ON INTERPOLATION AND CURVATURE VIA WASSERSTEIN GEODESICS

MARTIN KELL

Abstract. In this article, a proof of the interpolation inequality along geodesics in $p$-Wasserstein spaces is given. This interpolation inequality was the main ingredient to prove the Borel-Brascamp-Lieb inequality for general Riemannian and Finsler manifolds and led Lott-Villani and Sturm to define an abstract Ricci curvature condition. Following their ideas, a similar condition can be defined and for positively curved spaces one can prove a Poincaré inequality. Using Gigli’s recently developed calculus on metric measure spaces, even a $q$-Laplacian comparison theorem holds on $q$-infinitesimal convex spaces.

In the appendix, the theory of Orlicz-Wasserstein spaces is developed and necessary adjustments to prove the interpolation inequality along geodesics in those spaces are given.

The proof of the Borel–Brascamp–Lieb (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 [List06]. 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 maps maps to different points and (2) helps to move the terminal point away from the cut-locus. Using this, a proof

The author wants to thank Prof. Jürgen Jost and the MPI MiS for providing a stimulating research environment. The research was funded by the IMPRS “Mathematics in the Sciences”.
of the (almost) semiconcavity of solution potentials (Theorem 3.8) is given, which
is shorter than Ohta’s original 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 A.17.

The star-shapedness of $c_p$-concave functions, resp. pseudo star-shapedness of
$\mathcal{C}_L$-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 $m$
\[
t^{p-1}d^p(m,y) \leq d^p(m,z) + t^{p-1}(1-t)d^p(x,y)
\]
and
\[
t^{-1}L(d(m,y)) \leq L(d(m,z)/t) + t^{-1}(1-t)L(d(x,y))
\]
where $L$ is a strictly increasing convex function.

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 a second paper [Kel14] 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 $p \in (1,3)$.

In the end, we show how to prove the interpolation via geodesics in Orlicz-
Wasserstein spaces. Since Orlicz-Wasserstein spaces are notationally more involved
and we need some additional result, like the geodesic character of Orlicz-Wasserstein
spaces, we give those results in the appendix. However, the metric of Orlicz-
Wasserstein spaces is not defined via a single optimization problem. Thus there
is no natural dual problem and by now no "vertical dual" to the theory of Orlicz-
Wasserstein spaces, in particular, there is no Orlicz-Cheeger energy and no Orlicz-
Laplacian.

Now we will give an outline of the result and the structure of the paper: In
the first section, we will give an overview of the used concepts. The second
section will develop the theory of $c_p$-concave functions, their Lipschitz regularity
and star-shapedness. The next section will deal with the smooth setting, i.e. the
Brenier-McCann-Ohta solution, second order differentiability and the interpolation
inequality, which can be stated as
\[
J_t(x)^{1/n} \geq (1-t)v^>_t(x,y_1)^{1/n} + tv^<_t(x,y_1)^{1/n}J_1(x)^{1/n}
\]
where $J_t(x)$ is the Jacobian of the interpolation map and $v^>_t$ and $v^<_t$ are the (back-
ward, resp. forward) volume distortion coefficients. The convexity of functionals in
$\mathcal{DC}_N$ immediately follows along the by-now-standard lines and hence a version of
Lott-Villani’s curvature condition.

In the fourth section, we will define the curvature dimension condition $\text{CD}_p(K,N)$
along the lines of Lott-Villani-Sturm. Since most proofs, which don’t use Cauchy-
Schwary, can be easily adapted, we will only show a Poincaré inequality for posi-
tively curved $\text{CD}_p(K,\infty)$-spaces, i.e. $K > 0$. For those spaces, we get
\[
\left( \int (h - \bar{h})^2 d\mu \right)^{\frac{1}{2}} \leq \frac{1}{\sqrt{2K}} \left( \int |D^{-} h|^q d\mu \right)^{\frac{1}{q}}
\]
for any Lipschitz function $h$. In the end of this section we show a version of the
metric Brenier theorem and use Gigli’s recently developed calculus [Gig12] to give
a $q$-Laplacian comparison theorem, namely
\[ \Delta_q \phi \leq N \tilde{\sigma}_{K,N} (|\nabla \phi|_{u_0}^{q-1}) d\mu \]
for any $c_p$-concave function in the domain of the $q$-Laplacian.

In the appendix, we develop the theory of Orlicz-Wasserstein spaces and show how to adapt the proofs of the interpolations inequality.

1. Preliminaries

In this part, we will introduce the main concepts used in this work. For a general introduction to the theory of optimal transport and curvature via 2-Wasserstein spaces see [Vil09], especially its Chapter 6 on Wasserstein spaces. We follow Ohita's notation [Oht08, Oht09] for Finsler manifolds and otherwise refer to [BCS00, She01].

As a convention we will always assume that $(M,d,\mu)$ is a locally compact metric space equipped with a locally finite Borel measure $\mu$ and if not otherwise stated it is assumed to be geodesic (see below). Since we will also deal with spaces which are not locally compact (e.g. $(P_p(M), w_p)$ with $M$ non-compact), the sections below do not assume that $(X,d)$ is locally compact. And as an abbreviation define
\[ \mathbb{R} := \mathbb{R} \cup \{-\infty\}. \]

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

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 $AC^p(I,X) \subset C(I,X)$ where $C(I,X)$ is equipped with the sup distance $d^*$
\[ d^*(\gamma, \gamma') := \sup_{t \in I} d(\gamma_t, \gamma'_t). \]
For each $t \in I$ we can define the evaluation map $e_t : C(I,X) \to X$ 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 $\gamma_t = x_t$ and
\[ d(\gamma_s, \gamma_t) = |s - t| d(\gamma_0, \gamma_1). \]
In this case, $\gamma$ is called constant speed geodesic. The space of all constant speed geodesics $\gamma : [0, 1] \to X$ will be donated by $\text{Geo}(X)$. Using the triangle inequality it is not difficult to show the following:
Lemma 1.1. Assume $\gamma : [0, 1] \to X$ is a curve such that
\[ d(\gamma_s, \gamma_t) \leq |t - s|d(\gamma_0, \gamma_1) \]
then $\gamma$ is a geodesic from $\gamma_0$ to $\gamma_1$.

A weaker concept is the concept of 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 1.2. If $(X, d)$ is locallly compact then so is $(\text{Geo}(X), d^*)$ where $d^*$ is the sup-distance on $C(I, X)$.

Lipschitz constants and upper gradients. Given a function $f : X \to \mathbb{R} = [-\infty, \infty]$, 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 $x \in D(f) = \{ x \in X \mid f(x) \in \mathbb{R} \}$, otherwise $|Df|(x) = \infty$. The one-sided versions $|D^+ f|$ and $|D^- f|$, also called ascending slope (resp. descending slope)
\[ |D^+ f|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|_+}{d(y, x)} \]
\[ |D^- f|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|_-}{d(y, x)} \]
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.

We say that $g : X \to [0, \infty]$ is an upper gradient of $f : X \to \mathbb{R}$ 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
\[ |f(\gamma_1) - f(\gamma_0)| \leq \int_0^1 g(\gamma_t) dt. \]
It is not difficult to see that the local Lipschitz constant and the two slopes are upper gradients in case $f$ is (locally) Lipschitz.

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 $(\pi_1)_* \pi = \mu_0$ and $(\pi_2)_* \pi = \mu_1$ with $\pi_i$ being the projections to the $i$-th coordinate.

It is well-known that problem has a solution $\pi_{\text{opt}}$, i.e. a probability measure $\pi_{\text{opt}}$ in $\Pi(\mu_0, \mu_1)$ such that
\[ C(\mu_0, \mu_1) = \int c(x, y) d\pi_{\text{opt}}(x, y). \]
Given any such cost function one can define a dual problem
\[
\mathcal{C}(\mu_0, \mu_1) = \sup_{\phi(x) + \psi(y) \leq c(x,y)} \int \phi d\mu_0 + \int \psi d\mu_1.
\]
It is not difficult to see that \(\mathcal{C} \leq C\).

The solution to this problem can be described by a pair of \(c\)-concave potentials: if \(\psi : M \to \mathbb{R}\) then every bounded set in \(\pi\) optimal transport measure \(\phi\) with \(\psi\) \(\pi\) is \(c\)-concave if it is the \(\bar{c}\)-transform of a function \(\phi\).

Given a \(c\)-concave function \(\phi = \psi^c\) one can define the \(c\)-subdifferential \(\partial^c\phi\) by
\[
\partial^c\phi(x) = \{ y \in M \mid \phi(x) + \psi(y) = c(x,y) \}.
\]

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

**Theorem 1.3.** [Vil09, Theorem 5.11] One always has
\[
\mathcal{C}(\mu_0, \mu_1) = C(\mu_0, \mu_1)
\]
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 \(\pi_{\text{opt}}\) is supported on the graph of the \(c\)-subdifferential 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 \(\partial^c\phi(\cdot)\) is single-valued \(\mu_0\)-almost everywhere, then \(\pi_{\text{opt}}\) 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..

**\(p\)-Wasserstein spaces.** The \(p\)-Wasserstein space for \(1 < p < \infty\) is the space of all probability measures with finite \(p\)-Moments
\[
\mathcal{P}_p(M) = \{ \mu \in \mathcal{P}(M) \mid \int d^p(x, x_0) d\mu(x) < \infty \}
\]
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 known that \((\mathcal{P}_p(M), w_p)\) is a complete metric measure space if \((M, d)\) is, and 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 1.4** (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 \(\mathcal{P}_p(M) \subset \mathcal{P}(M)\).
Furthermore, if \( M \) is a geodesic space then so is \( \mathcal{P}_B(M) \) (see [Lis06]).

In the appendix, 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.

**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].

**Finsler structures.** Let \( M \) be a connected, \( n \)-dimensional \( C^\infty \)-manifold.

**Definition 1.5** (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 \( TM \setminus \{0\} \) where \( 0 \) 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 \((x^i)^n_{i=1}\) 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 \( v \in \pi^{-1}(U) \setminus 0 \) where \( \pi : TM \to M \) is the natural projection of the tangent bundle.

Strictly speaking, this is nothing more than defining a Minkowski norm \( F|_{T_x M} \) on each \( T_x M \) 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. \( \bar{F}(v) = F(v) \).

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 \( v = \mathcal{L}_x(\alpha) \in T_x M \) such that \( F(v) = F^*(\alpha) \) and \( \alpha(v) = F(v)^2 \), where \( F^* \) is the dual norm of \( F \) on \( T^*M \). This transform is \( C^\infty \) from \( T^*M \setminus \{0\} \) to \( TM \setminus \{0\} \) and is \( C^\infty \) in case \( F \) is a Riemannian structure, i.e. the parallelogram inequality holds on each \( T_x M \). The gradient \( \nabla f \) at \( x \) is now defined by \( \nabla f(x) = \mathcal{L}_x(df_x) \in T_x M \). Then we have for every unit speed \( C^1 \)-curve \( \eta : [0, l] \to M \) (i.e. \( F(d\eta/dt) \equiv 1 \))

\[
-\int_0^l F(\nabla(-f)(\eta))dt \leq f(\eta(l)) - f(\eta(0)) \leq \int_0^l F(\nabla f(\eta))dt.
\]

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

\[
d(x, y) = \sup_{f \in C^1, F(\nabla f) \leq 1} f(y) - f(x)
\]

which is symmetric iff \( F = \bar{F} \).

Similar to the gradient, there is no notion of (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

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

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

$$g_v\left(\sum_{i=1}^n w_i^i \frac{\partial}{\partial x^i}, \sum_{j=1}^n w_j^j \frac{\partial}{\partial x^j}\right) = \sum_{i,j=1}^n g_{ij}(v) w_i^i w_j^j.$$  

Using the definition of Legendre transform one sees that $L_x^{-1}(v)(w) = g_v(v, w)$ 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 $v \in TM\backslash\{0\}$).

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

Further, we can define the formal Christoffel symbol by

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

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

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

where $(g^{ij})$ is the inverse of $(g_{ij})$ and $A_j^k := \sum_l g^{ij} A_{lj}^k$.

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

$$\omega_j^i = \sum_k \Gamma_j^{ik} dx^k.$$  

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

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

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 connection two-forms

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

which can be also written as
\[
\Omega^j_i(v) = \frac{1}{2} \sum_{k,l=1}^{n} R^i_{jkl}(v) dx^k \wedge dx^l + \frac{1}{F(v)} \sum_{k,l=1}^{n} P^i_{jkl}(v) dx^x \wedge \delta v^l
\]
where we require \( R^i_{jkl} = -R^i_{jlk} \) and \( \delta v^k = dv^k + \sum_l N^k_l dx^l \).

