# On curvature conditions using Wasserstein spaces

With an application to the $q$-Laplace heat equation

Der Fakultät für Mathematik und Informatik
der Universität Leipzig
eingereichte

DISSERTATION

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM
(Dr.rer.nat.)

im Fachgebiet

Mathematik

vorgelegt

von Diplommathematiker Martin Kell

geboren am 19. September 1985 in Neubrandenburg

Die Annahme der Dissertation wurde empfohlen von:

1. Professor Dr. Luigi Ambrosio (Pisa, Italien)
2. Professor Dr. Jürgen Jost (Leipzig)

Die Verleihung des akademischen Grades erfolgt mit Bestehen
der Verteidigung am 22.07.2014 mit dem Gesamtprädikat summa cum laude.

To the Wei I found

# Acknowledgement

I thank Prof. Jürgen Jost for giving me the opportunity to do my PhD-studies at the Max-Planck-Institute for Mathematics in the Sciences in Leipzig. From the beginning to the end he gave me complete freedom to "look around and check what others do". With this I not only found my niche, but also worked with many other people and had interesting discussions and fruitful results.

I want to thank my officemate Miroslav Bačák for discussions and the many replies to minor questions which popped up from time to time. I want to thank Christian Rodrigues and Ilona Kosiuk for interesting discussions on dynamical systems and also Hua, Bobo and Xia, Chao for discussions on abstract curvature notions. In particular, many thanks to Alihan Kabalak who made me interested in game theory and formal logic by having so many mathematical questions on the subject.

I want to thank the IMPRS at our institute for continuous financial support during my PhD-studies and also for organizing workshops with challenging questioners. I also thank our library team for providing even the most difficult to find articles less than a day after the request. Even more, I'm indepted to Prof. Otto for leaving his personal copy of [AGS08] to the library. For almost 4 year this book is by now an almost inventory of our office and inspired many parts of this work. Without the hard copy, things might have happened differently.

I thank the IMPA in Rio de Janeiro for two interesting months and financial support during that time.

I thank Antje Vandenberg for organizing trips, dealing with paperwork and for making our section more than just offices for scientists.

I'm grateful to be able to do my studies in the Max-Planck-Institute with its kitchen sections inviting people from different disciplines to share idea, especially I want to thank the regular and not so regular participants of the T- and C-breaks and the weekly D-dinners: Wiktor, Nils, Pierre-Yves, Holger, Lily and many others. From them I even saw the world outside of mathematics.

Last but not least, I'm grateful to have found my wife Wei who happened to attend courses with collegues from the institute. Without her continuous efforts to make me find my way I wouldn't have learnt and produced so much. Thanks for going along.

# Contents

# 1 Introduction

This thesis has two big parts, one is a new proof of the Borel-Brascamp-Lieb (BBL) inequality via geodesics in general Wasserstein spaces which makes it possible to define an abstract curvature condition. In the second part, the observation of Jordan, Kinderlehrer and Otto [JKO98] to identify the heat flow and the gradient flow of the entropy functional in the 2-Wasserstein space is developed further to show that a similar identification holds for the $q$-heat flow and the gradient flow of the Renyi entropy in the $p$-Wasserstein space.

## 1.1 Borel-Brascamp-Lieb and the curvature dimension

The proof of the BBL inequality for Riemannian manifolds by Cordero-Erausquin-McCann-Schmuckenschläger [CEMS01], and later for Finsler manifolds by Ohta [Oht09], led Lott-Villani [LV09, LV07] and Sturm [Stu06a, Stu06b] to a new notion of a lower bound on the generalized Ricci curvature for metric measure spaces, called curvature dimension. Both, the BBL inequality and the curvature condition, rely on geodesics in the 2-Wasserstein space, which was a natural candidate because of its connection to convex analysis in the Euclidean setting.

Based on Ohta's proof [Oht09] we show how to prove the BBL inequality via geodesics in the p-Wasserstein spaces for any $p > 1$. Following Lott-Villani-Sturm, a new curvature condition can be defined via convexities along geodesics in the $p$-Wasserstein space, and many known results, like Poincaré inequality and Bishop-Gromov volume comparison, follow by similar arguments.

The proof of the BBL inequality relied on three ingredients: (1) a solution to the Monge problem and a prescription of the interpolation maps, (2) second order differentiability of the solution potential and a cut locus description, and (3) positive (semi-) definiteness of the Jacobian of the interpolation map. The solution to the Monge problem easily follows by combining [McC01] and [Oht09]. The interpolation maps already give the idea that optimal transport is along geodesics, which is well-known by Lisini's result [Lis06]. For the proof of second order differentiability, we rely on two observations: (1) Ohta [Oht09] noticed that the lack of $C^2$-smoothness of square distance $d^2(\cdot,\cdot)$ at the diagonal can be avoided by splitting the transport plan into a moving and a non-moving part, this actually works for all smooth functions of the distance, (2) the set of $c_p$-concave functions is star-shaped. (1) says we only need to check where the transport map maps to different points and (2) helps to move the terminal point away from the cut-locus. Using this, a proof of the (almost) semiconcavity of solution potentials (Theorem 4.10) is given, which is shorter than Ohta's orginial proof [Oht08], yet it doesn't show that $c_p$-concave functions are everywhere locally semiconcave. However, it easily adapts to the Orlicz case, see Theorem 7.17.

The star-shapedness of $c_p$-concave functions, resp. pseudo star-shapedness of $c_L$-concave functions, and positive (semi-) definiteness of the Jacobian rely on the following, quite innocent looking inequalities: if $z \in Z_t(x,y)$ then for any $m$

$$t^{p-1}d^p(m,y) \le d^p(m,z) + t^{p-1}(1-t)d^p(x,y)$$

and

$$t^{-1}L(d(m,y)) \le L(d(m,z)/t) + t^{-1}(1-t)L(d(x,y))$$

where $L$ is a strictly increasing convex function. The orginial inequality with $p = 2$ can be easily derived from the binormial formula. The more general cases need explicitly convexity of $L$, resp. $p > 1$. The form of these inequalities can be predicted by analyzing the original proof of the interpolation inequality for $p = 2$ backwards. For this one needs a precise description of the Breinier-McCann-Ohta solution and the interpolation solution. In the cases $1 < p < \infty$ this was long known. In this thesis, it will be shown that the statement *geodesics in Wasserstein spaces transport their mass along geodesics* also holds for Orlicz-Wasserstein spaces.

As a "vertical dual" one can use the recent theory and calculus developed around the $q$-Cheeger energy ($q$ is the Hölder conjugate of $p$) by Ambrosio-Gigli-Savaré [AGS13, AGS11a, Gig12] to even get a $q$-Laplacian comparison, which, however, is equivalent to the usual one in the smooth setting. In chapter 6, we will study the gradient flow of the $q$-Cheeger energy, called $q$-heat flow, and use the "duality" and curvature condition to identify it with the gradient flow of the $(3-p)$-Renyi entropy if $p \in (1,3)$.

## 1.2 Heat and gradient flows

In [JKO98] Jordan, Kinderlehrer and Otto showed in the Euclidean setting that one can identify the heat flow with the gradient flow of the entropy functional in the 2-Wasserstein space. The main idea was to show that the solution of the gradient flow problem solves also the heat equation. Uniqueness of the solution implies that the two flows are identical. The identification of the heat flow and the gradient flow of the entropy functional on manifolds was later accomplished by Erbar [Erb10].

Otto [Ott96, Ott01] also gave a formal proof of how to use gradient flows in the $p$-Wasserstein spaces modeled on $\mathbb{R}^n$ in order to solve other equations like the porous media equation and the parabolic $q$-Laplace equation, i.e. the $q$-heat flow. Rigorous proofs were later given by Agueh [Agu02, Agu05]. Only recently Ohta and Takatsu [OT11a, OT11b] also showed that a similar construction works on manifolds if the functionals are $K$-convex.

All proofs until then required the contraction property which follows, at least in the Riemannian setting, from the curvature dimension condition introduced by Lott-Villani and Sturm [LV09, LV07, Stu06a, Stu06b]. Since this condition can be defined on any metric measure spaces it was believed that a similar identification holds also under such a condition. In [Gig09] Gigli gave a proof which did not require the contraction property. This proof let Ambrosio-Gigli-Savaré [AGS13] to define a new generalized gradient from which one gets a natural heat flow associated to a metric space. With the help of a calculus of the heat flow and its mass preservation they could show that the heat flow is a solution of the gradient flow problem of the entropy functional in the 2-Wasserstein space. Using a convexity of the square of the upper gradient of the entropy functional one gets uniqueness and hence the two flows are identical.

One of the main ingredient of the proof was the Kuwada lemma, i.e. if $\mu_t = f_t\mu$ is a solution of the heat flow and $|\dot{\mu}_t|$ is the metric derivative of $t \mapsto \mu_t$ in the 2-Wasserstein space $\mathcal{P}_2(M)$ then

$$|\dot{\mu}_t|^2 \le \int \frac{|\nabla f_t|^2}{f_t} d\mu$$

where the write hand side is called the Fisher information of $f_t$. This was the "missing" ingredient, since it was long known that the derivative along the heat flow $t \mapsto f_t$ of the entropy functional is (minus) the Fisher information of $f_t$.

In [AGS11a] Ambrosio-Gigli-Savaré showed the Kuwada lemma for $q \neq 2$, namely if $t \mapsto f_t$ is the $q$-heat flow such that the density is bounded from above and away from zero from below (implying the measure $\mu$ is finite), they showed

$$|\dot{\mu}_t|^p \le \int \frac{|\nabla f_t|^q}{f_t^{p-1}} d\mu$$

where this time the metric derivative is taken in the $p$-Wasserstein space $\mathcal{P}_p(M)$, $t \mapsto \mu_t = f_t\mu$ is a solution of the $q$-heat flow and $p$ and $q$ are Hölder conjugates. A formal calculation reveals that the derivative of the following functional

$$f \mapsto \frac{1}{(3-p)(2-p)} \int f^{3-p} - f d\mu,$$

called $(3-p)$-Renyi entropy, along the $q$-heat flow in the $p$-Wasserstein space is exactly minus the right hand side of the previous inequality, which can be called the $q$-Fisher information.

In this paper, we will follow [AGS13] and first develop a calculus of the $q$-heat flow to show mass preservation in the non-compact setting and that the formal calculation above holds in an abstract setting. In case $q > 2$ there is almost no restriction on the measure to get mass preservation besides a "not too bad" growth of the measure of a ball. The cases $q < 2$ are more restrictive. Using generalized exponential functions already know from information theory [OT11a, Section 3] one of the conditions can be stated as

$$\int \exp_p(-V^p) d\mu < \infty$$

where $V(x) = Cd(x, x_0)$ for some $C > 0$ and $exp_p$ is the generalized exponential function which agrees with the usual exponential function and the condition with the condition stated in [AGS13]. In $\mathbb{R}^n$ this condition boils down to $q > \frac{2n}{n+1}$. However, the current proof requires the more restrictive condition

$$\int V^p \exp_p(-V^p) d\mu < \infty.$$

In the second part under some assumptions on the functional, which hold assuming a curvature condition defined in chapter 46, we show that the proof of [AGS13] can be adjusted to show that the $q$-heat flow solves the gradient flow problem of the Renyi entropy in $\mathcal{P}_p(M)$. For $q > 2$ we also get convexity of the $q$-the power of the upper gradient and hence uniqueness of the gradient flow. This implies that the $q$-heat flow and the gradient flow of the Renyi entropy can be identified. The current proof of the cases $q < 2$ requires the space to be compact and the measure be $n$-Ahlfors regular for some $n$ depending on $q$. However, this condition is satisfied on smooth manifolds if the the curvature condition $CD_p(0, N)$ holds for $N > n$.

The difficulty of the cases $p \neq 2$ rely on a lack of Cauchy-Schwary inequalty and that in the original proofs many times the binormial formula is used instead of Jensen's inequality or convexity. Furthermore, the entropy functional and the heat equation are already well-studied objects . In particular, the condition $\int \exp(-V^2)d\mu < \infty$ was known to be sufficient to obtain mass preservation of the heat flow long before the paper [AGS13] whereas the author couldn't find any known condition on the mass preservation of the $q$-heat flow.

## A note on the notation and prepublished results

### Notation

In this work we try to use a context-sensitive notation in order to avoid overcomplicated names with $\tilde{}$, $\hat{}$, $\bar{}$. Below we will define distinct functionals $U_N$,$U_p$ and $U_m$ and assume $m, p, N \in \mathbb{R}$. Whereas $U_3$ is ambiguous, we hope that statements about $U_p, U_N$ or $U_m$ are unambiguous in the embedded context. Furthermore, whenever possible we will assume that $p$ and $q$ are Hölder conjugates and if not otherwise stated $1 < p, q < \infty$.

### Preprints

Chapter 3 to 5 and chapter 7 are already available as a preprint [Kel13] on arXiv and chapter 6 as [Kel14]. All results are sole work of the author; only the remark on page 48 was given by Shin-ichi Ohta on an early version of [Kel13].

# 2 Notation and Preliminaries

In this part, we will introduce the main concepts used in this work. Most of the notation and concepts of abstract metric spaces and gradients are taken from [AGS13, Gig12] (see also [AGS08]). For an introduction to the theory of optimal transport see [Vil09].

As a convention we will always assume that $(M, d)$ is a locally compact metric space and if not otherwise stated it is assumed to be geodesic (see below). Since we will also deal with non-locally compact spaces (e.g. $(\mathcal{P}_p(M), w_p)$ with $M$ non-compact), the sections below do not assume that $(X, d)$ is locally compact.

## 2.1 Calculus in metric spaces

Let $(X, d)$ be a (complete) metric space and for simplicity we assume that $X$ has no isolated points.

### 2.1.1 Lipschitz constants and upper gradients

Given a function $f : X \to \overline{\mathbb{R}} = [-\infty, \infty]$, the local Lipschitz constant $|Df| : X \to [0, \infty]$ is given by

$$|Df|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|}{d(y, x)}$$

for $x \in D(f) = \{x \in X \,|\, f(x) \in \mathbb{R}\}$, otherwise $|Df|(x) = \infty$. The one sided versions $|D^+ f|$ and $|D^- f|$, also called ascending slope (resp. descending slope)

$$\begin{aligned} |D^+ f|(x) &:= \limsup_{y \to x} \frac{[f(y) - f(x)]_+}{d(y, x)} \\ |D^- f|(x) &:= \limsup_{y \to x} \frac{[f(y) - f(x)]_-}{d(y, x)} \end{aligned}$$

for $x \in D(f)$ and $\infty$ otherwise, where $[r]_+ = \max\{0, r\}$ and $[r]_- = \max\{0, -r\}$. It is not difficult to see that $|Df|$ is (locally) bounded iff $f$ is (locally) Lipschitz.

The following lemma will be crucial to calculate the derivative of functionals along the gradient flow of the Cheeger energy.

**Lemma 2.1** ([AGS13, Lemma 2.5])**.** *Let $f, g : X \to \mathbb{R}$ be (locally) Lipschitz functions, $\phi : \mathbb{R} \to \mathbb{R}$ be a $C^1$-function with $0 \le \phi' \le 1$ and $\psi : [0, \infty) \to \mathbb{R}$ be a convex nondecreasing function. Setting*

$$\tilde{f} := f + \phi(g - f), \qquad \tilde{g} = g - \phi(g - f)$$

*we have for every $x \in X$*

$$\psi(|D\tilde{f}|)(x) + \psi(|D\tilde{g}|)(x) \le \psi(|Df|)(x) + \psi(|Dg|)(x).$$

We say that $g : X \to [0, \infty]$ is an upper gradient of $f : X \to \overline{\mathbb{R}}$ if for any absolutely continuous curve $\gamma : [0, 1] \to D(f)$ the curve $t \mapsto g(\gamma_s)|\dot{\gamma}_s|$ is measurable in $[0, 1]$ (with convention $0 \cdot \infty = 0$) and

$$|f(\gamma_1) - f(\gamma_0)| \le \int_0^1 g(\gamma_t)dt.$$

It is not difficult to see that the local Lipschitz constant and the two slopes are upper gradients in case $f$ is (locally) Lipschitz.

### 2.1.2 Relaxed slopes and the Cheeger energy

In a metric space there is no natural gradient of $L^r$-functions which are not Lipschitz. Cheeger defined in [Che99] a gradient via a relaxation procedure using slopes of Lipschitz function. In [AGS13, AGS11a] Ambrosio-Gigli-Savaré used a more restrictive version of Cheeger's original definition.

**Definition 2.2** ($q$-relaxed slope)**.** A function $g \in L^q$ is a $q$-relaxed slope of $f \in L^2$ if there is a sequence of Lipschitz functions $f_n$ strongly converging to $f$ in $L^2$ such that $|Df_n|$ converges weakly (in $L^q$) to some $\tilde{g} \in L^q$ with $g \le \tilde{g}$. We denote by $|\nabla f|_{*,q}$ the element of minimal $L^q$-norm among all $q$-relaxed slopes.

*Remark.* In order to apply the gradient flow theory of Hilbert spaces, we divert from the approach in [AGS11a] and use approximations of $f$ in $L^2$ instead of $L^q$. Note that the proofs of [AGS11a] also work in this setting if appropriate changes are made.

It was shown in [AGS11a] that this definition, Cheeger's original and two other definitions agree almost everywhere. However, if the space does not satisfy a local doubling condition and a local Poincaré inequality, then the $q$-relaxed slope might be different from the $q'$-relaxed slope if $q \neq q'$, see [DS13]. Nevertheless, we will drop the dependency on $q$ and just write $|\nabla f|_*$.

One can show that the relaxed slope is sublinear, i.e. $|\nabla(f + g)|_* \le |\nabla f|_* + |\nabla g|_*$ almost everywhere, and satisfies a weak form of the chain rule, i.e. for any $C^1$-function $\phi : \mathbb{R} \to \mathbb{R}$, which is Lipschitz on the image of $f$, we have $|\nabla\phi(f)|_* \le |\phi'(f)||\nabla f|_*$ with equality if $\phi$ is non-decreasing [AGS13, Proposition 4.8]. This can be easily proven for Lipschitz functions and their slopes, and follows by a cut-off argument also for functions and their relaxed slopes.

Now the $q$-Cheeger energy of the metric measure space $(M, d, \mu)$ is defined as

$$\mathrm{Ch}_q(f) = \frac{1}{q} \int |\nabla f|_*^q d\mu$$

for all $f$ admitting a relaxed slopes, otherwise $\mathrm{Ch}_q(f) = \infty$. Similarly, given a convex increasing function $L : [0, \infty) \to [0, \infty)$ with $L(0) = 0$, the $L$-Cheeger energy is defined as $\mathrm{Ch}_L(f) = \int L(|\nabla f|)d\mu$. Then the $q$-Cheeger energy is nothing but the $L$-Cheeger energy for $L(r) = r^q/q$.

**Proposition 2.3.** *Let $f, g \in D(\mathrm{Ch}_q)$ and $\phi : \mathbb{R} \to \mathbb{R}$ be a nondecreasing contraction (with $\phi(0) = 0$ if $\mu(M) = \infty$) then $\mu$-almost everywhere in $M$*

$$|\nabla(f + \phi(g - f))|_*^q + |\nabla(g - \phi(g - f))|_*^q \leq |\nabla(f)|_*^q + |\nabla(g)|_*^q.$$

*Proof.* The proof follows along the lines of the proof of [AGS13, Proposition 4.8] using Lemma 2.1 (see [AGS13, Lemma 2.5]). □

### 2.1.3 Fisher information

The Fisher information is the derivative of the entropy functional along the heat flow. The Kuwada lemma, a key tool of [AGS13] to identify the heat flow and the gradient flow of the entropy functional, shows that the square of the metric derivative in the 2-Wasserstein space along the heat flow is bounded from above by the Fisher information. In a different paper [AGS11a] they showed that in the compact setting with density of the measure bounded from below and above, there is also a version of this along the $q$-heat flow in the $p$-Wasserstein space (see Lemma 6.9 for a precise version). For that reason we define the following $q$-Fisher information as follows.

**Definition 2.4** ($q$-Fisher information)**.** Let $q \in (\frac{1+\sqrt{5}}{2}, \infty)$. For a Borel function $f : M \to [0, \infty]$ we define the $q$-Fisher information $\mathsf{F}_q(f)$ as

$$\mathsf{F}_q(f) := r^{-q} \int |\nabla f^r|_*^q d\mu = q r^{-q} \,\mathrm{Ch}_q(f^r)$$

where $q \neq \frac{1+\sqrt{5}}{2}$ and

$$r = 1 - \frac{p-1}{q} = 1 - \frac{(p-1)^2}{p}.$$

In case $q = \frac{1+\sqrt{5}}{2}$, note $q = p - 1$ and thus we define

$$\mathsf{F}_q(f) = \int |\nabla \log f|_w^q d\mu = q \,\mathrm{Ch}_q(\log f).$$

*Remark.* For $q \in (\frac{1+\sqrt{5}}{2}, \infty)$, we also have $r \in (0, 1)$, which will be our main interest for technical reasons. Nevertheless, all case $q \geq 2$ are covered. In the following, we will just write $r > 0$. Furthermore, notice that $N \geq 2$ and $1 - \frac{1}{N} = 3 - p$ implies $p = 2 + \frac{1}{N} \leq 2.5 < \frac{3+\sqrt{5}}{2}$. Thus only the cases $N \in (1, 2)$ remain to be covered. In the smooth setting $CD_p(K, N)$ with $N \in (1, 2)$ can only hold for 1-dimensional spaces.

**Proposition 2.5.** *Let $r > 0$. Then for every Borel function $f : M \to [0, \infty]$ we have the equivalence*

$$f \in D(\mathsf{F}_q) \iff f \in L^{2r}(M, \mu) \text{ and } \int_{\{f>0\}} \frac{|\nabla f|_*^q}{f^{p-1}} d\mu < \infty$$

*and in this case we have*

$$\mathsf{F}_q(f) = \int_{\{f>0\}} \frac{|\nabla f|_*^q}{f^{p-1}} d\mu.$$

*In addition, the functional is sequentially lower semicontinuous w.r.t. the strong convergence in $L^{2r}(M,\mu)$ and $L^2(M,\mu)$. If $p < 2$ then the functional is also convex.*

*Remark.* Compare this to [AGS11a, Remark 6.2] and [AGS13, Lemma 4.10]. And note that the statement $|\nabla f|_w \in L^1$ follows already from $f \in L^1$ and $\int \frac{|\nabla f|_w^2}{f} d\mu < \infty$ by applying the reverse Hölder inequality.

*Proof.* Similar to [AGS13, Lemma 4.10] first assume $f$ is bounded. Then note that $f \in D(\mathsf{F}_q)$ requires $f^r \in L^2(M,\mu)$, i.e. $f \in L^{2r}(M,\mu)$ and by chain rule

$$|\nabla f^r|_*^q = r^q \frac{|\nabla f|_*^q}{f^{p-1}}.$$

Conversely, just use $\phi(r) = \sqrt{r+\epsilon} - \sqrt{\epsilon}$, apply the chain rule and let $\epsilon \to 0$.

Convexity for $p < 2$ follows from [Bor97]: Since in that case $q \geq p$, we know $(x,y) \mapsto x^q/y^{p-1}$ is convex in $\mathbb{R}^2$. □

Later we will see that the $q$-Fisher information is the derivative of the Renyi entropy

$$f \mapsto \frac{1}{(3-p)} \int f \ln_p f d\mu$$

along the $q$-heat flow. In case $q = p = 2$ we see that this boils down to the classical case.

### 2.1.4 Absolutely continuous curves and geodesics

If $I \subset \mathbb{R}$ is an open interval then we say that a curve $\gamma : I \to X$ is in $AC^p(I,X)$ (we drop the metric $d$ for simplicity) for some $p \in [1,\infty]$ if

$$d(\gamma_s,\gamma_t) \leq \int_s^t g(r)dr \quad \forall s,t \in J : s < t$$

for some $g \in L^p(J)$. In case $p = 1$ we just say that $\gamma$ is absolutely continuous. It can be shown [AGS08, Theorem 1.1.2] that in this case the metric derivative

$$|\dot\gamma_t| := \limsup_{s\to t} \frac{d(\gamma_s,\gamma_t)}{|s-t|}$$

with lim for a.e. $t \in I$ is a minimal representative of such a $g$. We will say $\gamma$ has constant (unit) speed if $|\dot\gamma_t|$ is constant (resp. 1) almost everywhere in $I$.

It is not difficult to see that $AC^p(I,X) \subset C(\bar I,X)$ where $C(\bar I,X)$ is equipped with the sup distance $d^*$

$$d^*(\gamma,\gamma') := \sup_{t\in\bar I} d(\gamma_t,\gamma'_t).$$

For each $t \in \bar{I}$ we can define the evaluation map $e_t : C(\bar{I}, X) \to X$ by

$$e_t(\gamma) = \gamma_t.$$

We will say that $(X, d)$ is a geodesic space if for each $x_0, x_1 \in X$ where is a constant speed curve $\gamma : [0,1] \to X$ with $\gamma_i = x_i$ and

$$d(\gamma_s, \gamma_t) = |t - s| d(\gamma_0, \gamma_1).$$

In this case, we say that $\gamma$ is a constant speed geodesic. The space of all constant speed geodesics $\gamma : [0,1] \to X$ will be donated by $\mathrm{Geo}(X)$. Using the triangle inequality it is not difficult to show the following.

**Lemma 2.6.** *Assume $\gamma : [0,1] \to X$ is a curve such that*

$$d(\gamma_s, \gamma_t) \le |t - s| d(\gamma_0, \gamma_1)$$

*then $\gamma$ is a geodesic from $\gamma_0$ to $\gamma_1$.*

A weaker concept is a length space: In such spaces the distance between point $x_0$ and $x_1 \in X$ is given by

$$d(x_0, x_1) = \inf \int_0^1 |\dot{\gamma}_t| dt$$

where the infimum is taken over all absolutely continuous curves connecting $x_0$ and $x_1$. In case $X$ is complete and locally compact, the two concepts agree. Furthermore, Arzela-Ascoli also implies:

**Lemma 2.7.** *If $(X, d)$ is locally compact then so is $(\mathrm{Geo}(X), d^*)$ where $d^*$ is the* sup-*distance on $C(\bar{I}, X)$.*

### 2.1.5 Geodesically convex functionals and gradient flows

A functional $E : X \to \mathbb{R} \cup \{+\infty\}$ is said to be $K$-geodesically convex for some $K \in \mathbb{R}$ if for each $x_0, x_1 \in D(E)$ there is a geodesic $\gamma \in \mathrm{Geo}(X)$ connecting $x_0$ and $x_1$ such that

$$E(\gamma_t) \le (1-t)E(\gamma_0) + tE(\gamma_1) - \frac{K}{2}(1-t)td^2(\gamma_0, \gamma_1).$$

In such a case it can be shown ([AGS08, Section 2.4] that the descending slope is an upper gradient of $E$ and can be expressed as

$$|D^- E|(x) = \sup_{y \in X \setminus \{x\}} \left( \frac{E(x) - E(y)}{d(x,y)} + \frac{K}{2} d(x,y) \right)$$

In particular, it is lower semicontinuous if $E$ is. Furthermore, if $x : [0, \infty) \to D(E)$ is a locally absolutely continuous curve then

$$E(x_t) \ge E(x_s) - \int_s^t |\dot{x}_r| |D^- E|(y_r) dr$$

for every $s,t \in [0,\infty)$ and $s < t$. Note by Young's inequality we also have for any $p \in (1,\infty)$

$$E(x_t) \geq E(x_0) - \frac{1}{p}\int_0^t |\dot{x}_t|^p dt - \frac{1}{q}\int_0^t |D^- E|^q(x_r)dr.$$

**Definition 2.8** ($(E,p)$-dissipation inequality and metric gradient flows)**.** Let $E : X \to \mathbb{R} \cup \{\infty\}$ be a functional on $X$ then we say that a locally absolutely continuous curve $t \mapsto y_t \in D(E)$ satisfies the $(E,p)$-dissipation inequality if for all $t \geq 0$

$$E(x_0) \geq E(x_t) + \frac{1}{p}\int_0^t |\dot{x}_t|^p dt + \frac{1}{q}\int_0^t |D^- E|^q(x_r)dr.$$

$t \mapsto x_t$ is a gradient flow of $E$ starting at $y_0 \in D(E)$ if

$$E(y_0) = E(x_t) + \frac{1}{p}\int_0^t |\dot{x}_t|^p dt + \frac{1}{q}\int_0^t |D^- E|^q(x_r)dr.$$

In the geodesically convex case we immediately see that if $t \mapsto x_t$ satisfies the $(E,p)$-dissipation inequality then it is a (generalized) gradient flow and

$$\frac{d}{dt}E(x_t) = -|\dot{x}_t|^p = -|D^- E|^q(x_t)$$

for almost all $t \in (0,1)$.

*Remark.* The theory developed in [AGS08] covers mainly the case $p = 2$ and only mentioned the required adjustments. For a comprehensive treatment of the case $p \neq 2$ and even more general situations see [RMS08].

### 2.1.6 Optimal Transport

Let $(M,d)$ be a proper metric space. Given two probability measure $\mu_0, \mu_1 \in \mathcal{P}(M)$ and a (non-negative) cost function $c : M \times M \to [0,\infty)$ one can define the following Kantorovich problem

$$C(\mu_0,\mu_1) = \inf_{\pi \in \Pi(\mu_0,\mu_1)} \int c(x,y) d\pi(x,y)$$

where $\Pi(\mu_0,\mu_1)$ is the set of all $\pi \in \mathcal{P}(M \times M)$ such that $(p_1)_*\pi = \mu_0$ and $(p_2)_*\pi = \mu_1$ with $p_i$ being the projections to the $i$-th coordinate.

It is well-known that the problem has a solution $\pi_{opt}$, i.e. a probability measure $\pi_{opt}$ in $\Pi(\mu_0,\mu_1)$ such that

$$C(\mu_0,\mu_1) = \int c(x,y) d\pi_{opt}(x,y).$$

Given any such cost function one can define a dual problem

$$\tilde{C}(\mu_0,\mu_1) = \sup_{\phi(x)+\psi(y) \leq c(x,y)} \int \phi d\mu_0 + \int \psi d\mu_1.$$

It is not difficult to see that $\tilde{C} \leq C$.

The solution to this problem can be described by a pair of $c$-concave potentials: if $\psi : M \to \underline{\mathbb{R}}$ then one can define the $c$-transform as

$$\psi^c(y) = \inf_{x\in M} c(x,y) - \phi(x).$$

We say that $\phi$ is $c$-concave if it is the $c$-transform of some function $\psi$. Similarly, with $c$ replaced by $\bar{c}(x,y) = c(y,x)$ one defines the $\bar{c}$-transform of $\phi$ and says $\psi$ is $\bar{c}$-concave if it is the $\bar{c}$-transform of a function $\phi$.

Given a $c$-concave function $\phi = \psi^c$ one can define the $c$-superdifferential $\partial^c\phi$ by

$$\partial^c\phi(x) = \{y \in M \,|\, \phi(x) + \psi(y) = c(x,y)\}.$$

One of the major results in optimal transport theory is the following:

**Theorem 2.9.** *[Vil09, Theorem 5.11]One always has*

$$\tilde{C}(\mu_0,\mu_1) = C(\mu_0,\mu_1)$$

*and the dual problem is attained by a pair $(\phi,\psi)$ of c-concave/$\bar{c}$-concave functions with $\phi = \psi^c$ and $\psi = \phi^{\bar{c}}$. Assuming, for simplicity, that c is continuous, then the optimal transport measure $\pi_{opt}$ is supported on the graph of the c-superdifferential which is c-cyclically monotone, i.e. given n couples $(x_i,y_i) \in \partial^c\phi$ one has*

$$\sum_{i=0}^{n-1} c(x_i,y_i) \leq \sum_{i=1}^{n-1} c(x_i,y_{i+1}).$$

*Furthermore, if $\partial^{c_p}\phi(\cdot)$ is single-valued $\mu_0$-almost everywhere, then $\pi_{opt}$ is concentrated on the graph of a measurable function $T$ where $T$ is a measurable selection of $x \mapsto \partial^c\phi(x)$ which is uniquely defined $\mu_0$-a.e..*

## Wasserstein spaces

In this section, we will give a short introduction to the Wasserstein spaces $\mathcal{P}_p(M)$, for an overview see [Vil09, Chapter 6].

Fix some $x_0 \in M$ and let $\mathcal{P}(M)$ be the set of probability measures on $M$. The $p$-Wasserstein space for $1 < p < \infty$ is the space of all probability measures with finite $p$-moments

$$\mathcal{P}_p(M) = \{\mu \in \mathcal{P}(M) \,|\, \int d^p(x,x_0)d\mu(x) < \infty\}$$

equipped with the metric

$$w_p(\mu_0,\mu_1) = (C_p(\mu_0,\mu_1))^{\frac{1}{p}}$$

where the cost function is given by $c_p(x,y) = d^p(x,y)/p$.

It is well know that $(\mathcal{P}_p(M), w_p)$ is a complete metric measure space if $(M,d)$ is and is geodesic if $M$ is. Furthermore, it is compact iff $M$ is (see [Vil09, Chapter 6]). However, it is not locally compact if $M$ is just locally compact. Nevertheless, in case $M$ is a proper metric space there is a sufficiently nice weak topology induced by the subspace topology of $\mathcal{P}(M)$ with its weak topology.

**Lemma 2.10** (see e.g. [Kel11, Theorem 6]). *Let $(M,d)$ be a proper metric space, then every bounded set in $\mathcal{P}_p(M)$ is precompact w.r.t. to the weak topology induced by $\mathcal{P}_p(M) \subset \mathcal{P}(M)$.*

*Proof.* Let $x_0$ be some fixed point in $M$. By [Vil09, Lemma 4.3] we know that the $w_p(\delta_{x_0},\cdot)$ is weakly lower semicontinuous. Thus we only need to prove tightness of every $w_p$-ball $B_R^p(\delta_{x_0}) \subset \mathcal{P}_p(M)$, i.e. for every $\epsilon > 0$ there is a compact set $K_\epsilon \subset M$ such that every $\mu \in B_R^p(\delta_{x_0})$

$$\mu(M\backslash K_\epsilon) \le \epsilon.$$

If we set $K_\epsilon = B_{\frac{1}{\epsilon}}(x_0)$ then

$$\begin{aligned}\mu(M\backslash K_\epsilon) &\le \epsilon^p \int_{d(x,x_0)\ge\frac{1}{\epsilon}} d^p(x,x_0)d\mu(x)\\ &\le \epsilon^p p w_p^p(\delta_{x_0},\mu) \le \epsilon p R^p\end{aligned}$$

which implies tightness since any ball in $M$ is compact. □

In the chapter 7, we introduce more general Wasserstein spaces, called Orlicz-Wasserstein space. For those the distance is not given by a single optimization problem and so far there is no nicely defined dual problem. However, the lemma above also holds in a similar way (see Proposition 7.5).

We say that a function $E : \mathcal{P}_p(M) \to \mathbb{R} \cup \{\infty\}$ is weakly lower semi-continuous if it is lower semicontinuous w.r.t. the weak topology on $\mathcal{P}_p(M) \subset \mathcal{P}(M)$. In particular, the weak closure of bounded subset of sublevels of $E$ are contained in that sublevel.

**Theorem 2.11.** *Let $(M,d)$ be a proper geodesic metric space and $E$ be a functional on $\mathcal{P}_p(M)$ such that $E$ and $|D^-E|$ are weakly lower semicontinuous. Then for all $\mu_0 \in D(E)$ there exists a gradient flow $t \mapsto \mu_t$ of $E$ starting at $\mu_0$.*

*Proof.* Just note by the previous lemma the assumptions [AGS08, Assumption 2.4a,c] hold and thus [AGS08, Corollary 2.4.12] can be applied. □

*Remark.* The requirement $|D^-E|$ to be weakly lower semicontinuous is rather restrictive in the non-compact case. Note, however, below we only need lower semicontinuity, which follows from $K$-convexity. Existence will follow from existence of the $q$-heat equation.

## 2.2 Finsler manifolds

In this section, we recall some notation and facts from Finsler geometry. We will mainly follow the notation of [Oht09, Oht08] and otherwise refer to [BCS00, She01].

### 2.2.1 Finsler structures

Let $M$ be a connected, $n$-dimensional $C^\infty$-manifold.

**Definition 2.12** (Finsler structure)**.** A $C^\infty$-Finsler structure on $M$ is a function $F : TM \to [0,\infty)$ such that the following holds

1. (Regularity) $F$ is $C^\infty$ on $TM\backslash\{\mathbf{0}\}$ where $\mathbf{0}$ stands for the zero section,
2. (Positive homogeneity) for any $v \in TM$ and any $\lambda > 0$, it holds $F(\lambda v) = \lambda F(v)$,
3. (Strong convexity) In local coordinates $(x^i)_{i=1}^n$ on $U \subset M$ the matrix
$$(g_{ij}(v)) := \left(\frac{1}{2}\frac{\partial^2(F^2)}{\partial v^i \partial v^j}(v)\right)$$
is positive-definite at every $v \in \pi^{-1}(U)\backslash 0$ where $\pi : TM \to M$ is the natural projection of the tangent bundle.

Strictly speaking, this is nothing more than defining a Minkowski norm $F|_{T_xM}$ on each $T_xM$ with some regularity requirements depending on $x$. We don't require $F$ to be absolutely homogeneous, i.e. $F(v) \neq F(-v)$ is possible. In such a case the "induced" distance (see below) is not symmetric. As an abbreviation we let $\bar{F}$ denote the reverse Finsler structure, i.e. $\bar{F}(v) = F(-v)$.

*Remark.* Most of the statements below and in Chapter 4 only require $C^2$-regularity of the Finsler structure. The $C^\infty$-regularity is only assumed for convenience.

