t, \partial \rho_t \rangle_H
\]

\[
\leq \int_{[0, \infty)^2} \dot{C}_n(s, t) \, d\langle e(s, t) \partial \rho_t, \partial \rho_t \rangle_H
\]

\[
= \mathcal{E}(\rho_t, C_n(\rho_t))
\]

\[
\leq \mathcal{I}(\rho_t).
\]

On the other hand, \( \dot{C}_n \cdot LM \not\to 1 \) implies

\[
\langle \xi_t^n, \partial a \rangle_{\rho_t} = -\int_{[0, \infty)^2} \dot{C}_n(s, t)LM(s, t) \, d\langle e(s, t) \partial \rho_t, \partial a \rangle_H
\]

\[
\to -\langle \partial \rho_t, \partial a \rangle_H
\]

\[
= \langle -L(\rho_t, a) \rangle_{L^2(M, \tau)}
\]

\[
= \tau(a \rho_t)
\]

for all \( a \in A_\theta \).

Let \( \xi_t^n \) be the projection of \( \xi_t^n \) onto \( H_{\rho_t} \). From the computations above we conclude that \( (\xi_t^n) \) converges weakly to some \( \xi_t \in H_{\rho_t} \) with \( \|\xi_t\|^2_{\rho_t} \leq \mathcal{I}(\rho_t) \) for a.e. \( t > 0 \) and

\[
\langle \xi_t, \partial a \rangle_{\rho_t} = \tau(a \rho_t)
\]

for all \( a \in A_\theta. \)
Proposition 5.23 (Fisher information equals entropy dissipation). Assume that $\tau$ is finite. If $\rho \in \mathcal{D}(\mathcal{M}, \tau) \cap L^2(\mathcal{M}, \tau)$, then

$$\text{Ent}(\rho) - \text{Ent}(P_t\rho) = \int_0^t I(P_s\rho) \, ds$$

for all $t \geq 0$. In particular, Ent is decreasing along $(P_t\rho)_{t \geq 0}$.

Proof. Let $C_n = (\log n) \vee (-n) + n$ and

$$f_n : [0, \infty) \to \mathbb{R}, f_n(t) = -e^{-t} + \int_{e^{-t}}^t (C_n(r) + 1 - n) \, dr.$$  

Then $f_n \in C^1((-e^{-n}, \infty))$, $f_n(t) \nearrow t \log t$ for all $t \geq 0$, and $f_n' = C_n - n + 1$ is bounded and Lipschitz. Define furthermore

$$F_n : [0, \infty) \to \mathbb{R}, s \mapsto \tau(f_n(P_s\rho)).$$

Since $s \mapsto P_s\rho$ is $L^1$-differentiable, we can apply Corollary 5.10 to see that $F_n$ is locally absolutely continuous and $F_n'(s) = -\tau(f_n'(P_s\rho)\mathcal{L}P_s\rho)$.

Thus

$$F_n(\rho) - F_n(P_t\rho) = \int_0^t \tau(f_n'(P_s\rho)\mathcal{L}P_s\rho) \, s \, ds = \int_0^t \mathcal{E}(C_n(P_s\rho), P_s\rho) \, ds \to \int_0^t I(P_s\rho) \, ds, n \to \infty.$$

Here we used conservativeness for the fact that $\tau(\mathcal{L}\bar{\rho}) = 0$ for all $\bar{\rho} \in L^1(\mathcal{M}, \tau)$.

Since $\rho \in \mathcal{D}(\mathcal{M}, \tau) \cap L^2(\mathcal{M}, \tau) \subset \mathcal{D}(\text{Ent})$, the monotone convergence theorem gives the convergence of the left-hand side to $\text{Ent}(\rho) - \text{Ent}(P_t\rho)$. \hfill \square

Corollary 5.24. Assume that $\theta$ is the logarithmic mean. If $\rho \in \mathcal{D}(\text{Ent})$, then

$$\int_0^\infty I(P_t\rho) \, dr \leq \text{Ent}(\rho),$$

$(P_t\rho)_{t \geq 0}$ is an admissible curve, Ent is decreasing along $(P_t\rho)_{t \geq 0}$ and $\mathcal{W}(\rho, P_t\rho) \to 0$ as $t \to 0$.

Proof. Let $\rho_n = \frac{\rho \wedge n}{\tau(\rho \wedge n)}$. Since $\rho_n \to \rho$ in $L^1(\mathcal{M}, \tau)$, one has $P_t\rho_n \to P_t\rho$ in $L^1(\mathcal{M}, \tau)$ for all $t \geq 0$. Moreover,

$$\text{Ent}(\rho_n) = \tau(\rho_n \log \rho_n) = \frac{1}{\tau(\rho \wedge n)} \tau((\rho \wedge n) \log(\rho \wedge n)) - \log \tau(\rho \wedge n) \to \text{Ent}(\rho).$$

It follows from the lower semicontinuity of the entropy and Proposition 5.23 that

$$\text{Ent}(P_t\rho) \leq \liminf_{n \to \infty} \text{Ent}(P_t\rho_n) \leq \lim_{n \to \infty} \text{Ent}(\rho_n) = \text{Ent}(\rho)$$

for all $t \geq 0$. Thus Ent is decreasing along $(P_t\rho)_{t \geq 0}$.

The lower semicontinuity of $I$ (Lemma 5.20) and Fatou's lemma imply

$$\int_0^\infty I(P_t\rho) \, dr \leq \int_0^\infty \liminf_{n \to \infty} I(P_t\rho_n) \, dr \leq \liminf_{n \to \infty} \int_0^\infty I(P_t\rho_n) \, dr.$$  

From Proposition 5.23 we deduce

$$\liminf_{n \to \infty} \int_0^\infty I(P_t\rho_n) \, dr \leq \lim_{n \to \infty} \text{Ent}(\rho_n) = \text{Ent}(\rho).$$
Moreover,
\[
\liminf_{n \to \infty} \int_0^\infty \|D(P_t \rho_n)\|^2_{P_t \rho_n} \, dr \leq \liminf_{n \to \infty} \int_0^\infty T(P_t \rho_n) \, dr
\]
by Proposition 5.22. Thus \((P_t \rho)_{t \geq 0}\) is admissible by Theorem 4.30.

Finally, \(W(\rho, P_t \rho) \to 0\) as \(t \to 0\) is a direct consequence of the admissibility of \((P_t \rho)_{t \geq 0}\).

6. Gradient flow of the entropy

In this section we give the announced characterization of the quantum Markov semigroup as metric gradient flow of the entropy under suitable conditions.

To be more precise, in Subsection 6.1 we first introduce a Bakry–Émery-type gradient estimate for the semigroup \((P_t)\) and prove that it implies a kind of Feller regularization for the semigroup (Proposition 6.7) as well as contraction/expansion estimates with respect to \(W\) (Theorem 6.13).

A central difficulty in the proof of the gradient flow characterization, as already in the case of metric measure spaces, lies in the fact that we are working on \(L^1\), so that the heat flow curves may fail to be differentiable and Hilbert space methods are not directly applicable. To overcome this problem, Subsection 6.2 is devoted to a fine analysis of standard semigroup mollification in our setting. In particular, we prove an entropy regularization estimate in Proposition 6.21.

In the last Subsection 6.3 we review the evolution variational inequality (EVI) formulation of gradient flows in metric spaces and complete the proof of the characterization of the Markovian quantum master equation as EVI gradient flow of the entropy (Theorem 6.26).

In part our proof strategy is a careful adaptation to the noncommutative setting of the paths taken by Ambrosio et al. (see [AGS15] in the case of infinitesimally Hilbertian metric measure spaces and [AES16] in the case of abstract local Dirichlet forms). However, both proofs rely strongly on duality (either the dual problem of the Monge–Kantorovich or the Benamou–Brenier formulation). Little is known on the dual formulation in the present setting, so we avoid it altogether. In this way, our approach gives a new proof variant even when restricted to the case of infinitesimally Hilbertian metric measure spaces (with finite measure).

As usual, let \((\mathcal{M}, \tau)\) be a tracial von Neumann algebra, \(\mathcal{E}\) a quantum Dirichlet form on \(L^2(\mathcal{M}, \tau)\) such that \(\tau\) is energy dominant, and \((\partial, \mathcal{H}, L, R, J)\) the associated first order differential calculus. We further assume that \(\tau\) is a state and \(L^1(\mathcal{M}, \tau)\) is separable. In the beginning we only assume that \(\theta\) can be represented by a symmetric operator mean, but starting from Subsection 6.2 we take the logarithmic mean for \(\theta\). We do not make any density assumptions on \(A_\theta\) – these follow automatically from the gradient estimate we introduce in the first subsection.

6.1. The gradient estimate \(GE(K, \infty)\). In this subsection we introduce the gradient estimate
\[
\|\partial P_t a\|^2 \leq e^{-2Kt}\|\partial a\|^2_{P_t \rho},
\]
which is inspired by the classical Bakry–Émery estimate (see the original paper [BE85] or the monograph [BGL14])
\[
\Gamma(P_t u) \leq e^{-2Kt} P_t \Gamma(u).
\]
In the case of the heat semigroup on a complete Riemannian manifold, the Bakry–Émery gradient estimate is equivalent to a lower bound on the Ricci curvature. Recently, it has been shown that this equivalence between gradient estimates and lower Ricci curvature bounds (in the sense of Lott–Villani–Sturm) extends to a
large class of metric measure spaces [AGS14b, AGS15, EKS15]. Thus the gradient estimate from above can be interpreted as the noncommutative version of a lower Ricci curvature bound.

One of the reasons the gradient estimate is important for the gradient flow characterization is that it provides certain regularizing effects of the semigroup. In this subsection we investigate some of these consequences of the gradient estimates, among them an $L^\infty$-to-$A_\theta$-regularization property of the semigroup (Proposition 6.7) and an exponential contraction (or expansion) bound for the metric $W$ (Theorem 6.13).

Throughout this subsection let $(\mathcal{M}, \tau)$ be a tracial von Neumann algebra, $\mathcal{E}$ a quantum Dirichlet form on $L^2(\mathcal{M}, \tau)$, $(\partial, \mathcal{H}, L, R, J)$ the associated first-order differential calculus and assume that $\tau$ is energy dominant. We further assume that $\theta$ can be represented by a symmetric operator mean. Denote by $(P_t)_{t \geq 0}$ the quantum Markov semigroup associated with $\mathcal{E}$ and by $\mathcal{L} = \partial^* \partial$ its generator.

**Definition 6.1 (GE($K, \infty$)).** Let $K \in \mathbb{R}$. The quantum Dirichlet form $\mathcal{E}$ satisfies the gradient estimate $GE(K, \infty)$ (for the mean $\theta$) if

$$\|\partial P_t a\|_\rho^2 \leq e^{-2Kt}\|\partial a\|_{P_t \rho}^2$$

for all $a \in D(\mathcal{E})$, $\rho \in D(\mathcal{M}, \tau)$ and $t \geq 0$.

**Remark 6.2.** The gradient estimate $GE(K, \infty)$ is a modification of the classical Bakry–Émery gradient estimate. Indeed, if $\mathcal{E}$ is a strongly local (commutative) Dirichlet form on $L^2(X, m)$, then the gradient estimate $GE(K, \infty)$ reads

$$\int_X \Gamma(P_t u) \rho \, dm \leq e^{-2Kt} \int_X P_t \Gamma(u) \rho \, dm,$$

which is just a weak formulation of the Bakry–Émery gradient estimate

$$\Gamma(P_t u) \leq e^{-2Kt} P_t \Gamma(u).$$

If $\mathcal{E}$ is the standard Dirichlet energy on a complete Riemannian manifold $(M, g)$, then $\mathcal{E}$ satisfies $GE(K, \infty)$ if and only if the Ricci curvature of $(M, g)$ is bounded below by $K$.

**Remark 6.3.** In the classical Bakry–Émery gradient estimate there is an additional dimension parameter $N$, and $BE(K, \infty)$ corresponds to the case $N = \infty$. That is why we keep the parameter $\infty$ in our notation although we do not introduce any finite-dimensional variant of $GE(K, \infty)$.

**Remark 6.4.** For finite graphs, the gradient estimate $GE(K, \infty)$ was introduced in [EF18], where it was shown to be equivalent to $K$-convexity of the entropy along $\mathcal{W}$-geodesics ([EF18, Theorem 3.1]). The latter was taken as definition for a lower Ricci curvature bound $K$ of a graph in [EMT14] (see also [Mie13]).

Some examples of quantum Dirichlet forms satisfying the gradient estimate will be studied more systematically in upcoming work. Here we just give a simple (commutative) example.

**Example 6.5.** Endow $\mathbb{Z}^d$ with the natural graph structure given by

$$b(m, n) = \begin{cases} 1 & \text{if } |m - n| = 1, \\ 0 & \text{otherwise}, \end{cases}$$

and the uniform measure $m = 1$.

Since $(\mathbb{Z}^d, b, m)$ has bounded degree, the Dirichlet forms $\mathcal{E}^{(D)}$ and $\mathcal{E}^{(N)}$ introduced in Example 1.13 coincide, and we simply denote them by $\mathcal{E}$. We will show that $\mathcal{E}$
satisfies $\text{GE}(0, \infty)$ for the logarithmic mean. By a simple approximation argument it suffices to prove
\[
\|\partial P_t u\|_\rho^2 \leq \|\partial u\|_{P_t\rho}^2
\]
for $u \in C_c(\mathbb{Z}^d)$ and $\rho \in C_c(\mathbb{Z}^d)_+$. For this, we will approximate $\mathbb{Z}^d$ by finite graphs and use known results for finite Markov chains. Let $K_j = \{-j, \ldots, j\}^d$ and
\[
L_j: \ell^2(K_j) \rightarrow \ell^2(K_j), L_j u(m) = \sum_{n \in K_j} b(m, n) (u(m) - u(n)).
\]
This is the generator of the Dirichlet form $\mathcal{E}_j$ associated with the weighted graph $(K_j, b|_{K_j \times K_j}, 1)$. Let $(P^j_t)$ denote the associated semigroup. It follows from [FM16, Theorem 4.1] in combination with [EM14, Theorem 6.2] or [Mic13, Theorem 5.1] that $\mathcal{E}_j$ satisfies $\text{GE}(0, \infty)$ for the logarithmic mean.

We extend $L_j$ to an operator on $\ell^2(\mathbb{Z}^d)$ by setting it zero on $\ell^2(\mathbb{Z}^d \setminus K_j)$. Clearly, if $u \in C_c(\mathbb{Z}^d)$ with $\text{supp} \, u \subset K_j$, then $L_{j+1} u = L u$, where $L$ is the generator of $\mathcal{E}$. Since the sequence $(L_j)$ is uniformly bounded, it follows that $L_j \rightarrow L$ strongly. Thus also $e^{-t L_j} \rightarrow e^{-t L}$ strongly.