With the help of \( R^i_{jkl} \) one can define the Riemannian tensor with reference vector \( v \in TM \)
\[
R^i(v,w,v,w) := \sum_{i,j,k,l=1}^{n} v^j R^i_{jkl}(v) w^k v^l \partial \partial x^i | x
\]
which enjoys the following
\[
g_v(R^i(v,w,v,w)) = g_v(R^i(w',v,v)) \text{ and } R^i(v,v) = 0.
\]

Given all those definition we finally have the flag curvature
\[
\mathcal{K}(v,w) := \frac{g_v(R^i(v,w,v,w))}{g_v(v,v)g_v(w,w) - g_v(v,w)^2}
\]
and the Ricci curvature
\[
\text{Ric}(v) := \sum_{i=1}^{n-1} \mathcal{K}(v,e_i)
\]
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 \( (M,F,\mu) \) has Ricci curvature bounded from below if \( \text{Ric}(v) \geq K \) for every unit vector \( v \in TM \). For weighted manifolds we need the following: Let \( \mu \) be the reference measure and \( \text{vol}_{g_v} \) be the Lebesgue measure on \( T_xM \) induced by \( g_v \). If \( \mu_x \) denotes the measure \( T_xM \) induced by \( \mu \) define
\[
\mathcal{V}(v) := \log \left( \frac{\text{vol}_{g_v}(B^+_{T_xM}(0,1))}{\mu_x(B^+_{T_xM}(0,1))} \right)
\]
where \( B^+_{T_xM}(0,1) \) denotes the (forward) unit ball of radius 1 w.r.t. the norm \( F|_{T_xM} \).

Further, let
\[
\partial_v \mathcal{V} := \left. \frac{d}{dt} \mathcal{V}(\eta(t)) \right|_{t=0}, \partial^2_v \mathcal{V} := \left. \frac{d}{dt} \partial_v \mathcal{V}(\eta(t)) \right|_{t=0}
\]
where \( \eta : (-\epsilon, \epsilon) \to M \) is a geodesic with \( \eta(0) = v \).

**Definition 1.6 (Weighted Ricci curvature).** Define the following objects:

1. \( \text{Ric}_n(v) := \begin{cases} \text{Ric}(v) + \partial^2_v \mathcal{V} & \text{if } \partial_v \mathcal{V} = 0 \\ -\infty & \text{otherwise} \end{cases} \)
2. \( \text{Ric}_N(v) := \text{Ric}(v) + \partial^2_v \mathcal{V} + \frac{\partial \mathcal{V}}{N} \) for \( N \in (n, \infty) \).
3. \( \text{Ric}_{\infty}(v) := \text{Ric}(v) + \partial^2_v \mathcal{V} \)

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 $v \in TM$.

**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 $D^\dot{\eta}_t \dot{\eta} = 0$ on $(0, r)$. Note however that the reverse curve $\overline{\eta}_t = \eta(t-r)$ 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 $\eta_0 = v$. 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 defined on all of $TM$, i.e. if we can extend any constant speed geodesic $\eta$ to geodesic $\eta : [0, \infty) \to M$. For such a 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 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 $\text{Cut}(x)$ the set of all cut points of $x$, also called the cut locus of $x$. One can show that the exponential map is a $C^\infty$-diffeomorphism from $\{tv \mid v \in T_xM, F(v) = 1, t \in (0, r(v))\}$ to $M \setminus (\text{Cut}(x) \cup \{x\})$. 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 0 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 $\sigma(t, 0) = \eta(t)$. We abbreviate the derivatives as $T(t, s) = \partial_\tau \sigma(t, s), U(t, s) = \partial_s \sigma(t, s)$.

The first variation of the arclength is given by
\[ \frac{\partial \mathcal{L}(\sigma_s)}{\partial s} = \left[ \frac{g_{\tau T}(U, T)}{F(T)} \right]_{t=0}^r - \int_0^r g_T \left( U, D^\tau T \left[ \frac{T}{F(T)} \right] \right) \, dt. \]
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
\[ \frac{\partial^2 \mathcal{L}(\sigma_s)}{\partial s^2} \bigg|_{s=0} = I(U, U) + \left[ \frac{g_{\tau T}(U, T)}{F(T)} \right]_{t=0}^r - \int_0^r \frac{1}{F(T)} \left( \frac{\partial F(T)}{\partial s} \right)^2 \, dt \]
where
\[ I(V, W) := \frac{1}{F(\dot{\eta})} \int_0^r \left\{ g_{\dot{\eta}}(D^\dot{\eta}_t V, D^\dot{\eta}_t W) - g_{\dot{\eta}}(R^\dot{\eta}(V, \dot{\eta}) \dot{\eta}, W) \right\} \, dt. \]
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
\[ I(V, W) = \frac{1}{F(\dot{\eta})} \left[ g_{\dot{\eta}}(D^\dot{\eta}_t V, W) \right]_{t=0}^r. \]
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
\[ D^2_\eta D^2_\eta J + R^0(J, \eta)\eta = 0.\]
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.

2. $c_p$-CONCAVE FUNCTIONS

Assume throughout $M$ 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 $\phi : X \to \mathbb{R}$ 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 $c_L$ will be an abbreviation for $c_{L}$, 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 2.1** ($c_p$-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 $\phi : X \to \mathbb{R}$ 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 $\psi : Y \to \mathbb{R}$ as
\[ \psi^{\bar{c}_p}(x) = \inf_{y \in Y} c_p(x, y) - \psi(y).\]

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

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 $\psi : Y' \to \mathbb{R}$ is such that $\phi = (\psi')^{c_p}$ then let
\[ \psi(y) = \begin{cases} 
\psi'(y) & \text{if } y \in Y' \\
-\infty & \text{if } y \in Y \setminus Y'.
\end{cases}\]

Then obviously $\phi = (\psi')^{c_p} = \psi^{c_p}$ 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 $\psi : Y \to \mathbb{R}$ relative to $(Y, X)$.

The following is easy to show:

**Lemma 2.2.** 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) $\phi \leq \phi^c_p \phi$ and $\phi^c_p \phi = \phi^c_p \phi^c_p$

(2) If $\phi$ is not identically $-\infty$ then $\phi$ is $c_p$-concave iff $\phi = \phi^c_p \phi$

(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 2.3. 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 $c_p/c_p$-transform is actually achieved and the following sets are non-empty for each $c_p/c_p$-concave functions.

Definition 2.4 ($c_p$-subdifferential). Let $X$ and $Y$ be two compact subsets of $M$ and $\phi : X \rightarrow \mathbb{R}$ be a $c_p$-concave function relative to $(X, Y)$ then the $c_p$-subdifferential 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 $c_p$-subdifferential of a $c_p$-concave function $\psi : Y \rightarrow \mathbb{R}$ as the non-empty set

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

It is not difficult to see that

$$y \in \partial^c_p \phi(x) \iff x \in \partial^c_p \phi^c_p(y)$$

whenever $\phi$ is $c_p$-concave. Furthermore, $y \in \partial^c_p \phi^c_p(y)$.

Lemma 2.5 (Semicontinuity of the $c_p$-subdifferential). 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 $(x_n, y_n) \rightarrow (x, y)$, 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

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

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 2.6. 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

$$L(R) - L(R - \epsilon) \rightarrow \infty.$$
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 \( \phi = (\phi^{\alpha})^\bullet \) 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

\[
\phi(x_n) \geq c_p(x_n, y_n) - \phi^{\alpha}(y_n) - 1.
\]

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

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

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

\[
\sup_{\bar{B}_1(\gamma^n_1)} \phi \to -\infty \text{ as } n \to \infty.
\]

In order to prove this, note that for \( x \in \bar{B}_1(\gamma^n_1) \) we have \( d(x, \gamma^n_1) \leq 1 = d(x_n, \gamma^n_1) \) and thus

\[
\phi(x) \leq c_p(x, y_n) - \phi^{\alpha}(y_n) \leq \frac{(d(x, \gamma^n_1) + d(\gamma^n_1, y_n))^p}{p} - \phi^{\alpha}(y_n)
\]

\[
\leq \frac{(d(x_n, \gamma^n_1) + d(\gamma^n_1, y_n))^p}{p} - \phi^{\alpha}(y_n)
\]

\[
= c_p(x_n, y_n) - \phi^{\alpha}(y_n) \leq \phi(x_n) + 1.
\]

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

Since \( M \) is proper, we can assume \( \gamma^n_1 \to z \) 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

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

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

\[
\lim_{n \to \infty} \sup \phi(x) - \phi(\gamma^n_1) \geq \lim_{n \to \infty} \sup \phi^{\alpha}.c_p(x, y_n) - c_p(\gamma^n_1, y_n)
\]

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

\[
c_p(x, y_n) - c_p(\gamma^n_1, y_n) = \int_{R_n-r}^{R_n} l(s)ds \geq r \cdot l(R_n - r) \to \infty
\]

which is a contradiction. Hence \( y_n \) is bounded and by properness has accumulation points which all belong to \( \partial^{\alpha}\phi(x) \). Similarly, we can show that \( \cup_{x \in K} \partial^{\alpha}\phi(x) \) is bounded for any compact \( K \).
Finally, for all $x \in B_r(\bar{x})$
\begin{align*}
\phi(x) &= \inf_{y \in M} c_p(x, y) - \phi^p(y) \\
&= \min_{B_C(\bar{x})} c_p(x, y) - \phi^p(y).
\end{align*}
Since for $y \in B_C(\bar{x})$ the functions $x \mapsto c_p(x, y) - \phi^p(y)$ are uniformly Lipschitz on $B_r(x)$, $\phi$ 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 3.14 below). It generalizes [CEMS01, Claim 2.4] and will be used in Lemma 2.8 (see [Oht09, (3.1) p. 221] for the case $p = 2$). Lemma 2.9 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 A.7).

Lemma 2.7. If $x,y \in M$ and $z \in Z_t(x,y)$ for some $t \in [0,1]$. Then for all $m \in M$
\[ ip^{-1}d_p(m,y) \leq d_p(m,z) + tp^{-1}(1-t)d_p(x,y). \]
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{align*}
    tp^{-1}d_p(m,y) &\leq \left\{ t \cdot \frac{1}{t}d(m,z) + (1-t)d(x,y) \right\}^p \\
    &\leq \left\{ t \left( \frac{1}{t}d(m,z) \right) + (1-t)d_p(x,y) \right\}^p \\
    &= d_p(m,z) + tp^{-1}(1-t)d_p(x,y).
\end{align*}
Furthermore, choosing $m = x$ we see that each inequality is actually an equality.

Lemma 2.8. 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 $h(t) := f_t(m) - tp^{-1}f_1(m)$ has a minimum at $x$.
Proof. Using Proposition 2.7 above we have for \( z = \eta_t \in Z_t(x, y) \)
\[
-\phi_t(m) = t^{p-1}d^p(m, y) - d^p(m, z) \leq t^{p-1}(1-t)d^p(x, y) = t^{p-1}d^p(x, y) - d^p(x, \eta_t) = -\phi_t(x).
\]

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

**Lemma 2.9** \((c_p\text{-concave functions form a star-shaped set})\). 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 \( \phi_t(x) := c_p(x, y) = d^p(x)/p \). We claim that the following representation holds
\[
t^{p-1}d^p_y(m)/p = \inf_{z \in Z_t(x,y)} \left\{ d^p_z(m)/p + \inf_{x \in X, z \in Z_t(x,y)} t^{p-1}(1-t)d^p_y(x)/p \right\}.
\]

Indeed, by Lemma 2.7 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 \( c_{p\phi} \)-transform of the function
\[
\psi(z) = -\inf_{x \in X, z \in Z_t(x,y)} t^{p-1}(1-t)d^p_y(x)/p
\]
(real-valued on \( Z_t(X,Y) \)) and therefore \( t^{p-1}\phi \) is \( c_{p\phi} \)-concave relative to \( (X, Z_t(x,y)) \).

Since \( \mathcal{I}^{c_p}(X, Z_t(x,y)) \subset \mathcal{I}^{c_p}(X, X) \) we see that each \( t^{p-1}d^p_y/p \) 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 \( c_{p\phi} \)-concave relative to \( (X, Z_t(X,Y)) \). Since \( \phi = \phi^{c_p}_{\phi} \) 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 2.10** \((\text{non-branching spaces})\). 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
\[
Z_t(x, y) \cap Z_t(x, y') \neq \emptyset \text{ for some } t \in (0, 1) \implies y = y'.
\]

**Lemma 2.11.** 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 for \( (x_i, y_i) \in U \), 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 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 $(x_i, y_i) \in U$

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

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

$$d(x_1, y_2)^p + d(x_2, y_1)^p \leq (d(x_1, z) + d(z, y_2))^p + (d(x_2, z) + d(z, y_1))^p$$

$$= (td(x_1, y_1) + (1 - t)d(x_2, y_2))^p$$

$$+ (td(x_2, y_2) + (1 - t)d(x_1, y_1))^p$$

$$\leq td(x_2, y_2)^p + (1 - t)d(x_2, y_2)^p$$

$$+ td(x_2, y_2)^p + (1 - t)d(x_1, y_1)^p$$

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

Because $U$ is $c_p$-cyclically monotone we see that the inequality actually must be an equality. Since $p > 1$ we must have

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

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 $y_1 = y_2$. 

\[\square\]

3. Interpolation in the Smooth Setting

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 get a lower (weighted) Ricci curvature bound from the curvature dimension condition.