On any $C^\infty$-manifold one can define the differential $df$ of a $C^1$-function $f$. In order to define the gradient of $f$ one needs the following: let $\mathcal{L} : T^*M \to TM$ be the Legendre transform associating to each co-vector $\alpha \in T_x^*M$ the unique vector $v = \mathcal{L}_x(\alpha) \in T_xM$ such that $F(v) = F^*(v)$ and $\alpha(v) = F(v)^2$, where $F^*$ is the dual norm of $F$ on $T^*M$. This transform is $C^\infty$ from $T^*M\backslash\{0\}$ to $TM\backslash\{0\}$ and is $C^\infty$ in case $F$ is a Riemannian structure, i.e. the parallelogram inequality holds on each $T_xM$. The gradient $\nabla f$ at $x$ of $f$ is now defined by $\nabla f(x) = \mathcal{L}_x(df_x) \in T_xM$. Then we have for every unit speed $C^1$-curve $\eta : [0,l] \to M$ (i.e. $F(d\eta/dt) \equiv 1$)

$$-\int_0^l F(\nabla(-f)(\eta_t))dt \le f(\eta(l)) - f(\eta(0)) \le \int_0^l F(\nabla f(\eta_t))dt.$$

Thus one can define an intrinsic metric of the Finsler manifold by

$$d(x,y) = \sup_{f \in C^1, F(\nabla f) \le 1} f(y) - f(x)$$

which is symmetric iff $F = \bar{F}$.

In the Finsler setting, there is no good notion of a (Finsler) Hessian of a $C^2$-function $f$, so that we will use the well-defined differential of $df : M \to T^*M$ which can be written in local coordinates as

$$d(df)_x = \sum_{i,j=1}^n \left( \delta_j^i \frac{\partial}{\partial x^i}\bigg|_{df_x} + \frac{\partial^2 f}{\partial x^i \partial x^j}(x) \frac{\partial}{\partial v^i}\bigg|_{df_x} \right) dx^j\big|_x.$$

Note, however, that this expression is not coordinate free.

### 2.2.2 Chern connection, covariant derivatives and curvature

In contrast to Riemannian manifolds there is no "unique" canonical connection defined on a Finsler manifold. As in [Oht09] we will only use the Chern connection in this article which is the same as the Levi-Civita connection in the Riemannian case. In order to reduce the notation we will only use the Chern connection and denote it by $\nabla$ without stating its exact property ([Oht09, Definition 2.2]). For a thorough introduction see [Oht09, BCS00, She01].

Recall that by strong convexity of $F$ the matrix $(g_{ij}(v))$ is positive definite for every $v \in T_xM\backslash\{0\}$ and hence defines a scalar product on $T_xM$ which will be denoted by $g_v(\cdot,\cdot)$, i.e.

$$g_v(\sum_{i=1}^n w_1^i \frac{\partial}{\partial x^i}\bigg|_x, \sum_{j=1}^n w_2^j \frac{\partial}{\partial x^j}\bigg|_x) = \sum_{i,j=1}^n g_{ij}(v) w_1^i w_2^j.$$

Using the definition of Legendre transform one sees that $\mathcal{L}_x^{-1}(v)(w) = g_v(v,w)$ for $w \in T_xM$ and thus $g_v(v,v) = F(v)^2$. Different from Riemannian metrics, $g_v$ is non-constant and the following tensor, called Cartan tensor is non-zero (at least for some $v \in TM\backslash\{0\}$).

$$A_{ijk}(v) := \frac{F(v)}{2}\frac{\partial g_{ik}}{\partial v^k}(v) = \frac{F(v)}{4}\frac{\partial^3(F^2)}{\partial v^i \partial v^j \partial v^k}(v).$$

Further, we can define the formal Christoffel symbol by

$$\gamma_{jk}^i(v) := \frac{1}{2}\sum_{l=1}^n g^{il}(v)\left\{\frac{\partial g_{lj}}{\partial x^k}(v) - \frac{\partial g_{jk}}{\partial x^l}(v) + \frac{\partial g_{kl}}{\partial x^j}(v)\right\}$$

for $v \in TM\backslash 0$ and also

$$N_j^i(v) := \sum_{k=1}^n \gamma_{jk}^i(v)v^k - \frac{1}{F(v)}\sum_{k,l,m=1} A_{jk}^i(v)\gamma_{lm}^k(v)v^l v^m$$

where $(g^{ij})$ is the inverse of $(g_{ij})$ and $A_{jk}^i := \sum_l g^{il}A_{ljk}$.

Given the Chern connection $\nabla$ let $\omega_j^i$ be its connection one-forms which are defined by

$$\nabla_v \frac{\partial}{\partial x^j} = \sum_{i=1}^n \omega_j^i(v)\frac{\partial}{\partial x^i}, \nabla_v dx^i = \sum_{j=1}^n -\omega_j^i(v)dx^j$$

and by torsion-freeness can be written as

$$\omega_j^i = \sum_k \Gamma_{jk}^i dx^k.$$

Given two non-zero vectors $v, w \in T_xM\backslash\{0\}$, a $C^\infty$-vector field $X$ and the connection one-forms, one can define the covariant derivative $D_v^w X$ with reference vector $w$ as

$$(D_v^w X)(x) := \sum_{i,j=1}^n \left\{v^j \frac{\partial X^i}{\partial x^j} + \sum_{k=1}^n \Gamma_{jk}^i(w)v^j X^k\right\}\frac{\partial}{\partial x^i}\bigg|_x.$$

In the Riemannian case, the covariant derivative does not depend on the vector $w$ and is just the usual covariant derivative.

From the Chern connection one can also define its curvature two-forms

$$\Omega_i^j := dw_j^i - \sum_{k=1}^{n} \omega_j^k \wedge \omega_k^i$$

which can be also written as

$$\Omega_i^j(v) = \frac{1}{2}\sum_{k,l=1}^{n} R_{jkl}^i(v) dx^k \wedge dx^l + \frac{1}{F(v)} \sum_{k,l=1}^{n} P_{jkl}^i(v) dx^x \wedge \delta v^l$$

where we require $R_{jkl}^i = -R_{jlk}^i$ and $\delta v^k = dv^k + \sum_l N_l^k dx^l$.

With the help of $R_{jkl}^i$ one can define the Riemannian tensor with reference vector $v \in TM$

$$R^v(w,v)v := \sum_{i,j,k,l=1}^{n} v^j R_{jkl}^i(v) w^k v^l \frac{\partial}{\partial x^i}\bigg|_x$$

which enjoys the following

$$g_v(R^v(w,v)v, w') = g_v(R^v(w',v)v, w) \text{ and } R^v(v,v) = 0.$$

Given all those definition we finally have the flag curvature

$$\mathcal{K}(v,w) := \frac{g_v(R^v(w,v)v, w)}{g_v(v,v)g_v(w,w) - g_v(v,w)^2}$$

and the Ricci curvature

$$Ric(v) := \sum_{i=1}^{n-1} \mathcal{K}(v, e_i)$$

where $e_1, e_2, \cdots, e_{n-1}, v/F(v)$ form an orthonormal basis of $T_xM$ w.r.t. $g_v$.

On unweighted Finsler manifolds we say that $(M,F)$ has Ricci curvature bounded from below if

$$Ric(v) \geq K$$

for every unit vector $v \in TM$. For weighted manifolds we need the following: Let $\mu$ be the reference measure and $\mathrm{vol}_{g_v}$ be the Lebesgue measure on $T_xM$ induced by $g_v$. If $\mu_x$ denotes the measure $T_xM$ induced by $\mu$ define

$$\mathcal{V}(v) := \log\left(\frac{\mathrm{vol}_{g_v}(B_{T_xM}^+(0,1)}{\mu_x(B_{T_xM}^+(0,1)}\right)$$

where $B_{T_xM}^+(0,1)$ denotes the (forward) unit ball of radius 1 w.r.t. the norm $F|_{T_xM}$. Further, let

$$\partial_v \mathcal{V} := \frac{d}{dt}\bigg|_{t=0} \mathcal{V}(\dot\eta(t)), \partial_v^2 \mathcal{V} := \frac{d}{dt}\bigg|_{t=0} \mathcal{V}(\dot\eta(t))$$

where $\eta : (-\epsilon, \epsilon) \to M$ is a geodesic with $\dot\eta(0) = v$.

**Definition 2.13** (Weighted Ricci curvature)**.** Define the following objects:

1. $Ric_n(v) := \begin{cases} Ric(v) + \partial_v^2 \mathcal{V} & \text{if } \partial_v \mathcal{V} = 0 \\ -\infty & \text{otherwise} \end{cases}$
2. $Ric_N(v) := Ric(v) + \partial_v^2 \mathcal{V} + \frac{\partial_v \mathcal{V}}{N-n}$ for $N \in (n, \infty)$.
3. $Ric_\infty(v) := Ric(v) + \partial_v^2 \mathcal{V}$

Which is called the (weighted) $n$-Ricci curvature, resp. $N$- and $\infty$-Ricci curvature of the weighted Finsler manifold $(M, F, \mu)$.

*Remark.* By a recent paper of Ohta [Oht13a] it also makes sense to define the $N$-Ricci curvature for negative $N$.

Now a lower curvature bound $K$ on the $N$-Ricci curvature (resp. $n$-, $\infty$-Ricci curvature) is nothing but

$$Ric_N(v) \geq K$$

for all unit vector $v \in TM$. In the Riemannian setting $Ric_\infty$ is the so called Bakry-Émery Ricci tensor and a lower curvature bound $K$ is equivalent to the Bakry-Émery curvature condition on the heat flow, i.e.

$$|\nabla P_t f|^2 \leq e^{-2Kt} P_t(|\nabla f|^2).$$

Similarly, a lower bound on $Ric_N$ is equivalent to a more general Bakry-Émery conditions involving $N$. However, the Bakry-Émery calculus requires the space to be Riemannian, resp. infinitesimal Hilbertian. Nevertheless, one could define a Finsler version of the Bakry-Émery condition similar to the Finslerian Bochner inequality of [OS11]. It is not difficult to see that those curvature bounds are equivalent. The lack of a well-defined linearization of the heat flow makes it difficult to even define such a generilized Bakry-Émery condition for more abstract spaces.

### 2.2.3 Geodesics and first and second variation formula

Given a $C^1$-curve $\eta : [0, r] \to M$ its arclength is defined by

$$\mathcal{L}(\eta) := \int_0^r F(\dot{\eta}_t) dt$$

where $\dot{\eta}_t = \frac{d}{dt}\eta_t$. We say that a $C^\infty$-curve $\eta$ is a geodesic (of constant speed) if $D_{\dot{\eta}}^{\dot{\eta}} \dot{\eta} = 0$ on $(0, r)$. Note however that the reverse curve $\bar{\eta}_t = \eta_{(r-t)}$ may not be a geodesic (not even w.r.t. the reverse Finsler structure $\bar{F}$).

The exponential map is given by $exp(v) = \exp_{\pi(v)} v := \eta(1)$ if there is a geodesic $\eta : [0, 1] \to M$ with $\dot{\eta}_0 = v$. Note however, that the exponential map is only $C^1$ at the zero section. We say that $(M, F)$ is forward geodesically complete if the exponential map is define on all of $TM$, i.e. if we can extend any constant speed geodesic $\eta$ to geodesic $\eta : [0, \infty) \to M$. For such case, we can connect any two points of $M$ by a minimal geodesic, i.e. for every $x, y \in M$ there is a geodesic $\eta$ from $x$ to $y$ such that $\mathcal{L}(\eta) = d(x, y)$.

Given a unit vector $v \in T_xM$, let $r(v) \in (0, \infty]$ be the the supremum of all $r > 0$ such that $t \mapsto exp_x tv$ is a minimal geodesic. If $r(v) < \infty$ then we say that $exp_x(r(v)v)$ is a cut-point of $x$ and denote by $\mathrm{Cut}(x)$ the set of all cut points of $x$, also called the cut locus of $x$. One can show that the exponential map is a $C^\infty$-diffeomorpism from $\{tv \,|\, v \in T_xM, F(v) = 1, t \in (0, r(v))\}$ to $M \backslash (\mathrm{Cut}(x) \cup \{x\})$. This also shows that the distance $d(x, \cdot)$ is $C^\infty$ away from $x$ and the cut locus of $x$. In particular, if $L : [0, \infty) \to [0, \infty)$ is $C^\infty$ away from 0 then $L(d(x, \cdot))$ is $C^\infty$ away from $x$ and the cut locus of $x$.

A variation of a $C^\infty$-curve $\eta : [0, r] \to M$ is a $C^\infty$-map $\sigma : [0, r] \times (-\epsilon, \epsilon) \to M$ such that $\eta(t) = \sigma(t, 0)$. We abbreviate the derivatives as

$$T(t, s) = \partial_t \sigma(t, s), U(t, s) = \partial_s \sigma(t, s).$$

The first variation of the arclenth is given by

$$\frac{\partial \mathcal{L}(\sigma_s)}{\partial s} = \left[\frac{g_T(U, T)}{F(T)}\right]_{t=0}^{r} - \int_0^r g_T\left(U, D_T^T\left[\frac{T}{F(T)}\right]\right) dt.$$

where we dropped the dependency on $t$ and $s$. In case $\eta$ is a geodesic, the second term is zero. Furthermore, the second variation along a geodesic has the form

$$\left.\frac{\partial^2 \mathcal{L}(\sigma_s)}{\partial s^2}\right|_{s=0} = I(U, U) + \left[\frac{g_T(U, T)}{F(T)}\right]_{t=0}^{r} - \int_0^r \frac{1}{F(T)}\left(\frac{\partial F(T)}{\partial s}\right)^2 dt$$

where

$$I(V, W) := \frac{1}{F(\dot\eta)} \int_0^r \left\{g_{\dot\eta}(D_{\dot\eta}^{\dot\eta} V, D_{\dot\eta}^{\dot\eta} W) - g_{\dot\eta}(R^{\dot\eta}(V, \dot\eta)\dot\eta, W)\right\} dt.$$

Since the tensor $R^{\dot\eta}$ enjoys some symmetry, we easily see that $I(V, W) = I(W, V)$. And if $V$ is a Jacobi field then the second term is zero and one can show

$$I(V, W) = \frac{1}{F(\dot\eta)} \left[g_{\dot\eta}(D_{\dot\eta}^{\dot\eta} V, W)\right]_{t=0}^{r}.$$

And finally, we say that a $C^\infty$-vector field $J$ along a geodesic $\eta : [0, r] \to M$ is a Jacobi field if it satisfies

$$D_{\dot\eta}^{\dot\eta} D_{\dot\eta}^{\dot\eta} J + R^{\dot\eta}(J, \dot\eta)\dot\eta = 0.$$

Any Jacobi field can be represented as a variational vector field of some geodesic variation $\sigma$ (each $\sigma_s$ is a geodesic) and vice versa.

# 3 $c_p$-Concave Functions

Assume throughout that $M$ is a proper geodesic space.

Define for $1 < p < \infty$

$$c_p(x,y) = \frac{d^p(x,y)}{p}.$$

We say that a function $\phi : X \to \underline{\mathbb{R}}$ is proper if it is not identically -$\infty$.

*Remark.* Almost all results about $c_p$-concave functions also hold for $c_L$-concave functions by exchanging $c_p$ with $c_L$ where $L$ is a strictly convex, increasing, function differentiable in $(0,\infty)$ and

$$c_L(x,y) = L(d(x,y)).$$

If $L$ is fixed then $c_t$ will be an abbreviation for $c_{L_t}$ where $L_t(r) = L(r/t)$.

The definition of $c_p$-transform can be localized. This has the advantage to give properness of the function and Lipschitz regularity on the domain also in the non-compact setting.

**Definition 3.1** ($c_p$-transform and the subset $\mathcal{I}_p^c(X,Y)$)**.** Let $X$ and $Y$ be two subsets of $M$. The $c_p$-transform relative to $(X,Y)$ of a function $\phi : X \to \underline{\mathbb{R}}$ is defined as

$$\phi^{c_p}(y) = \inf_{x\in M} c_p(x,y) - \phi(x).$$

In case $X = Y = M$ we just write $c_p$-transform. Similarly, we define the $\bar{c}_p$-transform relative to $(Y,X)$ of a function $\psi : Y \to \underline{\mathbb{R}}$ as

$$\psi^{\bar{c}_p}(x) = \inf_{y\in Y} c_p(x,y) - \psi(y).$$

We say that a proper function $\phi : X \to \underline{\mathbb{R}}$ is $c_p$-concave (relative to $(X,Y)$) if there is a function $\psi : Y \to \underline{\mathbb{R}}$ such that $\phi = \psi^{\bar{c}_p}$. Similarly, we define $\bar{c}_p$-concave function relative to $(Y,X)$ as those proper function $\psi$ such that $\psi = \phi^{c_p}$ for some function $\phi : X \to \underline{\mathbb{R}}$.

Let $\mathcal{I}^{c_p}(X,Y)$ (resp. $\mathcal{I}^{\bar{c}_p}(Y,X)$) denote the set of all $c_p$-concave functions relative to $(X,Y)$ (resp. the set of all $\bar{c}_p$-concave functions relative to $(Y,X)$).

Note that $\mathcal{I}^{c_p}(X,Y') \subset \mathcal{I}^{c_p}(X,Y)$ for all $Y' \subset Y$. Indeed, if $\phi \in \mathcal{I}^{c_p}(X,Y')$ and $\psi' : Y' \to \underline{\mathbb{R}}$ is such that $\phi = (\psi')^{\bar{c}_p}$ then let

$$\psi(y) = \begin{cases} \psi'(y) & \text{if } y \in Y' \\ \text{-}\infty & \text{if } y \in Y\backslash Y'. \end{cases}$$

Then obviously $\phi = (\psi')^{\bar{c}_p} = \psi^{\bar{c}_p}$ and thus $\phi \in \mathcal{I}^{c_p}(X,Y)$. Similarly, if $X' \subset X$, we can extend any function $\phi \in \mathcal{I}^{c_p}(X',Y)$ to a $c_p$-concave $\phi \in \mathcal{I}^{c_p}(X,Y)$ by letting $\phi$ be the $\bar{c}_p$-transform of $\psi : Y \to \underline{\mathbb{R}}$ relative to $(Y,X)$.

The following is easy to show:

**Lemma 3.2.** *Let $\phi : M \to \mathbb{R} \cup \{-\infty\}$ and let all statement be relative to some pair $(X,Y)$ of compact subsets. Then the following holds:*

1. *$\phi \le \phi^{c_p\bar{c}_p}$ and $\phi^{c_p} = \phi^{c_p\bar{c}_pc_p}$*
2. *if $\phi$ is not identically $-\infty$ then $\phi$ is $c_p$-concave iff $\phi = \phi^{c_p\bar{c}_p}$*
3. *if $\{\phi_i\}_{i\in I} \subset \mathcal{I}^{c_p}(X,Y)$ for some index set $I$ and $\phi(x) := \inf_I \phi_i(x)$ is a proper function, then $\phi \in \mathcal{I}^{c_p}(X,Y)$.*
4. *If $\phi$ is $c_p$-concave, then it is Lipschitz continuous and its Lipschitz constant is bounded from above by a constant depending only on $X, Y$ and $p$.*

**Corollary 3.3.** *If $M$ is compact and $\phi$ is $c_p$-concave then $\phi$ is Lipschitz continuous with Lipschitz constant bounded from above by a constant only depending on $M$ and $p$. In particular, the set of $c_p$-concave functions with $\phi(x_0) = 0$ is a precompact subset of $C^0(M,\mathbb{R})$ with bounded Lipschitz constant only depending on $M$.*

Since $X$ and $Y$ are compact, the inf in the definition of $c_p/\bar{c}_p$-transform is actually achieved and the following sets are non-empty for each $c_p/\bar{c}_p$-concave functions.

**Definition 3.4** ($c_p$-superdifferential)**.** Let $X$ and $Y$ be two compact subsets of $M$ and $\phi : X \to \underline{\mathbb{R}}$ be a $c_p$-concave function relative to $(X,Y)$ then the $c_p$-superdifferential of $\phi$ at $x \in X$ is the non-empty set

$$\partial^{c_p}\phi(x) = \{y \in Y \,|\, \phi(x) = c_p(x,y) - \phi^{c_p}(y)\}.$$

Similarly, we define $\bar{c}_p$-superdifferential of a $\bar{c}_p$-concave function $\psi : Y \to \underline{\mathbb{R}}$ as the non-empty set

$$\partial^{\bar{c}_p}\psi(y) = \{x \in X \,|\, \psi(y) = c_p(x,y) - \phi^{c_p}(x)\}.$$

It is not difficult to see that

$$y \in \partial^{c_p}\phi(x) \iff x \in \partial^{\bar{c}_p}\phi^{c_p}(y)$$

whenever $\phi$ is $c_p$-concave. Furthermore, $y \in \partial^{c_p}\phi\left(\partial^{\bar{c}_p}\phi^{c_p}(y)\right)$.

**Lemma 3.5** (Semicontinuity of the $c_p$-superdifferential)**.** *Let $X, Y$ be two compact subsets of $M$ and $\phi$ be a $c_p$-concave function relative to $(X,Y)$. Then, whenever $y_n \in \partial^{c_p}\phi(x_n)$ for some sequence $(x_n,y_n) \in X \times Y$ such that $(x_n,y_n) \to (x,y)$, we have $y \in \partial^{c_p}\phi(x)$. In particular, if $\partial^{c_p}\phi(x) = \{y\}$ is single-valued, then for every neighborhood $V$ of $y$, the set $(\partial^{c_p}\phi)^{-1}(V)$ contains a neighborhood $U$ of $x$ (relative to $X$), in particular, for any $x' \in U \cap X$ there is a $y' \in \partial^{c_p}\phi(x) \cap V \cap Y$.*

*Proof.* Note that $\phi$ and $\phi^{c_p}$ are Lipschitz continuous on $X$, resp. $Y$. Since $X$ and $Y$ are closed we have $(x,y) \in X \times Y$ and hence

$$0 = \phi(x_n) + \phi^{c_p}(y_n) - c_p(x_n,y_n) \to \phi(x) + \phi^{c_p}(y) - c_p(x,y) = 0,$$

i.e. $y \in \partial^{c_p}\phi(x)$.

The second statement directly follows from the set-wise continuity of $x' \mapsto \partial^{c_p}\phi(x')$ at $x$ in case $\partial^{c_p}\phi(x)$ is single-valued. □

In case $M$ is non-compact and $X = Y = M$ we can show the following.

**Lemma 3.6.** *Let $\phi$ be a $c_p$-concave function and $\Omega \subset X$ the interior of $\{\phi > -\infty\}$. Then $\phi$ is locally bounded and locally Lipschitz on $\Omega$ and for every compact set $K \subset \Omega$ the set $\cup_{x\in K}\partial^{c_p}\phi$ is bounded and not empty.*

*Remark.* This lemma extends [GRS13, Lemma 3.3] to all cases $p \neq 2$. The same result also holds for $c_L$-concave functions if we assume that $L$ is strictly increasing and convex and satisfies the following

$$L(R) - L(R-\epsilon) \to \infty$$

as $R \to \infty$ for any $\epsilon > 0$, i.e. if $L(R) = \int_0^R l(r)dr$ with $l$ increasing and unbounded.

*Proof.* By definition $\phi = (\phi^{c_p})^{\bar{c}_p}$ and thus $\phi$ is the infimum of a family of continuous functions and therefore upper semicontinuous and locally bounded from above.

As in [GRS13], we prove that $\phi$ is locally bounded from below by contradiction. Assuming $\phi$ is not locally bounded near a point $x_\infty \in \Omega$, there is a sequence $\Omega \ni x \to x_\infty$ such that $\phi(x_n) \to -\infty$.

Furthermore, for every $n \in \mathbb{N}$ we can find $y_n \in M$ such that

$$\phi(x_n) \geq c_p(x_n, y_n) - \phi^{c_p}(y_n) - 1.$$

This immediately yields $\phi^{c_p}(y_n) \to \infty$. Because

$$\mathbb{R} \ni \phi(x_\infty) \leq c_p(x_\infty, y_n) - \phi^{c_p}(y_n),$$

we must have $c_p(x_\infty, y_n) \to \infty$, i.e. $y_n$ is an unbounded sequence. In addition, also note $c_p(x_n, y_n) \to \infty$.

So w.l.o.g. we can assume $c_p(x_n, y_n) \geq 1$. Now let $\gamma^n : [0, d(x_n, y_n)] \to M$ be a unit speed minimal geodesic between $x_n$ and $y_n$. We will show that

$$\sup_{\bar{B}_1(\gamma_1^n)} \phi \to -\infty \text{ as } n \to \infty.$$

In order to prove this, note that for $x \in \bar{B}_1(\gamma_1^n)$ we have $d(x, \gamma_1^n) \leq 1 = d(x_n, \gamma_1^n)$ and thus

$$\begin{aligned}
\phi(x) &\leq c_p(x, y_n) - \phi^{c_p}(y_n) \leq \frac{(d(x,\gamma_1^n) + d(\gamma_1^n, y_n))^p}{p} - \phi^{c_p}(y_n) \\
&\leq \frac{(d(x_n,\gamma_1^n) + d(\gamma_1^n, y_n))^p}{p} - \phi^{c_p}(y_n) \\
&= c_p(x_n, y_n) - \phi^{c_p}(y_n) \leq \phi(x_n) + 1.
\end{aligned}$$

Because $\phi(x_n) \to -\infty$, we proved our claim.

Since $M$ is proper, we can assume $\gamma_1^n \to z$ such that $d(x_\infty, z) = 1$. In addition, the claim implies that $\phi$ is identically $-\infty$ in the interior of $B_1(z)$. But this contradicts $x_\infty \in \Omega$. Therefore, $\phi$ is locally bounded in $\Omega$.

It remains to show that $\phi$ is locally Lipschitz. Choose $\bar{x} \in \Omega$ and $r > 0$ such that $B_{2r}(\bar{x}) \subset \Omega$. Choose $x \in B_r(\bar{x})$ and let $y_n$ be a sequence such that

$$\phi(x) = \lim_{n\to\infty} c_p(x, y_n) - \phi^{c_p}(y_n).$$

We will show that $y_n \in B_C(\bar{x})$ for some $C$ only depending on $\bar{x}, r$ and $\phi$. We may assume $d(x, y_n) > r$ otherwise we are done. Let $\gamma^n : [0, d(x, y_n)] \to M$ a minimal unit speed geodesic from $x$ to $y_n$. We have

$$\limsup_{n\to\infty} \phi(x) - \phi(\gamma_r^n) \geq \limsup_{n\to\infty} c_p(x, y_n) - c_p(\gamma_r^n, y_n)$$

and we know already that the left hand side is bounded. If $R_n := d(y_n, x) \to \infty$ then for $l(r) = r^{p-1}$

$$c_p(x, y_n) - c_p(\gamma_r^n, y_n) = \int_{R_n - r}^{R_n} l(s)ds \geq r \cdot l(R_n - r) \to \infty$$

which is a contradiction. Hence $y_n$ is bounded and by properness has accumulation points which all belong to $\partial^{c_p}\phi(x)$. Similarly, we can show that $\cup_{x\in K}\partial^c\phi(x)$ is bounded for any compact $K$.

Finally, for all $x \in B_r(\bar{x})$

$$\begin{aligned} \phi(x) &= \inf_{y\in M} c_p(x, y) - \phi^{c_p}(y) \\ &= \min_{B_C(\bar{x})} c_p(x, y) - \phi^{c_p}(y). \end{aligned}$$

Since for $y \in B_C(\bar{x})$ the functions $x \mapsto c_p(x, y) - \phi^{c_p}(y)$ are uniformly Lipschitz on $B_r(\bar{x})$, $\phi$ is locally Lipschitz as well. □

For $x, y \in M$ and $t \in [0, 1]$ define $Z_t(x, y) \subset M$ as

$$Z_t(x, y) := \{z \in M \mid d(x, z) = td(x, y) \text{ and } d(z, y) = (1 - t)d(x, y)\}.$$

If there is a unique geodesic between $x$ and $y$ then obviously $Z_t(x, y) = \{\gamma(t)\}$. Furthermore, for general set $X, Y \subset M$ define

$$Z_t(x, Y) := \bigcup_{y\in Y} Z_t(x, y)$$

and $Z_t(X, Y)$ as

$$Z_t(X, Y) := \bigcup_{x\in X} Z_t(x, Y).$$

The following three results are crucial ingredients to show absolute continuity of the interpolation measure in the smooth setting (see Lemma 4.17 below). It generalizes [CEMS01, Claim 2.4] and will be used in Lemma 3.8 (see [Oht09, (3.1) p. 221] for the case $p = 2$). Lemma 3.9 will also help to prove "almost everywhere" second order differentiability of $c_p$-concave functions. This proof is much easier than the original one given in [CEMS01, Oht08]. There is also a counterpart in the Orlicz-Wasserstein case which is stated and proved in the appendix (see Lemma 7.7).

**Lemma 3.7.** *If $x, y \in M$ and $z \in Z_t(x,y)$ for some $t \in [0,1]$. Then for all $m \in M$*

$$t^{p-1}d^p(m,y) \leq d^p(m,z) + t^{p-1}(1-t)d^p(x,y).$$

*Furthermore, choosing $x = m$ this becomes an equality.*

*Proof.* Using the triangle inequality, the fact that $d(z,y) = (1-t)d(x,y)$ and that $r \mapsto r^p$ is convex for $p > 1$, we get

$$\begin{aligned} t^{p-1}d^p(m,y) &\leq t^{p-1}\left\{t \cdot \frac{1}{t}d(m,z) + (1-t)d(x,y)\right\}^p \\ &\leq t^{p-1}\left\{t \cdot \left(\frac{1}{t}d(m,z)\right)^p + (1-t)d^p(x,y)\right\} \\ &= d^p(m,z) + t^{p-1}(1-t)d^p(x,y). \end{aligned}$$

Furthermore, choosing $m = x$ we see that each inequality is actually an equality. □

**Lemma 3.8.** *Let $\eta : [0,1] \to M$ be a geodesic between two distinct points $x$ and $y$. For $t \in (0,1]$ define*

$$f_t(m) := -c_p(m, \eta_t).$$

*Then for some fixed $t \in [0,1]$ the function $h(m) := f_t(m) - t^{p-1}f_1(m)$ has a minimum at $x$.*

*Proof.* Using Proposition 3.7 above we have for $z = \eta_t \in Z_t(x,y)$

$$\begin{aligned} -ph(m) = t^{p-1}d^p(m,y) - d^p(m,z) &\leq t^{p-1}(1-t)d^p(x,y) \\ &= t^{p-1}d^p(x,y) - d^p(x,\eta_t) = -ph(x). \end{aligned}$$

□

The following lemma will be useful to describe the interpolation potential of the optimal transport map. It generalizes [CEMS01, 5.1] to the cases $p \neq 2$.

**Lemma 3.9** ($c_p$-concave functions form a star-shaped set)**.** *Let $X$ and $Y$ be compact subsets of $M$ and let $t \in [0,1]$. If $\phi \in \mathcal{I}^{c_p}(X,Y)$ then $t^{p-1}\phi \in \mathcal{I}^{c_p}(X, Z_t(X,Y))$.*

*Proof.* Note that the cases $t = 0$ and $t = 1$ are trivial since $0 \in \mathcal{I}^{c_p}(X,X)$. For the rest we follow the strategy of [CEMS01, Lemma 5.1]. Let $t \in [0,1]$ and $y \in Y$ and define $\phi(x) := c_p(x,y) = d_y^p(x)/p$. We claim that the following representation holds

$$t^{p-1}d_y^p(m)/p = \inf_{z \in Z_t(X,y)}\left\{d_z^p(m)/p + \inf_{\{x \in X \,|\, z \in Z_t(x,y)\}} t^{p-1}(1-t)d_y^p(x)/p\right\}.$$

Indeed, by Lemma 3.7 the left hand side is less than or equal to the right hand side for any $z \in Z_t(X,y)$. Furthermore, choosing $x = m$ we get an equality and thus showing the representation.

Now note that the claim implies that $t^{p-1}\phi$ is the $\bar{c}_p$-transform of the function

$$\psi(z) = -\inf_{\{x\in X\,|\,z\in Z_t(x,y)\}} t^{p-1}(1-t)d_y^p(x)/p$$

(real-valued on $Z_t(X,Y)$) and therefore $t^{p-1}\phi$ is $c_p$-concave relative to $(X, Z_t(X,y))$. Since $\mathcal{I}^{c_p}(X, Z_t(X,y)) \subset \mathcal{I}^{c_p}(X, Z_t(X,Y))$ we see that each $t^{p-1}d_y^p/p$ is in $\mathcal{I}^{c_p}(X, Z_t(X,Y))$.

It remains to show that for an arbitrary $c_p$-concave function $\phi$ and $t\in[0,1]$ the function $t^{p-1}\phi$ is $c_p$-concave relative to $(X, Z_t(X,Y))$. Since $\phi = \phi^{c_p\bar{c}_p}$ we have

$$t^{p-1}\phi(x) = \inf_y t^{p-1}c_p(x,y) - t^{p-1}\phi^{c_p}(y).$$

But each function

$$\psi_y(x) = t^{p-1}c_p(x,y) - t^{p-1}\phi^{c_p}(y)$$

is $c_p$-concave relative to $(X, Z_t(X,Y))$ and $\phi$ is proper, thus also the infimum is $c_p$-concave relative to $(X, Z_t(X,Y))$, i.e. $t^{p-1}\phi \in \mathcal{I}^{c_p}(X, Z_t(X,Y))$. □

Finally, assuming the space is non-branching, e.g. a Riemannian or Finsler manifold, we want to show the well-known result that the optimal transport rays cannot intersect at intermediate times. The proof is easily adaptable to Orlicz-Wasserstein spaces and will give positivity of the Jacobian for the interpolation measures.

**Definition 3.10** (non-branching spaces)**.** A geodesic space $(M,d)$ is said to be non-branching, if for all $x,y,y'\in M$ with $d(x,y)=d(x,y')>0$ one always has

$$Z_t(x,y)\cap Z_t(x,y') \neq \varnothing \text{ for some } t\in(0,1) \implies y=y'.$$

**Lemma 3.11.** *Assume $M$ is non-branching and $\mu_0$ and $\mu_1$ two measures in $\mathcal{P}_p(M)$. If $\pi$ is an optimal transport plan between $\mu_0$ and $\mu_1$ then there is a subset $U$ of $M\times M$ of $\pi$-measure 1 such that for $i=1,2$ let $\gamma_i$ be a geodesic between $x_i, y_i \in U$, then $\gamma_1(t)=\gamma_2(t)$ for some $t\in[0,1]$ implies $(x_1,y_1)=(x_2,y_2)$.*

*Remark.* Exactly the same results hold for the optimal transport plan with cost function $L(d(\cdot,\cdot))$. In particular, it holds for Orlicz-Wasserstein spaces using [Stu11, Proposition 3.1] and $c_\lambda$-cyclicity of the support where $\lambda = w_L(\mu_0,\mu_1)$ (see appendix for definition of $w_L$).

*Proof.* According to [Vil09, Theorem 5.10] there is a subset $U$ of $M\times M$ of $\pi$-measure 1 such that for each $(x_i,y_i)\in U$

$$\frac{d(x_1,y_1)^p}{p} + \frac{d(x_2,y_2)^p}{p} \le \frac{d(x_1,y_2)^p}{p} + \frac{d(x_2,y_1)^p}{p},$$

this property is called $c_p$-cyclically monotone (of order 2) (see [Vil09, Definition 5.1]).

Now assume for some $(x_i, y_i) \in U$ there is a $t \in (0,1)$ such that we have $z = \gamma_1(t) = \gamma_2(t)$. Then

$$\begin{aligned} d(x_1,y_2)^p + d(x_2,y_1)^p &\leq (d(x_1,z)+d(z,y_2))^p + (d(x_2,z)+d(z,y_1))^p \\ &= (td(x_1,y_1)+(1-t)d(x_2,y_2))^p \\ &\quad + (td(x_2,y_2)+(1-t)d(x_1,y_1))^p \\ &\leq td(x_2,y_2)^p + (1-t)d(x_2,y_2)^p \\ &\quad + td(x_2,y_2)^p + (1-t)d(x_1,y_1)^p \\ &= d(x_1,y_1)^p + d(x_2,y_2)^p. \end{aligned}$$

Because $U$ is $c_p$-cyclically monotone we see that the inequality actually must be an equality. Since $p > 1$ we must have $d(x_1,y_1) = d(x_2,y_2)$ and

$$d(x_1,y_2)^p + d(x_2,y_1)^p = d(x_1,y_1)^p + d(x_2,y_2)^p.$$

This also implies that $d(x_1,y_2) = d(x_1,y_1) = d(x_2,y_1)$. Because $z$ is the common $t : (1-t)$ fraction point and there are no branching geodesics, we must have $x_1 = x_2$ and $y_1 = y_2$. $\square$

# 4 Optimal Transport on Finsler manifolds

In this section we will assume throughout that $M$ is a $C^\infty$-Finsler manifold. We are going to show that the interpolation inequality can be proven along $p$-Wasserstein geodesics. From this inequality and a lower Ricci curvature bound, one can easily derive the curvature dimension condition as defined in the next section. Furthermore, it turns out to be equivalent to the lower Ricci curvature bound. As Ohta [Oht09] noted, in the Finsler setting one needs additional assumptions on the background measure to get a lower (weighted) Ricci curvature bound from the curvature dimension condition.

## 4.1 Notation and technical ingredients

Let $q$ be the Hölder conjugate of $1 < p < \infty$, i.e. $\frac{1}{q}+\frac{1}{p} = 1$ or equivalently $(p-1)(q-1) = 1$.

In order to get a nice description of the interpolation maps we need to define the following $q$-gradient

$$\nabla^q \phi := |\nabla\phi|^{q-2}\nabla\phi.$$

Note that for $v \in T_xM$

$$\nabla\phi(x) = |v|^{p-2}v$$

iff

$$\nabla^q\phi = v.$$

Also note that $\nabla\phi = 0$ iff $\nabla^q\phi = 0$, and $x \mapsto \nabla^q\phi(x)$ is continuous iff $x \mapsto \nabla\phi(x)$ is. For $t > 0$ we have

$$\nabla^q(t^{p-1}\phi) = t\nabla^q\phi.$$

In addition, we use the abbreviation $\mathcal{K}d\phi = \nabla^q\phi$ (note that $\mathcal{L}d\phi = \nabla\phi$). This is indeed invertible, continuous from $T^*M \to TM$ and $C^\infty$ away from the zero section. Furthermore,

$$\mathcal{K}_x t^{p-1} d\phi_x = t\nabla^q\phi(x).$$

*Remark.* $\mathcal{K}_x$ can actually be seen as the Legendre transform from $T^*_xM \to T_xM$ that associates to each cotangent vector $\alpha \in T^*M$ the unique tangent vector $v = \mathcal{K}(\alpha) \in TM$ such that $F(v)^p = F^*(\alpha)^q$ and $\alpha(v) = F^*(\alpha)^q$ where $F^*$ denotes the dual norm of $F$ on $T^*M$.