Since $\mathcal{E}_j$ satisfies $\text{GE}(0, \infty)$, we have
\[
\frac{1}{2} \sum_{m,n \in K_j, m \overset{c}{=} n} \rho(m, n) (P^j_t u(m) - P^j_t u(n))^2 \leq \frac{1}{2} \sum_{m,n \in K_j, m \overset{c}{=} n} \overline{P^j_t \rho}(m, n) (u(m) - u(n))^2
\]
for all $u \in C_c(\mathbb{Z}^d)$, $\rho \in C_c(\mathbb{Z}^d)_+$ and $j \in \mathbb{N}$ such that $\text{supp} \, u, \text{supp} \, \rho \subset K_j$. By Fatou’s lemma, the limes inferior as $j \rightarrow \infty$ of the left-hand side is bounded below by $\|\partial P_t u\|_\rho^2$, while the right-hand side converges to $\|\partial u\|_{P_t\rho}^2$ by the dominated convergence theorem. Hence
\[
\|\partial P_t u\|_\rho^2 \leq \|\partial u\|_{P_t\rho}^2.
\]

More generally, this method deducing a gradient estimate for a graph from gradient estimates for an exhaustion of finite subgraphs can be applied if the Laplacians of the subgraphs converge to the Laplacian of the original graph in the strong resolvent sense. This is the case for instance if the graph is locally finite and stochastically complete (in this case one can easily adapt the arguments from [KL12, Proposition 2.7]).

**Remark 6.6.** If $\theta$ is the arithmetic mean, then $\text{GE}(K, \infty)$ reads
\[
\Gamma(P_t a) \leq e^{-2Kt} P_t \Gamma(a)
\]
for $a \in D(\mathcal{E})$ self-adjoint. This noncommutative form of the Bakry–Émery gradient estimate was used for example in [JZ15] in the study of noncommutative Poincaré inequalities.

In general, the gradient estimate $\text{GE}(K, \infty)$ is not equivalent to the (noncommutative) Bakry–Émery gradient estimate in this form, even in the simplest (commutative) examples. For instance, if $\mathcal{E}$ is the Dirichlet form associated with the weighted graph $(\{0, 1\}, b, m)$ with $b(0, 1) > 0$, then one can show that the best possible constant $K$ in $\text{GE}(K, \infty)$ for the logarithmic mean coincides with the best possible constant for the arithmetic mean if and only if $m(0) = m(1)$.

**Proposition 6.7** (Feller property). If $\mathcal{E}$ satisfies $\text{GE}(K, \infty)$ for some $K \in \mathbb{R}$, then $P_t$ maps $L^2(\mathcal{M}, \tau) \cap \mathcal{M}$ into $\mathcal{A}_\theta$ for $t > 0$.
Proof. Let $a \in L^2(M, \tau) \cap M_h$ and $t > 0$. Since $P_t$ maps $L^2(M, \tau)$ into $D(L)$, we can assume $a \in D(L) \cap M$. For $\rho \in D(L) \cap L^1(M, \tau) \cap M$ define

$$\varphi : [0, t] \to \mathbb{R}, \varphi(s) = \int_0^s \|\partial P_{t-r}a\|_{P_r}^2 dr.$$ 

Note that $r \mapsto \hat{P}_r \rho$ is strongly continuous by Lemma 3.14 and $r \mapsto \partial P_{t-r}a$ is continuous in $H$, so that the integrand is continuous in $r$.

Since $E$ satisfies $\text{GE}(K, \infty)$, the map $s \mapsto e^{-2Ks} \varphi'(s)$ is increasing. It follows from a comparison argument (see [AGS15, Lemma 2.2 and Equation (2.30)]) that

$$I_{2K}(t) \|\partial P_t a\|_{\rho}^2 = I_{2K}(t) \varphi'(0) \leq \int_0^t \varphi'(s) ds = \int_0^t \|\partial P_{t-s}a\|_{P_{s-r}}^2 ds,$$

where $I_{s}(t) = \int_0^s e^{\alpha s} ds$.

By Lemma 3.24 and a direct calculation we have

$$\|\partial P_{t-s}a\|_{P_{s-r}}^2 \leq \tau(\Gamma(P_{t-s}a) P_{s-r}) = \frac{1}{2} \frac{d}{ds} \tau((P_{t-s}a)^2 P_{s-r}).$$

If we plug this into (6.1), we get

$$I_{2K}(t) \|\partial P_t a\|_{\rho}^2 \leq \frac{1}{2} \tau((P_t(a^2) - (P_t a)^2) P_{s-r}).$$

Thus

$$\|\partial P_t a\|_{\rho}^2 \leq \frac{1}{2I_{2K}(t)} \|a\|_{\text{Ht}}^2 \|\rho\|_1.$$ 

In the general case $\rho \in L^1(M, \tau)$, one can use Lemma 3.3 and once more $\text{GE}(K, \infty)$ to also get

$$\|\partial P_t a\|_{\rho}^2 = \lim_{m \to \infty} \|\partial P_t a\|_{\rho \wedge m}^2 \leq \liminf_{m \to \infty} \liminf_{\delta \to 0} e^{-2K\delta} \|\partial P_{t-\delta} a\|_{P_{\delta}(\rho \wedge m)}^2 \leq \liminf_{m \to \infty} \liminf_{\delta \to 0} \frac{e^{-2K\delta}}{2I_{2K}(t - \delta)} \|a\|_{\text{Ht}}^2 \|\rho \wedge m\|_1 = \frac{1}{2I_{2K}(t)} \|a\|_{\text{Ht}}^2 \|\rho\|_1.$$ 

Hence $P_t a \in A_{\theta}$. 

\[\square\]

Remark 6.8. If $E$ is a strongly local (commutative) Dirichlet form, then Proposition 6.7 recovers the $L^\infty$-to-Lipschitz Feller property of the heat flow on RCD spaces from [AGS15] Theorem 6.8] (with essentially the same proof). If $E$ is not strongly local or not commutative, the correct analogue of the algebra of bounded Lipschitz functions seems to be $A_{AM}$ (see Example 3.8), so that Proposition 6.7 is in general weaker than (noncommutative) $L^\infty$-to-Lipschitz regularization unless $\theta$ is the arithmetic mean.

Corollary 6.9. If $E$ satisfies $\text{GE}(K, \infty)$, then $A_{\theta} \cap M_1$ is dense in $D(E) \cap M_1$ with respect to $\|\cdot\|_E$ and strongly dense in $M_1$.

Proof. The density of $A_{\theta} \cap M_1$ in $D(E) \cap M_1$ follows directly from Proposition 6.7.

The strong density in $M_1$ is then a consequence of Kaplansky’s density theorem. 

We now fulfill the promise made in Remark 4.9 and introduce a regularity property that ensures that the duality in the definition of admissible curves can be extended to $A_{\theta}$.

Definition 6.10 (Regular mean). The mean $\theta$ is called regular for $E$ if for all $a \in A_{\theta}$ there exists a sequence $(a_n)$ in $A_{AM}$ such that $a_n \to a$ $\sigma$-weakly, $(a_n)$ is bounded in $A_{\theta}$ and $\partial a_n \to \partial a$ weakly in $H_\rho$ for all $\rho \in D(M, \tau)$. 

The next lemma then follows immediately from this definition.

**Lemma 6.11.** If the mean $\theta$ is regular for $E$, then $\partial A_\theta \subset H_\rho$ for all $\rho \in D(M, \tau)$ and if $(\rho_i)_{i \in I}$ is an admissible curve in $D(M, \tau)$, then

$$\tau(a(\rho_t - \rho_s)) = \int_s^t \langle \partial a, D\rho_r \rangle_{\rho_r} \, dr$$

for all $a \in A_\theta$ and $s, t \in I$.

Of course, the arithmetic mean is regular for every quantum Dirichlet form. For other means it may not be easy to check whether it is regular for a given quantum Dirichlet form. However, the gradient estimate $GE(K, \infty)$ yields a sufficient condition for regularity of means. More precisely, we have the following result.

**Proposition 6.12.** Let $\theta$ be a mean and $K \in \mathbb{R}$. If $E$ satisfies $GE(K, \infty)$ for $\theta$ and the arithmetic mean, then $\theta$ is regular for $E$.

**Proof.** For $a \in A_\theta$ let $a_n = P_{t_n} a$ for some null sequence $(t_n)$. Since $E$ satisfies $GE(K, \infty)$ for AM, we have $a_n \in A_{AM}$ by Proposition 6.7.

The $\sigma$-weak convergence of $(a_n)$ to $a$ is a consequence of the continuity of $(P_t)$. Moreover, $GE(K, \infty)$ for $\theta$ implies

$$||\partial a_n||_{\rho} \leq e^{-Kt_n}||\partial a||_{P_{t_n}} \leq \max \{1, e^{-K\sup_n t_n}\} ||a||_{\theta}$$

for all $\rho \in D(M, \tau)$, where the norms $||\cdot||_{\rho}$ and $||\cdot||_{P_{t_n}}$ are to be understood with respect to the mean $\theta$. Hence $(a_n)$ is bounded in $A_\theta$.

Since $||\partial a_n||_{\rho} + ||\partial a_n||_{H}$ is bounded and $(D(\hat{\rho}^{1/2}), \langle \cdot, \cdot \rangle_H + \langle \cdot, \cdot \rangle_{\rho})$ is complete, every subsequence of $(\partial a_n)$ has a subsequence that converges weakly in $D(\hat{\rho}^{1/2})$. Moreover, since $\partial a_n \to \partial a$ in $H$, the weak limit is $\partial a$. Finally, since $D(\hat{\rho}^{1/2})$ is dense in $H_\rho$ and $||\partial a_n||_{\rho}$ is bounded, we also have $\partial a_n \to \partial a$ weakly in $H_\rho$. \hfill $\square$

**Theorem 6.13** (Contraction estimate). Assume that $\theta$ is regular for $E$. If $E$ satisfies $GE(K, \infty)$ for $\theta$ and $(\rho_i)_{i \in I}$ is an admissible curve in $D(M, \tau)$, then $(P_T \rho_i)_{i \in I}$ is an admissible curve and $\|DP_T \rho_i\|_{P_{T \rho_i}} \leq e^{-KT}\|D\rho_i\|_{\rho_i}$ for a.e. $t \in I$.

In particular, $\mathcal{W}(P_T \rho_0, P_T \rho_1) \leq e^{-KT} \mathcal{W}(\rho_0, \rho_1)$ for $\rho_0, \rho_1 \in D(M, \tau)$ and $T \geq 0$.

**Proof.** Let $(\rho_i)_{i \in I}$ be an admissible curve in $D(M, \tau)$. For all $s, t \in I$ and $a \in A_{AM}$ we have

$$|\tau(a(P_T \rho_t - P_T \rho_s))| = |\tau(P_T a(\rho_t - \rho_s))|$$

$$\leq \int_s^t ||\partial P_T a||_{\rho_r} \|D\rho_r\|_{\rho_r} \, dr$$

$$\leq e^{-KT} \int_s^t ||\partial a||_{P_{T \rho_r}} \|D\rho_r\|_{\rho_r} \, dr,$$

where we used Proposition 6.7 and Lemma 6.11 for the first and $GE(K, \infty)$ for the second inequality.

Thus, $(P_T \rho_r)_{r \in [0,1]}$ is an admissible curve with $\|DP_T \rho_r\|_{P_{T \rho_r}} \leq e^{-KT}\|D\rho_r\|_{\rho_r}$ for a.e. $r \in I$. Minimizing over all admissible curves connecting $\rho_0$ and $\rho_1$ yields the second claim. \hfill $\square$

**Corollary 6.14.** Assume that the logarithmic mean is regular for $E$ and $E$ satisfies $GE(K, \infty)$ for the logarithmic mean. If $\rho \in D(\text{Ent})^W$, then $\mathcal{W}(P_t \rho, \rho) \to 0$ as $t \to 0$.

**Proof.** In Corollary 5.23 we have already seen that this convergence holds for $\rho \in D(\text{Ent})$. If $\rho \in D(\text{Ent})^W$, let $(\rho_k)$ be a sequence in $D(\text{Ent})$ such that $\mathcal{W}(\rho_k, \rho) \to 0$. By Theorem 6.13 we have

$$\mathcal{W}(P_t \rho, \rho) \leq (1 + e^{-KT})\mathcal{W}(\rho, \rho_k) + \mathcal{W}(P_t \rho_k, \rho_k).$$
Letting first $t \to 0$ and then $k \to \infty$ yields the claimed convergence. □

6.2. Mollification. A central difficulty in the proof of the gradient flow characterization, as already in the case of metric measure spaces, lies in the fact that we are working on $L^1$, so that the heat flow curves may fail to be differentiable and Hilbert space methods are not directly applicable. To overcome this problem, Section 6.2 is devoted to a fine analysis of semigroup mollification techniques in our setting. In particular, we prove an entropy regularization estimate in Proposition 6.21 that will be crucial in the proof of the main theorem.

As usual, let $(\mathcal{M}, \tau)$ be a tracial von Neumann algebra, $\mathcal{E}$ a quantum Dirichlet form on $L^2(\mathcal{M}, \tau)$ such that $\tau$ is energy dominant, and $(\partial, \mathcal{H}, L, R, J)$ the associated first-order differential calculus. We further assume that $\tau$ is a state, $L^1(\mathcal{M}, \tau)$ is separable and $\theta$ is the logarithmic mean.

We first introduce a mollified version of $(P_t)$, which is a standard tool in the theory of operator semigroups. Let $\kappa \in C_0^\infty((0, \infty))$ be a positive function with support in $(1, 2)$ and $\int_0^\infty \kappa(r) \, dr = 1$. For $\varepsilon > 0$ and $p \in [1, \infty]$ define

$$p^\varepsilon : L^p(\mathcal{M}, \tau) \to L^p(\mathcal{M}, \tau), \quad p^\varepsilon a = \frac{1}{\varepsilon^2} \int_0^\infty \kappa \left( \frac{r}{\varepsilon} \right) P_r a \, dr,$$

where the integral is to be understood as Bochner integral if $p < \infty$ and as Pettis integral for the $\sigma$-weak topology if $p = \infty$. It is clear that $p^\varepsilon$ is positive and contractive on all $L^p$ spaces.

The following Lemma is standard, see for example the proof of [EN00, Proposition 1.8].

**Lemma 6.15.** Let $p \in [1, \infty]$ and $a \in L^p(\mathcal{M}, \tau)$. For all $\varepsilon > 0$ one has $p^\varepsilon a \in D(L^{(p)})$ and

$$L^{(p)} p^\varepsilon a = \frac{1}{\varepsilon^2} \int_0^\infty \kappa \left( \frac{r}{\varepsilon} \right) P_r a \, dr.$$

If $p < \infty$, then $p^\varepsilon a \to a$ in $L^p(\mathcal{M}, \tau)$, and if $p = \infty$, then $p^\varepsilon a \to a$ $\sigma$-weakly as $\varepsilon \to 0$.