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 $v \in T_x M$

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

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
\[
\nabla^q(t^{-1}\phi) = t\nabla^q \phi.
\]
In addition, we use the abbreviation $Kd\phi = \nabla^q \phi$ (note that $Ld\phi = \nabla \phi$). This is indeed invertible, continuous from $T^*M \to TM$ and $C^\infty$ away from the zero section. Furthermore,
\[
K_x t^{-1}d\phi_x = t\nabla^q \phi(x).
\]

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

Remark. $K_x$ can actually be seen as the Legendre transform from $T^*_xM \to T^*_xM$ that associates to each cotangent vector $\alpha \in T^*_xM$ the unique tangent vector $v = K(\alpha) \in TM$ 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 3.1** (Cut locus characterization). 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 $y_\epsilon \notin \text{Cut}(x) \cup \{x\}$ and using the first variation formula we get for $t > 0$
\[
f(\xi_v(-t)) - f(x) \leq \{d(\xi_v(-t), y_\epsilon) + \epsilon\}^p / p - \{d(x, y_\epsilon) + \epsilon\}^p / p = t \{d(x, y_\epsilon) + \epsilon\}^{p-1} g_{\eta(0)}(v, \dot{\eta}(0)) + \mathcal{O}(t^2)
\]
\[
= td^{p-1}(x, y) g_{\eta(0)}(v, \dot{\eta}(0)) + \mathcal{O}(t^2).
\]
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 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 $v = D^q_\eta f(0) \in T^*_xM \setminus \{0\}$ let $V_1$ be the parallel vector field along $\eta$ (i.e. $D^q_\eta V_1 \equiv 0$) 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 $g_{\eta(0)}(v, v) > 0$ 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 $\sigma(t,s) = \sigma_s(t) := \xi_{J_s(t)}(s)$. Because $J_s \neq 0$ 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])

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

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

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

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

$$T_{\eta(0)}(v) = g_\eta(D^2_{\eta} v - D^0_{\eta} v, \dot{\eta}) = \epsilon^{-2} g_{\eta(0)}(D^2_{J_s} J_s - D^0_{J_s} J_s, \dot{\eta}) = -\epsilon^{-2} g_{\eta(0)}(D^0_{J_s} J_s, \dot{\eta})$$

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

$$\frac{\partial^2 \mathcal{L}(\sigma_s)}{\partial s^2} \bigg|_{s=0} \leq I(J,J) + 2\epsilon I(J,V) + \epsilon^2 I(V,V) + \epsilon^2 T_{\eta(0)}(v)/d(x,y)$$

$$= \left\{ \left[ g_\eta(D^0_{\eta} J_s, J_s) \right]_{t=0}^{1} + 2\epsilon \left[ g_\eta(D^2_{\eta} J_s J_s, \dot{\eta}) \right]_{t=0}^{1} + \epsilon^2 T_{\eta(0)}(v) \right\} /d(x,y) + \epsilon^2 I(V,V).$$

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

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

So that we get

$$\lim_{s \to 0} \frac{\mathcal{L}(\sigma_s)^p + \mathcal{L}(\sigma_{-s})^p - 2\mathcal{L}(\sigma_0)^p}{s^2} = p\mathcal{L}^{p-2}(\sigma_0) \left[ \mathcal{L}(\sigma_0) \frac{\partial^2 \mathcal{L}(\sigma_s)}{\partial s^2} \bigg|_{s=0} \right]$$

$$+ (p-1) \left( \frac{\partial \mathcal{L}(\sigma_s)}{\partial s} \bigg|_{s=0} \right)^2$$

$$\leq p\mathcal{L}^{p-2}(x,y) \left\{ -2\epsilon g_\eta(v, v) + \epsilon^2 \left[ T_{\eta(0)}(v) + d(x,y)I(V,V) + (p-1)F(v)^2 \right] \right\}.$$ 

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

$$\lim_{s \to 0} \frac{f(\xi_\epsilon(\epsilon s)) + f(\xi_\epsilon(-\epsilon s)) - 2f(x)}{\epsilon^2 s^2} \leq \lim_{s \to 0} \frac{\mathcal{L}(\sigma_s)^p + \mathcal{L}(\sigma_{-s})^p - 2\mathcal{L}(\sigma_0)^p}{p\epsilon^2 s^2}$$

$$\leq \mathcal{L}^{p-2}(x,y) \left\{ -2\epsilon^{-1} g_\eta(v, v) + T(v) + d(x,y)I(V,V) + (p-1)F(v)^2 \right\}.$$ 

Letting $\epsilon$ tend to zero completes the proof. □
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 3.2. Let $\phi : M \to \mathbb{R}$ be a $c_p$-concave function. If $\phi$ is differentiable at $x$ then $\partial^\phi \phi(x) = \{\exp_x(\nabla q(-\phi)(x))\}$. 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 $y \in \partial^\phi \phi(x)$ be arbitrary and define $f(z) := c_p(z, y) = d^p(z, y)/p$. By definition of $\partial^\phi \phi(x)$ we have for any $v \in T_xM$

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

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

$$f(\exp_x v) - f(x) \leq \frac{1}{p} \{(d(\exp_x v, y_\epsilon) + \epsilon)^p - (d(x, y_\epsilon) + \epsilon)^p\}$$

$$= - (d(x, y_\epsilon) + \epsilon)^{p-1} g_{\eta(0)}(v, \dot{\eta}(0)) + o(F(v)).$$

Therefore, $d\phi_x(v) \leq -d^{p-1}(x, y)\nabla_x^{-1}(\dot{\eta}(0))(v)$ 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. \hfill \□

Let $\text{Lip}_{c_p}(X, Y)$ 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 3.3. 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 $(\phi, \psi) \in \text{Lip}_{c_p}(X, Y)$ then $(\phi^\phi + \phi^\psi) \in \text{Lip}_{c_p}(X, Y)$ and $\phi^\phi \geq \psi$ and $\phi^\psi \geq \phi$.

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

It remains to show that this solution is unique: Let $(\phi_1, \psi_1), (\phi_2, \psi_2) \in \text{Lip}_{c_p}(X, Y)$ be two solutions of the problem. Now setting $\phi = (\phi_1 + \phi_2)/2$, we see that
\[ \phi \geq \frac{1}{2}(\phi_{r}^1 + \phi_{r}^2) \] and thus \((\phi, \phi r) \in \text{Lip}_r (X, Y)\) and hence, by maximality, \(\phi = \frac{1}{2}(\phi_{r}^1 + \phi_{r}^2)\) and \(\phi\) is \(c_r\)-concave.

Now if \(y \in \partial \phi\phi(x)\) then \(y \in \partial \phi\phi_{1}(x) \cap \partial \phi\phi_{2}(x)\). Thus, using Lemma 3.2 above and the absolute continuity of \(\mu_0\) we see that

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

\[ \square \]

**Theorem 3.4.** 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 \(c_r\)-concave function \(\phi\) such that \(\pi = (\text{Id} \times \mathcal{F})_*\mu_0\) 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 to 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 \(\psi \epsilon = \phi_{r} + c \epsilon\) for \(\epsilon \in \mathbb{R}\) close to 0. Let \(x \in M\) be arbitrary, then we can find \(y_\epsilon \in M\) such that

\[
(\psi \epsilon)(x) = c_p(x, y_\epsilon) - \psi \epsilon(y_\epsilon).
\]

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

\[
\phi(x) - \epsilon h(y_\epsilon) \leq c_p(x, y_\epsilon) - \phi_{r} - \epsilon h(y_\epsilon) = (\psi \epsilon)(x)
\]

and thus

\[
(\psi \epsilon)(x) = \phi(x) - \epsilon h(\mathcal{F}(x)) + o(\epsilon) \quad \text{and} \quad |o(\epsilon)| \leq 2\epsilon \|h\|_{\infty}.
\]

Now set \(J(\epsilon) = \int (\psi \epsilon)^{\phi_{r}} d\mu + \int \psi \epsilon d\nu\) and by maximality of \((\phi, \phi_{r})\) we have

\[
0 = \lim_{\epsilon \to 0} \frac{J(\epsilon) - J(0)}{\epsilon} = -\int h d\mathcal{F}_\mu + \int h d\mu_1
\]

and hence \(\mathcal{F}_\mu \mu_0 = \mu_1\).

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

**Corollary 3.5.** If \(\mu_0\) is absolutely continuous and \(\phi\) is \(c_r\)-concave, then the map \(\mathcal{F}(x) := \exp_x (\nabla^q (-\phi))\) is the unique optimal transport map from \(\mu_0\) to \(\mathcal{F}_\mu \mu_0\).

Furthermore, we will see in Lemma 3.14 below that the interpolation measures are absolutely continuous if \(\mu_0\) and \(\mathcal{F}_\mu \mu_0\) are.

**Corollary 3.6.** If \(\phi\) is \(c_r\)-concave and \(\mu_0\) is absolutely continuous, then the map \(\mathcal{F}_t(x) := \exp_x (\nabla^q (-t^{\phi_{r}} - \phi))\) is the unique optimal transport map from \(\mu_0\) to \(\mu_t = (\mathcal{F}_t)_\mu \mu_0\) 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
\[ w_p(\mu_s, \mu_t) \leq |s - t| w_p(\mu_0, \mu_1). \]
Let \( \pi \) be the plan on \( \text{Geo}(M) = \{ \gamma : [0, 1] \to M \mid \gamma \) is a geodesic in \( M \} \) give by \( \mu_0 \), the map \( F \) and the unique geodesic connecting \( \mu \)-almost every \( x \in M \) to a point \( F_t(x) \) (see e.g. [Lis06, Theorem 4.2] and [Vil09, Chapter 7]), in particular, \( \mu_t = (F_t)_* \mu_0 \). Since \( (c_s, c_t)_* \pi \) is a plan between \( \mu_s \) and \( \mu_t \) for \( s, t \in [0, 1] \), we have
\[
w_p(\mu_s, \mu_t) \leq \left( \int d^p(\gamma_s, \gamma_t) d\pi(\gamma) \right)^{1/p} = (s - t) \left( \int d^p(\gamma_0, \gamma_1) d\pi(\gamma) \right)^{1/p} = |s - t| w_p(\mu_0, \mu_1).
\]

\[ \square \]

Almost Semiconcavity of \( c_p \)-concave functions. This section will be one of the main ingredients to show Theorem 3.13. In [Oht08] Ohta showed that every \( c_p \)-concave function is almost everywhere second order differentiable. He proved this by showing the the square of the distance function \( d^2_\mu = d^2(\cdot, y) \) 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 3.7 (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 second order 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 second order differentiable.

Instead of following the arguments in [Oht08] (which is done in the author’s thesis), we give a new, shorter proof using star-shapedness of the \( c_p \)-concave functions (Lemma 2.9).

The proof for the following: If the Finsler metric \( F \) is \( C^\infty \) then the function \( d^p_\mu(z) = d(z, y)^p \) is \( C^\infty \) in \( U_y \setminus \{y\} \) for some sufficiently small neighborhood \( U_y \) of \( y \). This follows from smoothness of the exponential map \( \exp_y \) in \( V \setminus \{0\} \subset T_y M \) for some neighborhood \( V \) of \( 0 \in T_x M \), see [She97, p. 315]. In particular, for \( x \in U_y \setminus \{y\} \) 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 \( \{d^p_\mu : U_1 \to \mathbb{R}\}_{\gamma_0 \in V_1} \) 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 A.9 is used instead of Lemma 2.9.

Theorem 3.8. Let \( \phi \) be a \( c_p \)-concave function. Let \( \Omega_{id} \) be the the points \( x \in M \) where \( \phi \) is differentiable and \( d\phi_x = 0 \), or equivalently \( d^{p+} \phi(x) = \{x\} \). Then \( \phi \) is locally semiconcave on an open subset \( U \subset M \setminus \Omega_{id} \) of full measure (relative to \( M \setminus \Omega_{id} \)). In particular, it is second order differentiable almost everywhere in \( U \).
Proof. Since $\partial^\nu \phi(x)$ 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 $x \in \text{int}(M \setminus \Omega_{id})$. 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 2.9 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 $\partial^\nu \phi(x) = \{y\}$ and a unique geodesic $\eta : [0, 1] \to M$ between $x$ and $y$ (see Lemma 3.2). Also note that $\phi_s$ is differentiable at $x$ and

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

Let $s \in [0, 1]$ be such that $d(x, \eta(s)) < \frac{\epsilon}{\min \lambda}$. Because $x \neq \eta(s)$ and $z \mapsto \partial^\nu \phi_s(z)$ is continuous and single-valued at $x$, we can find a neighborhood $V_1 \subset U$ of $x$ such that $(\partial^\nu \phi_s)^{-1}(V_1) \cap U$ contains some ball $B_{2r}(x)$ disjoint from $V_1$ and thus the functions $\{d^p_y : B_2r(x) \to \mathbb{R}\}_{y \in V_1}$ are semiconcave with constant $C$.