In order to show that optimal transport is almost everywhere away from the cut locus we need to following result. Its proof is based on [Oht09, Lemma 3.1].

**Lemma 4.1** (Cut locus charaterization)**.** *If $y \neq x$ is a cut point of $x$, then $f(z) := d^p(z,y)/p$ satisfies*

$$\liminf_{v\to 0\in T_xM} \frac{f(\xi_v(1)) + f(\xi_v(-1)) - 2f(x)}{F(v)^2} = -\infty$$

*where $\xi_v : [-1,1] \to M$ is the geodesic with $\dot{\xi}_v(0) = v$.*

*Proof.* First recall that $y$ is a cut point of $x$ if either there are two minimal geodesics from $x$ to $y$, or $y$ is the first conjugate point along a unique geodesic $\eta$ from $x$ to $y$, i.e. there is a Jacobi field along $\eta$ vanishing only at $x$ and $y$ (see [BCS00, Corollary 8.2.2]).

So let's first assume there are two distinct unit speed geodesics $\eta, \zeta : [0, d(x,y)] \to M$ from $x$ to $y$ and let $v = \dot{\zeta}(0)$ and $w = \dot{\eta}(0)$. For fixed small $\epsilon > 0$ set $y_\epsilon = \eta(d(x,y) - \epsilon)$ then $y_\epsilon \notin \mathrm{Cut}(x) \cup \{x\}$ and using the first variation formula we get for $t > 0$

$$\begin{aligned} f(\xi_v(-t)) - f(x) &\leq \{d(\xi_v(-t), y_\epsilon) + \epsilon\}^p/p - \{d(x, y_\epsilon) + \epsilon\}^p/p \\ &= t\{d(x,y_\epsilon) + \epsilon\}^{p-1} g_{\dot{\eta}(0)}(v, \dot{\eta}(0)) + \mathcal{O}(t^2) \\ &= td^{p-1}(x,y) g_{\dot{\eta}(0)}(v, \dot{\eta}(0)) + \mathcal{O}(t^2). \end{aligned}$$

The term $\mathcal{O}(t^2)$ is ensured by smoothness of $\xi_v$ and by the fact that $x \neq y_\epsilon$. We also get by the Taylor formula

$$f(\xi_v(t)) - f(x) = \{d(x,y) - t\}^p/p - d^p(x,y)/p = -td^{p-1}(x,y) + \mathcal{O}(t^2).$$

Combining these two facts with $g_w(v,w) < 1$ ($\eta$ and $\xi$ are distinct), we get

$$\frac{f(\xi_v(-t)) + f(\xi_v(t)) - 2f(x)}{t^2} \leq \frac{1 - g_w(v,w)}{t} d^{p-1}(x,y) + t^{-2}\mathcal{O}(t^2) \to -\infty \text{ as } t \to 0.$$

Next we will treat the case that $y$ is the first conjugate point of $x$ along a unique minimal geodesic $\eta : [0,1] \to M$ from $x$ to $y$. By definition, let $J$ be a Jacobi field along $\eta$ vanishing only at $x$ and $y$. For $v = D^{\dot{\eta}}_{\dot{\eta}} J(0) \in T_xM\backslash\{0\}$ let $V_1$ be the parallel vector field along $\eta$ (i.e. $D^{\dot{\eta}}_{\dot{\eta}} V_1 \equiv 0$) such that $V_1(0) = v$. Furthermore, define for $t \in [0,1]$ the vector field $V(t) := (1-t)V_1(t)$ and $J_\epsilon = J + \epsilon V$ for small $\epsilon > 0$. Note that $J_\epsilon(0) = \epsilon v$ and $J_\epsilon(1) = 0$, and since $g_{\dot{\eta}(0)}(v,v) > 0$ also $J_\epsilon \neq 0$ on $[0,1)$ for sufficiently small $\epsilon > 0$.

We define a variation $\sigma : [0,1] \times [-1,1] \to M$ by $\sigma(t,s) = \sigma_s(t) := \xi_{J_\epsilon(t)}(s)$. Because $J_\epsilon \neq 0$ on $[0,1)$ this variation is $C^\infty$ on $(0,1) \times (-1,1)$. According to the second variation formula we get (see [Oht09, Proof of 3.1])

$$\left.\frac{\partial^2 \mathcal{L}(\sigma_s)}{\partial s^2}\right|_{s=0} = I(J_\epsilon, J_\epsilon) - \frac{g_{\dot{\eta}}(D^{\dot{\eta}}_{J_\epsilon} J_\epsilon, \dot{\eta})}{d(x,y)} - \frac{1}{d(x,y)} \int \left\{ \frac{\partial F(\partial_t \sigma)}{\partial s}(t) \right\}^2 dt$$

where $\mathcal{L}$ is the length functional

$$\mathcal{L}(\sigma_s) = \mathrm{length}(\sigma_s(\cdot)).$$

By definition of tangent curvature $\mathcal{T}$ (see [Oht09]), we have

$$\mathcal{T}_{\dot\eta(0)}(v) = g_{\dot\eta}(D^v_v v - D^{\dot\eta}_v v, \dot\eta) = \epsilon^{-2} g_{\dot\eta(0)}(D^{J_\epsilon}_{J_\epsilon} J_\epsilon - D^{\dot\eta}_{J_\epsilon} J_\epsilon, \dot\eta) = -\epsilon^{-2} g_{\dot\eta(0)}(D^{\dot\eta}_{J_\epsilon} J_\epsilon, \dot\eta)$$

where the last equality follows from the fact that $\sigma_0 = \xi_{J_\epsilon(0)}$ is a geodesic. Combining these we get

$$\begin{aligned} \left.\frac{\partial^2 \mathcal{L}(\sigma_s)}{\partial s^2}\right|_{s=0} &\le I(J,J) + 2\epsilon I(J,V) + \epsilon^2 I(V,V) + \epsilon^2 \mathcal{T}_{\dot\eta(0)}(v)/d(x,y) \\ &= \left\{ \left[ g_{\dot\eta}(D^{\dot\eta}_{\dot\eta} J, J)\right]_{t=0}^1 + 2\epsilon \left[ g_{\dot\eta}(D^{\dot\eta}_{\dot\eta} J, V)\right]_{t=0}^1 + \epsilon^2 \mathcal{T}_{\dot\eta(0)}(v) \right\} / d(x,y) + \epsilon^2 I(V,V) \\ &= \left\{ -2\epsilon g_{\dot\eta(0)}(v,v) + \epsilon^2 \mathcal{T}_{\dot\eta(0)}(v)\right\} / d(x,y) + \epsilon^2 I(V,V). \end{aligned}$$

Furthermore, note by the first variations formula and the fact that $\sigma_0$ is a geodesic

$$\left.\frac{\partial \mathcal{L}(\sigma_s)}{\partial s}\right|_{s=0} = [g_{\dot\eta}(J_\epsilon, \dot\eta)]_{t=0}^1 = [\epsilon g_{\dot\eta}(V, \dot\eta)]_{t=0}^1 = -\epsilon g_{\dot\eta(0)}(v, \dot\eta(0)) \ge -\epsilon F(v).$$

So that we get

$$\begin{aligned} \lim_{s\to 0} \frac{\mathcal{L}(\sigma_s)^p + \mathcal{L}(\sigma_{-s})^p - 2\mathcal{L}(\sigma_0)^p}{s^2} &= p\mathcal{L}^{p-2}(\sigma_0) \Bigg[ \mathcal{L}(\sigma_0) \left.\frac{\partial^2}{\partial s^2} L(\sigma_s)\right|_{s=0} \\ &\qquad + (p-1)\left( \left.\frac{\partial \mathcal{L}(\sigma_s)}{\partial s}\right|_{s=0}\right)^2 \Bigg] \\ &\le pd^{p-2}(x,y)\Bigg( -2\epsilon g_{\dot\eta}(v,v) \\ &\qquad + \epsilon^2 \left\{ \mathcal{T}_{\dot\eta(0)}(v) + d(x,y) I(V,V) + (p-1)F(v)^2 \right\} \Bigg). \end{aligned}$$

Using the fact that $f(\xi_v(\epsilon s)) \le \mathcal{L}(\sigma_s)^p/p$ we obtain

$$\begin{aligned} \liminf_{s\to 0} \frac{f(\xi_v(\epsilon s)) + f(\xi_v(-\epsilon s)) - 2f(x)}{\epsilon^2 s^2} &\le \liminf_{s\to 0} \frac{\mathcal{L}(\sigma_s)^p + \mathcal{L}(\sigma_{-s})^p - 2\mathcal{L}(\sigma_0)^p}{p\epsilon^2 s^2} \\ &\le d^{p-2}(x,y)\Bigg( -2\epsilon^{-1} g_{\dot\eta}(v,v) \\ &\qquad + \mathcal{T}(v) + d(x,y) I(V,V) + (p-1)F(v)^2 \Bigg). \end{aligned}$$

Letting $\epsilon$ tend to zero completes the proof. □

## 4.2 The Brenier-McCann-Ohta solution

The first step to prove the interpolation inequality is showing the existence of a transport map. This was first done by Brenier [Bre91] in the Euclidean setting and later by McCann [McC01] for Riemannian manifolds and any cost function $c_L$. Later Ohta proved it for Finsler manifolds for the cost function $c_2$. The proof easily adapts to any $p \in (1, \infty)$.

**Lemma 4.2.** *Let $\phi : M \to \mathbb{R}$ be a $c_p$-concave function. If $\phi$ is differentiable at $x$ then $\partial^{c_p}\phi(x) = \{\exp_x(\nabla^q(-\phi)(x))\}$. Moreover, the curve $\eta(t) := \exp_x(t\nabla^q(-\phi)(x))$ is a unique minimal geodesic from $x$ to $\exp_x(\nabla^q(-\phi)(x))$.*

*Proof.* Let $y \in \partial^{c_p}\phi(x)$ be arbitrary and define $f(z) := c_p(z,y) = d^p(z,y)/p$. By definition of $\partial^{c_p}\phi(x)$ we have for any $v \in T_xM$

$$f(\exp_x v) \geq \phi^{c_p}(y) + \phi(\exp_x v) = f(x) - \phi(x) + \phi(\exp_x v) = f(x) + d\phi_x(v) + o(F(v)).$$

Now let $\eta : [0, d(x,y)] \to M$ be a minimal unit speed geodesic from $x$ to $y$. Given $\epsilon > 0$, set $y_\epsilon = \eta(d(x,y) - \epsilon)$ and note that $\eta|_{[0,d(x,y)-\epsilon]}$ does not cross the cut locus of $x$. By the first variation formula, we have

$$\begin{aligned} f(\exp_x v) - f(x) &\leq \frac{1}{p}\{(d(\exp_x v, y_\epsilon) + \epsilon)^p - (d(x, y_\epsilon) + \epsilon)^p\} \\ &= -(d(x,y_\epsilon) + \epsilon)^{p-1}\, g_{\dot\eta(0)}(v, \dot\eta(0)) + o(F(v)). \\ &= -d^{p-1}(x,y)\mathcal{L}_x^{-1}(\dot\eta(0))(v) + o(F(v)). \end{aligned}$$

Therefore, $d\phi_x(v) \leq -d^{p-1}(x,y)\mathcal{L}_x^{-1}(\dot\eta(0))(v)$ for all $v \in T_xM$ and thus $\nabla(-\phi) = d^{p-1}(x,y)\cdot\dot\eta(0).$, i.e. $\nabla^q(-\phi) = d(x,y)\cdot\dot\eta(0)$. In addition, note that $\eta(t) = \exp_x(t\nabla^q(-\phi)(x))$, which is uniquely defined. □

Let $\mathrm{Lip}_{c_p}(X,Y)$ be the set of pairs of Lipschitz function tuples $\phi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$ such that

$$\phi(x) + \psi(y) \leq c_p(x,y).$$

**Lemma 4.3.** *Let $\mu_0$ and $\mu_1$ be two probability measures on $M$. Then there exists a unique (up to constant) $c_p$-concave function $\phi$ that solves the Monge-Kantorovich problem with cost function $c_p$. Moreover, if $\mu_0$ is absolutely continuous, then the vector field $\nabla^q(-\phi)$ is unique among such minimizers.*

*Proof.* Note that if $(\phi,\psi) \in \mathrm{Lip}_{c_p}(X,Y)$ then $(\phi^{c_p\bar{c}_p}, \phi^{c_p}) \in \mathrm{Lip}_{c_p}(X,Y)$ and $\phi^{c_p} \geq \psi$ and $\phi^{c_p\bar{c}_p} \geq \phi$.

Now fix some $x_0 \in X$ and let $\{(\phi_n,\psi_n)\}_{n\in\mathbb{N}} \subset \mathrm{Lip}_{c_p}(X,Y)$ be a maximizing sequence of the Kantrovich problem. By the remark just stated, it is easy to see that also $(\hat\phi_n, \hat\psi_n) = (\phi_n^{c_p\bar{c}_p} - \phi_n^{c_p\bar{c}_p}(x_0), \phi_n^{c_p} - \phi_n^{c_p\bar{c}_p}(x_0))$ is maximizing and in addition $\phi_n^{c_p}$ is $c_p$-concave. Since the sequence has uniform bound on the Lipschitz constant and $\hat\phi_n(x_0) = 0$, the sequence is precompact and thus we can assume w.l.o.g. that $(\hat\phi_n)_{n\in\mathbb{N}}$ converges to a Lipschitz function $\phi : X \to \mathbb{R}$. By similar arguments, we can also assume that $(\hat\psi_n)_{n\in\mathbb{N}}$ converges to a function $\psi : Y \to \mathbb{R}$. In addition, note that $\phi^{c_p} = \psi$ and that because each $\hat\phi_i$ is $c_p$-concave also $\phi$ is, in particular, a solution of the Monge-Kantorovich problem exists and each solution is a pair $(\phi, \phi^{c_p}) \in \mathrm{Lip}_{c_p}(X,Y)$.

It remains to show that this solution is unique: Let $(\phi_1,\psi_1), (\phi_2,\psi_2) \in \mathrm{Lip}_{c_p}(X,Y)$ be two solutions of the problem. Now setting $\phi = (\phi_1+\phi_2)/2$, we see that $\phi^{c_p} \geq (\phi_1^{c_p}+\phi_2^{c_p})/2$ and thus $(\phi,\phi^{c_p}) \in \mathrm{Lip}_{c_p}(X,Y)$ and hence, by maximality, $\phi^{c_p} = (\phi_1^{c_p} + \phi_2^{c_p})/2$ and $\phi$ is $c_p$-concave.

Now if $y \in \partial^{c_p}\phi(x)$ then $y \in \partial^{c_p}\phi_1(x) \cap \partial^{c_p}\phi_2(x)$. Thus, using Lemma 4.2 above and the absolute continuity of $\mu_0$ we see that

$$\nabla^q\phi(x) = \nabla^q\phi_i(x) \qquad \mu_0\text{-almost every } x \in X.$$

□

**Theorem 4.4.** *Let $\mu_0$ and $\mu_1$ be two probability measure on $M$ and assume $\mu_0$ is absolutely continuous with respect to $\mu$. Then there is a $c_p$-concave function $\phi$ such that $\pi = (\mathrm{Id} \times \mathcal{F})_*\mu_0$ is the unique optimal coupling of $(\mu_0, \mu_1)$, where $\mathcal{F}(x) = \exp_x(\nabla^q(-\phi))$. Moreover, $\mathcal{F}$ is the unique optimal transport map from $\mu_0$ to $\mu_1$.*

*Remark.* The proof follows line by line from [Oht09, Theorem 4.10]. For convenience, we include the whole proof.

*Proof.* Let $\phi$ be given by the Lemma above. Define $\mathcal{F}(x) = \exp_x(\nabla^q(-\phi))$ for all points where $\phi$ is differentiable. Since $\mu_0$ is absolutely continuous, $\mathcal{F}$ is well-defined and continuous on some $\Omega$ with $\mu_0(\Omega) = 1$, in particular it is measurable.

Now let $h$ be a continuous function and put $\psi_\epsilon = \phi^{c_p} + \epsilon h$ for $\epsilon \in \mathbb{R}$ close to 0. Let $x \in M$ be arbitrary, then we can find $y_\epsilon \in M$ such that

$$(\psi_\epsilon)^{\bar{c}_p}(x) = c_p(x, y_\epsilon) - \psi_\epsilon(y_\epsilon).$$

Furthermore, whenever $\phi$ is differentiable at $x$ then $y_\epsilon$ converges to $y_0 = \mathcal{F}(x)$. In addition, we have

$$\begin{aligned} \phi(x) - \epsilon h(y_\epsilon) &\leq c_p(x, y_\epsilon) - \phi^{c_p} - \epsilon h(y_\epsilon) = (\psi_\epsilon)^{\bar{c}_p}(x) \\ &\leq c_p(x, \mathcal{F}(x)) - \psi_\epsilon(\mathcal{F}(x)) = \phi(x) - \epsilon h(\mathcal{F}(x)) \end{aligned}$$

and thus $(\psi_\epsilon)^{\bar{c}_p}(x) = \phi(x) - \epsilon h(\mathcal{F}(x)) + o(|\epsilon|)$ and $|o(|\epsilon|)| \leq 2\epsilon\|h\|_\infty$.

Now set $J(\epsilon) = \int (\psi_\epsilon)^{\bar{c}_p} d\mu + \int \psi_\epsilon d\nu$ and by maximality of $(\phi, \phi^{c_p})$ we have

$$0 = \lim_{\epsilon \to 0} \frac{J(\epsilon) - J(0)}{\epsilon} = -\int h d\mathcal{F}_*\mu_0 + \int h d\mu_1$$

and hence $\mathcal{F}_*\mu_0 = \mu_1$.

Obviously we have for $\pi_\phi := (\mathrm{Id} \times \mathcal{F})_*\mu_0$ that $c_p(x, y) = \phi(x) + \phi^{c_p}(y)$ holds $\pi_\phi$-almost everywhere and thus $\int c_p d\pi_\phi = \int \phi d\mu_0 + \int \phi^{c_p} d\mu_1$, which implies that $\pi_\phi$ is optimal. Conversely, if $\pi$ is an optimal coupling of $(\mu_0, \mu_1)$ then $c_p(x, y) = \phi(x) + \phi^{c_p}(y)$ holds $\pi$-almost everywhere, therefore $\pi\left(\bigcup_{x\in M}(x, \mathcal{F}(x))\right) = 1$ which implies $\pi = \pi_\phi$. □

**Corollary 4.5.** *If $\mu_0$ is absolutely continuous and $\phi$ is $c_p$-concave, then the map $\mathcal{F}(x) := \exp_x(\nabla^q(-\phi))$ is the unique optimal transport map from $\mu_0$ to $\mathcal{F}_*\mu_0$.*

Furthermore, we will see in Lemma 4.17 below that the interpolation measures are absolutely continuous if $\mu_0$ and $\mathcal{F}_*\mu_0$ are.

**Corollary 4.6.** *If $\phi$ is $c_p$-concave and $\mu_0$ is absolutely continuous, then the map $\mathcal{F}_t(x) := exp_x(\nabla^q(-t^{p-1}\phi))$ is the unique optimal transport map from $\mu_0$ to $\mu_t = (\mathcal{F}_t)_*\mu_0$ and $t \mapsto \mu_t$ is a constant geodesic from $\mu_0$ to $\mu_1$ in $\mathcal{P}_p(M)$.*

*Proof.* We only need to show that

$$w_p(\mu_s, \mu_t) \le |s-t| w_p(\mu_0, \mu_1).$$

Let $\pi$ be the plan on $\mathrm{Geo}(M) = \{\gamma : [0,1] \to M \,|\, \gamma \text{ is a geodesic in } M\}$ give by $\mu_0$, the map $\mathcal{F}$ and the unique geodesic connecting $\mu$-almost every $x \in M$ to a point $\mathcal{F}_1(x)$ (see e.g. [Lis06, Theorem 4.2] and [Vil09, Chapter 7]), in particular, $\mu_t = (\mathcal{F}_t)_*\mu_0$. Since $(e_s, e_t)_*\pi$ is a plan between $\mu_s$ and $\mu_t$ for $s, t \in [0,1]$, we have

$$\begin{aligned} w_p(\mu_s, \mu_t) &\le \left(\int d^p(\gamma_s, \gamma_t) d\pi(\gamma)\right)^{1/p} \\ &= |s-t| \left(\int d^p(\gamma_0, \gamma_1) d\pi(\gamma)\right)^{1/p} = |s-t| w_p^p(\mu_0, \mu_1). \end{aligned}$$

□

## 4.3 Almost Semiconcavity of $c_p$-concave functions

This section will be one of the main ingredients to show Theorem 4.16. In [Oht08] Ohta showed that every $c_p$-concave function is almost everywhere twice differentiable. He proved this by showing the the square of the distance function $d_y^2 = d^2(\cdot, y)$ for fixed $y \in M$ is semiconcave [Oht08, Corollary 4.4] and extending the Alexandrov-Bangert Theorem [Oht08, Theorem 6.6] to Finsler manifolds.

**Theorem 4.7** (Alexandrov-Bangert [Vil09, 14.1] [Oht08, 6.6])**.** *Let $M$ be a smooth symmetric Finsler manifold, then every function $\phi : M \to \mathbb{R}$ which is locally semiconvex in some open subset $U$ of $M$ is almost everywhere twice differentiable in $U$.*

Even though for general $1 < p < \infty$ we cannot show that every $c_p$-concave function is semiconcave, we show that almost all points $x$ of differentiability of a $c_p$-concave function $\phi$ with $d\phi_x \neq 0$ are twice differentiable.

We will show this result in two different ways: The first one will show, similar to [Oht08], that $d_x^p$ is semiconcave on $M \backslash \{x\}$. In the second one we will use Lemma 3.9 and the fact that $d^p(\cdot, y)$ is $C^\infty$ in $U \backslash \{y\}$ for some sufficiently small neighborhood $U$ of $y$. In both cases we conclude that $\phi$ is semi-convex in an open set of full measure relative to $M \backslash \{x \in M \,|\, x \in \partial^{c_p}\phi(x)\}$.

**Proposition 4.8** (Almost Smoothness of $d_x^p$)**.** *Let $M$ be a connected, forward geodesically complete $C^\infty$-Finsler manifold. Then for any distinct points $x$ and $y$ in $M$ and $w \in T_yM$ with $F(w) = 1$ we have*

$$\limsup_{s \to 0} \frac{1}{2s^2} \{d_x^p(\xi_w(-s)) + d_x^p(\xi_w(s)) - 2d_x^p(y)\} \le C < \infty$$

*where* $\xi_w : (-\epsilon, \epsilon) \to M$ *is a geodesic with* $\dot{\xi}_w(0) = w$ *and* $d_x = d(x, \cdot)$*. Furthermore,* $C > 0$ *can be chosen uniformly in a neighborhood of* $x$ *and* $y$*. In a similar way it is possble to prove*

$$\limsup_{s\to 0} \frac{1}{2s^2}\left\{d_y^p(\xi_v(-s)) + d_y^p(\xi_v(s)) - 2d_y^p(x)\right\} \le C < \infty.$$

*Remark.* In [Oht08, Theorem 4.2,5.1] showed that the same holds for $d^2(x, \cdot)$. He used the semiconcavity (see below) to show that $c_2$-concave functions are semiconcave [Oht08, Theorem 7.4]. Note that the second statement follows by using the inverse Finsler structure.

*Proof.* We will only indicate where Ohta's proof needs adjustment:

Set $r = d(x, y)$ and note that if $\sigma : (-\epsilon, \epsilon) \times [0, r]$ is a $C^\infty$-variation such that $\sigma_0 = \sigma(0, \cdot)$ is a minimal, unit speed geodesic between $x$ and $y$ then the first variation of the length function $s \mapsto L(\sigma_s)$ at $s = 0$ has the following form

$$\frac{\partial}{\partial s} L(\sigma_s)|_{s=0} = g_{\dot{\sigma}_0(r)}(U(r), \dot{\sigma}_0(r)) - g_{\dot{\sigma}_0(0)}(U(0), \dot{\sigma}_0(0)),$$

where

$$U(t) = \frac{\partial}{\partial s}\sigma|_{s=0}.$$

Note because $\mathcal{L}_x^{-1}(v)(w) = g_v(w, v)$ for every $v, w \in T_xM$ (see [Oht09, p.215]) we easily see that

$$g_{\dot{\sigma}_0(t)}(U(t), \dot{\sigma}_0(t))$$

is bounded as long as $U(t)$ is.

Now let $\sigma$ be a $C^\infty$-variation such that $\sigma_0$ is a geodesic then

$$\begin{aligned}
\frac{\partial^2}{\partial s}\left(L(\sigma_s)\right)^p|_{s=0} &= p\left(L(\sigma_0)\right)^{p-2}\left\{L(\sigma_0)\frac{\partial^2}{\partial s}L(\sigma_s)|_{s=0} + (p-1)\frac{\partial}{\partial s}L(\sigma_s)|_{s=0}\right\} \\
&= pr^{p-2}\left\{r\frac{\partial^2}{\partial s}L(\sigma_s)|_{s=0} + \frac{\partial}{\partial s}L(\sigma_s)|_{s=0} + (p-2)\frac{\partial}{\partial s}L(\sigma_s)|_{s=0}\right\} \\
&= pr^{p-2}\left\{I_1 + (p-2)I_2\right\}.
\end{aligned}$$

In the first case we choose the variation as in [Oht08, Theorem 4.2]) to get the following bound of $I_1$:

$$\frac{S^2 r\sqrt{k}\cosh(\sqrt{k}r)}{\sinh(\sqrt{k}r)} + r\delta$$

where $\delta > 0$, $S \ge 1$ and $k > 0$ can be chosen uniformly in a neighborhood of $(x, y)$ depending only on $(M, F)$. Note that $-k$ represents a lower bound on the flag curvature which is locally bounded by $C^\infty$-regularity of the Finsler structure. Since the norm of $U(t)$ is bounded we easily see that the second term $I_2$ is bounded from below and above.

Furthermore, we note that this bound can be chosen unformly in a neighborhood of $y$. Thus, we see that $pr^{p-2}\{I_1 + (p-2)I_2\}$ is bounded and noting that

$$\limsup_{s\to 0} \frac{1}{2s^2}\{d_x^p(\xi_v(-s)) + d_x^p(\xi_v(s)) - 2d_x^p(y)\} \quad \le \quad \frac{\partial^2}{\partial s}(L(\sigma_s))^p\,|_{s=0}$$

we can conclude the proof. □

Following the proof of [Oht08, Corollary 4.4] we immediately get the following.

**Corollary 4.9.** *For every distict $x, y \in M$ there is a ball neighborhood $B_R(x)$ of $x$ with $R = R(x, y)$ such that $d_y^p$ satisfies*

$$d_y^p(\gamma(t)) \ge (1-t)d_y^p(\gamma(0)) + td_y^p(\gamma(1)) - C(1-t)td^2(\gamma(0), \gamma(1))$$

*for any $\gamma(0), \gamma(1) \in B_{\frac{R}{2}}(x)$ and some $K = K(x, y)$. In particular, each $d_y^p$ is locally semi-convex on $M\backslash\{y\}$.*

*Remark.* Note that we can actually prove that there are neighborhoods $U$ of $x$ and $V$ of $y$ such that each $d_{y'}^p$ is semiconcave on $V$ with constant $C$.

*Proof.* Let $U$ be a neighborhood such that the previous lemma holds for some constant $C > 0$. Then, in particular, $V$ contains a ball $B_R(x)$.

Now let $\eta : [0, 1] \to M$ be a minimal geodesic between two arbitrary point $z, z' \in B_{\frac{R}{2}}(x)$. Then $\eta(t) \in B_R(x)$ and therefore for each $\tau \in (0, 1)$

$$\limsup_{\epsilon\to 0} \frac{1}{2\epsilon^2}\{d^p(\eta(\tau-\epsilon), y) + d^p(\eta(\tau-\epsilon), y) - 2d^p(\eta(\tau), y)\} \le C \cdot d^2(z, z').$$

Thus $\tau \mapsto d^p(\eta(\tau), y) - \tau^2 C d^2(z, z')$ is concave on $[0, 1]$ and hence we have

$$d^p(\eta(\tau), y) - \tau^2 C d^2(z, z') \ge (1-\tau)d^p(z, y) + \tau\left\{d^p(z', y) - Cd^2(z, z')\right\}.$$

The desired inequality is obtained after rearrangement. □

Finally, we can prove that any $c_p$-concave function is semiconcave away from the points with $x \in \partial^{c_p}\phi(x)$.

**Theorem 4.10** (Almost semiconcavity of $c_p$-concave functions)**.** *Assume $M$ is a connected, forward geodesically complete $C^\infty$-Finsler manifold. Let $Y \subset M$ be compact and $U$ be an open, precompact set. Then any $c_p$-concave function $\phi : \mathrm{Cl}\,U \to \mathbb{R}$ relative to $(\mathrm{Cl}\,U, Y)$ is locally semiconcave in $U\backslash\Omega_{id}$ where $\Omega_{id} = \{x \in \mathrm{Cl}\,U \,|\, x \in \partial^{c_p}\phi(x)\}$. In particular, $\phi$ is twice differentiable almost everywhere $U\backslash\Omega_{id}$. The same result holds for $\bar{c}_p$-concave functions.*

*Remark.* Note by continuity of $x \mapsto \partial^{c_p}\phi(x)$ the set $\Omega_{id}$ is closed and $U\backslash\Omega_{id}$ contains all point $x$ of differentiability of $\phi$ with $d\phi_x \neq 0$.

*Proof.* The second statement immediately follows from Ohta's extension of the Alexandrov-Bangert theorem [Oht08, Theorem 6.6]. Thus we only need to show that $\phi$ is semiconcave on $U\backslash\Omega_{id}$.

Choose some $x \in U$ with $y \in \partial^{c_p}\phi(x)$. Then for some sufficiently small neighborhood $B_R(x)$ of $x$ and and $V'$ of $y$ the functions $\{d^p_{y'} = d^p(\cdot, y')\}_{y' \in V'}$ are semiconcave on $U'$ with some constant $C > 0$. By continuity of $z \mapsto \partial^{c_p}\phi(z)$ at $x$ (see Lemma 3.5) we can also assume that $\partial^{c_p}\phi(z) \cap V'$ is non-empty for every $z \in B_R(x)$.

Now let $\gamma : [0,1] \to B_{2\epsilon}(x)$ be a minimal geodesic and set $x_t = \gamma(t)$. Choose $y_t \in \partial^{c_p}\phi_s(x_t) \cap V_1$. By the definition of $c_p$-concavity we have

$$\begin{aligned}
\phi_s(x_0) &\leq \phi_s(x_t) + \frac{1}{p}d^p(x_0, y_t) - \frac{1}{p}d^p(x_t, y_t) \\
\phi_s(x_1) &\leq \phi_s(x_t) + \frac{1}{p}d^p(x_1, y_t) - \frac{1}{p}d^p(x_t, y_t).
\end{aligned}$$

Further, because $y_t \in V_1$ we also have

$$d^p(x_t, y_t) \geq (1-t)d^p(x_0, y_t) + td^p(x_1, y_t) - C(1-t)td^2(x_0, x_1).$$

Therefore, taking the $(1-t), t$ convex combination of the first two inequality we obtain

$$\begin{aligned}
\phi_s(x_t) &\geq (1-t)\phi_s(x_0) + t\phi_s(x_1) + \frac{d^p(x_t, y_t)}{p} - (1-t)\frac{d^p(x_0, y_t)}{p} - t\frac{d^p(x_1, y_t)}{p} \\
&\geq (1-t)\phi_s(x_0) + t\phi_s(x_1) - \frac{C}{p}(1-t)td^2(x_0, x_1).
\end{aligned}$$

□

For the alternative proof note the following: If the Finsler metric $F$ is $C^\infty$ then the function $d^p_y(z) = d(z, y)^p$ is $C^\infty$ in $U_y\backslash\{y\}$ for some sufficiently small neighborhood $U_y$ of $y$. This follows from smoothness of the exponential map $exp_y$ in $V\backslash\{0\} \subset T_yM$ for some neighborhood $V$ of $0_x \in T_xM$ , see [She97, p. 315]. In particular, for $x \in U_y\backslash\{y\}$ we can choose a small neighborhood $U_1 \subset U$ of $x$ and an open set $V_1 \subset U$ disjoint from $U_1$ such that $\{d^p_{y'} : U_1 \to \mathbb{R}\}_{y' \in V_1}$ are uniformly bounded in $C^2$, in particular the functions are uniformly semiconcave. In addition, note that since $M$ is compact, $U_y$ can be chosen to contain a ball $B_{r_{min}}(y)$ where $r_{min} > 0$ can be chosen locally uniformly on $M$, in case $M$ is compact even uniformly.

*Remark.* Note that we only need a $C^2$-bounds so that $F$ only needs to be locally $C^2$. Also note that the same argument holds for any convex function of the distance which is smooth enough away from the origin. Furthermore, the theorem below holds for any $c_L$-concave function if Lemma 7.9 is used instead of Lemma 3.9.

**Theorem 4.11.** *Let $\phi$ be a $c_p$-concave function. Let $\Omega_{id}$ be the set of points $x \in M$ where $\phi$ is differentiable and $d\phi_x = 0$, or equivalently $\partial^{c_p}\phi(x) = \{x\}$. Then $\phi$ is locally semiconcave on an open subset $U \subset M\backslash\Omega_{id}$ of full measure (relative to $M\backslash\Omega_{id}$). In particular, it is second order differentiable almost everywhere in $U$.*

*Proof.* Since $\partial^{c_p}\phi(x)$ is non-empty for every $x \in M$ and semicontinuous in $x$, we have the following: if $\phi$ is differentiable in $x$ with $d\phi_x \neq 0$ then $x \in \mathrm{int}(M \backslash \Omega_{id})$. Thus it suffices to show that each such point has a neighborhood $U_1$ in which $\phi$ is uniformly semiconcave.

So fix such an $x$ with $d\phi(x) \neq 0$ and note that $\phi$ is semiconcave on $U_1$ iff $\lambda\phi$ is for an arbitrary $\lambda > 0$. Furthermore, by Lemma 3.9 we know that $\phi_s = s^{p-1}\phi$ is $c_p$-concave for any $s \in [0,1]$.

Since $d\phi(x) \neq 0$, there is a unique $y \in M$ with $\partial^{c_p}\phi(x) = \{y\}$ and a unique geodesic $\eta : [0,1] \to M$ between $x$ and $y$ (see Lemma 4.2). Also note that $\phi_s$ is differentiable at $x$ and

$$\partial^{c_p}\phi_s(x) = \{\eta(s)\}.$$

Let $s \in [0,1]$ be such that $d(x, \eta(s)) < \frac{r_{min}}{2}$. Because $x \neq \eta(s)$ and $z \mapsto \partial^{c_p}\phi_s(z)$ is continuous and single-valued at $x$, we can find a neighborhood $V_1 \subset U$ of $y$ such that $(\partial^{c_p}\phi_s)^{-1}(V_1) \cap U$ contains some ball $B_{2\epsilon}(x)$ disjoint from $V_1$. Thus the functions $\{d_y^p : B_{2\epsilon}(x) \to \mathbb{R}\}_{y \in V_1}$ are semiconcave with constant $C$.

Now let $\gamma : [0,1] \to B_{2\epsilon}(x)$ be a minimal geodesic and set $x_t = \gamma(t)$. Choose $y_t \in \partial^{c_p}\phi_s(x_t) \cap V_1$. By the definition of $c_p$-concavity we have

$$\begin{aligned}
\phi_s(x_0) &\leq \phi_s(x_t) + \frac{1}{p}d^p(x_0, y_t) - \frac{1}{p}d^p(x_t, y_t) \\
\phi_s(x_1) &\leq \phi_s(x_t) + \frac{1}{p}d^p(x_1, y_t) - \frac{1}{p}d^p(x_t, y_t).
\end{aligned}$$

Further, because $y_t \in V_1$ we also have

$$d^p(x_t, y_t) \geq (1-t)d^p(x_0, y_t) + td^p(x_1, y_t) - C(1-t)td^2(x_0, x_1).$$

Therefore, taking the $(1-t), t$ convex combination of the first two inequality we obtain

$$\begin{aligned}
\phi_s(x_t) &\geq (1-t)\phi_s(x_0) + t\phi_s(x_1) + \frac{d^p(x_t, y_t)}{p} - (1-t)\frac{d^p(x_0, y_t)}{p} - t\frac{d^p(x_1, y_t)}{p} \\
&\geq (1-t)\phi_s(x_0) + t\phi_s(x_1) - \frac{C}{p}(1-t)td^2(x_0, x_1).
\end{aligned}$$

□

## 4.4 Volume distortion

In order to describe the interpolation density, one needs to have a proper definition of determinant of the differential of the transport map. We follow Ohta's idea to describe the volume distortion as a proper replacement.

If $Q : T_xM \to T_yM$ we define $\mathbf{D}[Q] = \mu_y(Q(A))/\mu_x(A)$ where $\mu_x$ and $\mu_y$ are the measure on $T_xM$ induced by $\mu$ and $A$ is a nonempty, open and bounded Borel subset of $T_xM$. Note that $\mathbf{D}$ satisfies the classical Brunn-Minkowski inequality, i.e. if $Q_0, Q_1 : T_xM \to T_yM$ then for $Q_t = (1-t)Q_0 + tQ_1$

$$\mathbf{D}[Q_t] \geq (1-t)\mathbf{D}[Q_0] + t\mathbf{D}[Q_1].$$

Now if $B_r^+(x)$ denotes the forward ball of radius $r$ around $x$, i.e. all $y \in M$ with $d(x,y) < r$ and $B_r^-(x)$ the backward ball around $x$, i.e. all $y \in M$ with $d(y,x) < r$. then define the volume distortion coefficients as follows

$$\mathfrak{v}_t^<(x,y) = \lim_{r\to 0} \frac{\mu(Z_t(x, B_r^+(y))}{\mu(B_{tr}^+(y))} \text{ and } \mathfrak{v}_t^>(x,y) = \lim_{r\to 0} \frac{\mu(Z_t(B_r^-(x), y)}{\mu(B_{(1-t)r}^-(x))}.$$

*Remark.* In the symmetric setting one has $\mathfrak{v}_t^>(x,y) = \mathfrak{v}_{1-t}^<(y,x)$.