If $a \in D(\mathcal{E})$, then $p^\varepsilon a \in D(\mathcal{E})$ and $p^\varepsilon a \to a$ with respect to $\|\cdot\|_{\mathcal{E}}$ as $\varepsilon \to 0$.

**Lemma 6.16.** Assume that $\mathcal{E}$ satisfies $\text{GE}(K, \infty)$ for some $K \in \mathbb{R}$. If $a \in \mathcal{M}$, then $p^\varepsilon a \in \mathcal{A}_{LM}$ for $\varepsilon > 0$. Moreover, if $a \in \mathcal{A}_{AM}$, then

$$\|\partial p^\varepsilon a\|_\rho^2 \leq \|\partial a\|_\rho^2 \sup_{r \in [0, 2\varepsilon]} e^{-2Kr}$$

and $p^\varepsilon a \to a$ in $\mathcal{H}_\rho$ for all $\rho \in D(\mathcal{M}, \tau)$.

**Proof.** Let $\rho \in D(\mathcal{M}, \tau) \cap \mathcal{M}$ be invertible. By $\text{GE}(K, \infty)$ we have

$$\|\partial P_r a\|_\rho^2 \leq e^{-2K(r-\varepsilon)} \|P_{\varepsilon} a\|_{P_{\varepsilon} - \varepsilon}^2$$

whenever $r \geq \varepsilon$. The right-hand side is bounded above by $e^{-2K(r-\varepsilon)} \|P_{\varepsilon} a\|_{LM}^2$, which is finite by Proposition 6.7. By Lemma 3.3 the map $a \mapsto \|\partial a\|_\rho^2$ is lower semicontinuous. Thus we can apply Jensen’s inequality (compare the proof of Lemma 4.33) to see that

$$\|\partial p^\varepsilon a\|_\rho^2 \leq \frac{1}{\varepsilon^2} \int_0^{2\varepsilon} \kappa \left( \frac{r}{\varepsilon} \right) \|\partial P_r a\|_\rho^2 \, dr \leq \frac{\|P_{\varepsilon} a\|_{LM}^2}{\varepsilon} \int_0^\infty e^{-2K(r-\varepsilon)} \kappa \left( \frac{r}{\varepsilon} \right) \, dr.$$

The right-hand side is clearly bounded independently of $\rho$. Hence $p^\varepsilon a \in \mathcal{A}_{LM}$.

Now let $\rho \in D(\mathcal{M}, \tau)$ be arbitrary and assume that $a \in \mathcal{A}_{AM}$. By $\text{GE}(K, \infty)$ we have

$$\|\partial p^\varepsilon a\|_\rho^2 \leq \frac{1}{\varepsilon} \int_0^\infty \kappa \left( \frac{r}{\varepsilon} \right) \|\partial P_r a\|_\rho^2 \, dr \leq \frac{1}{\varepsilon} \int_0^\infty \kappa \left( \frac{r}{\varepsilon} \right) e^{-2Kr} \|\partial a\|_{P_{\varepsilon} - \varepsilon}^2 \, dr.$$
It follows from the upper semicontinuity and concavity of \( \rho \mapsto \| \partial a \|_\rho^2 \) by an application of Jensen’s inequality that

\[
\| \partial p^e a \|_\rho^2 \leq \left( \sup_{r \in [0, 2\varepsilon]} e^{-2Kr} \right) \frac{1}{\varepsilon} \int_0^\infty \kappa \left( \frac{r}{\varepsilon} \right) \| \partial a \|_{\rho, r}^2 \, dr \leq \| \partial a \|_{p^e, r}^2 \sup_{r \in [0, 2\varepsilon]} e^{-2Kr}.
\]

Hence

\[
\limsup_{\varepsilon \to 0} \| \partial p^e a \|_\rho^2 \leq \limsup_{\varepsilon \to 0} \| \partial a \|_{p^e, r}^2 \leq \| \partial a \|_\rho^2
\]

by Theorem 3.15.

Thus \((\partial p^e a)_{\varepsilon > 0}\) is a bounded net in the Hilbert space \(D(\rho^{1/2})\) with inner product

\[
\langle \cdot, \cdot \rangle_H + \langle \rho^{1/2} \cdot, \rho^{1/2} \cdot \rangle_H.
\]

On the other hand, since \(\partial p^e a \to \partial a\) in \(H\), every weak limit point of \((p^e a)_{\varepsilon > 0}\) in \(D(\rho^{1/2})\) coincides with \(\partial a\). Thus \(\partial p^e a \to \partial a\) also weakly in \(H_\rho\). Finally, (6.2) implies that the convergence is indeed strong.

**Lemma 6.17.** Let \(\rho \in D(M, \tau)\). For all \(\varepsilon > 0\) one has \(\text{Ent}(p^e(\rho)) \leq \text{Ent}(P_\varepsilon \rho)\). If \(\rho \in D(\text{Ent})\), then \(p^e(\rho) \in D(\text{Ent})\) and \(\mathcal{W}(\rho, p^e(\rho)) \to 0\) as \(\varepsilon \to 0\).

**Proof.** Since the entropy is a convex lower semicontinuous functional and decreasing along heat flow trajectories, Jensen’s inequality implies

\[
\text{Ent}(p^e(\rho)) \leq \frac{1}{\varepsilon} \int_\varepsilon^{2\varepsilon} \kappa \left( \frac{r}{\varepsilon} \right) \text{Ent}(P_r \rho) \, dr \leq \text{Ent}(P_\varepsilon \rho).
\]

If \(\rho \in D(\text{Ent})\), then \((P_t \rho)_{t \geq 0}\) is admissible by Corollary 5.24.

Let

\[
\sigma : [0, 1] \longrightarrow D(M, \tau), \quad \sigma_t = \frac{1}{\varepsilon} \int_0^\infty \kappa \left( \frac{r}{\varepsilon} \right) P_{rt} \rho \, dr
\]

and

\[
\sigma^n : [0, 1] \longrightarrow D(M, \tau), \quad \sigma^n_t = \sum_{k=1}^{2n} P_{\frac{k-1}{n} \varepsilon t} \rho \cdot \int_{\frac{k}{n}}^{\frac{k+1}{n}} \kappa(r) \, dr.
\]

Since \((P_{rt} \rho)_{r \geq 0}\) is \(L^1\) continuous, we have \(\sigma^n_t \to \sigma_t\) in \(L^1\) for all \(t \in [0, 1]\). As \(\sigma^n\) is a convex combination of admissible curves, it is itself admissible and

\[
\| D_t \sigma^n_t \|_{\sigma^n_t}^2 \leq \sum_{k=1}^{2n} \| D_t P_{\frac{k-1}{n} \varepsilon t} \rho \|_{p^e, r}^2 \cdot \int_{\frac{k-1}{n}}^{\frac{k+1}{n}} \kappa(r) \, dr
\]

by Lemma 4.24.

It follows from Theorem 4.30 and Corollary 5.24 that \(\sigma\) is an admissible curve connecting \(\rho\) and \(p^e(\rho)\) with

\[
\int_0^1 \| D \sigma_t \|_{\sigma_t}^2 \, dt \leq \liminf_{n \to \infty} \sum_{k=1}^{2n} \int_0^1 \| D_t P_{\frac{k-1}{n} \varepsilon t} \rho \|_{p^e, r}^2 \, dt \cdot \int_{\frac{k-1}{n}}^{\frac{k+1}{n}} \kappa(r) \, dr
\]

\[
\leq 2\varepsilon \int_0^{2\varepsilon} \| DP_t \rho \|_{p^e, r}^2 \, dt \cdot \int_0^2 \kappa(r) \, dr
\]

\[
\to 0
\]

as \(\varepsilon \to 0\). \(\square\)

**Corollary 6.18.** The space \(D(L^{(1)}) \cap D(\text{Ent})\) is dense in \((D(\text{Ent}), \mathcal{W})\).
Lemma 6.19. Assume that the logarithmic mean is regular for $E$ and $E$ satisfies GE($K, \infty$). If $(\rho_t)_{t \in [0,1]}$ is an admissible curve in $D(M, \tau)$, then $(p^\tau\rho_t)_{t \in [0,1]}$ is admissible for all $\varepsilon > 0$ and
\[
\limsup_{\varepsilon \to 0} \int_0^1 \| D_p^\tau(\rho_t) \|^2_{p(\rho_t)} dr \leq \int_0^1 \| D\rho_t \|^2_{\rho_t} dr,
\]
\[
\limsup_{\varepsilon \to 0} \esssup_{t \in [0,1]} \| D_p^\tau\rho_t \|^2_{p(\rho_t)} \leq \esssup_{t \in [0,1]} \| D\rho_t \|_{\rho_t}.
\]

If additionally $(\rho_t) \in C^1([0,1]; L^1(M, \tau))$, then $(p^\tau\rho_t)_{t \in [0,1]; L^1(M, \tau))$ and \( \frac{d}{dt} p^\tau\rho_t = p^\tau\rho_t \).

Proof. Let $\rho_s^\varepsilon = p^\varepsilon\rho_s$. For $a \in A_{AM}$ and $s, t \in [0,1]$ we have
\[
|\tau(a(\rho_s^\varepsilon - \rho_t^\varepsilon))| \leq \frac{1}{\varepsilon} \int_0^\infty \kappa \left( \frac{r}{\varepsilon} \right) |\tau(P_\tau a(\rho_t - \rho_s))| dr
\]
\[
\leq \frac{1}{\varepsilon} \int_0^t \kappa \left( \frac{r}{\varepsilon} \right) \int_s^t \| P_\tau a \|_{P_s} \| D\rho_u \|_{\rho_u} du dr
\]
\[
\leq \int_s^t \| D\rho_u \|_{\rho_u} \frac{1}{2\varepsilon} \int_0^\varepsilon e^{-Kr} \kappa \left( \frac{r}{\varepsilon} \right) \| a \|_{P_s} \rho_u dr du,
\]
where we used Proposition [6.7] and Lemma [6.11] in the second and GE($K, \infty$) in the third inequality. Let $C(\varepsilon) = \sup_{t \in [0,2\varepsilon]} e^{-Kr}$ and note that $C(\varepsilon) \to 1$ as $\varepsilon \to 0$. An application of Jensen’s inequality yields
\[
\frac{1}{\varepsilon} \int_0^\infty \kappa \left( \frac{r}{\varepsilon} \right) \| \rho_u \|_{P_s} \rho_u dr \leq \left( \frac{1}{\varepsilon} \int_0^\infty \kappa \left( \frac{r}{\varepsilon} \right) \| \rho_u \|_{P_s}^{2} dr \right)^{1/2} \leq \| \rho_s \|_{\rho_u}.
\]
Hence
\[
|\tau(a(\rho_s^\varepsilon - \rho_t^\varepsilon))| \leq C(\varepsilon) \int_s^t \| D\rho_u \|_{\rho_u} \| \rho_u \|_{\rho_s} du.
\]
Thus $(\rho_t^\varepsilon)_{t \in [0,1]}$ is admissible with $\| D\rho_t^\varepsilon \|_{p^\tau} \leq C(\varepsilon) \| D\rho_t \|_{\rho_t}$ for a.e. $t \in [0,1]$. This settles both of the claimed inequalities.

Finally, the claim concerning the differentiability follows easily from an application of the dominated convergence theorem.

Lemma 6.20. Assume that the logarithmic mean is regular for $E$ and $E$ satisfies GE($K, \infty$) for some $K \in \mathbb{R}$. For every admissible curve $(\rho_s) \in C^1([0,1]; L^1(M, \tau))$ there exists a sequence $\varepsilon_n \searrow 0$ such that the curves $(\rho_s^{\varepsilon_n})_{s \in [0,1]}$ defined by $\rho_s^{\varepsilon_n} = (1 + 1/n)^{-1} p^\varepsilon a(\rho_s + 1/n)$ satisfy
(a) $(\rho_s^{\varepsilon_n})_{s \in [0,1]} \in C^1([0,1]; L^1(M, \tau))$ for $n \in \mathbb{N}$,
(b) $\rho_s^{\varepsilon_n} \in D(L^{(1)})$ for $s \in [0,1], n \in \mathbb{N}$,
(c) $\rho_s^{\varepsilon_n} \geq \frac{1}{2\varepsilon_n}$ for $s \in [0,1], n \in \mathbb{N}$,
(d) $\rho_s^{\varepsilon_n} \to \rho_s$ in $L^1(M, \tau)$ for $s \in [0,1]$,
(e) $\text{Ent}(\rho_s^{\varepsilon_n}) \leq \text{Ent}(\rho_s)$,
(f) $\limsup_{n \to \infty} \int_0^1 \| D\rho_s^{\varepsilon_n} \|^2_{\rho_s} ds \leq \int_0^1 \| D\rho_s \|^2_{\rho_s} ds$,
\[
\text{lim sup}_{n \to \infty} \text{ess sup}_{s \in [0,1]} \| D\rho_s^{\varepsilon_n} \|^2_{\rho_s} \leq \text{ess sup}_{s \in [0,1]} \| D\rho_s \|^2_{\rho_s}.
\]

Proof. Let $\tilde{\rho}_s^{\varepsilon_n} = (1 + 1/n)^{-1}(\rho_s + 1/n)$. Clearly the curves $(\tilde{\rho}_s^{\varepsilon_n})$ satisfy (a), (c) and (d). Property (e) follows from the convexity of Ent. Moreover,
\[
\tau(a(\tilde{\rho}_s^{\varepsilon_n} - \rho_s^{\varepsilon_n})) = (1 + 1/n)^{-1} \tau(a(\rho_s - \rho_s))
\]
implies $\| D\tilde{\rho}_s^{\varepsilon_n} \|_{\rho_s} = (1 + 1/n)^{-1} \| D\rho_s \|_{\rho_s}$.\]
Now let $\rho^n_s = p^n_s \rho^n_s$ for a strictly positive null sequence $(\varepsilon_n)$. The curve $(\rho^n_s)$ satisfies (a) by Lemma 6.19, (b) by Lemma 6.15 and (c) as a direct consequence of the positivity of $(P_t)$. Moreover,

$$\|\rho^n_s - \rho_s\|_1 \leq \|p^n_s \rho^n_s - p^n_s \rho_s\|_1 + \|p^n_s \rho_s - \rho_s\|_1 \leq \|\rho^n_s - \rho_s\|_1 + \|p^n_s \rho_s - \rho_s\|_1 \to 0.$$  

Thus (d) is satisfied. Property (e) follows from Lemma 6.17. Finally, by Lemma 6.19 we can achieve (f) if we choose $(\varepsilon_n)$ appropriately.

Lemma 6.20 gives entropy estimates at the endpoints of the connecting curves, but we will need entropy estimates for the entire curve. In [AGS15] these are established via a logarithmic Harnack inequality. Since the proof relies on the second-order chain rule for the Laplacian, there seems to be little hope to generalize it beyond the local setting. Indeed, obtaining Harnack inequalities from Bakry–Émery-type Ricci curvature bounds has turned out to be exceptionally challenging in the non-local case. For the gradient estimate used here, there seem to be no results in this direction even in the case of finite graphs (see however [CLY14, BHL+15, DKZ17, Mũ18] for Harnack inequalities on graphs under related assumptions).

Instead we adopt a different approach. The kind of entropy estimate we need is a consequence of the EVI gradient flow characterization (see [DS08, Theorem 3.1]) and it turns out that one can run the portion of the proof needed to show only this consequence with the weaker regularity estimates already established in Lemma 6.20. This is done in the next proposition.

**Proposition 6.21 (Entropy regularization).** Assume that $\mathcal{E}$ satisfies GE$(K, \infty)$. If $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{M}, \tau)$, then

$$\text{Ent}(P_t \rho_1) \leq \text{Ent}(\rho_0) + \frac{1}{2t} \left( \int_0^1 e^{-2Kst} ds \right) \mathcal{W}^2(\rho_0, \rho_1)$$

for $t > 0$.

**Proof.** Let

$$C(K, t) = \int_0^1 e^{-2Kst} ds.$$  

We can assume that $\text{Ent}(\rho_0) < \infty$ and $\mathcal{W}(\rho_0, \rho_1) < \infty$. By Lemmas 4.22 and 4.33 for every $\varepsilon > 0$ there exists an admissible curve $(\rho_s) \in C^1([0, 1]; L^1(\mathcal{M}, \tau))$ such that

$$\text{ess sup}_{s \in [0, 1]} \|D\rho_s\|_{\rho_s}^2 \leq \mathcal{W}(\rho_0, \rho_1)^2 + \varepsilon$$

Let $(\rho^n_s)_{s \in [0, 1]}$, $n \in \mathbb{N}$, be curves defined in Lemma 6.20. If we can show

$$\text{Ent}(P_t \rho^n_1) \leq \text{Ent}(\rho^n_0) + \frac{C(K, t)}{2t} \text{ess sup}_{s \in [0, 1]} \|D\rho^n_s\|_{\rho^n_s}^2,$$  

then the claim of the proposition follows by taking the limit $n \to \infty$ and then the limit $\varepsilon \to 0$.

Let $(C_k)$ be an increasing sequence in $C^1((0, \infty))$ such that each $C_k$ is increasing, 1-Lipschitz, $C_k(s) = s$ if $s \leq k - 1$ and $C_k(s) = k$ if $s \geq n$. Let

$$f_k = C_k \circ \log$$

$$F_k : (0, \infty) \to \mathbb{R}, t \mapsto \int_0^t (f_k(s) + 1) \, ds.$$

Note that $f_k(s) \nrightarrow \log s$ and by monotone convergence also $F_k(t) \nrightarrow t \log t$.

Thus, in order to prove (6.3), it suffices to show

$$\tau(F_k(P_t \rho_1)) \leq \tau(F_k(\rho_0)) + \frac{C(K, t)}{2t} \int_0^1 \|D\rho^n_s\|_{\rho^n_s}^2 \, ds.$$
Let \( \sigma_s = P_s \rho^s_n \). Since \( \rho^s_n \in D(L^{(1)}) \) for all \( s \in [0,1] \), the curve \( (\sigma_s) \) is \( L^1 \)-differentiable with derivative

\[
\dot{\sigma}_s = P_s \dot{\rho}_s^n - tL^{(1)}(\sigma_s).
\]

Hence, by Lemma 5.10

\[
(6.5) \quad \tau(F_k(P_t\rho_1) - F_k(\rho_0)) = \int_0^1 \tau(f_k(\sigma_s)(P_s \dot{\rho}_s^n - tL^{(1)}(\sigma_s))) \, ds.
\]

Since \( (P_s) \) maps \( M \) into \( A_{LM} \) by Proposition 6.7 we have

\[
|\tau(f_k(\sigma_s)P_s \dot{\rho}_s^n)| \leq \|P_s \dot{f}_k(\sigma_s)\|_{\rho^s_n} \|D\rho^s_n\|_{\rho^s_n} \\
\leq e^{-Kst}\|\dot{f}_k(\sigma_s)\|_{\sigma_s} \|D\rho^s_n\|_{\rho^s_n} \\
\leq e^{-2Kst} \frac{2t}{t} \|D\rho^s_n\|_{\rho^s_n}^2 + \frac{t}{2} \|\dot{f}_k(\sigma_s)\|_{\sigma_s}^2,
\]

where we used the admissibility of \( (\rho^s_n) \) and Lemma 6.11 for the first inequality, \( GE(K,\infty) \) for the second and Young’s inequality for the third.

We are now going to estimate the second summand. Note that by definition \( 0 \leq f_k(s) \leq 1/s \) and for each fixed \( k \in \mathbb{N} \) there exists \( l \in \mathbb{N} \) such that \( f_k(s) = f_k(s \wedge l) \) for all \( s > 0 \).

Since \( \sigma_s \in D(L^{(1)}) \), we have \( \sigma_s \wedge m \in D(E) \) for all \( m \in \mathbb{N} \) by Lemma 5.15. Thus, writing \( e \) for the joint spectral measure of \( L(\sigma_s \wedge m) \) and \( R(\sigma_s \wedge m) \),

\[
\|\dot{f}_k(\sigma_s)\|_{\sigma_s \wedge m}^2 = \|\dot{f}_k(\sigma_s \wedge m)\|_{\sigma_s \wedge m}^2 \\
= \int_{(0,\infty)^2} \tilde{f}_k(s,t)\log(s,t) \, d(e(s,t)\partial(\sigma_s \wedge m),\partial(\sigma_s \wedge m))_H \\
\leq \sum_k (\tilde{f}_k(L(\sigma_s \wedge m),R(\sigma_s \wedge m)))\partial(\sigma_s \wedge m)\partial(\sigma_s \wedge m))_H \\
= (\dot{f}_k(\sigma_s),\partial(\sigma_s \wedge m))_H \\
\leq \tau(f_k(\sigma_s)L^{(1)}(\sigma_s)),
\]

where we used \( \tilde{f}_k \leq \log^2 \) and Lemma 5.15. Now we can let \( m \) go to infinity to obtain

\[
\|\dot{f}_k(\sigma_s)\|_{\sigma_s}^2 \leq \tau(f_k(\sigma_s)L^{(1)}(\sigma_s)).
\]

If we plug this inequality into (6.5), we get

\[
|\tau(f_k(\sigma_s)P_s \dot{\rho}_s^n)| \leq \frac{e^{-2Kst}}{2t} \|D\rho^s_n\|_{\rho^s_n}^2 + \frac{t}{2} \tau(f_k(\sigma_s)L^{(1)}(\sigma_s)),
\]

which we can then apply to (6.5) to obtain

\[
(6.7) \quad \tau(F_k(P_t\rho_1) - F_k(\rho_0)) \leq \int_0^1 \frac{e^{-2Kst}}{2t} \|D\rho^s_n\|_{\rho^s_n}^2 \, ds \\
\leq \frac{C(K,t)}{2t} \sup_s \|D\rho^s_n\|_{\rho^s_n}^2.
\]

This settles 6.2.4. \qed

**Corollary 6.22.** Assume that \( E \) satisfies \( GE(K,\infty) \) for some \( K \in \mathbb{R} \). For every admissible curve \( (\rho_s)_{s \in [0,1]} \in C^1([0,1]; L^1(\mathcal{M},\tau)) \) with \( \text{Ent}(\rho_0) < \infty \) the curves \( (\rho^s_n)_{s \in [0,1]} \) defined in Lemma 6.20 satisfy

\[\begin{align*}
&\text{g) } \sup_{s \in [0,1]} \text{Ent}(\rho^s_n) < \infty \text{ for } n \in \mathbb{N}, \\
&\text{h) } \sup_{t \geq 0} \sup_{s \in [0,1]} I(P_t \rho^s_n) < \infty \text{ for } n \in \mathbb{N}.
\end{align*}\]
Proof. As in the proof of Lemma 6.20 let \( \tilde{\rho}_s^n = (1 + 1/n)^{-1}(\rho_s + 1/n) \) and \( \rho_s^n = \rho^n \tilde{\rho}_s^n \) for a suitably chosen strictly positive null sequence \((\varepsilon_n)\). By Lemma 6.17 we have \( \text{Ent}(\rho_s^n) \leq \text{Ent}(\rho_s^n) \). We can apply Theorem 6.21 to the right-hand side to get

\[
\text{Ent}(P_{\varepsilon_n}, \rho_s^n) \leq \text{Ent}(\rho_s^n) + C(K, \varepsilon_n) W(\rho_s^n, \rho^n) \]

\[
\leq \frac{n}{n + 1} \left( \text{Ent}(\rho_0) + C(K, \varepsilon_n) \int_0^1 \|D\rho_r\|^2 dr \right).
\]

The latter is clearly bounded independently of \( s \in [0, 1] \). This proves (g).

To establish (h), we use Jensen’s inequality (which is applicable according to Lemma 5.20 and Proposition 5.21) to see that

\[
\mathcal{I}(P_t, \rho_s^n) \leq \frac{1}{\varepsilon_n} \int_{\varepsilon_n}^\infty \kappa \left( \frac{r}{\varepsilon_n} \right) \mathcal{I}(P_{t+r}, \rho_s^n) dr \leq \frac{\|\kappa\|\infty}{\varepsilon_n} \int_{\varepsilon_n}^\infty \mathcal{I}(P_t, P_t \rho_s^n) dr.
\]

By Corollary 5.21 we have

\[
\int_{\varepsilon_n}^\infty \mathcal{I}(P_t, P_t \rho_s^n) dr \leq \text{Ent}(P_{t+\varepsilon_n}, \rho_s^n) \leq \text{Ent}(P_{\varepsilon_n}, \rho_s^n).
\]

We have already seen in the first part that the right-hand side is bounded independently of \( s \in [0, 1] \). \( \square \)

6.3. Proof of the gradient flow characterization. With preparations done in the previous subsection, we can now launch the proof of the gradient flow characterization. We start with some basic facts from the theory of gradient flows in metric spaces.

Recall that we mean by an extended metric space a pair \((X, d)\) that satisfies all axioms of a metric space, except that \( d \) may take the value \( \infty \). Let \( S: X \rightarrow (-\infty, \infty] \) be a proper lower semicontinuous functional and let \( D(S) \) denote its proper domain, that is,

\[
D(S) = \{x \in X \mid S(x) < \infty\}.
\]

The descending and ascending slope of \( S \) are defined by

\[
|D^+ S|(x) = \limsup_{y \rightarrow x} \frac{(S(y) - S(x))^+}{d(x, y)}
\]

\[
|D^- S|(x) = \limsup_{y \rightarrow x} \frac{(S(y) - S(x))^-}{d(x, y)}
\]

if \( x \in D(S) \) is not isolated. For isolated points \( x \in D(S) \) one sets

\[
|D^+ S|(x) = |D^- S|(x) = 0,
\]

and furthermore \( |D^+ S| = |D^- S| = \infty \) on \( X \setminus D(S) \).

As a further piece of notation we need the upper right derivative (or upper Dini derivative) \( \frac{d^+}{dt} \) of a function \( f \) on a right-open interval \( I \), which is defined by

\[
\frac{d^+}{dt} f(t) = \limsup_{h \searrow 0} \frac{f(t + h) - f(t)}{h}
\]

for \( t \in I \).

Definition 6.23 (EDE and EVI gradient flow curves). Let \((X, d)\) be an extended metric space and \( S: X \rightarrow (-\infty, \infty] \) a proper lower semicontinuous functional. A locally absolutely continuous curve \((\gamma_t)_{t \geq 0} \) in \( X \) is called EDE gradient flow of \( S \) if it satisfies the energy dissipation equality

\[
(S(\gamma_0) = S(\gamma_t) + \frac{1}{2} \int_0^t |\dot{\gamma}_s|^2 ds + \frac{1}{2} \int_0^t |D^- S|^2(\gamma_s) ds)
\]

(EDE)
for all $t \geq 0$.

Let $K \in \mathbb{R}$. The curve $\gamma$ is called EVI$_K$ gradient flow curve of $S$ if it satisfies the evolution variational inequality

$$(6.8) \quad \frac{1}{2} \frac{d^+}{dt} d(\gamma_t, x)^2 + \frac{K}{2} d(\gamma_t, x)^2 + S(\gamma_t) \leq S(x)$$

for all $t \geq 0$ and $x \in X$ with $d(x, \gamma_0) < \infty$.

A semigroup of continuous maps $T_t : D(S) \to D(S)$, $t \geq 0$, is called EVI$_K$ gradient flow of $S$ if

1. $d(T_t x, x) \to 0$ as $t \to 0$ for all $x \in X$,
2. $S$ is decreasing along $(T_t x)_{t \geq 0}$ for all $t \geq 0$,
3. $(T_t x)_{t \geq 0}$ is an EVI$_K$ gradient flow for all $x \in X$.

**Remark 6.24.** If $\mathbb{R}^d$ is endowed with the Euclidean metric and $S \in C^1(\mathbb{R}^d)$, then a $C^1$-curve $\gamma$ is an EDE gradient flow curve of $S$ if and only if it satisfies the classical gradient flow equality

$$\dot{\gamma}_t = -\nabla S(\gamma_t).$$

If moreover $S - \frac{K}{2} |\cdot|^2$ is convex, then the notion of EDE gradient flow curve, EVI$_K$ gradient flow curve and classical gradient flow curve all coincide for $C^1$-curves. Conversely, if $S$ admits an EVI$_K$ gradient flow, then $S - \frac{K}{2} |\cdot|^2$ is convex.

While the existence of EVI$_K$ gradient flow curves for a given functional is not guaranteed, the uniqueness is a consequence of the defining property (see e.g. [DS08, Proposition 3.1]):

**Lemma 6.25.** Let $(X, d)$ be an extended metric space and $S : X \to (-\infty, \infty]$ a proper lower semicontinuous functional. If $\gamma$, $\tilde{\gamma}$ are EVI$_K$ gradient flow curves of $S$ starting in $\gamma_0$, $\tilde{\gamma}_0$ respectively, then

$$d(\gamma_t, \tilde{\gamma}_t) \leq e^{-Kt} d(\gamma_0, \tilde{\gamma}_0)$$

for all $t \geq 0$.

In particular, there is at most one EVI$_K$ gradient flow curve with a given starting point.

**Theorem 6.26.** Assume that $\tau$ is finite, $L^1(\mathcal{M}, \tau)$ is separable and the logarithmic mean is regular with respect to $E$. If $E$ satisfies GE($K, \infty$), then $(P_t)$ is an EVI$_K$ gradient flow of Ent.

**Proof.** The continuity of $(P_t)$ with respect to $\mathcal{W}$ is a consequence of Theorem 6.13. Moreover, it was proven in Corollary 5.24 that $(P_t)_{t \geq 0}$ is strongly continuous with respect to $\mathcal{W}$ and that Ent is decreasing along $(P_t \rho)$. It remains to show that

$$(6.8) \quad \frac{1}{2} \frac{d^+}{dt} \mathcal{W}(P_t \rho_1, \rho_0)^2 + \frac{K}{2} \mathcal{W}(\rho_0, \rho_1)^2 + \text{Ent}(P_t \rho_1) \leq \text{Ent}(\rho_0)$$

for $\rho_0, \rho_1 \in D(\text{Ent})$ with $\mathcal{W}(\rho_0, \rho_1) < \infty$ and $t \geq 0$. Since $(P_t)$ is a semigroup, it suffices to check (6.8) at $t = 0$.

If we can prove

$$(6.9) \quad \mathcal{W}(P_t \rho_1, \rho_0)^2 \leq \left( \int_0^1 e^{-2Ks} \right) \mathcal{W}(\rho_0, \rho_1)^2 - 2t(\text{Ent}(P_t \rho_1) - \text{Ent}(\rho_0)),$$

then

$$\frac{\mathcal{W}(P_t \rho_1)^2 - \mathcal{W}(\rho_0, \rho_1)^2}{2t} \leq \text{Ent}(\rho_0) - \text{Ent}(P_t \rho_1) + \frac{1}{2} \int_0^1 e^{-2Ks} ds - \frac{1}{2} \mathcal{W}(\rho_0, \rho_1)^2,$$

from which (6.8) at $t = 0$ follows in the limit $t \to 0$. 

In order to prove \([6.9]\), let \(\varepsilon > 0\) and let \((\rho_s) \in C^1([0, 1]; L^1(\mathcal{M}, \tau))\) be an admissible curve with

\[
\text{ess sup}_{s \in [0, 1]} \| D\rho_s \|_{\rho_s} \leq W(\rho_0, \rho_1) + \varepsilon.
\]

Let \((\rho^n_s)_{s \in [0, 1]}\) be the curves defined in Lemma 6.20 and let \(s, t = P_{st} \rho^n_s\). Since \(\rho^n_s \in D(\mathcal{L}^{(1)})\), for each \(t \geq 0\) the curve \((\sigma_{s,t})_{s \in [0, 1]}\) is \(L^1\)-differentiable with derivative

\[
(6.10) \quad \frac{d}{ds} \sigma_{s,t} = P_{st} \rho^n_s - t \mathcal{L}^{(1)} \sigma_{s,t}.
\]

We will now show that the curve \((\sigma_{s,t})_{s \in [0, 1]}\) is admissible by evaluating both summands separately.

Since \(\mathcal{E}\) satisfies \(\text{GE}(K, \infty)\), the logarithmic mean is regular for \(\mathcal{E}\) and \((\rho^n_s)\) is admissible, Proposition 6.11 and Lemma 6.21 imply

\[
|\tau(aP_{st} \rho^n_s)| = |\tau(\rho^n_s P_{st} a)| \leq \| D_{\mathcal{E}} \|_{\rho^n_s} \| D\rho^n_s \|_{\rho^n_s} \leq e^{-Kst} \| \partial a \|_{\sigma_{s,t}} \| D\rho^n_s \|_{\rho^n_s}
\]

for all \(a \in D(\mathcal{E}) \cap \mathcal{M}\) with \(\partial a \in D(\sigma_{s,t}^{1/2})\). Thus there exists a unique \(\xi_s\) in \(\mathcal{H}_{\sigma_{s,t}}\) such that

\[
(6.11) \quad \tau(aP_{st} \rho^n_s) = \langle \partial a, \xi_s \rangle_{\sigma_{s,t}}
\]

for all \(a \in D(\mathcal{E}) \cap \mathcal{M}\) with \(\partial a \in D(\sigma_{s,t}^{1/2})\). Moreover, \(\| \xi_s \|_{\sigma_s} \leq e^{-Kst} \| D\rho^n_s \|_{\rho^n_s}^2\).

Some more work is necessary for the second summand in \((6.10)\). If \(\sigma_{s,t} \in D(\mathcal{L}^{(2)})\), then

\[
\tau(aL^{(1)} \sigma_{s,t}) = \langle \partial a, \partial \log \sigma_{s,t} \rangle_{\mathcal{H}} = \langle \partial a, \partial \log \sigma_{s,t} \rangle_{\sigma_{s,t}}.
\]

To show the equality of the left- and right-hand side in the general case (the middle is of course not well-defined), we argue by approximation.

Let \(\tilde{\rho}_n = (1 + 1/n)^{-1}(\rho_n + 1/n)\) and recall that \(\rho^n_s = \rho^n_s \in \mathcal{E}\). Moreover, let \(\sigma^n_{s,t} = P_{st} \rho^n_s \in \mathcal{M}\). If \(\sigma_{s,t} \in D(\mathcal{L}^{(2)})\), then

\[
(6.12) \quad \tau(a L^{(1)} \sigma_{s,t}) = \lim_{N \to \infty} \tau((\mathcal{L}^{(2)} \rho^n_s a) P_{st} (\tilde{\rho}_{s,t}^n \land N))
\]

\[
= \lim_{N \to \infty} \langle \partial \rho^n_s a, \partial P_{st} (\tilde{\rho}_{s,t}^n \land N) \rangle_{\mathcal{H}}
\]

\[
= \lim_{N \to \infty} \langle \partial a, \partial \log \sigma^n_{s,t} \rangle_{\sigma^n_{s,t}}.
\]

Since \(\sigma^n_{s,t} \leq \sigma_{s,t}\) and \(\sigma^n_{s,t} \to \sigma_{s,t}\) in \(L^1(\mathcal{M}, \tau)\) as \(N \to \infty\), we have

\[
(6.13) \quad \tilde{\sigma}_{s,t}^{1/2} \partial a \to \sigma_{s,t}^{1/2} \partial a
\]

strongly in \(\mathcal{H}\) as \(N \to \infty\) by Lemma 3.26.

We will now show that \(\tilde{\sigma}_{s,t}^{1/2} \partial \log \sigma^n_{s,t} \to \sigma_{s,t}^{1/2} \partial \log \sigma_{s,t}\) weakly in \(\mathcal{H}\). By Jensen’s inequality and Corollary 5.24,

\[
I \left( \frac{\sigma^n_{s,t}}{(\rho^n_s \land N)} \right) \leq \frac{\| \kappa \|_{\varepsilon_n}}{\varepsilon_n} \int_{\varepsilon_n}^{\infty} I \left( P_{\tau} P_{st} \tilde{\rho}_s^n \land N \right) \, dr
\]

\[
\leq \frac{\| \kappa \|_{\varepsilon_n}}{\varepsilon_n} \text{Ent} \left( P_{\varepsilon_n} \tilde{\rho}_s^n \land N \right)
\]

\[
\leq \frac{\| \kappa \|_{\varepsilon_n}}{\varepsilon_n} \tau((P_{\varepsilon_n} (\tilde{\rho}_s^n \land N) \log P_{\varepsilon_n} (\tilde{\rho}_s^n \land N))_+)
\]

\[
- \frac{\| \kappa \|_{\varepsilon_n}}{\varepsilon_n} \log \tau(\tilde{\rho}_s^n \land N).
\]
For the first summand observe that $t \mapsto (t \log t)_+$ is increasing, which implies by [BK90, Lemma 4] that
\[
\tau((P_{\varepsilon_n} \bar{\rho}_n^a \wedge N) \log P_{\varepsilon_n} (\bar{\rho}_n^a \wedge N))_+ \leq \tau((P_{\varepsilon_n} \bar{\rho}_n^a \log P_{\varepsilon_n} \bar{\rho}_n^a)_+).
\]
The right-hand side is finite by Proposition [6.21]. Since $\tau(\bar{\rho}_n^a) \to 1$, we infer from (6.14) that $\sup_N \mathcal{I}(\sigma_{s,t}^N) < \infty$. The lower bound $\sigma_{s,t}^N \geq 1/n$ then implies
\[
\sup_N \mathcal{E}(\log \sigma_{s,t}^N) \leq n \sup_N \mathcal{I}(\sigma_{s,t}^N) < \infty.
\]

From $\sigma_{s,t}^N \to \sigma_{s,t}$ in $L^1(\mathcal{M}, \tau)$ we infer $\log \sigma_{s,t}^N \to \log \sigma_{s,t}$ in $L^2(\mathcal{M}, \tau)$ by [Tik87, Theorem 3.2]. Together with the bound on the energy this implies $\log \sigma_{s,t}^N \to \log \sigma_{s,t}$ weakly in $(D(\mathcal{E}), \langle \cdot, \cdot \rangle_{\mathcal{E}})$.

If $\xi \in D(\partial^{1/2}_{s,t})$, then
\[
\langle \sigma_{s,t}^{N \frac{1}{2}}, \partial \log \sigma_{s,t}^{N \frac{1}{2}}, \xi \rangle_{\mathcal{H}} = \langle \partial \log \sigma_{s,t}^{N \frac{1}{2}}, \sigma_{s,t}^{N \frac{1}{2}}, \xi \rangle_{\mathcal{H}} \to \langle \partial \log \sigma_{s,t}, \sigma_{s,t}^{1/2}, \xi \rangle_{\mathcal{H}}
\]
as $N \to \infty$ by Lemma [3.20]. Since $D(\sigma_{s,t}^{1/2})$ is dense in $\mathcal{H}$ and $\sup_N \mathcal{I}(\sigma_{s,t}^N) < \infty$, this implies
\[
\sigma_{s,t}^{N \frac{1}{2}} \partial \log \sigma_{s,t}^{N \frac{1}{2}} \to \sigma_{s,t}^{1/2} \partial \log \sigma_{s,t}
\]
weakly in $\mathcal{H}$ as $N \to \infty$.

If we combine this convergence with (6.13), then we can deduce from (6.12) that
\[
\tau(a \mathcal{L}^{(1)}(\sigma_{s,t})) = \langle \partial a, \partial \log \sigma_{s,t} \rangle_{\sigma_{s,t}}
\]
for all $a \in D(\mathcal{E}) \cap \mathcal{M}$ with $\partial a \in D(\sigma_{s,t}^{1/2})$.

Let $\eta_{s,t} = \partial \log \sigma_{s,t}$. If we combine the results (6.11) and (6.15), we see that $(\sigma_{s,t})_{s \in [0,1]}$ is admissible and
\[
\| D_s \sigma_{s,t} \|_{\sigma_{s,t}}^2 \leq \| \xi_{s,t} - t \eta_{s,t} \|_{\sigma_{s,t}}^2.
\]

Thus
\[
\int_0^1 \| D_s \sigma_{s,t} \|_{\sigma_{s,t}}^2 \, ds \leq \int_0^1 \| \xi_{s,t} - t \eta_{s,t} \|_{\sigma_{s,t}}^2 \, ds
\]
\[
= \int_0^1 (\| \xi_{s,t} \|_{\sigma_{s,t}}^2 - 2 t \langle \eta_{s,t}, \xi_{s,t} - t \eta_{s,t} \rangle_{\sigma_{s,t}} - t^2 \| \eta_{s,t} \|_{\sigma_{s,t}}^2) \, ds
\]
\[
\leq \int_0^1 (\| \xi_{s,t} \|_{\sigma_{s,t}}^2 - 2 t \langle \eta_{s,t}, \xi_{s,t} - t \eta_{s,t} \rangle_{\sigma_{s,t}}) \, ds
\]
\[
\leq \left( \int_0^1 e^{-2Kst} \, ds \right) \text{ess sup} \| D_p \rho_s^a \|_{\rho_s^a}^2
\]
\[
- 2 t \int_0^1 \langle \eta_{s,t}, \xi_{s,t} - t \eta_{s,t} \rangle_{\sigma_{s,t}} \, ds.
\]

Let $(C_k)$ be an increasing sequence in $C^1((0, \infty))$ such that each $C_k$ is increasing, 1-Lipschitz, $C_k(s) = s$ if $s \leq k - 1$ and $C_k(s) = k$ if $s \geq n$. Let $f_k = C_k \circ \log$ and
\[
F_k: (0, \infty) \to \mathbb{R}, \ t \mapsto \int_0^t (f_k(s) + 1) \, ds.
\]

As in the proof of Proposition [6.21] one can show
\[
\tau(F_k(P_t \rho_0^a) - F_k(\rho_0^a)) = \int_0^1 \tau(f_k(\sigma_{s,t})(P_{s,t} \rho_s^a - t \mathcal{L}^{(1)}(\sigma_{s,t}))) \, ds.
\]
Since \( f_k(\sigma_{s,t}) \in D(\mathcal{E}) \cap \mathcal{M} \) by Lemma 5.15 and \( \| \partial f_k(\sigma_{s,t}) \|_{\sigma_{s,t}}^2 \leq I(\sigma_{s,t}) \), we can apply (6.11) and (6.15) to get

\[
(6.17) \quad \tau(F_k(P_t\rho^n_k) - F_k(\rho^n_0)) = \int_0^1 \langle \partial f_k(\sigma_{s,t}), \xi_{s,t} - t\eta_{s,t} \rangle_{\sigma_{s,t}} \, ds
\]

Since \( f_k(\sigma_{s,t}) \to \log \sigma_{s,t} \) in \( L^2(\mathcal{M}, \tau) \) as \( k \to \infty \) and \( \| \partial f_k(\sigma_{s,t}) \|_{\sigma_{s,t}}^2 \leq I(\sigma_{s,t}) \), we have \( \partial f_k(\sigma_{s,t}) \to \partial \log \sigma_{s,t} = \eta_{s,t} \) weakly in \( \mathcal{H}_{\sigma_{s,t}} \) and the integrand on the right-hand side of (6.17) is pointwise bounded by \( I(\sigma_{s,t})^{1/2} \| \xi_{s,t} - t\eta_{s,t} \|_{\sigma_{s,t}} \).

By the dominated convergence theorem,

\[
\lim_{k \to \infty} \int_0^1 \langle \partial f_k(\sigma_{s,t}), \xi_{s,t} - t\eta_{s,t} \rangle_{\sigma_{s,t}} \, ds = \int_0^1 \langle \eta_{s,t}, \xi_{s,t} - t\eta_{s,t} \rangle_{\sigma_{s,t}} \, ds.
\]

On the other hand, the left-hand side of (6.17) converges to \( \text{Ent}(P_t\rho_1) - \text{Ent}(\rho_0) \) by the monotone convergence theorem.

Then (6.16) becomes

\[
(6.18) \quad \int_0^1 \| D_s \sigma_{s,t} \|_{\sigma_{s,t}}^2 \, ds \leq \left( \int_0^1 e^{-2Kst} \, ds \right) \text{ess sup}_{s \in [0,1]} \| D\rho^n_s \|_{\rho^n_s}^2 \]

\[
- 2t(\text{Ent}(P_t\rho^n_1) - \text{Ent}(\rho^n_0)).
\]

Since \( \sigma_{s,t} = P_{st}\rho^n_s \to P_{st}\rho_s \) in \( L^1(\mathcal{M}, \tau) \) as \( n \to \infty \), Theorem 4.30 can be used to see that the curve \( (P_{st}\rho_s)_s \) is admissible and

\[
\mathcal{W}(\rho_0, P_t\rho_1)^2 \leq \int_0^1 \| D_s (P_{st}\rho_s) \|_{P_{st}\rho_s}^2 \leq \liminf_{n \to \infty} \int_0^1 \| D_s (P_{st}\rho^n_s) \|_{P_{st}\rho^n_s}^2 \, ds.
\]

By Lemma 6.20 we have

\[
\limsup_{n \to \infty} \text{ess sup}_{s \in [0,1]} \| D\rho^n_s \|_{\rho^n_s}^2 \leq \text{ess sup}_{s \in [0,1]} \| D\rho_s \|_{\rho_s}^2 \leq (\mathcal{W}(\rho_0, \rho_1) + \varepsilon)^2.
\]

Furthermore, using the lower semicontinuity of the entropy and Lemma 6.20 we obtain \( \text{Ent}(P_t\rho^n_1) \to \text{Ent}(P_t\rho_1) \), \( \text{Ent}(\rho^n_0) \to \text{Ent}(\rho_0) \).

These inequalities allow to pass to the limit \( n \to \infty \) in (6.18) to get

\[
\mathcal{W}(\rho_0, P_t\rho_1)^2 \leq \left( \int_0^1 e^{-2Kst} \, ds \right) (\mathcal{W}(\rho_0, \rho_1) + \varepsilon)^2 - 2t(\text{Ent}(P_t\rho_1) - \text{Ent}(\rho_0)),
\]

which yields (6.9) as \( \varepsilon \to 0 \).

As a consequence of the uniqueness of \( \text{EVI}_K \) gradient flow curves (Lemma 6.25) we note the following corollary.

**Corollary 6.27.** A curve \( (\rho_t)_{t \geq 0} \) with \( \rho_0 \in D(\text{Ent}) \) is an \( \text{EVI}_K \) gradient flow curve of \( \text{Ent} \) if and only if \( \rho_t = P_t\rho_0 \) for all \( t \geq 0 \).

**7. Geodesic Convexity**

In this section we will study an important consequence of the gradient flow characterization, namely the (semi-) convexity of the entropy along geodesics in \( (\mathcal{D}(\mathcal{M}, \tau), \mathcal{W}) \). This property served as definition for synthetic Ricci curvature bounds by Lott–Villani [LV09] and Sturm [Stu06a, Stu06b] and could therefore also be an entrance gate to the study of Ricci curvature in noncommutative geometry.

As discussed in a previous section, even the existence of \( \mathcal{W} \)-geodesics is not clear in general. The situation is much better if \( (P_t) \) satisfies the gradient estimate \( \text{GE}(K, \infty) \) and we restrict our attention to the domain of the entropy (Theorem 7.7). This is due to two ingredients, which we will study next: First, the sublevel sets of the entropy are compact in the weak \( L^1 \)-topology (Lemma 7.1). Together with the lower semicontinuity of the the action functional with respect to pointwise
weak convergence in $L^1$, this can be employed for the standard existence proof of minimizers for a variational problem, provided one can always find a minimizing sequence with uniformly bounded entropy. As we will see, the latter is essentially a consequence of the evolution variational inequality (Proposition 7.3).

Once the existence of geodesics is proven, the semi-convexity of the entropy along them follows from abstract results on gradient flows (Theorem 7.11). Finally, we summarize the relations between the gradient estimate $\mathrm{GE}(K, \infty)$, the evolution variational inequality $\mathrm{EVI}_K$ and $K$-convexity of the entropy in Theorem 7.12.

As usual, let $(\mathcal{M}, \tau)$ be a tracial von Neumann algebra, $\mathcal{E}$ a quantum Dirichlet form on $L^2(\mathcal{M}, \tau)$ such that $\tau$ is energy dominant, and $(\partial, \mathcal{H}, L, R, J)$ the associated first order differential calculus. We further assume that $\tau$ is a state, $L^1(\mathcal{M}, \tau)$ is separable, $\theta$ is the logarithmic mean and $\theta$ is a regular mean for $\mathcal{E}$.

**Lemma 7.1.** If $F : [0, \infty) \to \mathbb{R}$ is a lower semicontinuous function such that $f(t)/t \to \infty$ as $t \to \infty$, then the sublevel sets of

$$F : \mathcal{D}(\mathcal{M}, \tau) \to (-\infty, \infty], \quad F(\rho) = \tau(f(\rho))$$

are compact in the weak $L^1$-topology.

**Proof.** The sublevel sets are closed since $F$ is lower semicontinuous by Lemma 5.6 so it suffices to show that they are relatively weakly compact. We assume that $f \geq 0$; otherwise one can replace it by $f_+$ and use that $f_+ \leq f - \inf f$.

The proof of relative weak compactness is a noncommutative version of the Vallée Poussin theorem (see [Bog07, Theorem 4.5.9]). By the noncommutative version of the Dunford-Pettis theorem ([Hak02, Theorem III.5.4]), it suffices to show that $\tau(p_n \rho) \to 0$ uniformly in $\rho \in \mathcal{F}$ whenever $(p_n)$ is a decreasing sequence of projections in $\mathcal{M}$ such that $p_n \searrow 0$.

Let $C = \sup_{\rho \in \mathcal{F}} \tau(f(\rho))$. For $\varepsilon > 0$ let $M = \frac{2C}{\varepsilon}$. By assumption there exists $T > 0$ such that $f(t) \geq Mt$ for all $t \geq T$. Moreover, since $\tau$ is normal, we can choose $N \in \mathbb{N}$ such that $\tau(p_n) < \frac{\varepsilon}{2T}$ for $n \geq N$.

It follows that

$$\tau(p_n \rho) = \tau(p_n \rho(1_{[0,T]}) + 1_{[T,\infty]}(\rho))) \leq T \tau(p_n) + \frac{\varepsilon}{2T} \tau(p_n f(\rho)) < \varepsilon$$

for all $\rho \in \mathcal{F}$ and $n \geq N$. \hfill $\square$

**Remark 7.2.** By the Eberlein-Šmulian theorem ([DSSS, Theorem V.6.1]), weak compactness and weak sequential compactness are equivalent for weakly closed subsets of a Banach space.

**Proposition 7.3.** For all $K, \alpha, D > 0$ there exists a constant $C(K, \alpha, D) > 0$ such that the following holds:

If the form $\mathcal{E}$ satisfies $\mathrm{GE}(K, \infty)$, $\rho_0, \rho_1 \in D(\mathrm{Ent})$ with $\mathrm{Ent}(\rho_0), \mathrm{Ent}(\rho_1) \leq \alpha$, and $\mathcal{W}(\rho_0, \rho_1) \leq D$, then there is a sequence of admissible curves $(\rho^n_t)_{t \in [0,1]}$ connecting $\rho_0$ and $\rho_1$ such that

$$\sup_{n \in \mathbb{N}} \sup_{t \in [0,1]} \mathrm{Ent}(\rho^n_t) \leq C(K, \alpha, D)$$

and

$$\int_0^1 \|D \rho^n_t\|_{\rho^n_t}^2 \, dt \to \mathcal{W}(\rho_0, \rho_1)^2.$$  

**Proof.** For $n \in \mathbb{N}$ let $(\sigma^n_t)_{t \in [0,1]}$ be an admissible $L^\infty$-Lipschitz curve in $(\mathcal{D}(\mathcal{M}, \tau), \mathcal{W})$ connecting $\rho_0$ and $\rho_1$ such that $L^2_n \leq \mathcal{W}(\rho_0, \rho_1)^2 + \frac{1}{n^2}$.  

Let $\tilde{\sigma}_t^n = P_{t/n} \sigma_t^n$. Since $(P_t)$ is an EVI$_K$ gradient flow of Ent by Theorem 6.26 of [DS08], Theorem 3.2 of [DS08] asserts

$$\text{Ent}(\tilde{\sigma}_t^n) \leq (1-t)\text{Ent}(\rho_0) + t\text{Ent}(\rho_1) - \frac{K}{2} \left(1-t\right) W(\rho_0, \rho_1)^2 + \frac{1}{2n^2 I_K(1/n)},$$

where $I_K(t) = \int_0^t e^{Kr} \, dr$.

As $n^2 I_K(1/n) \to \infty$ as $n \to \infty$, the supremum

$$c(K) = \sup_{n \in \mathbb{N}} \frac{1}{2n^2 I_K(1/n)}$$

is finite. Thus

$$\sup_{n \in \mathbb{N}} \sup_{t \in [0, 1]} \text{Ent}(\tilde{\sigma}_t^n) \leq \alpha + \frac{|K|}{2} D^2 + c(K).$$

Furthermore, Theorem 6.13 implies $\|D\tilde{\sigma}_t^n\|_{\sigma_t^n} \leq e^{-K/n}\|D\sigma_t^n\|_{\sigma_t^n}$ for a.e. $t \in [0, 1]$.

Moreover, $\text{Ent}(P_t \rho_0) \leq \text{Ent}(\rho_0) \leq \alpha$ and $(P_t \rho_0)_{t \geq 0}$ is admissible by Corollary 5.24, hence

$$\int_{0}^{1/n} \|D_{s} P_t \rho_0\|_{P_t \rho_0}^2 \, ds \to 0$$

as $n \to \infty$. Of course, the same holds for $\rho_0$ replaced by $\rho_1$.

Hence one can concatenate the curves $(P_t \rho_0)_{t \in [0, 1/n]}$, $(\tilde{\sigma}_t^n)_{t \in [0, 1]}$ and $(P_{\frac{1}{n} - t} \rho_1)_{t \in [0, 1/n]}$ to get a curve $(\rho_t^n)$ with the desired properties.

**Definition 7.4.** We say that the entropy has **regular sublevel sets** if every curve $(\rho_t) \in AC^2_{\text{loc}} ([0, 1]; (\mathcal{D}(\mathcal{M}, \tau), \mathcal{W}))$ with uniformly bounded entropy is admissible and $\|D\rho_t\|_{\rho_t} = |\rho_t|_{\mathcal{W}}$ for a.e. $t \in I$.

**Proposition 7.5.** If $\mathcal{E}$ satisfies GE$(K, \infty)$, then the entropy has regular sublevel sets.

**Proof.** First assume that $(\rho_t) \in AC^2([0, 1]; (\mathcal{D}(\mathcal{M}, \tau), \mathcal{W}))$. Since $(\rho_t)$ is continuous on a compact interval, it is uniformly continuous. Thus, for every $\varepsilon > 0$ there exists a partition $0 = t_0 < t_1 < \cdots < t_n = 1$ of $[0, 1]$ such that $\mathcal{W}(\rho_{t_{k-1}}, \rho_t) < \varepsilon$ for all $t \in [t_{k-1}, t_k]$, $1 \leq k \leq n$.

For $k \in \{1, \ldots, n\}$ let $\sigma^{k, \varepsilon} : [t_{k-1}, t_k] \to \mathcal{D}(\mathcal{M}, \tau)$ be an admissible curve with $\sigma_{t_{k-1}} = \rho_{t_{k-1}}$, $\sigma_{t_k} = \rho_{t_k}$ and

$$\int_{t_{k-1}}^{t_k} \|D\sigma_{r}^{k, \varepsilon}\|_{\sigma_{r}^{k, \varepsilon}}^2 \, dr \leq \frac{\mathcal{W}(\rho_{t_{k-1}}, \rho_{t_k})^2}{t_k - t_{k-1}} + \varepsilon.$$

Moreover, by Proposition 7.3, the curves $\sigma^{k, \varepsilon}$ can be chosen such that

$$\sup_{\varepsilon \in (0, 1)} \sup_{k \in \mathbb{N}} \sup_{t \in [t_{k-1}, t_k]} \text{Ent}(\sigma_t^{k, \varepsilon}) < \infty.$$
Denote by $\rho^\varepsilon$ the concatenation of $\sigma^{1,\varepsilon}, \ldots, \sigma^{n,\varepsilon}$. Then
\[
\int_0^1 \|D\rho_t^\varepsilon\|_{\rho_t^\varepsilon}^2 \, dr = \sum_{k=1}^n \int_{t_{k-1}}^{t_k} \|D\sigma_t^\varepsilon\|_{\rho_t^\varepsilon}^2 \, dr \\
\leq \varepsilon + \sum_{k=1}^n \frac{\mathcal{W}(\rho_{t_{k-1}}, \rho_{t_k})^2}{t_k - t_{k-1}} \\
\leq \varepsilon + \sum_{k=1}^n \int_{t_{k-1}}^{t_k} |\dot{\rho}_r|_W \, dr \\
= \varepsilon + \int_0^1 |\dot{\rho}_t|_W^2 \, dr.
\]
Moreover, for every $t \in [0,1]$ there is a $k \in \{1, \ldots, n\}$ such that $\mathcal{W}(\rho_{t_{k-1}}, \rho_{t_k}) \leq \varepsilon$, hence
\[
\mathcal{W}(\rho_t^\varepsilon, \rho_t) \leq \mathcal{W}(\rho_t^\varepsilon, \rho_{t_k}) + \mathcal{W}(\rho_{t_k}, \rho_t) < 3\varepsilon.
\]
Thus, $\mathcal{W}(\rho_t^\varepsilon, \rho_t) \to 0$ as $\varepsilon \searrow 0$.

By Lemma 7.1 and the uniform bound on the entropy $\mathcal{L}_1$, for every $t \in [0,1]$ and every sequence $(\varepsilon_n)$ converging to 0 there is a subsequence $(\varepsilon_{n(k)})$ and $\tilde{\rho}_t \in \mathcal{D}(\mathcal{M}, \tau)$ such that $\rho_t^\varepsilon_{n(k)} \to \tilde{\rho}_t$ weakly in $L^1$ as $k \to \infty$. In particular, $\tau(\rho_t^\varepsilon_{n(k)} a) \to \tau(\tilde{\rho}_t a)$ for all $a \in \mathcal{A}_L$.

On the other hand, $\mathcal{W}(\rho_t^\varepsilon_{n(k)}, \rho_t) \to 0$ implies $\tau(\rho_t^\varepsilon_{n(k)} a) \to \tau(\rho_t a)$ for all $a \in \mathcal{A}_L$ by Proposition 4.20. Since $\mathcal{A}_L \subseteq \mathcal{M}$ is $\sigma$-weakly dense by Corollary 6.3, it follows that $\tilde{\rho}_t = \rho_t$ for all $t \in [0,1]$. Therefore, $\rho_t^\varepsilon \to \rho_t$ weakly in $L^1$ as $\varepsilon \searrow 0$ for all $t \in [0,1]$.