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

$$\phi_s(x_0) \leq \phi_s(x_t) + \frac{1}{p} d^p(x_0, y_t) - \frac{1}{p} d^p(x_t, y_t),$$

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

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

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

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

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

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

$$D|Q_t \geq (1-t)D|Q_0 + tD|Q_1.$$  

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

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

Remark. In the symmetric setting one has $v^>_t(x, y) = v^<_t(1-t)(x, y)$.  

Theorem 3.10 (Volume distortion for $d^2$ [Oht09, 3.2]). For point $x \neq y \in M$ with $y \notin \text{Cut}(x)$, let $\eta : [0,1] \rightarrow M$ 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

\[
\psi_t^>(x,y) = D \left[ d(\exp_x)\nabla_{g_t(x)} \circ [d(\exp_x)\nabla_{g_t(x)}]^{-1} \right]
\]

\[
\psi_t^<(x,y) = (1-t)^{-n}D \left[ d(\exp_x \circ L_x)_{d(g_1)_x} \circ [d (d(tg_1))_x - d (dg_1)_x] \right].
\]

Proof. The first equation follows from the fact that

\[
\nabla^g f_t(x) = \nabla \left( -d(x,\eta(t))^2/2 \right).
\]

For the second part note that

\[
L_z(d(tg_1)_z) = K_z(d(t^{p-1}f_1)_z)
\]

and thus

\[
\psi_t^>(x,y) = (1-t)^{-n}D \left[ d(\exp \circ L \circ (d(tg_1)_z)) \right]
\]

\[
= (1-t)^{-n}D \left[ d(\exp \circ K)_{d(t^{p-1}f_1)_x} \circ d (d(t^{p-1}f_1))_x \right].
\]

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

\[
d(\exp_x \circ K_z)_x \circ d(df_1)_x = 0.
\]

Note that

\[
L_z(d(g_1)_z) = K_z(d(f_1)_z)
\]

and thus

\[
L(z) = \exp_z \circ K_z(d(f_1)_z) = \exp_z \circ L_z(d(g_1)_z)
\]

\[
= \eta(t).
\]

Which immediately implies $dL = 0$. \qed

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

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

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

in any local coordinate system $(x^i)_{i=1}^n$ around $x$.\hfill
(3) Define \( f_y(z) := -c_p(z, y) \) and

\[
dF_x := d(\exp \circ K_x)_d(-\phi)_x \circ [d(d(-\phi))_x - d(df_y)_x] : T_xM \to T_yM
\]

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 \( v \in T_xM \)

\[
\sup \{|u - dF_x(v)| \mid exp_y u \in \partial^{\text{v}} \phi(exp_x y), |u| = d(y, exp_y u)\} = o(|v|).
\]

Proof. As \( \phi \) is differentiable at \( x \) we have \( \partial^{\text{v}} \phi(x) = \{y\} \) 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 \( \phi \)

\[
h(x) = c_p(x, y) - \phi(x) = \partial^{\text{v}} \phi(y) \leq c_p(\xi_t(\pm t), y) - \phi(\xi_t(\pm t)) = h(\xi_t(\pm t))
\]

where \( \xi_t : (-\varepsilon, \varepsilon) \to M \) is a geodesic with \( \dot{\xi}_t(0) = v \). Thus, we have

\[
\frac{\phi(\xi_t(t)) + \phi(\xi_t(-t)) - 2\phi(x)}{t^2} \leq \frac{f_y(\xi_t(t)) + f_y(\xi_t(-t)) - 2f_y(x)}{t^2}.
\]

Since \( \phi \) is second order differentiable at \( x \) we have

\[
-\infty < \frac{\partial^2(\phi \circ \xi_t)}{\partial t^2}(0) = \lim_{t \to 0^+} \frac{f_y(\xi_t(t)) + f_y(\xi_t(-t)) - 2f_y(x)}{t^2}
\]

and hence \( y \) is not a cut point of \( x \) (Lemma 3.1).

Now the second statement follows immediately from the inequality above and the fact that \( y \notin \text{Cut}(x) \cup \{x\} \) 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 \( d(f_y)_x = \alpha d\phi_x \) and thus the difference \( d(d(-\phi))_x - d(df_y)_x \) 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 \( y_t := \exp_y u_t \in \partial^{\text{v}} \phi(x_t) \) and \( d(y, y_t) = F(u_t) \). In addition, we have

\[
-\phi(\exp_x w) \geq -\phi(x_t) - f_{y_t}(x_t) + f(\exp_x w) = -\phi(x_t) + d(f_{y_t}, x_t)(w) + o(F(w))
\]

for \( w \in T_xM \). Differentiating \( y_t = \exp \circ K(d(f_{y_t})_x)_t \) at \( t = 0 \) we get

\[
\frac{\partial y_t}{\partial t} \big|_{t=0} = d(\exp \circ K)(d(-\phi)_x) \circ d(d(-\phi)_x)(v).
\]

Moreover, we have \( \exp \circ K(d(f_y)_x)_t = y \) and thus \( d(\exp \circ K)(d(f_y)_x)_t \circ d(d(f_y)_x)(v) = 0 \). Therefore,

\[
\frac{\partial y_t}{\partial t} \big|_{t=0} = d(\exp \circ K)(d(-\phi)_x) \circ [d(-\phi)_x - d(df_y)_x] (v) = dF_x(v).
\]

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

\[\square\]

**Proposition 3.12.** Let \( \mu_0 \) and \( \mu_1 \) be absolutely continuous measure with density \( f_0 \) and \( f_1 \) resp. and assume that there are open set \( U_i \) with compact closure \( X = \overline{U}_0 \) and \( Y = \overline{U}_1 \) such that \( \text{supp} \mu_i \subset U_i \). Let \( \phi \) be the unique \( c_p \)-concave Kantorovich potential and define \( F(z) = \exp_z(\nabla^q (-\phi)(z)) \). Then \( F \) is injective \( \mu_0 \)-almost everywhere and for \( \mu_0 \)-almost every \( x \in M \), \( \Omega_{id} \)
The proof follows without any change from \[ \text{Theorem 5.2}, \] see also \[ \text{[Oht09, Chapter 11]}. \]

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

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

**Theorem 3.13.** Let \( \phi : M \to \mathbb{R} \) be a \( c_p \)-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 \( y_t := \exp_x(\nabla^q(-t^{p-1}\phi)) \), \( f_t(z) = -c_p(z,y_t) \) and \( J_t(x) = D[d\mathcal{F}_x] \) where
\[
d(\mathcal{F}_x) := d(\exp_x \circ K_x)d(-\tau_{t\phi})_x \circ |d(t^{p-1}\phi)|_x - d(d(f_t))_x | T_x M \to T_{y_t} M. \]

Then for any \( t \in (0,1) \)
\[
J_t(x)^{1/n} \geq (1-t)\nu^\phi_\tau(x,y_t)^{1/n} + tv^\phi_\tau(x,y_t)^{1/n} J_1(x)^{1/n}. \]

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

**Proof.** Note first that
\[
d(d(-t^{p-1}\phi)_x - d(df_t)_x = \{d(d(-t^{p-1}\phi))_x - d(t^{p-1}f_1)_x \} + \{d(d(t^{p-1}f_1))_x - d(df_t)_x \}
\]
and
\[
d(f_t)_x = d(-t^{p-1}\phi)_x - d(t^{p-1}f_1)_x.
\]

Now define \( \tau_x : T_x M \to T_x M \) as \( \tau_x(v) = s^{p^{-1}}v \) and note
\[
d(\exp_x \circ K_x)_x d(-\tau_{t\phi})_x \circ (d(d(-t^{p-1}\phi))_x - d(t^{p-1}f_1)_x) = d(\exp_x \circ K_x)_x d(-\tau_{t\phi})_x \circ d(\tau_t)_x d(-\phi)_x \circ [d(d(-\phi))_x - d(df_t)_x] = d(\exp_x \circ K_x)_x d(-\tau_{t\phi})_x \circ [d(\exp_x \circ K_x)_x d(-\phi)_x]^{-1} \circ d(\mathcal{F}_x) = d(\exp_x)_{\nabla^q(-\tau_{t\phi})} \circ K_x \circ \tau_t \circ K_x^{-1} \circ \nabla^q(-\phi)_x \circ [d(\exp_x)_{\nabla^q(-\phi)}]^{-1} \circ d(\mathcal{F}_x) = t \cdot d(\exp_x)_{\nabla^q(-\tau_{t\phi})} \circ [d(\exp_x)_{\nabla^q(-\phi)}]^{-1} \circ d(\mathcal{F}_x),
\]
because \( K_x \circ \tau_t \circ K_x^{-1} \) is linear and, for \( v \in T_x M \)
\[
K_x \circ \tau_t \circ K_x^{-1}(v) = K_x(t^{p-1}K_x^{-1}(v)) = tv,
\]
i.e. \( d(K_x \circ \tau_t \circ K_x^{-1})_{\nabla^q(-\phi)} = t \cdot \text{Id} \). Note that we identified \( T_{\nabla^q(-\tau_{t\phi}(\phi))(\mathcal{T}_x M)} \) with \( T_{\nabla^q(-\phi)(\mathcal{T}_x M)} \) to get the last inequality.
Because $D$ is concave we get
\[
J_t(x)^{1/n} = D[d(F_t)_x]^{1/n}
\]
\[
= D\left[ d(\exp_x \circ K_x)_{d(\cdot^{-t-1} \phi)_x} \circ \left[ d(d(t^{t-1} f_1))_x - d(df_1)_x \right] \\
+ d(\exp_x \circ K_x)_{d(\cdot^{-t-1} \phi)_x} \circ \left( d(d(-t^{t-1} \phi))_x - d(d(t^{t-1} f_1))_x \right) \right]^{1/n}
\]
\[
= D\left[ d(\exp_x \circ K_x)_{d(\cdot^{-t-1} \phi)_x} \circ \left( d(d(t^{t-1} f_1))_x - d(df_1)_x \right) \\
+ t \cdot d(\exp_x \circ K_x)_{d(\cdot^{-t-1} \phi)_x} \circ \left[ d(\exp_x \circ K_x_{d(\cdot^{-t-1} \phi)_x}) \circ (d(\exp_x \circ K_x \circ (\cdot^{-t-1} \phi)_x) - d(df_1)_x) \right] \right]^{1/n}
\]
\[
\geq (1 - t)D\left[ (1 - t)^{-1} d(\exp_x \circ K_x)_{d(\cdot^{-t-1} \phi)_x} \circ \left( d(d(t^{t-1} f_1))_x - d(df_1)_x \right) \right]^{1/n}
\]
\[
+ tD\left[ d(\exp_x \circ K_x)_{d(\cdot^{-t-1} \phi)_x} \circ \left[ d(\exp_x \circ K_x_{d(\cdot^{-t-1} \phi)_x}) \circ (d(\exp_x \circ K_x \circ (\cdot^{-t-1} \phi)_x) - d(df_1)_x) \right] \right]^{1/n}
\]
\[
= (1 - t)\nu_t^\gamma(x, y_1)^{1/n} + t\nu_t^\gamma(x, y_1)^{1/n}J_1(x)^{1/n}.
\]

Combing this with Lemma 2.11 and Lemma 3.15 below we get similar to [Oht09, 6.2]:

**Lemma 3.14.** Given two absolutely continuous measures $\mu_t = \rho_t \mu$ on $M$, let $\phi$ be the unique $c_p$-concave optimal Kantorovich potential. Define $F_t(x) := \exp_x(\nabla x (-t^{-1} \phi))$ for $t \in [0, 1]$. Then $\mu_t = (F_t)_{\ast} \mu_0$ is absolutely continuous for any $t \in [0, 1]$.

**Proof.** By Lemma 2.11 the map $F_t$ 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
\[
\mu_t|_{\Omega_{id}} = (F_t)_{\ast}\mu_0 |_{\Omega_{id}} = \mu_0 |_{\Omega_{id}}.
\]

By Theorem 3.8 the potential $\phi$ is second order differentiable in a subset $\Omega \subset M \setminus \Omega_{id}$ of full measure. In addition, $D[d(F_t)] > 0$ for all $x \in \Omega$ (see Proposition 3.12) and $F_t$ is continuous in $\Omega$ for any $t \in [0, 1]$. The map $d(F_t)_x : T_x M \to T_{F_t(x)} M$ defined in Proposition 3.11 as
\[
d(F_t)_x := d(\exp_x \circ K_x)_{d(\cdot^{-t-1} \phi)_x} \circ \left( d(d(-t^{t-1} \phi))_x - d(df_1)_x \right)
\]
where $f_t(z) := -c_p(z, F_t(x))$ for $t \in (0, 1]$. Also note that for $x \in \Omega$
\[
d\left( d(t^{-t-1} \phi)_x - d(df_1)_x \right) = \left\{ d\left( d(t^{-t-1} \phi)_x - d(df_1)_x \right) \right\} + \left\{ d\left( d(t^{-t-1} \phi)_x - d(df_1)_x \right) \right\}.
\]
Which implies $D[d(F_t)_x] > 0$ because $D[d(F_t)_x] > 0$ and the lemma below.