**Theorem 4.12** (Volume distortion for $d^2$ [Oht09, 3.2])**.** *For point $x \neq y \in M$ with $y \notin \mathrm{Cut}(x)$, let $\eta : [0,1] \to M$ be the unique minimal geodesic from $x$ to $y$. For $t \in (0,1]$ define $g_t(z) = -d(z,\eta(t))^2/2$.*

*Then we have*

$$\begin{aligned} \mathfrak{v}_t^<(x,y) &= \mathbf{D}\left[d(\exp_x)_{\nabla g_t(x)} \circ [d(\exp_x)_{\nabla g_1(x)}]^{-1}\right] \\ \mathfrak{v}_t^>(x,y) &= (1-t)^{-n}\mathbf{D}\left[d(\exp_x \circ \mathcal{L}_x)_{d(g_1)_x} \circ [d\left(d(tg_1)\right)_x - d\left(dg_t\right)_x]\right]. \end{aligned}$$

**Theorem 4.13** (Volume distortion for $d^p$)**.** *Let $x \neq y$ with $y \notin \mathrm{Cut}(x)$ and $\eta$ be as above. For $t \in (0,1]$ define $f_t(z) = -d(z,\eta(t))^p/p$.*

*Then we have*

$$\begin{aligned} \mathfrak{v}_t^<(x,y) &= \mathbf{D}\left[d(\exp_x)_{\nabla^q f_t(x)} \circ [d(\exp_x)_{\nabla^q f_1(x)}]^{-1}\right] \\ \mathfrak{v}_t^>(x,y) &= (1-t)^{-n}\mathbf{D}\left[d(\exp_x \circ \mathcal{K}_x)_{d(t^{p-1}f_1)_x} \circ [d\left(d(t^{p-1}f_1)\right)_x - d\left(df_t\right)_x]\right]. \end{aligned}$$

*Proof.* The first equation follows from the fact that

$$\nabla^q f_t(x) = \nabla\left(-d(x,\eta(t))^2/2\right).$$

For the second part note that

$$\mathcal{L}_z(d(tg_1)_z) = \mathcal{K}_z(d(t^{p-1}f_1)_z)$$

and thus

$$\begin{aligned} \mathfrak{v}_t^>(x,y) &= (1-t)^{-n}\mathbf{D}\left[d\left(\exp \circ \mathcal{L} \circ (d(tg_1)_z)\right)\right] \\ &= (1-t)^{-n}\mathbf{D}\left[d(\exp \circ \mathcal{K})_{d(t^{p-1}f_1)_x} \circ d\left(d(t^{p-1}f_1)\right)_x\right]. \end{aligned}$$

Similar to [Oht09, Proof of 3.2] since $d(f_t)_x = d(t^{p-1}f_1)_x$ it suffices to show that

$$d(exp_x \circ \mathcal{K}_x)_{d(f_t)_x} \circ d(df_t)_x = 0.$$

Note that

$$\mathcal{L}_z(d(g_t)_z) = \mathcal{K}_z(d(f_t)_z)$$

and thus

$$\begin{aligned} L(z) = \exp_z \circ \mathcal{K}_z(d(f_t)_z) &= \exp_z \circ \mathcal{L}_z(d(g_t)_z) \\ &= \eta(t). \end{aligned}$$

Which immediately implies $dL = 0$. □

## 4.5 Interpolation inequality in the $p$-Wasserstein space

The following proposition is a generalization of [Oht09, 5.1] to the case $p \neq 2$. The proof is up to some changes in notation and changes of powers the same as Ohta's.

**Proposition 4.14.** *Let $\phi : M \to \mathbb{R}$ be a $c_p$-concave function and define $\mathcal{F}(z) = exp_z(\nabla^q(-\phi)(z))$ at all point of differentiability of $\phi$. Fix some $x \in M$ such that $\phi$ is twice differentiable at $x$ and $d\phi_x \neq 0$. Then the following holds:*

1. *$y = \mathcal{F}(x)$ is not a cut point of $x$.*
2. *The function $h(z) = c_p(z, y) - \phi(z)$ satisfies $dh_x = 0$ and*

$$\left(\frac{\partial^2 h}{\partial x^i \partial x^j}(x)\right) \geq 0$$

   *in any local coordinate system $(x^i)_{i=1}^n$ around $x$.*

3. *Define $f_y(z) := -c_p(z, y)$ and*

$$d\mathcal{F}_x := d(\exp_x \circ \mathcal{K}_x)_{d(-\phi)_x} \circ [d(d(-\phi))_x - d(df_y)_x] : T_xM \to T_yM$$

   *where the vertical part of $T_{d(-\phi)_x}(T^*M)$ and $T_{d(-\phi)_x}(T^*M)$ are identified. Then the following holds for all $v \in T_xM$*

$$\sup\left\{|u - d\mathcal{F}_x(v)| \mid \exp_y u \in \partial^{c_p}\phi(\exp_x y), |u| = d(y, \exp_y u)\right\} = o(|v|).$$

*Proof.* As $\phi$ is differentiable at $x$ we have $\partial^{c_p}\phi(x) = \{y\}$ and hence for any vector $v \in T_xM$ with $F(v) = 1$ and sufficiently small $t > 0$, we have by $c_p$-concavity of $\phi$

$$h(x) = c_p(x, y) - \phi(x) = \phi^{c_p}(y) \leq c_p(\xi_v(\pm t), y) - \phi(\xi_v(\pm t)) = h(\xi_v(\pm t))$$

where $\xi_v : (-\epsilon, \epsilon) \to M$ is a geodesic with $\dot{\xi}_v(0) = v$. Thus, we have

$$\frac{\phi(\xi_v(t)) + \phi(\xi_v(-t)) - 2\phi(x)}{t^2} \leq \frac{f_y(\xi_v(t)) + f_y(\xi_v(-t)) - 2f_y(x)}{t^2}.$$

Since $\phi$ is twice differentiable at $x$ we have

$$-\infty < \frac{\partial^2(\phi \circ \xi_v)}{\partial t^2}(0) = \limsup_{t \to 0^+} \frac{f_y(\xi_v(t)) + f_y(\xi_v(-t)) - 2f_y(x)}{t^2}$$

and hence $y$ is not a cut point of $x$ (Lemma 4.1).

Now the second statement follows immediately from the inequality above and the fact that $y \notin \mathrm{Cut}(x) \cup \{x\}$ implies that $f_y$ is $C^\infty$ at $x$ and $\nabla^q f_y(x) = \nabla^q \phi(x)$, i.e. $h$ takes its minimum at $x$.

The last part follows from the fact that $dh_x = 0$ implies $d(f_y)_x = d\phi_x$ and thus the difference $d(d(-\phi))_x - d(df)_x$ makes sense. Putting $x_t = \exp_x tv$ for some $v \in T_xM$ and small $t \geq 0$ we can find $u_t \in T_yM$ such that $y_t := \exp_y u_t \in \partial^{c_p}\phi(x_t)$ and $d(y, y_t) = F(u_t)$. In addition, we have

$$-\phi(\exp_{x_t} w) \geq -\phi(x_t) - f_{y_t}(x_t) + f(\exp_{x_t} w) = -\phi(x_t) + d(f_{y_t})_{x_t}(w) + o(F(w))$$

for $w \in T_{x_t}M$. Differentiating $y_t = \exp \circ \mathcal{K}(d(f_{y_t})_{x_t})$ at $t = 0$ we get

$$\frac{\partial y_t}{\partial t}\Big|_{t=0} = d(\exp \circ \mathcal{K})_{d(-\phi)_x} \circ d(d(-\phi)_x)(v).$$

Moreover, we have $\exp \circ \mathcal{K}(d(f_y)_{x_t}) \equiv y$ and thus $d(\exp \circ \mathcal{K})_{d(f_y)_x} \circ d(df_y)_x(v) = 0$. Therefore,

$$\frac{\partial y_t}{\partial t}\Big|_{t=0} = d(\exp \circ \mathcal{K})_{d(-\phi)_x} \circ \left[d(d(-\phi)_x - d(df_y)_x\right](v) = d\mathcal{F}_x(v).$$

Note that, because $d(d(-\phi)_x) - d(df_y)$ contains only vertical terms (see also [Oht09, Proof of 5.1]) we regard it as living in $T_{d(-\phi)_x}(T^*_xM)$ and thus replace $d(exp \circ \mathcal{K})_{d(-\phi)_x}$ by $d(exp \circ \mathcal{K}_x)_{d(-\phi)_x}$. The last part follows immediately by noticing that $\phi$ is second order differentiable and thus $y_t = \exp_y u_t$ with $u_t = d\mathcal{F}_x(tv) + o(t)$ where $o(t)$ can be chosen uniformly in $v$. □

**Proposition 4.15.** *Let $\mu_0$ and $\mu_1$ be absolutely continuous measures with density $f_0$ and $f_1$ resp. and assume that there are open set $U_i$ with compact closure $X = \bar{U}_0$ and $Y = \bar{U}_1$ such that $\operatorname{supp} \mu_i \subset U_i$. Let $\phi$ be the unique $c_p$-concave Kantorovich potential and define $\mathcal{F}(z) = \exp_z(\nabla^q(-\phi)(z))$. Then $\mathcal{F}$ is injective $\mu_0$-almost everywhere and for $\mu_0$-almost every $x \in M \backslash \Omega_{id}$*

1. *The function $h(z) = c_p(z, \mathcal{F}(z)) - \phi(z)$ satisfies*

$$\left(\frac{\partial^2 h}{\partial x^i \partial x^j}(x)\right) > 0$$

*in any local coordinate system $(x^i)_{i=0}^n$ around $x$.*

2. *In particular, $\mathbf{D}[d\mathcal{F}_x] > 0$ holds for the map $d\mathcal{F}_x : T_xM \to T_{\mathcal{F}(x)}M$ defined as above and*

$$\lim_{r\to 0} \frac{\mu(\partial^{c_p}\phi(B_r^+(x)))}{\mu(B_r^+(x))} = \mathbf{D}[d\mathcal{F}_x]$$

*and*

$$f(x) = g(\mathcal{F}(x))\mathbf{D}[d\mathcal{F}_x].$$

*Remark.* defining $d\mathcal{F}_x = \mathrm{Id}$ for points $x$ of differentiability of $\phi$ with $d\phi_x = 0$, we see that the second statement above holds $\mu$-a.e.

*Proof.* The proof follows without any change from [Oht09, Theorem 5.2], see also [Vil09, Chapter 11]. □

**Theorem 4.16.** *Let $\phi : M \to \mathbb{R}$ be a $c_p$-concave function and $x \in M$ such that $\phi$ is twice differentiable with $d\phi_x \neq 0$. For $t \in (0,1]$, define $y_t := \exp_x(\nabla^q(-t^{p-1}\phi))$, $f_t(z) = -c_p(z, y_t)$ and $\mathbf{J}_t(x) = \mathbf{D}[d(\mathcal{F}_t)_x]$ where*

$$d(\mathcal{F}_t)_x := d(\exp_x \circ \mathcal{K}_x)_{d(-t^{p-1}\phi)_x} \circ \left[d(d(-t^{p-1}\phi))_x - d(d(f_t))_x\right] : T_xM \to T_{y_t}M.$$

*Then for any $t \in (0,1)$*

$$\mathbf{J}_t(x)^{1/n} \geq (1-t)\mathfrak{v}_t^{>}(x,y_1)^{1/n} + t\mathfrak{v}_t^{<}(x,y_1)^{1/n}\mathbf{J}_1(x)^{1/n}.$$

*Remark.* The proof is based on the proof of [Oht09, Proposition 5.3].

*Proof.* Note first that

$$\begin{aligned} d(d(-t^{p-1}\phi))_x - d(df_t)_x &= \left\{d(d(-t^{p-1}\phi))_x - d(d(t^{p-1}f_1))_x\right\} \\ &\quad + \left\{d(d(t^{p-1}f_1))_x - d(df_t)_x\right\} \end{aligned}$$

and

$$d(f_t)_x = d(-t^{p-1}\phi)_x = d(-t^{p-1}f_1)_x.$$

Now define $\tau_s : T^*M \to T^*M$ as $\tau_s(v) = s^{p-1}v$ and note

$$\begin{aligned} & d(\exp_x \circ \mathcal{K}_x)_{d(-t^{p-1}\phi)_x} \circ \left(d(d(-t^{p-1}\phi))_x - d(d(t^{p-1}f_1))_x\right) \\ = \;& d(\exp_x \circ \mathcal{K}_x)_{d(-t^{p-1}\phi)_x} \circ d(\tau_t)_{d(-\phi)_x} \circ [d(d(-\phi))_x - d(d(f_1))_x] \\ = \;& d(\exp_x \circ \mathcal{K}_x)_{d(-t^{p-1}\phi)_x} \circ d(\tau_t)_{d(-\phi)_x} \circ \left[d(\exp_x \circ \mathcal{K}_x)_{d(-\phi)_x}\right]^{-1} \circ d(\mathcal{F}_1) \\ = \;& d(\exp_x)_{\nabla^q(-t^{p-1}\phi)_x} \circ d(\mathcal{K}_x \circ \tau_t \circ \mathcal{K}_x^{-1})_{\nabla^q(-\phi)_x} \circ [d(\exp_x)_{\nabla^q(-\phi)_x}]^{-1} \circ d(\mathcal{F}_1) \\ = \;& t \cdot d(\exp_x)_{\nabla^q(-t^{p-1}\phi)_x} \circ [d(\exp_x)_{\nabla^q(-\phi)_x}]^{-1} \circ d(\mathcal{F}_1), \end{aligned}$$

because $\mathcal{K}_x \circ \tau_t \circ \mathcal{K}_x^{-1}$ is linear and for $v \in T_xM$

$$\mathcal{K}_x \circ \tau_t \circ \mathcal{K}_x^{-1}(v) = \mathcal{K}_x(t^{p-1}\mathcal{K}_x^{-1}(v)) = tv,$$

i.e. $d(\mathcal{K}_x \circ \tau_t \circ \mathcal{K}_x^{-1})_{\nabla^q(-\phi)_x} = t \cdot \mathrm{Id}$. Note that we identified $T_{\nabla^q(-t^{p-1}\phi)(x)}(T_xM)$ with $T_{\nabla^q(-\phi)(x)}(T_xM)$ to get the last inequality.

Because $\mathbf{D}$ is concave we get

$$
\begin{aligned}
&\mathbf{J}_t(x)^{1/n} = \mathbf{D}[d(\mathcal{F}_t)_x]^{1/n} \\
= \;& \mathbf{D}\Big[d(\exp_x \circ \mathcal{K}_x)_{d(-t^{p-1}\phi)_x} \circ \big[d(d(t^{p-1}f_1))_x - d(df_t)_x\big] \\
&+ d(\exp_x \circ \mathcal{K}_x)_{d(-t^{p-1}\phi)_x} \circ \big(d(d(-t^{p-1}\phi))_x - d(d(t^{p-1}f_1))_x\big)\Big]^{1/n} \\
= \;& \mathbf{D}\Big[d(\exp_x \circ \mathcal{K}_x)_{d(-t^{p-1}\phi)_x} \circ \big(d(d(t^{p-1}f_1))_x - d(df_t)_x\big) \\
&+ t \cdot d(\exp_x)_{\nabla^q(-t^{p-1}\phi)_x} \circ [d(\exp_x)_{\nabla^q(-\phi)_x}]^{-1} \circ d(\mathcal{F}_1)\Big]^{1/n} \\
\geq \;& (1-t)\mathbf{D}\Big[(1-t)^{-1} d(\exp_x \circ \mathcal{K}_x)_{d(-t^{p-1}\phi)_x} \circ \big(d(d(t^{p-1}f_1))_x - d(df_t)_x\big)\Big]^{1/n} \\
&+ t\mathbf{D}\Big[d(\exp_x)_{\nabla^q(-t^{p-1}\phi)_x} \circ [d(\exp_x)_{\nabla^q(-\phi)_x}]^{-1} \circ d(\mathcal{F}_1)\Big]^{1/n} \\
= \;& (1-t)\mathfrak{v}_t^{>}(x,y_1)^{1/n} + t\mathfrak{v}_t^{<}(x,y_1)^{1/n}\mathbf{J}_1(x)^{1/n}.
\end{aligned}
$$

□

Combing this with Lemma 3.11 and Lemma 4.18 below we get similar to [Oht09, 6.2]:

**Lemma 4.17.** *Given two absolutely continuous measures $\mu_i = \rho_i\mu$ on $M$, let $\phi$ be the unique $c_p$-concave optimal Kantorovich potential. Define $\mathcal{F}_t(x) := \exp_x(\nabla^q(-t^{p-1}\phi))$ for $t \in [0,1]$. Then $\mu_t = (\mathcal{F}_t)_*\mu_0$ is absolutely continuous for any $t \in [0,1]$.*

*Proof.* By Lemma 3.11 the map $\mathcal{F}_t$ is injective $\mu_0$-almost everywhere. Let $\Omega_{id}$ be the points $x \in M$ of differentiability of $\phi$ with $d\phi_x = 0$. Then

$$\mu_t\big|_{\Omega_{id}} = (\mathcal{F}_t)_*(\mu_0\big|_{\Omega_{id}}) = \mu_0\big|_{\Omega_{id}}.$$

By Theorem 4.10 the potential $\phi$ is second order differentiable in a subset $\Omega \subset M\backslash\Omega_{id}$ of full measure. In addition, $\mathbf{D}[d(\mathcal{F}_1)] > 0$ for all $x \in \Omega$ (see Proposition 4.15) and $\mathcal{F}_t$ is continuous in $\Omega$ for any $t \in [0,1]$. The map $d(\mathcal{F}_t)_x : T_xM \to T_{\mathcal{F}_t(x)}M$ defined in Proposition 4.14 as

$$d(\mathcal{F}_t)_x := d(\exp_x \circ \mathcal{K}_x)_{d(-t^{p-1}\phi)} \circ \big[d(d(-t^{p-1}\phi))_x - d(d(f_t)_x\big]$$

where $f_t(z) := -c_p(z, \mathcal{F}_t(x))$ for $t \in (0,1]$. Also note that for $x \in \Omega$

$$d(d(-t^{p-1}\phi))_x - d(df_t)_x = \big\{d(d(-t^{p-1}\phi))_x - d(d(t^{p-1}f_1))_x\big\} + \big\{d(d(t^{p-1}f_1))_x - d(f_t)_x\big\}.$$

Which implies $\mathbf{D}[d(\mathcal{F}_t)_x] > 0$ because $\mathbf{D}[d(\mathcal{F}_1)_x) > 0$ and the lemma below.

The result then immediately follows by [CEMS01, Claim 5.6]. □

**Lemma 4.18.** *Let $y \notin \mathrm{Cut}(x) \cup \{x\}$ and $\eta : [0,1] \to M$ be the unique minimal geodesic from $x$ to $y$. Define*

$$f_t(z) = -c_p(z, \eta(t)).$$

*Then the function $h(z) = t^{p-1} f_1(z) - f_t(z)$ satisfies*

$$\left( \frac{\partial^2 h}{\partial x^i \partial x^j}(x) \right) \geq 0$$

*in any local coordinate system around $x$.*

*Proof.* This follows directly from 3.8. $\square$

# 5 Abstract curvature condition

In this chapter we define a curvature condition à la Lott-Villani-Sturm ([LV07, LV09] and [Stu06b, Stu06a]) with respect to geodesics in $\mathcal{P}_p(M)$ with $p \in (1,\infty)$. For simplicity, throughout this chapter, we assume that $M$ is a proper geodesic space.

## 5.1 Curvature dimension

In [LV09] (see also [Vil09, Part II-III]) Lott and Villani introduced the following set of real-valued functions.

**Definition 5.1** ($\mathcal{DC}_N$)**.** For $N \in [1,\infty]$ let $\mathcal{DC}_N$ all convex functions $U : [0,\infty) \to \mathbb{R}$ with $U(0) = 0$ such that for $N < \infty$ the function

$$\psi(\lambda) = \lambda^N U(\lambda^{-N})$$

is convex on $(0,\infty)$. In case $N = \infty$ we require

$$\psi(\lambda) = e^{\lambda} U(e^{-\lambda})$$

to be convex on $(-\infty,\infty)$.

**Lemma 5.2** ([LV09, Lemma 5.6])**.** *If $N \le N'$ then $\mathcal{DC}_{N'} \subset \mathcal{DC}_N$.*

**Example 5.3.** Note the following examples

1. if $m = 1 - \frac{1}{N}$ for $N \in (1,\infty)$ then $U_m : x \mapsto \frac{1}{m(m-1)} x^m$ is in $\mathcal{DC}_N$
2. the classical entropy functional $U_\infty : x \mapsto x \log x$ is in $\mathcal{DC}_\infty$
3. if $m > 1$ then $U_m \in \mathcal{DC}_\infty$

Given a function $U \in \mathcal{DC}_N$ for $N \in [1,\infty]$ we write $U'(\infty) = \lim_{r\to\infty} \frac{U(r)}{r}$. Given some reference measure $\mu \in \mathcal{P}(M)$ we define the functional $\mathcal{U}_\mu : \mathcal{P}(M) \to \mathbb{R} \cup \{\infty\}$ by

$$\mathcal{U}_\mu(\nu) = \int U(\rho) d\mu + U'(\infty)\nu_s(M)$$

where $\nu = \rho\mu + \mu_s$ the the Lebesgue decomposition of $\nu$ w.r.t. $\mu$.

*Remark.* In the following we usually fix a metric measure space $(M, d, \mu)$ and drop the subscript $\mu$ from the functional $\mathcal{U}_\mu$. In addition, we use $\mathcal{U}_m$, $\mathcal{U}_\alpha$ etc. to denote the functional generated by $U_m$, $U_\alpha$, etc.

In [LV07, Section 4] Lott and Villani defined for each $K \in \mathbb{R}$ and $N \in (1, \infty]$ the functions $\beta_t : M \times M \to \mathbb{R} \cup \{\infty\}$ and $t \in [0,1]$ as follows

$$\beta_t(x_1, x_2) = \begin{cases} e^{\frac{1}{6}K(1-t^2)d(x_0,x_1)^2} & \text{if } N = \infty, \\ \infty & \text{if } N < \infty, K > 0 \text{ and } \alpha > \pi, \\ \left(\frac{\sin(t\alpha)}{t \sin \alpha}\right)^{N-1} & \text{if } N < \infty, K > 0 \text{ and } \alpha \in [0, \pi], \\ 1 & \text{if } N < \infty \text{ and } K = 0, \\ \left(\frac{\sinh(t\alpha)}{t \sinh \alpha}\right)^{N-1} & \text{if } N < \infty \text{ and } K < 0, \end{cases}$$

where

$$\alpha = \sqrt{\frac{|K|}{N-1}} d(x_0, x_1)$$

and for $N = 1$

$$\beta_t(x_0, x_1) = \begin{cases} \infty & \text{if } K > 0, \\ 1 & \text{if } K \leq 0. \end{cases}$$

Note that $\beta$ and $\alpha$ depend implicitly on an a priori chosen $K$ and $N$ which will be suppressed to keep the notation simple.

*Remark.* In [BS10] Bacher and Sturm defined a reduced curvature dimension condition with a different weight function $\sigma_t$ instead of $\beta_t$. Because of the localization and tensorization property this weight function turned out to be powerful ([AGS11b, AGMR12, Raj12, Raj11, GM13, EKS13, Gig13, HKX13]). Using the inequalities of the proof of Lemma 3.11 most of the things proven in [BS10] will also hold for localized version $CD_p^*(K, N)$.

**Definition 5.4** ((strong) $CD_p(K, N)$)**.** We say $(M, d, \mu)$ satisfies the strong $CD_p(K, N)$ condition if the following holds: Given two measures $\mu_0, \mu_1 \in \mathcal{P}(M)$ with Lebesgue decomposition $\mu_i = \rho_i \mu + \mu_{i,s}$. Then there exists some optimal dynamical transfer plan $\Pi \in \mathcal{P}(\mathrm{Geo})$ such that $\mu_t = (e_t)_* \Pi$ is a geodesic from $\mu_0$ to $\mu_1$ in $\mathcal{P}_p(M)$ such that for all $U \in \mathcal{DC}_N$ and $t \in [0, 1]$

$$\begin{aligned} \mathcal{U}(\mu_t) \quad \leq \quad & (1-t) \int_{M \times M} \beta_{1-t}(x_0, x_1) U\left(\frac{\rho_0(x_0)}{\beta_{1-t}(x_0, x_1)}\right) d\pi(x_1|x_0) d\mu(x_0) \\ & + t \int_{M \times M} \beta_t(x_0, x_1) U\left(\frac{\rho_1(x_i)}{\beta_t(x_0, x_1)}\right) d\pi(x_0|x_1) d\mu(x_1) \\ & + U'(\infty)\left((1-t)\mu_{0,s}(M) + t\mu_{1,s}(M)\right), \end{aligned}$$

where $\pi = (e_0, e_1)_* \Pi$ is the optimal transference plan of $(\mu_0, \mu_1)$ w.r.t. $c_p$ associated to $\Pi$. Furthermore, in case $\beta_s(x_0, x_1) = \infty$ we interpret $\beta_s(x_0, x_1) U\left(\frac{\rho_i(x_i)}{\beta_s(x_0,x_1)}\right)$ as $U'(0)\rho_i(x_i)$.

In addition, we say that the very strong $CD_p(K, N)$ condition holds if the inequality holds for all optimal dynamical transference plans (and thus all geodesics).

Note that this definition is Lott-Villani's [LV07, Defnition 4.7] by just requiring the geodesic $t \mapsto \mu_t$ to be in $\mathcal{P}_p(M)$ instead of $\mathcal{P}_2(M)$. And, in case both $\mu_i$ are absolutely continuous it looks like

$$\begin{aligned}\mathcal{U}(\mu_t) \quad &\le \quad (1-t)\int \frac{\beta_{1-t}(x_0,x_1)}{\rho_0(x_0)} U\left(\frac{\rho_0(x_0)}{\beta_{1-t}(x_0,x_1)}\right) d\pi(x_0,x_1) \\ &+t\int \frac{\beta_t(x_0,x_1)}{\rho_1(x_1)} U\left(\frac{\rho_1(x_1)}{\beta_t(x_0,x_1)}\right) d\pi(x_0,x_1).\end{aligned}$$

An immediate consequence of the curvature condition is the following:

**Lemma 5.5.** *Assume $(M,d,\mu)$ satisfies the strong $CD_p(K,N)$ and $\mu_0$ and $\mu_1$ are absolutely continuous, if $t \mapsto \mu_t$ satisfies the functional inequality then $\mu_t$ is absolutely continuous.*

*Proof.* The proof follows from [LV09, Theorem 5.52] (see also [LV07, Theorem 4.30]) by noting that [LV09, Lemma 5.43] does not need $\mu_i$ to be in $\mathcal{P}_2^{ac}(M)$. □

Furthermore, we will also define a variant of Sturm's curvature condition [Stu06a, Stu06b]:

**Definition 5.6** ((weak) $CD_p(K,N)$)**.** We say $(M,d,\mu)$ satisfies the weak $CD_p(K,N)$ condition if for $N \in (1,\infty)$ the above inequality holds only for the functionals

$$U_{N'}(r) = N'r(1-r^{-1/N'})$$

for any $N' \ge N$. In case $N' = \infty$ the functional $\mathcal{U}_\infty$ generated by

$$U_\infty(r) = r\log r$$

and has to be $K$-convex along a geodesic $t \mapsto \mu_t$ in $\mathcal{P}_p(M)$, i.e.

$$\mathcal{U}_\infty(\mu_t) \le (1-t)\mathcal{U}_\infty(\mu_0) + t\mathcal{U}_\infty(\mu_1) - \frac{K}{2}t(1-t)w_p^2(\mu_0,\mu_1).$$

The following follows immediately from Theorem 4.16 by similar statements to the case $CD_2(K,N)$ (see e.g. [Oht09, Vil09]).

**Corollary 5.7.** *Any $n$-dimensional Finsler manifold with $N$-Ricci curvature bounded from below by $K$ and $N > n$ satisfies the very strong $CD_p(K,N)$ condition for all $p \in (1,\infty)$.*

*Remark.* Note[1] that in contrast to the case $p = 2$ the strong $CD_p(K,\infty)$-condition does not imply the weak one. Indeed the strong $CD_p(K,\infty)$-condition [LV07, Lemma 4.14] only gives

$$\mathcal{U}_\infty(\mu_t) \le (1-t)\mathcal{U}_\infty(\mu_0) + t\mathcal{U}_\infty(\mu_1) - \frac{1}{2}\lambda(U)t(1-t)\int d^2(x,y)d\pi_{opt}(x,y),$$

[1]We thank Shin-ichi Ohta for making this remark on an early version of a preprint

where $\pi_{opt}$ is the $d^p$-optimal coupling between $\mu_0$ and $\mu_1$. However, using Hölder inequality we get for $p > 2$

$$\int d^2(x,y)d\pi_{opt}(x,y) \le \left(\int d^p(x,y)d\pi_{opt}(x,y)\right)^{\frac{2}{p}} = C_p w_p^2(\mu_0,\mu_1)$$

and for $p < 2$

$$c_p w_p^2(\mu_0,\mu_1) \le \left(\int d^p(x,y)d\pi_{opt}(x,y)\right)^{\frac{2}{p}} \le \int d^2(x,y)d\pi_{opt}(x,y).$$

Thus we get $K'$-convexity for some $K'$ depending only on $p$ and $K$ follows if either $\lambda(U) > 0$ and $p < 2$ or $\lambda(U) < 0$ and $p > 2$.

In the negatively curved case with bounded diameter one can also do the following: the function

$$\lambda \mapsto e^{\lambda}U_\infty(e^{-\lambda})$$

is convex and non-increasing. This means, if we take some $\beta_t^{'}(\cdot,\cdot) \le \beta_t(\cdot,\cdot)$ then we still have

$$\begin{aligned}\mathcal{U}(\mu_t) \quad &\le \quad (1-t)\int \frac{\beta_{1-t}^{'}(x_0,x_1)}{\rho_0(x_0)}U\left(\frac{\rho_0(x_0)}{\beta_{1-t}^{'}(x_0,x_1)}\right)d\pi(x_0,x_1)\\ &\quad +t\int \frac{\beta_t^{'}(x_0,x_1)}{\rho_1(x_1)}U\left(\frac{\rho_1(x_1)}{\beta_t^{'}(x_0,x_1)}\right)d\pi(x_0,x_1),\end{aligned}$$

assuming $\mu_0$ and $\mu_1$ are absolutely continuous. Now choose for $r < 2$ and $D_r = (\operatorname{diam} M)^{2-r}$ then $d^2(x,y) \le D_r d^r(x,y)$ and define the following function

$$\beta_t^{'}(x,y) = e^{\frac{1}{6}D_r K(1-t^2)d^r(x,y)}.$$

If $K < 0$ then obviously $\beta_t^{'} \le \beta_t$ and the interpolation inequality above holds. As above we conclude that the functional is $K'$-convex for some $K'$ depending on $D_r K$ and $p > r$.

## 5.2 Positive curvature and global Poincaré inequality

In this section we will show a Poincaré inequality for positively curved spaces first proven by Lott and Villani in [LV07] for the case $p = 2$.

For that fix a metric measure space $(M,d,\mu)$ and let $q$ be the Hölder conjugate of $p$. Then for a given $U \in C^2(\mathbb{R})$ we define the generalized $q$-Fisher information (associated to $(U,\mu)$)

$$\begin{aligned} I_q(\nu) &= \int U''(\rho)^q|D^-\rho|^q d\nu\\ &= \int \rho U''(\rho)^q|D^-\rho|^q d\mu\end{aligned}$$

where $\nu$ is an absolutely continuous measure w.r.t. $\mu$.

*Remark.* The Fisher information defined here is more general then the one define in section 2.1.3. In case $U(r) = \frac{1}{(3-p)(2-p)}(r^{3-p} - r)$ where $p$ is the Hölder conjugate of $q$ one has equality.

In case the $CD_p(K,N)$ holds for $K > 0$ and $N \in (1,\infty)$ the following directly follows from [LV07, Theorem 5.34] without changing the proofs.

**Lemma 5.8.** *Let $(M,d,\mu)$ be a metric measure space satisfying $CD_p(K,N)$ for $K > 0$ and $N \in (1,\infty)$. Then for any Lipschitz function $f$ on $M$ with $\int f d\mu = 0$ it holds*

$$\int f^2 d\mu \leq \frac{N-1}{KN} \int |D^- f|^2 d\mu.$$

However in case $N = \infty$ we need to adjust the proof using the Lemma below.

**Lemma 5.9.** *Let $(M,d,\mu)$ be a compact geodesic metric measure space and $U$ be a continuous convex function on $[0,\infty)$ with $U(0) = 0$. Let $\nu \in \mathcal{P}_p(M)$ and assume $t \mapsto \mu_t$ is a geodesic in $\mathcal{P}_p(M)$ from $\mu_0 = \nu$ to $\mu_1 = \mu$ such that the functional $\mathcal{U}$ (associated to $(U,\mu)$) is $K$ convex along $\mu_t$, i.e.*

$$\mathcal{U}(\mu_t) \leq (1-t)\mathcal{U}(\mu_0) + t\mathcal{U}(\mu_1) - \frac{K}{2}t(1-t)w_p^2(\mu_0,\mu_1).$$

*Then*

$$\frac{K}{2}w_p(\nu,\mu) \leq \mathcal{U}(\nu) - \mathcal{U}(\mu).$$

*If $U$ is $C^2$-regular on $(0,\infty)$, $\nu = \rho\mu$ for some positive Lipschitz function $\rho$ on $M$ with $\mathcal{U}(\nu) < \infty$ and $\mu_t$ is absolutely continuous for each $t \in [0,1]$ then*

$$\mathcal{U}(\nu) - \mathcal{U}(\mu) \leq w_p(\nu,\mu)\sqrt[q]{I_q(\nu)} - \frac{K}{2}w_p(\nu,\mu)^2.$$

*Proof.* The proof follows from [LV09, Proposition 3.36] by making some minor adjustments. We will include the whole proof, since it can also be used to generalize [LV07, Theorem 5.3] (note that $\mathcal{U}$ with $U \in \mathcal{DC}_N$ is not necessarily $K$-convex).

The first part follows directly from the $K$-convexity: Let $\phi(t) = \mathcal{U}(\mu_t)$, then

$$\phi(t) \leq t\phi(1) + (1-t)\phi(0) - \frac{1}{2}t(1-t)w_p(\nu,\mu)^2.$$

If the inequality does not hold then $\phi(0) - \phi(1) < \frac{1}{2}w_p(\nu,\mu)^2$ and hence

$$\phi(t) - \phi(1) \leq (1-t)\left(\phi(0) - \phi(1) - \frac{K}{2}tw_p(\nu,\mu)^2\right)$$

which implies that $\phi(t) - \phi(1)$ is negative for $t$ close to 1. But this contradicts [LV09, Lemma 3.36], i.e. $\mathcal{U}(\mu) \geq \mathcal{U}(\nu) = U(1)$. Therefore, the first inequality holds.

To prove the second part, let $\rho_t$ be the density of $\mu_t$. Then $\phi(t) = \int U(\rho_t)d\mu$ and from above we have

$$\phi(0) - \phi(1) \leq -\frac{\phi(t) - \phi(1)}{t} - \frac{K}{2}(1-t)w_p(\nu,\mu)^2.$$

So to prove the second inequality we just need to show

$$\liminf_{t\to\infty}\left(-\frac{\phi(t)-\phi(0)}{t}\right)\le w_p(\nu,\mu)\sqrt[q]{I_q(\nu)}.$$

Since $U$ is convex we have

$$U(\rho_t)-U(\rho_0)\ge U'(\rho_0)(\rho_t-\rho_0).$$

Integrating w.r.t. $\mu$ and dividing by $-t<0$ we get

$$\begin{aligned}-\frac{1}{t}\left(\phi(t)-\phi(0)\right) &\le -\frac{1}{t}\int U'(\rho_0(x))\left(d\mu_t(x)-d\mu(x)\right)\\ &= -\frac{1}{t}\int U'(\rho_0(\gamma_t))-U'(\rho_0(\gamma_0))d\Pi(\gamma)\end{aligned}$$

where $\Pi$ is the optimal transference plan in $\mathcal{P}(\mathrm{Geo})$ associated to $t\mapsto\mu_t$.

Since $U'$ is non-decreasing and $d(\gamma_t,\gamma_0)=td(\gamma_0,\gamma_1)$ we obtain for

$$\begin{aligned}-\frac{1}{t}\int U'(\rho_0(\gamma_t))-U'(\rho_0(\gamma_0))d\Pi(\gamma) &\le -\frac{1}{t}\int_{\rho_0(\gamma_t)\le\rho_0(\gamma_0)}\left[U'(\rho_0(\gamma_t))-U'(\rho_0(\gamma_0))\right]d\Pi(\gamma)\\ &\le \int\frac{U'(\rho_0(\gamma_t))-U'(\rho_0(\gamma_0))}{\rho_0(\gamma_t)-\rho_0(\gamma_0)}\\ &\qquad\times\frac{[\rho_0(\gamma_t)-\rho_0(\gamma_0)]_-}{d(\gamma_t,\gamma_0)}d(\gamma_1,\gamma_0)d\Pi(\gamma).\end{aligned}$$

Applying Hölder inequality we get

$$\begin{aligned}&\sqrt[q]{\int\frac{[U'(\rho_0(\gamma_t))-U'(\rho_0(\gamma_0))]^q}{[\rho_0(\gamma_t)-\rho_0(\gamma_0)]^q}\frac{[\rho_0(\gamma_t)-\rho_0(\gamma_0)]_-^q}{d(\gamma_t,\gamma_0)^q}d\Pi(\gamma)}\\ &\qquad\times\sqrt[p]{\int d(\gamma_0,\gamma_1)^q d\Pi(\gamma)}.\end{aligned}$$

where the second factor is just $w_p(\nu,\mu)$. Taking continuity of $\rho_0$ and the definition of $|D-\rho_0|$ into account we conclude as in the proof of [LV09, Proposition 3.36] that the first factor equals

$$\sqrt[q]{\int U''(\rho_0)^q|D^-\rho_0|^q d\nu}=\sqrt[q]{I_q(\nu)}.$$

□

**Corollary 5.10.** *Assume that the (weak) $CD_p(K,\infty)$ condition holds for $K>0$ and some $N\in[1,\infty]$. Then for all $\nu\in\mathcal{P}_p(M)$*

$$\frac{K}{2}w_p(\nu,\mu)^2\le\mathcal{U}_\infty(\nu).$$

*If $\nu$ is absolutely continuous with positive Lipschitz density $\rho$ then*

$$\mathcal{U}_\infty(\nu)\le w_p(\nu,\mu)\sqrt[q]{I_q(\nu)}-\frac{K}{2}w_p(\nu,\mu)^2\le\frac{1}{2K}\left(I_q(\nu)\right)^{\frac{2}{q}}.$$

*Proof.* Just note that if $\mathcal{U}_\infty$ is $K$-convex along a geodesic $t \mapsto \mu_t$ between absolutely continuous measures, then each $\mu_t$ is absolutely continuous. □

Note that in this case

$$I_q(\rho\mu) = \int \rho \frac{1}{\rho^q} |D^-\rho|^q d\mu = \int \frac{|D^-\rho|^q}{\rho^{q-1}} d\mu.$$

Similar to [LV09, Section 6.2] we will show that the $(2,q)$-log-Sobolev inequality

$$\mathcal{U}_\infty(\rho\mu) \leq \frac{1}{2K} \left(I_q(\rho\mu)\right)^{\frac{2}{q}}.$$

implies a global $(2,q)$-Poincaré inequality. Note that the $(2,q)$-log-Sobolev inequality is different from the one defined in [GRS12].

**Corollary 5.11.** *Assume for $K > 0$ and all positive Lipschitz functions*

$$\mathcal{U}_\infty(\rho\mu) \leq \frac{1}{2K} \left(I_q(\rho\mu)\right)^{\frac{2}{q}}.$$

*Then the $(2,q)$-Poincaré inequality holds with factor independent of $q$, i.e.*

$$\left(\int (h - \bar{h})^2 d\mu\right)^{\frac{1}{2}} \leq \frac{1}{\sqrt{2K}} \left(\int |D^-h|^q d\mu\right)^{\frac{1}{q}}$$

*for $h \in \mathrm{Lip}(M)$. In particular, this holds if $(M,d,\mu)$ satisfies the weak $CD_p(K,\infty)$ condition.*

*Proof.* We will first prove

*Claim.* If $f \in \mathrm{Lip}(M)$ satisfies $\int f^p d\mu = 1$ then

$$\left(\int f^q \log f^q d\mu\right)^{\frac{1}{2}} \leq \frac{q}{\sqrt{2K}} \left(\int |D^-f|^q d\mu\right)^{\frac{1}{q}}.$$

*Proof of the claim.* For any $\epsilon > 0$ let $\rho_\epsilon = \frac{f^p+\epsilon}{1+\epsilon}$ then from the previous corollary

$$\int \rho_\epsilon \log \rho_\epsilon d\mu \leq \frac{1}{2K} \left(\int \frac{|D^-\rho_\epsilon|^q}{\rho_\epsilon^{q-1}} d\mu\right)^{\frac{2}{q}}.$$

By chain rule we have

$$\frac{|D^-\rho_\epsilon|^q}{\rho_\epsilon^{q-1}} = \frac{1}{1+\epsilon} \frac{(qf^{q-1})^q}{(f^q+\epsilon)^{q-1}} |\nabla^- f|^q \to q^q |D^-f|^q$$

as $\epsilon \to 0$, which implies the claim. □

Assume w.l.o.g. $\int h = 0$. For $\epsilon \in [0, \frac{1}{\|h\|_\infty})$ set $f_\epsilon = \sqrt[q]{1 + \epsilon h} > 0$. Then by chain rule

$$|D^- f_\epsilon| = \frac{\epsilon |D^- h|}{q\,(1+\epsilon h)^{\frac{q-1}{q}}}$$

and thus

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left( \int |D^- f_\epsilon|^q d\mu \right)^{\frac{1}{q}} = \frac{1}{q} \left( \int |D^- h|^q d\mu \right)^{\frac{1}{q}}.$$

Note that the Taylor expansion of $x \log x - x + 1$ around $x_0 = 1$ is given by $\frac{1}{2}(x-1)^2 + \ldots$, and thus

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \int f_\epsilon^q \log f_\epsilon^q d\mu = \int h^2 d\mu.$$

Combining this we get

$$\left( \int h^2 d\mu \right)^{\frac{1}{2}} \le \frac{1}{\sqrt{2K}} \left( \int |D^- h|^q d\mu \right)^{\frac{1}{q}}.$$

□

## 5.3 Metric Brenier

**Lemma 5.12** ([Gig12, 5.4])**.** *Let $(M, d, \mu)$ be a metric measure space and $(\mu_n)_{n \in \mathbb{N}}$ be a sequence $\mathcal{P}(M)$ and let $\mu_0 \in \mathcal{P}(M)$ be such that $\mu_0 \ll \mu$ . Assume for some bounded closed set $B \subset M$ with $\mu(B) < \infty$ we have $\operatorname{supp} \mu_n \cup \operatorname{supp} \mu_0 \subset B$, $\mu_n$ converges weakly to $\mu$ and*

$$\mathcal{U}_N(\mu_n) \to \mathcal{U}_N(\mu_0) \quad \text{as } n \to \infty.$$

*Then for every bounded Borel function $f : B \to \mathbb{R}$ it holds*

$$\lim_{n \to \infty} \int f d\mu_n = \int f d\mu$$

**Proposition 5.13.** *Let $(M, d, \mu)$ be a metric measure space and $B$ be a bounded closed subset of $M$ with $\mu(B) < \infty$. Assume $\mu_0$ and $\mu_1$ are two probability measures in $\mathcal{P}_p(M)$ such that $\mu_0 \ll \mu$ and there is an optimal coupling $\pi \in \mathrm{OptGeo}_p(\mu_0, \mu_1)$ such that*

$$\lim_{t \to 0} \mathcal{U}_N(\mu_t) = \mathcal{U}_N(\mu_0)$$

*and $\operatorname{supp}(\mu_t) \subset B$, where $\mu_t = (e_t)_* \pi$. If $\phi$ is the associated Kantorovich potential of the pair $(\mu_0, \mu_1)$ and $\phi$ is Lipschitz on bounded subsets of $X$. Then for every $\tilde{\pi} \in \mathrm{OptGeo}_p(\mu_0, \mu_1)$*

$$d(\gamma_0, \gamma_1)^p = \left( |D^+ \phi|(\gamma_0) \right)^q \quad \tilde{\pi}\text{-a.e. } \gamma.$$

*Remark.* The proof follows by similar arguments as in [Gig12, 5.5] and [AGS13, 10.3].

*Proof.* Let $x \in M$ be arbitrary and choose any $y \in \partial^{c_p}\phi(x)$, then for all $z \in M$

$$\begin{aligned}\phi(x) &= c_p(x,y) - \phi^{c_p}(y),\\ \phi(z) &\leq c_p(z,y) - \phi^{c_p}(y).\end{aligned}$$

Thus

$$\begin{aligned}\phi(z) - \phi(y) &\leq \frac{(d(z,x) + d(x,y))^p - d^p(x,y)}{p}\\ &= (d(z,x) + h_1(d(z,x)) \cdot d(x,y)^{p-1}\end{aligned}$$

where $h_1 : \mathbb{R} \to \mathbb{R}$ is such that $h_1(r) = o(r)$ as $r \to 0$ depending only on $p > 1$. Therefore, dividing by $d(x,z)$ and letting $z \to x$ we see that

$$|D^+\phi|(x) \leq \inf_{y \in \partial^{c_p}\phi(x)} d(x,y)^{p-1}.$$

In particular, since for an arbitrary $\tilde{\pi} \in \mathrm{OptGeo}_p(\mu_0, \mu_1)$ we have $\gamma_1 \in \partial^{c_p}\phi(\gamma_0)$ for $\tilde{\pi}$-almost every $\gamma$, we also have

$$|D^+\phi|(\gamma_0) \leq d(\gamma_0, \gamma_1)^{p-1} \quad \tilde{\pi}\text{-a.e. } \gamma.$$

Note that $q \cdot (p-1) = p$ and thus

$$\int |D^+\phi|^q d\mu_0 \geq w_p^p(\mu_0, \mu_1).$$

So it suffices to show the opposite inequality. For that let $\pi \in \mathrm{OptGeo}(\mu_0, \mu_1)$ as in the hypothesis. Because $\phi$ is a Kantorovich potential we have for $t \in (0,1]$

$$\begin{aligned}\phi(\gamma_0) - \phi(\gamma_t) &\geq \frac{d(\gamma_0,\gamma_1)^p}{p} - \frac{d(\gamma_t,\gamma_1)^p}{p}\\ &= \frac{d(\gamma_0,\gamma_1)^p}{p}(1 - (1-t)^p) = d(\gamma_0,\gamma_1)^p(t + o(t)).\end{aligned}$$

Thus dividing by $d(\gamma_0, \gamma_t) = td(\gamma_0, \gamma_1)$ and integrating to the $q$-th power we get

$$\liminf_{t \to 0} \int \left(\frac{\phi(\gamma_0) - \phi(\gamma_t)}{d(\gamma_0,\gamma_t)}\right)^q d\pi(\gamma) \geq \int d(\gamma_0,\gamma_1)^p d\pi(\gamma) = w_p^p(\mu_0, \mu_1).$$

Because $\phi$ is locally Lipschitz, $|D^+\phi|$ is an upper gradient for $\phi$, we also have

$$\begin{aligned}\int \left(\frac{\phi(\gamma_0) - \phi(\gamma_t)}{d(\gamma_0,\gamma_t)}\right)^q d\pi(\gamma) &\leq \int \frac{1}{t^q}\left(\int_0^t |D^+\phi|(\gamma_s)ds\right)^q d\pi(\gamma)\\ &\leq \int \frac{t^{\frac{q}{p}}}{t^q}\int_0^t |D^+\phi|^q(\gamma_s)ds d\pi(\gamma)\\ &= \frac{1}{t}\int_0^t \int |D^+\phi|^q(\gamma_s)d\pi(\gamma)ds\end{aligned}$$

because $\frac{q}{p} = \frac{1}{p-1} = q - 1$.

Now our assumptions imply that $|D^+\phi|^q$ is a bounded Borel functions thus we can apply the previous lemma to get (see also [Gig12, 5.5]

$$\lim_{t\to 0} \frac{1}{t} \int_t^t \int |D^+\phi|^q(\gamma_s) d\pi(\gamma) ds = \int |D^+\phi|^q d\mu_0.$$

□

In order to avoid the introduction of complicated notation, we just remark that one can also prove [Gig12, Corollary 5.8] and show that the plan $\pi$ above weakly $q$-represents $\nabla(-\phi)$ (for definition see [Gig12, Definition 5.7]).

## 5.4 Laplacian comparison

As an application to the metric Brenier theorem we get the following. Since we do not prove the theorem, we refer to [Gig12] for a precise definition of infinitesimal strictly convex spaces.

**Theorem 5.14** (Comparison estimates)**.** *Let $K \in \mathbb{R}$ and $N \in (1, \infty)$ and $(M, d, \mu)$ be an infinitesimal strictly convex $CD_p(K, N)$-space. If $\phi : X \to \mathbb{R}$ is a $c_p$-concave function. Then*

$$\phi \in D(\Delta_q) \qquad \textit{and} \qquad \Delta^q \phi \le N\tilde{\sigma}_{K,N}(|\nabla\phi|_w^{q-1}) d\mu$$

*where*

$$\tilde{\sigma}_{K,N}(\theta) = \begin{cases} \frac{1}{N}\left(1 + \theta\sqrt{K/(N-1)}\operatorname{cotan}\left(\theta\sqrt{\frac{K}{N-1}}\right)\right) & \textit{if } K > 0 \\ 1 & \textit{if } K = 0 \\ \frac{1}{N}\left(1 + \theta\sqrt{K/(N-1)}\operatorname{cotanh}\left(\theta\sqrt{\frac{K}{N-1}}\right)\right) & \textit{if } K < 0 \end{cases}$$

*Proof.* Follow [Gig12, Theorem 5.14] and just note that the metric Brenier theorem implies $d(\gamma_0, \gamma_1) = |\nabla\phi|_w^{q-1}$. □

**Corollary 5.15** (Laplacian comparison of the distance)**.** *For any $x_0$ one has*

$$\frac{d_{x_0}^p}{p} \in D(\Delta_q) \qquad \textit{with} \qquad \Delta_q \frac{d_{x_0}^p}{p} \le N\tilde{\sigma}_{K,N}(d_{x_0}) d\mu \quad \forall x_0 \in X$$

*and*

$$d_{x_0} \in D(\Delta_q, X\backslash\{x_0\}) \qquad \textit{with} \qquad \Delta^q d_{x_0}\big|_{X\backslash\{x_0\}} \le \frac{N\tilde{\sigma}_{K,N}(d_{x_0})}{d_{x_0}^{p-1}} d\mu.$$

*Remark.* Note that formally

$$\begin{aligned}\Delta^q \frac{d_{x_0}^p}{p} &= \nabla \cdot \left( |\nabla \frac{d_{x_0}^p}{p}|^{q-2} \nabla \frac{d_{x_0}^p}{p} \right) \\ &= \nabla \cdot \left( (d_{x_0}^{p-1})^{q-1} \nabla d_{x_0} \right) \\ &= \nabla \cdot (d_{x_0} \nabla d_{x_0}) = \Delta \frac{d_{x_0}^2}{2},\end{aligned}$$

thus the result might not give any new results in the smooth setting.

*Proof of the Remark.* Note first that $d_{x_0}^p/p$ is $c_p$-concave and because $|\nabla d_{x_0}| = 1$ almost everywhere and by the chain rule $|\nabla(d_{x_0}^p/p)| = d_{x_0}^{p-1}$. □

## 5.5 $c_p$-concavity of Busemann functions

In [Gig13] Gigli used, beside many other things, $c_2$-concavity of the Busemann function and linearity of the Laplacian to prove the splitting theorem for $RCD(K,N)$-spaces, i.e. $CD(K,N)$-spaces with a linear Laplacian. We will show that the Busemann function is $c_p$-concave for any $p \in (1,\infty)$, even more general it is $c_L$-concave. In the non-linear setting and the case $p = 2$, Ohta [Oht13b] used a comparison principle to show that Busemann functions on Finsler manifolds are harmonic. The author believes that such a principle also holds in a more general non-linear setting and even for the case $p \neq 2$ so that one can conclude harmonicity (resp. $p$-harmonicity) of Busemann functions.

A function $\gamma : [0,\infty) \to M$ is called geodesic ray if for any $T > 0$ the restriction to $[0,T]$ is a minimal geodesic. Furthermore, we will always assume that a geodesic rays are parametrized by arc length. We can define the Busemann function $b$ associated to $\gamma$ by

$$b(x) = \lim_{t\to\infty} b_t(x) \qquad \text{where } b_t(x) = d(x,\gamma_t) - t.$$

Note

$$t \mapsto b_t(x) \qquad \text{is non-increasing}$$

**Lemma 5.16.** *Let $(M,d)$ be a geodesic space and $b$ be the Busemann functions associated to some geodesic ray $\gamma : [0,\infty) \to X$. Then $b$ is $c_p$-concave.*

*Proof.* From Lemma 3.2 we know $b^{c_p\bar{c}_p} \geq b$, so that we only need to show the opposite inequality.

Fix an arbitrary $x \in X$ and $t \geq 0$ and let $\gamma^{t,x} : [0, d(x,\gamma_t)] \to X$ be a unit speed geodesic connecting $x$ and $\gamma_t$. Then for any $t \geq t_x$ we have $d(x,\gamma_t) \geq 1$ and

$$b^{c_p\bar{c}_p}(x) = \inf_{y\in X} \sup_{\tilde{x}\in X} \frac{d^p(x,y)}{p} - \frac{d^p(\tilde{x},y)}{p} + b(\tilde{x}) \leq \sup_{\tilde{x}\in X} \frac{1}{p} - \frac{d^p(\tilde{x},\gamma_1^{t,x})}{p} + b_t(\tilde{x}).$$

Furthermore, for any $\tilde{x} \in X$ and $t \geq t_x$ we also have

$$\begin{aligned}
\frac{1}{p} - \frac{d^p(\tilde{x}, \gamma_1^{t,x})}{p} + b_t(\tilde{x}) &= \frac{1}{p} - \frac{d^p(\tilde{x}, \gamma_1^{t,x})}{p} + d(\tilde{x}, \gamma_t) - t \\
&\leq \frac{1}{p} - \frac{d^p(\tilde{x}, \gamma_1^{t,x})}{p} + d(\tilde{x}, \gamma_1^{t,x}) + d(\gamma_1^{t,x}, \gamma_t) - t \\
&= -\frac{p-1}{p} - \frac{d^p(\tilde{x}, \gamma_1^{t,x})}{p} + d(\tilde{x}, \gamma_1^{t,x}) + d(x, \gamma_t) - t \\
&\leq d(x, \gamma_t) - t = b_t(x)
\end{aligned}$$

where we used Young's inequality and $(p-1)/p = 1/q$. Therefore,

$$b^{c_p \bar{c}_p}(x) \leq \lim_{t \to \infty} b_t(x) = b(x).$$

□

Actually, we can also show that the Busemann function is $c_L$-concave for any convex functional $L$ such that $c_L(x, y) = L(d(x, y))$ (see chapter on Orlicz-Wasserstein spaces).

**Lemma 5.17.** *Let $(M, d)$ be a geodesic space and $b$ be the Busemann functions associated to some geodesic ray $\gamma : [0, \infty) \to X$. Then $b$ is $c_L$-concave where such that $c_L(x, y) = L(d(x, y))$ for some convex function $L : [0, \infty) \to [0, \infty)$ such that $L^*(1) = r - L(r)$ for some $r \geq 0$.*

*Remark.* The condition for such an $r$ to exist rather weak, e.g. superlinearity of $L$ is sufficient.

*Proof.* Let $L^*$ be the Legendre transform of $L$, then Young's inequality holds

$$xy \leq L(x) + L^*(y),$$

in particular $x \leq L(x) + L^*(1)$.

Let $r$ be such that $L^*(1) = r - L(r)$. As above, we only need to show that $b^{c_L \bar{c}_L} \leq b$. We have

$$b^{c_L \bar{c}_L}(x) = \inf_{y \in X} \sup_{\tilde{x} \in X} L(d(x, y)) - L(d(\tilde{x}, y)) + b(\tilde{x}) \leq \sup_{\tilde{x} \in X} L(r) - d(\tilde{x}, \gamma_r^{t,x}) + b_t(\tilde{x}).$$

Furthermore, for all $\tilde{x} \in M$ and $t \geq t_x$ such that $d(x, \gamma_t) \geq r$ we get

$$\begin{aligned}
L(r) - L(d(\tilde{x}, \gamma_r^{t,x})) + b_t(\tilde{x}) &= L(r) - L(d(\tilde{x}, \gamma_r^{t,x})) + d(\tilde{x}, \gamma_t) - t \\
&\leq L(r) - L(d(\tilde{x}, \gamma_r^{t,x})) + d(\tilde{x}, \gamma_1^{t,x}) + d(\gamma_1^{t,x}, \gamma_t) - t \\
&= L(r) - r - L(d(\tilde{x}, \gamma_r^{t,x})) + d(\tilde{x}, \gamma_r^{t,x}) + d(x, \gamma_t) - t \\
&= -L^*(1) - L(d(\tilde{x}, \gamma_r^{t,x})) + d(\tilde{x}, \gamma_r^{t,x}) + d(x, \gamma_t) - t \\
&\leq d(x, \gamma_t) - t = b_t(x).
\end{aligned}$$

where we used Young's inequality to get the last inequality. Therefore,

$$b^{c_L \bar{c}_L}(x) \leq \lim_{t \to \infty} b_t(x) = b(x).$$

□

# 6 Gradient flow identification

Let $\mu \in \mathcal{P}(M)$ be some reference measure, we define the functional $\mathcal{U}_\mu : \mathcal{P}(M) \to \mathbb{R} \cup \{\infty\}$ by

$$\mathcal{U}_\mu(\nu) = \int U(\rho) d\mu + U'(\infty)\nu_s(M)$$

where $\nu = \rho\mu + \mu_s$ the the Lebesgue decomposition of $\nu$ w.r.t. $\mu$.

In the following we usually fix a metric measure space $(M, d, \mu)$ and drop the subscript $\mu$ from the functional $\mathcal{U}_\mu$. In addition, we use $\mathcal{U}_m$, $\mathcal{U}_\alpha$ etc. to denote the functional generated by $U_m$, $U_\alpha$, etc.

Now let

$$U_p(x) = \frac{1}{(3-p)(2-p)}(x^{3-p} - x)$$

and let $\mathcal{U}_p$ be the associated functional.

*Remark.* The linear term in $U_p$ is just for cosmetic reasons, it does not have any influence: Take $U = c \cdot x$ with $c > 0$ and let $\mathcal{U}$ be the associated functional, then $U'(\infty) = c$ for $p \in (2, 3)$ and thus

$$\mathcal{U}(\nu) = c \int f d\mu + c \cdot \nu^s(M) = c$$

where $\nu = \rho\mu + \nu^s$ is the Lebesgue decomposition w.r.t. $\mu$. Therefore, we have $\mathcal{U}_p(\nu) = \tilde{\mathcal{U}}_p(\nu) - \frac{1}{(3-p)(2-p)}$ with $\tilde{U}_p(x) = \frac{1}{(3-p)(2-p)} x^{3-p}$. For $p \in (1, 2)$ we have $\mathcal{U}_p(\nu) < \infty$ iff $\nu^s = 0$ and hence the linear term is constant as well.

Following the strategy in [AGS13, Section 7.2 and 8] we will show that under a curvature condition the $q$-heat flow can be identified with the gradient flow of the function $\mathcal{U}_p$ in the $p$-Wasserstein space: More precisely, if $p \in (1, 2)$ then $3 - p \in (1, 2)$ and the functional is displacement convex if the strong version of $CD_p(K, \infty)$ holds for some $K \geq 0$. If $p \in (2, 3)$ we have $3 - p \in (0, 1)$ so that $\mathcal{U}_p$ is displacement convex if $CD_p(0, N)$ holds with $1 - \frac{1}{N} = 3 - p$.

*Remark.* Note, that in contrast to the case $p = 2$, the strong version of $CD_p(K, \infty)$ does not imply $K$-convexity of functionals in $\mathcal{DC}_\infty$ for $K < 0$ and $p < 2$. We get $K'$-convexity in those cases if the space is bounded (see remark on page 48). Also Ohta and Takatsu could show that on a weighted Riemannian manifold of non-negative Ricci curvature the functional $\mathcal{U}_p$ with $p \in (2, 3)$ is $K$-convex in $\mathcal{P}_2(M)$ if $CD(K, N)$ holds (see [OT11a, Theorem 4.1]). This is, however, not enough for $p \neq 2$.

Recall from the introduction that $r > 0$ will be an abbreviation for $q \in (\frac{1+\sqrt{5}}{2}, \infty)$, or equivalently $p \in (1, \frac{3+\sqrt{5}}{2})$.

**Lemma 6.1.** *Assume $r > 0$ then*

$$\nu \mapsto \mathcal{U}_p(\nu)$$

*is lower semicontinuous in $\mathcal{P}_p(M)$.*

*Proof.* Just note that $U_p$ is convex and for $r > 0$ we have $p \in (1, \frac{3+\sqrt{5}}{2}) \subset (1,3)$ and thus $3 - p > 0$. □

*Remark.* The functional $\mathcal{U}_p$ appeared in a similar form already in [Gig12, Proof of Lemma 3.13] and Otto's preprint [Ott96] and also Augeh's thesis [Agu02, Agu05]. Gigli used the functional and the gradient flow of the $q$-Cheeger energy to show that all gradients of $q$-Sobolev functions can be weakly represented by a plan. In the Euclidean case, Otto and Augeh showed that the parabolic $q$-Laplace equation, which is the $q$-heat flow for smooth solutions, can be solved using the gradient flow of $\mathcal{U}_p$ in the $p$-Wasserstein case. This should also be compared to [OT11a, OT11b], where the (parabolic) porous media equation is solved via a gradient flow of a similar functional in the 2-Wasserstein space for Riemannian manifolds of non-negative Ricci curvature. Note, however, no identification is done. Furthermore, our approach shows that the abstract solution of the $q$-heat flow solves the gradient flow problem in the $p$-Wasserstein space.

## 6.1 Gradient flow of the Cheeger energy in $L^2$

We assume now that $\mathrm{Ch}_q$ is the $q$-Cheeger energy on $(M, d, \mu)$ where $(M, d)$ is a proper metric space and $\mu$ is a $\sigma$-finite measure. From [AGS13, Proposition 4.1] we know that the domain of $\mathrm{Ch}_q$ is dense in $L^2(M, \mu)$.

Since $L^2(M, \mu)$ is Hilbert and $\mathrm{Ch}_q$ is convex and lower semicontinuous, we can apply the classical theory of gradient flows developed in [Bré73] (see also [AGS08]). For that recall that the subdifferential $\partial^- \mathrm{Ch}_q$ at $f \in D(\mathrm{Ch}_q)$ is defined as (possibly empty) set of functions $\ell \in L^2(M, \mu)$ such that for all $g \in L^2(M, \mu)$

$$\int \ell(g - f) d\mu \le \mathrm{Ch}_q(g) - \mathrm{Ch}_q(f).$$

If $f \notin D(\mathrm{Ch}_q)$ then $\partial \mathrm{Ch}_q(f) = \varnothing$. The domain $D(\partial \mathrm{Ch}_q)$ of $\partial^- \mathrm{Ch}_q$ will be all $f \in L^2(M, \mu)$ such that $\partial^- \mathrm{Ch}_q \neq \varnothing$, which is dense in $L^2(M, \mu)$ (see [Bré73, Proposition 2.11]).

By [Bré73] the gradient flow of $\mathrm{Ch}_q$ gives for all $f_0 \in L^2(M, \mu)$ a locally Lipschitz map $t \mapsto f_t = H_t(f_t)$ (we drop $q$ if no confusion arises) from $(0, \infty)$ to $L^2(M, \mu)$ and $f_t \to f_0$ in $L^2(M, \mu)$ as $t \to 0$ and the derivative satisfies

$$\frac{d}{dt} f_t \in -\partial^- \mathrm{Ch}_q(f_t) \qquad \text{for a.e. } t \in (0, \infty).$$

**Definition 6.2** ($q$-Laplacian)**.** Let $f \in D(\partial \mathrm{Ch}_q)$ then $\Delta_q f$ is defined as the element $\ell \in -\partial^- \mathrm{Ch}_q(f)$ of minimal $L^2$-norm.

By [Bré73, Theorem 3.2] we have the regularization effect that $\frac{d^+}{dt} f_t$ exists everywhere in $(0, \infty)$ and is the element $\ell \in -\partial^- \operatorname{Ch}_q(f_t)$ with minimal $L^2$-norm, i.e. $\frac{d^+}{dt} f_t = \Delta_q f_t$.

*Remark.* We can also define the $L$-Laplacian using the same theory where $L$ is a convex increasing function with $L(0) = 0$. Since such flows might be interesting in combination with Orlicz-Wasserstein spaces, we will analyze these flows in the future.

**Proposition 6.3** (Properties of the Laplacian)**.** *If $f \in D(\Delta_q)$ and $g \in D(\operatorname{Ch}_q)$ then*

$$-\int g \Delta_q f d\mu \leq \int |\nabla g|_* |\nabla f|_*^{q-1} d\mu.$$

*Equality holds if $g = \phi(f)$ for some Lipschitz function $\phi : J \to \mathbb{R}$ with $J$ a closed interval containing the image of $f$ (and $\phi(0) = 0$ if $\mu(M) = \infty$). In that case one also has*

$$-\int \phi(f) \Delta_q f d\mu = \int \phi'(f) |\nabla f|_*^q d\mu.$$

*If, in addition, $g \in D(\Delta_q)$ and $\phi$ is nondecreasing and Lipschitz on $\mathbb{R}$ with $\phi(0) = 0$ then*

$$\int (\Delta_q g - \Delta_q f) \phi(g - f) d\mu \leq 0.$$

*Proof.* The first two parts were already proven in [AGS11a, Proposition 6.5] for $C^1$-functions $\phi$. However, using the proof of [AGS13, Proposition 4.15], adapted to $p \neq 2$, this can be proven in the same way. For convenience we include the full proof: Since $-\Delta_q f \in \partial^- \operatorname{Ch}_q(f)$ we have for all $\epsilon > 0$

$$\operatorname{Ch}_q(f) - \int \epsilon g \Delta_q f d\mu \leq \operatorname{Ch}_q(f + \epsilon g).$$

Furthermore, $|\nabla f|_* + \epsilon |\nabla g|_*$ is a relaxed slope of $f + \epsilon g$, we get

$$\begin{aligned} -\int \epsilon g \Delta f &\leq \frac{1}{q} \int (|\nabla f|_* + \epsilon |\nabla g|_*)^q - |\nabla f|_*^q d\mu \\ &= \epsilon \int |\nabla g|_* |\nabla f|_*^{q-1} d\mu + o(\epsilon). \end{aligned}$$

Dividing by $\epsilon$ and letting $\epsilon \to 0$ we obtain the result.

In case $g = \phi(f)$ we apply the chain rule and get $|\nabla(f + \epsilon\phi(f))|_* = (1 + \epsilon\phi'(f))|\nabla f|_*$ and thus

$$\begin{aligned} \operatorname{Ch}_q(f + \epsilon\phi(f)) - \operatorname{Ch}_q(f) &= \frac{1}{q} \int |\nabla f|_*^q ((1 + \epsilon\phi(f))^q - 1) d\mu \\ &= \epsilon \int \phi'(f) |\nabla f|_*^q d\mu + o(\epsilon). \end{aligned}$$

For the third part, just set $h = \phi(g - f)$, then $h \in D(\operatorname{Ch}_q)$ and for $\epsilon > 0$

$$\begin{aligned} -\epsilon \int (\Delta_q f - \Delta_q g) h d\mu &= -\epsilon \int \Delta_q f \cdot h d\mu - \epsilon \int \Delta_q g \cdot (-h) d\mu \\ &\leq \operatorname{Ch}_q(f + \epsilon h) - \operatorname{Ch}_q(f) + \operatorname{Ch}_q(g - \epsilon h) - \operatorname{Ch}_q(g). \end{aligned}$$

Taking $\epsilon$ sufficiently small such that $\epsilon\phi$ is a contraction, we can apply Proposition 2.3 and conclude. □

Actually with the help of Proposition 2.3 we can also prove:

**Proposition 6.4.** *If $f, g \in D(\Delta_L)$ and $\phi$ is nondecreasing and Lipschitz on $\mathbb{R}$ with $\phi(0) = 0$ then*

$$\int (\Delta_L g - \Delta_L f)\phi(g - f)d\mu \leq 0.$$

*Proof.* As above set $h = \phi(g - f)$, then $h \in D(\mathrm{Ch}_q)$ and for $\epsilon > 0$

$$\begin{aligned} -\epsilon \int (\Delta_L f - \Delta_L g)h d\mu &= -\epsilon \int \Delta_L f \cdot h d\mu - \epsilon \int \Delta_L g \cdot (-h) d\mu \\ &\leq \mathrm{Ch}_L(f + \epsilon h) - \mathrm{Ch}_L(f) + \mathrm{Ch}_L(g - \epsilon h) - \mathrm{Ch}_L(g). \end{aligned}$$

Then conclude by taking $\epsilon$ sufficiently small and applying Proposition 2.3. □

Using these results we can generalize [AGS13, Theorem 4.16] to the case $p \neq 2$ (and also [AGS11a, Proposition 6.6] where $0 < c \leq f_0 \leq C < \infty$ is required).

**Theorem 6.5** (Comparison principle and contraction)**.** *Let $f_t = H_t(f_0)$ and $g_t = H_t(g_0)$ be the gradient flows of $\mathrm{Ch}_q$ starting from $f_0, g_0 \in L^2(M, \mu)$ respectively. Then the following holds:*

1. *(Comparison principle) Assume $f_0 \leq C$ (resp. $f_0 \geq c$). Then $f_t \leq C$ (resp. $f_t \geq c$) for every $t \geq 0$. Similarly, if $f_0 \leq g_0 + C$ for some constant $C \in \mathbb{R}$, then $f_t \leq g_t + C$.*
2. *(Contraction) If $e : \mathbb{R} \to [-l, \infty]$ is a convex lower semicontinuous function and $E(f) = \int e(f)d\mu$ is the associated convex and lower semicontinuous functional in $L^2(M, \mu)$ then*
$$E(f_t) \leq E(f_0) \qquad \text{for every } t \geq 0,$$
*and*
$$E(f_t - g_t) \leq E(f_0 - g_0) \qquad \text{for every } t \geq 0.$$
*In particular, $H_t : L^2(M, \mu) \to L^2(M, \mu)$ is a contraction on $L^2(M, \mu) \cap L^r(M, \mu)$ w.r.t. the $L^r(M, \mu)$-norm for all $r \geq 1$, i.e. for all $f_0, g_0 \in L^2(M, \mu) \cap L^r(M, \mu)$ then*
$$\|H_t(f_0) - H(g_0)\|_r \leq \|f_0 - g_0\|_r.$$
3. *If $e : \mathbb{R} \to [0, \infty]$ is locally Lipschitz in $\mathbb{R}$ and $E(f_0) < \infty$ then*
$$E(f_t) + \int_0^t \int e''(f_t)|\nabla f_t|_*^q d\mu ds = E(f_0) \ \forall t \geq 0.$$
4. *When $\mu(M) < \infty$ we have*
$$\int f_t d\mu = \int f_0 d\mu \qquad \text{for every } t \geq 0.$$

*Remark.* (1) The first two assertions also hold for the gradient flow of the $L$-Cheeger energy, we will leave the details to the reader.

*Proof.* The proof follows along the lines of [AGS13, Theorem 4.16]. We will only show the result assuming $e'$ is bounded and globally Lipschitz. By the same approximation as in [AGS13, Theorem 4.16] the result follows.

Note first that the first statement follows by choosing $e(r) = \max\{r - C, 0\}$ (resp. $e(r) = \max\{c - r, 0\}$).

So let $e'$ be bounded and Lipschitz on $\mathbb{R}$ then for $x, y \in \mathbb{R}$ we have

$$|e'(x)| \leq |e'(0)| + \mathrm{Lip}(e')|x|,$$

$$|e(y) - e(x) - e'(x)(y - x)| \leq \frac{1}{2}\,\mathrm{Lip}(e')|y - x|^2$$

$$|e(y) - e(x)| \leq \left(|e'(0)| + \mathrm{Lip}(e')\right)\left(|x| + |y - x|\right)|y - x|,$$

where we assume $e'(0) = e(0) = 0$ if $\mu(M) = \infty$. Furthermore, we will assume w.l.o.g. $E(f_0 - g_0) < \infty$ (which forces $e(0) = 0$ if $\mu(M) = \infty$).

By convexity of $\mathrm{Ch}_q$ the maps $t \mapsto f_t$ and $t \mapsto g_t$ are locally Lipschitz continuous in $(0, \infty)$ with values in $L^2(M, \mu)$ (see [AGS08, Theorem 2.4.15] and [Bré73, Theorem 3.2]). Thus, the map $t \mapsto e(f_t - g_t)$ is locally Lipschitz in $(0, \infty)$ with values in $L^1(M, \mu)$, in particular, wherever $t \mapsto f_t$ and $t \mapsto g_t$ are commonly differentiable, we have

$$\begin{aligned} \frac{d}{dt}e(f_t - g_t) &= e'(f_t - g_t)\frac{d}{dt}(f_t - g_t) \\ &= e'(f_t - g_t)(\Delta_q f_t - \Delta_q g_t) \leq 0. \end{aligned}$$

Hence the function is $t \mapsto E(f_t - g_t)$ is locally Lipschitz in $(0, \infty)$. Integrating we see that the second assertion holds.

For the third statement, set $g_0 = g_t = 0$. Absolute continuity of $t \mapsto E(f_t)$ and the previous theorem yields for $\phi = e'$

$$\frac{d}{dt}\int e(f_t)d\mu = \int e'(f_t)\Delta_q f_t d\mu = -\int e''(f_t)|\nabla f_t|_*^q d\mu.$$

In case $\mu(M) < \infty$ we can choose $e(r) = r$ and thus

$$\frac{d}{dt}\int f_t d\mu = -\int 0 \cdot |\nabla f_t|_*^q d\mu$$

and hence $\int f_t d\mu = \int f_0 d\mu$. □

In order to prove mass preservation for $\mu(M) = \infty$ we adjust [AGS13, Section 4.4]. First we recall some facts about the $p$-logarithm (see also [OT11a, Section 3]) which will make the notation below easier.

**Lemma 6.6.** *The following inequality holds for $p \in (2,3)$, $x \geq 0$ and $V \geq 0$*

$$\begin{aligned} x \ln_p x &\geq x - \exp_p(-V^p) - (p-2)V^p \exp_p(-V^p) \\ &\quad +(p-3)V^p x \end{aligned}$$

*where*

$$\exp_p(t) = \{1 + (2-p)t\}^{\frac{1}{2-p}}$$

*and*

$$\ln_p(s) = \frac{s^{2-p} - 1}{2-p}$$

*which are inverse of each other for $t \in (-\infty, \frac{1}{p-2}]$. Note also that $\exp_p$ is monotone on its domain and for sufficiently small $h$*

$$\exp_p(h) \cdot \exp_p(-h) \leq 2.$$

*Proof.* Note first that $x \ln_p x$ is convex and thus

$$\begin{aligned} x \ln_p x &\geq x_0 \ln_p x_0 + (\ln_p x_0 + x_0^{2-p})(x - x_0) \\ &= x \ln_p x_0 + x_0^{2-p}(x - x_0). \end{aligned}$$

Now choosing $x_0 = \exp_p(-V^p) \geq 0$ then

$$\begin{aligned} x_0^{2-p}(x - x_0) &= \{1 + (p-2)V^p\}^{\frac{2-p}{2-p}} (x - \exp_p(-V^p)) \\ &= x - exp_p(-V^p) - (p-2)V^p \exp_p(-V^p) + (p-2)V^p x \end{aligned}$$

Since $x \ln_p x_0 = -V^p x$ we see that

$$x \ln_p x \geq x - \exp_p(-V^p) - (p-2)V^p \exp_p(-V^p) + (p-3)V^p x.$$

$\square$

**Lemma 6.7** (Momentum-entropy estimate)**.** *Assume $p \in (1,3)$. Let $\mu$ be a finite measure and $V : X \to [0, \infty)$ be a Lipschitz function with $V \geq \epsilon > 0$ such that*

$$I_p := \begin{cases} 0 & \text{if } p \in (1,2) \\ \frac{p-2}{3-p} \int V^p \exp_p(-V^p) d\mu & \text{if } p \in (2,3) \end{cases}$$

*is finite and if $p \in (2,3)$ assume in addition*

$$\int \exp_p(-V^p) d\mu \leq 1$$

*Let $f_0 \in L^2(X, \mu)$ be non-negative with*

$$\int V^p f_0 d\mu < \infty$$

*and for some $z > 0$*

$$z \int \exp_p(-V^p) d\mu \leq \int f_0 d\mu$$

*if $p \in (2,3)$ and otherwise choose $z \leq 1$. Then $t \mapsto \int V^p f_t d\mu$ is locally absolutely continuous in $[0,\infty)$ and for every $t \geq 0$*

$$\int V^p f_t d\mu \leq S_t$$

*and*

$$\int_0^t \int_{\{f_s > 0\}} \frac{|\nabla f_s|_*^q}{f_s^{p-1}} d\mu ds \leq \frac{4}{3-p} S_t$$

*where*

$$S_t = e^{C_p \operatorname{Lip}(V)^q t} \left( I_p + \int \frac{1}{(2-p)} (f_0^{3-p} - f_0) + (pV^p + z^{-1} l_p) f_0 d\mu \right)$$

*with $C_p = (p \cdot (3-p)^{-1})^q / q$ and $l_p = \max\{\frac{1}{2-p}, 1\}$.*

*Proof.* Define the following

$$M^q(t) := \int V^p f_t d\mu, \qquad E(t) := \tfrac{1}{(3-p)(2-p)} \int f_t^{3-p} - f_t d\mu,$$
$$F^p(t) := \int_{\{f_t > 0\}} \tfrac{|\nabla f_t|_*^q}{f_t^{p-1}} d\mu.$$

Applying Theorem 6.5 (see remark below that theorem) to $(f_t + \epsilon) = H_t(f_t + \epsilon)$ and letting $\epsilon \to 0$ we see that $F \in L^p(0,T)$ for every $T > 0$ and

$$\frac{d}{dt} E(t) = -F^p(t) \text{ a.e. in } (0,T).$$

Furthermore, by the Lemma above, conservation of mass and the assumption $\int \exp_p(-V^p) d\mu \leq 1$, we have for $p \in (2,3)$

$$\begin{aligned}
(3-p)E(t) &= \int f_t \ln_p f_t d\mu \\
&\geq \int f_t - \exp_p(-V^p) d\mu - (p-2) \int V^p \exp_p(-V^p) d\mu \\
&\quad + (p-3) M^q(t) \\
&\geq \int f_0 - \exp_p(-V^p) d\mu - I_p - (p-1) M^q(t) \\
&\geq (1 - z^{-1}) \int f_0 d\mu - I_p + (p-3) M^q(t) \\
&\geq -z^{-1} l_p \int f_0 d\mu - I_p + (p-3) M^q(t)
\end{aligned}$$

For $p \in (1,2)$ note that $\frac{1}{(2-p)} x^{3-p} \geq 0$ and hence

$$\begin{aligned}(3-p)E(t) &= \frac{1}{(2-p)} \int f_t^{3-p} - f_t d\mu \\ &\geq -\frac{1}{(2-p)} \int f_0 d\mu \\ &\geq -z^{-1} l_p \int f_0 d\mu - I_p + (p-3) M^q(t).\end{aligned}$$

In order to estimate the derivative of $M(t)$ we introduce a truncated weight $V_k(x) = \min\{V(x), k\}$ and the corresponding functional $M_k^q(t)$ as above. We know that the function $t \mapsto M_k^q(t)$ is locally Lipschitz continuous and thus for a.e. $t > 0$

$$\begin{aligned}\left|\frac{d}{dt} M_k^q(t)\right| &= \left|\int V_k^p \Delta_q f_t d\mu\right| \\ &\leq p \int V_k^{p-1} |\nabla V_k|_* |\nabla f_t|_*^{q-1} d\mu \\ &\leq pL \int \left(V_k^{p-1} f_t^{\frac{1}{q}}\right) \cdot \left(\frac{|\nabla f_t|_*^{q-1}}{f_t^{\frac{1}{q}}}\right) d\mu \\ &\leq pLF(t) M_k(t)\end{aligned}$$

using $\operatorname{Lip} V_k \leq L$ and Hölder inequality (note $(p-1)q = p$).

Since by mass preservation $M_k(t) \geq \tilde{\epsilon} := \epsilon \int f_0 d\mu$, we can apply Gronwall's inequality and get

$$M_k^q(t) \leq M_k^q(0) \exp\left(\int_0^t \frac{pLF(s)}{M_k^{q-1}(s)} ds\right) \leq M^q(0) \exp\left(\int_0^t \frac{pLF(s)}{\tilde{\epsilon}^{q-1}} ds\right)$$

for $t \in [0, N]$. Thus $M_k^q(t)$ is uniformly bounded and by monotone convergence, we obtain the same differential inequality for $M^q(t)$, i.e. for $t \in [0, \infty)$

$$\left|\frac{d}{dt} M^q(t)\right| = pLF(t)M(t).$$

Now combining this with the result above we get

$$\frac{d}{dt}((3-p)E + pM^q) + (3-p)F^p \leq p^2 LFM \leq (3-p)F^p + C_p L^q M^q$$

where

$$C_p = (p \cdot (3-p)^{-1})^q / q.$$

Combining this with the inequality above, we get by the Gronwall inequality

$$\begin{aligned}-z^{-1}l_p\int f_0 d\mu - I_p + (2p-3)M^q(t) &\leq (3-p)E(t) + pM^q(t)\\ &\leq e^{C_pL^qt}\left((3-p)E(0)+pM^q(0)\right).\\ &\leq e^{C_pL^qt}\Big((3-p)E(0)+pM^q(0)\\ &\qquad +z^{-1}l_p\int f_0 d\mu + I_p\Big).\end{aligned}$$

Furthermore, we have

$$(3-p)\int_0^t F^p(s)ds \leq (3-p)(E(0)-E(t)) \leq (3-p)E(0)+I_q+z^{-1}l_p\int f_0 d\mu+(3-p)M^q(t).$$

□

Having established this, similar to [AGS13, Theorem 4.20] we can show that the gradient flow of the $q$-Cheeger energy is mass preserving even if the measure $\mu$ is just $\sigma$-finite. The proof relies on an approximation procedure developed in [AGS13, Section 4.3]. We will freely use the concepts and results during the proof. The reader may consult [AGS13, Section 4.3] for further reference.

**Theorem 6.8.** *Assume $p \in (1,\infty)$. If $\mu$ is a $\sigma$-finite measure such that for some Lipschitz function $V : X \to [\epsilon,\infty]$ for some $\epsilon > 0$ such that for $p \in (2,3)$*

$$\int \exp_p(-V^p)d\mu \leq 1$$

*and*

$$\int V^p \exp_p(-V^p)d\mu < \infty$$

*and for $p \in (1,2)$ there is an increasing function $\Phi : \mathbb{R} \to [0,\infty]$ such that*

$$\int \Phi(-V^p)d\mu \leq 1.$$

*Then the gradient flow $H_t$ of the q-Cheeger energy is mass preserving, i.e. for $f_t = H_t(f_0)$ with $\int f_0 d\mu < \infty$*

$$\int f_t d\mu = \int f_0 d\mu.$$

*Moreover, if $f_0 \in L^2(M,\mu)$ is nonnegative and*

$$\int V^p f_0 d\mu, \int f_0 d\mu < \infty$$

*then the bound of the previous Lemma hold.*

*Proof.* We will use the construction of [AGS13, Theorem 4.20], see in particular, [AGS13, Proposition 4.17]. By homogeneity of the $H_t$, i.e. $H_t\lambda f = \lambda^q H_t f$, we can assume $\int f_0 d\mu \leq 1$ if $\int f_0 d\mu < \infty$. In case $\int f_0 d\mu = \infty$, we can find a sequence $f_0^n \leq f_0$ such that $n \leq \int f_0^n d\mu < \infty$. Since mass preservation holds for those functions, we can use the comparison principle to show that $\int f_t d\mu \geq n$ for all $n$ and hence it also holds in the case $\int f_0 d\mu = \infty$. So w.l.o.g. $\int f_0 d\mu \leq 1$.

We first show the case $p \in (2,3)$. For that use the following approximation: $\mu^0 := \exp_p(-V^p)\mu$ and $\mu^k := \exp_p(-V_k^p)\mu^0 =$ for $V_k := \min(V,k)$. Then $\mu^k$ is an increasing family of finite measures and

$$\lim \mu^k(B) = \mu(B) \quad \forall B \in \mathcal{B}(M).$$

Since $V$ is Lipschitz we see that the density of $\mu$ w.r.t. $\mu^0$ is bounded from below and above on any bounded set.

For each $\mu^k$ let $f_t^k = H_t^k(f_0)$ be the gradient flow starting at $f_0$. Then since $\int \exp_p(-V^p)d\mu^k \leq 1$ we can apply the previous lemma with $z_k = \int f_t^k d\mu^k$ for all $t \geq 0$ and obtain

$$\int V^p f_t^k d\mu^k \leq e^{2\,\mathrm{Lip}(V)^q t}\left(I_p + \int \frac{1}{(2-p)}(f_0^{3-p} - f_0) + (pV^p + z^{-1}l_p)f_0 d\mu^k\right).$$

Since $f_t^k \to f_t$ strongly in $L^2(X,\mu^0)$ (see [AGS13, Proposition 4.17]) we can assume up to changing to a subsequence $f_t^k \to f_t$ $\mu$-almost every where, and thus Fatou's lemma and monotonicity of $\mu^k$ implies

$$\int V^p f_t d\mu \leq \liminf_{k\to\infty} \int V^p f_t^k d\mu^k$$

and the bound of the previous lemma holds since $z_k \nearrow z = \int f_0 d\mu = 1$.

Now consider $A_h = \{x \in M \,|\, V(x) \leq h\}$. Since we assume $\int \exp_p(-V^p)d\mu \leq 1$ we can choose $h$ such that $\exp_p(h)\exp_p(-h) \leq 2$ and get by monotonicity

$$\mu(A_h) \leq \int 2\exp_p(h^p)\exp_p(-V^p)d\mu \leq 2exp_p(h^p) < \infty$$

and thus by (4.42) of [AGS13, Proposition 4.17]

$$\int_{A_h} f_t d\mu = \lim_{k\to\infty} \int_{A_h} f_t^k d\mu^k.$$

From the bound on the $p$-th moment we obtain for every $t > 0$ a constant $C > 0$ such that

$$h^q \int_{X\backslash A_h} f_t^k d\mu^k \leq C$$

for every $h > 0$ and hence

$$\begin{aligned}\int f_t d\mu &\geq \int_{A_h} f_t d\mu = \lim_{k\to\infty}\int_{A_h} f_t^k d\mu^k \\ &\geq z - \limsup_{k\to\infty}\int_{X\backslash A_h} f_t^k d\mu^k \geq z - C/h^p.\end{aligned}$$

Since $h$ is arbitrary and the integral of $f_t$ does not exceed $z$ we see that $\int f_t d\mu = z$. The second inequality of the previous lemma follows by lower semicontinuity of the Cheeger energy (see 2.5).

Mass preservation for signed initial data $f_0$ follows by the same arguments as in [AGS13, Theorem 4.20].

In order to treat the case $p \in (1,2)$ let $\Phi$ be increasing such that $\int \Phi(-V)d\mu \leq 1$ and construct a monotone approximation $\mu^k = \Phi(-V_k)\mu^0$ and proceed as above. □

*Remark.* Let $p \in (2,3)$ if $p \to 2$ then the condition

$$\int \exp_p(-V^p)d\mu \leq 1$$

converges to

$$\int \exp(-V^2)d\mu \leq 1$$

which is precisely the condition used in [AGS13, (4.2)]. Note, however, it is stronger: Assuming $p \in (2,3)$ and $(p-2)V^p \geq 1$ we have

$$\begin{aligned} \exp_p(-V^p) &= \{1 + (2-p)(-V^p)\}^{\frac{1}{2-p}} \\ &\leq \{2(p-2)V^p\}^{\frac{1}{2-p}} \\ &= CV^{\frac{p}{2-p}} \geq C\exp(-V^2) \end{aligned}$$

if $V$ is sufficiently large. In the Euclidean setting with $V(x) \approx \|x\|$ we get

$$\begin{aligned} \int_{\mathbb{R}^n \backslash B_1(0)} \exp_p(-V^p)d\lambda &\approx \int_{\mathbb{R}\backslash B_1(0)} \|x\|^{-\frac{p}{p-2}} d\lambda \\ &\approx \int_1^\infty r^{-\frac{p}{p-2}} r^{n-1} dr \end{aligned}$$

which is finite if $p < \frac{2n}{(n-1)}$, i.e. $q > \frac{2n}{n+1}$. However, note that we currently need the more restrictive condition

$$\int V^p \exp_p(-V^p)d\mu \approx \int_1^\infty r^{p-\frac{p}{p-2}} r^{n-1} d\mu$$

which is finite iff

$$p - \frac{p}{p-2} + n < 1,$$

i.e. $p < \frac{1}{2}\left(3 - n + \sqrt{n^2+2n+9}\right) \approx 2 + \frac{2}{n} - \frac{2}{n^2} - \frac{2}{n^3} + \mathcal{O}(\frac{1}{n^4})$ as $n \to \infty$.

## 6.2 Gradient flow in the Wasserstein space

Throughout this section we will assuming that the gradient flow of the $q$-Cheeger energy is mass preserving, i.e. the conditions of Theorem 6.8 hold. Furthermore, we assume that all slopes of Lipschitz functions are equal almost everywhere, i.e.

$$|Df| = |D^{\pm} f| \quad \mu\text{-almost everwhere.}$$

This condition holds if the space satisfies a local doubling and Poincaré condition, in particular if $CD_p(K,N)$ holds with $N < \infty$.

Our motivation for the functional $\mathcal{U}_p$ and the identification is the Kuwada lemma. It appeared the first time in [Kuw10] for $p = 2$ and was extended by Ambrosio-Gigli-Savaré to $p \neq 2$ for finite measures and $0 < c \le f_0 \le C < \infty$.

**Lemma 6.9** (Kuwada lemma)**.** *Let $f_0 \in L^q(M,\mu)$ be non-negative and $(f_t)_{t\in[0,\infty)}$ be the gradient flow of the $q$-Cheeger energy starting from $f_0$. Assume $\int f_0 d\mu = 1$. Then the curve $t \mapsto d\mu_t = f_t d\mu$ is absolutely continuous in $\mathcal{P}_p(M)$ and*

$$|\dot{\mu}_t|^p \le \int \frac{|\nabla f_t|_*^q}{f_t^{p-1}} d\mu \qquad \text{for almost every } t \in (0,1).$$

*Proof.* The proof follows from [AGS11a, Lemma 7.2] because using Theorem 6.5 above the requirement $0 < c \le f_0 \le C < \infty$ can be easily dropped. □

*Remark.* Formally this lemma can be extended to cover $\partial_t f_t = \Delta\phi(f_t)$, which includes the porous media equation, $\phi(r) = c_m \cdot r^m$. The theorems below hold with minor adjustments as well. However, since a general existence theory of such equations on abstract metric spaces is not available, an identification is difficult using our approach. This is exactly why Ohta-Takatsu [OT11a, OT11b] can only use the gradient flows in $\mathcal{P}_2$ to get a solution, but they do not identify the two flows.

We say that the measure $\mu$ is local $n$-Ahlfors if for every $R > 0$ there are constants $0 < c_R < C_R$ such that for all $x \in M$ and $0 < r < R$ we have

$$c_R r^n \le \mu(B_r(x)) \le C_R r^n.$$

**Proposition 6.10.** *Let $M$ be a proper metric measure space. In case $p \in (2,3)$ assume, in addition, that $M$ is compact and local $n$-Ahlfors regular for $3 - p > 1 - \frac{1}{n}$, i.e. $n(p-2) < 1$. If $r > 0$ and $\mathcal{U}_p(\mu_0) < \infty$, then $|D^-\mathcal{U}_p|(\mu_0) < \infty$ implies $\mu_0$ is absolutely continuous w.r.t. $\mu$ and if there is a sequence of absolutely continuous measure $\mu_n$ such that $w_p(\mu_0,\mu_n) \to 0$ and*

$$|D^-\mathcal{U}_p|(\mu_0) = \lim_{n\to 0} \frac{\mathcal{U}_p(\mu_0) - \mathcal{U}_p(\mu_n)}{w_p(\mu_0,\mu_n)}.$$

*Remark.* (1) The proof is extracted from [OT11a, Proof of Claim 7.7 and Remark 7.8]. It is stated in the smooth setting but also works in the Ahlfors regular case. The proof depends on the Ahlfors regularity to show that $\mathcal{U}_p(\hat{\mu}_r) < \mathcal{U}_p(\mu_0)$, but it might be interesting to know if Ahlfors regularity is really needed.

(2) The only time where this proposition is needed is during the proof of Theorem 6.12 which is based on [AGS13, Theorem 7.5]. In order to use the coupling technique and convexity absolute continuity of $\mu_n$ is essential.

*Proof.* Let $m = 3 - p$. In case $m > 1$ the measures $\mu_0$ and $\mu_n$ must be absolutely continuous. So we are left to show the cases $0 < m < 1$.

First assume $\mu_0$ has non-trivial singular part, i.e. $\mu_0 = f_0\mu + \mu^s$ where $\mu^s$ and $\mu$ are mutually singular. Define for each $r > 0$ a measure $\hat{\mu}_r$ as follows

$$d\hat{\mu}_r(x) = \rho_r(x)d\mu(x) := \left\{ f_0(x) + \int \frac{\chi_{B_r(y)}(x)}{\mu(B_r(y))} d\mu^s(y) \right\} d\mu(x).$$

Then we have

$$\begin{aligned}
\int \rho_r(x)^m d\mu &= \int \left[ \int \left\{ \frac{f_0(x)}{\mu^s(M)} + \frac{\chi_{B_r(y)}(x)}{\mu(B_r(y))} \right\} d\mu^s(y) \right]^m d\mu(x) \\
&\geq \mu^s(M)^{m-1} \int \left[ \int \left\{ \frac{f_0(x)}{\mu^s(M)} + \frac{\chi_{B_r(y)}(x)}{\mu(B_r(y))} \right\}^m d\mu^s(y) \right] d\mu(x) \\
&\geq \mu^s(M)^{m-1} \int \left[ \int_{M\backslash B_r(y)} \frac{f_0}{\mu^s(M)} d\mu + \int_{B_r(y)} \frac{1}{\mu(B_r(y))^m} d\mu \right] d\mu^s(y) \\
&= \int f_0^m d\mu - \mu^s(M)^{-1} \int \left( \int_{B_r(y)} f_0^m d\mu \right) d\mu^s(y) \\
&\quad + \mu^s(M)^{m-1} \int \mu(B_r(y))^{1-m} d\mu^s(y).
\end{aligned}$$

Ahlfors regularity implies that for some $C, c > 0$

$$c \cdot r^n \leq \mu(B_r(y)) \leq C \cdot r^n$$

and thus

$$\mu^s(M)^{m-1} \int \mu(B_r(y))^{1-m} d\mu^s(y) \geq c^{1-m} \mu^s(M)^m \cdot r^{n(1-m)}.$$

Furthermore, notice

$$\begin{aligned}
\int_{B_r(y)} f_0^m d\mu &\leq \left( \int_{B_r(y)} f_0 d\mu \right)^m \left( \int_{B_r(y)} d\mu \right)^{1-m} \\
&\leq \left( \int_{B_r(y)} f_0 d\mu \right)^m C^{1-m} r^{n(1-m)}.
\end{aligned}$$

Since $\lim_{r\to 0} \sup_{y \in M} \int_{B_r(y)} f_0 d\mu = 0$ we see that for sufficiently small $r > 0$ (note $(m-1) < 1$)

$$\mathcal{U}_m(\hat{\mu}_r) \leq \mathcal{U}_m(\mu_0) - \tilde{C}\mu^s(M)^m \cdot r^{n(1-m)}.$$

Furthermore, by our assumption

$$n(1-m) = n(p-2) < 1.$$

To estimate $w_p(\mu_0, \hat{\mu}_r)$ note that the density of $\hat{\mu}_r$ is defined as follows

$$\rho_r^s(x) := \int \frac{\chi_{B_r(y)}(x)}{\mu(B_r(y))} d\mu^s(y).$$

Now choose the following coupling $\pi$ between $\mu^s$ and $\rho_r^s \mu$

$$d\pi(x,y) = \int \frac{\chi_{B_r(z)}(x)}{\mu(B_r(z))} d\mu(x) d(\mathrm{Id} \times \mathrm{Id})_* \mu^s(z,y)$$

Then

$$\begin{aligned} w_p^p(\mu_0, \hat{\mu}_r) &\leq \frac{1}{p} \int d^p(x,y) d\pi(x,y) \\ &\leq \frac{1}{p} \int \int d^p(x,y) \frac{\chi_{B_r(z)}(x)}{\mu(B_r(z))} d\mu(x) d(\mathrm{Id} \times \mathrm{Id})_* \mu^s(z,y) \\ &\leq \frac{1}{p} \int r^p d\mu^s \end{aligned}$$

and thus

$$w_p(\mu_0, \hat{\mu}_r) \leq r \left( \frac{\mu^s(M)}{p} \right)^{\frac{1}{p}}.$$

Combining these we get

$$\begin{aligned} |D^- \mathcal{U}_m|(\mu_0) &\geq \limsup_{r \to 0} \frac{\mathcal{U}_N(\mu_0) - \mathcal{U}_N(\hat{\mu}_r)}{w_p(\mu_0, \hat{\mu}_r)} \\ &\geq \limsup_{r \to 0} \frac{\tilde{C} \mu^s(M)^m \cdot r^{n(1-m)}}{r \left( \frac{\mu^s(M)}{p} \right)^{\frac{1}{p}}} = \infty \end{aligned}$$

since $n(1-m) < 1$. Which implies that $\mu_0$ must be absolutely continuous.

For the second part, a similar argument works. Given $\mu_n$ we can construct $\hat{\mu}_n^r$ similar to $\hat{\mu}_r$. The estimates for $\mathcal{U}_m$ hold without any change. For the rest just note

$$\begin{aligned} \frac{w_p(\mu_0, \hat{\mu}_n^r)}{w_p(\mu_0, \mu_n)} &\leq \frac{1}{w_p(\mu_0, \mu_n)} \{ w_p(\mu_0, \mu_n) + w_p(\mu_n, \hat{\mu}_n^r) \} \\ &\leq 1 + \frac{1}{w_p(\mu_0, \mu_n)} r \left( \frac{\mu^s(M)}{p} \right)^{\frac{1}{p}}. \end{aligned}$$

Thus

$$\frac{\mathcal{U}_m(\mu_0) - \mathcal{U}_m(\hat{\mu}_n^r)}{w_p(\mu_0, \hat{\mu}_n^r)} \geq \frac{\mathcal{U}_m(\mu_0) - \mathcal{U}_m(\mu_n) + \tilde{C} \mu^s(M) r^{n(1-m)}}{w_p(\mu_0, \mu_n)} \cdot \left( 1 + \frac{r \mu_n^s(M)^{\frac{1}{p}}}{p^{\frac{1}{p}} w_p(\mu_0, \mu_n)} \right)^{-1}.$$

Now choosing $r_n$ such that $r_n/w_p(\mu_0,\mu_n)\to 0$ we see that

$$\limsup_{n\to\infty}\frac{\mathcal{U}_m(\mu_0)-\mathcal{U}_m(\hat{\mu}_n^{r_n})}{w_p(\mu_0,\hat{\mu}_n)}\geq\limsup\frac{\mathcal{U}_m(\mu_0)-\mathcal{U}_m(\mu_n)}{w_p(\mu_0,\mu_n)}$$

By maximality of $\mu_n$ we see that this has to be an equality, so up to extracting a subsequence we get

$$|D^-\mathcal{U}_p|(\mu_0)=\lim_{n\to 0}\frac{\mathcal{U}_p(\mu_0)-\mathcal{U}_p(\hat{\mu}_n^{r_n})}{w_p(\mu_0,\hat{\mu}_n)}.$$

□

**Theorem 6.11.** *Assume $r>0$ and let $\mu_0\in D(\mathcal{U}_p)$ with $|D^-\mathcal{U}_p|(\mu_0)<\infty$. Then $\mu_0=\rho\mu$, $\rho^r\in D(\mathrm{Ch}_q)$ and*

$$r^{-q}\int|\nabla\rho^r|_*^q d\mu\leq|D^-\mathcal{U}_p|^q(\mu_0).$$

*Proof.* We will follow the strategy of [AGS13, Theorem 7.4]. First assume $\rho\in L^2(M,\mu)$ and let $(\rho_t)_{t\in(0,\infty)}$ be the gradient flow of the $q$-Cheeger energy starting from $\rho$. Let $\mu_t=\rho_t\mu$ then according to the definition of the $q$-Fisher information we have by Lemma 6.5 and 6.9

$$\begin{aligned}\mathcal{U}_p(\mu_0)-\mathcal{U}_p(\mu_t) &\geq \frac{1}{q}\int_0^t\mathsf{F}_q(\rho_s)ds+\frac{1}{p}\int_0^t|\dot{\mu}_s|^p d\mu\\ &\geq \frac{1}{q}\left(\frac{1}{t^{\frac{1}{p}}}\int_0^t\sqrt[q]{\mathsf{F}_q(\rho_s)}ds\right)^q+\frac{1}{p}\left(\frac{1}{t^{\frac{1}{q}}}\int_0^t|\dot{\mu}_s|ds\right)^p\\ &\geq \frac{1}{t}\left(\int_0\sqrt[q]{\mathsf{F}_q(\rho_s)}ds\right)w_p(\mu_0,\mu_t).\end{aligned}$$

Thus dividing by $w_p(\mu_0,\mu_t)$ and letting $t\to 0^+$ we get the result, since lower-semicontinuity of $\mathsf{F}_q$ implies

$$\sqrt[q]{\mathsf{F}_q(\rho_0)}\leq\liminf_{t\to 0^+}\frac{1}{t}\int_0^t\sqrt[q]{\mathsf{F}_q(\rho_s)}ds.$$

In case just $\mathcal{U}_p(\mu_0)<\infty$ holds we prove the result by approximation: Let $\rho^n=\min\{\rho,n\}$ and $(\rho_t^n)$ be the corresponding gradient flow of the $q$-Cheeger energy. Using the comparison principle we see that $\rho_t=\lim_{n\to\infty}\rho_t^n$ almost everywhere. Thus using the fact that $z_n=\int\rho^n d\mu=\int\rho_t^n d\mu$ we deduce that $\mu_t^n=\frac{1}{z_n}\rho_t^n\mu$ converges to $\mu_t=\rho_t\mu$ in $\mathcal{P}_p(M)$. Now using the lower semicontinuity properties of $\mathcal{U}_p$ we deduce

$$\mathcal{U}_p(\mu_0)-\mathcal{U}_p(\mu_t)\geq\frac{1}{t}\left(\int_0^t\sqrt[q]{\mathsf{F}_q}ds\right)w_p(\mu_0,\mu_t)$$

and conclude as above. □

**Theorem 6.12.** *Assume $\mu$ is finite and, in addition if $p > 2$, assume also that $(M, d, \mu)$ is as in Proposition 6.10. Let $\mu_0 = \rho\mu \in D(\mathcal{U}_p)$ and assume $\rho$ is a bounded Lipschitz continuous map with $\rho \geq \epsilon$. Then*

$$|D^-\mathcal{U}_p|^q(\mu_0) \leq \int \frac{|D\rho|^q}{\rho^{p-1}} d\mu = r^{-q} \int |D\rho^r|^q d\mu,$$

*where $|D\rho|(x) = \max\{|D^+\rho|(x), |D^-\rho|(x)\}$.*

*Remark.* For $p < 2$, we have $2 - p > 0$ and the idea of [AGS13, Theorem 7.5] can be followed in a similar way using the approximation function $\Phi$ (see Theorem 6.8) so that a similar version to that theorem follows. For $p > 2$, we have $2 - p < 0$, so that an appropriate version requires further work. Note, however, that Proposition 6.10 requires $M$ to be compact and hence $\mu$ to be finite.

*Proof.* Recall that

$$\begin{aligned} U_p(r) &= \frac{1}{(3-p)(2-p)}(x^{3-p} - x) \\ \tilde{U}_p(r) &= \frac{1}{(3-p)(2-p)}x^{3-p}. \end{aligned}$$

Define

$$L(x, y) := \begin{cases} \frac{\left(\frac{1}{2-p}\rho^{2-p}(x) - \frac{1}{2-p}\rho^{2-p}(y)\right)^+}{d(x,y)} & \text{if } x \neq y \\ \frac{|D\rho|}{\rho^{p-1}} & \text{if } x = y. \end{cases}$$

Note that $L$ is measurable and for fixed $x \in M$ the map $y \mapsto L(x, y)$ is upper semicontinuous. Furthermore, since $\rho$ is Lipschitz and $\epsilon \leq \rho \leq M$, $L$ is bounded.

Now take a sequence of absolutely continuous measures $\mu_n$ with $w_p(\mu_0, \mu_n) \to 0$ and

$$|D^-\mathcal{U}_p|(\mu_0) = \lim_{n\to 0} \frac{\mathcal{U}_p(\mu_0) - \mathcal{U}_p(\mu_n)}{w_p(\mu_0, \mu_n)}.$$

Let $\rho_n$ be the density of $\mu_n$ w.r.t. $\mu$ and $\pi_n$ be some $c_p$-optimal transport plan of $(\mu_0, \mu_n)$. Because $r \mapsto U_p(r)$ is convex we have

$$\begin{aligned} \mathcal{U}_p(\mu_0) - \mathcal{U}_p(\mu_n) &= \int (U_p(\rho) - U_p(\rho_n))\, d\mu \leq \int U_p^{'}(\rho)(\rho - \rho_n) d\mu \\ &= \int U_p^{'}(\rho) d\mu_0 - \int U_p^{'}(\rho) d\mu_n = \int \left(\tilde{U}_p^{'}(\rho(x)) - \tilde{U}_p^{'}(\rho(y))\right) d\pi_n(x, y) \\ &\leq \int L(x, y) d(x, y) d\pi_n(x, y) \leq w_p(\mu_0, \mu_n) \left(\int L^q(x, y) d\pi_n(x, y)\right)^{1/q} \\ &= w_p(\mu_0, \mu_n) \left(\int \left(\int L^q(x, y) d\pi_{n,x}(y)\right) d\mu_0(x)\right)^{1/q} \end{aligned}$$

where $\pi_{n,x}$ is the disintegration of $\pi_n$ w.r.t. the first marginal $\mu_0$ and $\tilde{U}_p(x) = \frac{1}{(3-p)(2-p)}x^{2-p}$. Since $\int(\int d^p(x,y)d\pi_{n,x}(y))d\mu_0(x) \to 0$ we can assume w.l.o.g. that for $\mu_0$-a.e. $x \in M$

$$\lim_{n\to\infty} \int d^p(x,y)d\pi_{n,x}(y) = 0$$

and in particular

$$\int_{M\backslash B_r(x)} L^q(x,y)d\pi_{n,x}(y) \to 0$$

for all $r > 0$. Furthermore, notice

$$\begin{aligned} \limsup_{n\to\infty} \int L^q(x,y)d\pi_{n,x}(y) &\le \limsup_{n\to\infty} \int_{B_r(x)} L^q(x,y)d\pi_{n,x}(y) \\ &\quad + \limsup_{n\to\infty} \int_{M\backslash B_r(x)} L^q(x,y)d\pi_{n,x}(y) \\ &\le \limsup_{n\to\infty} \int_{B_r(x)} L^q(x,y)d\pi_{n,x}(y) \le \sup_{y\in B_r(x)} L^q(x,y). \end{aligned}$$

By upper semicontinuity of $L(x,\cdot)$ we immediately get $\limsup_n \int L^q(x,y)d\pi_{n,x}(y) \le L^q(x,x)$ for $\mu_0$-almost every $x \in M$. Since $L$ is bounded, we can use Fatou's lemma and conclude

$$\begin{aligned} |D^-\mathcal{U}_p|(\mu_0) &= \lim_{n\to 0} \frac{\mathcal{U}_p(\mu_0) - \mathcal{U}_p(\mu_n)}{w_p(\mu_0,\mu_n)} \\ &\le \int \limsup_{n\to\infty} \left(\int L^q(x,y)d\pi_{n,x}(y)\right)^{1/q} d\mu_0(x) \\ &\le \left(\int L^q(x,x)d\mu_0(x)\right) \\ &= \left(\int \frac{|D\rho|^q}{\rho^{(p-1)q}}\rho d\mu\right)^{1/q} = \left(\int \frac{|D\rho|^q}{\rho^{p-1}}\rho d\mu\right)^{1/q}. \end{aligned}$$

□

**Proposition 6.13.** *If $|D^-\mathcal{U}_p|$ is sequentially lower semicontinuous w.r.t. $\mathcal{P}_p(M)$ then*

$$|D^-\mathcal{U}_p|^q(\mu_0) = r^{-q}\int |\nabla\rho^r|_*^q d\mu \qquad \forall \mu_0 = \rho\mu \in D(\mathcal{U}_p).$$

*Remark.* In [AGS13, Theorem 7.6] Ambrosio-Gigli-Savaré proved also that the converse holds for the entropy functional. We are not able to prove the converse in case $2r > 1$, i.e. $p > 2$.

*Proof.* By the above results we only need to show that $|D^-\mathcal{U}_p|(\mu_0) \le r^{-q}\int |\nabla\rho^r|_*^q d\mu$. First assume $\rho$ is bounded and find a sequence of measures $\mu_n \in \mathcal{P}_p(M)$ with Lipschitz densities $\rho_n$ bounded from below by $\frac{1}{n}$ converging in $L^{2r}(M)$ to $\rho$ (by compactness $\rho_n^r \to \rho^r$ in $L^2$) such that

$$\lim_{n\to\infty} \frac{1}{q} \int |\nabla \rho_n^r|_*^q d\mu = \mathrm{Ch}_q(\rho^r).$$

Since $|\nabla \rho_n^r|_w = |D\rho_n^r|$ almost everywhere, we see that

$$\begin{aligned} |D^-\mathcal{U}_p|(\mu_0) &\leq \liminf_{n\to\infty} |D^-\mathcal{U}_p|(\mu_n) \\ &\leq \liminf r^{-q} \int |\nabla \rho_n^r|_*^q d\mu = r^{-q} \int |\nabla \rho^r|_*^q d\mu. \end{aligned}$$

In case $\rho$ is unbounded we can truncate $\rho$ without increasing the $q$-Cheeger energy use the lower semicontinuity again to conclude the result. □

**Corollary 6.14.** *Assume one of the following holds:*

- $p \in (1,2)$ *and the strong* $CD_p(K,\infty)$ *condition holds for some* $K \geq 0$
- $p \in (2, \frac{3+\sqrt{5}}{2})$*, the* $CD_p(0,N)$ *condition holds such that* $p = \frac{2N+1}{N}$ *and* $M$ *is* $n$*-Ahlfors regular for some* $n < N$*.*

*Then* $|D^-\mathcal{U}_p|$ *is lower semicontinuous and an upper gradient of* $\mathcal{U}_p$*.*

*Proof.* In case $p \in (1,2)$ note that $3-p \in (1,2)$ and thus $U_p \in \mathcal{DC}_\infty$. In case $p \in (2, \frac{3+\sqrt{5}}{2})$ we have $3-p \in (0,1)$ and thus $U_p \in \mathcal{DC}_N$ for $3-p = 1-\frac{1}{N}$. In both cases displacement convexity, i.e. $K$-convexity with $K=0$, follows. Which implies that $|D^-\mathcal{U}_p|$ is lower semicontinuous and an upper gradient of $\mathcal{U}_p$. □

The conclusion holds equally if $\mathcal{U}_p$ is just $K$-convex. Since $K$-convexity neither follows from the strong $CD_p(K,\infty)$-condition in case $p \in (1,2)$ nor from $CD_p(K,N)$, we use those conditions to imply convexity. Nevertheless, we hope that it is possible to show that $|D^-\mathcal{U}_p|$ is lower semicontinuous and an upper gradient of $\mathcal{U}_p$ if one of the curvature condition holds.

**Theorem 6.15** (Uniqueness of the gradient flow of $\mathcal{U}_p$)**.** *Let* $r > 0$ *and assume that* $|D^-\mathcal{U}_p|^q$ *is lower semicontinuous and convex w.r.t. linear interpolation. Then for every* $\mu_0 \in \mathcal{P}_p(M)$ *there exists at most one gradient flow of* $\mathcal{U}_p$ *starting from* $\mu_0$*.*

*Remark.* By Lemma 2.5 and [AGS13, Theorem 7.8] convexity of $|D^-\mathcal{U}_p|^q$ holds if $p \leq 2 \leq q$.