By Theorem 4.30, the curve $(\rho_t^\varepsilon)$ is admissible and
\[
\int_0^1 \|D\rho_t^\varepsilon\|_{\rho_t^\varepsilon}^2 \, dr \leq \liminf_{\varepsilon \searrow 0} \int_0^1 \|D\rho_t^\varepsilon\|_{\rho_t^\varepsilon}^2 \, dr \leq \int_0^1 |\dot{\rho}_t|_W^2 \, dr.
\]
As the reverse inequality is obvious, we conclude $\|D\rho_t^\varepsilon\|_{\rho_t^\varepsilon} = |\dot{\rho}_t|_W$ for a.e. $r \in [0,1]$.

In the general case $(\rho_t) \in \mathcal{A}^{2}_{loc}(I; (\mathcal{D}(\mathcal{M}, \tau), \mathcal{W}))$ one can simply partition $I$ into countably many compact intervals to obtain the same result. □

Let $(X, d)$ be an extended metric space. A curve $(\gamma_t)_{t \in [0,1]}$ in $X$ is called (constant speed) geodesic if $d(\gamma_0, \gamma_1) < \infty$ and $d(\gamma_s, \gamma_t) = |s - t|d(\gamma_0, \gamma_1)$ for all $s, t \in [0,1]$. The extended metric space $(X, d)$ is called geodesic space if any two $x, y \in X$ with $d(x, y) < \infty$ can be joined by a geodesic.

**Lemma 7.6.** For $\alpha > 0$ let $S_\alpha = \{ \rho \in \mathcal{D}(\mathcal{M}, \tau) \mid \text{Ent}(\rho) \leq \alpha \}$. If $L > 0$ and $((\rho_{t}^\varepsilon)_{t \in [0,1]})_n$ is a sequence of admissible curves in $S_\alpha$ such that
\[
\int_s^t \|D\rho_t^\varepsilon\|_{\rho_t^\varepsilon}^2 \, dr \leq L^2 |t - s|
\]
for all $s, t \in [0,1]$ and $n \in \mathbb{N}$, then there exists an admissible curve $(\rho_t)$ in $S_\alpha$ and a subsequence $(\rho_{t}^{n_k})_k$ of $(\rho^n)$ such that $\rho_{t}^{n_k} \to \rho_t$ weakly in $L^1$ for all $t \in [0,1]$, and
\[
\int_0^1 \|D\rho_t^\varepsilon\|_{\rho_t^\varepsilon}^2 \, dt \leq \liminf_{n \to \infty} \int_0^1 \|D\rho_t^n\|_{\rho_t^n}^2 \, dt.
\]
Proof. Otherwise passing to a subsequence, we can assume that \(\int_0^1 \|D\rho^a_t\|_{\rho^a_t}^2\,dt\) converges. If \(a \in A_{LM}\), then

\[
|\tau((\rho^n_t - \rho^n_s)a)| \leq \int_s^t \|\partial a\|_{\rho^a_t} \|D\rho^a_t\|_{\rho^a_t} \,dr \leq L\|a\|_{LM}^2|t - s|
\]

for all \(s, t \in [0, 1]\) and \(n \in \mathbb{N}\). Thus \((\rho^n_t)\) is uniformly equicontinuous with respect to the metric

\[
d: \mathcal{D}(\mathcal{M}, \tau) \times \mathcal{D}(\mathcal{M}, \tau) \to [0, \infty), \quad d(\rho, \sigma) = \sup_{\|a\|_{LM} \leq 1} |\tau(a(\rho - \sigma))|.
\]

By Lemma 7.1, the set \(S_a\) is sequentially compact with respect to the weak topology on \(L^1\). Hence we can apply [AGS08] Proposition 3.3.1 (the lower semicontinuity property of \(d\) is obvious, while the completeness of \((S_a, d)\) follows from the weak compactness and the lower semicontinuity) to get a subsequence \((\rho^{n_k})\) and a curve \((\rho_t)\) in \(S_a\) such that \(\rho^{n_k}_t \to \rho_t\) weakly in \(L^1\) for all \(t \in [0, 1]\).

The remaining assertions follow from Theorem 4.30. \(\square\)

Theorem 7.7. If the entropy has regular sublevel sets and \((P_t)\) is an EVI-K gradient flow of \(\text{Ent}\), then for all \(\rho_0, \rho_1 \in D(\text{Ent})\) with \(W(\rho_0, \rho_1) < \infty\) there exists a geodesic \((\rho_t)_{t \in [0, 1]}\) connecting \(\rho_0\) and \(\rho_1\) such that \(L_n^2 \leq e^{-2K/n}(W(\rho_0, \rho_1)^2 + \frac{1}{n})\) and

\[
\sup_{n \in \mathbb{N}} \sup_{t \in [0, 1]} \text{Ent}(\rho^n_t) < \infty.
\]

Since the entropy has regular sublevel sets, the curves \((\rho^n_t)\) are admissible and

\[\|D\rho^n_t\|_{\rho^n_t} = |\rho^n_t|_{W} \leq L_n.\]

As \((L_n)\) is bounded, we can apply Lemma 7.6 to get an admissible curve \((\rho_t)_{t \in [0, 1]}\) with uniformly bounded entropy connecting \(\rho_0\) and \(\rho_1\) such that

\[
\int_0^1 \|D\rho_t\|_{\rho_t}^2\,dt \leq \liminf_{n \to \infty} \int_0^1 \|D\rho^n_t\|_{\rho^n_t}^2\,dt \leq \liminf_{n \to \infty} e^{-2K/n}(W(\rho_0, \rho_1)^2 + \frac{1}{n^2}) = W(\rho_0, \rho_1)^2.
\]

Hence \((\rho_t)\) is a geodesic connecting \(\rho_0\) and \(\rho_1\). \(\square\)

Corollary 7.8. If the entropy has regular sublevel sets and \((P_t)\) is an EVI-K gradient flow of \(\text{Ent}\), then the metric \(W\) is lower semicontinuous with respect to weak \(L^1\)-convergence on sublevel sets of the entropy.

Although Theorem 7.7 guarantees the existence of geodesics connecting density matrices with finite entropy provided their distance is finite, it does not rule out the possibility that density matrices with finite entropy have infinite distance. In the next proposition we will see that this cannot happen if \(\mathcal{E}\) is ergodic and satisfies \(\text{GE}(K, \infty)\) for strictly positive \(K\).

Here \(\mathcal{E}\) is called ergodic if for all \(a \in L^2(\mathcal{M}, \tau)\) one has \(\lim_{t \to \infty} P_t(a) = \tau(a)\) in \(L^2(\mathcal{M}, \tau)\). Equivalently, there are no projections \(p \in \mathcal{M}\) other than 0 and 1 such that \(P_t p \leq p\) for all \(t \geq 0\).
Proposition 7.9. Assume that \( \tau(1) = 1 \). If \( \mathcal{E} \) is ergodic and satisfies \( \text{GE}(K, \infty) \) for some \( K > 0 \), then the Talagrand inequality

\[
\mathcal{W}(\rho, 1)^2 \leq \frac{2}{K} \text{Ent}(\rho)
\]

holds for all \( \rho \in D(\text{Ent}) \). In particular, \( \mathcal{W} \) is finite on \( D(\text{Ent}) \times D(\text{Ent}) \).

Proof. First we show that for \( \rho \in D(\mathcal{M}, \tau) \cap L^2(\mathcal{M}, \tau) \) we have

\[
(7.2) \quad \text{Ent}(P_t \rho) \leq \frac{1}{2K} \mathcal{I}(P_t \rho)
\]

for a.e. \( t \geq 0 \).

Since \( (P_t \rho)_{t \geq 0} \) is an admissible curve with \( \|DP_t \rho\|_{P_t \rho}^2 \leq \mathcal{I}(P_t \rho) \) for a.e. \( t \geq 0 \), we have

\[
\limsup_{h \searrow 0} \frac{1}{h} \mathcal{W}(P_{t+h} \rho, P_t \rho) \leq \limsup_{h \searrow 0} \frac{1}{h} \int_t^{t+h} \|DP_r \rho\|_{P_r \rho} \, dr \leq \mathcal{I}(P_t \rho)^{1/2}
\]

for a.e. \( t \geq 0 \).

Thus

\[
-\frac{1}{2} \frac{d}{dt} \mathcal{W}(P_t \rho, P_{s} \rho)^2 = \limsup_{h \to 0} \frac{1}{2h} (\mathcal{W}(P_t \rho, P_s \rho)^2 - \mathcal{W}(P_{t+h} \rho, P_s \rho)^2)
\]

\[
\leq \limsup_{h \searrow 0} \frac{1}{2h} \left( \mathcal{W}(P_{t+h} \rho, P_t \rho)^2 + 2 \mathcal{W}(P_{t+h} \rho, P_t \rho) \mathcal{W}(P_{t+h} \rho, P_s \rho) \right)
\]

\[
\leq \mathcal{I}(P_t \rho)^{1/2} \mathcal{W}(P_t \rho, P_s \rho)
\]

for a.e. \( t \geq 0 \) and all \( s > 0 \).

The evolution variational inequality from Theorem 6.26 implies

\[
\text{Ent}(P_t \rho) \leq -\frac{1}{2} \frac{d}{dt} \mathcal{W}(P_t \rho, \rho)^2 - \frac{K}{2} \mathcal{W}(P_t \rho, \rho)^2 + \text{Ent}(P_s \rho)
\]

\[
\leq \mathcal{I}(P_t \rho)^{1/2} \mathcal{W}(P_t \rho, \rho) - \frac{K}{2} \mathcal{W}(P_t \rho, \rho)^2 + \text{Ent}(P_s \rho)
\]

\[
\leq \frac{1}{2K} \mathcal{I}(P_t \rho) + \text{Ent}(P_s \rho)
\]

for a.e. \( t \geq 0 \) and all \( s \geq 0 \).

Since \( \mathcal{E} \) is ergodic and convergence in \( L^2 \) implies convergence of the entropy, we have \( \text{Ent}(P_s \rho) \to 0 \) as \( s \to \infty \) and (7.2) follows.

Since \( t \to \text{Ent}(P_t \rho) \) is a locally absolutely continuous function with derivative a.e. equal to \( -\mathcal{I}(P_t \rho) \) by Proposition 5.23, Grönwall’s lemma asserts

\[
\text{Ent}(P_t \rho) \leq e^{-2Kt} \text{Ent}(\rho).
\]

By the same arguments used in the proof of Corollary 5.24, this inequality remains true if we only assume \( \rho \in D(\text{Ent}) \). In particular, \( \text{Ent}(P_t \rho) \to 0 \) as \( t \to \infty \).

If \( \hat{\rho} \) is any limit point in the weak \( L^1 \)-topology of \( (P_t \rho) \) as \( t \to \infty \), then

\[
\text{Ent}(\hat{\rho}) \leq \limsup_{t \to \infty} \text{Ent}(P_t \rho) = 0,
\]

which implies \( \hat{\rho} = 1 \). On the other hand, since the sublevel sets of the entropy are compact in the weak \( L^1 \)-topology by Lemma 7.1, there exist limit points in the weak \( L^1 \)-topology of \( (P_t \rho) \) as \( t \to \infty \). Both facts combined give \( P_t \rho \to 1 \) weakly as \( t \to \infty \).

Once again assume that \( \rho \in D(\mathcal{M}, \tau) \cap L^2(\mathcal{M}, \tau) \) and let \( \vartheta: [0, 1) \to [0, \infty) \) be a strictly increasing differentiable function with \( \lim_{t \to 1} \vartheta(t) = \infty \). Let \( \rho_t = P_{\vartheta(t)} \rho \).
Since \((P_t \rho)_{t \geq 0}\) is admissible and
\[
\int_0^\infty \|DP_t \rho\|_{P_t \rho}^2 dt \leq \int_0^\infty \mathcal{I}(P_t \rho) dt \leq \text{Ent}(\rho) < \infty
\]
by Propositions 5.22 and 5.23, the reparametrized curve \((\rho_t)_{t \in [0,1]}\) is also admissible. Since

\[
-2 \text{Ent}(P_t \rho)^{1/2} \frac{d}{dt} \text{Ent}(P_t \rho)^{1/2} = \mathcal{I}(P_t \rho)
\]

for a.e. \(t \geq 0\) such that \(P_t \rho \neq 1\), we can use the inequality \(\text{Ent}(P_t \rho) \leq \frac{1}{2K} \mathcal{I}(P_t \rho)\) proven above to see that

\[
\mathcal{I}(P_t \rho)^{1/2} \leq -\left(\frac{2}{K}\right)^{1/2} \frac{d}{dt} \text{Ent}(P_t \rho)^{1/2}
\]

for a.e. \(t \geq 0\) (clearly the inequality holds if \(P_t \rho = 1\), since \(\mathcal{I}(1) = 0\)).

Thus

\[
\mathcal{W}(\rho, 1) \leq \int_0^1 \|D \rho_t\|_{\rho_t} dt \\
= \int_0^\infty \|DP_t \rho\|_{P_t \rho} dt \\
\leq \int_0^\infty \mathcal{I}(P_t \rho)^{1/2} dt \\
\leq -\left(\frac{2}{K}\right)^{1/2} \int_0^\infty \frac{d}{dt} \text{Ent}(\rho_t)^{1/2} dt \\
= \left(\frac{2}{K}\right)^{1/2} \text{Ent}(\rho)^{1/2}.
\]

In the general case \(\rho \in D(\text{Ent})\), we can argue by approximation. Let \(\rho^n = \frac{\rho^n \wedge \tau}{\|\rho^n \wedge \tau\|_1}\).

Since \(\text{Ent}(\rho) < \infty\), it is easy to see that the sequence \((\rho^n)\) has bounded entropy. By Proposition 7.3 we can choose admissible curves \((\rho^n_t)_{t \in [0,1]}\) connecting \(\rho^n\) and 1 with uniformly bounded entropy such that

\[
\int_s^t \|D \rho^n_t\|^2_{\rho^n_t} dt \leq \left(\mathcal{W}(\rho^n, 1) + \frac{1}{n^2}\right) |t - s|
\]

for all \(s, t \in [0,1]\) and \(n \in \mathbb{N}\).

It follows from Lemma 7.5 that there exists a strictly increasing sequence \((n_k)\) in \(\mathbb{N}\) and an admissible curve \((\rho_t)_{t \in [0,1]}\) such that \(\rho^{n_k}_t \to \rho_t\) weakly in \(L^1(\mathcal{M}, \tau)\) for all \(t \in [0,1]\) and

\[
\mathcal{W}(\rho, 1)^2 \leq \int_0^1 \|DP_t \rho\|_{P_t \rho}^2 dt \\
\leq \liminf_{n \to \infty} \int_0^1 \|DP_t \rho^n\|_{P_t \rho^n}^2 dt \\
\leq \liminf_{n \to \infty} \mathcal{W}(\rho^n, 1)^2 \\
\leq \frac{2}{K} \lim_{n \to \infty} \text{Ent}(\rho^n) \\
= \frac{2}{K} \text{Ent}(\rho). \qed
The last property discussed in this section is $K$-convexity of the entropy. Let $(X,d)$ be an extended metric space. A functional $S: X \rightarrow (-\infty, \infty]$ is called $K$-convex along the geodesic $(\gamma_t)_{t \in [0,1]}$ in $(D(S),d)$ if

$$S(\gamma_t) \leq (1-t)S(\gamma_0) + tS(\gamma_1) - \frac{K}{2} t(1-t)d(\gamma_0,\gamma_1)^2$$

for all $t \in [0,1]$.