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

**Lemma 3.15.** Let $y \notin \text{Cut}(x) \cup \{x\}$ 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 $h(z) = t^{t-1} f_1(z) - f_t(z)$ satisfies
\[
\frac{\partial^2 h}{\partial x^i \partial x^j}(x) \geq 0
\]
in any local coordinate system around x.

Proof. This follows directly from 2.8. □

4. Abstract curvature condition

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

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

**Definition 4.1** ($DC_N$). For $N \in [1, \infty]$ let $DC_N$ 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

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

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

**Lemma 4.2** ([LV09, Lemma 5.6]). If $N \leq N'$ then $DC_{N'} \subset DC_N$.

**Example 4.3.** 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 $DC_N$

2. The classical entropy functional $U_{\infty} : x \mapsto x \log x$ is in $DC_\infty$

3. If $m > 1$ then $U_m \in DC_\infty$

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

$$
U_\mu(\nu) = \int U(\rho) d\mu + U'(\infty) \nu_{\infty}(M)
$$

where $\nu = \rho \mu + \mu_\infty$ 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 $U_\mu$. In addition, we use $U_m, U_\alpha$ etc. to denote the functional generated by $U_m, U_\alpha, \alpha$ 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

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

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 $\sigma_i$ instead of $\beta_t$. Because of the localization and tensorization property this weight function turned out to be powerless ([AGS11b, AGMR12, Raj12, Raj11, GM13, EKS13, Gig13, HKX13]). Using the inequalities of the proof of Lemma 2.11 most of the things proven in [BS10] will also hold for localized version $CD^*_p(K, N)$.

**Definition 4.4** (strong $CD_p(K, N)$). We say $(M, d, \mu)$ satisfies the strong $CD_p(K, N)$ condition if the following holds: Given two measure $\mu_0, \mu_1 \in P(M)$ with Lebesgue decomposition $\mu_i = \rho_i \mu + \mu_i$. Then there exists some optimal dynamical transference plan $\Pi \in P(\text{Geo})$ such that $\mu_t = (e_t)\Pi$ is a geodesic from $\mu_0$ to $\mu_1$ in $P_p(M)$ such that for all $U \in D\mathcal{C}_N$ and $t \in [0, 1]$

$$U(\mu_t) \leq (1 - t) \int_{M \times M} \beta_{1-t}(x_0, x_1) U \left( \frac{\rho_0(x_0)}{\beta_{1-t}(x_0, x_1)} \right) d\pi(x_1|x_0) d\mu(x_0) + t \int_{M \times M} \beta_t(x_0, x_1) U \left( \frac{\rho_1(x_1)}{\beta_t(x_0, x_1)} \right) d\pi(x_0|x_1) d\mu(x_1) + U'(\infty) \left( (1 - t)\mu_0, (M) + t\mu_1, (M) \right),$$

where $\pi = (e_0, e_1)_\ast \Pi$ 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 $\beta_s(x_0, x_1) U \left( \frac{\rho_i(x_i)}{\beta_i(x_0, x_1)} \right)$ 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, Definition 4.7] by just requiring the geodesic $t \mapsto \mu_t$ to be in $P_p(M)$ instead of $P_2(M)$. And, in case both $\mu_i$ are absolutely continuous looks like

$$U(\mu_t) \leq (1 - t) \int \frac{\beta_{1-t}(x_0, x_1)}{\rho_0(x_0)} U \left( \frac{\rho_0(x_0)}{\beta_{1-t}(x_0, x_1)} \right) d\pi(x_0, x_1) + t \int \frac{\beta_t(x_0, x_1)}{\rho_1(x_1)} U \left( \frac{\rho_1(x_1)}{\beta_t(x_0, x_1)} \right) d\pi(x_0, x_1).$$

An immediate consequence of the curvature condition is the following:

**Lemma 4.5.** 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 $P^c_2(M)$.

Furthermore, we will also define a variant of Sturm’s curvature condition [Stu06a, Stu06b]:
**Definition 4.6 ((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

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

for any $N' \geq N$. In case $N' = \infty$ the functional $U_\infty$ 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.

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

The following follows immediately from Theorem 3.13 by similar statements to the case $CD_2(K, N)$ (see e.g. [Oht09, Vil09]).

**Corollary 4.7.** 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

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

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

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

and for $p < 2$

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

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 $\beta'_k(\cdot, \cdot) \leq \beta_k(\cdot, \cdot)$ then we still have

$$U(\mu_t) \leq (1 - t)\int \frac{\beta'_{1-t}(x_0, x_1)}{\rho_0(x_0)} U \left( \frac{\rho_0(x_0)}{\beta'_{1-t}(x_0, x_1)} \right) d\pi(x_0, x_1)$$

$$+ t\int \frac{\beta'_{1}(x_0, x_1)}{\rho_1(x_1)} U \left( \frac{\rho_1(x_1)}{\beta'_{1}(x_0, x_1)} \right) d\pi(x_0, x_1),$$

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

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

\(^1\)We thank Shin-ichi Ohta for making this remark on an early version of the paper.
If $K < 0$ then obviously $\beta' \leq \beta_t$ and the interpolation inequality above holds. As above we conclude that the functional is $K'$-convex for some $K'$ depending on $D_t, K$ and $p > r$.

**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 $q$-Fisher information (associated to $(U, \mu)$)

$$I_q(\nu) = \int U''(\rho)|D^{-}\rho|^q d\nu$$

$$= \int \rho U''(\rho)|D^{-}\rho|^q d\mu$$

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

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

**Lemma 4.8.** Let $(M, d, \mu)$ be a metric measure space satisfying $\text{CD}_p(K, N)$ 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 4.9.** Let $(M, d, \mu)$ be compact geodesic metric measure space and $U$ be 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 $U$ (associated to $(U, \mu)$) is $K$ convex along $\mu_t$, i.e.

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

Then

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

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

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

**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 $U$ with $U \in \mathcal{D}_{CN}$ is not necessarily $K$-convex).

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

$$\phi(t) - \phi(1) \leq (1 - t) \left( \phi(0) - \phi(1) - \frac{K}{2} t w_p(\nu, \mu)^2 \right)$$
which implies that $\phi(t) - \phi(1)$ is negative for $t$ close to 1. But this contradicts [LV09, Lemma 3.36], i.e. $\mathcal{U}(\mu) \geq \mathcal{U}(\nu) = \mathcal{U}(1)$. 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

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

Since $U$ is convex we have

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

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

$$-\frac{1}{t}(\phi(t) - \phi(0)) \leq -\frac{1}{t} \int U'(\rho_0(x)) \left(d\mu(x) - d\mu(x)\right)$$

$$= -\frac{1}{t} \int U'(\rho_0(\gamma_t)) - U'(\rho_0(\gamma_0)) d\Pi(\gamma)$$

where $\Pi$ is the optimal transference plan in $\mathcal{P}(\text{Geo})$ 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

$$-\frac{1}{t} \int U'(\rho_0(\gamma_t)) - U'(\rho_0(\gamma_0)) d\Pi(\gamma) \leq -\frac{1}{t} \int_{\rho_0(\gamma_t) \leq \rho_0(\gamma_0)} [U'(\rho_0(\gamma_t)) - U'(\rho_0(\gamma_0))] d\Pi(\gamma)$$

$$\leq \int \frac{U'(\rho_0(\gamma_t)) - U'(\rho_0(\gamma_0))}{\rho_0(\gamma_t) - \rho_0(\gamma_0)} \frac{d\rho_0(\gamma_t) - d\rho_0(\gamma_0)}{d(\gamma_t, \gamma_0)} d(\gamma_t, \gamma_0) d\Pi(\gamma).$$

Applying Hölder inequality we get

$$\sqrt[q]{\int \frac{[U'(\rho_0(\gamma_t)) - U'(\rho_0(\gamma_0))]^q [\rho_0(\gamma_t) - \rho_0(\gamma_0)]^q}{d(\gamma_t, \gamma_0)^q} d\Pi(\gamma)}$$

$$\times \int d(\gamma_t, \gamma_0)^q d\Pi(\gamma).$$

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

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

\[\Box\]

**Corollary 4.10.** Assume that the (weak) $CD_p(K, \infty)$ condition holds for $K > 0$ and some $N \in [1, \infty]$. Then for all $\nu \in \mathcal{P}(\mathcal{M})$

$$\frac{K}{2} w_p(\nu, \mu)^2 \leq \mathcal{U}_{\infty}(\nu).$$
If $\nu$ is absolutely continuous with positive Lipschitz density $\rho$ then
\[ U_\infty(\nu) \leq w_p(\nu, \mu) \sqrt{I_q(\nu) - \frac{K}{2} w_p(\nu, \mu)^2} \leq \frac{1}{2K} (I_q(\nu))^{rac{q}{2}}. \]

**Proof.** Just note that if $U_\infty$ is $K$-convex along a geodesic $t \mapsto \mu_t$ between absolutely continuous measures, then each $\mu_t$ is absolutely continuous. \[\square\]

Note that in this case
\[ I_q(\rho \mu) = \int \rho^q |D^- \rho|^q d\mu = \int \frac{|D^- \rho|^q}{\rho^{q-1}} d\mu. \]

Similar to [LV09, Section 6.2] we will show that the $(2, q)$-log-Sobolev inequality
\[ U_\infty(\rho \mu) \leq \frac{1}{2K} (I_q(\rho \mu))^{rac{q}{2}}. \]
implies a global $(2, q)$-Poincaré inequality. Note that the $(2, q)$-log-Sobolev inequality is different from the one defined in [GRS12].

**Corollary 4.11.** Assume for $K > 0$ and all positive Lipschitz functions
\[ U_\infty(\rho \mu) \leq \frac{1}{2K} (I_q(\rho \mu))^{rac{q}{2}}. \]
Then the $(2, q)$-Poincaré inequality holds with factor independent of $q$, i.e.
\[ \left( \int (h - \bar{h})^2 d\mu \right)^{rac{q}{2}} \leq \frac{1}{\sqrt{2K}} \left( \int |D^- h|^q d\mu \right)^{rac{q}{2}} \]
for $h \in \text{Lip}(M)$. In particular, this holds if $(M, d, \mu)$ satisfies the weak CD$_p(K, \infty)$ condition.

**Proof.** We will first prove

**Claim.** If $f \in \text{Lip}(M)$ satisfies $\int f^q d\mu = 1$ then
\[ \left( \int f^q \log f^q d\mu \right)^{rac{q}{2}} \leq \frac{q}{\sqrt{2K}} \left( \int |D^- f|^q d\mu \right)^{rac{q}{2}}. \]

**Proof of the claim.** For any $\epsilon > 0$ let $\rho_\epsilon = \frac{f^q + \epsilon}{1 + \epsilon}$ then from the previous corollary
\[ \int \rho_\epsilon \log \rho_\epsilon d\mu \leq \frac{1}{2K} \left( \int |D^- \rho_\epsilon|^q d\mu \right)^{rac{q}{2}}. \]
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. \[\square\]

Assume w.l.o.g. $\int h = 0$. For $\epsilon \in [0, \frac{1}{\|h\|_{\infty}})$ set $f_\epsilon = \sqrt{1 + \epsilon} h > 0$. Then by chain rule
\[ |D^- f_\epsilon| = \frac{\epsilon |D^- h|}{q (1 + \epsilon)^{\frac{q-1}{q}}} \]
and thus
\[ \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left( \int |D^- f_\epsilon|^q d\mu \right)^{rac{q}{2}} = \frac{1}{q} \left( \int |D^- h|^q d\mu \right)^{rac{q}{2}}. \]
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

\[
\lim_{\epsilon \to 0} \frac{1}{\epsilon} \int f' f^2 \mu = \int h^2 \mu.
\]

Combining this we get

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

\[\square\]

**Metric Brenier.**

**Lemma 4.12** ([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 \(\text{supp} \mu_n \cup \text{supp} \mu_0 \subset B\), \(\mu_n\) converges weakly to \(\mu\) and

\[U_N(\mu_n) \to U_N(\mu_0) \quad \text{as } n \to \infty.\]

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 4.13.** 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 measure in \(\mathcal{P}_p(M)\) such that \(\mu_0 \ll \mu\) and there is an optimal coupling \(\pi \in \text{OptGeo}_p(\mu_0, \mu_1)\) such that

\[
\lim_{t \to 0} U_N(\mu_t) = U_N(\mu_0)
\]

and \(\text{supp}(\mu_t) \subset B\), 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 \(\tilde{\pi} \in \text{OptGeo}_p(\mu_0, \mu_1)\)

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

**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 \(y \in \partial^{p} \phi(x)\), then for all \(z \in M\)

\[
\phi(x) = c_p(x,y) - \phi^{c_p}(y),
\]

\[
\phi(z) \leq c_p(z,y) - \phi^{c_p}(y).
\]