*Proof.* Assume that $(\mu_t^1)$ and $(\mu_t^2)$ are two distinct gradient flows starting from $\mu_0$. Then we have for $i=1,2$ and all $T \geq 0$

$$\begin{aligned} \mathcal{U}_p(\mu_0) &= \mathcal{U}_p(\mu_T^i) + \frac{1}{p}\int_0^T |\dot{\mu}_t^i|^q dt \\ &\quad + \frac{1}{q}\int_0^T |D^-\mathcal{U}_p|^q(\mu_t^i) dt. \end{aligned}$$

Note that the curve $t \mapsto \mu_t = (\mu_t^1 + \mu_t^2)/2$ is absolutely continuous in $\mathcal{P}_p(M)$ and

$$|\dot{\mu}_t|^p \leq \frac{|\dot{\mu}_t^1|^p + |\dot{\mu}_t^2|^p}{2}.$$

Using the strict convexity of $\mathcal{U}_p$ and the convexity of $|D^-\mathcal{U}_p|^q$ we conclude

$$\begin{aligned}\mathcal{U}_p(\mu_0) &> \mathcal{U}_p(\mu_T) + \frac{1}{p}\int_0^T |\dot{\mu}_t|^q dt \\ &\quad + \frac{1}{q}\int_0^T |D^-\mathcal{U}_p|^q(\mu_t)dt \\ &\geq \mathcal{U}_p(\mu_T) + \int_0^T |\dot{\mu}_t||D^-\mathcal{U}_p|(\mu_t)dt\end{aligned}$$

But this is a contradiction to

$$\mathcal{U}_p(\mu_t) \geq \mathcal{U}_p(\mu_s) - \int_s^t |\dot{\mu}_t||D^-\mathcal{U}|(\mu_t)dt$$

for $s, t \in [0, \infty)$ (note $|D^-\mathcal{U}_p|$ is an upper gradient). □

Finally we can identify the two flows. The theorem and its proof is similar to [AGS13, Theorem 8.5].

**Theorem 6.16** (Identification of the gradient flows)**.** *Let $r > 0$ and assume that $\mathcal{U}_p$ is $K$-convex in $\mathcal{P}_p(M)$. Then for all $f_0 \in L^2(M, \mu)$ such that $\mu_0 = f_0\mu \in \mathcal{P}_p(M)$ the following is equivalent:*

1. *If $f_t$ is the gradient flow of $\mathrm{Ch}_q$ in $L^2(M, \mu)$ starting from $f_0$, then $\mu_t = f_t\mu$ is the gradient flow of $\mathcal{U}_p$ in $\mathcal{P}_p(M)$ starting from $\mu_0$, the map $t \mapsto \mathcal{U}_p(\mu_t)$ is absolutely continuous in $(0, \infty)$ and*

$$-\frac{d}{dt}\mathcal{U}_p(\mu_t) = |\dot{\mu}_t|^p = |D^-\mathcal{U}_p|^q \qquad \text{for a.e.} t \in (0, \infty).$$

2. *Conversely, if we assume in addition that $|D^-\mathcal{U}_p|^q$ is convex w.r.t. linear interpolation, then whenever $\mu_t$ is the gradient flow of $\mathcal{U}_p$ in $\mathcal{P}_p(M)$ starting from $\mu_0$, then $\mu_t$ is absolutely continuous and its density $f_t$ w.r.t. $\mu$ is the gradient flow of $\mathrm{Ch}_q$ in $L^2(M, \mu)$ starting from $f_0$. The same holds if the gradient flow of $\mathcal{U}_p$ in $\mathcal{P}_p(M)$ starting at $\mu_0$ is unique.*

*Proof.* By $K$-convexity of $\mathcal{U}_p$ we know that $|D^-\mathcal{U}_p|$ is an upper gradient and

$$|D^-\mathcal{U}_p|^q(\rho\mu) = \mathsf{F}_q(\rho)$$

thus by the Kuwada lemma we know that if $f_t$ is the gradient flow of the $q$-Cheeger energy then

$$|\dot{\mu}_t|^p \leq \int \frac{|\nabla f_t|^q}{f_t^{p-1}} d\mu = \mathsf{F}_q(f_t)$$

and

$$t \mapsto \mathcal{U}_p(\mu_t)$$

is absolutely continuous with

$$\frac{d}{dt}\mathcal{U}_p(\mu_t) = -\int \frac{|\nabla f_t|^q}{f_t^{p-1}} d\mu.$$

Hence

$$\int \frac{|\nabla f_t|^q}{f_t^{p-1}} d\mu \geq \frac{1}{p}|\dot{\mu}_t|^p + \frac{1}{q}|D^- \mathcal{U}_p|^p$$

so that $\mu_t$ satisfies the $\mathcal{U}_p$-dissipation inequality, i.e.

$$\begin{aligned}\mathcal{U}_p(\mu_0) - \mathcal{U}_p(\mu_t) &= \int_0^t \int \frac{|\nabla f_s|^q}{f_s^{p-1}} d\mu ds \\ &\geq \frac{1}{p}\int_0^t |\dot{\mu}_t|^p ds + \frac{1}{q}\int_0^t |D^- \mathcal{U}_p|^q ds\end{aligned}$$

and $\mu_t$ is the gradient flow of $\mathcal{U}_p$ in $\mathcal{P}_p(M)$ starting at $\mu_0$. Absolute continuity of $t \mapsto \mathcal{U}_p(\mu_t)$ in $(0,\infty)$ implies

$$\begin{aligned}\frac{d}{dt}\mathcal{U}_p(\mu_t) &= -|\mu_t||D^- \mathcal{U}_p| \\ &= -|\mu_t|^p \\ &= -|D^- \mathcal{U}_p|^q.\end{aligned}$$

For the second part, assume that $t \mapsto \tilde{f}_t$ is the gradient flow of the $q$-Cheeger energy starting at $f_0$. By the previous part we know that $\tilde{\mu}_t = \tilde{f}_t \mu$ is also a gradient flow of $\mathcal{U}_p$. Uniqueness (Theorem 6.15 above) implies that $\mu_t = \tilde{\mu}_t$ for all $t \geq 0$. □

# 7 Orlicz-Wasserstein spaces

In this chapter we show that the interpolation inequality can be proven also for Orlicz-Wasserstein spaces using similar arguments. Before that we will define and investigate Orlicz-Wasserstein spaces. The main difference between a general convex and increasing function $L$ and a homogeneous function is that there is no well-defined dual problem. However, one can use $c_L$-concave function and the geodesic structure to determine the interpolation potentials.

## 7.1 General Results

Let $L : [0,\infty) \to [0,\infty)$ be a strictly convex increasing function with $L(0) = 0$. Assume further there is an increasing function $l : (0,\infty) \to (0,\infty)$ with $\lim_{r\to 0} l(r) = 0$ and

$$L(r) = \int_0^r l(s)ds$$

and hence $L'(s) = l(s)$.

Define $L_\lambda(r) = L(r/\lambda)$ and note

$$\begin{aligned} L_\lambda(r) &= \int_0^r l_\lambda(s)ds \\ &= \int_0^{r/\lambda} l(s)ds \end{aligned}$$

and thus

$$l_\lambda(s) = \frac{1}{\lambda} l\left(\frac{s}{\lambda}\right)$$

and

$$l_\lambda^{-1}(t) = \lambda l^{-1}(\lambda t).$$

We denote by $c_L$ the cost function given by $c_L(x,y) = L(d(x,y))$ and as an abbreviation $c_\lambda = c_{L_\lambda}$.

The $c_L$-transform of a function $\phi : X \to \underline{\mathbb{R}}$ relative to $(X,Y)$ is defined as

$$\phi^{c_L}(y) = \inf_x c_L(x,y) - \phi(x)$$

and similarly the $\bar{c}^L$-transform.

**Definition 7.1** (Orlicz-Wasserstein space)**.** Let $\mu_i$ be two probability measures on $M$ and define

$$w_L(\mu_0,\mu_1) = \inf\left\{\lambda > 0 \,|\, \inf_{\pi\in\Pi(\mu_0,\mu_1)} \int L_\lambda\left(d(x,y)\right) d\pi(x,y) \le 1\right\}.$$

With convention $\inf\varnothing = \infty$.

According to Sturm [Stu11, Proposition 3.2], $w_L$ is a complete metric on

$$\mathcal{P}_L(M) := \{\mu_1 \in \mathcal{P}(M) \,|\, w_L(\mu_1,\delta_{x_0}) < \infty\}$$

where $x_0$ is some fixed point.

Even though the following lemma is not needed, it makes many proofs below easier.

**Lemma 7.2** ([Stu11, Proposition 3.1])**.** *For every $\mu_i \in \mathcal{P}_L(M)$ there is an optimal coupling $\pi_{opt}$ of $(\mu_0,\mu_1)$ such that*

$$\lambda_{min} = w_L(\mu_0,\mu_1) \Rightarrow \int L_{\lambda_{min}}(d(x,y))d\pi_{opt}(x,y) = 1.$$

Actually the Lemma shows that the whole theory of Kantorovich potentials will depend on the distance. Furthermore, the $c_L$-convace functions are not necessarily star-shaped. Nevertheless, we will show that $\mathcal{P}_L(M)$ is a geodesic space iff $M$ is and that a similar property to the star-shapedness holds.

**Proposition 7.3.** *Let $\Phi$ be a convex increasing function with $\Phi(1) = 1$, then*

$$w_L \le w_{\Phi\circ L}.$$

*Remark.* This just uses Sturm's idea to show the same inequality for the Luxemburg norm of Orlicz spaces. Compare this also to [Vil09, Remark 6.6], but note that Villani defines $w_p$ without the factor $\frac{1}{p}$.

*Proof.* This follows easily from Jensen's inequality. Let $\mu_0,\mu_1$ be two measures and $\lambda > 0$ and $\pi$ be a coupling such that $\int(\Phi\circ L)_\lambda(d(x,y))d\pi(x,y) \le 1$ then since $(\Phi\circ L)_\lambda = \Phi\circ L$

$$\Phi(\int L_\lambda(d(x,y))d\pi(x,y)) \le \int \Phi\circ L_\lambda(d(x,y))d\pi(x,y) \le 1$$

Since $\Phi(1) \le 1$ and $\Phi$ is increasing, we see that $\int L_\lambda(d(x,y))d\pi(x,y) \le 1$ which implies $w_L(\mu_0,\mu_1) \le w_{\Phi\circ L}(\mu_0,\mu_1)$. ☐

**Proposition 7.4.** *Assume for all $\lambda > 0$*

$$\sup_{R\to\infty} \frac{L(\lambda R)}{L(R)} < \infty.$$

*If $\mu_n,\mu_\infty \in \mathcal{P}_L(M)$ and $\mu_n$ converges weakly to $\mu_\infty$, then*

$$w_L(\mu_n,\mu_\infty) \to 0 \iff \lim_{R\to\infty}\limsup_{n\to\infty}\int_{M\backslash B_R(x_0)} L_\lambda(d(x,x_0))d\mu_n = 0$$

*for all $0 < \lambda < \lambda_0$.*

*Remark.* This generalizes [Vil03, Theorem 7.12]. The other equivalences in Villani's theorem can be proven similarly. We, however, only need the one stated above.

*Proof.* Fix some $x_0 \in M$. It is not difficult to see that for any $\lambda > 0$ and any $\mu' \in \mathcal{P}_L(M)$

$$\lim_{R\to\infty} \int_{M\backslash B_R(x_0)} L_\lambda(d(x,x_0))d\mu'(x) = 0 \iff \lim_{R\to\infty} \int_{M\backslash B_R(x_0)} L(d(x,x_0))d\mu'(x) = 0.$$

First assume $w_L(\mu_n, \mu_\infty)$ and let $\pi_n$ be the optimal plans with $l_n = w_L(\mu_n, \mu_\infty)$ and

$$\int L_{l_n}(d(x,y))d\pi_n(x,y) = 1.$$

For $n$ large, for any $\lambda > 0$ choose a sequence $r_n \le \frac{1}{2}$ such that $l_n = r_n\lambda$. Then using the triangle inequality and convexity of $L$ we get

$$\begin{aligned}
\int L_\lambda\left(d(x,x_0)\right) d\mu_n(x) &= \int L_\lambda\left(d(x,x_0)\right) d\pi_n(x,y) \\
&\le r_n \int L_{r_n\lambda}\left(d(x,y)\right) d\pi_n(x,y) + (1-r_n)\int L_{(1-r_n)\lambda}\left(d(y,x_0)\right) d\pi_n(x,y) \\
&\le r_n + (1-r_n)\int L_{\frac{1}{2}\lambda}\left(d(y,x_0)\right) d\mu_\infty(y).
\end{aligned}$$

since $L_{(1-r_n)\lambda} \le L_{\frac{1}{2}\lambda}$. Therefore,

$$\lim_{R\to\infty} \limsup_{n\to\infty} \int_{M\backslash B_R(x_0)} L_\lambda(d(x,x_0))d\mu_n(x) \le \lim_{R\to\infty} \int_{M\backslash B_R(x_0)} L_{\frac{1}{2}\lambda}(d(x,x_0))d\mu_\infty(x) = 0.$$

Now assume that

$$\lim_{R\to\infty} \limsup_{n\to\infty} \int_{M\backslash B_R(x_0)} L_\lambda(d(x,x_0))d\mu_n(x) = 0$$

for any $0 < \lambda < \lambda_0$ and $\mu_n$ converges weakly to $\mu_\infty$. This bound ensures that $\mu_\infty$ is in $\mathcal{P}_L(M)$.

Take any $\lambda > 0$ and an optimal coupling $\pi_n$ of $(\mu_n, \mu_\infty)$ w.r.t. $L_\lambda$. For $R > 0$ and $A \wedge B = \min\{A, B\}$ we have

$$d(x,y) \le d(x,y) \wedge R + 2d(x,x_0)\chi_{B_{R/2}(x_0)}(x) + 2d(x_0,y)\chi_{B_{R/2}(x_0)}(y)$$

and thus by convexity of $L$ and $L(0) = 0$

$$L_\lambda(d(x,y)) \le \frac{1}{3}L_{\frac{\lambda}{3}}(d(x,y)\wedge R) + \frac{1}{3}L_{\frac{\lambda}{6}}(d(x,x_0)\chi_{B_{R/2}(x_0)}(x)) + \frac{1}{3}L_{\frac{\lambda}{6}}(d(x_0,y)\chi_{B_{R/2}(x_0)}(y)).$$

Thus integrating over $\pi_n$ we get

$$\begin{aligned}
3\int L_\lambda(d(x,y))d\pi_n(x,y) &\le \int L_{\frac{\lambda}{3}}(d(x,y)\wedge R)d\pi_n(x,y) \\
&\quad + \int_{M\backslash B_{R/2}(x_0)} L_{\frac{\lambda}{6}}(d(x,x_0))d\mu_n(x) \\
&\quad + \int_{M\backslash B_{R/2}(x_0)} L_{\frac{\lambda}{6}}(d(x_0,y))d\mu_\infty(y).
\end{aligned}$$

we first take the $\limsup$ with $n \to \infty$ and then $R \to \infty$ and conclude that the last two terms converges to zero by our assumption and since $L_{\frac{\lambda}{3}}(d(x,y) \wedge R)$ is a bounded continuous function and $\pi_n$ converges weakly to the trivial coupling $(\mathrm{Id} \times \mathrm{id})_* \mu_\infty$, the first term converges to zero as well. In particular, for $n \geq N(\lambda)$ we have

$$\int L_\lambda(d(x,y)) d\pi_n(x,y) \leq 1.$$

and thus

$$w_L(\mu_n, \mu_\infty) \leq \lambda.$$

Since $\lambda$ was arbitrary we conclude $w_L(\mu_n, \mu_\infty) \to 0$. □

**Proposition 7.5.** *Assume $M$ is a proper metric space and $\Phi$ is convex, increasing, $\Phi(1) = 1$ and $L(r) \to \infty$ and $r/\Phi(r) \to 0$ as $r \to \infty$. In addition, assume for all $\lambda > 0$*

$$\sup_{R \to \infty} \frac{L(\lambda R)}{L(R)} < \infty.$$

*Suppose $A$ is closed subset of $\mathcal{P}_L(M)$ such that $w_{\tilde{L}}$ is bounded where $\tilde{L} = \Phi \circ L$. Then $A$ is precompact in $\mathcal{P}_L(M)$.*

*Remark.* Compare this to [Kel11, Theorem 6] for the case $L(t) = t^p$, $\Phi(t) = t^r$ for $p \geq 1$ and $r > 1$.

*Proof.* It suffices to show that each $w_{\tilde{L}}$-ball is compact in $\mathcal{P}_L(M)$

So for some $r > 0$ and $\mu_0 \in \mathcal{P}_{\tilde{L}}(M) \subset \mathcal{P}_L(M)$ let

$$\tilde{B} := \tilde{B}_r(\mu_0) = \{\mu_1 \in \mathcal{P}_L(M) \,|\, w_{\tilde{L}}(\mu_0, \mu_1) \leq r\}.$$

and let $(\mu_n)_{n \in \mathbb{N}}$ be a sequence in $\tilde{B}$. Then there are (optimal) couplings $\pi_n$ such that

$$\int \tilde{L}_r(d(x,y)) d\pi_n(x,y) \leq 1$$

(for $w_{\tilde{L}}(\mu_n, \mu_0) < r$ just take the definition. Using the proposition above,we see

$$\int L_r(d(x,y)) d\pi_n(x,y) \leq 1.$$

Because of the stability of optimal couplings and lower semicontinuity of the cost [Vil09, Theorem 5.20, Lemma 4.3], we only need to show that $(\mu_n)_{n \in \mathbb{N}}$ is weakly precompact and

$$\lim_{R \to \infty} \limsup_{n \to \infty} \int_{M \backslash B_R(x_0)} L_\lambda(d(x, x_0)) d\mu_n = 0$$

i.e. it is precompact in $\mathcal{P}_L(M)$ by the lemma above.

Since $\tilde{B}$ is bounded w.r.t. $w_{\tilde{L}}$ we can assume that for some $R > 0$

$$w_{\tilde{L}}(\mu_n, \delta_{x_0}) \leq \lambda_0.$$

Now set $\lambda_0 = 0$. For $c\lambda = \lambda_0$ and $c \in (0,1)$ we have

$$\begin{aligned}\int_{M\backslash B_R(x_0)} L_\lambda(d(x,x_0))d\mu(x) &\le \frac{L_\lambda(R)}{\Phi(L_{\lambda_0}(R))}\int_{M\backslash B_R(x_0)} \tilde{L}_{\lambda_0}(d(x,x_0))d\mu_n(x)\\ &\le \frac{L_{\lambda_0}(R)}{\Phi(L_{\lambda_0}(R))}\frac{L_{\lambda_0}(c^{-1}R)}{L_{\lambda_0}(R)} \le C\frac{L_{\lambda_0}(R)}{\Phi(L_{\lambda_0}(R))}\end{aligned}$$

for some $C > 0$ depending only on $\lambda_0, c$ and $L$. Hence by the fact that $L(R), \Phi(R) \to \infty$ as $R \to \infty$ we conclude

$$\lim_{R\to\infty} \limsup_{n\to\infty} \int_{M\backslash B_R(x_0)} L_\lambda(d(x,x_0))d\mu_n = 0.$$

In order to show weak precompactness notice that $L(R) \ge 1$ for $R \ge r_0 = r_0(L)$ implies tightness, which is equivalent to precompactness by the classical Prokhorov theorem. Indeed, $B_R(x_0)$ is compact and for $r_0 \le R \to \infty$

$$\int_{M\backslash B_R(x_0)} d\mu_n \le C\frac{L_{\lambda_0}(R)}{\Phi(L_{\lambda_0}(R))} \to 0$$

uniformly in $n$. □

**Proposition 7.6.** *Assume $M$ is a geodesic space. Let $\pi_{opt}$ be the optimal coupling of $(\mu_0, \mu_1)$ then there is a $\Pi$ supported on the geodesics such that for $i = 0, 1$*

$$(e_i)_*\Pi = \mu_i.$$

*Furthermore, let $\mu_t = (e_t)_*\Pi$ then*

$$w_L(\mu_s, \mu_t) = |s-t| w_L(\mu_0, \mu_1).$$

*In particular, $\mathcal{P}_L(M)$ is a geodesic space.*

*Proof.* The first part follows from using the measurable selection theorem for

$$(x,y) \mapsto \{\gamma : [0,1] \to M \mid \gamma \text{ is a geodesic from } x \text{ to } y\}$$

similar to [Lis06] in case of $p$-Wasserstein spaces.

For the second part note for $\lambda_{min} = w_L(\mu_0, \mu_1)$

$$\int L\left(\frac{d(\gamma_s,\gamma_t)}{|s-t|\lambda_{min}}\right) d\Pi(\gamma) = \int L_{\lambda_{min}}\left(d(\gamma_0,\gamma_1)\right) d\Pi(\gamma) = 1.$$

Hence

$$w_L(\mu_t, \mu_s) \le |s-t|\lambda_{min}.$$

So $t \mapsto \mu_t$ is absolutely continuous in $\mathcal{P}_L(M)$ and $|\dot{\mu}_t| \le \lambda_{min}$. But we also have

$$\lambda_{min} = w_L(\mu_0, \mu_1) = \int_0^1 |\dot{\mu}_t| dt.$$

Therefore, $|\mu_t| = \lambda_{min}$ and

$$w_L(\mu_s, \mu_t) = \left|\int_s^t |\dot{\mu}_r| dr\right| = |s-t| w_L(\mu_s, \mu_t).$$

$\square$

It is also possible to define a dual problem by

$$\sup\{\lambda > 0 \,|\, \sup_{\phi \in L^1(\mu_0)} \left\{\int \phi d\mu_0 + \int \phi^{c_\lambda} d\mu_1\right\} \geq 1\}.$$

However, we will not go into this dual problem and directly deal with the $c_\lambda$-transform whenever Kantorovich potentials are needed. Main "problem": the restriction property does not hold for $w_L$ and many results depend on (the number) $w_L(\mu_0, \mu_1)$.

The following inequality will help to show that $c_L$-conave functionals enjoy a similar property to star-shapedness. It will also show that the Jacobians of the interpolation measures are positive semidefinite.

**Lemma 7.7.** *If $x, y \in M$ and $z \in Z_t(x, y)$ for some $t \in [0, 1]$. Then for all $m \in M$*

$$t^{-1} L(d(m, y)) \leq L_t(d(m, z)) + t^{-1}(1-t) L(d(x, y)).$$

*Furthermore, choosing $x = m$, this becomes an equality.*

*Remark.* This extends Lemma 3.7.

*Proof.* Since $L$ is convex and increasing

$$\begin{aligned} L(d(m, y)) &\leq L(t \cdot t^{-1} d(m, z) + (1-t) d(x, y)) \\ &\leq t L_t(d(m, z)) + (1-t) L(d(x, y)). \end{aligned}$$

Dividing by $t$ we get the inequality and choosing $x = m$ we see that all inequalities are actually equalities. $\square$

**Lemma 7.8.** *Let $\eta : [0, 1] \to M$ be a geodesic between two distinct points $x$ and $y$. For $t \in (0, 1]$ define*

$$f_t(m) := -c_t(m, \eta_t).$$

*Then for some fixed $t \in [0, 1]$ the function $h(m) := f_t(m) - t^{-1} f_1(m)$ has a minimum at $x$.*

*Proof.* Using Proposition 3.7 above for $t \in (0, 1)$ we have for $z = \eta_t \in Z_t(x, y)$

$$\begin{aligned} -h(m) = t^{-1} L(d(m, y)) - L_t(d(m, z)) &\leq t^{-1}(1-t) L(d(x, y)) \\ &= t^{-1} L(d(x, y)) - L_t(d(x, \eta_t)) = -h(x). \end{aligned}$$

$\square$

**Lemma 7.9.** *Let $X$ and $Y$ be compact subsets of $M$ and let $t \in (0,1]$. If $\phi \in \mathcal{I}^{c_L}(X,Y)$ then $t^{-1}\phi \in \mathcal{I}^{c_t}(X, Z_t(X,Y))$.*

*Proof.* For $t=1$ there is nothing to prove. For the rest we follow the strategy of [CEMS01, Lemma 5.1]. Set $L_y(x) = L(d(x,y))$ and let $t \in (0,1]$ and $y \in Y$ and define $\phi(x) := c_L(x,y) = L_y(x)$. We claim that the following representation holds

$$t^{-1}L_y(m) = \inf_{z \in Z_t(X,y)} \left\{ (L_t)_z(m) + \inf_{\{x \in X \,|\, z \in Z_t(x,y)\}} t^{-1}(1-t)L_y(x) \right\}.$$

Indeed, by Lemma 3.7 the left hand side is less than or equal to the right hand side for any $z \in Z_t(X,y)$. Furthermore, choosing $x = m$ we get an equality and thus showing the representation.

Now note that the claim implies that $t^{-1}\phi$ is the $\bar{c}_p$-transform of the function

$$\psi(z) = - \inf_{\{x \in X \,|\, z \in Z_t(x,y)\}} t^{-1}(1-t)L_y(x)$$

and therefore $t^{-1}\phi$ is $c_t$-concave relative to $(X, Z_t(X,y))$. Since $\mathcal{I}^{c_t}(X, Z_t(X,y)) \subset \mathcal{I}^{c_t}(X, Z_t(X,Y))$ we see that each $t^{-1}L_y$ is in $\mathcal{I}^{c_t}(X, Z_t(X,Y))$.

It remains to show that for an arbitrary $c_L$-concave function $\phi$ and $t \in (0,1]$ the function $t^{-1}\phi$ is $c_t$-concave relative to $(X, Z_t(X,Y))$. Since $\phi = \phi^{c_L \bar{c}_L}$ we have

$$t^{-1}\phi(x) = \inf_y t^{-1}L(d(x,y)) - t^{-1}\phi^{c_L}(y).$$

But each function

$$\psi_y(x) = t^{-1}L_y(x) - t^{-1}\phi^c(y)$$

is $c_p$-concave relative to $(X, Z_t(X,Y))$ and $\phi$ is proper, thus also the infimum is $c_t$-concave relative to $(X, Z_t(X,Y))$, i.e. $t^{-1}\phi \in \mathcal{I}^{c_t}(X, Z_t(X,Y))$. □

## 7.2 Orlicz-Wassterstein spaces on Finsler manifolds

### 7.2.1 Technical ingredients

For simplicity, assume throughout the section that $L$ is smooth away from 0.

For $L_x = L(d(x,\cdot))$ and $x \neq y$

$$\nabla L_x(y) = l(d(x,y))\nabla d_x(y).$$

Define

$$\nabla^L \phi := \frac{l^{-1}(|\nabla \phi|)}{|\nabla \phi|}\nabla \phi.$$

Note that for $v \in T_xM$ with $|v| = 1$ and $r \geq 0$

$$\nabla \phi(x) = l(r)v$$

iff

$$\nabla^L \phi = rv.$$

We also use the abbreviation

$$\nabla^\lambda \phi = \nabla^{L_\lambda} \phi.$$

It is easy to see that under our assumptions that $\phi \mapsto \nabla^L \phi$ is continuous and (as) smooth (as $L$) wherever $\nabla \phi(x) \neq 0$.

Similar to the $c_p$-case we will use the abbreviation $\mathcal{K}_x^L d\phi_x$ (resp. $\mathcal{K}_x^\lambda d\phi_x$) for $\nabla^L \phi(x)$ (resp. $\nabla^\lambda \phi(x)$). As mentioned above, this can also be seen as a Legendre transform from $T^*M$ to $TM$.

**Lemma 7.10** (Cut locus charaterization)**.** *If $y \neq x$ is a cut point of $x$, then $f(z) := L(d(x,y))$ satisfies*

$$\liminf_{v \to 0 \in T_x M} \frac{f(\xi_v(1)) + f(\xi_v(-1)) - 2f(x)}{F(v)^2} = -\infty$$

*where $\xi_v : [-1,1] \to M$ is the geodesic with $\dot{\xi}_v(0) = v$.*

*Proof.* The proof follows in the same fashion as Lemma 4.1. We will show the necessary adjustments.

As above, let's first assume there are two distinct unit speed geodesics $\eta, \zeta : [0, d(x,y)] \to M$ from $x$ to $y$ and let $v = \dot{\zeta}(0)$ and $w = \dot{\eta}(0)$. For fixed small $\epsilon > 0$ set $y_\epsilon = \eta(d(x,y) - \epsilon)$ then $y_\epsilon \notin \mathrm{Cut}(x) \cup \{x\}$ and using the first variation formula we get for $t > 0$

$$\begin{aligned}
f(\xi_v(-t)) - f(x) &\leq L(d(\xi_v(-t), y_\epsilon) + \epsilon) - L(d(x, y_\epsilon) + \epsilon) \\
&= tl(d(x, y_\epsilon) + \epsilon) g_{\dot{\eta}(0)}(v, \dot{\eta}(0)) + \mathcal{O}(t^2) \\
&= tl(d(x,y)) g_{\dot{\eta}(0)}(v, \dot{\eta}(0)) + \mathcal{O}(t^2).
\end{aligned}$$

The term $\mathcal{O}(t^2)$ is ensured by smoothness of $\xi_v$ and by the facts that $x \neq y_\epsilon$ and that $L(d(\cdot,\cdot))$ is bounded in a neighborhood $(x,y)$. We also get by Taylor formula

$$f(\xi_v(t)) - f(x) = L(d(x,y) - t) - L(d(x,y)) = -tl(d(x,y)) + \mathcal{O}(t^2).$$

Combining these two facts with $g_w(v,w) < 1$ ($\eta$ and $\xi$ are distinct), we get

$$\frac{f(\xi_v(-t)) + f(\xi_v(t)) - 2f(x)}{t^2} \leq \frac{1 - g_w(v,w)}{t} l(d(x,y)) + t^{-2}\mathcal{O}(t^2) \to -\infty \text{ as } t \to 0.$$

For the conjugate point case, we use the same construction and notation as in the proof of Lemma 4.1. Note that

$$\begin{aligned}
\lim_{s\to 0}\frac{L(\mathcal{L}(\sigma_s))+L(\mathcal{L}(\sigma_{-s}))-2L(\mathcal{L}(\sigma_0))}{s^2} &= \Bigg(l(\mathcal{L}(\sigma_0))\frac{\partial^2}{\partial s^2}\mathcal{L}(\sigma_s)\Big|_{s=0} \\
&\qquad +l'(\mathcal{L}(\sigma_0))\left(\frac{\partial L(\sigma_s)}{\partial s}\Big|_{s=0}\right)^2\Bigg) \\
&\le l(d(x,y))\Bigg(-2\epsilon g_{\dot\eta}(v,v)/d(x,y) \\
&\qquad +\epsilon^2\left\{\mathcal{T}_{\dot\eta(0)}(v)/d(x,y)+I(V,V))^2\right\}\Bigg) \\
&\quad +l'(d(x,y))F(v).
\end{aligned}$$

Using the fact that $f(\xi_v(\epsilon s))\le L(\mathcal{L}(\sigma_s))$ we obtain

$$\begin{aligned}
\liminf_{s\to 0}\frac{f(\xi_v(\epsilon s))+f(\xi_v(-\epsilon s))-2f(x)}{\epsilon^2 s^2} &\le \liminf_{s\to 0}\frac{L(\mathcal{L}(\sigma_s))+L(\mathcal{L}(\sigma_{-s}))-2L(\mathcal{L}(\sigma_0))}{\epsilon^2 s^2} \\
&\le l(d(x,y))\Bigg(-2\epsilon^{-1}g_{\dot\eta}(v,v)/d(x,y) \\
&\qquad +\mathcal{T}(v)/d(x,y)+d(x,y)I(V,V)\Bigg) \\
&\quad +l'(d(x,y))F(v)^2.
\end{aligned}$$

Letting $\epsilon$ tend to zero completes the proof. □

### 7.2.2 The Brenier-McCann-Ohta solution

**Lemma 7.11.** *Let $\phi : M \to \mathbb{R}$ be a $c_L$-concave function. If $\phi$ is differentiable at $x$ then $\partial^{c_L}\phi(x) = \{exp_x(\nabla^L(-\phi)(x))\}$. Moreover, the curve $\eta(t) := exp_x(t\nabla^L(-\phi)(x))$ is a unique minimal geodesic from $x$ to $exp_x(\nabla^L(-\phi)(x))$.*

*Remark.* See also [McC01, Theorem 13] for the Riemannian case.

*Proof.* Let $y \in \partial^{c_L}\phi(x)$ be arbitrary and define $f(z) := c_L(z,y) = L(d(z,y))$. By definition of $\partial^{c_L}\phi(x)$ we have for any $v \in T_xM$

$$f(exp_x v) \ge \phi^{c_L}(y)+\phi(exp_x v) = f(x)-\phi(x)+\phi(exp_x v) = f(x)+d\phi_x(v)+o(F(v)).$$

Now let $\eta : [0,d(x,y)] \to M$ be a minimal unit speed geodesic from $x$ to $y$. Given $\epsilon > 0$, set $y_\epsilon = \eta(d(x,y)-\epsilon)$ and note that $\eta|_{[0,d(x,y)-\epsilon]}$ does not cross the cut locus of $x$. By the first variation formula we have

$$\begin{aligned}
f(exp_x v)-f(x) &\le L\left(d(exp_x v,y_\epsilon)+\epsilon\right)-L\left(d(x,y_\epsilon)+\epsilon\right) \\
&= -l\left(d(x,y_\epsilon)+\epsilon\right)g_{\dot\eta(0)}(v,\dot\eta(0))+o(F(v)). \\
&= -l(d(x,y))\mathcal{L}_x^{-1}(\dot\eta(0))(v)+o(F(v)).
\end{aligned}$$

Therefore, $d\phi_x(v) \le -l(d(x,y))\mathcal{L}_x^{-1}(\dot{\eta}(0))(v)$ for all $v \in T_xM$ and thus $\nabla(-\phi) = l(d(x,y))\cdot\dot{\eta}(0).$, i.e. $\nabla^L(-\phi) = d(x,y)\cdot\dot{\eta}(0)$. In addition, note that $\eta(t) = exp_x(t\nabla^L(-\phi)(x))$, which is uniquely defined. □

**Lemma 7.12.** *Let $t \mapsto \mu_t$ be a geodesic between $\mu_0$ and $\mu_1$, i.e. $w_L(\mu_0, \mu_t) = t\lambda$. If $\mu_0$ is absolutely continuous and the unique $\phi_t$ the Kantorovich potential of $(\mu_0, \mu_t)$ w.r.t. $L_{t\lambda}$ such that $\phi_t(x_0) = 0$. Then $\phi_t = t^{-1}\phi$.*

*Proof.* For $x \neq y \in \partial^{c_\lambda}\phi_1(x)$ define $x_t = exp_x(t\nabla^L(-\phi)(x))$. Since $x_t \in \partial^{c_{t\lambda}}\phi_t(x)$, we have for $t \in (0,1]$

$$\begin{aligned} x_t &= exp_x\left(t\nabla^{L_\lambda}(-\phi)(x)\right) \\ &= exp_x\left(\frac{t \cdot l_\lambda^{-1}(t \cdot t^{-1}|\nabla(-\phi)|(x))}{|\nabla(-\phi)|(x)}\nabla(-\phi)(x)\right) \\ &= exp_x\left(\frac{l_{t\lambda}^{-1}(t^{-1}|\nabla(-\phi)|(x))}{|\nabla(-\phi)|(x)}\nabla(-\phi)(x)\right) \\ &= exp_x\left(\frac{l_{t\lambda}^{-1}(|\nabla(-t^{-1}\phi)|(x))}{|\nabla(-t^{-1}\phi)|(x)}\nabla(-t^{-1}\phi)(x)\right). \end{aligned}$$

Since $t^{-1}\phi$ is $c_t$-concave and $t^{-1}\phi(x_0) = 0$, uniqueness implies $\phi_t = t^{-1}\phi$. □

*Remark.* Note that this agrees with the cases $L(r) = r^p/p$: Assume for simplicity that $w_p(\mu_0, \mu_1) = 1$ then $\phi^L = \phi^{c_p}$ and $L_t = t^p d^p/p$. Hence

$$\begin{aligned} \phi_t^{c_t}(y) &= \inf t^p \frac{d^p(x,y)}{p} - t^{-1}\phi(x) \\ &= t^{-p} \inf \frac{d^p(x,y)}{p} - t^{p-1}\phi(x) = t^{-p}(t^{p-1}\phi)^{c_p}(y) \end{aligned}$$

Thus up to a factor the interpolation potentials are the same (recall that $t^{p-1}\phi$ gives the potential of $(\mu_0, \mu_t)$ w.r.t. $c_p$).

The next results follow using exactly the same arguments as for $c_p$.

**Lemma 7.13.** *Let $\mu_0$ and $\mu_1$ be two probability measures on $M$. Then there exists a unique (up to constant) $c_L$-concave function $\phi$ that solves the Monge-Kantorovich problem w.r.t. $L$. Moreover, if $\mu_0$ is absolutely continuous, then the vector field $\nabla^L(-\phi)$ is unique among such minimizers.*

*Remark.* At this point we do not work with $\mathcal{P}_L(M)$ directly. However all statements make sense also for $L_\lambda$ and any $\lambda > 0$ and we will see later that Lemma 7.9 can be used to show that the interpolation inequality in Theorem 7.21 is actually an interpolation inequality w.r.t. the geodesic $t \mapsto \mu_t$ in $\mathcal{P}_L(M)$ if the function $L_\lambda$ is used with $\lambda = w_L(\mu_0, \mu_1)$.

**Theorem 7.14.** *Let $\mu_0$ and $\mu_1$ be two probability measures on $M$ and assume $\mu_0$ is absolutely continuous with respect to $\mu$. Then there is a $c_L$-concave function $\phi$ such that $\pi = (\mathrm{Id} \times \mathcal{F})_* \mu_0$ is the unique optimal coupling of $(\mu_0, \mu_1)$ w.r.t. $L$, where $\mathcal{F}(x) = exp_x(\nabla^L(-\phi))$. Moreover, $\mathcal{F}$ is the unique optimal transport map from $\mu_0$ to $\mu_1$.*

**Corollary 7.15.** *If $\phi$ is $c_L$-concave and $\mu_0$ is absolutely continuous, then the map $\mathcal{F}(x) := exp_x(\nabla^L(-\phi))$ is the unique optimal transport map from $\mu_0$ to $\mathcal{F}_*\mu_0$ w.r.t. the cost function $c_L(x, y) = L(d(x, y))$.*

**Corollary 7.16.** *Assume $\mu_0$ is absolutely continuous and $\phi$ is $c_\lambda$-concave with $\lambda = w_L(\mu_0, (\mathcal{F}_1)_*\mu_0)$ where $\mathcal{F}_t(x) := exp_x(\nabla^\lambda(-t^{-1}\phi))$, then $\mathcal{F}_t$ is the unique optimal transport map from $\mu_0$ to $\mu_t = (\mathcal{F}_t)_*\mu_0$ w.r.t. $L_\lambda$ and $t \mapsto \mu_t$ is a constant geodesic from $\mu_0$ to $\mu_1$ in $\mathcal{P}_L(M)$.*

*Remark.* We will see in Lemma 7.22 below that the interpolation measures are absolutely continuous if $\mu_0$ and $(\mathcal{F})_{t*}\mu_0$ are.

*Proof.* We only need to show that

$$w_L(\mu_s, \mu_t) \le |s - t| w_L(\mu_0, \mu_1).$$

Let $\pi$ be the plan on $\mathrm{Geo}(M) = \{\gamma : [0, 1] \to M \,|\, \gamma \text{ is a geodesic in } M\}$ give by $\mu_0$, the map $\mathcal{F}_1$ and the unique geodesic connecting $\mu$-almost every $x \in M$ to a point $\mathcal{F}_1(x)$ (existence follows from [Lis06, Proof of Prop. 4.1], see also [Vil09, Chapter 7]), in particular, $\mu_t = (\mathcal{F}_t)_*\mu_0$. We also have

$$\int L\left(\frac{d(\gamma_0, \gamma_1}{\lambda}\right) d\pi(\gamma) = 1$$

for $\lambda = w_L(\mu_0, \mu_1)$ by definition $w_L$. Since $(e_s, e_t)_*\pi$ is a plan between $\mu_s$ and $\mu_t$ for $s, t \in [0, 1]$ we have

$$\int L\left(\frac{d(\gamma_s, \gamma_t)}{|t - s|\lambda}\right) d\pi(\gamma) = \int L\left(\frac{d(\gamma_0, \gamma_0)}{\lambda}\right) d\pi(\gamma) = 1.$$

Therefore, $w_L(\mu_s, \mu_t) \le |t - s|\lambda$. □

### 7.2.3 Almost Semiconcavity of Orlicz-concave functions

The proof of almost semiconcavity of $c_L$-concave functions follows along the lines of the proof of Theorem 4.10 by noticing that $\phi_s = s^{-1}\phi$ will be $c_s$-concave instead of $c_L$-concave, i.e. the type of concavity changes since the "distance changes".

**Theorem 7.17.** *Let $\phi$ be a $c_L$-concave function. Let $\Omega_{id}$ be the the points $x \in M$ where $\phi$ is differentiable and $d\phi_x = 0$, or equivalently $\partial^{c_L}\phi(x) = \{x\}$. Then $\phi$ is locally semiconcave on an open subset $U \subset M \backslash \Omega_{id}$ of full measure (relative to $M \backslash \Omega_{id}$). In particular, it is second order differentiable almost everywhere in $U$.*

*Proof.* Since $\partial^{c_L}\phi(x)$ is non-empty for every $x \in M$ and semicontinuous in $x$, we have the following: if $\phi$ is differentiable in $x$ with $d\phi_x \neq 0$ then $x \in \operatorname{int}(M\backslash\Omega_{id})$. Thus it suffices to show that each such points has a neighborhood $U_1$ in which $\phi$ is uniformly semiconcave.

So fix such an $x$ with $d\phi(x) \neq 0$ and note that $\phi$ is semiconcave on $U_1$ iff $\lambda\phi$ is for an arbitrary $\lambda > 0$. Furthermore, by Lemma 7.12 we know that $\phi_s = s^{-1}\phi$ is $c_s$-concave for any $s \in [0,1]$.

Since $d\phi(x) \neq 0$, there is a unique $y \in M$ with $\partial^{c_L}\phi(x) = \{y\}$ and a unique geodesic $\eta : [0,1] \to M$ between $x$ and $y$ (see Lemma 7.11). Note that $\phi_s$ is also differentiable at $x$ and

$$\partial^{c_s}\phi_s(x) = \{\eta(s)\}.$$

Let $s \in [0,1]$ be such that $d(x,\eta(s)) < \frac{r_{min}}{2}$. Because $x \neq \eta(s)$ and $z \mapsto \partial^{c_s}\phi_s(z)$ is continuous and single valued at $x$, we can find a neighborhood $V_1 \subset U$ of $y$ such that $(\partial^{c_s}\phi_s)^{-1}(V_1) \cap U$ contains some ball $B_{2\epsilon}(x)$ disjoint from $V_1$. Thus the functions $\{L_y : B_{2\epsilon}(x) \to \mathbb{R}\}_{y\in V_1}$ are semiconvave with constant $C$.

Now let $\gamma : [0,1] \to B_{2\epsilon}(x)$ be a minimal geodesic and set $x_t = \gamma(t)$. Choose $y_t \in \partial^{c_s}\phi_s(x_t) \cap V_1$. By the definition of $c_s$-concavity we have

$$\begin{aligned}\phi_s(x_0) &\leq \phi_s(x_t) + L(d(x_0,y_t)) - L(d(x_t,y_t))\\ \phi_s(x_1) &\leq \phi_s(x_t) + L(d(x_1,y_t)) - L(d(x_t,y_t)).\end{aligned}$$

Further, because $y_t \in V_1$ we also have

$$L(d(x_t,y_t)) \geq (1-t)L(d(x_0,y_t)) + tL(d(x_1,y_t)) - C(1-t)td^2(x_0,x_1).$$

Therefore, taking the $(1-t), t$ convex combination of the first two inequality we obtain

$$\begin{aligned}\phi_s(x_t) &\geq (1-t)\phi_s(x_0) + t\phi_s(x_1) + L(d(x_t,y_t)) - (1-t)L(d(x_0,y_t)) - tL(d(x_1,y_t))\\ &\geq (1-t)\phi_s(x_0) + t\phi_s(x_1) + C(1-t)td^2(x_0,x_1).\end{aligned}$$

□

### 7.2.4 Interpolation inequality in the Orlicz-Wasserstein space

**Theorem 7.18** (Volume distortion for $L$)**.** *Let $x \neq y$ with $y \notin \operatorname{Cut}(x)$ and $\eta$ be the unique minimal geodesic from $x$ to $y$. For $t \in (0,1]$ define $f_t(z) = -L_t(d(z,\eta(t)))$.*

*Then we have*

$$\begin{aligned}\mathfrak{v}_t^{<}(x,y) &= \mathbf{D}\left[d(exp_x)_{\nabla^{L_t}f_t(x)} \circ [d(exp_x)_{\nabla^{L}f_1(x)}]^{-1}\right]\\ \mathfrak{v}_t^{>}(x,y) &= (1-t)^{-n}\mathbf{D}\left[d(exp_x \circ \mathcal{K}_x^t)_{d(t^{-1}f_1)_x} \circ [d\left(d(t^{-1}f_1)\right)_x - d\left(df_t\right)_x]\right].\end{aligned}$$

*Remark.* The statements hold equally if one takes $L_\lambda$ and $L_{t\lambda}$, they only depend on the smoothness of $L$.

*Proof.* Recall Theorem 4.12 and the function $g_t(z) = -d^2(x, \eta(t))/2$.

We have for $L_{\eta(t)} = L_t(d(\cdot, \eta(t))$

$$\nabla L_{\eta(t)}(x) = l_t(d(x, \eta(t))\nabla d(x, \eta(t))$$

and thus

$$\nabla^t f_t(z) = l_t^{-1}(l_t(d(z, \eta(t))\nabla(-d(z, \eta(t)) = \nabla g_t(z)$$

which implies the first equation.

For the second part note that (see calculations in the proof of Lemma 7.12)

$$\begin{aligned}\mathcal{K}_z^t(d(t^{-1}f_1)_z) &= \frac{l_t^{-1}(t^{-1}|\nabla f_1|(z))}{|\nabla f_1|(z)}\nabla f_1(z) \\ &= t\frac{l^{-1}(|\nabla f_1|(z))}{|\nabla f_1|(z)}\nabla f_1(z) \\ &= t\nabla^L f_1(z) = \mathcal{L}_z(d(tg_1)_z)\end{aligned}$$

and hence

$$\begin{aligned}\mathfrak{v}_t^{>}(x,y) &= (1-t)^{-n}\mathbf{D}\left[d\left(exp \circ \mathcal{L} \circ (d(tg_1)_z)\right)\right] \\ &= (1-t)^{-n}\mathbf{D}\left[d(exp \circ \mathcal{K}^t)_{d(t^{-1}f_1)_x} \circ d\left(d(t^{-1}f_1)\right)_x\right].\end{aligned}$$

We have $d(f_t)_x = d(t^{-1}f_1)_x$. Indeed, since $l_t(r) = t^{-1}l(t^{-1}r)$ and $d(d(\cdot, \eta(t))_x = d(d(\cdot, y))_x$

$$\begin{aligned}-d(f_t)_x &= d(L_t(d(\cdot, \eta(t)))_x \\ &= l_t(d(x, \eta(t))d(d(\cdot, \eta(t))_x \\ &= t^{-1}l(t^{-1}td(x, y))d(d(\cdot, y))_x = -d(t^{-1}f_1)_x\end{aligned}$$

Similar to [Oht09, Proof of 3.2] it suffices to show that

$$d(exp_x \circ \mathcal{K}_x^t)_{d(f_t)_x} \circ d(t^{-1}df_t)_x = 0.$$

Now since $\nabla f_t(z) = l_t(d(z, \eta(t)))\nabla d_{\eta(t)}(z)$ we get in a neighborhood $U$ of $x$ not containing $\eta(t)$,

$$\begin{aligned}\mathcal{K}_z^t(d(f_t)_z) &= \nabla^{L_t}(t^{-1}f_t)(z) \\ &= l_t^{-1}(l_t(d(z, \eta(t))))\nabla d_{\eta(t)}(z) \\ &= \mathcal{L}_z(d(g_t)_z)\end{aligned}$$

and thus the function $D: U \to M$ defined as

$$\begin{aligned}D(z) = exp_z \circ \mathcal{K}_z(d(f_t)_z) &= exp_z \circ \mathcal{L}_z(d(g_t)_z) \\ &= \eta(t),\end{aligned}$$

is constant in a neighborhood of of $x$. This immediately implies $dL_x = 0$. □

**Proposition 7.19.** *Let $\phi : M \to \mathbb{R}$ be a $c_L$-concave function and define $\mathcal{F}(z) = exp_z(\nabla^L(-\phi)(z))$ at all points of differentiability of $\phi$. Fix some $x \in M$ such that $\phi$ is twice differentiable at $x$ and $d\phi_x \neq 0$. Then the following holds:*

1. *$y = \mathcal{F}(x)$ is not a cut point of $x$.*
2. *The function $h(z) = c_L(z, y) - \phi(z)$ satisfies $dh_x = 0$ and*
$$\left(\frac{\partial^2 h}{\partial x^i \partial x^j}(x)\right) \geq 0$$
*in any local coordinate system $(x^i)_{i=1}^n$ around $x$.*
3. *Define $f_y(z) := -c_L(z, y)$ and*
$$d\mathcal{F}_x := d(\exp_x \circ \mathcal{K}_x^L)_{d(-\phi)_x} \circ [d(d(-\phi))_x - d(df_y)_x] : T_x M \to T_y M$$
*where the vertical part of $T_{d(-\phi)_x}(T^*M)$ and $T_{d(-\phi)_x}(T^*M)$ are identified. Then the following holds for all $v \in T_x M$*
$$\sup\left\{|u - d\mathcal{F}_x(v)| \mid \exp_y u \in \partial^{c_L}\phi(\exp_x y), |u| = d(y, \exp_y u)\right\} = o(|v|).$$

*Proof.* The proof follows without any change from the proof of Proposition 4.14 but using Lemma 7.10 instead and the fact that $y \notin \mathrm{Cut}(x) \cup \{x\}$ implies that $f_y$ is $C^\infty$ at $x$ and $\nabla^L f_y(x) = \nabla^L \phi(x)$. $\square$

Similarly the Jacobian equation holds:

**Proposition 7.20.** *Let $\mu_0$ and $\mu_1$ be absolutely continuous measures with density $f_0$ and $f_1$ and $\lambda = w_L(\mu_0, \mu_1)$. Also assume that there are open sets $U_i$ with compact closed $X = \bar{U}_0$ and $Y = \bar{U}_1$ such that* $\operatorname{supp} \mu_i \subset U_i$*. Let $\phi$ be the unique $c_\lambda$-concave Kantorovich potential and define $\mathcal{F}(z) = exp_z(\nabla^\lambda(-\phi)(z))$. Then $\mathcal{F}$ is injective $\mu_0$-almost everywhere and for $\mu_0$-almost every $x \in M \backslash \Omega_{id}$*

1. *The function $h(z) = c_\lambda(z, \mathcal{F}(z)) - \phi(z)$ satisfies*
$$\left(\frac{\partial^2 h}{\partial x^i \partial x^j}(x)\right) > 0$$
*in any local coordinate system $(x^i)_{i=0}^n$ around $x$.*
2. *In particular, $\mathbf{D}[d\mathcal{F}_x] > 0$ holds for the map $d\mathcal{F}_x : T_x M \to T_{\mathcal{F}(x)} M$ defined as above and*
$$\lim_{r \to 0} \frac{\mu(\partial^{c_\lambda}\phi(B_r^+(x)))}{\mu(B_r^+(x))} = \mathbf{D}[d\mathcal{F}_x]$$
*and*
$$f(x) = g(\mathcal{F}(x))\mathbf{D}[d\mathcal{F}_x].$$

*Remark.* defining $d\mathcal{F}_x = \mathrm{Id}$ for points $x$ of differentiability of $\phi$ with $d\phi_x = 0$ we see that the second statement above holds $\mu$-a.e.

*Proof.* Similar to Proposition 4.15, the proof follows without any change from [Oht09, Theorem 5.2], see also [Vil09, Chapter 11]. □

**Theorem 7.21.** *Let $\phi : M \to \mathbb{R}$ be a $c_L$-concave function and $x \in M$ such that $\phi$ is second order differentiable with $d\phi_x \neq 0$. For $t \in (0,1]$, define $y_t := \exp_x(\nabla^t(-t^{-1}\phi))$, $f_t(z) = -c_t(z, y_t)$ and $\mathbf{J}_t(x) = \mathbf{D}[d(\mathcal{F}_t)_x]$ where*

$$d(\mathcal{F}_t)_x := d(\exp_x \circ \mathcal{K}_x^t)_{d(-t^{-1}\phi)_x} \circ \left[d(d(-t^{-1}\phi))_x - d(d(f_t))_x\right] : T_xM \to T_{y_t}M.$$

*Then for any $t \in (0,1)$*

$$\mathbf{J}_t(x)^{1/n} \geq (1-t)\mathfrak{v}_t^{>}(x, y_1)^{1/n} + t\mathfrak{v}_t^{<}(x, y_1)^{1/n}\mathbf{J}_1(x)^{1/n}.$$

*Remark.* The proof is based on the proof of [Oht09, Proposition 5.3] but is notationally slightly more involved then the proof of Theorem 4.16.

*Proof.* Note first that

$$d(d(-t^{-1}\phi))_x - d(df_t)_x = \left\{d(d(-t^{-1}\phi))_x - d(d(t^{-1}f_1))_x\right\} + \left\{d(d(t^{-1}f_1))_x - d(df_t)_x\right\}$$

and

$$d(f_t)_x = d(-t^{-1}\phi)_x = d(-t^{-1}f_1)_x.$$

Now define $\tau_s : T^*M \to T^*M$ as $\tau_s(v) = s^{-1}v$ and note for $\nabla\phi(x) \neq 0$

$$\begin{aligned}
\mathcal{K}_x^t(t^{-1}d\phi_x) &= \frac{l_{t\lambda}^{-1}(|\nabla t^{-1}\phi(x)|)}{|\nabla t^{-1}\phi(x)|}\nabla t^{-1}\phi(x) \\
&= t\frac{l_\lambda(|\nabla\phi(x)|)}{|\nabla\phi(x)|}\nabla\phi(x) = t\mathcal{K}_x^L(d\phi_x)
\end{aligned}$$

and thus

$$\mathcal{K}_x^t \circ \tau_t \circ (\mathcal{K}_x^L)^{-1} = t\,\mathrm{Id}_{T_xM}$$

which implies

$$\begin{aligned}
& d(\exp_x \circ \mathcal{K}_x^t)_{d(-t^{-1}\phi)_x} \circ \left(d(d(-t^{-1}\phi))_x - d(d(t^{-1}f_1))_x\right) \\
= \;& d(\exp_x \circ \mathcal{K}_x^t)_{d(-t^{-1}\phi)_x} \circ d(\tau_t)_{d(-\phi)_x} \circ [d(d(-\phi))_x - d(d(f_1))_x] \\
= \;& d(\exp_x \circ \mathcal{K}_x^t)_{d(-t^{-1}\phi)_x} \circ d(\tau_t)_{d(-\phi)_x} \circ \left[d(exp_x \circ \mathcal{K}_x^L)_{d(-\phi)_x}\right]^{-1} \circ d(\mathcal{F}_1) \\
= \;& d(\exp_x)_{\nabla^t(-t^{-1}\phi)_x} \circ d(\mathcal{K}_x \circ \tau_t \circ \mathcal{K}_x^{-1})_{\nabla^L(-\phi)_x} \circ [d(\exp_x)_{\nabla^L(-\phi)_x}]^{-1} \circ d(\mathcal{F}_1) \\
= \;& t \cdot d(\exp_x)_{\nabla^t(-t^{-1}\phi)_x} \circ [d(\exp_x)_{\nabla^L(-\phi)_x}]^{-1} \circ d(\mathcal{F}_1).
\end{aligned}$$

where we identified $T_{\nabla^t(-t^{-1}\phi)(x)}(T_xM)$ with $T_{\nabla^L(-\phi)(x)}(T_xM)$ to get the last inequality (remember $t\nabla^t(-t^{-1}\phi) = \nabla^L(-\phi)$).

Because $\mathbf{D}$ is concave we get

$$
\begin{aligned}
& \mathbf{J}_t(x)^{1/n} = \mathbf{D}[d(\mathcal{F}_t)_x]^{1/n} \\
= \; & \mathbf{D}\Big[d(\exp_x \circ \mathcal{K}^t_x)_{d(-t^{-1}\phi)_x} \circ \big[d(d(t^{-1}f_1))_x - d(df_t)_x\big] \\
& + d(\exp_x \circ \mathcal{K}^t_x)_{d(-t^{-1}\phi)_x} \circ \big(d(d(-t^{-1}\phi))_x - d(d(t^{-1}f_1))_x\big)\Big]^{1/n} \\
= \; & \mathbf{D}\Big[d(\exp_x \circ \mathcal{K}^t_x)_{d(-t^{-1}\phi)_x} \circ \big(d(d(t^{-1}f_1))_x - d(df_t)_x\big) \\
& + t \cdot d(\exp_x)_{\nabla^t(-t^{-1}\phi)_x} \circ [d(\exp_x)_{\nabla^L(-\phi)_x}]^{-1} \circ d(\mathcal{F}_1)\Big]^{1/n} \\
\geq \; & (1-t)\mathbf{D}\Big[(1-t)^{-1} d(\exp_x \circ \mathcal{K}^t_x)_{d(-t^{-1}\phi)_x} \circ \big(d(d(t^{-1}f_1))_x - d(df_t)_x\big)\Big]^{1/n} \\
& + t\mathbf{D}\Big[d(\exp_x)_{\nabla^t(-t^{-1}\phi)_x} \circ [d(\exp_x)_{\nabla^L(-\phi)_x}]^{-1} \circ d(\mathcal{F}_1)\Big]^{1/n} \\
= \; & (1-t)\mathfrak{v}^{>}_t(x,y_1)^{1/n} + t\mathfrak{v}^{<}_t(x,y_1)^{1/n}\mathbf{J}_1(x)^{1/n}.
\end{aligned}
$$

□

Combing this with Lemma 3.11 (see remark after that lemma) and Lemma 7.23 below we get similar to Lemma 4.17 and [Oht09, 6.2]:

**Lemma 7.22.** *Given two absolutely continuous measures $\mu_i = \rho_i\mu$ on $M$, let $\phi$ be the unique $c_\lambda$-concave optimal Kantorovich potential with $\lambda = w_L(\mu_0,\mu_1)$. Define $\mathcal{F}_t(x) := \exp_x(\nabla^{t\lambda}(-t^{-1}\phi))$ for $t \in (0,1]$. Then $\mu_t = \rho_t d\mu$ is absolutely continuous for any $t \in [0,1]$.*

*Proof.* By Lemma 3.11 the map $\mathcal{F}_t$ is injective $\mu_0$-almost everywhere. Let $\Omega_{id}$ be the points $x \in M$ of differentiability of $\phi$ with $d\phi_x = 0$. Then

$$\mu_t\big|_{\Omega_{id}} = (\mathcal{F}_t)_*(\mu_0\big|_{\Omega_{id}}) = \mu_0\big|_{\Omega_{id}}.$$

By Theorem 4.10 the potential $\phi$ is second order differentiable in a subset $\Omega \subset M\backslash\Omega_{id}$ of full measure. In addition, $\mathbf{D}[d(\mathcal{F}_1)] > 0$ for all $x \in \Omega$ (see Proposition 4.15) and $\mathcal{F}_t$ is continuous in $\Omega$ for any $t \in [0,1]$. The map $d(\mathcal{F}_t)_x : T_xM \to T_{\mathcal{F}_t(x)}M$ defined in Proposition 4.14 as

$$d(\mathcal{F}_t)_x := d(\exp_x \circ \mathcal{K}^t_x)_{d(-t^{-1}\phi)} \circ \big[d(d(-t^{-1}\phi))_x - d(d(f_t)_x\big]$$

where $f_t(z) := -c_{t\lambda}(z, \mathcal{F}_t(x))$ for $t \in (0,1]$. Also note that for $x \in \Omega$

$$d(d(-t^{-1}\phi))_x - d(df_t)_x = \big\{d(d(-t^{-1}\phi))_x - d(d(t^{-1}f_1))_x\big\} + \big\{d(d(t^{-1}f_1))_x - d(f_t)_x\big\}.$$

Which implies $\mathbf{D}[d(\mathcal{F}_t)_x] > 0$ because $\mathbf{D}[d(\mathcal{F}_1)_x) > 0$ and the lemma below.

The result then immediately follows by [CEMS01, Claim 5.6]. □

**Lemma 7.23.** *Let $y \notin \mathrm{Cut}(x) \cup \{x\}$ and $\eta : [0,1] \to M$ be the unique minimal geodesic from $x$ to $y$. Define*

$$f_t(z) = -c_t(z, \eta(t)).$$

*Then the function $h(z) = t^{-1} f_1(z) - f_t(z)$ satisfies*

$$\left( \frac{\partial^2 h}{\partial x^i \partial x^j}(x) \right) \geq 0$$

*in any local coordinate system around $x$.*

*Proof.* This follows directly from 7.8. □

Using this interpolation inequality, one can show that a curvature dimension condition $CD_L(K,N)$ holds on any $n$-dimensional ($n < N$) Finsler manifold $M$ with (weighted) Ricci curvature bounded from below by $K$. The condition $CD_L(K,N)$ is nothing but a convexity property of functionals in $\mathcal{DC}_N$ along geodesics in $\mathcal{P}_L(M)$. Most geometric properties (Brunn-Minkowski, Bishop-Gromov, local Poincaré and doubling) also hold under such a condition. However, the lack of an "easy-to-understand" dual theory makes it difficult to prove statements involving (weak) upper gradients.

**Corollary 7.24.** *Any $n$-dimensional Finsler manifold with $N$-Ricci curvature bounded from below by $K$ and $N > n$ satisfies the very strong $CD_L(K,N)$ condition for all strictly convex, increasing functional $L : [0,\infty) \to [0,\infty)$ which is smooth away from zero.*

# Conclusion and outlook

## Curvature dimension and Orlicz-Wasserstein spaces

In this thesis a proof of the Borel-Brascamp-Lieb (BBL) inequality along geodesics in $p$-Wasserstein spaces and along geodesics in Orlicz-Wasserstein spaces was given. This led to a new definition of an abstract curvature condition $CD_p(K, N)$ and resp. $CD_L(K, N)$ along the lines of Lott-Villani [LV09, LV07] and Sturm [Stu06a, Stu06b]. The conditions can be defined on metric measure spaces and similar to the case $p = 2$ one gets nice geometric and analytical properties of the space. In particular, a metric variant of Brenier's theorem (Lemma 4.2) and a Laplacian comparison theorem (Theorem 5.14) were derived.

To prove the interpolation inequality leading to the BBL inequality the Brenier-McCann-Ohta solution was extened to cover the cases $p \neq 2$ and weak regularity properties of the transport map are given. For this the author showed how Ohta's orginial idea *to avoid the diagonal* can be used to give a nice and short proof of the almost everywhere second order differentiability of the transport map.

For positivly curved spaces $K$-convexity of the entropy functional follows and a new global Poincaré inequality can be derived. This global Poincaré inequality holds for manifolds with positive Ricci curvature, because manifolds with Ricci curvature bounded from below by $K > 0$ satisfy the $CD_p(K_p, \infty)$-condition for any $p$, where $0 < c_K \leq K_p \leq C_K$ depends on $K$ and $p$ (see remark on page 48).

In the future, it might be interesting to see what the strong $CD_p(K, \infty)$-condition for negative $K$ and unbounded spaces means. As seen on page 48, the strong $CD_p(K, \infty)$-condition does not imply $K$-convexity of the functional generated by $\mathcal{DC}_\infty$. However, one always gets $K'$-convexity if the space is bounded. Hence, the strong $CD_p(K, N)$-condition behaves differently for unbounded spaces.

In Chapter 7 the theory of Orlicz-Wasserstein spaces was developed. In particular, their geodesic character and a structure of weak topologies among different Orlicz-Wasserstein spaces similar to the cases $1 \leq p < \infty$ was shown.

Having a description of the geodesics of Orlicz-Wasserstein spaces, the author could follow along the lines of $1 < p < \infty$ to show the interpolation inequality which can be used to show that the condition $CD_L(K, N)$ holds for $n$-dimensional (Finsler) manifolds with Ricci curvature bounded from below by $K$ and $N > n$.

In contrast to $p$-Wasserstein space, Orlicz-Wasserstein spaces are defined via two optimization problems and there is no nice dual theory for those spaces. Furthermore, it is not clear what the analytical "vertical dual" of the theory is. In this thesis we could show that the analyticial dual of $p$-Wasserstein spaces is the theory of the $q$-Cheeger functional. Only from this, one gets the $q$-Laplacian which is required to formulate the

Laplacian comparison theorm. By now it is not clear what the Orlicz-Cheeger functional looks like. But by its nonlinear character it is clear that it must be more general than $f \mapsto \int L(|\nabla f|_*)d\mu$.

## Heat and gradient flows

In Chapter 6 the theory of the $q$-heat flow as a gradient flow of the $q$-Cheeger energy was developed further. On the one hand, a more general comparison theorem and a calculus along the $q$-heat flow (Theorem 6.5) was given which does not require bounds for the density from above and away from zero. This made it possible to prove mass preservation with a natural growth condition on the measure of balls of the background measure. In the furture the author hopes to be able to drop the condition $\int V^p \exp_p(-V^p)d\mu < \infty$, so that only $\int \exp_p(-V^p)d\mu < \infty$ is required to obtain mass preservation.

In the second part of Chapter 6 the developed theory was used to show that the $q$-heat flow is a solution of the gradient flow problem of the $(3-p)$-Renyi entropy functional in the $p$-Wasserstein space. For this, one needed to show that all measures at which the descending slope is finite are absolutely continuous. The current proof for $p \in (2,3)$ requires the space to be compact and $n$-Ahlfors regular. Finding a proof which drops one of these conditions might be challenging.

The purpose of the chapter was to identify the $q$-heat flow with the gradient flow of the Renyi entropy functional. In the cases $p \in (1,2)$ convexity of the function $(x,y) \mapsto x^q/y^{p-1}$ are used to give uniqueness which implies immediately the identification of the two flows. However, one can easily show that this function is no convex in case $p \in (2,3)$.

In the future, the author plans to further analyze the $q$-heat flow and also the heat flow on spaces which are not infinitesimal Hilbertian, i.e. the heat flow is not linear. Currently it is not know if the $q$-heat flow is contractive in the $p$-Wasserstein space. However, Ohta and Sturm used in [OS11] a linearization of the heat flow on Finsler manifolds to give pseudo-contraction property. With the help of this one can show that for $t > 0$ the (nonlinear) heat flow $P_t$ maps into the Lipschitz functions. Furthermore, the author hopes that if a similar (semi)linearization can be done for the $q$-heat flow, then it might be possible to use it to show uniqueness of the gradient flow problem of the Renyi entropy functional in the $p$-Wasserstein space.

# Bibliography


[AGMR12] L. Ambrosio, N. Gigli, A. Mondino, and T. Rajala, *Riemannian Ricci curvature lower bounds in metric measure spaces with $\sigma$-finite measure*, arxiv:1207.4924 (2012).

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[AGS11a] ———, *Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces*, arxiv:1111.3730 (2011).

[AGS11b] ———, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, arxiv:1109.0222 (2011).

[AGS13] ———, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Inventiones mathematicae (2013).

[Agu02] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, PhD-Thesis (2002), 1–96.

[Agu05] ———, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, Advances in Differential Equations **10** (2005), no. 3, 309–360.

[BCS00] D. Bao, S. Chern, and Z. Shen, *An Introduction to Riemann-Finsler Geometry*, 2000.

[Bor97] J. M. Borwein, *A Generalization Of Young's $\ell^p$ Inequality*, Mathematical Inequalities and Applications 1, 1997.

[Bré73] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, Elsevier, 1973.

[Bre91] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), no. 4, 375–417.

[BS10] K. Bacher and K.-Th. Sturm, *Localization and tensorization properties of the curvature-dimension condition for metric measure spaces*, Journal of Functional Analysis **259** (2010), no. 1, 28–56.

[CEMS01] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Inventiones Mathematicae **146** (2001), no. 2, 219–257.

[Che99] J. Cheeger, *Differentiability of Lipschitz Functions on Metric Measure Spaces*, Geometric And Functional Analysis **9** (1999), no. 3, 428–517.

[DS13] S. Di Marino and G. Speight, *The p-Weak Gradient Depends on p*, arxiv:1311.4171 (2013).

[EKS13] M. Erbar, K. Kuwada, and K.-Th. Sturm, *On the Equivalence of the Entropic Curvature-Dimension Condition and Bochner's Inequality on Metric Measure Spaces*, arxiv:1303.4382 (2013).

[Erb10] M. Erbar, *The heat equation on manifolds as a gradient flow in the Wasserstein space*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **46** (2010), no. 1, 1–23.

[Gig09] N. Gigli, *On the heat flow on metric measure spaces: existence, uniqueness and stability*, Calculus of Variations and Partial Differential Equations **39** (2009), no. 1-2, 101–120.

[Gig12] ———, *On the differential structure of metric measure spaces and applications*, arxiv:1205.6622 (2012).

[Gig13] ———, *The splitting theorem in non-smooth context*, arxiv:1302.5555 (2013).

[GM13] N. Garofalo and A. Mondino, *Li-Yau and Harnack type inequalities in $RCD^*(K,N)$ metric measure spaces*, arxiv:1306.0494 (2013).

[GRS12] N. Gozlan, C. Roberto, and P.-M. Samson, *Hamilton Jacobi equations on metric spaces and transport-entropy inequalities*, arxiv:1203.2783 (2012).

[GRS13] N. Gigli, T. Rajala, and K.-Th. Sturm, *Optimal maps and exponentiation on finite dimensional spaces with Ricci curvature bounded from below*, arxiv:1305.4849 (2013).

[HKX13] B. Hua, M. Kell, and Ch. Xia, *Harmonic functions on metric measure spaces*, arxiv:1308.3607 (2013).

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto, *The Variational Formulation of the Fokker–Planck Equation*, SIAM Journal on Mathematical Analysis **29** (1998), no. 1, 1–17.

[Kel11] M. Kell, *Stability of the global attractor under Markov-Wasserstein noise*, arxiv:1103.3401 (2011).

[Kel13] ———, *On Interpolation and Curvature via Wasserstein Geodesics*, arxiv:1311.5407 (2013).

[Kel14] ______, *q-heat flow and the gradient flow of the Renyi entropy in the p-Wasserstein space*, arxiv:1401.0840 (2014).

[Kuw10] K. Kuwada, *Duality on gradient estimates and Wasserstein controls*, Journal of Functional Analysis **258** (2010), no. 11, 3758–3774.

[Lis06] St. Lisini, *Characterization of absolutely continuous curves in Wasserstein spaces*, Calculus of Variations and Partial Differential Equations **28** (2006), no. 1, 85–120.

[LV07] J. Lott and C. Villani, *Weak curvature conditions and functional inequalities*, Journal of Functional Analysis **245** (2007), no. 1, 311–333.

[LV09] ______, *Ricci curvature for metric-measure spaces via optimal transport*, Annals of Mathematics **169** (2009), no. 3, 903–991.

[McC01] R. J. McCann, *Polar factorization of maps on Riemannian manifolds*, Geometric and Functional Analysis **11** (2001), no. 3, 589–608.

[Oht08] S. Ohta, *Uniform convexity and smoothness, and their applications in Finsler geometry*, Mathematische Annalen **343** (2008), no. 3, 669–699.

[Oht09] ______, *Finsler interpolation inequalities*, Calculus of Variations and Partial Differential Equations **36** (2009), no. 2, 211–249.

[Oht13a] ______, *$(K, N)$-convexity and the curvature-dimension condition for negative $N$*, arxiv:1310.7993 (2013).

[Oht13b] ______, *Splitting theorems for Finsler manifolds of nonnegative Ricci curvature*, Journal für die reine und angewandte Mathematik (2013).

[OS11] S. Ohta and K.-Th. Sturm, *Bochner-Weitzenboeck formula and Li-Yau estimates on Finsler manifolds*, arxiv:1105.0983 (2011).

[OT11a] S. Ohta and A. Takatsu, *Displacement convexity of generalized relative entropies*, Advances in Mathematics **228** (2011), no. 3, 1742–1787.

[OT11b] ______, *Displacement convexity of generalized relative entropies.II*, arxiv:1112.5554 (2011).

[Ott96] F Otto, *Double degenerate diffusion equations as steepest descent*, preprint, SFB256 (1996).

[Ott01] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Communications in Partial Differential Equations **26** (2001), no. 1-2, 101–174.

[Raj11] T. Rajala, *Local Poincaré inequalities from stable curvature conditions on metric spaces*, Calculus of Variations and Partial Differential Equations **44** (2011), no. 3-4, 477–494.

[Raj12] ______, *Interpolated measures with bounded density in metric spaces satisfying the curvature-dimension conditions of Sturm*, Journal of Functional Analysis **263** (2012), no. 4, 896–924.

[RMS08] R. Rossi, A. Mielke, and G. Savaré, *A metric approach to a class of doubly monlinear evolution equations and applications*, 2008, pp. 97–170.

[She97] Z. Shen, *Volume Comparison and Its Applications in Riemann-Finsler Geometry*, Advances in Mathematics **128** (1997), no. 2, 306–328.

[She01] ______, *Lectures on Finsler Geometry*, World Scientific, 2001.

[Stu06a] K.-Th. Sturm, *On the geometry of metric measure spaces*, Acta Mathematica **196** (2006), no. 1, 65–131.

[Stu06b] ______, *On the geometry of metric measure spaces. II*, Acta Mathematica **196** (2006), no. 1, 133–177.

[Stu11] ______, *Generalized Orlicz spaces and Wasserstein distances for convex-concave scale functions*, Bulletin des Sciences Mathématiques **135** (2011), no. 6-7, 795–802.

[Vil03] C. Villani, *Topics in Optimal Transportation*, 2003.

[Vil09] ______, *Optimal transport: old and new*, Springer Verlag, 2009.

**Informationen über den Autor**

---

| | |
|---|---|
| **Name:** | Martin Kell |
| **Geburtsdatum:** | 19. September 1985 in Neubrandenburg |
| | |
| **10/2005 - 09/2010** | Studium der Mathematik und Informatik<br>an der *Universität Rostock*, Rostock |
| **09/2007** | Vordiplom in Informatik |
| **08/2010** | Diplom in Mathematik |
| **07/2008 - 12/2009** | Studium der Mathematik<br>an der *Michigan State University*, East Lansing (USA) |
| **12/2009** | Master of Science in Mathematics |
| **seit 09/2010** | Doktorand an der IMPRS<br>des *Max Planck Institutes für Mathematik in den Naturwissenschaften*, Leipzig |
| **01/2012 - 02/2012** | Summer PhD-Fellow<br>am *Instituto Nacional de Matemtica Pura e Aplicada (IMPA)*,<br>Rio de Janeiro (Brasilien) |

## Bibliographische Daten

*On curvature conditions using Wasserstein spaces*
(*Über Krümmungsbedingungen mittels Wassersteinräumen*)
Kell, Martin
Universität Leipzig, Dissertation, 2014
100 Seiten, 52 Referenzen, S-Nummer 0

**Selbstständigkeitserklärung**

---

Hiermit erkläre ich, die vorliegende Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

Leipzig, den 17. Februar 2014

. . . . . . . . . . . . . . . . . . . . . . . . . . .
(Martin Kell)