The functional $S$ is called strongly geodesically $K$-convex if it is $K$-convex along every geodesic in $(D(S),d)$. It is called geodesically $K$-convex if every pair $x_0,x_1 \in D(S)$ can be joined by a geodesic $(\gamma_t)$ such that $S$ is $K$-convex along $(\gamma_t)$.

**Remark 7.10.** If $(D(S),d)$ is a geodesic space, then every strongly geodesically $K$-convex functional is geodesically $K$-convex. If $(D(S),d)$ is not a geodesic space, it does not make too much sense to talk about geodesic convexity at all.

If $(P_t)$ is an EVI$^K$ gradient flow of the entropy and the sublevel sets of the entropy are regular, the strong $K$-convexity follows from abstract results on gradient flows in metric spaces.

**Theorem 7.11.** If $(P_t)$ is an EVI$^K$ gradient flow of $\text{Ent}$ and $\text{Ent}$ has regular sublevel sets, then $(D(\text{Ent}),W)$ is a geodesic space and $\text{Ent}$ is strongly geodesically $K$-convex.

**Proof.** By Theorem 7.7, the space $(D(\text{Ent}),W)$ is geodesic, and by [AGS14b, Proposition 2.23] the entropy is $K$-convex along all geodesics in $(D(\text{Ent}),W)$. □

Let us summarize the results of this section.

**Theorem 7.12.** Assume that $\tau$ is a state, $L^1(M,\tau)$ is separable and $\theta$ is the logarithmic mean. For $K \in \mathbb{R}$ consider the following properties.

(i) The semigroup $(P_t)$ satisfies the gradient estimate $\text{GE}(K,\infty)$.

(ii) The semigroup $(P_t)$ is an EVI$^K$ gradient flow of $\text{Ent}$, the sublevel sets of $\text{Ent}$ are regular and $W$ is non-degenerate.

(iii) The pseudo metric $W$ is non-degenerate, $(D(\text{Ent}),W)$ is geodesic and $\text{Ent}$ is strongly geodesically $K$-convex.

Then (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii).

**Remark 7.13.** The properties (i), (ii) and (iii) can all be understood as lower Ricci curvature bounds for the geometry determined by $\mathcal{E}$. This approach has been studied intensively for metric measure spaces (see e.g. [AGS14b, LY09, Stu06a, Stu06b]) and, more recently, also for graphs (see e.g. [EM12, EHMT17]). We hope that the present framework allows to address the highly interesting question of introducing a concept of Ricci curvature (bounds) in noncommutative geometry. First steps in this direction was already taken by Hornshaw [Hor18].

**Remark 7.14.** The properties (i), (ii) and (iii) are equivalent for the Cheeger energy on infinitesimally Hilbertian length metric measure spaces ([AGS15], Theorem 1.1) and the Dirichlet form associated with a finite graph ([EM12, Theorem 4.5] and [EF18, Theorem 3.1]). It would be interesting to know if this is still true in this more general setting.

**REFERENCES**

[AES16] L. Ambrosio, M. Erbar, and G. Savaré. Optimal transport, Cheeger energies and contractivity of dynamic transport distances in extended spaces. *Nonlinear Anal.*, 137:77–134, 2016.

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
[AGS14a] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.*, 195(2):289–391, 2014.

[AGS14b] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.*, 163(7):1405–1490, 2014.

[AGS15] L. Ambrosio, N. Gigli, and G. Savaré. Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. *Ann. Probab.*, 43(1):339–404, 2015.

[AH77] S. Albeverio and R. Høegh-Krohn. Dirichlet forms and Markov semigroups on $C^*$-algebras. *Comm. Math. Phys.*, 56(2):173–187, 1977.

[AS18] L. Ambrosio and G. Stefani. Heat and entropy flows in Carnot groups. *ArXiv e-prints*, January 2018.

[BB00] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[BBI01] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BD58] A. Beurling and J. Deny. Espaces de Dirichlet. I. Le cas élémentaire. *Acta Math.*, 99:203–224, 1958.

[BD59] A. Beurling and J. Deny. Dirichlet spaces. *Proc. Nat. Acad. Sci. U.S.A.*, 45:208–215, 1959.

[BÉ85] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[BHL+15] F. Bauer, P. Horn, Y. Lin, G. Lippner, D. Mangoubi, and S.-T. Yau. Li-Yau inequality on graphs. *J. Differential Geom.*, 99(3):359–405, 2015.

[BK90] L. G. Brown and H. Kosaki. Jensen’s inequality in semi-finite von Neumann algebras. *J. Operator Theory*, 23(1):3–19, 1990.

[Bog07] V. I. Bogachev. *Measure theory. Vol. I, II*. Elements of Mathematics (Berlin). Springer-Verlag, Berlin, 2007.

[Bou89] N. Bourbaki. *General topology. Chapters 1–4*. Elements of Mathematics (Berlin). Springer-Verlag, Berlin, 1989. Translated from the French, Reprint of the 1966 edition.

[Bre17] Y. Brenier. Solution by convex minimization of the Cauchy problem for hyperbolic systems of conservation laws with convex entropy. *ArXiv e-prints*, October 2017.

[Bre18] Y. Brenier. The initial value problem for the Euler equations of incompressible fluids viewed as a concave maximization problem. *Comm. Math. Phys.*, 364(2):579–605, 2018.

[Car10] E. Carlen. Trace inequalities and quantum entropy: an introductory course. In *Entropy and the quantum*, volume 529 of *Contemp. Math.*, pages 73–140. Amer. Math. Soc., Providence, RI, 2010.

[CGGT17] Y. Chen, W. Gangbo, T. T. Georgiou, and A. Tannenbaum. On the Matrix Monge-Kantorovich Problem. *ArXiv e-prints*, January 2017.

[CGT18] Y. Chen, T. T. Georgiou, and A. Tannenbaum. Vector-Valued Optimal Mass Transport. *SIAM J. Appl. Math.*, 78(3):1682–1696, 2018.

[CHLZ12] S.-N. Chow, W. Huang, Y. Li, and H. Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. *Arch. Ration. Mech. Anal.*, 203(3):969–1008, 2012.

[Cip08] F. Cipriani. Dirichlet forms on noncommutative spaces. In *Quantum potential theory*, volume 1954 of *Lecture Notes in Math.*, pages 161–276. Springer, Berlin, 2008.

[Cip16] F. Cipriani. Noncommutative potential theory: a survey. *J. Geom. Phys.*, 105:25–59, 2016.

[CLLZ17] S.-N. Chow, W. Li, J. Lu, and H. Zhou. Population games and Discrete optimal transport. *ArXiv e-prints*, April 2017.

[CLSS10] J. A. Carrillo, S. Lisini, G. Savaré, and D. Slepčev. Nonlinear mobility continuity equations and generalized displacement convexity. *J. Funct. Anal.*, 258(4):1273–1309, 2010.

[CLY14] F. Chung, Y. Lin, and S.-T. Yau. Harnack inequalities for graphs with non-negative Ricci curvature. *J. Math. Anal. Appl.*, 415(1):25–32, 2014.

[CLZ18] S.-N. Chow, W. Li, and H. Zhou. Entropy dissipation of Fokker-Planck equations on graphs. *Discrete Contin. Dyn. Syst.*, 38(10):4929–4950, 2018.
S.-N. Chow, W. Li, and H. Zhou. A discrete Schrödinger equation via optimal transport on graphs. *J. Funct. Anal.*, 276(8):2440–2469, 2019.

E. A. Carlen and J. Maas. An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker-Planck equation is gradient flow for the entropy. *Comm. Math. Phys.*, 331(3):887–926, 2014.

E. A. Carlen and J. Maas. Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.*, 273(5):1810–1869, 2017.

A. Connes. Compact metric spaces, Fredholm modules, and hyperfiniteness. *Ergodic Theory Dynam. Systems*, 9(2):207–220, 1989.

E. A. Carlen and J. Maas. An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker-Planck equation is gradient flow for the entropy. *Comm. Math. Phys.*, 331(3):887–926, 2014.

E. A. Carlen and J. Maas. Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.*, 273(5):1810–1869, 2017.

A. Connes. Noncommutative geometry. Academic Press, Inc., San Diego, CA, 1994.

F. Cipriani and J.-L. Sauvageot. Derivations as square roots of Dirichlet forms. *J. Funct. Anal.*, 201(1):78–120, 2003.

D. Dier, M. Kassmann, and R. Zacher. Discrete versions of the Li-Yau gradient estimate. *ArXiv e-prints*, January 2017.

E. B. Davies and J. M. Lindsay. Noncommutative symmetric Markov semigroups. *Math. Z.*, 210(3):379–411, 1992.

N. Dunford and J. T. Schwartz. *Linear operators. Part I*. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1988. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

S. Daneri and G. Savaré. Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.*, 40(3):1104–1122, 2008.

J. Diestel and J. J. Uhl, Jr. *Vector measures*. American Mathematical Society, Providence, R.I., 1977. With a foreword by B. J. Pettis, Mathematical Surveys, No. 15.

M. Erbar and M. Fathi. Poincaré, modified logarithmic Sobolev and isoperimetric inequalities for Markov chains with non-negative Ricci curvature. *J. Funct. Anal.*, 274(11):3056–3089, 2018.

M. Erbar, M. Fathi, V. Laschos, and A. Schlichting. Gradient flow structure for McKean-Vlasov equations on discrete spaces. *Discrete Contin. Dyn. Syst.*, 36(12):6799–6833, 2016.

M. Erbar, C. Henderson, G. Menz, and P. Tetali. Ricci curvature bounds for weakly interacting Markov chains. *Electron. J. Probab.*, 22:Paper No. 40, 23, 2017.

M. Erbar, K. Kuwada, and K.-T. Sturm. On the equivalence of the entropic curvature-dimension condition and Bochner’s inequality on metric measure spaces. *Invent. Math.*, 201(3):993–1071, 2015.

M. Erbar, C. Henderson, G. Menz, and P. Tetali. Ricci curvature bounds for weakly interacting Markov chains. *Electron. J. Probab.*, 22:Paper No. 40, 23, 2017.

M. Erbar, K. Kuwada, and K.-T. Sturm. On the equivalence of the entropic curvature-dimension condition and Bochner’s inequality on metric measure spaces. *Invent. Math.*, 201(3):993–1071, 2015.

M. Erbar, J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.

M. Erbar, J. Maas. Gradient flow structures for discrete porous medium equations. *Discrete Contin. Dyn. Syst.*, 34(4):1355–1374, 2014.

M. Erbar, J. Maas, and M. Wirth. On the geometry of geodesics in discrete optimal transport. *Calc. Var. Partial Differential Equations*, 58(1):Art. 19, 19, 2019.

K.-J. Engel and R. Nagel. *One-parameter semigroups for linear evolution equations*, volume 194 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000. With contributions by S. Brendle, M. Campiti, T. Hahn, G. Metafune, G. Nickel, D. Pallara, C. Perazzoli, A. Rhandi, S. Romanelli and R. Schnaubelt.

M. Erbar. The heat equation on manifolds as a gradient flow in the Wasserstein space. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(1):1–23, 2010.

M. Erbar. Gradient flows of the entropy for jump processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 50(3):920–945, 2014.

M. Erbar. A gradient flow approach to the Boltzmann equation. *ArXiv e-prints*, March 2016.

M. Fathi and J. Maas. Entropic Ricci curvature bounds for discrete interacting systems. *Ann. Appl. Probab.*, 26(3):1774–1806, 2016.

M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet Forms and Symmetric Markov Processes*. De Gruyter Studies in Mathematics Series. De Gruyter, 1994.

N. Garcia Trillos. Gromov-Hausdorff limit of Wasserstein spaces on point clouds. *ArXiv e-prints*, February 2017.

N. Gigli. Nonsmooth differential geometry – An approach tailored for spaces with Ricci curvature bounded from below. *ArXiv e-prints*, July 2014.
[Pet88] D. Petz. A variational expression for the relative entropy. *Comm. Math. Phys.*, 114(2):345–349, 1988.

[PX03] G. Pisier and Q. Xu. Non-commutative $L^p$-spaces. In *Handbook of the geometry of Banach spaces, Vol. 2*, pages 1459–1517. North-Holland, Amsterdam, 2003.

[Rie99] M. A. Rieffel. Metrics on state spaces. *Doc. Math.*, 4:559–600, 1999.

[Roc97] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[RS78] M. Reed and B. Simon. *Methods of modern mathematical physics. IV. Analysis of operators*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1978.

[Rud91] W. Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.

[Sch12] K. Schmüdgen. *Unbounded self-adjoint operators on Hilbert space*, volume 265 of Graduate Texts in Mathematics. Springer, Dordrecht, 2012.

[Seg53a] I. E. Segal. Correction to “A non-commutative extension of abstract integration”. *Ann. of Math. (2)*, 58:595–596, 1953.

[Seg53b] I. E. Segal. A non-commutative extension of abstract integration. *Ann. of Math. (2)*, 57:401–457, 1953.

[SS64] D. Shale and W. F. Stinespring. States of the Clifford algebra. *Ann. of Math. (2)*, 80:365–381, 1964.

[Stu06a] K.-T. Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[Stu06b] K.-T. Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1):133–177, 2006.

[Tak02] M. Takesaki. *Theory of operator algebras. I*, volume 124 of Encyclopaedia of Mathematical Sciences. Springer-Verlag, Berlin, 2002. Reprint of the first (1979) edition, Operator Algebras and Non-commutative Geometry, 5.

[Tik87] O. E. Tikhonov. Continuity of operator functions in topologies connected with a trace on a von Neumann algebra. *Izv. Vyssh. Uchebn. Zaved., Mat.*, (1):77–79, 1987. In Russian.

[Vil09] C. Villani. *Optimal transport. Old and new*, volume 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2009.

[Zae15] D. Zaev. $L^p$-Wasserstein distances on state and quasi-state spaces of $C^*$-algebras. *ArXiv e-prints*, May 2015.

[Zae16] D. Zaev. On some topics of analysis on noncommutative spaces. *ArXiv e-prints*, December 2016.

**Institute of Mathematics, Department of Mathematics and Computer Science, Friedrich Schiller University Jena, 07737 Jena, Germany**

**Current address:** Institute of Science and Technology Austria (IST Austria), Am Campus 1, 3400 Klosterneuburg, Austria

**Email address:** melchior.wirth@ist.ac.at