Thus

\[
\phi(z) - \phi(y) \leq \frac{(d(z,x) + d(x,y))^p - d^p(x,y)}{p} = (d(z,x) + h_1(d(z,x))) \cdot d(x,y)^{p-1}
\]

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

\[
|D^+ \phi|(x) \leq \inf_{y \in \partial^{p} \phi(x)} d(x,y)^{p-1}.
\]
In particular, since for an arbitrary \( \tilde{\pi} \in \text{OptGeo}_p(\mu_0, \mu_1) \) we have \( \gamma_1 \in \partial^p \phi(\gamma_0) \) 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

\[
\int |D^+ \phi|^q d\mu_0 \geq w_\mu^p(\mu_0, \mu_1).
\]

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

\[
\phi(\gamma_0) - \phi(\gamma_t) \geq \frac{d(\gamma_0, \gamma_1)^p}{p} - \frac{d(\gamma_t, \gamma_1)^p}{p} = \frac{d(\gamma_0, \gamma_1)^p}{p} (1 - (1 - t)^p) = d(\gamma_0, \gamma_1)^p (t + o(t)).
\]

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

\[
\liminf_{t \to 0} \frac{1}{t^q} \left( \int_0^t \phi(\gamma_0) - \phi(\gamma_s) d\gamma_s \right) \geq \left( \int d(\gamma_0, \gamma_1)^p d\pi(\gamma) \right)^{q/p} = w_\mu^p(\mu_0, \mu_1).
\]

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

\[
\int \left( \frac{\phi(\gamma_0) - \phi(\gamma_1)}{d(\gamma_0, \gamma_1)} \right)^q d\pi(\gamma) \leq \int \frac{1}{t^q} \left( \int_0^t |D^+ \phi(\gamma_s)| d\gamma_s \right)^q d\pi(\gamma)
\]

\[
\leq \int \frac{t^q}{t^q} \int_0^t |D^+ \phi|^q(\gamma_s) d\gamma_s d\pi(\gamma)
\]

\[
= \frac{1}{t} \int_0^t \int |D^+ \phi|^q(\gamma_s) d\pi(\gamma) d\gamma
\]

because \( \frac{t^q}{t^q} = 1 = q - 1 \).

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])

\[
\lim_{t \to 0} \frac{1}{t} \int_0^t \int |D^+ \phi|^q(\gamma_s) d\pi(\gamma) d\gamma = \int |D^+ \phi|^q d\mu_0.
\]

\[ \square \]

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]).

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

\[
\phi \in D(\Delta_q) \quad \text{and} \quad \Delta^q \phi \leq N \tilde{\sigma}_{K_\mu}(|\nabla \phi|_{\mu_\mu}^{q-1}) d\mu
\]

\[ \square \]
where

\[ \hat{\sigma}_{K,N}(\theta) = \begin{cases} \frac{1}{N} \left( 1 + \theta \sqrt{K/(N-1)} \right) \cotan \left( \theta \sqrt{\frac{K}{N-1}} \right) & \text{if } K > 0 \\ 1 & \text{if } K = 0 \\ \frac{1}{N} \left( 1 + \theta \sqrt{K/(N-1)} \right) \coth \left( \theta \sqrt{\frac{K}{N-1}} \right) & \text{if } K < 0 \end{cases} \]

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

Corollary 4.15 (Laplacian comparision of the distance). For any \( x_0 \) one has

\[ \frac{dp_{x_0}}{p} \in D(\Delta_q) \quad \text{with} \quad \Delta_q \frac{dp_{x_0}}{p} \leq N \hat{\sigma}_{K,N}(d_{x_0})d\mu \quad \forall x_0 \in X \]

and

\[ d_{x_0} \in D(\Delta_q, X \setminus \{x_0\}) \quad \text{with} \quad \Delta^q d_{x_0} \big|_{X \setminus \{x_0\}} \leq N \hat{\sigma}_{K,N}(d_{x_0}) \frac{d_{x_0}^{p-1}}{d_{x_0}} d\mu. \]

Remark. Note that formally

\[ \Delta^q \frac{dp_{x_0}}{p} = \nabla \cdot \left( \frac{dp_{x_0}}{p} \left( |\nabla d_{x_0}|^{q-2} \nabla d_{x_0} \right) \right) = \nabla \cdot \left( (d_{x_0}^{p-1})^{q-1} \nabla d_{x_0} \right) = \nabla \cdot (d_{x_0} \nabla d_{x_0}) = \Delta d_{x_0}^2, \]

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

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

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. If such a principle holds in a more general non-linear setting and even for the case \( p \neq 2 \), one could also 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 geodesic rays are parametrized by arc length. We can the Busemann function \( b \) associated to \( \gamma \) by

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

Note

\[ t \mapsto b_t(x) \quad \text{is non-increasing} \]

Lemma 4.16. 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 2.2 we know $b^e_r \geq b$, 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

$$b^e_r(x) = \inf_{y \in X} \sup_{x \in X} \frac{d^p(x, y)}{p} - \frac{d^p(\tilde{x}, y)}{p} + b(\tilde{x}) \leq \sup_{x \in X} \frac{1}{p} - \frac{d^p(\tilde{x}, \gamma_t)}{p} + b(\tilde{x}).$$

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

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

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

$$b^e_r(x) \leq \lim_{t \to \infty} b_t(x) = b(x).$$

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 4.17.** 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 $b^{c_L \tilde{e}_L} \leq b$. We have

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

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

$$L(r) - L(d(\tilde{x}, \gamma_r^{t,x})) + b_t(\tilde{x}) = L(r) - L(d(\tilde{x}, \gamma_r^{t,x})) + d(\tilde{x}, \gamma_t) - t \leq L(r) - L(d(\tilde{x}, \gamma_r^{t,x})) + d(\tilde{x}, \gamma_1^{t,x}) + d(\gamma_1^{t,x}, \gamma_t) - t = L(r) - r - L(d(\tilde{x}, \gamma_r^{t,x})) + d(\tilde{x}, \gamma_r^{t,x}) + d(x, \gamma_t) - t = -L^*(1) - L(d(\tilde{x}, \gamma_r^{t,x})) + d(\tilde{x}, \gamma_r^{t,x}) + d(x, \gamma_t) - t \leq d(x, \gamma_t) - t = b_t(x).$$
where we used Young’s inequality to get the last inequality. Therefore,
\[ b^{p+r}(x) \leq \lim_{t \to \infty} b_t(x) = b(x). \]

\[ \square \]

Appendix A. Appendix

In this appendix 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.

**Orlicz-Wasserstein spaces.** Let \( L : [0, \infty) \to [0, \infty) \) be a strictly convex increasing functions 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
\[ L_\lambda(r) = \int_0^r l_\lambda(s)ds = \int_0^{r/\lambda} l(s)ds \]
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 \( \phi : X \to \mathbb{R} \) 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 A.1** (Orlicz-Wasserstein space). Let \( \mu_i \) be two probability measures on \( M \) and define
\[ w_L(\mu_0, \mu_1) = \inf \left\{ \lambda > 0 \mid \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int L_\lambda(d(x, y)) d\pi(x, y) \leq 1 \right\}. \]

With convention \( \inf \emptyset = \infty \).

According to Sturm [Stu11, Proposition 3.2], \( w_L \) is a complete metric on \( \mathcal{P}_L(M) := \{ \mu_1 \in \mathcal{P}(M) \mid w_L(\mu_1, \delta_{x_0}) < \infty \} \)
where \( x_0 \) is some fixed point.

Even though the following lemma is not needed, it makes many proofs below easier.
Lemma A.2 ([Stu11, Proposition 3.1]). For every \( \mu_i \in P_L(M) \) there is an optimal coupling \( \pi_{opt} \) of \( (\mu_0, \mu_1) \) such that
\[
\lambda_{\text{min}} = w_L(\mu_0, \mu_1) \Rightarrow \int L_{\lambda_{\text{min}}}(d(x,y))d\pi_{opt}(x,y) = 1.
\]

Actually the Lemma shows that the whole theory of Kantorovich potentials will depend on the distance. Furthermore, the \( c_L \)-convex functions are not necessarily star-shaped. Nevertheless, we will show that \( P_L(M) \) is a geodesic space iff \( M \) is and that a similar property to the star-shapedness holds.

Proposition A.3. Let \( \Phi \) be a convex increasing function with \( \Phi(1) = 1 \), then
\[
w_L \leq w_{\Phi \circ L}.
\]

Remark. This just uses Sturm’s idea to show the same inequality for 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 \( \int (\Phi \circ L)(d(x,y))d\pi(x,y) \leq 1 \) then since \( (\Phi \circ L)_{\lambda} = \Phi \circ L \)
\[
\Phi(\int L_{\lambda}(d(x,y))d\pi(x,y)) \leq \int \Phi \circ L_{\lambda}(d(x,y))d\pi(x,y) \leq 1
\]
Since \( \Phi(1) \leq 1 \) and \( \Phi \) is increasing, we see that \( \int L_{\lambda}(d(x,y))d\pi(x,y) \leq 1 \) which implies \( w_L(\mu_0, \mu_1) \leq w_{\Phi \circ L}(\mu_0, \mu_1) \). \( \square \)

Proposition A.4. Assume for all \( \lambda > 0 \)
\[
\sup_{R \to \infty} \frac{L(\lambda R)}{L(R)} < \infty.
\]
If \( \mu_n, \mu_\infty \in P_L(M) \) and \( \mu_n \) converges weakly to \( \mu_\infty \), then
\[
w_L(\mu_n, \mu_\infty) \to 0 \iff \lim_{R \to \infty} \limsup_{n \to \infty} \int_{M \setminus B_R(x_0)} L_{\lambda}(d(x,x_0))d\mu_n = 0
\]
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 \( \mu' \in P_L(M) \)
\[
\lim_{R \to \infty} \int_{M \setminus B_R(x_0)} L_{\lambda}(d(x,x_0))d\mu'(x) = 0 \iff \lim_{R \to \infty} \int_{M \setminus B_R(x_0)} L(d(x,x_0))d\mu'(x) = 0.
\]
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 \( r_n \leq \frac{1}{2} \) such that \( l_n = r_n\lambda \). Then using the triangle inequality and convexity of \( L \) we get

\[
\int L_\lambda (d(x, x_0)) d\mu_n(x) = \int L_\lambda (d(x, x_0)) d\pi_n(x, y)
\]

\[
\leq r_n \int L_{r_n \lambda} (d(x, x_0)) d\pi_n(x, y) + (1 - r_n) \int L_{(1-r_n)\lambda} (d(y, x_0)) d\pi_n(x, y)
\]

\[
\leq r_n + (1 - r_n) \int L_{\frac{1}{\lambda}} (d(y, x_0)) d\mu_\infty(y).
\]

since \( L_{(1-r_n)\lambda} \leq L_{\frac{1}{\lambda}} \). Therefore,

\[
\lim_{R \to \infty} \lim_{n \to \infty} \int_{M \setminus B_R(x_0)} L_\lambda(d(x, x_0))d\mu_n(x) \leq \lim_{R \to \infty} \int_{M \setminus B_R(x_0)} L_{\frac{1}{\lambda}}(d(x, x_0))d\mu_\infty(x) = 0.
\]

Now assume that

\[
\lim_{R \to \infty} \lim_{n \to \infty} \int_{M \setminus B_R(x_0)} L_\lambda(d(x, x_0))d\mu_n(x) = 0
\]

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

\[
d(x, y) \leq d(x, y) \wedge R + 2d(x, x_0)\chi_{B_{R/2}(x_0)}(x) + 2d(x_0, y)\chi_{B_{R/2}(x_0)}(y)
\]

and thus by convexity of \( L \) and \( L(0) = 0 \)

\[
L_\lambda(d(x, y)) \leq \frac{1}{3} L_{\frac{1}{\lambda}}(d(x, y) \wedge R) + \frac{1}{3} L_{\frac{1}{\lambda}}(d(x, x_0)\chi_{B_{R/2}(x_0)}(x)) + \frac{1}{3} L_{\frac{1}{\lambda}}(d(x_0, y)\chi_{B_{R/2}(x_0)}(y)).
\]

Thus integrating over \( \pi_n \) we get

\[
3 \int L_\lambda(d(x, y))d\pi_n(x, y) \leq \int L_{\frac{1}{\lambda}}(d(x, y) \wedge R)d\pi_n(x, y)
\]

\[
+ \int_{M \setminus B_{R/2}(x_0)} L_{\frac{1}{\lambda}}(d(x, x_0))d\mu_n(x)
\]

\[
+ \int_{M \setminus B_{R/2}(x_0)} L_{\frac{1}{\lambda}}(d(x_0, y))d\mu_\infty(y).
\]

we first take the lim sup with \( n \to \infty \) and then \( R \to \infty \) and conclude that the last two terms converges to zero by our assumption and since \( L_{\frac{1}{\lambda}}(d(x, y) \wedge R) \) is a bounded continuous function and \( \pi_n \) converges weakly to the trivial coupling \((\text{Id} \times \text{id}), \mu_\infty \), 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 \( w_L(\mu_n, \mu_\infty) \to 0 \). \( \square \)
Proposition A.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 $w_L$ 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 $w_L$-ball is compact in $\mathcal{P}_L(M)$

So for some $r > 0$ and $\mu_0 \in \mathcal{P}_L(M) \subset \mathcal{P}_L(M)$ let

$$\tilde{B} := \tilde{B}_r(\mu_0) = \{\mu_1 \in \mathcal{P}_L(M) \mid w_L(\mu_0, \mu_1) \leq r\}.$$  

and let $(\mu_n)_{n \in \mathbb{N}}$ be a sequence in $\tilde{B}$. Then there are (optimal) couplings $\pi_n$ such that

$$\int L_r(d(x,y))d\pi_n(x,y) \leq 1$$  

(for $w_L(\mu_n, \mu_0) < r$ 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 are stable 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

$$\lim_{R \to \infty} \limsup_{n \to \infty} \int_{M \setminus B_R(x_0)} L_\lambda(d(x, x_0))d\mu_n = 0$$  

i.e. it is precompact in $\mathcal{P}_L(M)$ by the lemma above.

Since $\tilde{B}$ is bounded w.r.t. $w_L$ we can assume that for some $R > 0$

$$w_L(\mu_n, \delta_{x_0}) \leq \lambda_0.$$  

Now set $\lambda_0 = 0$. For $c \lambda = \lambda_0$ and $c \in (0, 1)$ we have

$$\int_{M \setminus B_R(x_0)} L_\lambda(d(x, x_0))d\mu(x) \leq \frac{L_\lambda(R)}{\Phi(L_{\lambda_0}(R))} \int_{M \setminus B_R(x_0)} \tilde{L}_{\lambda_0}(d(x, x_0))d\mu_n(x)$$  

$$\leq \frac{L_{\lambda_0}(R)}{\Phi(L_{\lambda_0}(R))} \frac{L_{\lambda_0}(c^{-1}R)}{L_{\lambda_0}(R)} \leq C \frac{L_{\lambda_0}(R)}{\Phi(L_{\lambda_0}(R))}$$  

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

$$\lim_{R \to \infty} \limsup_{n \to \infty} \int_{M \setminus B_R(x_0)} L_\lambda(d(x, x_0))d\mu_n = 0.$$  

In order to show weak precompactness notice that $L(R) \geq 1$ for $R \geq r_0 = r_0(L)$ implies tightness, which is equivalent to precompactness by the classical Prokhorov theorem. Indeed, $B_R(x_0)$ is compact and for $r_0 \leq R \to \infty$

$$\int_{M \setminus B_R(x_0)} d\mu_n \leq C \frac{L_{\lambda_0}(R)}{\Phi(L_{\lambda_0}(R))} \to 0$$  

uniformly in $n$.  

$\square$
**Proposition A.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, $P_L(M)$ 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 $\lambda_{min} = w_L(\mu_0, \mu_1)$

$$ \int L(\frac{d(\gamma_s, \gamma_t)}{|s - t|\lambda_{min}})d\Pi(\gamma) = \int L_{\lambda_{min}}(d(\gamma_0, \gamma_1))d\Pi(\gamma) = 1. $$

Hence

$$ w_L(\mu_t, \mu_s) \leq |s - t|\lambda_{min}. $$

So $t \mapsto \mu_t$ is absolutely continuous in $P_L(M)$ and $|\dot{\mu}_t| \leq \lambda_{min}$. 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

$$ w_L(\mu_s, \mu_t) = \int_s^t |\dot{\mu}_r|dr = |s - t|w_L(\mu_s, \mu_t). $$

□

It is also possible to define a dual problem by

$$ \sup\{\lambda > 0 \mid \sup_{\phi \in L^1(\mu_0)} \left\{ \int \phi d\mu_0 + \int \phi \gamma \lambda \mu_1 \right\} \geq 1\}. $$

However, we will not go into this dual problem and directly deal with the $c_L$-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 $c_L$-concave functional enjoy a similar property to star-shapedness. It will also show that the Jacobians of the interpolation measures are positive semidefinite.

**Lemma A.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 2.7.

**Proof.** Since $L$ is convex and increasing

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

Dividing by $t$ we get the inequality and choosing $x = m$ we see that all inequalities are actually equalities. □
Lemma A.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 2.7 above for \( t \in (0,1) \) we have for \( z = \eta_t \in Z_t(x,y) \)
\[
-h(m) = t^{-1} L(d(m,y)) - L_t(d(m,z)) \leq t^{-1}(1 - t) L(d(x,y))
\]
\[
= t^{-1} L(d(x,y)) - L_t(d(x,\eta_t)) = -h(x).
\]
\( \Box \)

Lemma A.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 \( t^{-1} \phi \in \mathcal{I}^{c_l}(X,Z_t(X,Y)) \).

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 \( \phi(x) := c_t(x,y) = L_y(x) \). We claim that the following representation holds
\[
t^{-1} L_y(m) = \inf_{z \in Z_t(x,y)} \left\{ (L_t)_z(m) + \inf_{\{x \in X \mid z \in Z_t(x,y)\}} t^{-1}(1 - t)L_y(x) \right\}.
\]
Indeed, by Lemma 2.7 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
\[
\psi(z) = -\inf_{\{x \in X \mid z \in Z_t(x,y)\}} t^{-1}(1 - t)L_y(x)
\]
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 \( c_t \)-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 \( \phi = \phi^{c_t} \) we have
\[
t^{-1} \phi(x) = \inf_y t^{-1} L(d(x,y)) - t^{-1} \phi^{c_t}(y).
\]
But each function
\[
\psi_y(x) = t^{-1} L_y(x) - t^{-1} \phi^{c_t}(y)
\]
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. \( t^{-1} \phi \in \mathcal{I}^{c_t}(X,Z_t(X,Y)) \). \( \Box \)

Orlicz-Wasserstein spaces on Finsler manifolds.

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 \( r \geq 0 \)
\[
\nabla \phi(x) = l(r)v
\]
Combining these two facts with the proof of Lemma A.10 we get for \( t > 0 \)

\[
\eta \mapsto \nabla L \phi \quad \text{is bounded in a neighborhood} \quad \lim_{t \to 0} \frac{f(\xi_v(t)) + f(\xi_v(-1)) - 2f(x)}{F(v)^2} = -\infty
\]

where \( \xi_v : [-1,1] \to M \) is the geodesic with \( \xi_v(0) = v \).

**Proof.** The proof follows in the same fashion as Lemma 3.1. 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 \( v = \zeta(0) \) and \( w = \eta(0) \). For fixed small \( \epsilon > 0 \) set \( y_\epsilon = \eta(d(x,y) - \epsilon) \) then \( y_\epsilon, y \notin \text{Cut}(x) \cup \{x\} \) and using the first variation formula we get for \( t > 0 \)

\[
f(\xi_v(-t)) - f(x) \leq L(d(\xi_v(-t), y_\epsilon) + \epsilon) - L(d(x, y_\epsilon) + \epsilon)
= tl(d(x, y_\epsilon) + \epsilon)g_{\eta(0)}(v, \eta(0)) + O(t^2)
= tl(d(x, y))g_{\eta(0)}(v, \eta(0)) + O(t^2).
\]

The term \( O(t^2) \) 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

\[
f(\xi_v(t)) - f(x) = L(d(x,y) - t) - L(d(x,y)) = -tl(d(x,y)) + O(t^2).
\]

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

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

For the conjugate point case, we use the same construction and notation as in the proof of Lemma 3.1. Note that

\[
\lim_{s \to 0} \frac{L(\mathcal{L}(\sigma_s)) + L(\mathcal{L}(\sigma_{-s})) - 2L(\mathcal{L}(\sigma_0))}{s^2} = \left( l(\mathcal{L}(\sigma_0)) \frac{\partial^2}{\partial s^2} \mathcal{L}(\sigma_s) \right)_{s=0}
+ l'(\mathcal{L}(\sigma_0)) \left( \frac{\partial L(\sigma_s)}{\partial s} \right)_{s=0}^2
\leq l(d(x,y)) \left( -2\epsilon g_\eta(v, v)/d(x,y)
+ \epsilon^2 \{ T_{\eta(0)}(v)/d(x,y) + I(V, V)^2 \}
+ l'(d(x,y))F(v) \right).
\]
Using the fact that $f(\xi_\epsilon(\epsilon s)) \leq L(\mathcal{L}(\sigma_s))$ we obtain
\[
\liminf_{s \to 0} \frac{f(\xi_\epsilon(\epsilon s)) + f(\xi_\epsilon(-\epsilon s)) - 2f(x)}{\epsilon^2 s^2} \leq \liminf_{s \to 0} \frac{L(\mathcal{L}(\sigma_s)) + L(\mathcal{L}(\sigma_{-s})) - 2L(\mathcal{L}(\sigma_0))}{\epsilon^2 s^2} \\
\leq \ell(d(x, y))(\n-2\epsilon^{-1}g_\theta(v, v)/d(x, y) \\
+ T(v)/d(x, y) + d(x, y)I(V, V) \\
+l'(d(x, y))F(v)^2).
\]

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

The Brenier-McCann-Ohta solution.

**Lemma A.11.** Let $\phi : M \to \mathbb{R}$ be a $c_L$-concave function. If $\phi$ is differentiable at $x$ then $\partial^{c_L} \phi(x) = \{\text{exp}_x(\nabla^L(-\phi)(x))\}$. Moreover, the curve $\eta(t) := \text{exp}_x(t\nabla^L(-\phi)(x))$ is a unique minimal geodesic from $x$ to $\text{exp}_x(\nabla^L(-\phi)(x))$.

Remark. See also [McCo01, 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 $v \in T_x M$
\[
f(\text{exp}_x v) \geq \phi^{c_L}(y) + \phi(\text{exp}_x v) = f(x) - \phi(x) + \phi(\text{exp}_x v) = f(x) + d\phi_x(v) + o(F(v)).
\]

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
\[
f(\text{exp}_x v) - f(x) \leq L(d(\text{exp}_x v, y_\epsilon) + \epsilon) - L(d(x, y_\epsilon) + \epsilon) \\
= -l(d(x, y_\epsilon) + \epsilon)g_\theta(0)(v, \dot{\eta}(0)) + o(F(v)) \\
= -l(d(x, y))\mathcal{L}_x^{-1}(\dot{\eta}(0))(v) + o(F(v)).
\]

Therefore, $d\phi_x(v) \leq -l(d(x, y))\mathcal{L}_x^{-1}(\dot{\eta}(0))(v)$ for all $v \in T_x M$ 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 $\eta(t) = \text{exp}_x(t\nabla^L(-\phi)(x))$, which is uniquely defined. \qed

**Lemma A.12.** Let $t \mapsto \mu_t$ be a geodesic between $\mu_0$ and $\mu_1$, i.e. $w_L(\mu_0, \mu_1) = t\lambda$. If $\mu_0$ is absolutely continuous and the unique $\phi_0$ the Kantorovich potential of $(\mu_0, \mu_1)$ w.r.t. $L_{t\lambda}$ such that $\phi_1(x_0) = 0$. Then $\phi_t = t^{-1}\phi$.

**Proof.** For $x \neq y \in \partial^{c_L} \phi_1(x)$ define $x_t = \text{exp}_x(t\nabla^L(-\phi)(x))$. Since $x_t \in \partial^{c_L} \phi_1(x)$, we have for $t \in (0, 1]$
\[
x_t = \text{exp}_x(t\nabla^L(-\phi)(x)) \\
= \text{exp}_x \left( t \cdot \mathcal{L}_x^{-1}(t^{-1}\nabla(-\phi)(x)) \cdot \nabla(-\phi)(x) \right) \\
= \text{exp}_x \left( t \cdot \mathcal{L}_x^{-1}(t^{-1}\nabla(-\phi)(x)) \cdot \nabla(-\phi)(x) \right) \\
= \text{exp}_x \left( t \cdot \mathcal{L}_x^{-1}(t^{-1}\nabla(-\phi)(x)) \cdot \nabla(-\phi)(x) \right).
\]

Since $t^{-1}\phi$ is $c_L$-concave and $t^{-1}\phi(x_0) = 0$, uniqueness implies $\phi_t = t^{-1}\phi$. \qed
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 $\phi^L = \phi^r$ and $L_t = t^p d^p/p$. Hence
\[
\phi_t^c(y) = \inf \frac{d_p(x,y)}{p} - t^{-1}\phi(x) = t^{-p} \inf \frac{d_p(x,y)}{p} - t^{-1}\phi(x) = t^{-p}(t^{-1}\phi)^c_r(y)
\]
Thus up to a factor the interpolation potentials are the same (recall that $t^{-p-1}\phi$ gives the potential of $(\mu_0, \mu_1)$ w.r.t. $c_p$).

The next results follow using exactly the same arguments as for $c_p$.

**Lemma A.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 A.9 can be used to show that the interpolation inequality in Theorem A.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 A.14.** 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 $c_L$-concave function $\phi$ such that $\pi = (\text{Id} \times \mathcal{F})_*\mu_0$ 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 A.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 A.16.** Assume $\mu_0$ is absolutely continuous and $\phi$ is $c_L$-concave with $\lambda = w_L(\mu_0, (\mathcal{F}_t)_*\mu_0)$ 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 A.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
\[
w_L(\mu_s, \mu_t) \leq |s-t| w_L(\mu_0, \mu_1).
\]
Let $\pi$ be the plan on $\text{Geo}(M) = \{ \gamma : [0,1] \to M \mid \gamma \text{ is a geodesic in } M \}$ give by $\mu_0$, the map $\mathcal{F}_t$ and the unique geodesic connecting $\mu$-almost every $x \in M$ to a point $\mathcal{F}_t(x)$ (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, $w_L(\mu_s, \mu_t) \leq |t-s|\lambda$. \hfill \Box

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 3.8 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 A.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 $U \subset M\setminus\Omega_{id}$ of full measure (relative to $\mathcal{M}\setminus\Omega_{id}$). In particular, it is second order differentiable almost everywhere in $U$.

Proof of the interpolation inequality in the Orlicz case.

Theorem A.18 (Volume distortion for $L$). Let $x \neq y$ with $y \notin \text{Cut}(x)$ 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

$$v_t^x(y) = D \left[ d(\exp_x)\nabla_{L_t f_t(x)} \circ [d(\exp_x)\nabla_{L_t f_t(x)}]^{-1} \right]$$

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 for (see calculations in the proof of Lemma A.12)

$$\mathcal{K}_{t}(d(t^{-1}f_1)_z) = \frac{l_t^{-1}(t^{-1}\nabla f_1)_z}{|\nabla f_1|_z} \nabla f_1(z)$$

$$= \frac{t^{-1}(\nabla f_1)_z}{|\nabla f_1|_z} \nabla f_1(z)$$

and hence

$$v_t^x(y) = (1-t)^{-n} D \left[ d(\exp \circ \mathcal{L} \circ (d(tg_1)_z)) \right]$$
We have $d(f_1)_x = d(t^{-1}f_1)_x$. Indeed, since $l_t(r) = t^{-1}l(t^{-1}r)$ and $d(d(\cdot, \eta(t)))_x = d(d(\cdot, y))_x$
\[
-d(f_1)_x = d(L_t(d(\cdot, \eta(t)))_x = l_t(d(x, \eta(t)))d(d(\cdot, \eta(t)))_x = t^{-1}l(t^{-1}td(x, y))d(d(\cdot, y))_x = -d(t^{-1}f_1)_x
\]

Similar to [Oht09, Proof of 3.2] it suffices to show that
\[
d(exp_x \circ K^t_x)(d(f_1)_x) = 0.
\]

Now since $\nabla f_1(z) = l_t(d(z, \eta(t)))\nabla d_{\eta(t)}(z)$ we get in a neighborhood $U$ of $x$ not containing $\eta(t)$.
\[
K^t_x(d(f_1)_z) = \nabla L_t(t^{-1}f_1)(z) = l_t^{-1}(l_t(d(z, \eta(t)))\nabla d_{\eta(t)}(z) = L_z(d(g_t)_z)
\]

and thus the function $D : U \rightarrow M$ defined as
\[
D(z) = exp_x \circ K_x(d(f_1)_z) = exp_z \circ L_z(d(g_t)_z) = \eta(t).
\]
is constant in a neighborhood of of $x$. This immediately implies $dL_x = 0$. \hfill \Box

**Proposition A.19.** Let $\phi : M \rightarrow \mathbb{R}$ be a $c_L$-concave function and define $F(z) = exp_z(\nabla_L(-\phi)(z))$ at all point of differentiability of $\phi$. Fix some $x \in M$ such that $\phi$ is second order differentiable at $x$ and $d\phi_x \neq 0$. Then the following holds:

1. $y = F(x)$ is not a cut point of $x$.
2. The function $h(z) = c_L(z, y) - \phi(z)$ satisfies $dh_x = 0$ and $\left(\frac{\partial^2 h}{\partial x^i \partial x^j}(x)\right) \geq 0$
3. Define $f_y(z) := -c_L(z, y)$ and
\[
d\mathcal{F}_x := d(exp_x \circ K^t_x)(d(-\phi))_x \circ [d(d(-\phi))_x - d(df_y)_x] : T_xM \rightarrow T_yM
\]
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 $v \in T_xM$
\[
\sup \{ |u - d\mathcal{F}_x(v)| \mid exp_yu \in \nabla^L \phi(exp_x y), |u| = d(y, exp_y u) \} = o(|v|).
\]

**Proof:** The proof follows without any change from the proof of Proposition 3.11 but using Lemma A.10 instead and the fact that $y \notin \text{Cut}(x) \cup \{x\}$ implies that $f_y$ is $C^\infty$ at $x$ and $\nabla^L f_y(x) = \nabla^L \phi(x)$.

Similarly the Jacobian equation holds:

**Proposition A.20.** Let $\mu_0$ and $\mu_1$ be absolutely continuous measure with density $f_0$ and $f_1$ and $\lambda = \omega_L(\mu_0, \mu_1)$. Also assume that there are open set $U_i$ with compact closed $C = \bar{U}_0$ and $Y = \bar{U}_1$ such that $\text{supp} \mu_i \subset U_i$. Let $\phi$ be the unique $c_L$-concave Kantorovich potential and define $F(z) = exp_z(\nabla L(-\phi)(z))$. Then $F$ is injective $\mu_0$-almost everywhere and for $\mu_0$-almost every $x \in M \setminus \Omega_{ld}$,
Remark. The proof is based on the proof of Theorem A.21.

Proof. Similar to Proposition 3.12, the proof follows without any change from [Oht09, Theorem 5.2], see also [Vil09, Chapter 11].

Theorem A.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 \( J_t(x) = D[d(F_t)_x] \) where

\[
d(F_t)_x := d(\exp_x \circ K^+_t d(-t^{-1}\phi)_x \circ [d(d(-t^{-1}\phi))_x - d(d(f_t))_x]] : T_x M \to T_y M.
\]

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

\[
J_t(x)^{1/n} \geq (1-t)\mathbf{v}^+_t(x,y_1)^{1/n} + t\mathbf{v}^-_t(x,y_1)^{1/n} J_1(x)^{1/n}.
\]

Remark. The proof is based on the proof of [Oht09, Proposition 5.3] but is notationally slightly more involved then the proof of Theorem 3.13.

Proof. Note first that

\[
d(d(-t^{-1}\phi))_x - d(d(f_t))_x = \{d(d(-t^{-1}\phi))_x - d(d(f_t))_x\} + \{d(d(-1f_1))_x - d(d(f_1))_x\}
\]

and

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

Now define \( t : T^* M \to T^* M \) as \( t(v) = s^{-1}v \) and note for \( \nabla \phi(x) \neq 0 \)

\[
K^+_t(t^{-1}d\phi_x) = \frac{l_{t^{-1}}(|\nabla t^{-1}\phi(x)|)}{|\nabla t^{-1}\phi(x)|} \nabla t^{-1}\phi(x)
\]

\[
= \frac{t(l(|\nabla \phi(x)|)}{|\nabla \phi(x)|} \nabla \phi(x) = tK^+_t(d\phi_x)
\]

and thus

\[
K^+_t \circ t_1 \circ (K^+_x)^{-1} = tId_{T_x M}
\]

which implies

\[
d(\exp_x \circ K^+_t d(-t^{-1}\phi)_x \circ [d(d(-t^{-1}\phi))_x - d(d(-1f_1))_x]]
\]

\[
= d(\exp_x \circ K^+_t d(-t^{-1}\phi)_x \circ [d(\tau_1 d(-\phi)_x - d(\phi))_x - d(\tau_1 f_1)_x]]
\]

\[
= d(\exp_x \circ K^+_t d(-t^{-1}\phi)_x \circ [d(\tau_1 d(-\phi)_x - d(\phi))_x - d(\tau_1 f_1)_x]]
\]

\[
= d(\exp_x \circ K^+_t d(-t^{-1}\phi)_x \circ d(K_x \circ t_1 \circ K^{-1}_x) \nabla \phi(x) \circ [d(\exp_x \circ \nabla \phi(x))^{-1}]) \circ d(F_t)
\]

\[
= t \cdot d(\exp_x \circ \nabla \circ t(-t^{-1}\phi)_x \circ d(\exp_x \circ \nabla \phi(x))^{-1} \circ d(F_t).
\]
where we identified $T_{\nabla^t(-\phi)(x)}(T_xM)$ with $T_{\nabla^L(-\phi)(x)}(T_xM)$ to get the last inequality (remember $t\nabla^t(-\phi\cdot) = \nabla^L(-\phi)$).

Because $D$ is concave we get

$$J_t(x)^{1/n} = D[d(F_t)_{x}]^{1/n}$$

$$= D \left[ d(exp_x \circ K^t_x)_{d(-\phi)_x} \circ \left[ d(d(-t^{-1} f_1))_x - d(df_t)_x \right] + d(exp_x \circ K^l_x)_{d(-\phi)_x} \circ \left[ d(d(t^{-1} \phi))_x - d(df_t)_x \right] \right]^{1/n}$$

$$= D \left[ d(exp_x \circ K^t_x)_{d(-\phi)_x} \circ \left[ d(d(-t^{-1} f_1))_x - d(df_t)_x \right] + t \cdot d(exp_x \circ K^l_x)_{d(-\phi)_x} \circ \left[ d(exp_x \circ K^l_x)_{d(\phi)}_{x}^{-1} \circ d(F_t) \right] \right]^{1/n}$$

$$\geq (1 - t) D \left[ (1 - t)^{-1} d(exp_x \circ K^t_x)_{d(-\phi)_x} \circ \left[ d(d(t^{-1} f_1))_x - d(df_t)_x \right] \right]^{1/n}$$

$$+ t D \left[ d(exp_x \circ K^l_x)_{d(-\phi)_x} \circ \left[ d(exp_x \circ K^l_x)_{d(\phi)}_{x}^{-1} \circ d(F_t) \right] \right]^{1/n}$$

$$= (1 - t) J_t^\gamma(x, y)^{1/n} + t J_t^\gamma(x, y)^{1/n} J_1(x)^{1/n}.$$

\[\square\]

Combining this with Lemma 2.11 (see remark after that lemma) and Lemma A.23 below we get similar to Lemma 3.14 and [Oht09, 6.2]:

**Lemma A.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 $F_t(x) := exp_x(\nabla^L(\lambda(-\phi))$ for $t \in (0, 1]$. Then $\mu_t = \rho_t \mu$ is absolutely continuous for any $t \in [0, 1]$.

**Proof.** By Lemma 2.11 the map $F_t$ 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

$$\mu_t|_{\Omega_{id}} = (F_t)^*(\mu_0|_{\Omega_{id}}) = \mu_0|_{\Omega_{id}}.$$ 

By Theorem 3.8 the potential $\phi$ is second order differentiable in a subset $\Omega \subset M \setminus \Omega_{id}$ of full measure. In addition, $D[d(F_t)] > 0$ for all $x \in \Omega$ (see Proposition 3.12) and $F_t$ is continuous in $\Omega$ for any $t \in [0, 1]$.

The map $d(F_t) : T_xM \to T_{F_t(x)}M$ defined in Proposition 3.11 as

$$d(F_t)_{x} := d(exp_x \circ K^t_x)_{d(-\phi)} \circ \left[ d(d(-t^{-1} \phi))_x - d(df_t)_x \right]$$

where $f_t(z) := -c_\lambda(z, F_t(x))$ for $t \in (0, 1]$.

Also note that for $x \in \Omega$

$$d(d(-t^{-1} \phi))_x - d(df_t)_x = \left\{ d(d(-t^{-1} \phi))_x - d(df_t)_x \right\} + \left\{ d(df_t)_x - d(df_t)_x \right\}.$$

Which implies $D[d(F_t)] > 0$ because $D[d(F_t)_x] > 0$ and the lemma below.

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

**Lemma A.23.** Let $y \notin \text{Cut}(x) \cup \{x\}$ and $\eta : [0, 1] \to M$ be the unique minimal geodesic from $x$ to $y$. Define

$$f_t(z) := -c_\lambda(z, \eta(t)).$$
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 A.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 $DC_N$ 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 A.24.** 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 strictly convex, increasing functional $L : [0, \infty) \to [0, \infty)$ which is smooth away from zero.

**References**

[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).

[BCS00] D. Bao, S. Chern, and Z. Shen, *An Introduction to Riemann-Finsler Geometry*, 2000.

[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.

[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).

[Gig12] N. Gigli, *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. Gigli, 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).

[Ke11] M. Kell, *Stability of the global attractor under Markov-Wasserstein noise*, arxiv:1103.3401 (2011).

[Ke14] , *q-heat flow and the gradient flow of the Renyi entropy in the $p$-Wasserstein space*, arxiv (2014).

[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).

[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.

[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.

E-mail address: mkell@mis.mpg.de

Max-Planck-Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany