HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

Department of Mathematics

Ricci Curvature & Optimal Transport

Author:
Sotirios Simos

Supervisor:
Dr. Panagiotis Gianniotis

Master in Pure Mathematics:
Thesis

August 25, 2020

To my grandma.

Περίληψη

Αφού πρώτα υπενθυμίσουμε βασικά στοιχεία της γεωμετρίας Riemann, θα κάνουμε μια εισαγωγή στα βασικά της θεωρίας βέλτιστης μεταφοράς πάνω σε πολλαπλότητες Riemann. Το κύριο αποτέλεσμα που παρουσιάζεται είναι η ισοδυναμία ενός κάτω φράγματος για την καμπυλότητα Ricci με την K-κυρτότητα της σχετικής εντροπίας, ένα συναρτησοειδές πάνω στο χώρο των απόλυτα συνεχών (ως προς $\mathrm{vol_g}$) μέτρων πιθανότητας. Αυτή η ισοδυναμία επιτρέπει τον ορισμό ενός κάτω φράγματος για την καμπυλότητα Ricci σε μετρικούς χώρους με μέτρο, όπου η δομή Riemann απουσιάζει.

Abstract

After reminding some basic elements of Riemannian Geometry, we will make an introduction to the basics of optimal transport theory on Riemannian manifolds. The main result presented is the equivalence of a lower bound for Ricci curvature with the K-convexity of the relative entropy, a functional on the space of absolutely continuous (w.r.t. $\mathrm{vol_g}$) probability measures. This equivalence allows the definition of lower Ricci bounds on metric measure spaces, where the Riemannian structure is absent.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Panagiotis Gianniotis for his constant support, guidance and motivation. It would never have been possible for me to take this work to completion without his incredible support and encouragement. I am truly grateful to my parents for their immeasurable love and care. They have always encouraged me to explore my potential and pursue my dreams. They helped me a lot to reach this stage in my life. I would also like to thank my cousin, Sotirios Roussos, for raising me to love mathematics and making me who I am today. Last, but not least I would like to thank my friends, for their support and friendship, important ingredients for my ability to write this thesis.

Contents

Chapter 1

Introduction

The purpose of this thesis is twofold. First of all, it is to present the subject itself and cultivate a desire for further exploration and research. Secondly, it is to stress the fact and remind to the reader (as well as myself) that the combination of different areas of mathematics can yield fruitful results. Moreover, the journey itself, in doing so, is an interesting experience, since it can give a unique point of view of these areas. To be more precise, in our case, we will develop (to some satisfying point) the theory of **Optimal Transport**, assuming that we work on a riemannian manifold, hence we can use every tool from **Riemannian Geometry**, so that we can generalize a lower Ricci curvature bound on more abstract spaces. When we say "Ricci lower bound" we mean that there exists a $K \in \mathbb{R}$ such that $\mathrm{Ric}_p(v, v) \geq (n-1)Kg_p(v, v)$ $\forall x \in M, v \in T_pM$ and write $\mathrm{Ric}(M) \geq (n-1)K$.

There are various reasons to try to extend notions of curvature from smooth Riemannian manifolds to more general spaces. Let's consider the sectional curvature. It naturally controls the behavior of the distance along geodesics. For instance, the lower bound $\sec(M) \geq K$ for $K \in \mathbb{R}$ is equivalent to that every geodesic triangle in M is "thicker" than the triangle with the same side lengths in the 2-dimensional space form of constant sectional curvature K, known as Toponogov's theorem ([9]). It turns out, that this triangle comparison condition also makes sense in slightly more general spaces, the **geodesic spaces**. These spaces are just metric spaces in which the distance between two points equals the *minimum* of the lengths of curves joining the points. Geodesic spaces with "sectional curvature $\geq K$" are called *Alexandrov spaces* and are deeply investigated from both geometric and analytic viewpoints ([23], [27], [8]).

In view of Alexandrov's work, it is natural to ask whether there are metric space versions of other types of Riemannian curvature, such as Ricci curvature. Since Ricci curvature is just the trace of sectional curvature, it holds less information and controls only the behaviour of the measure $\mathrm{vol}_{\mathrm{g}}$. For example, one of the most important theorems in comparison Riemannian geometry, *Bishop-Gromov volume comparison theorem*, asserts that a lower Ricci curvature bound implies an upper bound of the volume growth. Also, positive lower Ricci bounds provide upper bounds on the diameter of M, due to *Bonnet-Myers theorem.* These two theorems give nice intuition of how spaces with lower Ricci curvature bounds look like and will be proved in the sequence, using Jacobi fields, just to capture the idea that Jacobi fields are controlled by Ricci curvature. Although bounding Ricci curvature from below is essential in many analytic applications, how to characterize such spaces without using differentiable structure had been a long standing important problem.

Our ultimate goal is to present a way to define a notion of *Ricci curvature* on metric spaces, which will be independent of the Riemannian setting and the dimension of the manifold, in the same fashion as with Alexandrov spaces. But, in order to talk about volumes, a measure has to be included, therefore introducing **metric measure spaces**. A first approach would be to "reverse" the Bishop-Gromov theorem, or something slightly more general. The Bishop-Gromov volume comparison theorem can be regarded as a concavity estimate of $\mathrm{vol}_{\mathrm{g}}^{1/n}$ along the contraction of a ball to its center, thus it is a special case of the well known, *Brunn-Minkowski inequality (BMI).*

Moreover, this contraction can be seen as the transportation of the measure of a ball to that of balls of smaller radii. This interpretation indicates that the tools of Optimal Transport theory, whose object of study is the optimal transportation of measures on metric measure spaces, should play a key role in understanding Ricci curvature. This approach was pioneered by Lott - Villani [22], von Renesse - Sturm [17] and Sturm [14], eventually leading to the definition of the so-called $CD(K, N)$ and $RCD(K, N)$ ($N \in [0, +\infty]$) spaces, which should be understood as metric measure spaces that satisfy a lower Ricci curvature bound and upper dimension bound, in a measure-theoretic sense. These spaces have been the object of intense research activity over the last 10 years, leading to generalizations of many classical results in Riemannian geometry involving lower Ricci curvature bounds, for instance the theorem of Bonnet-Myers [15], the Splitting Theorem [16] and many more.

Our approach will follow Ohta's ([21]) observation that we can prove BMI in the Euclidean case, using optimal transport between uniform distributions. So it would be logical if we study optimal transport between probability measures, in the general case. Since BMI and Ricci curvature both depend on the dimension of the manifold, we must find an *infinitesimal version* of BMI. After developing the main theory of optimal transport on Riemannian manifolds, according to [30], we follow very closely the work of [31] in which they prove an infinitesimal version of the BMI, called *Jacobian inequality*, through the differentiation of the optimal transport map.

If the reader has never heard of optimal transport, they can visualize the following problem: let's say that we're given a pile of sand with which we must fill up a hole of the same volume (which we will assume it's 1). Moving the sand around requires some effort, which we wish to minimize overall. It's, literally, a problem of *optimal transportation* of sand. We will model the sand and the hole by Borel probability measures μ, ν defined on some *complete and separable metric measure spaces* X and Y, respectively. We denote these relations by $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$. The effort is modeled by some measurable **cost function** $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. We shall model the **transport plan** to be a measure $\pi \in \mathscr{P}(X \times Y)$, so that $d\pi(x, y)$ denotes the mass transferred from location x to location y, with the natural constraint that all the mass taken from point x coincide with $d\mu(x)$ and all the mass transferred to point y coincides with $d\nu(y)$. This is described by

$$\int_Y d\pi(x, y) = d\mu(x), \quad \int_X d\pi(x, y) = d\nu(y)$$

or, more rigorously,

$$\pi(A \times Y) = \mu(A), \quad \pi(X \times B) = \nu(B)$$

for all Borel sets $A \subseteq X$, $B \subseteq Y$. Equivalently, we require that

$$\int_{X\times Y} \varphi(x) + \psi(y) d\pi(x, y) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y)$$

for every $\varphi \in L^1(\mu)$, $\psi \in L^1(\nu)$. We denote the set of measures π as above by $\Pi(\mu, \nu)$, which is always a non-empty set, since $\mu \otimes \nu \in \Pi(\mu, \nu)$, which amount for the most inefficient transportation of sand, since any piece of sand is distributed over the entire hole, proportionally to the depth, regardless of its location. **Kantorovich's problem** asks for minimization of the **total transportation cost**:

$$\mathscr{I}(\pi) := \int_{X\times Y} c(x, y) d\pi(x, y)$$

over $\pi \in \Pi(\mu, \nu)$. If a $\tilde{\pi} \in \Pi(\mu, \nu)$ exists such that

$$\mathscr{I}(\tilde{\pi}) = \inf_{\pi \in \Pi(\mu, \nu)} \mathscr{I}(\pi)$$

we will be calling it **optimal transference plan**.

Kantorovich's problem is a relaxed version of the original optimal transport problem, proposed by Monge in the 18^{th} century, which is the same as Kantorovich's, except one thing: it is additionally required that no mass be split. In terms of transference plans, it means that we ask for π to have the special form:

$$d\pi(x, y) = d\pi_F(x, y) = d\mu(x)\delta_{F(x)}(y)$$

where $F : X \to Y$ is a measurable map and δ_{x_0} denotes the Dirac measure on x_0. The measure π_F is characterized by the property

$$\int_{X \times Y} c(x, y) d\pi_F(x, y) = \int_X c(x, F(x)) d\mu(x)$$

so that the total transportation cost takes the form

$$\mathscr{I}(F) := \mathscr{I}(\pi_F) = \int_X c(x, F(x)) d\mu(x).$$

Furthermore, the condition $\pi_F \in \Pi(\mu, \nu)$ translates into:

$$\int_X (\varphi(x) + \psi \circ F(x)) d\mu(x) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y)$$

which turns into

$$\int_X \psi \circ F(x) d\mu(x) = \int_Y \psi(y) d\nu(y)$$

for every $\psi \in L^1(\nu)$ such that $\psi \circ F \in L^1(\nu)$. In terms of Borel sets, this condition can be written as

$$\nu(A) = \mu(F^{-1}(A)) := F_{\#}\mu(A)$$

for any Borel set $B \subseteq Y$. When

$$F_{\#}\mu = \nu$$

is satisfied we will abusively write $F \in \Pi(\mu, \nu)$ and say that **F pushes forward $\boldsymbol{\mu}$ to $\boldsymbol{\nu}$**. Eventually, **Monge's problem** asks for minimization of

$$\mathscr{I}(F) = \int_X c(x, F(x)) d\mu(x)$$

over all Borel maps $F \in \Pi(\mu, \nu)$.

These problems have been solved through the development of very interesting and sophisticated tools in measure theory and convex analysis and we redirect the reader to sources like [1], [2], [29] and [5] to learn more about the classical approach. Here we will explore the topic in the case where X and Y are compact subsets of a complete Riemannian manifold and μ and ν are absolutely continuous with respect to $\mathrm{vol}_{\mathrm{g}}$ and compactly supported in X and Y, respectively. We will solve Monge's problem through a *duality technique*, which is a common move in these types of problems. Moreover, we will give an explicit form of the transport map F and find a way to differentiate it. Its Jacobian will satisfy a change of variable formula and the infinitesimal BMI we talked about earlier. Note that this inequality reflects the relationship of optimal maps with volumes. It's more or less obvious why the Riemannian setting is beneficial. We need to find a way to "move in an optimal way" but this is exactly what geodesics describe. Thus, it's no surprise that the map F has something to do with the exponential map. In fact, we shall find a special type of good function φ, a *c-concave* function, such that its gradient $\nabla\varphi(x)$ at x dictates in what direction and how much we should move point x. As to why these functions are important as well as do the job is based in the duality technique and will become clear later.

It turns out that this Jacobian inequality isn't something that can be generalized into a metric measure space. And that's where *relative entropy* comes into play, a functional on probability measures. In many areas of science, entropy is considered to be a measure of information. In our case, entropy holds information about the volumes. In particular, we will prove that in case we have a uniform probability distribution μ on A we will also have

$$\mathrm{Ent}(\mu) = -\log \mathrm{vol}_{\mathrm{g}}(A)$$

which shows that as the volume of A tends to 0 Ent blows up to $+\infty$ but as we get larger sets Ent becomes more and more negative. Also, we will show that the more a measure is concentrated the bigger its entropy, i.e. $\mathrm{supp}(\mu) \subseteq \mathrm{supp}(\nu) \Rightarrow \mathrm{Ent}(\mu) \geq \mathrm{Ent}(\nu)$. From there, we will follow closely the work in [17] and find an equivalent condition of lower Ricci bound, without involving Riemannian structure. Our approach is in terms of a geodesic space $W_2 := (\mathscr{P}^2(M), d_{W_2})$ canonically associated to our manifold M. Here, $\mathscr{P}^2(M)$ is the set of Borel probability measures that have finite second moment:

$$\int_M d^2(x, y)d\mu(y) < +\infty$$

while d_{W_2} is the so-called **Wasserstein distance**, which is defined to be

$$d_{W_2}(\mu_0,\, \mu_1) := \sqrt{\inf_{\pi \in \Pi(\mu_0,\, \mu_1)} \int_{M \times M} d(x,\, y) d\pi(x,\, y)}.$$

These spaces are called **Wasserstein spaces** and are well studied ([1], [2], [5], [29]), so we won't get into details, as we will drift off our main subject. We will just mention, right now, that they are geodesic spaces whenever M is a geodesic space (not even a manifold) and that its geodesics are completely determined by the optimal transport along the geodesics of M. We urge the reader to keep the last claim in mind while reading the proof of the main theorem. If the reader wants to see a proof of this claim, they can check any of the above references.

The key ingredient is the relation between entropy and the Jacobian of the optimal transport map. If the Ricci curvature is bounded below by $(n-1)K$ Bishop-Gromov volume comparison controls the Jacobian in a way that produces the Jacobian inequality. This concavity estimate for the Jacobian implies K-*convexity* for entropy. Let (X, d) be a geodesic space, $K \in \mathbb{R}$ a number and $U : X \to \mathbb{R}$ a function. We say that U is ***K*-convex** iff for every geodesic $\gamma : [0,\, 1] \to X$ with $U(\gamma(0)), U(\gamma(1)) < \infty$ we have for every $t \in [0,\, 1]$:

$$U(\gamma(t)) \leq (1-t)U(\gamma(0)) + tU(\gamma(1)) - \frac{K}{2}(1-t)td^2(\gamma(0), \gamma(1)).$$

Here, we would like to make a convention. K-convexity on our space, W_2, will have the additional requirement that the geodesic γ has compactly supported endpoints. On the other hand, we will perform optimal transport between two uniform probability distributions to produce a geodesic on the Wasserstein space between them. Then the K-convexity of entropy on this geodesic will recover the lower Ricci bound $(n-1)K$.

Since the very definitions of entropy and K-convexity are formulated on (geodesic) metric measure spaces, with the absence of any Riemannian structure we can have a definition for lower "Ricci curvature" bounds on such spaces, called, more generally the **curvature-dimension condition**. We persuade the reader to check [21], [28], [22] to see how much further one can go with this condition.

Chapter 2

Riemannian Geometry

In this section we will recall the theory of Riemannian manifolds to some extent, omitting the too well known facts or some proofs, but proving facts that we use in our main work, which appear less frequently in standard courses. The very last part is the most important, regarding the rest of the text, and we pursue the reader to skip the in between, if they feel comfortable with Riemannian geometry. We will use Einstein's summation convention. Namely, if in a term the same index appears twice, both as upper and a lower index, that term is assumed to be summed over all possible values of that index (usually from 1 to the dimension). For example,

$$a_i b^i := \sum_i a_i b^i, \quad a^{ijkl} b_{il} c_j := \sum_{i,j,l} a^{ijkl} b_{il} c_j$$

2.1 Manifolds

A **topological manifold** is a second countable, Hausdorff topological space such that each point in the space is contained in an open set which is homeomorphic to some open set in $\mathbb{R}^n$. These homeomorphisms are called **(local) charts**. A **smooth manifold** is a topological manifold equipped with a collection of local charts (which is called an **atlas**), denoted by (U, φ), where $U \subseteq M$ is an open set and φ is a homeomorphism to an open set in $\mathbb{R}^n$ and the union of all U is M, such that the **transition maps** $\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$ are C^∞ for every pair of charts and the atlas is maximal with the above properties.

2.2 Vectors & Bundles

In every point $x \in M$ we define the **tangent space**, a linear space over $\mathbb{R}$, which consists of *tangent vectors* to the manifold, arising as velocities of curves that pass from the point x, interpreted as directions for directional differentiation. If (U, φ) is a chart around x, then the canonical basis of T_xM is denoted by $\{\partial_i : i = 1, 2, \ldots, n\}$ and its action is defined by

$$\partial_i f = \frac{\partial (f \circ \varphi^{-1})}{\partial x^i}\Big|_{\varphi(x)}$$

for $f \in C^\infty(U)$. The disjoint union of all the tangent spaces is called **tangent bundle**, denoted by TM on which the manifold induces a smooth structure, making it a $2n$-dimensional manifold. Each tangent space has a dual space, denoted by T_x^*M and called **cotangent space**, with canonical basis $\{dx^i : i = 1, 2, \ldots, n\}$ with the following action:

$$dx^i(\partial_j) = \delta^i{}_j$$

and their disjoint union, denoted by T^*M, is called **cotangent bundle**.

More generally, a (p, q) **tensor** is a linear form (bounded linear functional) on $\prod_1^p T_xM \times \prod_1^q T_x^*M$ and their set is denoted by $T_p^q(T_xM)$ with canonical basis $\{dx^{i_1} \otimes \cdots \otimes dx^{i_p} \otimes \partial_{j_1} \otimes \cdots \otimes \partial_{j_q}\}_{i_1, \ldots, i_p, j_1, \ldots, j_q}$ where $u \otimes v(x, y) = u(x)v(y)$. The disjoint union of all $T_p^q(T_xM)$, $x \in M$ is denoted by T_p^qM, for short, and is called a **tensor bundle**, which is, of course, a generalization of the (co)tangent bundle. Moreover, a smooth section (a smooth right inverse of the projection function $T_xM \mapsto x$) of T_p^qM is called a **tensor field**. In particular, if $p = 0$, $q = 1$ we call it a **vector field** and if $p = 1$, $q = 0$ we call it a **covector field**, a **1-form** or, simply, a **form**.

2.3 Push forward & Pull back

The **differential at x** of a smooth function $\varphi : M \to N$ is the linear map $(D\varphi)_x : T_xM \to T_{\varphi(x)}N$ which is defined by

$$(D\varphi)_x(v)f = v(f \circ \varphi), \; f \in C^\infty(N)$$

Sometimes, $(D\varphi)_x(v)$ is called the **push forward of v by φ**.

The **pull back by φ on $(p, 0)$ tensors** is the linear map $\varphi^* : T_p^0(T_{\varphi(x)}N) \to T_p^0(T_xM)$ which is defined by

$$(\varphi^*T)(X_1, \dots, X_p) = T((D\varphi)_x(X_1), \dots, (D\varphi)_x(X_p))$$

If φ is a diffeomorphism, then $\varphi^* : T_{\varphi(x)}N \to T_xM$ is defined to be $(D\varphi)_x^{-1}$, so that the **pull back by φ on (p, q) tensors**, $\varphi^* : T_p^q(T_{\varphi(x)}N) \to T_p^q(T_xM)$, is defined from the above alongside with the rule $\varphi^*(T \otimes S) = \varphi^*(T) \otimes \varphi^*(S)$.

2.4 Differential Forms & Integration

In order to define a convenient and coordinate-free way to integrate on manifolds we need to introduce the concept of **differential forms**, i.e. *anti-symmetric, covariant* tensors and their *exterior derivatives*. Every $(p, 0)$ tensor T comes with an anti-symmetric multilinear functional:

$$A(T)(v_1, \dots, v_p) := \frac{1}{k!}\sum_\sigma \mathrm{sign}(\sigma)T(v_{\sigma(1)}, \dots, v_{\sigma(p)})$$

where σ runs over all permutations of $\{v_1, \dots, v_p\} \subset T_xM$ (for $x \in M$). Taking all the stationary points of A, i.e. the anti-symmetric $(p, 0)$ tensors, we construct the *fiber bundle of anti-symmetric p-linear functionals on TM*:

$$\wedge^p T^*M := \{T \in T_p^0 M : A(T) = T\}$$

with canonical basis $\{dx^{i_1} \wedge \cdots \wedge dx^{i_p} : i_1 < \cdots < i_p\}$, where

$$T \wedge S := \frac{(i+j)!}{i!j!}A(T \otimes S)$$

is the **exterior product** of $T \in \wedge^i T^*M$ and $S \in \wedge^j T^*M$. A section of the above bundle is called a ***p*-form**, with the convention that the 0-forms are smooth functions on M, $C^\infty(M)$. Also, this is why a covector is also called a 1-form. Another operation, remaining to be defined, is the **exterior differentiation**, which is a map $d : \wedge^p T^*M \to \wedge^{p+1}T^*M$ defined, in local coordinates, by

$$df = \partial_i f dx^i$$

if $f \in C^\infty(M)$ and by

$$d\omega = \sum_{i_1<\dots<i_p} d\omega_{i_1\dots i_p} \wedge dx^{i_1} \wedge \dots \wedge dx^{i_p}$$

if $\omega = \sum_{i_1<\dots<i_p} \omega_{i_1\dots i_p} dx^{i_1} \wedge \dots \wedge dx^{i_p}$.

We are now ready to define integration (of differential forms) on a smooth *orientable* manifold, i.e. all transition maps have positive definite Jacobian. Let (U, φ) be a chart and ω a n-form on M such that $\omega|_U = f dx^1 \wedge \dots \wedge dx^n$ for $f \in C^\infty(U)$. We define

$$\int_U \omega := \int_{\varphi(U)} f \circ \varphi^{-1} dx^1 \dots dx^n$$

Now, let $\{(U_i, \varphi_i)\}$ be a family of local charts for M such that $\{(U_i, \eta_i)\}$ is a partition of unity. Then, as one expects, we define

$$\int_M \omega = \sum_i \int_{U_i} \eta_i \omega$$

Observe that, at this moment, integration is regarded as a linear functional on the space of smooth functions. Once we provide the manifold with a *Riemannian metric*, we shall choose a special form, the *volume form*, so that the resulting integration will be compatible with the metric.

2.5 Riemannian Manifolds

Now let's turn our focus on the Riemannian setting. On each tangent space T_pM, we assign a *smooth* (as a function of p), *symmetric, positive definite* $(0, 2)$ tensor field $g_p : T_pM \times T_xM \to \mathbb{R}$. The smoothness can be interpreted as follows: if X, Y are two smooth vector fields on an open subset $U \subseteq M$ then $f(p) = g_p(X_p, Y_p)$ is a smooth function on U. This assignment of an inner product $\langle\cdot, \cdot\rangle_p := g_p(\cdot, \cdot)$ on each T_pM is called a **Riemannian metric**. If $(U, (x^1, \dots, x^n))$ is a chart, we can represent the Riemannian metric locally as follows: Let $g_{ij}(p) := \langle\partial_i, \partial_j\rangle_p$. Then for any smooth vector fields $X = X^i\partial_i$ and $Y = Y^j\partial_j$ on U, $\langle X_p, Y_p\rangle_p = g_{ij}(p)X^i(p)Y^j(p)$, so we can write, locally, $g = g_{ij}dx^i \otimes dx^j$. It is clear that the matrix (g_{ij}), which constitutes of smooth functions, is symmetric and positive definite at any p. We denote its inverse by (g^{ij}), so that $g_{ij}g^{jk} = {\delta_i}^k$. If such metric exists we say that (M, g) is a **Riemannian manifold**. It turns out that many metrics exist for a smooth manifold.

2.6 Lengths & Distance

We must give focus to the fact that g is not a *distance metric*, but it induces a natural distance function on M. In order to define this distance, we first need to define the length of a curve. Let $\gamma : [a, b] \to M$ be a smooth immersed parametric curve in M. Then, for any $t \in [a, b]$, $\dot{\gamma}(t) := (d\gamma)_t(\frac{d}{dt}\big|_t)$ is a tangent vector in $T_{\gamma(t)}M$. We shall always assume that the parametrization is *regular*, i.e. $\dot{\gamma}(t) \neq 0$ for all t. We define the **length** of γ as

$$L(\gamma) = \int_a^b \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} dt$$

and sometimes we'll refer to it as the **length functional**. Due to the well-known change of variable formula one easily checks that the length of curve is independent of the choice of regular parametriazations, so that it is well defined. Especially, the length remains the same through isometries, which tells us that any regular curve can be reparametrized so that $|\dot{\gamma}(t)|_{\gamma(t)} := \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} = 1$, through the inverse of the *arc-length function* $s(t) = \int_a^t \sqrt{\langle \dot{\gamma}(\tau), \dot{\gamma}(\tau) \rangle_{\gamma(\tau)}} d\tau$. This parametrization is called the **arc-length parametrization**. All of the above can be easily extended to piecewise smooth curves in M.

We are ready to define the **Riemannian distance function**:

$$d(p, q) = \inf\{L(\gamma) \mid \gamma \in \Gamma_{pq}\}$$

where $\Gamma_{pq} := \{\gamma : [a, b] \to M \mid \gamma \text{ is piecewise smooth and } \gamma(a) = p, \gamma(b) = q\}$. Note that if $M = \mathbb{R}^n$ and $g(\cdot, \cdot) = \langle \cdot, \cdot \rangle$ is the standard inner product, we are talking about the Euclidean distance.

Theorem. *The Riemannian distance function makes M into a metric space.*

Proof. It is easy to check that all of the conditions for a distance function are met, except for that we must have $d(p, q) > 0$ for $p \neq q$. For this purpose we take a chart (U, φ) around q with $p \notin U$. We can apply a linear transform so that $\varphi(q) = 0$ and $\varphi(U) = B_1(0)$. Let $\lambda > 0$ (because of positive definiteness) be the smallest eigenvalue of

$$\overline{g} := (\varphi^{-1})^* g_U$$

at all points in $\overline{B_{1/2}(0)}$ (in the sense that $\overline{g}(v, v) = \lambda \langle v, v \rangle$). Let γ be an arbitrary curve starting from $0 = \varphi(q)$ and ending at some point on $\partial B_{1/2}(0)$

and $\tilde{\gamma}$ be the first piece of γ that sits totally in $B_{1/2}(0)$ (except the endpoint), reparametrized in $[0,\,1]$. Then

$$L_{\overline{g}}(\gamma) \geq L_{\overline{g}}(\tilde{\gamma}) = \int_0^1 \sqrt{\overline{g}(\dot{\tilde{\gamma}}(t),\,\dot{\tilde{\gamma}}(t))}dt \geq$$
$$\geq \sqrt{\lambda}\int_0^1 \sqrt{\langle\dot{\tilde{\gamma}}(t),\,\dot{\tilde{\gamma}}(t)\rangle}dt = \sqrt{\lambda}L(\tilde{\gamma}) \geq \frac{\sqrt{\lambda}}{2}$$

Since any curve from p to q must intersect $\varphi^{-1}(\partial B_{1/2}(0))$ at some point, we conclude that

$$d(p,\,q) \geq \frac{\sqrt{\lambda}}{2} > 0$$

as desired. □

From the above proof we can deduce that any open set in a manifold contains a metric open ball. If we show that the function $f(\cdot) := d(p,\,\cdot)$ is continuous in the manifold topology we have that:

Theorem. *The metric topology on M coincides with the manifold topology on M.*

Proof. We shall prove that the above function f is continuous on M. Let q_n be a sequence of points that converges to q, in the sense that for any k there exists $N = N(k)$ such that for all $n \geq N$, $\varphi(q_n) \in B_{1/k}(0)$. By triangle inequality, we have

$$|f(q_n) - f(q)| \leq d(q,\,q_n)$$

So it suffices to prove that $d(q,\,q_n) \xrightarrow[n\to\infty]{} 0$. To this end, let $(U,\,\varphi)$ and $\overline{g}$ be exactly as in the previous proof and Λ be the greatest eigenvalue of $\overline{g}$ at all points in $\overline{B_{1/2}(0)}$. Let

$$\tilde{\gamma}_n : [0,\,1] \to B_1(0), \quad \tilde{\gamma}_n(t) = t\varphi(q_n)$$

be the line segment from $0 = \varphi(q)$ to $\varphi(q_n)$. Then for $n \geq N$ we have

$$L_{\overline{g}}(\tilde{\gamma}_n) = \int_0^1 \sqrt{\overline{g}(\dot{\tilde{\gamma}}_n,\,\dot{\tilde{\gamma}}_n)}dt \leq \sqrt{\Lambda}\int_0^1 \sqrt{\langle\dot{\tilde{\gamma}}_n,\,\dot{\tilde{\gamma}}_n\rangle}dt = \sqrt{\Lambda}L(\tilde{\gamma}_n) \leq \frac{\sqrt{\Lambda}}{k}$$

Since $(U,\,g_U)$ is isometric to $(B_1(0),\,\overline{g})$, we conclude that

$$L_{g_U}(\gamma_n) = L_{\overline{g}}(\tilde{\gamma}_n) \leq \frac{\sqrt{\Lambda}}{k}$$

where $\gamma_n := \varphi^{-1}(\tilde{\gamma}_n)$ is a curve from q to q_n. It follows that $d(q,\,q_n) \leq \frac{\sqrt{\Lambda}}{k}$ for all $n \geq N$. Thus f is continuous and the theorem is proved. □

2.7 The Riemannian Measure

Using partitions of unity we can define the regular Borel measure on M

$$d\mathrm{vol_g} = \sqrt{\det(g_{ij})}dm$$

where m is the Lebesgue measure on $\mathbb{R}^n$. This definition is inspired by the fact that for K a compact subset in some coordinate patch (U, φ) such that $\varphi(K)$ is measurable, the quantity

$$\int_{\varphi(K)} \sqrt{\det(g_{ij})} \circ \varphi^{-1} dm$$

is independent of the choice of local charts and represents the volume of the set K. We will integrate functions on M with respect to $d\mathrm{vol_g}$ and every widely known norm, such as L_p norms will be defined by the obvious formula

$$||f||_p^p := \int_M |f|^p d\mathrm{vol_g}\,.$$

2.8 Connections & Curvatures

In order to differentiate vector fields we must "connect" nearby tangent spaces. Therefore we define a **connection** as a bilinear map

$$\nabla : T_0^1 M \times T_0^1 M \to T_0^1 M$$

denoted by $(X, Y) \mapsto \nabla_X Y$ which satisfies the following properties:

1. $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$
2. $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$
3. $\nabla_X(fY) = f\nabla_X Y + X(f)Y$

for every $X, Y, Z \in T_0^1 M$ and $f, g \in C^\infty M$. It can be proved that there exists a unique correspondance which associates to a vector field X along a smooth curve $\gamma : I \to M$ another vector field along γ, denoted by $\frac{DX}{dt}$ or more frequently by X', called the **covariant derivative of X along $\boldsymbol{\gamma}$**, such that:

- $\frac{D}{dt}(X+Y) = \frac{DX}{dt} + \frac{DY}{dt}$
- $\frac{D}{dt}(fX) = \frac{df}{dt}X + f\frac{DX}{dt}$
- If $X(t) = \tilde{X}(\gamma(t))$, for $\tilde{X} \in T_0^1 M$, then $\frac{DX}{dt} = \nabla_{\dot{\gamma}}\tilde{X}$

where X, Y are vector fields along γ and $f \in C^\infty(I)$. The third bullet makes sense, since $\nabla_X Y(p)$ depends on the value of $X(p)$ and the values of $\tilde{X}$ along a curve, tangent to X at p, since part (3) of the definition allow us to show that the notion of connection is actually a local notion. In articular, $\nabla_X Y(p)$ depends only on $X^i(p)$, $Y^j(p)$ and the derivatives $X(Y^j)(p)$ of Y^j by X. A well known theorem states that any Riemannian manifold has a *unique* connection satisfying two more (good) properties:

(1) **Symmetric**: $\nabla_X Y - \nabla_Y X = [X, Y]$

(2) **Metric compatible**: $\frac{d}{dt}\langle X, Y\rangle = \langle X', Y\rangle + \langle X, Y'\rangle$

where $[\cdot, \cdot] : T_0^1 M \times T_0^1 M \to T_0^1 M$ is the **Lie bracket**, defined by

$$[X, Y](f) := X(Y(f)) - Y(X(f))$$

for every $f \in C^\infty M$. This connection is called the **Levi-Civita connection** and we shall always use this by default.

The notion of connection is equivalent to the notion of **parallel transport along curves**. Let's call a vector field X along a curve $\gamma : I \to M$ **parallel** if $X' \equiv 0$. Let $V_0 \in T_{\gamma(t_0)}M$ for some $t_0 \in I$. Then there exists a unique parallel field V along γ, such that $V(t_0) = V_0$. This field is what we call parallel transport along γ. It turns out that if we define a map

$$P : T_{\gamma(t)}M \to T_{\gamma(t_0)}M$$

where $P(v) =$ parallel transported v from $\gamma(t)$ to $\gamma(t_0)$, then it is a linear isometry. Also, if γ is an integral curve, of a smooth vector field X, through p, i.e. $\gamma(t_0) = p$ and $\dot{\gamma}(t) = X(\gamma(t))$ and Y is another smooth vector field then one has

$$\nabla_X Y(p) = \frac{d}{dt}\Big|_{t=t_0} P(Y(\gamma(t))).$$

For a 1-form ω define its covariant derivative with respect to a vector field X as

$$(\nabla_X \omega)(Y) = X(\omega(Y)) - \omega(\nabla_X Y).$$

Now, the connection can be extended on any $(p,\, q)$-tensor field as follows: let $X_1,\, \ldots,\, X_p$ be vector fields and $\omega_1,\, \ldots,\, \omega_q$ be 1-forms and T be a $(p,\, q)$-tensor field. The **covariant derivative of T with respect to X** is defined by

$$\begin{aligned}(\nabla_X T)(X_1,\, \ldots,\, X_p,\, \omega_1,\, \ldots,\, \omega_q) = X(T(X_1,\, \ldots,\, X_p,\, \omega_1,\, \ldots,\, \omega_q))\\ - \sum_{i=1}^{p} T(X_1,\, \ldots,\, \nabla_X X_i,\, \ldots,\, X_p,\, \omega_1,\, \ldots,\, \omega_q)\\ - \sum_{j=1}^{q} T(X_1,\, \ldots,\, X_p,\, \omega_1,\, \ldots,\, \nabla_X \omega_j,\, \ldots,\, \omega_q)\end{aligned}$$

If T is a smooth function f then $\nabla_X f = df(X)$, but when we write ∇f we mean the vector field that satisfies $df(X) = \langle \nabla f,\, X\rangle$.

Now, we can define the **second covariant derivative**. Let $X,\, Y,\, Z$ be smooth vector fields on M. by definition ∇Z is a $(1,\, 1)$-tensor field satisfying

$$(\nabla Z)(Y) = \nabla_Y Z.$$

Hence

$$\nabla_X \big[(\nabla Z)(Y)\big] = \nabla_X(\nabla_Y Z).$$

A differentiation should follow the Leibniz rule, so that we would get:

$$(\nabla_X(\nabla Z))(Y) + (\nabla Z)(\nabla_X Y) = \nabla_X(\nabla_Y Z)$$

$$\Rightarrow (\nabla_X(\nabla Z))(Y) = \nabla_X(\nabla_Y Z) - (\nabla Z)(\nabla_X Y) = \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z$$

and we define the second covariant derivative of a smooth vector field Z as a $(2,\, 1)$-tensor field, denoted by $\nabla^2 Z$ and defined by

$$\nabla^2_{X,Y} Z = (\nabla_X(\nabla Z))(Y) = \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z.$$

If T is a smooth function f, then $\nabla^2_{X,\,Y} f := \operatorname{Hess} f(X,\, Y)$ is called the **Hessian of f**.

The **Riemann curvature tensor** is a $(3,\, 1)$ tensor field defined by

$$\begin{aligned}R(X,\, Y)Z &= \nabla^2_{X,\,Y} Z - \nabla^2_{Y,\,X} Z\\ &= \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X,Y]} Z\end{aligned}$$

for all vector fields X, Y, Z. It can also be written as a $(4, 0)$ tensor field as $R(X, Y, Z, W) = g(R(X, Y)Z, W)$. The $(2, 0)$ **Ricci curvature tensor** is the trace of the curvature tensor:

$$\mathrm{Ric}(X, Y) = \sum_{i=1}^{n} R(X, e_i, e_i, Y)$$

for $X, Y \in T_pM$, where $\{e_i \mid i = 1, \ldots, n\}$ is an orthonormal basis for T_pM. The **scalar curvature** is just the trace of the Ricci curvature tensor:

$$\mathrm{scal}(M) = \sum_{j=1}^{n} \mathrm{Ric}(e_j, e_j)$$

and last, but not least, the **sectional curvature** of the plane $\Pi \subseteq T_pM$ spanned by $X, Y \in T_pM$ is defined as

$$\sec(\Pi) = \sec(X, Y) = \frac{R(X, Y, Y, X)}{g(X, X)g(Y, Y) - g(X, Y)^2}.$$

Observe that if $\{e_1, \ldots, e_n\}$ is an orthonormal basis for T_pM one has

$$\mathrm{Ric}(e_j, e_j) = \sum_{i=1}^{n} \sec(e_j, e_i)$$

for $j = 1, \ldots, n$. So Ricci curvature is like a mean of sectional curvatures in every direction, so it has to have some relationship with volume. We will see that this is the case through the Bishop-Gromov volume comparison theorem and its consequences on optimal mass transport. Lastly, sectional curvature is closely related to the curvature tensor since it holds most of its information as one can see from the following:

Proposition. *For all vector fields* X, Y, Z, W *on* M

$$R(X, Y, Z, W) = K\big(\langle X, W\rangle\langle Y, Z\rangle - \langle Y, W\rangle\langle X, Z\rangle\big) \Leftrightarrow \sec(M) \equiv K$$

where $\sec(M) \equiv K$ *means* $\sec_p \equiv K$ *for every* $p \in M$*.*

2.9 Geodesics & Exponential map

As usual a **geodesic** is a curve γ on M which satisfies $\nabla_{\dot\gamma}\dot\gamma = 0$. This is expressed in local coordinates as

$$\frac{d^2x^i}{dt^2} + \Gamma^i_{jk}(c(t))\frac{dx^j}{dt}\frac{dx^k}{dt} = 0$$

where Γ^i_{jk} denote the Christoffel symbols, defined by $\nabla_{E_j} E_k = \Gamma^i_{jk} E_i$. This is a system of second order, quadratic, nonlinear ordinary differential equations and by standard theory of O.D.E., for any point $p \in M$ and $v \in T_x M$, there exists a unique geodesic $\gamma : [0, t_0] \to M$ for some $t_0 > 0$, such that $\gamma(0) = p$ and $\dot\gamma(0) = v$. If $t_0 < \infty$, we consider the curve $\gamma_0(t) := \gamma(t_0 t)$. Clearly $\gamma_0(0) = p$ and $\dot\gamma_0(0) = t_0 v$, while the curve is defined for $t \in [0, 1]$ and $\nabla_{\dot\gamma_0}\dot\gamma_0 = t_0^2 \nabla_{\dot\gamma}\dot\gamma = 0$. Since v is an arbitrary vector, the theory of O.D.E. an prove that there exists a maximal open set $\mathcal{D}_p \subseteq T_p M$ such that $0 \in \mathcal{D}_p$ and for each $v \in \mathcal{D}_p$, there exists a unique geodesic $\gamma_0 : [0, 1] \to M$, such that $\gamma_0(0) = p$ and $\dot\gamma_0(0) = v$.

This fact enables us to define the **exponential map at $\mathbf{p} \in \mathbf{M}$** as a map $\exp_p : \mathcal{D}_p \to M$, defined by

$$\exp_p(v) = \gamma(1)$$

where $\gamma : [0, 1] \to M$ is the geodesic with $\gamma(0) = p$ and $\dot\gamma(0) = v$. It is obvious that $\exp_p(tv) = \gamma(t)$ holds that gives the well known fact that $(D \exp_p)_0 = \mathrm{Id}_{T_p M}$ for the linear map $(D \exp_p)_v : T_v(T_p M) \to T_{\exp_p v} M$. Now, the inverse function theorem states that $\exp_p$ is a local diffeomorphism.

At each $p \in M$ find an open set $\mathcal{D}_p \subseteq T_p M$ such that $\exp_p : \mathcal{D}_p \to \exp_p(\mathcal{D}_p) := \mathcal{U}_p$ is a diffeomorphism and call $\mathcal{U}_p$ a **normal neighbourhood of p**. If $\overline{B^p_\varepsilon(0)} \subseteq \mathcal{V}_p$ then we call $\exp_p B^p_\varepsilon(0) = B_\varepsilon(p)$ the **normal** (or **geodesic**) **ball with center p and radius $\boldsymbol{\varepsilon}$>0**. Now, let $\{e_1, \ldots, e_n\}$ be an orthonormal basis for $T_p M$, i.e. $g_p(e_i, e_j) = \delta_{ij}$. Then the local chart $(\mathcal{U}_p, \exp_p^{-1})$ is called **local normal coordinates around p**. It can be proved that for any $p \in M$ there exists a neighbourhood W of p so that it is a normal neighbourhood of each of its points, such neighbourhood is called **totally normal neighbourhood**. If we consider a geodesic $\gamma : [0, 1] \to B$ where B is a geodesic ball with center $\gamma(0) = p$. Then any other smooth curve δ that joins $\gamma(0)$ to $\gamma(1)$ has the same or greater length, while equality holds if and only if $\gamma([0, 1]) = \delta([0, 1])$. So geodesics minimize locally the length functional. This property is not global, for example on the 2-dimensional sphere any geodesic from the north pole stops being minimizing once it passes the south pole. Later, using some variational calculus, we will talk about how minimizing curves must be geodesics and that for every geodesic is minimizing if we restrict their domain. It is a well known fact that in these coordinates one has at $p \in M$ that $g_{ij}(p) = \delta_{ij}$ and $\Gamma^i_{jk}(p) = 0$. In particular the Taylor expansion at $x = p$ gives

$$g_{ij}(x) = \delta_{ij} + \frac{1}{3} R_{iklj}(p) x^k x^l + O(|x|^3)$$

and

$$R_{ijkl}(p) = \frac{1}{2}(\partial_i\partial_l g_{jk} + \partial_j\partial_k g_{il} - \partial_i\partial_k g_{jl} - \partial_j\partial_l g_{ik})$$

which has as a consequence an estimate for the **volume element**

$$\sqrt{\det(g_{ij})} = 1 - \frac{1}{6}\operatorname{Ric}_{kl}(p)x^k x^l + O(|x|^3)$$

which is a sign that Ricci controls more than just volumes, since

$$\operatorname{vol}_{\mathrm{g}}(B_r(p)) = \omega_n r^n \Big(1 - \frac{\operatorname{scal}(p)}{6(n+2)} r^2 + O(|r|^3)\Big)$$

where $B_r(p) := \{q \in M \mid d(p,\, q) < r\}$ and ω_n denotes the euclidean volume of the unit n-ball. These two equations provide a geometric interpretation for Ricci and scalar curvatures, while for the sectional curvature it can be proven that, for unit vectors $v,\, w \in T_xM$ with (non-oriented) angle θ one has

$$d(\exp_x(tv),\, \exp_x(tw)) = \sqrt{2(1-\cos\theta)}\,t\left(1 - \frac{\cos^2(\theta/2)}{6}\sec(v,\, w)t^2 + O(t^4)\right)$$

which tells us that the sectional curvatures infinitesimally measure the tendency of geodesics to converge (>0) or diverge (<0).

2.10 Hopf-Rinow Theorem & Gauss' Lemma

The manifolds of our work need to be complete as a metric space with respect to the distance in order to exploit some useful properties of theirs, arising from the well-known Hopf Rinow theorem:

Theorem (Hopf-Rinow). *Let $(M,\, g)$ be a Riemannian manifold with geodesic distance d. The following are equivalent:*

(i) M is a complete metric space

(ii) M is geodesically complete (i.e. geodesics ar defined on all of $\mathbb{R}$)

(iii) $\exists p \in M$ such that $\exp_p$ is defined on all of T_pM

(iv) $\forall p \in M$ the $\exp_p$ is defined on all of T_pM

(v) Closed and bounded subsets of M are compact

while any of the above implies that any two points can be joined by a minimizing geodesic.

We will not prove the above theorem since its proof is enormous, out of our focus field and anyone who's interested can find it in multiple introductory coursebooks, though we will refer to one extremely important tool that not only is needed for the discussed proof but it is of separate interest at any rate.

Lemma (Gauss' Lemma)**.** *Let* $u,\, v \in T_pM$ *and* $\gamma(t) = \exp_p tv$*. Then*

$$\langle (D\exp_p)_{sv}(v),\, (D\exp_p)_{sv}(u)\rangle_{\gamma(t)} = \langle v,\, u\rangle_p.$$

In particular, $\dot{\gamma}(t)$ *is orthogonal to a smooth geodesic sphere centered at* p *with radius* $d(p,\, \gamma(t))$*.*

An immediate corollary of this lemma is that there exists $\varepsilon > 0$ such that $\exp_p$ is a diffeomorphism from the ball $B^p_\varepsilon(0) \subseteq T_pM$ onto the normal ball $B_\varepsilon(p)$, while for any $q \notin B_\varepsilon(p)$ there exists $z \in \partial B_\varepsilon(p)$ such that

$$d(p,\, q) = d(p,\, z) + d(z,\, q) = \varepsilon + d(z,\, q).$$

Moreover, for any $v \in T_pM$ the geodesic $\gamma : [0,\, \varepsilon] \to M$, defined by $\gamma(t) = \exp_p tv$, is minimizing. From now on, we will assume that any manifold M is complete.

2.11 Conjugate Points & Cut Locus

In order to measure how broad is the region where $\exp_p$ is a diffeomorphism we define the **injectivity radius at p** $\in$ **M**, denoted by $\mathrm{inj}(p)$, as the supremum of the radii of balls centered at $0 \in T_pM$ on which $\exp_p$ is a diffeomorphism. It can be proved, that inj as a function $M \to \mathbb{R}$ is continuous. The infimum of $\mathrm{inj}(p)$ over all $p \in M$ is called **the injectivity radius of M**. A kind of element that exists beyond the injectivity radius at p is a **conjugate point of p**: if there exists $v \in T_pM$ such that $q = \exp_p v$ and the linear map $(D\exp_p)_v : T_v(T_pM) \to T_qM$ has non-zero kernel, then q is called a conjugate point of p. If we consider the geodesic $\gamma(t) = \exp_p tv,\, t \geq 0$. If $q = \exp_p t_0 v$ is a conjugate point of p then γ is not distance minimizing for $t \in [0,\, t_0 + \varepsilon]$ for every $\varepsilon > 0$. That inspires the definition of the **cut locus of p**:

$$\begin{aligned}\mathrm{cut}(p) := \{\exp_p v \in M \mid\ & \exp_p tv \text{ is minimizing for } t \in [0,\, 1] \\ & \text{but stops being for } t \in [0,\, 1+\varepsilon]\ \forall \varepsilon > 0\}\end{aligned}$$

Every point in $q \in \mathrm{cut}(p)$ is called a **cut point of p**. There is a characterization of the cut locus which makes use of conjugate points:

Theorem. *Suppose that $\gamma(t_0)$ is the cut point of $p = \gamma(0)$ along γ. Then at least one of the following holds:*

(1) $\gamma(t_0)$ is the first conjugate point of $\gamma(0)$ along γ

(2) there exists a geodesic $\delta \neq \gamma$ from p to $\gamma(t_0)$ such that $L(\delta) = L(\gamma)$

Conversely, if one of the above is satisfied, then there exists $\tau \in (0, t_0]$ such that $\gamma(\tau)$ is the cut point of p along γ.

We have three important corollaries produced by this theorem:

- $q \in \mathrm{cut}(p) \Leftrightarrow p \in \mathrm{cut}(q)$
- $q \notin \mathrm{cut}(p) \Rightarrow$ there exists a unique minimizing geodesic joining p to q
- $\forall p \in M$ $\mathrm{cut}(p)$ is closed

It is immediate that

$$\mathrm{inj}(p) = d(p,\, \mathrm{cut}(p))$$

but what about the points that realize this distance? Suppose $q \in \mathrm{cut}(p)$ such that $d(p,\, q) = \mathrm{inj}(p)$. Then at least one of the following is true:

(1) q is a conjugate point of p along a minimizing geodesic joining p to q

(2) There exist exactly two minimizing geodesics γ_1, γ_2 joining p to q. In addition, $\dot{\gamma}_1(l) = \dot{\gamma}_2(l)$ and $l = d(p,\, q)$

2.12 Variations & Jacobi Fields

A **variation** of a smooth curve γ is a smooth map $f : (-\varepsilon,\, \varepsilon) \times [0,\, 1] \to M$ so that $f(0,\, t) = \gamma(t)$ (we use 0 and 1 instead of a and b for simplicity). Sometimes we write $\gamma_s(t) = f(s,\, t)$. Observe that $\frac{\partial f}{\partial t}(s,\, t) = \dot{\gamma}_s(t)$. We will call $V(t) := \frac{\partial f}{\partial s}(0,\, t)$ the **variational field of f along $\boldsymbol{\gamma}$**. A **proper** variation is a variation that doesn't move endpoints, while a variation is called **geodesic** if γ_s is a geodesic for every $s \in (-\varepsilon, \varepsilon)$. It is an immediate consequence of the definitions that

$$\frac{D}{ds}\frac{\partial f}{\partial t} = \frac{D}{dt}\frac{\partial f}{\partial s}$$

where $\frac{D}{ds}$, $\frac{D}{dt}$ denote the covariant differentiation along curves. From the **energy functional**

$$E(\gamma) = \frac{1}{2}\int_0^1 |\dot{\gamma}(t)|^2_{\gamma(t)} dt$$

we get the **first variation of energy formula**:

$$\frac{dE(\gamma_s)}{ds}(0) := -\int_0^1 \langle V(t),\, \nabla_{\dot{\gamma}}\dot{\gamma}\rangle dt + \langle V(1),\, \dot{\gamma}(1)\rangle - \langle V(0),\, \dot{\gamma}(0)\rangle$$

which proves that the functional's critical points are exactly the geodesics and from the **second variation of energy formula**:

$$\begin{aligned}\frac{d^2E(\gamma_s)}{ds^2}(0) := -\int_0^1 \langle V(t),\, V''(t) + R(V,\, \dot{\gamma})\dot{\gamma}(t)\rangle dt \\ + \Big[\langle V(t),\, V'(t)\rangle + \langle \nabla_{V(t)}V(t),\, \dot{\gamma}(t)\rangle\Big]_{t=0}^{t=1}\end{aligned}$$

we get that for a sufficiently small interval a geodesic will be a minimizing one, i.e. it minimizes the distance between its endpoints. That's why sometimes we will refer to the distance function as **geodesic distance**. Note that this is the case because a curve minimizes the length functional if and only if it minimizes the energy functional and it has constant speed. It can be proved that an alternative form of the second formula is the following

$$\begin{aligned}\frac{d^2E(\gamma_s)}{ds^2}(0) = \int_0^1 \langle V'(t),\, V'(t)\rangle - \langle V(t),\, R(V,\, \dot{\gamma})\dot{\gamma}(t)\rangle dt \\ + \Big[\langle \nabla_{V(t)}V(t),\, \dot{\gamma}(t)\rangle\Big]_{t=0}^{t=1}\end{aligned}$$

from which, the definition of the **index form** of a geodesic is inspired:

$$\begin{aligned}I(X,\, Y) :&= \int_0^1 \langle X'(t),\, Y'(t)\rangle - R(X,\, \dot{\gamma},\, \dot{\gamma},\, Y)(t) dt \\ &= \Big[\langle X'(t),\, Y(t)\rangle\Big]_{t=0}^{t=1} - \int_0^1 \langle X''(t) + R(X,\, \dot{\gamma})\dot{\gamma}(t),\, Y(t)\rangle dt\end{aligned}$$

where X, Y are smooth vector fields along γ. Note that I is symmetric and bilinear and if the variation is proper the last terms in the first two formulas vanish and in particular

$$\frac{d^2E(\gamma_s)}{ds^2}(0) = I(V,\, V) = -\int_0^1 \langle V(t),\, V''(t) + R(V,\, \dot\gamma)\dot\gamma(t)\rangle dt$$

which inspires, in turn, the definition of the **Jacobi fields** along geodesics. These are smooth vector fields J along a curve γ satisfying

$$J''(t) + R(J(t),\, \dot\gamma(t))\dot\gamma(t) = 0.$$

Note that if X is a Jacobi field then

$$I(X,\, Y) = \Big[\langle X'(t),\, Y(t)\rangle\Big]_{t=0}^{t=1}$$

which means that the index form is determined by the information on the endpoints. Actually one sees easily that X is a Jacobi field if and only if $I(X,\, Y) = 0$ for all Y that are zero on their endpoints. Consider the set of all smooth fields along a geodesic γ that vanish at their endpoints and denote it by $\mathcal{V}$. The index form is closely related to conjugate points in the sense that

(1) $\gamma(0)$ has no conjugate point along $\gamma \Leftrightarrow I > 0$ on $\mathcal{V}$

(2) $\gamma(1)$ is the first conjugate of $\gamma(0)$ along $\gamma \Leftrightarrow I \geq 0$ on $\mathcal{V}$ and $I(X,\, X) = 0$ for some $X \in \mathcal{V}$

(3) $\gamma(0)$ has a conjugate point midway along $\gamma \Leftrightarrow I(X,\, X) < 0$ for some $X \in \mathcal{V}$

The index form can help us see that Ricci curvature controls more than just volumes:

Theorem (Bonnet-Myers Theorem)**.** *Let $(M,\, g)$ be a geodesically complete Riemannian manifold whose Ricci curvature satisfies* $\mathrm{Ric}(M) \geq (n-1)K$ *for a $K > 0$. Then M is compact and its diameter is bounded by*

$$\operatorname{diam} M \leq \frac{\pi}{\sqrt{K}}$$

Proof. Let $p,\, q \in M$ and $\gamma : [0,\, 1] \to M$ a minimizing geodesic joining p to q. Suppose that

$$l := L(\gamma) \geq \frac{\pi}{\sqrt{K}}$$

where L is the length functional. Complete $e_1 := \frac{\dot\gamma(0)}{l}$ into an orthonormal basis $\{e_1, \dots, e_n\}$ of T_pM and consider their parallel translation $\{e_i(t)\}$ along γ. For $i = 2, \dots, n$ define

$$V_i(t) := \sin(\pi t) e_i(t)$$

which has the property that $V_i(0) = V_i(1) = 0$. Then

$$I(V_i, V_i) = -\int_0^1 \langle V_i, V_i'' + R(V_i, \dot\gamma)\dot\gamma\rangle dt = \int_0^1 \sin^2(\pi t)(\pi^2 - l^2 R(e_1, e_i, e_i, e_1))dt$$

and by summing over i we get

$$\begin{aligned}\sum_{i=2}^n I(V_i, V_i) &= \int_0^1 \sin^2(\pi t)((n-1)\pi^2 - l^2 \operatorname{Ric}_p(e_1, e_1))dt \\ &< (n-1)\int_0^1 \sin^2(\pi t)dt(\pi^2 - l^2 K) < 0\end{aligned}$$

so that $I(V_j, V_j) < 0$ for some $j \geq 2$. Thus, there exists $t_0 \in (0, 1)$ such that $\gamma(t_0)$ is conjugate to p along γ. Therefore γ is not minimizing, which is a contradiction. $\square$

Let's turn our focus on Jacobi fields that are important since they describe the derivative of the exponential map:

Proposition. *Let $p \in M$ and $u, v \in T_pM$. Consider the geodesic $\gamma(t) = \exp_p tv$ and V the Jacobi field along γ such that $V(0) = 0$ and $V'(0) = u$. Then*

$$(D\exp_p)_{tv}(tu) = V(t).$$

We say "the" with confidence because of the lifesaving theory of O.D.E., once again. Note that $\dot\gamma(t)$ and $t\dot\gamma(t)$ are Jacobi fields along γ. The first one has derivative zero but vanishes nowhere, while the second one is zero if and only if $t = 0$. For this reason we mainly consider Jacobi fields that are normal to $\dot\gamma$.

Jacobi fields contribute to the search of conjugate points. If γ is a geodesic whose endpoints are conjugate, then a non trivial Jacobi field along γ exists such that it vanishes at its endpoints. Also, if $\gamma : [0, 1] \to M$ is a geodesic joining p to q which are not conjugate then for $v_0 \in T_pM$, $v_1 \in T_qM$ we can find a unique Jacobi field J along γ such that $J(0) = v_0$ and $J(1) = v_1$. This leads to the following

Corollary. *Let $\gamma : [0, 1] \to M$ be a geodesic in M^n and let $\mathcal{J}^\perp$ be the space of Jacobi fields with $J(0) = 0$ and $J'(0) \perp \dot{\gamma}(0)$. Let $\{J_1, \ldots, J_{n-1}\}$ be a basis of $\mathcal{J}^\perp$. If $\gamma(t)$, $t \in (0, 1]$ is not conjugate to $\gamma(0)$, then $\{J_1(t), \ldots, J_{n-1}(t)\}$ is a basis for the orthogonal complement $\langle\dot{\gamma}(t)\rangle^\perp \subseteq T_{\gamma(t)}M$ of $\dot{\gamma}(t)$.*

We shall close this paragraph by producing a useful formula for computing volume of a manifold. When a geodesic ball does not intersect the cut-locus of its center, we can use the exponential map and associated Jacobi fields to construct such a formula. The formula below seems to make the matter wors by getting Jacobi fields involved. Howevr, the differential equations satisfied by Jacobi fields make the formula useful.

Proposition. *Let M be a complete Riemann manifold. Suppose the ball $B_r(p)$ does not intersect the cut-locus of p. For each unit vector $v \in T_pM$, let $\{e_1, \ldots, e_{n-1}, v\}$ be an orthonormal basis of T_pM. Then*

$$\mathrm{vol}_{\mathrm{g}}(B_r(p)) = \int_{\mathbb{S}^{n-1}} \int_0^r \sqrt{\det \mathcal{A}(t)}\,dt\,dv$$

where

$$\mathcal{A}(t) := \Big(\langle J_i(t),\, J_j(t)\rangle_{\gamma(t)}\Big)_{i,j=1}^{n-1}$$

and J_i is the Jacobi field along $\gamma(t) = \exp_p(tv)$ with $J_i(0) = 0$ and $J_i'(0) = e_i$, $i = 1, \ldots, n-1$ and dv is the canonical volume element of $\mathbb{S}^{n-1}$, regarded as the unit sphere of T_pM.

Proof. Since the ball $B_r(p)$ does not intersect the cut-locus, we will use the inverse exponential map $\exp_p^{-1}$ as the local chart φ in the definition of the volume form. Let $\{e_1, \ldots, e_{n-1}, v\}$ be an orthonormal basis of T_pM. Every point $q \in B_r(p)$ is represented in this chart by the coordinates $(x^1, \ldots, x^n)$ such that $q = \exp_p(x^1e_1 + x^2e_2 + \cdots + x^nv)$. For any fixed $t \in [0, r]$, let $q = \exp_p(tv)$, which s not a conjugate point of p. Hence $(D\exp_p)_{tv}$ is nonsingular and

$$\{(D\exp_p)_{tv}e_1,\, \ldots,\, (D\exp_p)_{tv}e_{n-1},\, (D\exp_p)_{tv}v\}$$

is a basis of T_qM. In fact it is the local basis $\{\partial_1, \ldots, \partial_n\}$, since for any smooth function $f \in C^\infty(B_r(p))$, one has

$$\partial_i|_q f = \left[\frac{d}{ds}\bigg|_{s=0} \exp_p(tv + se_i)\right] f = \big[(D\exp_p)_{tv}e_i\big] f$$

by definition. But, along the geodesic $\gamma(t) = \exp_p(tv)$, the Jacobi field J_i with $J_i(0) = 0$ and $J_i'(0) = e_i$, $i = 1, \ldots, n-1$, has the form

$$J_i(t) = (D\exp_p)_{tv} e_i$$

while, by the chain rule, we have $(D\exp_p)_{tv} v = \gamma'(t)$. Hence

$$\{\partial_1|_p, \ldots, \partial_n|_p\} = \left\{\frac{1}{t}J_1(t), \ldots, \frac{1}{t}J_{n-1}(t), \gamma'(t)\right\}$$

is the canonical basis of T_qM. In the local chart henerated by $\exp_p^{-1}$, the coordinates of q in the spherical system are (v, s), where v is regarded as an element in $\mathbb{S}^{n-1}$. Therefore the volume element of q, under this local chart is

$$d\mathrm{vol}_{\mathrm{g}} = \sqrt{\det(\langle \partial_i, \partial_j\rangle)}dx^1 \ldots dx^n = \sqrt{\det(\langle J_i(t), J_j(t)\rangle)}dtdv$$

where dv is the canonical volume element of $\mathbb{S}^{n-1}$, regarded as the unit sphere of T_pM. After integration, we obtain

$$\mathrm{vol}_{\mathrm{g}}(B_r(p)) = \int_{\mathbb{S}n-1}\int_0^r \sqrt{\det(\langle J_i(t), J_j(t)\rangle)}dtdv.$$

□

Let's see an example. Denote by $\mathbb{M}_n^K$ the space of constant sectional curvature equal to K and let J be a Jacobi field along γ that is normal to $\dot\gamma$. Then, the Jacobian equation becomes

$$J''(t) + KJ(t) = 0$$

which has as a solution, with initial conditions $J(0) = 0$, $J'(0) = w(0)$:

$$J(t) = t\,\mathrm{sn}_K(t)w(t)$$

where

$$\mathrm{sn}_K(r) := \begin{cases} \dfrac{\sin(\sqrt{K}r)}{\sqrt{K}r} & , \ K > 0 \quad , 0 < r < \frac{\pi}{\sqrt{K}} \\ 1 & , \ K = 0 \quad , r > 0 \\ \dfrac{\sinh\left(\sqrt{-K}r\right)}{\sqrt{-K}r} & , \ K < 0 \quad , r > 0 \end{cases}$$

with

$$\mathrm{sn}_K(0) := 1$$

and $w(t)$ is a parallel field along γ which has unit length and is normal to $\dot{\gamma}$. Note that the upper bound is natural and no generality is lost, due to Bonnet-Myers theorem. Moreover, denoting by $V_{K,n}(r)$ the volume of a ball of any center and radius r, we have

$$V_{K,n}(r) = \int_{\mathbb{S}^{n-1}} \int_0^r t^{n-1} \operatorname{sn}_K^{n-1}(t) dt dv = a_{n-1} \int_0^r t^{n-1} \operatorname{sn}_K^{n-1}(t) dt$$

where a_{n-1} is the area of $\mathbb{S}^{n-1}$.

2.13 Bishop-Gromov Volume Comparison Theorem

In this section we will formulate and prove the well-known Bishop-Gromov volume comparison theorem. Given a unit vector $v \in T_xM$, we fix a unit speed minimal geodesic $\gamma : [0, l] \to M$ with $\dot{\gamma}(0) = v$ and an orthonormal basis $\{e_1, \ldots, e_{n-1}, v\}$ of T_xM. Now, we consider the geodesic variations $f_i : (-\varepsilon, \varepsilon) \times [0, l] \to M$ of γ, defined by

$$f_i(s, t) := \exp_x \big(t(v + se_i)\big)$$

for $i = 1, \ldots, n-1$. Next, we introduce their respective Jacobi fields along γ:

$$J_i(t) := \frac{\partial f_i}{\partial s}(0, t) = (D \exp_x)_{tv}(te_i) \in T_{\gamma(t)}$$

for $i = 1, \ldots, n-1$. Note that $J_i(0) = 0$ and $J_i'(0) = e_i$ while the Gauss lemma asserts that $g_\gamma(J_i, \dot{\gamma}) \equiv 0$ and, hence, $g_\gamma(J_i', \dot{\gamma}) \equiv 0$, since J_i are Jacobi fields along the geodesic γ. Since γ is minimal, $\gamma(t)$ isn't conjugate to x, for $t \in (0, l)$, and thus $\{\frac{1}{t}J_1(t), \ldots, \frac{1}{t}J_{n-1}(t), \dot{\gamma}(t)\}$ is the canonical basis of $T_{\gamma(t)}M$. We define an $(n-1) \times (n-1)$ matrix $\mathcal{U}(t) = (u_{ij}(t))_{i,j=1}^{n-1}$ such that $J_i'(t) = \sum_{j=1}^{n-1} u_{ij} J_j(t)$, for $t \in (0, l)$. Also, define two more $(n-1) \times (n-1)$ matrices, which happen to be symmetric:

$$\mathcal{A}(t) := \Big(\langle J_i(t),\, J_j(t)\rangle_{\gamma(t)}\Big)_{i,j=1}^{n-1},\ \mathcal{R}(t) := \Big(\langle R(J_i(t), \dot{\gamma}(t))\dot{\gamma}(t),\, J_j(t)\rangle_{\gamma(t)}\Big)_{i,j=1}^{n-1}.$$

Lemma. *Let $\mathcal{U}$, $\mathcal{A}$, $\mathcal{R}$ be the above matrices. Then*

(i) $\operatorname{Ric}_{\gamma(t)}(\dot{\gamma}(t)) = \operatorname{tr}(\mathcal{R}(t)\mathcal{A}^{-1}(t))$

(ii) $\mathcal{U}\mathcal{A} = \mathcal{A}\mathcal{U}^T$ *and* $2\mathcal{U} = \mathcal{A}'\mathcal{A}^{-1}$

(iii) $\mathcal{U}$ *is also symmetric and* $\operatorname{tr}(\mathcal{U}^2) \geq \frac{(\operatorname{tr}\mathcal{U})^2}{n-1}$

Proof. (i) Let $\mathcal{C} = (c_{ij})_{i,j=1}^{n-1}$ be a matrix such that $\left\{ \sum_{j=1}^{n-1} c_{ij}(t)J_j(t) \right\}_{i=1}^{n-1}$ is an orthonormal basis of $\langle \dot\gamma(t) \rangle^\perp$. Then

$$\begin{aligned}
\operatorname{Ric}_{\gamma(t)}(\dot\gamma(t)) &= \sum_{i,j,k=1}^{n-1} \langle R(c_{ij}J_j(t), \dot\gamma(t))\dot\gamma(t),\, c_{ik}J_k(t)\rangle_{\gamma(t)} \\
&= \operatorname{tr}(\mathcal{C}(t)\mathcal{R}(t)\mathcal{C}^T(t)) \\
&= \operatorname{tr}(\mathcal{R}(t)\mathcal{A}^{-1}(t)\mathcal{C}(t)\mathcal{A}(t)\mathcal{C}^T(t)) \\
&= \operatorname{tr}(\mathcal{R}(t)\mathcal{A}^{-1}(t))
\end{aligned}$$

since $\mathcal{C}\mathcal{A}\mathcal{C}^T = I$, by construction.

(ii) Observe that

$$\begin{aligned}
\mathcal{U}\mathcal{A} &= \left(\sum_{k=1}^{n-1} u_{ik}\langle J_k,\, J_j\rangle_\gamma \right)_{i,j=1}^{n-1} \\
&= \left(\left\langle \sum_{k=1}^{n-1} u_{ik}J_k,\, J_j \right\rangle_\gamma \right)_{i,j=1}^{n-1} \\
&= \Big(\langle J_i',\, J_j\rangle_\gamma \Big)_{i,j=1}^{n-1}
\end{aligned}$$

and similarly $\mathcal{A}\mathcal{U}^T = \Big(\langle J_i,\, J_j'\rangle_\gamma \Big)_{i,j=1}^{n-1}$, so that

$$\mathcal{A}' = \mathcal{U}\mathcal{A} + \mathcal{A}\mathcal{U}^T.$$

Differentiating along the geodesic γ gives

$$\begin{aligned}
\frac{D}{dt}\Big(\langle J_i',\, J_j\rangle_\gamma - \langle J_i,\, J_j'\rangle_\gamma \Big) &= \langle J_i'',\, J_j\rangle_\gamma - \langle J_i,\, J_j''\rangle_\gamma \\
&= -\langle R(J_i, \dot\gamma)\dot\gamma,\, J_j\rangle_\gamma + \langle J_i,\, R(J_j, \dot\gamma)\dot\gamma\rangle_\gamma = 0
\end{aligned}$$

and, since $J_i(0) = 0$, $J_i'(0) = e_i$ we have that $\langle J_i',\, J_j\rangle_\gamma = \langle J_i,\, J_j'\rangle_\gamma$. Thus $\mathcal{U}\mathcal{A} = \mathcal{A}\mathcal{U}^T$ and $\mathcal{A}' = 2\mathcal{U}\mathcal{A}$, which proves our claim.

(iii) Define a map

$$\varphi : \exp_x\left(\left\{ t\left(v + \sum_{i=1}^{n-1} s_i e_i\right) \,\middle|\, t \in [0,\, l],\, |s_i| < \varepsilon \right\}\right) \to \mathbb{R}$$

so that $\varphi(\exp_x(t(v + \sum_{i=1}^{n-1} s_i e_i))) = t$. Hence

$$\nabla\varphi(f_i(s,\, t)) = \frac{\partial f_i}{\partial t}(s,\, t)$$

and since $\frac{D}{dt}$, $\frac{D}{ds}$ commute we have that

$$J_i'(t) = \frac{D}{ds}\frac{\partial f_i}{\partial t}(0,\, t) = \frac{D}{ds}\nabla\varphi(\gamma(t)) = \nabla_{J_i(t)}\nabla\varphi(\gamma(t)) = \operatorname{Hess}_{\gamma(t)}\varphi(J_i(t))$$

so that $\mathcal{U}(t)$ is the matrix representation of the symmetric $(1,\, 1)$-tensor $\operatorname{Hess}_{\gamma(t)}\varphi$ with respect to the basis $\{J_1,\, \dots,\, J_{n-1}\}$ of $\langle\dot\gamma\rangle^\perp$. Therefore, $\mathcal{U}$ is also symmetric and if $\lambda_1,\, \dots,\, \lambda_{n-1}$ are its eigenvalues, the Causchy-Schwartz inequality shows that

$$(\operatorname{tr}\mathcal{U})^2 = \left(\sum_{i=1}^{n-1}\lambda_i\right)^2 \leq (n-1)\sum_{i=1}^{n-1}\lambda_i^2 = (n-1)\operatorname{tr}(\mathcal{U}^2)$$

which is what we wanted to show. □

A lower Ricci bound yields an important differential inequality which plays a crucial role in proving the Bishop-Gromov comparison theorem

Proposition. *If* $\operatorname{Ric}(M) \geq (n-1)K$ *for a* $K \in \mathbb{R}$ *then*

$$\left[(\det\mathcal{A})^{1/2(n-1)}\right]'' \leq -K(\det\mathcal{A})^{1/2(n-1)}$$

and thus $\sqrt{\det\mathcal{A}(t)}/t^{n-1}\operatorname{sn}_K^{n-1}(t)$ *is a non-increasing function of* t.

Proof. Using *Jacobi's formula for the derivative of the determinant* and the second part of the previous lemma the chain rule yields

$$\begin{aligned}\left[(\det\mathcal{A})^{1/2(n-1)}\right]' &= \frac{1}{2(n-1)}(\det\mathcal{A})^{1/2(n-1)-1}\cdot\det\mathcal{A}\operatorname{tr}(\mathcal{A}'\mathcal{A}^{-1})\\ &= \frac{1}{n-1}(\det\mathcal{A})^{1/2(n-1)}\operatorname{tr}\mathcal{U}.\end{aligned}$$

Differentiating once more, the third part of the lemma gives

$$\begin{aligned}\left[(\det\mathcal{A})^{1/2(n-1)}\right]'' &= \frac{1}{(n-1)^2}(\det\mathcal{A})^{1/2(n-1)}(\operatorname{tr}\mathcal{U})^2 + \frac{1}{n-1}(\det\mathcal{A})^{1/2(n-1)}\operatorname{tr}\mathcal{U}' \\ &\leq \frac{1}{n-1}(\det\mathcal{A})^{1/2(n-1)}(\operatorname{tr}\mathcal{U}^2 + \operatorname{tr}\mathcal{U}').\end{aligned}$$

Making use of $2\mathcal{U} = \mathcal{A}'\mathcal{A}^{-1}$ once again and taking into account that

$$J_i'' = -R(J_i, \dot{\gamma})\dot{\gamma}$$

we deduce that

$$\begin{aligned}\mathcal{U}' &= \frac{1}{2}\mathcal{A}''\mathcal{A}^{-1} - \frac{1}{2}(\mathcal{A}'\mathcal{A}^{-1})^2 \\ &= \frac{1}{2}(-2\mathcal{R} + 2\mathcal{U}\mathcal{A}\mathcal{U}^T)\mathcal{A}^{-1} - 2\mathcal{U}^2 \\ &= -\mathcal{R}\mathcal{A}^{-1} + \mathcal{U}^2\mathcal{A}\mathcal{A}^{-1} - 2\mathcal{U}^2 \\ &= -\mathcal{R}\mathcal{A}^{-1} - \mathcal{U}^2\end{aligned}$$

Taking trace on $\mathcal{U}^2 + \mathcal{U}' = -\mathcal{R}\mathcal{A}^{-1}$ the first part of the previous lemma yields

$$\operatorname{tr}\mathcal{U}^2 + \operatorname{tr}\mathcal{U}' = -\operatorname{Ric}_\gamma(\dot{\gamma})$$

and thus

$$\begin{aligned}\left[(\det\mathcal{A})^{1/2(n-1)}\right]'' &\leq \frac{1}{n-1}(\det\mathcal{A})^{1/2(n-1)}(\operatorname{tr}\mathcal{U}^2 + \operatorname{tr}\mathcal{U}') \\ &= -\frac{\operatorname{Ric}_\gamma(\dot{\gamma})}{n-1}(\det\mathcal{A})^{1/2(n-1)} \\ &\leq -K(\det\mathcal{A})^{1/2(n-1)}.\end{aligned}$$

Recall that sn_K are such that the function $y(t) := t\operatorname{sn}_K(t)$ is the solution of the differential equation $y'' + Ky = 0$, $y(0) = 0$, $y'(0) = 1$. Taking into account that y, $\det\mathcal{A} > 0$, this means that

$$\frac{\left[(\det\mathcal{A})^{1/2(n-1)}\right]''}{(\det\mathcal{A})^{1/2(n-1)}} \leq \frac{y''}{y}$$

$$\Rightarrow \left[(\det\mathcal{A})^{1/2(n-1)}\right]'' y - (\det\mathcal{A})^{1/2(n-1)} y'' \leq 0$$

$$\Rightarrow \left\{\left[(\det\mathcal{A})^{1/2(n-1)}\right]' y - (\det\mathcal{A})^{1/2(n-1)} y'\right\}' \leq 0$$

$$\Rightarrow \left[(\det \mathcal{A})^{1/2(n-1)}\right]' y - (\det \mathcal{A})^{1/2(n-1)} y' \leq -(\det \mathcal{A}(0))^{1/2(n-1)} \leq 0$$

$$\Rightarrow \left(\frac{(\det \mathcal{A})^{1/2(n-1)}}{y}\right)' \leq 0$$

which means that $(\det \mathcal{A})^{1/2(n-1)}/y$ is non increasing. But, since the operation $x \mapsto x^{n-1}$ preserves ordering we have that

$$\sqrt{\det \mathcal{A}(t)}/t^{n-1} \operatorname{sn}_K^{n-1}(t)$$

is a non-increasing function of t □

Now we can formulate and prove the Bishop-Gromov comparison theorem:

Theorem (Bishop-Gromov). *Assume that* $\operatorname{Ric}(M) \geq (n-1)K$ *holds for some* $K \in \mathbb{R}$. *Then we have, for any* $x \in M$ *and* $0 < r < R$ $\left(\leq \frac{\pi}{\sqrt{K}} \text{ if } K > 0\right)$,

$$\frac{\operatorname{vol}_{\mathrm{g}}(B_r(x))}{\operatorname{vol}_{\mathrm{g}}(B_R(x))} \geq \frac{V_{K,n}(r)}{V_{K,n}(R)}$$

where $V_{K,n}(r) = a_{n-1} \int_0^r \operatorname{sn}_K^{n-1}(t)dt$ *is the volume of a ball of radius* r *in the space form of constant sectional curvature equal to* K. *Thus the function* $\operatorname{vol}_{\mathrm{g}}(B_r(x))/V_{K,n}(r)$ *is non-increasing in* r. *Moreover, it tends to 1, as* $r \to 0$.

Proof. Following our previous conversation, setting $\mathscr{A}(t) = \int_{\mathbb{S}^{n-1}} \sqrt{\det \mathcal{A}(t)} dv$ and $\mathcal{S}(t) = t^{n-1} \operatorname{sn}_K^{n-1}(t)$ we have

$$\begin{aligned}
\operatorname{vol}_{\mathrm{g}}(B_r(x)) \int_0^R t^{n-1} \operatorname{sn}_K^{n-1}(t)dt &= \int_0^r \mathscr{A}(t)dt \int_r^R \mathcal{S}(t)dt + \int_0^r \mathscr{A}(t)dt \int_0^r \mathcal{S}(t)dt \\
&\geq \frac{\mathscr{A}(r)}{\mathcal{S}(r)} \int_0^r \mathcal{S}(t)dt \int_r^R \mathcal{S}(t)dt + \int_0^r \mathscr{A}(t)dt \int_0^r \mathcal{S}(t)dt \\
&\geq \int_0^r \mathcal{S}(t)dt \int_r^R \mathscr{A}(t)dt + \int_0^r \mathscr{A}(t)dt \int_0^r \mathcal{S}(t)dt \\
&= \operatorname{vol}_{\mathrm{g}}(B_R(x)) \int_0^r t^{n-1} \operatorname{sn}_K^{n-1}(t)dt
\end{aligned}$$

Multiplying with the volume of the unit ball a_{n-1} we get

$$\operatorname{vol}_{\mathrm{g}}(B_r(x)) V_{K,n}(R) \geq \operatorname{vol}_{\mathrm{g}}(B_R(x)) V_{K,n}(r)$$

so that the function $\mathrm{vol}_g(B_r(x))/V_{K,n}(r)$ is non-increasing in r. Lastly, in normal coordinates we have

$$\lim_{r\to 0} \frac{\mathrm{vol}_g(B_r(x))}{V_{K,n}(r)} = \lim_{r\to 0} \frac{\omega_n r^n \Big(1 - \frac{\mathrm{scal}(p)}{6(n+2)} r^2 + O(|r|^3)\Big)}{\omega_n r^n \Big(1 - \frac{K}{6(n+2)} r^2 + O(|r|^3)\Big)} = 1$$

and the theorem is proved. □

2.14 Volume Distortion Coefficients

We would like to measure how much the volume of a ball changes as we transport it along geodesics. To this end, we define a notion of barycenter to play the role of $(1-t)x + ty$. For fixed $t \in [0,\, 1]$ define the locus of points lying partway between x and $y \in M$:

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

If we're given a *unique* minimizing geodesic $\gamma : [0,\, 1] \to M$, then $Z_t(x,\, y) = \{\gamma(t)\}$, as one can see. The above definition extends to sets $X,\, Y \subseteq M$ by taking the union for each point. Namely:

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

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

Now, for a ball $B_r(y) \subseteq M$ and $t \in (0,\, 1]$, we define the **volume distortion coefficients**

$$v_t(x,\, y) := \lim_{r\to 0} \frac{\mathrm{vol}_g(Z_t(x,\, B_r(y)))}{\mathrm{vol}_g(B_{tr}(y))}$$

for which we shall prove that exists and is positive, when $y \notin \mathrm{cut}(x)$, by linking it with the differential of the exponential map $Y(t) = (D\exp_x)_{tv}$ and the Hessian of the distance function $H(t) = \mathrm{Hess}_x\, d^2_{\gamma(t)}/2$ where γ is the minimal geodesic joining x to y:

Proposition. *Fix $x,\, y \in M$ with $y \notin \mathrm{cut}(x)$ and let $\gamma(t) := \exp_x(tv)$ be the minimal geodesic joining x to y. Then, for $t \in (0,\, 1]$*

$$v_t(x,\, y) = \frac{\det Y(t)}{\det Y(1)} = \det Y(t)Y(1)^{-1} > 0$$

and for $t \in [0,\, 1)$

$$v_{1-t}(y,\, x) = \det \frac{Y(t)\big(H(t) - tH(1)\big)}{1-t}$$

Proof. Observe that since $\exp_x$ is a local diffeomorphism for small enough r the set $D_r := \exp_x^{-1}(B_r(y))$, for $y \notin \mathrm{cut}(x)$, is open and contains v thus we can find a small neighbourhood of v inside it that we can scale proportionally to r in order to put D_r inside. Thus D_r shrinks nicely to $v \in T_xM$. The map $G_t(u) := \exp_x(tu)$ is a local diffeomorphism between a neighbourhood of v and a neighbourhood of $\gamma(t)$ and maps D_r onto $Z_t(x,\, B_r(y))$, as long as $B_r(y) \cap \mathrm{cut}(x) = \emptyset$. Also, note that $(DG_t)_v = tY(t)$ and so, a change of variables, along with Lebesque's differentiation theorem ([7]) (since D_r shrinks nicely to v) gives:

$$\det(tY(t)) = \det(DG_t)_v = \lim_{r\to 0} \frac{\mathrm{vol_g}(Z_t(x,\, B_r(y)))}{\mathrm{vol_g}(D_r)}$$

since $\det Y(t) > 0$ (changes continuously, while $Y(0) = I$). Thus, for $t = 1$ one has

$$\det Y(1) = \lim_{r\to 0} \frac{\mathrm{vol_g}(B_r(y))}{\mathrm{vol_g}(D_r)}$$

hence

$$\frac{\det Y(t)}{\det Y(1)} = \lim_{r\to 0} \frac{\mathrm{vol_g}(Z_t(x,\, B_r(y)))}{t^n\, \mathrm{vol_g}(B_r(y))} = v_t(x,\, y)$$

since $\lim_{r\to 0} \dfrac{\mathrm{vol_g}(B_{tr}(y))}{t^n\, \mathrm{vol_g}(B_r(y))} = 1.$

Observe that $(DG_t)_v(u) = tY(t)u = (D\exp_x)_{tv}(tu)$ which is Jacobi field along γ, so $tY(t)Y(1)^{-1}u$ will also be a Jacobi field along γ, as a linear combination of Jacobi fields along γ. But, since $y \notin \mathrm{cut}(x)$, Jacobi fields along $\gamma : x \mapsto y$ are determined by their endpoints, which in this case are 0 and u for $t = 0$ and $t = 1$ respectively. So the first equality can by reformulated as: if $A(t)$ $(= tY(t)Y(1)^{-1})$ is the unique matrix with Jacobi fields as its columns (expressed at an orthonormal frame parallel transported along γ) such that $A(0) = 0$ and $A(1) = I$ then

$$v_t(x,\, y) = \det \frac{A(t)}{t}$$

for $t \in (0,\, 1]$. Therefore, if $B(t)$ $(= A(1-t))$ is the unique matrix of Jacobi fields along γ such that $B(0) = I$ and $B(1) = 0$, then

$$v_{1-t}(x,\, y) = \det \frac{B(t)}{1-t}$$

for $t \in [0, 1)$. Fix $u \in T_xM$ and consider the following geodesic variation of γ:

$$f(t, s) := \exp_x\big(-\nabla d^2_{\gamma(t)}(x)/2 + s(H(t) - tH(1))(u)\big).$$

Its variational field $V(t) := \frac{\partial f}{\partial s}(t, s)$ is a Jacobi field along γ and

$$\frac{\partial f}{\partial s}(t, s) = Y(t)(H(t) - tH(1))(u)$$

by the chain rule. Since $V(0) = Y(0)H(0)u = Iu = u$ and $V(1) = 0$ we conlude that $B(t) = Y(t)(H(t) - tH(1))$ and the theorem is proved. $\square$

The previous characterization of v_t together with the Bishop's comparison theorem leads to an estimate of v_t in terms of Ricci curvature:

Corollary. *Assume that* $\mathrm{Ric}(M) \geq (n-1)K$ *throughout* M *for some* $K \in \mathbb{R}$. *Then for* $x, y \in M$ *with* $y \notin \mathrm{cut}(x)$ *and* $t \in (0, 1)$ *one has*

$$v_t(x, y) \geq \left(\frac{\mathrm{sn}_K(td(x, y))}{\mathrm{sn}_K(d(x, y))}\right)^{n-1}$$

and equality holds when M *has constant sectional curvature equal to* K.

Proof. Let $\gamma(t) = \exp_x(tv)$ be the minimal geodesic joining x to y as before. Since γ is minimal and $y \notin \mathrm{cut}(x)$ the Ricci bound yields:

$$\mathrm{Ric}_{\gamma(t)}(\dot\gamma(t), \dot\gamma(t)) \geq K|v|_x^2$$

where $|\dot\gamma(t)|_{\gamma(t)} = |v|_x = d(x, y)$. Recall from the last section the matrix

$$\mathcal{A}(t) = \Big(\langle J_i(t), J_j(t)\rangle_{\gamma(t)}\Big)_{i,j=1}^{n-1}$$

where $\frac{1}{t}J_i(t) = (D\exp_x)_{tv}(e_i)$ is the i^{th} column of an $(n-1) \times (n-1)$ matrix $\mathcal{Y}(t)$ with respect to the basis $\{e_1, \ldots, e_{n-1}\}$ of $\langle v\rangle^\perp$ which we parallel transport along γ, that satisfies

$$Y(t) = \left(\begin{array}{c|c} \mathcal{Y}(t) & \mathbf{0} \\ \hline \mathbf{0} & |v|_x \end{array}\right)$$

It's easy to see that $\mathcal{Y}\mathcal{Y}^T(t) = \frac{1}{t^2}\mathcal{A}(t)$ and thus

$$\det Y(t) = |v|_x \det \mathcal{Y}(t) = \frac{|v|_x}{t^{n-1}}\sqrt{\det \mathcal{A}(t)}.$$

Now, the Bishop's comparison theorem asserts that

$$t \mapsto (\det Y(t))(\mathrm{sn}_K(t|v|_x))^{-(n-1)}$$

is a non-increasing function of $t \in (0, 1]$. Thus

$$v_t(x, y) = \frac{\det Y(t)}{\det Y(1)} \geq \left(\frac{\mathrm{sn}_K(t|v|_x)}{\mathrm{sn}_K(|v|_x)}\right)^{n-1}.$$

Now, if M has constant sectional curvature then computing, as in the previous section, the Jacobi fields that make up $Y(t)$ we get equality. □

2.15 Superdifferentiability of the Distance Function

A function $f : M \to \mathbb{R}$ is called **Lipschitz continuous** on M if there exists a non-negative constant K such that for every $x, y \in M$ one has

$$|f(x) - f(y)| \leq Kd(x, y)$$

and **locally Lipschitz** if it is Lipschitz continuous in every compact subset of M. One example of a locally Lipschitz map is the distance function itself, restricted to the one variable, with constant $K = 1$. An other one would be the square of the distance function, also restricted to the one variable:

Lemma. *For every $y \in M$ the function $d_y^2 := d^2(\cdot, y)$ is locally Lipschitz on M.*

Proof. Fix $y \in M$ and let $K \subseteq M$ be a compact set. Compactness gives an upper bound D for d_y. Let $x, z \in K$. Then, the triangle inequality gives us:

$$\begin{aligned} 2|d_y^2(x) - d_y^2(z)| &= |d^2(x, y) - d(x, y)d(z, y) + d(x, y)d(z, y) - d^2(z, y)| \\ &\leq |d(x, y)||d(x, y) - d(z, y)| + |d(z, y)||d(x, y) - d(z, y)| \\ &\leq Dd(x, z) + Dd(x, z) \\ &= 2Dd(x, z) \end{aligned}$$

which gives:

$$|d_y^2(x) - d_y^2(z)| \leq Dd(x, z), \text{ for every } x, z \in X$$

therefore, d_y^2 is Lipschitz on K. Since K was arbitrary we have that d_y^2 is locally Lipschitz. □

In this last section we will explore when $d_y := d(\cdot,\, y)$ is differentiable or not and give a characterization of cut points in terms of its differentiability. Firstly, we shall establish that its differentiability points are quite a few:

Theorem (Rademacher's Theorem, Riemannian case)**.** *Let* $(M,\, g,\, d)$ *be any smooth connected Riemannian manifold with geodesic distance* d*. Any function* $f : M \to \mathbb{R}$ *which is locally Lipschitz, is differentiable* $\mathrm{vol}_{\mathrm{g}}$*-almost everywhere and its gradient* $\nabla f : M \setminus N_f \to TM$ *is a Borel map, where* $\mathrm{vol}_{\mathrm{g}}(N_f) = 0$*.*

Proof. We will not prove the Euclidean case, which appears in many textbooks, such as [6], which we closely follow in the second part of the proof.

Fix $p \in M$ and *normal coordinates* $\eta : U \to \mathbb{R}^n$ centered at $p = \eta^{-1}(0)$, so that $g_{ij}(p) = \delta_{ij}$. We know, that for the largest eigenvalue of the *quadratic form* $g_x(\cdot,\, \cdot)$ we have $\lambda(x) = \max_{v \in T_xM} \dfrac{1}{\langle v,\, v\rangle} v^T[(g_{ij}(x))_{i,j}]v$ which is continuous, since the coefficients $g_{ij}(x)$ depend continuously on these coordinates. So, there is a smaller neighbourhood of z, let's say $V \subseteq U$, in which we can bound λ by a finite positive number k and we can have that

$$g_x(v,\, v) \leq k\langle v,\, v\rangle$$

for all $x \in V$ and $v \in T_xM$. We choose an $\varepsilon > 0$ small enough, so that $B_\varepsilon(0) \subset \eta(V)$ and set $W := \eta^{-1}(B_\varepsilon(0))$. If we show that f is differentiable $\mathrm{vol}_{\mathrm{g}}$-almost everywhere on W then we can extend this result to M, since a connected Riemannian manifold is locally compact and second countable, thus it can be covered by countably many such neighbourhoods.

The geodesic distance between $x,\, y \in W$ is bounded by the length of the path $\alpha(t) := \eta^{-1}((1-t)\eta(x) + t\eta(y))$, so one has:

$$d(x,\, y) \leq \int_0^1 \sqrt{\langle \dot\alpha(t),\, \dot\alpha(t)\rangle_{\alpha(t)}}dt \leq \sqrt{k}|\eta(y) - \eta(x)|$$

after combining $\dot\alpha^i(t) = \eta^i(y) - \eta^i(x)$ with our previous bound on the metric. So, the function $\varphi := f \circ \eta^{-1}$ is also locally Lipschitz, since f is assumed to be:

$$\begin{aligned} |\varphi(u) - \varphi(v)| = |f \circ \eta^{-1}(u) - f \circ \eta^{-1}(v)| &\leq \\ \leq L_f d(\eta^{-1}(u),\, \eta^{-1}(v)) &\leq \\ \leq L_f\sqrt{k}|u - v| \end{aligned}$$

for u, $v \in B_\varepsilon(0)$.

By the (Euclidean) Rademacher's theorem it follows that φ is differentiable m-almost everywhere on $B_\varepsilon(0)$ (m is the *Lebesgue measure*). Now, if we call

$$\begin{aligned} N_f &:= \{x \in M : f \text{ is not differentiable at } x\} \\ N_\varphi &:= \{x \in \mathbb{R}^n : \varphi \text{ is not differentiable at } x\} \end{aligned}$$

one has:

$$\begin{aligned} \mathrm{vol}_g(N_f \cap W) &= \int_{\eta(N_f)\cap B_\varepsilon(0)} \sqrt{\det(g_{ij}) \circ \eta^{-1}} dm \leq \\ &\leq Cm(\eta(N_f) \cap B_\varepsilon(0)) \\ &= Cm(N_\varphi \cap B_\varepsilon(0)) \\ &= 0 \end{aligned}$$

where the inequality comes from the fact that the integrated function is continuous on $B_\varepsilon(0)$ and the quality $\eta(N_f) = N_\varphi$ is a simple exercise:

$$\begin{aligned} v \in \eta(N_f) &\Leftrightarrow \exists x \in N_f : f \text{ is not differentiable at } x = \eta^{-1}(v) \Leftrightarrow \\ &\Leftrightarrow \varphi = f \circ \eta^{-1} \text{ is not differentiable at } v \Leftrightarrow \\ &\Leftrightarrow v \in N_\varphi \end{aligned}$$

since η is a local diffeomorphism. We deduce that f is differentiable vol_g-almost everywhere on W, hence vol_g-almost everywhere on M.

For the second part we will follow [6] very closely. After extending $\varphi = f \circ \eta^{-1}$ continuously to all of $\mathbb{R}^n$, they observe that its **upper derivative**:

$$\overline{D}_v\varphi(x) := \lim_{k\to\infty} \sup_{|t|\in(0,\,1/k)\cap Q} \frac{\varphi(x+tv) - \varphi(x)}{t}$$

in direction $v \in \mathbb{R}^n$ is expressed as a limit of suprema of continuous functions, hence Borel. Similarly the **lower derivative** $\underline{D}_v\varphi$ is also Borel as the only difference is an infimum in the place of the supremum. Thus, the directional derivative $D_v\varphi$ is a Borel function on the set of full measure where $\overline{D_v}\varphi = \underline{D}_v\varphi$. In particular, the n partial derivatives $\partial_i\varphi$ are Borel, as is

$$F_v := \{x \in B^n_\varepsilon(0) \mid \overline{D_v}\varphi = \underline{D}_v\varphi = v^j\partial_j\varphi\}.$$

In [6] they show that φ is differentiable on any countable intersecction $\cap F_{v_i}$ over a dense set of directions $v_i \in \partial B^n_1(0)$. Outside F_{v_i} differentiability fails, so $\nabla\varphi$ must be Borel. Clearly $g^{kj}\partial_j\varphi$ is also Borel on $\cap F_{v_i}$ and gives the coordinates of ∇f on $U \setminus N_f$. We conclude that ∇f is a Borel map. $\square$

As we see d_y^2 is differentiable almost everywhere and ∇d_y^2 is a Borel map. The next step would be to check under what circumstances it is differentiable. Consider a totally normal neighbourhood W of x. Every $y \in W$ is joined to x by a unique minimizing geodesic γ. This uniqueness is what forces d_y^2 to be differentiable at $x = \exp_y \dot\gamma(0)$:

Proposition. *If y, $x \in M$ are joined by a unique minimizing geodesic, then $d_y^2/2$ is differentiable at x and*

$$\nabla(d_y^2/2)(x) = -(\exp_x)^{-1}(y).$$

Proof. Let $\gamma : [0, 1] \to M$ be the minimizing geodesic connecting $y = \gamma(0)$ with $x = \gamma(1)$ parametrized with consant speed. We compute d_y^2's derivative by linearizing $\exp_x$ around the origin and $\exp_y$ around $\dot\gamma(0)$:

$$\begin{aligned}
d_y^2(\exp_x v) &= d^2(y, \exp_y(\exp_y^{-1} \exp_x v)) \\
&= |\exp_y^{-1}(\exp_x v)|_y^2 \\
&= |\dot\gamma(0) + (D\exp_y^{-1})_x(D\exp_x)_0 v + o(|v|_x)|^2 \\
&= |\dot\gamma(0)|_y^2 + g_y(\dot\gamma(0), (D\exp_y)^{-1}_{\dot\gamma(0)} I v) + o(|v|_x)|^2 \\
&= d^2(y, \exp_y \dot\gamma(0)) + 2\langle \dot\gamma(1), v\rangle + o(|v|_x) \\
&= d_y^2(x) + \langle 2\dot\gamma(1), v\rangle + o(|v|_x)
\end{aligned}$$

where we exploited Gauss' Lemma along with $(D\exp_y)_{\dot\gamma(0)}\dot\gamma(0) = \dot\gamma(1)$. So, dividing by 2 gives $\nabla(d_y^2/2)(x) = \dot\gamma(1)$ Taking in mind that $\delta(t) := \gamma(1-t)$ is "the same" geodesic we can deduce that

$$\dot\gamma(1) = -(\exp_x)^{-1}(y)$$

and the statement is proved. □

What happens if we don't have uniqueness? We must introduce a new notion. A function $\varphi : M \to \mathbb{R}$ is said to be **superdifferentiable at $\boldsymbol{x \in M}$** with **supergradient $\mathbf{p} \in \mathbf{T_x M}$** if

$$\varphi(\exp_x v) \leq \varphi(x) + \langle \mathbf{p}, v\rangle_x + o(|v|_x)$$

holds for small $v \in T_x M$ where $|v|_x = \sqrt{\langle v, v\rangle_x}$ and $o(\lambda)/\lambda \xrightarrow{\lambda\to 0} 0$. The set of all such pairs $(\mathbf{p}, x) \in TM$ is denoted by $\overline{\partial}\varphi$ and is called the **superdifferential of $\boldsymbol{\varphi}$**. Note that, when we say $\mathbf{p} \in \overline{\partial}\varphi(x)$, we actually mean the above inequality. Changing the prefix "super" with "sub" we are talking about the reverse inequality, which is expressed by $\mathbf{q} \in \underline{\partial}\varphi(x)$, where $\underline{\partial}\varphi$ is the **subgradient of $\boldsymbol{\varphi}$**.

Note that, if φ is both super- and sub-differentiable, then it is differentiable at x and

$$\mathbf{p} = \mathbf{q} = \nabla\varphi(x)$$

since, in this case, we have for small $v \in T_xM$:

$$\langle \mathbf{p} - \mathbf{q},\, v\rangle_x + o(|v|_x) \geq 0$$

so that, if we take $v = s(\mathbf{q} - \mathbf{p})$ for small s we have:

$$-|\mathbf{p} - \mathbf{q}|_x + \frac{o(|s|)}{|s|} \geq 0 \overset{s\to 0}{\Longrightarrow} |\mathbf{p} - \mathbf{q}|_x = 0 \Rightarrow \mathbf{p} = \mathbf{q}$$

and, of course,

$$\varphi(\exp_x v) = \varphi(x) + \langle \mathbf{p},\, v\rangle_x + o(|v|_x) = \varphi(x) + \langle \nabla\varphi(x),\, v\rangle_x + o(|v|_x).$$

An example, to grasp the above idea, would be $\varphi : \mathbb{R} \to \mathbb{R}$, $\varphi(x) = |x|$, which is differentiable for $x \neq 0$ and

$$\overline{\partial}\varphi|_{x<0} = \underline{\partial}\varphi|_{x<0} = \{-1\}$$

$$\overline{\partial}\varphi|_{x>0} = \underline{\partial}\varphi|_{x>0} = \{1\}$$

but, at $x = 0$,

$$\overline{\partial}\varphi(0) = \emptyset \text{ and } \underline{\partial}\varphi(0) = [-1,\, 1]$$

To see why this is true, observe that $p \in \overline{\partial}\varphi(0) \Leftrightarrow |v| \leq pv$ for v small, but if we take $v_n = \frac{(-1)^n}{n}$ we see that such p doesn't exist. On the other hand $q \in \underline{\partial}\varphi(0) \Leftrightarrow |v| \geq pv$ for v small $\Leftrightarrow p \in [-1,\, 1]$.

It's natural to expect that a chain rule for supergradients must hold:

Lemma. *Let $\varphi : M \to \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ have supergradients $\boldsymbol{p} \in \overline{\partial}\varphi(x)$ and $\boldsymbol{\lambda} \in \overline{\partial}_{\varphi(x)} f$, respectively, at some point $x \in M$. If f is non-decreasing then*

$$\boldsymbol{\lambda} \cdot \boldsymbol{p} \in \overline{\partial} f \circ \varphi(x)$$

Proof. Note that the superdifferential of f is a subset of $T\mathbb{R} = \mathbb{R}$, so λ is a real number.

Since f is superdifferentiable at $\varphi(x)$ with supergradient $\boldsymbol{\lambda}$, definition yields:

$$\begin{aligned}\varphi(\exp_x v) &\le \varphi(x) + \langle \mathbf{p},\, v\rangle_x + o(|v|_x)\\ f(\varphi(x)+\varepsilon) &\le f(\varphi(x)) + \boldsymbol{\lambda}\varepsilon + o(\varepsilon)\end{aligned}$$

for small v, ε. Since f is non-decreasing, combining the above inequalities for $\varepsilon = \langle \mathbf{p},\, v\rangle_x + o(|v|_x)$ (observe that $\varepsilon \to 0$ as $|v|_x \to 0$) we have:

$$\begin{aligned}f(\varphi(\exp_x v)) &\le f(\varphi(x) + \langle \mathbf{p},\, v\rangle_x + o(|v|_x))\\ &\le f(\varphi(x)) + \boldsymbol{\lambda}\langle \mathbf{p},\, v\rangle_x + \boldsymbol{\lambda} o(|v|_x) + o(\langle \mathbf{p},\, v\rangle_x + o(|v|_x))\\ &= f(\varphi(x)) + \langle \boldsymbol{\lambda}\mathbf{p},\, v\rangle_x + o(|v|_x)\end{aligned}$$

since

$$\begin{aligned}\frac{o(\langle \mathbf{p},\, v\rangle_x + o(|v|_x))}{|v|_x} &= \frac{o(\langle \mathbf{p},\, v\rangle_x + o(|v|_x))}{\langle \mathbf{p},\, v\rangle_x + o(|v|_x)} \cdot \frac{\langle \mathbf{p},\, v\rangle_x + o(|v|_x)}{|v|_x}\\ &= \frac{o(\varepsilon)}{\varepsilon} \cdot \left(\left\langle \mathbf{p},\, \frac{v}{|v|_x} \right\rangle_x + \frac{o(|v|_x)}{|v|_x} \right)\\ &\le \frac{o(\varepsilon)}{\varepsilon} \cdot \left(|\mathbf{p}|_x + \frac{o(|v|_x)}{|v|_x} \right) \xrightarrow{|v|_x \to 0} 0\end{aligned}$$

thus $o(\langle \mathbf{p},\, v\rangle_x + o(|v|_x)) = o(|v|_x)$ and, since $\boldsymbol{\lambda} o(|v|_x) + o(|v|_x) = o(|v|_x)$, we conclude that $\boldsymbol{\lambda}\mathbf{p} \in \overline{\partial} f \circ \varphi(x)$. $\square$

Now, we can give a general result regarding the differentiability of d_y^2:

Theorem. *For every $y \in M$, $d_y^2/2$ is superdifferentiable at any $x \in M$, with supergradient*

$$\dot\gamma(1) \in \overline{\partial}(d_y^2/2)(x)$$

where $\gamma : [0,\, 1] \to M$ is a minimizing geodesic from $y = \gamma(0)$ to $x = \gamma(1)$. In particular, it is differentiable around any $x \notin \mathrm{cut}(y)$.

Proof. Let $x \in M$ and $\gamma : [0,\, 1] \to M$ be a minimizing geodesic from $y = \gamma(0)$ to $x = \gamma(1)$, parametrized with constant speed. Take $z \in W$ to be any point lying on γ near the endpoint $x = \gamma(1)$. Applying the previous proposition to z instead of y yields

$$\nabla d_z(x) = \frac{\dot\gamma(1)}{|\dot\gamma(1)|_x}$$

as the chain rule dictates, since $d_z(x) = \sqrt{2(d_z^2/2)(x)}$. The triangle inequality gives us:

$$\begin{aligned} d(y,\, \exp_x v) &\leq d(y,\, z) + d(z,\, \exp_x v) \\ &= d(y,\, z) + d(z,\, x) + \langle \dot\gamma(1)/|\dot\gamma(1)|_x,\, v\rangle + o(|v|_x) \\ &= d(y,\, x) + \langle \dot\gamma(1)/|\dot\gamma(1)|_x,\, v\rangle + o(|v|_x) \end{aligned}$$

so that $d_y = \sqrt{2(d_y^2/2)}$ is superdifferentiable at x. Applying the one-sided chain rule with $\varphi = d_y$ and $f(r) = r^2/2,\ r \geq 0$ we get:

$$\dot\gamma(1) = d(y,\, x) \cdot \dot\gamma(1)/|\dot\gamma(1)|_x \in \overline{\partial} f \circ \varphi(x) = \overline{\partial}(d_y^2/2)(x)$$

and that completes the proof, since the second part was proved earlier. □

We end the section with a characterization of the cut locus of a point, which we'll definitely use later. The cut locus $\mathrm{cut}(y)$ consists of two kinds of points: (i) those connected to y by multiple minimizing geodesics and (ii) those which are conjugate to y but do not fall into class (i). As we've just seen, in case (i), first order differentiability of $d_y^2/2$ fails, but it doesn't, in case (ii). Although, case (ii), is exactly where second order differentiability must fail:

Proposition. *At every $x \in \mathrm{cut}(y)$ we must have:*

$$\inf_{0<|v|_x<1} \frac{d_y^2(\exp_x v)/2 + d_y^2(\exp_x(-v))/2 - 2d_y^2(x)/2}{|v|_x^2} = -\infty$$

Proof. We treat each case separately:
Case (i): There exist two distinct minimal geodesics joining x and y, thus there also exist two distinct supergradients $\mathbf{p_1}, \mathbf{p_2} \in \overline{\partial}(d_y^2/2)(x)$. Let $N \in \mathbb{N}$ and $v \in T_xM$ be a small vector and assume, without loss of generality, that $\langle \mathbf{p_1} - \mathbf{p_2},\, v\rangle < 0$ (if that's not the case, replace $\mathbf{p_1} \leftrightarrow \mathbf{p_2}$). Superdifferentiability gives:

$$d_y^2(\exp_x v)/2 \leq d_y^2(x)/2 + \langle \mathbf{p_1},\, v\rangle + o(|v|_x)$$

and

$$d_y^2(\exp_x(-v))/2 \leq d_y^2(x)/2 + \langle \mathbf{p_1},\, -v\rangle + o(|v|_x)$$

so that:

$$\frac{d_y^2(\exp_x v)/2 + d_y^2(\exp_x(-v))/2 - 2d_y^2(x)/2}{|v|_x^2}$$

$$\leq \left(\left\langle \mathbf{p_1} - \mathbf{p_2}, \frac{v}{|v|_x} \right\rangle_x + \frac{o(|v|_x)}{|v|_x} \right) \frac{1}{|v|_x}$$

but the term in the brackets can be made negative, since $\frac{o(|v|_x)}{|v|_x} \to 0$ as $|v|_x \to 0$ and the fraction outside would blow up, so that the infimum diverges to $-\infty$, for small v.

Case (ii): Only one minimal geodesic joins x to y, but they're conjugate points, hence, some non-zero normal Jacobi field along this geodesic vanishes at both endpoints. Assume that there exists a constant $C > 0$ such that:

$$\liminf_{v \to 0} \frac{d_y^2(\exp_x v)/2 + d_y^2(\exp_x(-v))/2 - 2d_y^2(x)/2}{|v|_x^2} \geq -C$$

and let $\gamma(t) := \exp_x(tu)$ be the minimal geodesic joining $x = \gamma(0)$ to $y = \gamma(1)$. Let $Y(t)$ be a non-zero normal Jacobi field along γ vanishing at 0 and 1. By scaling the overall size of the manifold and the vector field independently, it costs no generality to normalize the length of the geodesic so that $d(x, y) = |u|_x = 1$ and take $v := Y'(0)$ to be a unit vector. Let Z_1 be a parallel vector field along γ with $Z_1(0) = Y'(0) = v$ and let $Z(t) := (1-t)Z_1(t)$. Fix $\varepsilon > 0$ small enough such that:

$$-\frac{2}{\varepsilon} + I(Z, Z) < -C$$

where I denotes the index form along γ, and for this ε consider

$$Y_\varepsilon(t) := Y(t) + \varepsilon Z(t)$$

which is a normal field along γ, satisfying $Y_\varepsilon(0) = \varepsilon v$ and $Y_\varepsilon(1) = 0$. Now, introduce the following variation of the geodesic γ:

$$\gamma_s(t); = f(s, t) := \exp_{\gamma(t)}(sY_\varepsilon(t))$$

for s close to 0. The curve γ_s joins the point $\exp_x(s\varepsilon v)$ to y. From H́older's inequality gives:

$$\frac{1}{2}d^2(\exp_x(s\varepsilon v), y) \leq \frac{1}{2}\left(\int_0^1 |\dot{\gamma}_s(t)|_{\gamma(t)}\right)^2 \leq \frac{1}{2}\left(\int_0^1 |\dot{\gamma}_s(t)|^2_{\gamma(t)}\right) = E(s)$$

where E is the energy functional of the curve γ_s. Note, that for $s = 0$, since γ is a minimizing geodesic, there is equality above. So, now, our assumption gives:

$$\liminf_{s\to 0} \frac{E(s) + E(-s) - 2E(0)}{(\varepsilon s)^2} \geq -C$$

while, on the other hand, since $s \mapsto f(s, t)$ is a geodesic, for fixed t, we have $\nabla_s \frac{\partial f}{\partial s}(s, t) = 0$ and thus:

$$\begin{aligned}\frac{d^2 E(\gamma_s)}{ds^2}(0) &= I(Y_\varepsilon, Y_\varepsilon) \\ &= I(Y, Y) + 2\varepsilon I(Z, Y) + \varepsilon^2 I(Z, Z) \\ &= 2\varepsilon I(Z, Y) + \varepsilon^2 I(Z, Z)\end{aligned}$$

since Y is a Jacobi field vanishing at endpoints, hence $I(Y, Y) = 0$. Additionally, $I(Z, Y) = \Big[\langle Z(t), Y'(t)\rangle\Big]_0^1 = -|v|_x^2 = -1$. Thus,

$$\begin{aligned}\frac{d^2 E(\gamma_s)}{ds^2}(0) &= -2\varepsilon + \varepsilon^2 I(Z, Z) \geq -C \\ \Leftrightarrow -\frac{2}{\varepsilon} &+ I(Z, Z) \geq -\frac{C}{\varepsilon^2} \geq -C\end{aligned}$$

which contradicts the choice of ε. Hence, the given infimum diverges to $-\infty$ in both cases. $\square$

Chapter 3

Optimal Transport

A classical approach to solve Kantorovich's problem is to turn it into a dual problem, which will be easier, hopefully. Let's describe a little bit the duality, without getting into details. If one's desire is more details we redirect the reader to [1], [2], [5], [29]. Recall that Kantorovich's problem asks for minimization of the **total transportation cost**:

$$\mathscr{I}(\pi) := \int_{X\times Y} c(x,\, y)d\pi(x,\, y)$$

over $\pi \in \Pi(\mu,\, \nu)$. Now, for a non-negative measure π on $X \times Y$ one has

$$\sup_{\varphi,\,\psi\in L^1}\left[\int_X \varphi d\mu + \int_Y \psi d\nu - \int_{X\times Y} \varphi(x) + \psi(y) d\pi\right] = \begin{cases} 0, & \text{if } \pi \in \Pi(\mu,\, \nu) \\ +\infty, & \text{otherwise} \end{cases}$$

Hence we can remove the constraints on π if we add the previous sup to $\inf_{\pi\in\Pi(\mu,\,n)} \mathscr{I}(\pi)$, since if they are satisfied nothing has been added and if they are not one gets $+\infty$ which will be avoided by the minimization. Now a minimax theorem by Rockafella ([10]) lets us write

$$\begin{aligned}\inf_{\pi}&\left\{\int c d\pi + \sup_{\varphi,\,\psi}\left[\left[\int_X \varphi d\mu + \int_Y \psi d\nu - \int_{X\times Y} \varphi(x) + \psi(y) d\pi\right]\right]\right\} \\ &= \sup_{\varphi,\,\psi}\left[\int_X \varphi d\mu + \int_Y \psi d\nu + \inf_{\pi}\int_{X\times Y} c(x,\, y) - \varphi(x) - \psi(y) d\pi\right]\end{aligned}$$

Since π is non-negative the last term is 0 if $\varphi(x) + \psi(y) \le c(x,\, y)$ for all $(x,\, y) \in X \times Y$ and $-\infty$ otherwise. This leads to the **dual problem** in which we need to maximize the quantity:

$$J(\varphi, \psi) := \int_X \varphi d\mu + \int_Y \psi d\nu$$

over all $\varphi \in L^1(\mu)$, $\psi \in L^1(\nu)$ such that $\varphi \oplus \psi \leq c$, where $\varphi \oplus \psi(x, y) := \varphi(x) + \psi(y)$.

Similarily we deduce a dual problem for Monge's problem. We will solve this problem in the case where $X, Y \subseteq M$ are compact, μ, ν are probability measures, compactly supported on X and Y, respectively and $c(x, y) = d^2(x, y)/2$, since it suffices for our purpose and has the advantage of following the work of McCann ([30]). We will denote by $X \subset\subset M$ an open subset of M with compact closure $\overline{X}$.

Our test functions can be limited to continuous functions, so we define:

$$\mathrm{Lip}_c := \{(\varphi, \psi) \in C(X) \times C(Y) \mid \varphi \oplus \psi \leq c\}$$

and begin our journey to maximize J over Lip_c and provide a solution to Monge's problem. As to why this restriction works, the answer is a special class of functions, which have many helpful properties that we will exploit. For example, if they are not infinite, they have the same modulus of continuity with the cost function.

3.1 c-transforms & c-concave functions

Let $X, Y \subseteq M$ be compact. The **c-transform** of a function $\varphi : X \to \overline{\mathbb{R}}$ is a function $\varphi^c : Y \to \overline{\mathbb{R}}$ defined as:

$$\varphi^c(y) := \inf_{x \in X}\{c(x, y) - \varphi(x)\}$$

for every $y \in Y$. We define the **c-transform**, $\psi^c : X \to \overline{\mathbb{R}}$, of a function $\psi : Y \to \overline{\mathbb{R}}$ in a similar fashion as:

$$\psi^c(x) := \inf_{y \in Y}\{c(x, y) - \psi(y)\}$$

for every $x \in X$. Moreover, we say that a function φ defined on X is **c-concave** if there exists ψ such that $\varphi = \psi^c$ and we denote their set by $\mathcal{I}^c(X, Y)$. Similarly for functions ψ defined on Y, as for their set, it will be denoted by $\mathcal{I}^c(Y, X)$. In what follows we will be omitting the proofs for functions $\psi : Y \to \overline{\mathbb{R}}$ as the arguments will be symmetrical. Not all functions have the property of being c-concave:

Proposition. *For any $\varphi : X \to \overline{\mathbb{R}}$ we have $\varphi^{cc} \geq \varphi$ and φ^{cc} is the smallest c-concave function greater than φ. Equality holds if and only if φ is c-concave.*

Proof. Let $x \in X$. Since

$$\varphi^c(y) = \inf_{z\in X}\{c(z, y) - \varphi(z)\} \leq c(x, y) - \varphi(x)$$

one gets:

$$\varphi^{cc}(x) = \inf_{y\in Y}\{c(x, y) - \varphi^c(y)\} \geq \inf_{y\in Y}\{c(x, y) - c(x, y) + \varphi(x)\} = \varphi(x)$$

hence $\varphi^{cc} \geq \varphi$. Now, take a c-concave $\tilde{\varphi} = \chi^c \geq \varphi$. Then one has

$$\chi^{cc} = \inf_{y\in Y}\{c(\cdot, y) - \chi^c(y)\} \leq \inf_{y\in Y}\{c(\cdot, y) - \varphi(y)\} = \varphi^c$$

so that $\chi \leq \chi^{cc} \leq \varphi^c$ and by the same argument $\tilde{\varphi} = \chi^c \geq \varphi^{cc} \geq \varphi$.

Now, if $\varphi = \varphi^{cc} = (\varphi^c)^c$, then $\varphi(x) = \inf_{y\in Y}\{c(x, y) - \varphi^c(y)\}$, thus φ is c-concave. Conversely, let $\varphi = \zeta^c$ for some function $\zeta : Y \to \overline{\mathbb{R}}$. Then one has $\varphi^c = \zeta^{cc} \geq \zeta$ so that $\varphi^{cc} \leq \zeta^c = \varphi$ by the same argument used before twice and since the reverse inequality always holds we have established equality. □

This repeated technique used in the above proof, that c-transforms reverse the inequality, when combined with the result itself gives us a

Corollary. *For any $\varphi : X \to \overline{\mathbb{R}}$*

$$\varphi^{ccc} = \varphi^c$$

Proof. On the one hand

$$\varphi^{cc} \geq \varphi \Rightarrow (\varphi^{cc})^c \leq \varphi^c \Rightarrow \varphi^{ccc} \leq \varphi^c$$

while on the other hand

$$(\varphi^c)^{cc} \geq \varphi^c \Rightarrow \varphi^{ccc} \geq \varphi^c$$

so that $\varphi^{ccc} = \varphi^c$. □

For $\varphi \in \mathcal{I}^c(X, Y)$ there exists some $\psi : Y \to \overline{\mathbb{R}}$ such that

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

If ψ is not bounded from above, then we will be taking infimum on arbitrarily large negative numbers, so that $\varphi = -\infty$. If it is and at least one $y \in M$ exists such that $\psi(y) \in \mathbb{R}$ then $\varphi(x)$ would be finite. But if we have $\psi = -\infty$, then $\varphi(x) = +\infty$ for every $x \in M$, so that $\varphi = +\infty$.

If we focus on functions who are bounded from above and not identically $-\infty$ the modulus of continuity of their c-transform will be completely determined by the cost function:

Proposition. *Any c-concave $\varphi \in \mathcal{I}^c(X, Y)$, not identically infinite, is Lipschitz continuous on X.*

Proof. Let $\varepsilon > 0$ and $x, z \in X$. By definition, there exists a $y \in Y$ such that $\varphi(z) + \varepsilon \geq c(z, y) - \varphi^c(y)$, while $\varphi(x) \leq c(x, y) - \varphi^c(y)$ holds trivially. Subtracting these two inequalities, the Lipschitz continuity of c with constant D, yields:

$$|\varphi(x) - \varphi(z)| \leq |c(x, y) - c(z, y)| + \varepsilon \leq Dd(x, z) + \varepsilon$$

where D is the upper bound of d_y on X. As $\varepsilon > 0$ and $x, z \in X$ was arbitrarily chosen we have that ψ is Lipschitz, with Lipschitz constant $\leq D$. $\square$

From now on, we will be talking only for not identically infinite c-concave functions.

Corollary. *In view of Rademacher's Theorem, any c-concave function $\varphi \in \mathcal{I}^c(\overline{X}, Y)$, where $X \subset\subset M$, is differentiable* $\mathrm{vol}_{\mathrm{g}}$*-almost everywhere on $\overline{X}$ and its gradient is a Borel map.*

Observe that, from the definition of c-transform one has:

$$\varphi(x) + \varphi^c(y) \leq \varphi(x) + c(x, y) - \varphi(x) \leq c(x, y)$$

for every $x \in X$ and $y \in Y$. We can get a criterion for equality:

Theorem. *Let $X \subset\subset M$ open and Y compact as always. Let $x \in X$ such that $\varphi \in \mathcal{I}^c(\overline{X}, Y)$ is differentiable at. Then*

$$\varphi(x) + \varphi^c(y) = c(x, y) \iff y = \exp_x(-\nabla\varphi(x))$$

Moreover $\nabla\varphi(x) = \nabla(d_y^2/2)(x)$.

Proof. ($\Rightarrow$) Suppose that a $y \in Y$ is given such that equality holds. Then, one has

$$c(z, y) - \varphi(z) - \varphi^c(y) \geq 0 = c(x, y) - \varphi(x) - \varphi^c(y)$$

for every $z \in X$. For $d_y^2(z)/2 = c(z, y)$ one has

$$\begin{aligned} d_y^2(\exp_x v)/2 = c(\exp_x v, y) &\geq c(x, y) - \varphi(x) - \varphi^c(y) + \varphi(\exp_x v) + \varphi^c(y) \\ &= c(x, y) - \varphi(x) + \varphi(\exp_x v) \\ &= c(x, y) - \varphi(x) + \varphi(x) + g_x(\nabla\varphi(x), v) + o(|v|_x) \\ &= d_y^2(x)/2 + g_x(\nabla\varphi(x), v) + o(|v|_x) \end{aligned}$$

since φ is differentiable at x. It follows that $d_y^2/2$ has subgradient $\nabla\varphi(x) \in \underline{\partial}(d_y^2/2)(x)$ at x. On the other hand, we've already shown that $d_y^2/2$ has supergradient $\dot\gamma(1) \in \overline{\partial}(d_y^2/2)(x)$, for a minimizing geodesic γ connecting $y = \gamma(0)$ with $x = \gamma(1)$. Thus $d_y^2/2$ is differentiable at x and

$$\begin{aligned} \nabla\varphi(x) = \dot\gamma(1) = \nabla(d_y^2/2)(x) &= -(\exp_x)^{-1}(y) \\ \Rightarrow \quad y &= \exp_x(-\nabla\varphi(x)) \end{aligned}$$

and we're done.

($\Leftarrow$) We have now established that, for our fixed $x \in X$, there exists at most one point $y \in Y$ such that the equality holds. We have to show that there exists at least one such y. The c-concavity hasn't played its role yet. From compactness, the infimum

$$\varphi^{cc}(x) = \inf_{y \in Y}\{c(x, y) - \varphi^c(y)\}$$

is attained at some point $y \in Y$. Since $\varphi = \varphi^{cc}$, the same point produces the equality in question. But our previous argument shows $y = \exp_x(-\nabla\varphi(x))$ and the statement is proved. ☐

3.2 Duality

If $(\varphi, \psi) \in \mathrm{Lip}_c$, since φ is continuous and X is compact, its c-transform φ^c is finite-valued, hence Lipschitz continuous. Additionally, the definition of c-transform gives $\varphi \oplus \varphi^c \leq c$ so that $(\varphi, \varphi^c) \in \mathrm{Lip}_c$. On the other hand, for every $x \in M$ one has

$$\begin{aligned} (\varphi, \psi) \in \mathrm{Lip}_c &\Rightarrow \psi(y) \leq c(x, y) - \varphi(x) \\ &\Rightarrow \psi(y) \leq \inf_{x \in M}\{c(x, y) - \varphi(x)\} = \varphi^c(y) \end{aligned}$$

hence

$$\begin{aligned} J(\varphi,\, \psi) &= \int_M \varphi(x)d\mu(x) + \int_M \psi(y)d\nu(y) \\ &\leq \int_M \varphi(x)d\mu(x) + \int_M \varphi^c(y)d\nu(y) = J(\varphi, \varphi^c) \end{aligned}$$

We've just proven:

Lemma. *If $(\varphi,\, \psi) \in \mathrm{Lip}_c$, then $(\varphi,\, \varphi^c) \in \mathrm{Lip}_c$ and $J(\varphi,\, \psi) \leq J(\varphi,\, \varphi^c)$.*

Observe that the symmetry $\varphi \leftrightarrow \psi$ lets us apply the same arguments and get that $(\psi^c,\, \psi) \in \mathrm{Lip}_c$ and $J(\varphi,\, \psi) \leq J(\psi^c,\, \psi)$. Applying them once more with respect to ψ^c, as before, we conclude that $(\psi^c,\, \psi^{cc}) \in \mathrm{Lip}_c$ and $J(\varphi,\, \psi) \leq J(\psi^c,\, \psi) \leq J(\psi^c,\, \psi^{cc})$.

We will prove that the duality problem admits a solution among the c-concave functions:

Proposition. *If μ, ν are Borel probability measures on M such that X and Y contain their support, respectively, then there exists some $\varphi \in \mathcal{I}^c(X,\, Y)$, such that*

$$J(\varphi,\, \varphi^c) = \max_{(u,\, v)\in \mathrm{Lip}_c} J(u,\, v).$$

Proof. Choose a sequence $(\varphi_n,\, \psi_n) \in \mathrm{Lip}_c$ such that

$$J(\varphi_n,\, \psi_n) \to \sup_{(u,\, v)\in \mathrm{Lip}_c} J(u,\, v).$$

According to the previous proposition, $J(\psi_n^c,\, \psi_n^{cc})$ also tends to $\sup_{(u,\, v)\in \mathrm{Lip}_c} J(u,\, v)$ in Lip_c. Observe that, because $\mu(X) = \nu(Y) = 1$, the sequence

$$(u_n,\, v_n) := (\psi_n^c - \psi_n(z),\, \psi_n^{cc} + \psi_n(z))$$

for some fixed $z \in M$, is in Lip_c and has J tend to its supremum. From before, we know that u_n and v_n will have the same modulus of continuity. Also, for every $n \in \mathbb{N}$,

$$|u_n(x)| = |u_n(x) - u_n(z)| \leq Dd(x,\, z) \leq D^2$$

and

$$\begin{aligned} |v_n(y)| = |\inf_{x\in M}\{c(x,\, y) - u_n(x)\}| &\leq |c(x,\, y)| + D^2 \\ &\leq \frac{D^2}{2} + D^2 = \frac{3}{2}D^2 \end{aligned}$$

because the c-transform of any constant is just its negative. As the above equicontinuous families of functions are uniformly bounded though X and Y respectively, the *Ascoli-Arzela theorem* ([4]) extracts a subsequence (u_{k_n}, v_{k_n}) that converges pointwise to some $(u_0, v_0) \in \mathrm{Lip}_c$. Since the measures are finite we can apply the Dominated Convergence theorem ([7]) and get

$$J(u_{k_n}, v_{k_n}) \to J(u_0, v_0)$$

so that

$$J(u_0, v_0) = \sup_{(u,v)\in \mathrm{Lip}_c} J(u, v) = \max_{(u,v)\in \mathrm{Lip}_c} J(u, v)$$

and, hence, $J(v_0^c, v_0^{cc}) = J(u_0, v_0)$, by maximality. Setting $\varphi := v_0^c$ completes the proof, as $v_0^{ccc} = v_0^c$. □

3.3 Monge's problem & McCann's Theorem

Now we can choose a c-concave φ such that $(\varphi, \psi) = (\varphi^{cc}, \varphi^c) \in \mathrm{Lip}_c$ maximizes $J(\varphi, \psi) = \int_M \varphi d\mu + \int_M \psi d\nu$. We have shown that such functions are both Lipschitz, so $\nabla\varphi : X \to \mathbb{R}$ is defined μ-almost everywhere on X and is a Borel map, as Rademacher's Theorem shows. For Monge's problem, we obviously have:

$$\begin{aligned} J(u, v) &= \int_M u(x)d\mu(x) + \int_M v(T(x))d\mu(x) \\ &\le \int_M c(x, T(x))d\mu(x) \\ &= \mathscr{I}(T) \end{aligned}$$

for any $(u, v) \in \mathrm{Lip}_c$ and $T \in \Pi(\mu, \nu)$. Since both sides are finite one has:

$$\sup_{(u,v)\in \mathrm{Lip}_c} J(u, v) = J(\varphi, \varphi^c) \le \inf_{T\in\Pi(\mu,v)} \int_M c(x, T(x))d\mu(x)$$

for some $\varphi \in \mathcal{I}^c(X, Y)$. As one can see the only obstacle is the inequality

$$\varphi(x) + \varphi^c(T(x)) \le c(x, T(x))$$

but we have shown that $\varphi(x) + \varphi^c(y) = c(x, y) \iff y = \exp_x(-\nabla\varphi(x))$ which proposes an idea for the minimizer. In particular, if we set $F(x) := \exp_x(-\nabla\varphi(x))$ for every x at which φ is differentiable we get from Rademacher's Theorem that it is a Borel map. All we need to show is that $F \in \Pi(\mu, \nu)$ since then, the almost-everywhere equality will turn all the above inequalities into equalities. We can see that this is exactly the case:

Proposition. *The map* $F(x) = \exp_x(-\nabla\varphi(x))$ *pushes* μ *forward to* ν*, i.e.* $F \in \Pi(\mu, \nu)$*. As a consequence, Monge's problem admits a solution.*

Proof. In order to show that $F \in \Pi(\mu, \nu)$ it suffices to check that, for every $f \in C(M)$, $\int_M f d\nu = \int_M f d(F_{\#}\mu)$. For this purpose, let $f \in C(M)$, $x_0 \in X$ where φ is differentiable and define perturbations

$$\begin{aligned}\psi_\varepsilon(y) &:= \varphi^c(y) + \varepsilon f(y)\\ \varphi_\varepsilon(x) &:= \psi_\varepsilon^c(x) = \inf_{y\in Y}\{c(x, y) - \varphi^c(y) - \varepsilon f(y)\}\end{aligned}$$

for $x, y \in M$ and $|\varepsilon| < 1$. Since the above are continuous, the compactness of Y assures us that the infimum is attained. For $\varepsilon = 0$ it is attained uniquely at $y = \exp_{x_0}(-\nabla\varphi(x_0)) = F(x_0)$, according to our previous theorem, so for small enough ε, it must be attained at some nearby point $y_\varepsilon = F(x_0) + \delta(\varepsilon)$ where $\delta(\varepsilon) \xrightarrow{\varepsilon\to 0} 0$. To see why this is true, set $g(y) := c(T(x_0), y) - \varphi^c(y)$ and $g_{1/n}(y) := c(T(x_0), y) - \varphi^c(y) - \frac{1}{n}f(y)$ and take a sequence $y_{1/n} \in \operatorname{argmin} g_{1/n}$. Because of compactness, there exists a subsequence which converges to some point y_0. But, this point must be $T(x_0)$ because $g_{1/n}(y_{1/n}) \leq g_{1/n}(y)$ for every $y \in Y$ and, since f is continuous and Y is compact, the uniform convergence of $g_{1/n}$ to g will give us $g(y_0) \leq g(y)$ for every $y \in Y$. Thus

$$c(x_0, F(x_0)) - \varphi^c(F(x_0)) - \varepsilon f(y_\varepsilon) \leq \varphi_\varepsilon(x_0) \leq c(x_0, y) - \varphi^c(y) - \varepsilon f(y)$$

for all $y \in M$. Setting $y = F(x_0)$ the continuity of f gives us

$$\varphi_\varepsilon(x) = \varphi(x) - \varepsilon f(F(x)) + o(\varepsilon)$$

for μ-almost every $x \in X$. Since $J(\varphi_\varepsilon, \psi_\varepsilon)$ attains its maximum at $\varepsilon = 0$, a simple application of the dominated convergence theorem yields:

$$\begin{aligned}0 = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} J(\varphi_\varepsilon, \psi_\varepsilon) &= \lim_{\varepsilon\to 0}\frac{J(\varphi_\varepsilon, \psi_\varepsilon) - J(\varphi, \psi)}{\varepsilon}\\ &= \lim_{\varepsilon\to 0}\int_M \frac{\varphi_\varepsilon - \varphi}{\varepsilon} d\mu + \int_M f d\nu\\ &= \int_M -f\circ F d\mu + \int_M f d\nu\\ &= -\int_M f d(F_{\#}\mu) + \int_M f d\nu\end{aligned}$$

so that $F_{\#}\mu = \nu$ is established by the *Riesz representation theorem* ([6]), since $f \in C(M)$ was arbitrary. $\square$

We will now gather these results and prove this map's uniqueness. As we've already said this whole work is by McCann ([30]) in which he generalized Brenier's theorem ([18], [19], [20]). In order to make it easier for us to refer to the following theorem, we will name it "McCann's Theorem".

Theorem (McCann's Theorem)**.** *Let (M, g) be a Riemannian manifold with two Borel probability measures $\mu << \mathrm{vol_g}$ and ν on M such that X and Y contain their support, respectively. Then, there exists a c-concave $\varphi \in \mathcal{I}^c(X, Y)$, such that the map*

$$F(x) := \exp_x(-\nabla\varphi(x))$$

is the unique minimizer of the Monge's transportation problem, among all Borel maps that push μ forward to ν (modulo discrepancies on sets of zero μ-measure).

Proof. Our previous conversation is enough proof of existence. We will prove uniqueness in two steps:

- First, suppose that $G \in \Pi(\mu, \nu)$ is also a minimizer of Monge's problem, so it satisfies $\mathscr{O}(G) = J(\varphi, \varphi^c)$, hence

  $$\int_M c(x, G(x)) - \varphi(x) - \varphi^c(G(x))d\mu(x) = 0.$$

  Since $(\varphi, \varphi^c) \in \mathrm{Lip}_c$ the integrand is non-negative, thus $\varphi(x) + \varphi^c(G(x)) = c(x, G(x))$ holds μ-almost everywhere on X and the criterion for this equality yields $G(x) = F(x)$ μ-almost everywhere on X.

- Now, if $u = u^{cc}$ is another c-concave function maximizing J for which the map $G(x) = \exp_x(-\nabla u(x))$ is in $\Pi(\mu, \nu)$, then G is also a minimizer of Monge's problem and our previous argument shows that G and F must be equal μ-almost everywhere on X.

□

Corollary. *In view of the previous theorem with the extra demand that $\nu << \mathrm{vol_g}$, one has:*

$$\tilde{F} \in \Pi(\nu, \mu) \text{ where } \tilde{F}(y) := \exp_y(-\nabla\varphi^c(y))$$

and satisfies $\tilde{F}(F(x)) = x$ μ-a.e. on X and $F(\tilde{F}(y)) = y$ ν-a.e. on Y.

Proof. Denoting $\psi := \varphi^c$ we shall exploit the symmetry $\mu \leftrightarrow \nu$ and $(\varphi,\, \varphi^c) \leftrightarrow (\psi,\, \psi^c)$. We know that ψ is differentiable on $\mathrm{dom}\nabla\psi$ where $\nu(\mathrm{dom}\nabla\psi) = 1$, so $F_{\#}\mu = \nu$ implies $U := \mathrm{dom}\nabla\varphi \cap F^{-1}(\mathrm{dom}\nabla\psi)$ is a Borel set of full measure $\mu(U) = 1$. For $x \in U$ we have that

$$c(F(x),\, x) - \psi(F(x)) - \psi^c(x) = c(x,\, F(x)) - \varphi^c(F(x)) - \varphi(x) = 0$$

Since $F(x) \in \mathrm{dom}\nabla\psi$ we must have

$$x = \exp_{F(x)}(-\nabla\psi(F(x)))$$

thus $x = \tilde{F}(F(x))$ on U, which means μ-a.e. on X. It follows that given any continuous function $f \in C(M)$ we have

$$\int_M f d(\tilde{F}_{\#}\nu) = \int_M f(\tilde{F}(y)))d\nu(y) = \int_M f(\tilde{F}(F(x))))d\mu(x) = \int_M f(x)d\mu(x)$$

so that $\tilde{F} \in \Pi(\nu,\, \mu)$. By a symmetric argument we have that

$$V := \mathrm{dom}\nabla\psi \cap \tilde{F}^{-1}(\mathrm{dom}\nabla\varphi)$$

is a Borel set of full measure $\nu(V) = 1$ and similarly $F(\tilde{F}(y)) = y$ ν-a.e. on Y. $\square$

3.4 c-superdifferential of a c-concave function

A c-concave function is not necessarily differentiable everywhere and thus the mass transport map F is not defined everyhwere. However for fixed $x \in X$, the Lipschitz continuity of φ^c on the compact set $Y \subseteq M$ guarantees that some $y \in Y$ provides equality in $\varphi(x) + \varphi^c(y) \leq c(x,\, y)$. This motivates a, possibly multivalued, extension of F to all of X:

Definition. Let $X,\, Y \subseteq M$ be compact sets. For $\mathcal{I}^c(X,\, Y)$ and $x \in X$, the **c-superdifferential of φ at x** is the non-empty set

$$\begin{aligned}\partial^c\varphi(x) :=& \{y \in Y \mid \varphi(x) + \varphi^c(y) = c(x,\, y)\}\\ =& \{y \in Y \mid \forall z \in X\ \varphi(z) \leq \varphi(x) + c(y,\, z) - c(x,\, y)\}\end{aligned}$$

It is an extension of F in the following sense: if $\varphi \in \mathcal{I}^c(\overline{X},\, Y)$ is differentiable at $x \in X \subset\subset M$, then $\partial^c\varphi(x) = \{F(x)\} = \{\exp_x(-\nabla\varphi(x))\}$.

Recall the definition of the superdifferential at x:

$$\overline{\partial}\varphi(x) = \{\mathbf{p} \in T_xM \mid \varphi(\exp_x v) \leq \varphi(x) + g_x(\mathbf{p},\, v) + o(|v|_x),\, v \in T_xM \text{ small}\}$$

and observe their relationship:

Proposition. *Let $(x,\, y) \in X \times Y$ where $X \subset\subset M$ open and $Y \subset M$ compact and $\varphi \in \mathcal{I}^c(\overline{X},\, Y)$. If $\boldsymbol{p} \in T_xM$ satisfies $\exp_x(-\boldsymbol{p}) = y$ and $|\boldsymbol{p}| = d(x,\, y)$, then*

$$y \in \partial^c\varphi(x) \Rightarrow \boldsymbol{p} \in \overline{\partial}\varphi(x).$$

Proof. Let $(x,\, y) \in X \times Y$ such that $y \in \partial^c\varphi(x)$. Then

$$\varphi(z) \leq \varphi(x) + c(z,\, y) - c(x,\, y)$$

holds for every $z \in X$. We know that $\mathbf{p} \in \overline{\partial}(d_y^2/2)(x)$, thus for $z = \exp_x v$ we have:

$$c(\exp_x v,\, y) - c(x,\, y) \leq g_x(\mathbf{p},\, v) + o(|v|_x)$$

for small $v \in T_xM$. Combining the above we get

$$\varphi(\exp_x v) \leq \varphi(x) + g_x(\mathbf{p},\, v) + o(|v|_x)$$

which means $\mathbf{p} \in \overline{\partial}\varphi(x)$. □

3.5 Semi-concavity

It turns out that c-concave functions are only a part of a much more general class of functions admitting non-empty superdifferentials. Moreover, these functions admit a Hessian almost everywhere, known as *Alexandrov's Theorem* in the Euclidean case and *Bangert's Theorem* in the Riemannian case. These are the semi-concave functions:

Definition. Let $U \subseteq M$ open. A function $\varphi : U \to \mathbb{R}$ is **semi-concave at $\mathbf{x} \in \mathbf{U}$** if there exists a convex normal neighbourhood $B_r(x) \in M$ and a smooth function $h : B_r(x) \to \mathbb{R}$ such that $\varphi + h$ is *geodesically concave* thoughout $B_r(x)$, i.e. $(\varphi + h) \circ \gamma$ is concave for any geodesic $\gamma : [0,\, 1] \to B_r(x)$. Obviously, φ is said to be **semi-concave on U** if it is semi-concave at each point $x \in U$.

3.6 Hessian of semi-concave functions

Definition. Let $\varphi : U \to \mathbb{R}$ be semi-concave on an open set $U \subseteq M$. We say that φ admits a **Hessian** $\mathrm{Hess}_{\mathbf{x}}\,\varphi := \mathbf{H}$ **at** $\mathbf{x} \in \mathbf{U}$ if φ is differentiable at x and there exists a self-adjoint operator $H : T_xM \to T_xM$ satisfying

$$\sup_{u \in \overline{\partial}\varphi(\exp_x v)} |P_{x,v}u - \nabla\varphi(x) - Hv| = o(|v|_x)$$

for small $v \in T_xM$, where $P_{x,v} : T_{\exp_x v}M \to T_xM$ denotes the parallel transport on $\gamma(t) := \exp_x tv$.

This definition coincides with the usual one for smooth functions. To understand intuitively the existence of Hessian note that it is equivalent to the existence of a second order Taylor expansion for φ around x:

$$\varphi(\exp_x v) = \varphi(x) + g_x(\nabla\varphi(x),\, v) + \frac{1}{2}g_x(Hv,\, v) + o(|v|_x^2)$$

Let us formulate Alexandrov-Bengert Theorem where its proof is omitted because of computational complexity and divergency of our subject, but we redirect the reader to [24] for the Euclidean case and to [26] for the Riemannian case.

Theorem (Alexandrov-Bengert)**.** *Every semi-concave function $\varphi : U \to M$ on an open set $U \subseteq M$ admits a Hessian almost everywhere on U.*

The observation enabling us to exploit the theory of semi-concave functions is that every c-concave function is also semi-concave, which we will prove only after establishing uniform semi-concavity for the squared distance function as a special case. In order to do so, we need a *local characterization of semi-concavity*:

Proposition. *Let $\varphi : U \to \mathbb{R}$ be a continuous function and fix $x_0 \in U$. Assume that there exists a neighbourhood V of x_0 and a positive constant C such that for every $x \in V$ and $v \in T_xM$ one has*

$$\limsup_{r \to 0} \frac{\varphi(\exp_x(rv)) + \varphi(\exp_x(-rv)) - 2\varphi(x)}{r^2} \leq C$$

Then φ is semi-concave at x_0.

Proof. The function $h := d_{x_0}^2$ is smooth around x_0 and has $\operatorname{Hess}_{x_0} h = 2I$. So there exists a neighbourhood W of x_0 such that $H = \operatorname{Hess}_x h > I$ for every $x \in W$. Set $\psi := \varphi - Ch$ and take a convex neighbourhood $B \subset V \cap W$ centered at x_0. Now, every $x \in B$ and $v \in T_xM$ satisfy

$$\limsup_{r \to 0} \frac{\psi(\exp_x(rv)) + \psi(\exp_x(-rv)) - 2\psi(x)}{r^2} < 0$$

To see why this is true we can take $|v|_x = 1$, since $C' := |v|_x C$ is also a positive constant that bounds hypothesis' $\limsup$ from above, so we could work with that. Now, since $-H < -I$ the Taylor expansion of h gives us

$$\begin{aligned}
&\limsup_{r \to 0} \frac{\psi(\exp_x(rv)) + \psi(\exp_x(-rv)) - 2\psi(x)}{r^2} \leq \\
&\leq \limsup_{r \to 0} \frac{\varphi(\exp_x(rv)) + \varphi(\exp_x(-rv)) - 2\varphi(x)}{r^2} - Cg_x(Hv,\, v) \leq \\
&\leq C - Cg_x(Hv,\, v) < C - Cg_x(v,\, v) = C - C|v|_x = 0
\end{aligned}$$

Now let $\gamma : [0,\, 1] \to M$ be a geodesic contained in B and set $f(t) := \psi(\gamma(t))$. The function $f : [0,\, 1] \to \mathbb{R}$ is continuous. Applying the previous inequality to $x = \gamma(t)$ with $v = \dot\gamma(t)$ we get

$$\limsup_{r \to 0} \frac{f(t+r) + f(t-r) - 2f(t)}{r^2} < 0$$

for every $t \in (0,\, 1)$, which implies concavity for f: let $t_0,\, t_1 \in [0,\, 1]$ and $s \in (0,\, 1)$. Translating f by the affine function $g(s) = (f(t_1) - f(t_0)s + f(t_0)$ we can assume $f(t_0) = f(t_1) = 0$ without affecting the above inequality, at all. So the problem is reduced at showing that $f|_{(t_0,\, t_1)} \geq 0$. To the contrary, suppose that f has a negative minimum $f(t) > 0$ at $t \in (t_0,\, t_1)$, so for small enough r one has

$$2f(t) \leq f(t+r) + f(t-r) \Rightarrow \limsup_{r \to 0} \frac{f(t+r) + f(t-r) - 2f(t)}{r^2} \geq 0$$

which is a contradiction. Since $t_0,\, t_1 \in [0,\, 1]$ were arbitrary, f is concave on $[0,\, 1]$. Therefore ψ is geodesically concave on B, which tells us that φ is semi-concave at x_0. $\square$

Taking advantage of this charaterization we can prove:

Corollary. *For every $x \in X$, $y \in Y$, where $X,\, Y \subset M$ are compact sets, d_y^2 is semi-concave at x.*

Proof. Since M is complete, every point $x \in X$ is linked to every point $y \in Y$ be a minimizing geodesic. Since X and Y are compact, the union of all such minimizing segments is a closed and bounded, therefore compact. Thus, one can find a uniform lower bound $-k < 0$ for sectional curvatures on this set.

As before we can take an arbitrary unit vector $v \in T_xM$. Now, let $x \in X$, $y \in Y$ linked by a minimizing geodesic γ, parametrized by arc-length, with $|\dot\gamma(t)| = 1$ for every $t \in [0, l]$. Denoting $l := d(x, y)$ and letting $v(t)$ to be the parallel transport of v along γ we introduce the vector field along γ:

$$X(t) := a(t)v(t), \quad a(t) := \frac{\sinh(\sqrt{k}(l-t))}{\sinh(\sqrt{k}l)}$$

that satisfies $X(0) = v$ and $X(l) = 0$. Take a variation of the geodesic γ:

$$\gamma_r(t) := \exp_{\gamma(t)}(rX(t))$$

that satisfies $\gamma_r(0) = \exp_x(rv)$ and $\gamma_r(l) = y$. Now Hölder's inequality gives us:

$$d_y^2(\exp_x(rv)) \leq \left(\int_0^l |\partial_t \gamma_r(t)|_{\gamma(t)} dt \right)^2 \leq l \int_0^l |\partial_t \gamma_r(t)|^2_{\gamma(t)} dt =: lE(\gamma_r)$$

where E is the *energy functional* of the variation. For $r = 0$ we have equality $d_y^2(x) = E(\gamma_0)$, since $\gamma = \gamma_0$ is a constant-speed minimizing geodesic, thus:

$$\frac{d_y^2(\exp_x(rv)) + d_y^2(\exp_x(-rv)) - 2d_y^2(x)}{r^2} \leq l\frac{E(\gamma_r) + E(\gamma_{-r}) - 2E(\gamma_0)}{r^2}$$

Since $r \mapsto \gamma_r(t)$ is a geodesic for each t, the second variation formula gives us:

$$\begin{aligned}
\frac{1}{2}\frac{d^2E(\gamma_r)}{dr^2}\bigg|_{r=0} &= \int_0^l |X'(t)|^2 - R(X(t), \dot{\gamma(t)}, X(t), \dot\gamma(t))dt \\
&= \int_0^l (a'(t))^2 - (a(t))^2 R(P^\gamma_{x,v}(t), \dot{\gamma(t)}, P^\gamma_{x,v}(t), \dot\gamma(t))dt \\
&\leq \int_0^l (a'(t))^2 + k(a(t))^2 dt \\
&= \sinh^{-2}(\sqrt{k}l) k \int_0^l \cosh(2\sqrt{k}(t-l)))dt \\
&= \sinh^{-2}(\sqrt{k}l) \frac{\sqrt{k}}{2} \sinh(2\sqrt{k}l) \\
&= \frac{\sqrt{k}\cosh(\sqrt{k}l)}{\sinh(\sqrt{k}l)} = \sqrt{k}\coth(\sqrt{k}l)
\end{aligned}$$

after some calculations with our usual trigonometric identities, where the inequality came from the bound on sectional curvature. Remember that $l = d(x, y)$ and by denoting $h(t) := 2\sqrt{k}t \coth(\sqrt{k}t)$ we have proven that:

$$\limsup_{r\to 0} \frac{d_y^2(\exp_x(rv)) + d_y^2(\exp_x(-rv)) - 2d_y^2(x)}{r^2} \leq h(d(x, y)) \leq C$$

for some positive constant C which exists as an upper bound of $h \circ d(x, y)$ since it is a continuous function on a compact set and $d \geq 0$. Now, the previous proposition gives us what we want: d_y^2 is semi-concave at x. $\square$

Now, we are in a position to prove that c-concavity implies semi-concavity:

Theorem. *Let $X \subset\subset M$ be open and $Y \subset M$ be compact. Every c-concave function $\varphi \in \mathcal{I}^c(\overline{X}, Y)$ is semi-concave on X. Hence, it admits a Hessian almost everywhere on X.*

Proof. Fix $x \in X$. The continuity of φ^c on the compact $Y \subseteq M$ provides us with a $y \in \partial^c\varphi(x)$, such that

$$\varphi(\exp_x v) \leq \varphi(x) + d_y^2(\exp_x v)/2 - d_y^2(x)/2$$

for every small $v \in T_xM$ which implies

$$\limsup_{r\to 0} \frac{\varphi(\exp_x(rv)) + \varphi(\exp_x(-rv)) - 2\varphi(x)}{r^2} \leq$$
$$\leq \frac{1}{2} \limsup_{r\to 0} \frac{d_y^2(\exp_x(rv)) + d_y^2(\exp_x(-rv)) - 2d_y^2(x)}{r^2}$$

Uniform semi-concavity of the squared distance function together with the characterization of semi-concavity yields the result. Then the Alexandrov-Bangert theorem provides φ with a Hessian almost everywhere on X. $\square$

3.7 Differentiating the Optimal Transport Map

Let's turn our focus again on the c-superdifferential $\partial^c\varphi \subseteq X \times Y$ of a $\varphi \in \mathcal{I}^c(X, Y)$, which provides a multivalued extension of $F(x) = \exp_x(-\nabla\varphi(x))$ to points $x \in X$ where φ is not differentiable. The next proposition uses this to define a differential dF_x for such optimal maps F. From the chain rule for smooth fucntions, it is much expected that this differential should involve the derivative of the exponential map and the Hessian of φ.

Proposition. *Fix $X \subset\subset M$ open and $Y \subseteq M$ compact. Let $\varphi \in \mathcal{I}^c(\overline{X}, Y)$ and set $F(z) := \exp_z(-\nabla\varphi(z))$. Fix a point $x \in X$ where φ admits a Hessian and set $Y := (D\exp_x)_{-\nabla\varphi(x)}$ and $H := \operatorname{Hess}_x d_y^2/2$. Also, define $dF_x : T_xM \to T_yM$ by $dF_x := Y(H - \operatorname{Hess}_x \varphi)$. Then:*

(i) $y := F(x) \notin \operatorname{cut}(x)$ and $H - \operatorname{Hess}_x \varphi \geq 0$

(ii) As $u \to 0$ in T_xM: $\displaystyle\sup_{\substack{\exp_y v \in \partial^c\varphi(\exp_x u)\\ |v|_y = d(y, \exp_y v)}} |v - dF_x(u)|_y = o(|u|_x)$

Proof. (i) Since φ admits a Hessian at $x \in X$, it is differentiable there and $\partial^c\varphi(x) = \{F(x)\} = \{y\}$. Thus, for every $z \in X$ one has:

$$\varphi(z) \leq \varphi(x) + d_y^2(z)/2 - d_y^2(x)/2$$

which, for $z = \exp_x(\pm u)$, gives:

$$\frac{\varphi(\exp_x(rv)) + \varphi(\exp_x(-rv)) - 2\varphi(x)}{r^2} \leq \frac{d_y^2(\exp_x(rv))/2 + d_y^2(\exp_x(-rv))/2 - 2(d_y^2(x)/2)}{r^2}$$

which provides a lower bound of the right hand side, as $|u|_x \to 0$, since the left hand side tends to $g_x(\operatorname{Hess}_x \varphi(u), u)$. We have shown that if $x \in \operatorname{cut}(y)$ the right hand side is unbouded from below, so it must be the case that $x \notin \operatorname{cut}(y)$, or equivalently $y \notin \operatorname{cut}(x)$.

For the positive definiteness, set

$$h(z) := d_y^2(z)/2 - \varphi(z)$$

and observe that $\nabla h(x) = 0$ at $x \notin \operatorname{cut}(y)$ and that it has a minimum at $z = x$:

$$\varphi(z) \leq \varphi(x) + d_y^2(z)/2 - d_y^2(x)/2$$
$$\Leftrightarrow d_y^2(x)/2 - \varphi(x) \leq d_y^2(z)/2 - \varphi(z)$$
$$\Leftrightarrow h(x) \leq h(z)$$

Since $x \notin \operatorname{cut}(y)$ the Taylor expansion at x gives

$$g_x(\operatorname{Hess}_x h(u), u) \geq 0$$

hence $H - \operatorname{Hess}_x \varphi \geq 0$.

(ii) Fix a unit tangent vector $u \in T_xM$ and set $x_s := \exp_x(su)$, which for small s doesn't belong to $\mathrm{cut}(y)$. Since φ^c is continuous and X is compact, there exists a $y_s \in \partial^c\varphi(x_s)$. Let $u_s \in T_{x_s}M$ such that $y_s = \exp_{x_s} u_s$, with $u_s = d(x_s, y_s)$ and set $w_s := u_s + \nabla d_y^2(x_s)/2$ so that

$$y_s = \exp_{x_s}(-\nabla d_y^2(x_s)/2 + w_s)$$

From the relationship between c-superdifferential and superdifferential we get that $-u_s \in \overline{\partial}\varphi(x_s)$, since $y_s \in \partial^c\varphi(x_s)$. This means that $-w_s \in \overline{\partial}(-h)(x_s)$, where $h = d_y^2/2 - \varphi$ as before, which implies

$$P_{x,su}w_s = s\,\mathrm{Hess}_x\, h(u) + o(s)$$

by Hessian's definition, since $P_{x,su}$ and Hess_x are linear and $\nabla h(x) = 0$. We define three geodesic variations:

$$\begin{aligned}
f_1(t,\, s) &:= \exp_{x_s}\left(t(-\nabla d_y^2(x_s)/2)\right)\\
f_2(t,\, s) &:= \exp_{x_s}\left(t(-\nabla d_y^2(x_s)/2 + w_s)\right)\\
g(t,\, s) &:= \exp_{x}\left(t(-\nabla d_y^2(x)/2 + P_{x,su}w_s)\right)
\end{aligned}$$

and let

$$\begin{aligned}
J_1(t) &:= \frac{\partial f_1}{\partial s}(t,\, 0), \quad J_2(t) := \frac{\partial f_2}{\partial s}(t,\, 0)\\
V(t) &:= \frac{\partial g}{\partial s}(t,\, 0) = (D\exp_x)_{t(-\nabla d_y^2(x)/2)}\left(t\frac{d}{ds}\Big|_{s=0} P_{x,su}w_s\right)\\
&= t(D\exp_x)_{t(-\nabla d_y^2(x)/2)}(\mathrm{Hess}_x\, h(u))
\end{aligned}$$

be their respective variational fields. The fields V and $J := J_2 - J_1$ are Jacobi fields along the geodesic joining x to y, since J_1 and J_2 are such Jacobi fields and the Jacobi equation is linear. For their respective initial values we have that $J(0) = \frac{dx_s}{ds}|_0 - \frac{dx_s}{ds}|_0 = 0$ and obviously $V(0) = 0$, since g doesn't move the initial point. Also, $V'(0) = \mathrm{Hess}_x\, h(u)$, while

$$\begin{aligned}
J'(0) &= \frac{D}{ds}\Big|_{s=0}\left(\frac{\partial f_2}{\partial t} - \frac{\partial f_1}{\partial t}\right)(0,\, s)\\
&= \frac{D}{ds}\Big|_{s=0}(D\exp_{x_s})_0(-\nabla d_y^2(x_s)/2 + w_s + \nabla d_y^2(x_s)/2)\\
&= \frac{D}{ds}\Big|_{s=0} w_s
\end{aligned}$$

where $\frac{D}{ds}$ denotes the covariant differentiation along a curve. Let $e_i(0)$, $i = 1, \ldots, n$ be an orthonormal basis for T_xM and let $e_i(s)$, $i = 1, \ldots, n$ be its parallel transport to $T_{x_s}M$, hence $\frac{D}{ds}e_i = 0$. Then we can write $w_s = w^i(s)e_i(s)$ and $P_{x,su}w_s = w^i(s)e_i(0)$ for some smooth functions w^i. Computing gives:

$$\begin{aligned} J'(0) = \left.\frac{D}{ds}\right|_{s=0} w_s = \left.\frac{D}{ds}\right|_{s=0} (w^i(s)e_i(s)) &= \frac{dw^i}{ds}(0)e_i(0) \\ &= \left.\frac{d}{ds}\right|_{s=0} (w^i(s)e_i(0)) = \left.\frac{d}{ds}\right|_{s=0} P_{x,su}w_s \\ &= \operatorname{Hess}_x h(u) \end{aligned}$$

Thus $V \equiv J$. In particular $V(1) = J(1)$ which means

$$\left.\frac{d}{ds}\right|_{s=0} y_s = (D\exp_x)_{-\nabla d_y^2(x)/2}(\operatorname{Hess}_x h(u))$$

since $f_1(1,\, s)$ is constant and equal to y ($x_s \notin \operatorname{cut}(y)$). Now, since $\nabla h(x) = 0$ we, eventually, get

$$\left.\frac{d}{ds}\right|_{s=0} y_s = (D\exp_x)_{-\nabla\varphi(x)}(\operatorname{Hess}_x h(u)) = Y(H - \operatorname{Hess}_x \varphi)(u) = dF_x(u)$$

and by computing the Taylor expansion of $\exp_y^{-1}(y_s)$ at $s = 0$ we have:

$$y_s = \exp_y\left(s\left.\frac{d}{ds}\right|_{s=0} y_s + o(s)\right) = \exp_y(s dF_x(u) + o(s))$$

where the error term is independent of u as we have already shown. Therefore, if $v_s \in T_yM$ is the shortest vector such that $y_s = \exp_y(v_s)$ we must have

$$|v_s - dF_x(su)|_y = o(s)$$

since $\exp_y$ is a local diffeomorphism and the proof is complete. $\square$

Having established an almost everywhere notion for the differential dF_x of an optimal map $F : M \to M$, we can deduce some useful properties that we expect a differential to behold. For this purpose, let $\mu,\, \nu << \mathrm{vol_g}$ be two compactly supported measures on $X,\, Y \subset\subset M$, respectively, and denote their densities by f and g, respectively. Fix $\varphi \in \mathcal{I}^c(\overline{X},\, \overline{Y})$ that induces $F : \overline{X} \to \overline{Y}$ defined by $F(x) = \exp_x(-\nabla\varphi(x))$ which pushes μ forward to ν. Recall that, since $\varphi^c \in \mathcal{I}^c(\overline{Y},\, \overline{X})$ and ν is absolutely continuous, we can define an almost everywhere inverse of the optimal map F, that pushes ν forward to μ, by $\tilde{F}(y) = \exp_y(-\nabla\varphi^c(y))$. Introduce the sets of full measure

$$\begin{aligned} E_\varphi &:= \{x \in X \mid \operatorname{Hess}_x \varphi \text{ exists}\} \\ E_{\varphi^c} &:= \{y \in Y \mid \operatorname{Hess}_y \varphi^c \text{ exists}\} \end{aligned}$$

and note that the maps F and $\tilde{F}$ are well-defined on them, respectively. For $x \in E_\varphi$ we set, as before, $Y_x = (D\exp_x)_{-\nabla\varphi(x)}$ and $H_x = \operatorname{Hess}_x d^2_{F(x)}$ and $dF_x = Y_x(H_x - \operatorname{Hess}_x \varphi)$. Similarly, for $y \in E_{\varphi^c}$, define $Y_y = (D\exp_y)_{-\nabla\varphi^c(y)}$ and $H_y = \operatorname{Hess}_y d^2_{\tilde{F}(y)}$ and $d\tilde{F}_y = Y_y(H_y - \operatorname{Hess}_y \varphi^c)$. The first part of the previous proposition states that $F(x) \notin \operatorname{cut}(x)$ and $H_x - \operatorname{Hess}_x \varphi \geq 0$ and we can say the same thing for $\tilde{F}$ and y in the place of F and x, respectively. Now, define the following sets:

$$\tilde{E}_\varphi := \{x \in E_\varphi \mid H_x - \operatorname{Hess}_x \varphi > 0\}$$
$$\tilde{E}_{\varphi^c} := \{y \in E_{\varphi^c} \mid H_y - \operatorname{Hess}_y \varphi^c > 0\}$$

and

$$\Omega := \{x \in \tilde{E}_\varphi \mid F(x) \in \tilde{E}_{\varphi^c}\}$$

For $x \in \tilde{E}_\varphi$, the linear map $dF_x : T_xM \to T_{F(x)}M$ is a bijection, since Y is injective when $\exp_x(-\nabla\varphi(x)) \notin \operatorname{cut}(x)$. The same is true for $y \in \tilde{E}_\varphi$ and the map $d\tilde{F}_y$. We are in a position to derive an *inverse function theorem for optimal maps*:

Theorem. *Let $x \in E_\varphi$ such that $F(x) \in E_{\varphi^c}$. Then*

(a) $(dF_x)^{-1} = d\tilde{F}_{F(x)}$

(b) $x \in \Omega$

(c) $\det dF_x > 0$

(d) $\mu(\tilde{E}_\varphi) = \mu(\Omega) = 1$

Proof. Fix $u \in T_xM$. Let $x_s = \exp_x(su)$ for small s and $y_s \in M$ such that $y_s \in \partial^c\varphi(x_s)$. If v_s is the smallest vector of T_yM such tht $y_s = \exp_y v_s$, the second part of the previous proposition gives $v_s = s dF_x(u) + o(s)$. Applying the same argument for φ^c at $F(x)$ we get $su = d\tilde{F}_{F(x)}v_s + o(s) = s d\tilde{F}_{F(x)} dF_x u + o(s)$ which gives

$$d\tilde{F}_{F(x)} dF_x u = u + \frac{o(s)}{s}$$

for small s. Taking $s \to 0$ shows that $d\tilde{F}_{F(x)}$ is a left inverse to dF_x and, similarly, one can show that it is also a right inverse. Thus $(dF_x)^{-1} = d\tilde{F}_{F(x)}$ and $x \in \tilde{E}_\varphi$, $F(x) \in \tilde{E}_{\varphi^c}$, so that $x \in \Omega$. Also, since dF_x is invertible, we also have that $\det dF_x > 0$, since Y has positive determinant (it varies continuously; at zero is 1) and $H - \mathrm{Hess}_x\, \varphi > 0$. We can now write

$$\Omega = \{x \in E_\varphi \mid F(x) \in E_{\varphi^c}\}$$

and so

$$\mu(\Omega) = \mu\big(E_\varphi \cap F^{-1}(E_{\varphi^c})\big) = 1$$

since $\mu(F^{-1}(E_{\varphi^c})) = \nu(E_{\varphi^c}) = 1 = \mu(E_\varphi)$ which results from the absolute continuity of the measures. $\square$

A second property would be *the equivalence of algebraic and geometric Jacobians*:

Theorem. *Let $x \in \Omega$. Then $\partial^c\varphi(B_r(x))$ shrinks nicely to $F(x)$ when $r \to 0$ and*

$$\lim_{r\to 0} \frac{\mathrm{vol_g}[\partial^c\varphi(B_r(x))]}{\mathrm{vol_g}[B_r(x)]} = \det dF_x$$

Here, shrinks nicely to y *means that there exist $R(r) \xrightarrow{r\to 0} 0$ and $\alpha \in \mathbb{R}$ independent of r such that (i) $\partial^c\varphi(B_r(x)) \subseteq B_{R(r)}(y)$ and (ii)* $\mathrm{vol_g}[\partial^c\varphi(B_r(x))] > \alpha\, \mathrm{vol_g}[B_{R(r)}(y)]$.

Proof. Fix $x \in \Omega$ and set $y = F(x)$. For $z \in M$ let $B_r^z(0)$ denote the ball of radius r centered at the origin $0 \in T_zM$ and set $c_1 := ||dF_x||$ and $c_2 := ||dF_x^{-1}||$ where $||\cdot||$ is the operator norm. We shall prove that for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $r < \delta$ one has

$$\exp_y\big((1+\varepsilon c_1)^{-1} dF_x B_r^x(0)\big) \subseteq \partial^c\varphi(B_r(x)) \subseteq \exp_y\big((1+\varepsilon c_2) dF_x B_r^x(0)\big)$$

thus, for

$$S(r) := (c_2(1+\varepsilon c_1))^{-1} r \ \big(\xrightarrow{r\to 0} 0\big)$$

and

$$R(r) := c_1(1+\varepsilon c_2) r \ \big(\xrightarrow{r\to 0} 0\big)$$

one has, for small enough r, in normal coordinates:

$$\begin{aligned}
B_{S(r)}(y) &= \exp_y\big(B^y_{S(r)}(0)\big)\\
&\subseteq \exp_y\big((1+\varepsilon c_1)^{-1} dF_x B_r^x(0)\big)\\
&\subseteq \partial^c\varphi(B_r(x))\\
&\subseteq \exp_y\big((1+\varepsilon c_2) dF_x B_r^x(0)\big)\\
&\subseteq \exp_y\big(B^y_{R(r)}(0)\big) = B_{R(r)}(y)
\end{aligned}$$

where the first inclusion is due to the fact that for $w \in B^y_{S(r)}(0)$ one has

$$|dF_x^{-1}((1+\varepsilon c_1)w)|_x \leq ||dF_x^{-1}||(1+\varepsilon c_1)|w|_y < c_2(1+\varepsilon c_1)S(r) = r$$

while the last inclusion is due to the fact that for $v \in B^x_r(0)$ one has

$$|dF_x v|_y \leq ||dF_x|| \cdot |v|_x < c_1 r$$

Observe that

$$B_{S(r)}(y) \subseteq \partial^c\varphi(B_r(x)) \subseteq B_{R(r)}(y)$$

is exactly what we need in order to show that $\partial^c\varphi(B_r(x))$ shrinks nicely to y. In particular, the rightmost inclusion indicates that we're halfway there. For the other half, the leftmost inclusion gives us:

$$\begin{aligned}\lim_{r\to 0}\frac{\mathrm{vol_g}[\partial^c\varphi(B_r(x))]}{\mathrm{vol_g}[B_{R(r)}(y)]} &\geq \lim_{r\to 0}\frac{\mathrm{vol_g}[B_{S(r)}(y)]}{\mathrm{vol_g}[B_{R(r)}(y)]} = \left(\frac{S(r)}{R(r)}\right)^n \\ &= \frac{1}{(1+\varepsilon c_1)^n(1+\varepsilon c_2)^n} \xrightarrow{\varepsilon\to 0} 1\end{aligned}$$

which means that for every $\alpha \in (0,\, 1)$ one has

$$\mathrm{vol_g}[\partial^c\varphi(B_r(x))] > \alpha\,\mathrm{vol_g}[B_{R(r)}(y)]$$

thus $\partial^c\varphi(B_r(x))$ shrinks nicely to $y = F(x)$. Moreover, the inclusions to be proven allow us to calculate the Jacobian of F, since letting $\varepsilon \to 0$, in normal coordinates, one gets

$$\begin{aligned}\lim_{r\to 0}\frac{\mathrm{vol_g}\left(\partial^c\varphi(B_r(x))\right)}{\mathrm{vol_g}(B_r(x))} &= \lim_{r\to 0}\frac{\mathrm{Vol}_{\mathbb{R}^n}[\exp_y^{-1}(\partial^c\varphi(B_r(x)))]}{\mathrm{Vol}_{\mathbb{R}^n}(B^x_r(0))} \\ &= \lim_{r\to 0}\frac{\mathrm{Vol}_{\mathbb{R}^n}(dF_x B^x_r(0))}{\mathrm{Vol}_{\mathbb{R}^n}(B^x_r(0))} \\ &= \det dF_x\end{aligned}$$

since $\det dF_x > 0$ as we've shown at the previous theorem. Thus, it only remains to prove the inclusions.

For a given $\varepsilon > 0$, there exists $\delta > 0$ such that for every $u \in T_xM$ with $|u|_x < \delta$, if $v \in T_yM$ such that $|v|_y = d_y(\exp_y v)$ and $\exp_y v \in \partial^c\varphi(\exp_x u)$, then

$$|v - dF_x(u)|_y \leq \varepsilon|u|_x$$

. So, for $r < \delta$, pick $u \in B_r^x(0)$ and, if v is as above, then

$$\begin{aligned} |v - dF_x(u)|_y \leq \varepsilon |u|_x < \varepsilon r &\Rightarrow v - dF_x(u) \in \varepsilon B_r^y(0) \\ &\Rightarrow v \in dF_x(u) + \varepsilon B_r^y(0) \\ &\Rightarrow v \in dF_x(u + \varepsilon dF_x^{-1} B_r^y(0)) \\ &\Rightarrow v \in (1 + \varepsilon c_2) dF_x B_r^x(0) \end{aligned}$$

where the last conclusion came from the fact that for $w \in B_r^y(0)$ one has

$$|u + \varepsilon dF_x^{-1} w|_x \leq |u|_x + \varepsilon ||dF_x^{-1}|| \cdot |w|_y \leq (1 + \varepsilon c_2) r$$

and that dF_x is linear. Now, since $u \in B_r^x(0)$ and $\exp_y v \in \partial^c \varphi(\exp_x u)$ were arbitrary, exponentiation yields the second desired inclusion:

$$\partial^c \varphi(B_r(x)) \subseteq \exp_y \big((1 + \varepsilon c_2) dF_x B_r^x(0) \big)$$

On the other hand, taking $u \in dF_x^{-1} B_r^y(0)$ and r small enough such that $|u|_x < c_2 r < \delta$ then the shortest vector $v \in T_y M$: $\exp_y v \in \partial^c \varphi(\exp_x u)$ satisfies

$$\begin{aligned} |v - dF_x(u)|_y \leq \varepsilon |u|_x < \varepsilon c_2 r &\Rightarrow v - dF_x(u) \in \varepsilon c_2 B_r^y(0) \\ &\Rightarrow v \in dF_x(u) + \varepsilon c_2 B_r^y(0) \\ &\Rightarrow v \in (1 + \varepsilon c_2) B_r^y(0) \end{aligned}$$

since $dF_x(u) \in B_r^y(0)$. Again, exponentiation yields:

$$\partial^c \varphi \big(\exp_x \big(dF_x^{-1} B_r^y(0) \big) \big) \subseteq \exp_y \big((1 + \varepsilon c_2) B_r^y(0) \big)$$

while, applying the same argument to φ^c at $y = F(x)$, yields, for small enough r:

$$\partial^c \varphi^c \big(\exp_y \big(dF_x B_r^x(0) \big) \big) \subseteq \exp_x \big((1 + \varepsilon c_1) B_r^x(0) \big)$$

since $(d\tilde{F}_{F(x)})^{-1} = dF_x$ as we've shown at the previous theorem. Replacing r with $(1 + \varepsilon c_1) r$ we get

$$\partial^c \varphi^c \big(\exp_y \big((1 + \varepsilon c_1)^{-1} dF_x B_r^x(0) \big) \big) \subseteq \exp_x \big(B_r^x(0) \big) = B_r(x)$$

and, since $A \subseteq \partial^c \varphi(\partial^c \varphi^c(A))$, we get:

$$\exp_y \big((1 + \varepsilon c_1)^{-1} dF_x B_r^x(0) \big) \subseteq \partial^c \varphi(B_r(x))$$

which is the first desired inclusion. $\square$

A third property would be the *Jacobian identity*:

Theorem. *There exists a Borel set $K \subseteq X$ of full μ-measure such that, at each $x \in K$, φ admits a Hessian, thus $F(x) \notin \mathrm{cut}(x)$, and*

$$f(x) = g(F(x)) \det dF_x \neq 0$$

Proof. We set

$$K := \Omega \cap \{f \neq 0\} \cap \mathrm{Leb}(f) \cap F^{-1}(\mathrm{Leb}(g))$$

where $\mathrm{Leb}(h)$ denotes the set of Lebesgue points of the function h. The set Ω was put in the definition of K to ensure that a Hessian of φ exists. The other three sets are needed for proviing the identity. However, every single one of them has μ-full measure. We have already checked that $\mu(\Omega) = 1$ and since f, $g \in L^1(M, \mathrm{vol_g})$ are density functions they are non-zero $\mathrm{vol_g}$-a.e. and have Lebesgue points $\mathrm{vol_g}$-a.e.. Now, absolute continuity of measures and that fact that $\mu(F^{-1}(\mathrm{Leb}(g))) = \nu(\mathrm{Leb}(g))$ gives us $\mu(K) = 1$.

In order to prove the Jacobian identity, fix $x \in K$. Since x is a Lebesgue point of f, $F(x)$ is a Lebesgue point of g and $\partial^c\varphi(B_r(x))$ shrinks nicely to $F(x)$ we have that

$$f(x) = \lim_{r\to 0} \frac{\mu(B_r(x))}{\mathrm{vol_g}(B_r(x))} \quad \text{and} \quad g(F(x)) = \lim_{r\to 0} \frac{\nu(\partial^c\varphi(B_r(x)))}{\mathrm{vol_g}(\partial^c\varphi(B_r(x)))}$$

invoking the Lebesgue differentiation theorem ([7]). Fix $z \in B_r(x)$ at which φ is differentiable. Then $F(z) \in \partial^c\varphi(z)$ and since z was (almost) arbitrary we have that $F(B_r(x)) \subseteq \partial^c\varphi(B_r(x))$ and their difference consists of points on which φ is not differentiable. Since φ is differentiable $\mathrm{vol_g}$-a.e. the absolute continuity of ν gives

$$\nu(\partial^c\varphi(B_r(x))) = \nu(F(B_r(x))) = \nu(\tilde{F}^{-1}(B_r(x)))) = \mu(B_r(x))$$

where $\tilde{F}$ is the almost everywhere inverse of F, that pushes ν forward to μ. Thus

$$g(F(x)) = \lim_{r\to 0} \frac{\mu(B_r(x))}{\mathrm{vol_g}(\partial^c\varphi(B_r(x)))}$$

and by taking into account the previous theorem we get:

$$\begin{aligned} f(x) &= \lim_{r\to 0} \frac{\mu(B_r(x))}{\mathrm{vol_g}(B_r(x))} \\ &= \lim_{r\to 0} \frac{\mu(B_r(x))}{\mathrm{vol_g}(\partial^c\varphi(B_r(x)))} \frac{\mathrm{vol_g}(\partial^c\varphi(B_r(x)))}{\mathrm{vol_g}(B_r(x))} \\ &= g(F(x)) \det dF_x \end{aligned}$$

and the identity is proved. $\square$

From now on, we will be denoting by $\mathrm{J}(x)$ the quantity $\det dF_x$ and we will be calling it **the Jacobian of F at x**. As a byproduct of the previous theorem we derive the fourth and last property, a *change of variables formula*:

Corollary. *If $U : [0, +\infty) \to \mathbb{R} \cup \{+\infty\}$ is a Borel map with $U(0) = 0$, then*

$$\int_M U(g(y))d\mathrm{vol_g}(y) = \int_K U\left(\frac{f(x)}{\mathrm{J}(x)}\right)\mathrm{J}(x)d\mathrm{vol_g}(x)$$

where either both integrals are undefined or both take the same value in $\overline{\mathbb{R}}$.

Proof. Since $U(0) = 0$ and $F_{\#}\mu = \nu << \mathrm{vol_g}$ we have that:

$$\begin{aligned}\int_M U(g(y))d\mathrm{vol_g} &= \int_{\{g>0\}} \frac{U(g(y))}{g(y)}d\nu(y) \\ &= \int_{F^{-1}(\{g>0\})} \frac{U(g(F(x)))}{g(F(x))}d\mu(x) \\ &= \int_K U\left(\frac{f(x)}{\mathrm{J}(x)}\right)\frac{\mathrm{J}(x)}{f(x)}d\mu(x) \\ &= \int_K U\left(\frac{f(x)}{\mathrm{J}(x)}\right)\mathrm{J}(x)d\mathrm{vol_g}(x)\end{aligned}$$

since $K \subseteq F^{-1}(\{g > 0\})$ and $\mu(K) = 1$. □

3.8 Optimal interpolating maps

In what follows we shall investigate the family of maps

$$F_t(x) = \exp_x(-t\nabla\varphi(x))$$

which interpolate along geodesics from the identity map $x = F_0(x)$ to the optimal map $y = F_1(x) = F(x)$. Ultimately, we wish to extend the properties of F for every $t \in (0, 1)$ and this is accomplished after proving that the set of c-concave functions is star shaped in some sense. Then, McCann's theorem implies that F_t is the optimal map pushing $\mu << \mathrm{vol_g}$ forward to $\mu_t := (F_t)_{\#}\mu$. However, if we want the preceding results to hold we shall also show that $\mu_t << \mathrm{vol_g}$. We begin with the star-shapedness of c-concave functions:

Lemma. *Fix $t \in [0,\, 1]$ and compact sets $X,\, Y \subseteq M$. If $\varphi \in \mathcal{I}^c(X,\, Y)$ then $t\varphi \in \mathcal{I}^c(X,\, Z_t(X,\, Y))$.*

Proof. The lemma is trivial for $t = 0$ and $t = 1$, since $0 \in \mathcal{I}^c(X,\, X)$ and $t\varphi \in \mathcal{I}^c(X,\, Y)$. Therefore, fix $t \in (0,\, 1)$ and $y \in Y$. We'll treat a spacial case first, that of $d_y^2/2$. Let $x \in X$ and $z \in Z_t(x,\, y)$. The triangle inequality alongside the arithmetic-geometric mean inequality gives, for every $m \in M$ and any $\varepsilon > 0$:

$$\begin{aligned}
d^2(m,\, y) &\le (d(m,\, z) + d(z,\, y))^2 \\
&= d^2(m,\, z) + d^2(z,\, y) + 2d(m,\, z)d(z,\, y) \\
&= d^2(m,\, z) + d^2(z,\, y) + 2\sqrt{(\varepsilon^{-1}d^2(m,\, z))(\varepsilon d^2(z,\, y))} \\
&\le d^2(m,\, z) + d^2(z,\, y) + \varepsilon^{-1}d^2(m,\, z) + \varepsilon d^2(z,\, y) \\
&= (1+\varepsilon^{-1})d^2(m,\, z) + (1+\varepsilon)d^2(z,\, y)
\end{aligned}$$

and by choosing $\varepsilon = t/(1-t)$ we get:

$$\begin{aligned}
d^2(m,\, y) &\le \frac{1}{t}d^2(m,\, z) + \frac{1}{1-t}d^2(z,\, y) \\
&= \frac{1}{t}d^2(m,\, z) + \frac{1}{1-t}(1-t)^2 d^2(x,\, y) \\
&= \frac{1}{t}d^2(m,\, z) + (1-t)d^2(x,\, y) \\
\Rightarrow td^2(m,\, y)/2 &\le d^2(m,\, z)/2 + t(1-t)d^2(x,\, y)/2
\end{aligned}$$

since $z \in Z_t(x,\, y)$. Note that $m = x$ produces equality, since $d(x,\, z) = td(x,\, y)$. We define $\psi_y : Z_t(X,\, y) \to \mathbb{R}$ as follows: if $z \in Z_t(X,\, y)$ then there exists $x \in X$ such that $z \in Z_t(x,\, y)$, so the function

$$\psi_y(z) := -t(1-t) \inf_{\substack{x\in X:\\ z\in Z_t(x,\, y)}} \{d_y^2(x)/2\}$$

is well defined. Since the above inequality is true for every $x \in X$ we have, for every $z \in Z_t(X,\, y)$:

$$\begin{aligned}
td^2(m,\, y)/2 &\le d^2(m,\, z)/2 - \psi_y(z) \\
\Rightarrow td_y^2(m)/2 &\le c(m,\, z) - \psi_y(z) \\
\Rightarrow td_y^2(m)/2 &\le \inf_{z\in Z_t(X,\, y)} \{c(m,\, z) - \psi_y(z)\}
\end{aligned}$$

but, since equality is achieved if $m \in X$, we must have:

$$td_y^2(x)/2 = \inf_{z\in Z_t(X,\, y)} \{c(x,\, z) - \psi_y(z)\}$$

for $x \in X$, so that $td_y^2/2 \in \mathcal{I}^c(X,\, Z_t(X,\, y))$.

Now take a $\varphi \in \mathcal{I}^c(X,\, Y)$. Then $\varphi = \varphi^{cc}$ gives:

$$\begin{aligned}
\varphi(x) &= \inf_{y\in Y}\{c(x,\, y) - \varphi^c(y)\} \\
\Rightarrow\ t\varphi(x) &= \inf_{y\in Y}\{tc(x,\, y) - t\varphi^c(y)\} \\
&= \inf_{y\in Y}\ \inf_{z\in Z_t(X,\, y)}\{c(x,\, z) - \psi_y(z) - t\varphi^c(y)\} \\
&= \inf_{z\in Z_t(X,\, Y)}\left\{c(x,\, z) - \inf_{\substack{y\in Y:\\ z\in Z_t(X,\, y)}}\{\psi_y(z) + t\varphi^c(y)\}\right\}
\end{aligned}$$

and define $\zeta : Z_t(X,\, Y) \to \mathbb{R}$ as follows: for $z \in Z_t(X,\, Y)$ there exists $y \in Y$ such that $z \in Z_t(X,\, y)$. Define ψ_y as before and set:

$$\zeta(z) := \inf_{\substack{y\in Y:\\ z\in Z_t(X,\, y)}}\{\psi_y(z) + t\varphi^c(y)\}$$

which is well defined. We have found a function $\zeta : Z_t(X,\, Y) \to \mathbb{R}$ such that

$$t\varphi(x) = \inf_{z\in Z_t(X,\, Y)}\{c(x,\, z) - \zeta(z)\}$$

which means $t\varphi \in \mathcal{I}^c(X,\, Z_t(X,\, Y))$. $\square$

We establish two immediate corollaries. One is about Hessian semi-positivity relating distance functions:

Corollary. *Let $\gamma(t) = \exp_x(tv)$ be the minimal geodesic joining $x \in M$ to $\gamma(1) \notin \operatorname{cut}(x)$. The self-adjoined operator $H(t) - tH(1)$, defined on T_xM by $H(t) := \operatorname{Hess}_x d^2_{\gamma(t)}/2$, is positive semi-definite.*

Proof. Since $\gamma(t) \in Z_t(x,\, y)$ we get:

$$\begin{aligned}
&td^2(m,\, y)/2 \le d^2(m,\, \gamma(t))/2 + t(1-t)d^2(x,\, y)/2 \\
\Leftrightarrow\quad &\varphi(m) := d^2_{\gamma(t)}(m)/2 - td^2_y(m)/2 + t(1-t)d^2_y(x)/2 \ge 0
\end{aligned}$$

for every $m \in M$, with equality being achieved at $m = x$. So φ is non-negative and attains the minimum value $0 = \varphi(x)$. From the Taylor expansion of φ at $m = x$ we get that $g_x(\operatorname{Hess}_x \varphi(v),\, v) \ge 0$ for every $v \in T_xM$. $\square$

While the other one is about the optimality of the interpolant map:

Corollary. *Let $\mu << \mathrm{vol_g}$ be a Borel probability measure which is compactly supported in $X \subset\subset M$ and $Y \subseteq M$ compact. Fix $\varphi \in \mathcal{I}^c(\overline{X}, Y)$, $t \in [0, 1]$ and set*

$$F_t(x) := \exp_x(-t\nabla\varphi(x))$$

for $x \in X$. The map F_t coincides with the optimal map pushing μ forward to $\mu_t := (F_t)_\# \mu$.

Proof. We saw that $t\varphi$ is c-concave, since $t\varphi \in \mathcal{I}^c(\overline{X}, Z_t(\overline{X}, Y))$. From Rademacher's Theorem, we have that $F_t(x)$ is defined μ-almost everywhere ($\mu << \mathrm{vol_g}$) on X and $F_t(x) \in Z_t(\overline{X}, Y)$ which is a compact set, since $\overline{X}$ and Y are and distance function is continuous. The map pushes μ to μ_t by construction, so that we also have that μ_t is compactly supported in $Z_t(\overline{X}, Y)$. McCann's Theorem uniqueness gives the desired optimality among all maps pushing μ forward to μ_t. $\square$

In order to show $\mu_t << \mathrm{vol_g}$ we need the interpolant map to be injective:

Lemma. *Let $X \subset\subset M$ be open and $Y \subseteq M$ compact. Fix $\varphi \in \mathcal{I}^c(\overline{X}, Y)$ and $t \in (0, 1)$. If $F_t(x_1) = F_t(x_2)$ at two points $x_1, x_2 \in X$, at which φ is differentiable, then $x_1 = x_2$.*

Proof. If $x_1, x_2 \in X$ are two points of differentiability for φ, with $F_t(x_1) = F_t(x_2) := z$, then $z \in Z_t(x_1, y_1) \cap Z_t(x_2, y_2)$, where $y_i := F_1(x_i)$, $i \in \{1, 2\}$, for simplicity. This is true, since, for $i \in \{1, 2\}$, x_i, z, y_i all lie on the minimizing geodesic $t \mapsto F_t(x_i)$. For the distance between x_1, y_2 and x_2, y_1 the triangle inequalities gives us:

$$\begin{aligned}
d(x_1, y_2) &\leq d(x_1, z) + d(z, y_2) \\
d(x_2, y_1) &\leq d(x_2, z) + d(z, y_1)
\end{aligned}$$

which we square and sum, to get a parallelogram inequality:

$$\begin{aligned}
d^2(x_1, y_2) + d^2(x_2, y_1) &\leq d^2(x_1, z) + d^2(z, y_2) + d^2(x_2, z) + d^2(z, y_1) \\
&\quad + 2d(x_1, z)d(z, y_2) + 2d(x_2, z)d(z, y_1) \\
&= (d^2(x_1, z) + d^2(z, y_1)) + (d^2(x_2, z) + d^2(z, y_2)) \\
&\quad + 2d_{x_1}d_{y_2}(z) + 2d_{x_2}d_{y_1}(z) \\
&= (d(x_1, z) + d(z, y_1))^2 + (d(x_2, z) + d(z, y_2))^2 \\
&\quad + 2(d_{x_1}d_{y_2} + d_{x_2}d_{y_1} - d_{x_1}d_{y_1} - d_{x_2}d_{y_2})(z) \\
&= (d(x_1, y_1))^2 + (d(x_2, y_2))^2 \\
&\quad - 2(d_{x_1} - d_{x_2})(d_{y_1} - d_{y_2})(z) \\
&= d^2(x_1, y_1) + d^2(x_2, y_2) \\
&\quad - 2(d(x_1, z) - d(x_2, z))(d(y_1, z) - d(y_2, z)) \\
&= d^2(x_1, y_1) + d^2(x_2, y_2) \\
&\quad - 2t(1-t)(d(x_1, y_1) - d(x_2, y_2))^2 \\
&\leq d^2(x_1, y_1) + d^2(x_2, y_2)
\end{aligned}$$

where in the equalities we used the fact that $z \in Z_t(x_1, y_1) \cap Z_t(x_2, y_2)$.

On the other hand, since $y_1 \in \partial^c\varphi(x_1)$ and $y_2 \in \partial^c\varphi(x_2)$, we have that:

$$\begin{aligned}
\varphi(x_2) &\leq \varphi(x_1) + c(x_2, y_1) - c(x_1, y_1) \\
\varphi(x_1) &\leq \varphi(x_2) + c(x_1, y_2) - c(x_2, y_2)
\end{aligned}$$

respectively. Adding them and multiplying by 2, gives:

$$d^2(x_1, y_2) + d^2(x_2, y_1) \geq d^2(x_1, y_1) + d^2(x_2, y_2)$$

which contradicts the parallelogram inequality, unless equalities hold everywhere. The equalities concerning the triangle inequalities tell us that z separates x_1, x_2 from y_1, y_2, while all five points are lying on the same geodesic. The equality concerning the last inequality tells us that $d(x_1, y_1) = d(x_2, y_2)$, so that:

$$\begin{aligned}
d(x_1, x_2) &= d(x_1, z) - d(x_2, z) = td(x_1, y_1) - td(x_2, y_2) = 0 \\
d(y_1, y_2) &= d(z, y_1) - d(z, y_2) = (1-t)d(x_1, y_1) - (1-t)d(x_2, y_2) = 0
\end{aligned}$$

concluding that $x_1 = x_2$ and $y_1 = y_2$. □

Now, fix μ, $\nu << \mathrm{vol}_{\mathrm{g}}$ compactly supported, in fixed open sets $X, Y \subset\subset M$, and choose $\varphi \in \mathcal{I}^c(\overline{X}, \overline{Y})$ such that $F := F_1 \in \Pi(\mu, \nu)$, where $F_t(x) =$

$\exp_x(-t\nabla\varphi(x))$. Recall the sets

$$E_\varphi = \{x \in X \mid \operatorname{Hess}_x \varphi \text{ exists}\}$$

$$\tilde{E}_\varphi := \{x \in E_\varphi \mid \operatorname{Hess}_x \left(d^2_{F(x)}/2 - \varphi\right) > 0\}$$

where the latter has full measure (since μ, $\nu << \mathrm{vol_g}$) and the maps F and F_t are well defined on both of them. Also, for $x \in E_\varphi$ one has $F(x) \notin \mathrm{cut}(x)$, therefore $F_t(x) \notin \mathrm{cut}(x)$. Now, consider the following set:

$$\tilde{E}_t := \{x \in E_\varphi \mid \operatorname{Hess}_x \left(d^2_{F_t(x)} - t\varphi\right) > 0\}$$

which has, also, full measure, since $\tilde{E}_\varphi \subseteq \tilde{E}_t$. Indeed, observe that

$$\begin{aligned}\operatorname{Hess}_x \left(d^2_{F_t(x)}/2 - t\varphi\right) &= \operatorname{Hess}_x \left(d^2_{F_t(x)}/2 - td^2_{F(x)}/2 + td^2_{F(x)}/2 - t\varphi\right) \\ &= H(t) - tH(1) + t\operatorname{Hess}_x h\end{aligned}$$

where $H(t) = \operatorname{Hess}_x d^2_{F_t(x)}/2$ and since $H(t) - tH(1) \geq 0$, $\operatorname{Hess}_x h \geq 0$, we conclude that

$$\tilde{E}_\varphi \subseteq \tilde{E}_t \Rightarrow 1 = \mu(\tilde{E}_\varphi) \leq \mu(\tilde{E}_t) \Rightarrow \mu(\tilde{E}_t) = 1$$

Let $K \subseteq \tilde{E}_t$ be a compact set and denote by $K' := F_t(K)$ its image though F_t. The injectivity of interpolant tells us that the map $F_t : K \to K'$ is a bijection. The star-shapedness of c-concave functions yields a first order Taylor expansion for F_t at each $x \in K \subseteq \tilde{E}_t$, hence F_t is continuous. Therefore, $F_t : K \to K'$ is a homeomorphism as a continuous bijection from a compact to a Hausdorff space. Moreover, setting $Y(t) := (D\exp_x)_{-t\nabla\varphi(x)}$, the fact that $K \subseteq \tilde{E}_t$ tells us that the derivative

$$(dF_t)_x := Y(t)(H(t) - t\operatorname{Hess}_x \varphi)$$

has positive Jacobian $\det(dF_t)_x$, thus it's invertible, for every $x \in K$.

It will prove helpful to impose a *Lipschitz control of the inverse interpolant*:

Lemma. *For every $x \in K$, there exists a constant $k_x > 0$ such that*

$$d(F_t(x),\, F_t(z)) \geq k_x d(x,\, z)$$

for every $z \in K$.

Proof. Assume that there exists a sequence x_k such that

$$d(F_t(x),\, F_t(x_k)) < \frac{1}{k} d(x,\, x_k)$$

which we can assume, by taking a subsequence, that converges to some $z \in K$. By continuouity and injectivity of F_t we must have $z = x$ and so $x_k \to x$. Let $u_k \in T_x M$ and $w_k \in T_{F_t(x)} M$ be the smallest vectors for which $x_k = \exp_x u_k$ and $F_t(x_k) = \exp_{F_t(x)} w_k$, so that

$$d_{F_t(x)}(\exp_{F_t(x)} w_k) < \frac{1}{k} d_x(\exp_x u_k)$$

which means that $|w_k|_{F_t(x)} = o(|u_k|_x)$ (we will omit the subscripts for clarity). Now, the first order Taylor expansion asserts

$$|w_k - (dF_t)_x(u_k)| = o(|u_k|)$$

but, on the other hand, since $(dF_t)_x$ is invertible, we have that

$$||(dF_t)_x^{-1}|| \cdot |u_k| \leq |(dF_t)_x(u_k)|$$

which gives

$$0 < ||(dF_t)_x^{-1}|| \leq \frac{|w_k|}{|u_k|} + \frac{o(|u_k|)}{|u_k|} \to 0$$

which is a contradiction. $\square$

Finally, we are in a position to prove the *absolute continuity of the interpolant measure*:

Theorem. *If* $\mu,\, \nu << \mathrm{vol_g}$ *are as above, then for each* $t \in (0,\, 1)$*:*

$$\mu_t << \mathrm{vol_g}$$

where $\mu_t = (F_t)_\# \mu$.

Proof. Let $t \in (0,\, 1)$ and $A \subseteq M$ a Borel set. Invoking the regularity of measure μ we can find an increasing sequence of compact sets $K_i \subseteq \tilde{E}_t$, such that

$$\mu\left(\bigcup_{i=1}^{\infty} K_i\right) = \lim_{i \to +\infty} \mu(K_i) = \mu(\tilde{E}_t) = 1$$

so that, for any Borel set A one has

$$\mu_t(A) = \mu(F_t^{-1}(A)) = \mu\left(\bigcup_{i=1}^{\infty}(F_t^{-1}(A) \cap K_i)\right) \leq \sum_{i=1}^{\infty} \mu(F_t^{-1}(A) \cap K_i).$$

In order to show that $\mu_t << \mathrm{vol}_g$ we must consider a Borel set A satisfying $\mathrm{vol}_g(A) = 0$ and show that $\mu_t(A) = 0$, but, according to the above, we only need to show that $\mu(F_t^{-1}(A) \cap K_i) = 0$ for every i. However, we know that $\mu << \mathrm{vol}_g$, therefore it suffices to show that $\mathrm{vol}_g(F_t^{-1}(A) \cap K_i) = 0$ for every i. To sum up, we need to prove:

$$\mathrm{vol}_g(A) = 0 \Rightarrow (F_t)_\# \,\mathrm{vol}_g|_K(A) = 0$$

where A is a Borel set and $K \subseteq \tilde{E}_t$ is compact. In other words, we need to show that if $K \subseteq \tilde{E}_t$ is compact, then the restriction $(F_t)_\# \,\mathrm{vol}_g|_K$ is absolutely continuous on $K' := F_t(K)$, with respect to vol_g. The previous lemma asserts that $K = \bigcup_{k>0} K_k$ where

$$K_k := \{x \in K \mid \forall z \in K,\ d(F_t(x),\, F_t(z)) \geq \frac{1}{k} d(x,\, z)\}.$$

which are closed sets, since F_t is continuous, hence compact. But, the map $F_t : K_k \to K'_k := F_t(K_k)$ has Lipschitz continuous inverse F_t^{-1}, by the definition of K_k, hence K'_k is compact as well. Recall that the injectivity radius $\mathrm{inj} : M \to [0, \infty)$ is a continuous function ([3], Prop. 10.37). Hence, it attains a minimum on the compact $K_k \cup K'_k$, let's call it $r_0 > 0$. Now, since

$$K_k \cup K'_k \subseteq \bigcup_{z \in K_k \cup K'_k} B_{\frac{r_0}{2}}(z)$$

we can use compactness to find $m \in \mathbb{N}$ and $z_1, \ldots, z_m \in K_k \cup K'_k$ such that

$$K_k \cup K'_k \subseteq \bigcup_{i=1}^{m} B_{\frac{r_0}{2}}(z_i).$$

The continuity of the metric tensor g implies the continuity of the functions g_{ij} on every normal neighbourhood $B_{r_0}(z_i)$ (the reader must not be confused by the fact that i is used abusively as indexes of sums, unions and their terms and is different from the index of g_{ij}). Hence, the function $\sqrt{\det(g_{ij})}$ is also continuous on every normal neighbourhood $B_{r_0}(z_i)$ and consequentially on the compact sets $(K_k \cup K'_k) \cap \overline{B_{\frac{r_0}{2}}(z_i)} \subseteq B_{r_0}(z_i)$ attaining a minimum $a_i > 0$ and a maximum b_i. Hence

$$a := \max_{i=1,\ldots,m} a_i \leq \sqrt{\det(g_{ij})} \leq \min_{i=1,\ldots,m} b_i =: b$$

on all of $K_k \cup K'_k$. If we take $\lambda > 0$, $z \in K_k \cup K'_k$ and $0 < r < \frac{r_0}{\lambda}$ such that $B_r(z) \subseteq K_k \cup K'_k$ we have

$$am(B_r^z(0)) \leq \mathrm{vol_g}(B_r(z)) \leq bm(B_r^z(0))$$

so that

$$a\lambda^n m(B_r^z(0)) \leq \mathrm{vol_g}(B_{\lambda r}(z)) \leq b\lambda^n m(B_r^z(0)).$$

Combining the above we deduce, that for any $\lambda > 0$, $z \in K_k \cup K'_k$ and $0 < r < \frac{r_0}{\lambda}$ such that $B_r(z) \subseteq K_k \cup K'_k$, one has

$$\mathrm{vol_g}(B_{\lambda r}(z)) \leq C_g \lambda^n \,\mathrm{vol_g}(B_r(z))$$

where $C_g := \frac{b}{a}$. The Lipschitzian character of F_t^{-1} implies, immediately, that for every $0 < r < \frac{r_0}{k}$ and $x \in K'_k$,

$$F_t^{-1}(B_r(x)) \subseteq B_{kr}(F_t^{-1}(x))$$

so that

$$\begin{aligned} \mathrm{vol_g}(F_t^{-1}(B_r(x))) &\leq \mathrm{vol_g}(B_{kr}(F_t^{-1}(x))) \\ &\leq C_g k^n \,\mathrm{vol_g}(B_r(F_t^{-1}(x))) \\ &\leq C_g^2 k^n \,\mathrm{vol_g}(B_r(x)) \end{aligned}$$

where the last inequality is a product of the previous calculations and the fact that the Lebesgue measure is invariant under translation.

If $U \subseteq K'_k$ is open, for any $x \in U$ consider an $r_x \leq \min\{\frac{r_0}{5}, \frac{r_0}{k}\}$ such that $B_{r_x}(x) \subseteq U$. Now, we can write U as

$$U = \bigcup_{x \in U} B_{r_x}(x)$$

and the well-known Vitali's covering lemma ([25], [7], [6]) provides us with countable many points $x_i \in U$ such that

$$\bigcup_{i=1}^{\infty} B_{r_i}(x_i) \subseteq U \subseteq \bigcup_{i=1}^{\infty} B_{5r_i}(x_i)$$

where the balls are chosen from the family above and are disjoint, which, in turn, gives

$$\begin{aligned}
\mathrm{vol_g}(F_t^{-1}(U)) &\leq \sum_{i=1}^{\infty} \mathrm{vol_g}(F_t^{-1}(B_{5r_i}(x_i))) \\
&\leq C_g k^n \sum_{i=1}^{\infty} \mathrm{vol_g}(B_{5r_i}(x_i)) \\
&= C_g^2 (5k)^n \sum_{i=1}^{\infty} \mathrm{vol_g}(B_{r_i}(x_i)) \\
&= C_g^2 (5k)^n \,\mathrm{vol_g}\left(\bigcup_{i=1}^{\infty} B_{r_i}(x_i)\right) \\
&\leq C_g^2 (5k)^n \,\mathrm{vol_g}(U).
\end{aligned}$$

Now, let $\varepsilon > 0$ and $A \subseteq K'_k$ be any Borel set. The regularity of measure provides us with an open $A \subseteq U \subseteq K'_k$ such that $\mathrm{vol_g}(U) \leq \mathrm{vol_g}(A) + \varepsilon$. Combining the above we have

$$\begin{aligned}
\mathrm{vol_g}(F_t^{-1}(A)) &\leq \mathrm{vol_g}(F_t^{-1}(U)) \\
&\leq C_g^2 (5k)^n \,\mathrm{vol_g}(U) \\
&\leq C_g^2 (5k)^n (\mathrm{vol_g}(A) + \varepsilon)
\end{aligned}$$

and letting $\varepsilon \to 0^+$ we conclude

$$\mathrm{vol_g}(F_t^{-1}(A)) \leq C_{g,k,n} \,\mathrm{vol_g}(A)$$

for any Borel set $A \subseteq K'_k$, where $C_{g,k,n} > 0$ is some constant independent of the choice of A.

Now, let $A \subseteq M$ be any Borel set of zero measure, i.e. $\mathrm{vol_g}(A) = 0$. Observe that

$$\begin{aligned}
(F_t)_{\#} \,\mathrm{vol_g}|_K(A) &= \mathrm{vol_g}(F_t^{-1}(A) \cap K) \\
&= \mathrm{vol_g}\left(\bigcup_{k\in\mathbb{N}} (F_t^{-1}(A) \cap K_k)\right) \\
&\leq \sum_{k\in\mathbb{N}} \mathrm{vol_g}(F_t^{-1}(A \cap K'_k)) \\
&\leq \sum_{k\in\mathbb{N}} C_{g,n,k} \,\mathrm{vol_g}(A \cap K'_k) \\
&\leq \sum_{k\in\mathbb{N}} C_{g,n,k} \,\mathrm{vol_g}(A) = 0.
\end{aligned}$$

Hence, $(F_t)_\# \operatorname{vol_g}|_K$ is absolutely continuous with respect to $\operatorname{vol_g}$, which implies, as we argued before, that $\mu_t << \operatorname{vol_g}$ and the proof is complete. □

Since, now, μ_t is absolutely continuous and $t\varphi \in \mathcal{I}^c(\overline{X},\, Z_t(\overline{X},\, \overline{Y}))$ for any $t \in [0,\, 1]$ and $\varphi \in \mathcal{I}^c(\overline{X},\, \overline{Y})$ there exists a set of full measure for μ, $K_t \subseteq X$, such that φ has a Hessian at each $x \in K_t$, $f(x) \neq 0$ and the Jacobian identity holds:

$$f_t = f_0 \circ F_t \cdot \mathrm{J}_t$$

where $\mathrm{J}_t(x) := \det(dF_t)_x = \det Y(t)(H(t) - t \operatorname{Hess}_x \varphi)$ as above and $d\mu_t = f_t d\operatorname{vol_g}$ (note that $\mu_0 = \mu$). A change of variables formula is just an immediate...

Corollary. *If $U : [0,\, +\infty) \to \mathbb{R} \cup \{+\infty\}$ is a Borel map with $U(0) = 0$, then*

$$\int_M U(f_t(z)) d\operatorname{vol_g}(z) = \int_{K_t} U\left(\frac{f(x)}{\mathrm{J}_t(x)}\right) \mathrm{J}_t(x) d\operatorname{vol_g}(x)$$

where either both integrals are undefined or both take the same value in $\overline{\mathbb{R}}$.

3.9 Convexity of the Jacobian

Now, we can prove some sort of convexity for $\mathrm{J}_t^{1/n}$, known as *the Jacobian inequality*, which will prove very useful as it projects the fact that Ricci curvature controls the optimal mass transport.

Proposition. *Fix $X \subset\subset M$ open, $Y \subseteq M$ compact and a c-concave function $\varphi \in \mathcal{I}^c(\overline{X},\, Y)$ so that for each $t \in [0,\, 1]$ the map $F_t : X \to Z_t(\overline{X},\, Y)$ is defined by $F_t(z) = \exp_z(-t\nabla\varphi(z))$. If φ admits a Hessian at $x \in X$, then $Y(t)$, $H(t)$ exist and the Jacobian $\mathrm{J}_t(x)$ at x satisfies:*

$$\mathrm{J}_t^{1/n}(x) \geq (1-t) v_{1-t}^{1/n}(F(x),\, x) + t v_t^{1/n}(x,\, F(x))\, \mathrm{J}^{1/n}(x)$$

where $F = F_1$, $\mathrm{J} = \mathrm{J}_1$.

Proof. Let $x \in X$ be the point at which φ admits a Hessian. Then $F(x) \notin \operatorname{cut}(x)$ and since $\nabla d^2_{F(x)}(x)/2 = \nabla\varphi(x)$ we have $d(x,\, F(x)) = |\nabla\varphi(x)|_x$ so that $F_t(x) \notin \operatorname{cut}(x)$ for $t \in [0,\, 1]$. Thus $Y(t) = (D\exp_x)_{-t\nabla\varphi(x)}$ and

$H(t) = \operatorname{Hess}_x d^2_{F_t(x)}/2$ are both well defined. The inequality is trivial for $t = 1$, so fix $t \in [0, 1)$ and write

$$J_t(x) = \det Y(t) \det \Big[(1-t)\frac{H(t) - tH(1)}{1-t} + t(H(1) - \operatorname{Hess}_x \varphi)\Big].$$

The matrices $H(t) - tH(1)$ and $H(1) - \operatorname{Hess}_x \varphi$ are symmetric, while their positive definiteness has been established by the "triangle inequality" for $d^2/2$ and the c-concavity of φ, respectively. Moreover, differentiating the *Jacobi's formula for the derivative of the determinant* on the set of symmetric positive definite matrices yields

$$\operatorname{Hess}_A \det{}^{1/n}(X, X) = \frac{1}{n^2} \det{}^{1/n}(A)\big(\operatorname{tr}^2(A^{-1}X) - n \operatorname{tr}(A^{-2}X^2)\big) \le 0$$

due to the Cauchy-Schwartz inequality, after we diagonalize $A^{-1}X$, which we can do since both matrices are symmetric and one of them is positive definite. Now, the concavity of $\det^{1/n}$ yields

$$\begin{aligned} \mathrm{J}_t^{1/n}(x) &\ge (1-t)\det{}^{1/n} \frac{Y(t)(H(t) - tH(1))}{1-t} + t\det{}^{1/n} Y(t)(H(1) - \operatorname{Hess}_x \varphi) \\ &\ge (1-t)\det{}^{1/n} \frac{Y(t)(H(t) - tH(1))}{1-t} + t\det{}^{1/n} Y(t)Y(1)^{-1}\,\mathrm{J}_t^{1/n}(x) \\ &= (1-t)v_{1-t}^{1/n}(F(x), x) + tv_t^{1/n}(x, F(x))\,\mathrm{J}^{1/n}(x) \end{aligned}$$

since $t \mapsto F_t(x)$ is the minimal geodesic linking x to $F(x)$. $\square$

Chapter 4

Ricci Curvature vs. Entropy

In this last chapter we will define the entropy functional, explore it a bit and use it to characterize lower Ricci bounds.

4.1 Entropy

The **relative entropy** is defined as a functional on $\mathscr{P}^2(M)$ by

$$\mathrm{Ent}(\nu) := \begin{cases} \displaystyle\int_M \frac{d\nu}{d\mathrm{vol_g}} \log \frac{d\nu}{d\mathrm{vol_g}} d\mathrm{vol_g} & , \text{ if } \nu << \mathrm{vol_g} \text{ \& } \mathrm{Ent}_+(\nu) < \infty \\ +\infty & , \text{ otherwise} \end{cases}$$

where by $\mathrm{Ent}_+(\nu) < \infty$ we mean $\int_M \left[\frac{d\nu}{d\mathrm{vol_g}} \log \frac{d\nu}{d\mathrm{vol_g}}\right]_+ d\mathrm{vol_g} < \infty$.

We will prove the two properties of entropy we talked about in the introduction. It gives us a nice picture of how entropy behaves.

Lemma. *Let $A \subseteq M$ be a Borel set and μ, $\nu << \mathrm{vol_g}$ be probability measures such that*

$$\nu := \frac{\chi_A}{\mathrm{vol_g}(A)} \mathrm{vol_g}$$

is the normalized uniform distribution on A and $\mathrm{supp}(\mu) \subseteq A$. Then

(i) $\mathrm{Ent}(\nu) = -\log \mathrm{vol_g}(A)$

(ii) $\mathrm{Ent}(\mu) \geq \mathrm{Ent}(\nu)$

Proof. (i) A simple calculation yields:

$$\begin{aligned}\mathrm{Ent}(\nu) &= \int_M \frac{\chi_A}{\mathrm{vol_g}(A)} \log \frac{\chi_A}{\mathrm{vol_g}(A)} d\mathrm{vol_g} = \\ &= \frac{1}{\mathrm{vol_g}(A)} \int_A 0 - \log \mathrm{vol_g}(A) d\mathrm{vol_g} = \\ &= -\log \mathrm{vol_g}(A).\end{aligned}$$

(ii) Since $\log t \leq t - 1$, for $t > 0$, plugging in $t = \frac{x}{y}$, for x, $y > 0$, yields

$$y - y \log y \leq x - y \log x.$$

Since $\frac{d\mu}{d\mathrm{vol_g}}, \frac{d\nu}{d\mathrm{vol_g}}$ are probability densities, putting them in the place of y and x, respectively, and integrating on A gives:

$$\begin{aligned}\mathrm{Ent}(\mu) &\geq \int_A \frac{d\mu}{d\mathrm{vol_g}} \log \frac{\chi_A}{\mathrm{vol_g}(A)} d\mathrm{vol_g} \\ &= -\log \mathrm{vol_g}(A) \int_A \frac{d\mu}{d\mathrm{vol_g}} d\mathrm{vol_g} \\ &= \mathrm{Ent}(\nu) \cdot \mu(A) \\ &= \mathrm{Ent}(\nu).\end{aligned}$$

□

4.2 Main Theorem

In this last section we will gather all our knowledge and tools developed in the previous sections and finally prove an equivalence of lower Ricci bounds with entropy's K-convexity. We urge the reader to read the introduction (or at least the last parts of it) to recall the definition of K-convexity and some conventions made about it. Before we state and prove the main theorem we will stress the fact that the functions sn_K, defined as

$$\mathrm{sn}_K(r) = \begin{cases} \dfrac{\sin(\sqrt{K}r)}{\sqrt{K}r} & , \ K > 0 \quad , \ 0 < r < \frac{\pi}{\sqrt{K}} \\ 1 & , \ K = 0 \quad , \ r > 0 \\ \dfrac{\sinh\left(\sqrt{-K}r\right)}{\sqrt{-K}r} & , \ K < 0 \quad , \ r > 0 \end{cases}$$

with
$$\mathrm{sn}_K(0) = 1$$
are playing an important role in our transition from the infinitesimal BMI, the Jacobian inequality, to the K-convexity of entropy, by proving convexity for a function related to sn_K, that will come up later:

Lemma. *Let $r \in \left(0, \frac{\pi}{\sqrt{K}}\right)$ if $K > 0$ and $r > 0$ otherwise. The function*
$$\lambda(r) := \log \mathrm{sn}_K(r) + \frac{K}{6}r^2$$
satisfies
$$\lambda(r) \leq (1-t)\lambda((1-t)r) + t\lambda(tr)$$
for every $t \in [0, 1]$.

Proof. Since, for $t = 0, 1$ is obvious, let $t \in (0, 1)$. We observe that, since $(1-t)t > 0$, the inequality in question is equaivalent to
$$\frac{\lambda((1-t)r) - \lambda(r)}{(1-t)-1} + \frac{\lambda(tr) - \lambda(r)}{t-1} \leq 0$$
which we can achieve through calculus' mean value theorem, if we show that $\lambda'(r) \leq 0$ for every r. We will check that this is true for different values of K. However, if we define the function $\tilde{\lambda}(r) := \lambda\left(\frac{r}{\sqrt{\pm K}}\right)$ one easily sees that $\tilde{\lambda}'(r) \leq 0 \Leftrightarrow \lambda'(r) \leq 0$. So we might assume that $K = 0,\ 1,\ -1$, without loss of generality.

- If $K = 0$, then $\lambda'(r) = 0 \leq 0$.
- If $K = 1$, then for $r \in (0, \pi)$ we have
$$\lambda(r) = \log\sin(r) - \log r + \frac{r^2}{6}$$
and
$$\lambda'(r) = \frac{\cos r}{\sin r} - \frac{1}{r} + \frac{r}{3}$$
Then $\lambda'(r) \leq 0 \Leftrightarrow 3r\cos r - 3\sin r + r^2 \sin r \leq 0$. We define two new functions:
$$f(r) = 3r\cos r - 3\sin r + r^2 \sin r$$
$$g(r) = r\cos r - \sin r$$

and with basic calculus we see that:

$$\begin{aligned}\lambda'(r) \le 0 \Leftrightarrow f(r) \le 0 = f(0) \Leftrightarrow \\ \Leftrightarrow f'(r) \le 0 \Leftrightarrow \\ \Leftrightarrow r(r\cos r - \sin r) \le 0 \overset{r>0}{\Longleftrightarrow} \\ \Leftrightarrow g(r) \le 0 = g(0) \Leftrightarrow \\ \Leftrightarrow g'(r) \le 0 \Leftrightarrow \\ \Leftrightarrow -r\sin r \le 0\end{aligned}$$

which is true for $r \in (0, \pi)$ (the second and fifth equivalence is due to the fact that $f'(r) = 0 \Leftrightarrow r = 0 \Leftrightarrow g'(r) = 0$).

- If $K = -1$ the calculations are somewhat similar:

$$\begin{aligned}\lambda'(r) \le 0 \Leftrightarrow 3r\cosh r - 3\sinh r - r^2\sinh r \le 0 \Leftrightarrow \\ \Leftrightarrow r(\sinh r - r\cosh r) \le 0 \overset{r>0}{\Longleftrightarrow} \\ \Leftrightarrow \sinh r - r\cosh r \le 0 \Leftrightarrow \\ \Leftrightarrow -r\sinh r \le 0\end{aligned}$$

which is, again, true for $r > 0$.

□

This last theorem would signify the end of our journey. We follow the proof of [17]:

Theorem. *For any smooth complete Riemannian manifold M and $\tilde{K} \in \mathbb{R}$ we have the following equivalence:*

$$\mathrm{Ric}(M) \ge \tilde{K} \iff \mathrm{Ent}(\cdot) \textit{ is } \tilde{K}\textit{-convex on } W_2.$$

Proof. First assume that $\mathrm{Ric}(M) \ge (n-1)K$, where $K \in \mathbb{R}$ is such that $\tilde{K} = (n-1)K$, and let $t \mapsto \mu_t : [0,1] \to \mathscr{P}^2(M)$ be a geodesic. We observe that if $\mathrm{Ent}(\mu_0) = +\infty$ or $\mathrm{Ent}(\mu_1) = +\infty$ then $\mathrm{Ent}(\mu_t) \le +\infty$ and the statement will be true. So we can assume that $\mathrm{Ent}(\mu_0)$, $\mathrm{Ent}(\mu_1) < \infty$ and consequentially μ_0, $\mu_1 << \mathrm{vol_g}$. Thus μ_t is the unique geodesic connecting μ_0 and μ_1 given by $\mu_t = (F_t)_\#\mu_0$ where $F_t(x) = \exp_x(-t\nabla\varphi(x))$ and φ c-concave.

The change of variables formula yields

$$\begin{aligned}\operatorname{Ent}(\mu_t) &= \int_M \frac{d\mu_t}{d\mathrm{vol_g}} \log \frac{d\mu_t}{d\mathrm{vol_g}} d\mathrm{vol_g} \\ &= \int_M \frac{d\mu_0}{d\mathrm{vol_g}} \left(\log \frac{d\mu_0}{d\mathrm{vol_g}} - \log \mathrm{J}_t \right) d\mathrm{vol_g} \\ &= \operatorname{Ent}(\mu_0) - \int_M \log \mathrm{J}_t \, d\mu_0\end{aligned}$$

and from this we can deduce:

$$\begin{aligned}&(1-t)\operatorname{Ent}(\mu_0) + t\operatorname{Ent}(\mu_1) - \operatorname{Ent}(\mu_t) \\ =&(\operatorname{Ent}(\mu_0) - \operatorname{Ent}(\mu_t)) - t(\operatorname{Ent}(\mu_0) - \operatorname{Ent}(\mu_1)) \\ =&\int_M \log \mathrm{J}_t \, d\mu_0 - t\int_M \log \mathrm{J}_1 \, d\mu_0 \\ =&\int_M \log \mathrm{J}_t - t \log \mathrm{J}_1 \, d\mu_0\end{aligned}$$

Due to the fact that logarithm is an increasing concave function, the Jacobian inequality yields:

$$\begin{aligned}\log \mathrm{J}_t(x) &\geq n \log \left[(1-t) v_{1-t}^{1/n}(F_1(x),\, x) + t v_t^{1/n}(x,\, F_1(x))\, \mathrm{J}_1^{1/n}(x) \right] \\ &\geq n \left[(1-t) \log v_{1-t}^{1/n}(F_1(x),\, x) + t \log \left(v_t(x,\, F_1(x))\, \mathrm{J}_1(x) \right)^{1/n} \right] \\ &= (1-t) \log v_{1-t}(F_1(x),\, x) + t \log v_t(x,\, F_1(x)) + t \log \mathrm{J}_1(x)\end{aligned}$$

Since Ricci curvature is bounded from below by K we can bound the volume coefficients from below, too. Thus

$$\begin{aligned}&\log \mathrm{J}_t(x) - t \log \mathrm{J}_1(x) \\ \geq\ & (1-t) \log v_{1-t}(F_1(x),\, x) + t \log v_t(x,\, F_1(x)) \\ \geq\ & (1-t) \log \left(\frac{\mathrm{sn}_K((1-t) d(F_1(x),\, x))}{\mathrm{sn}_K(d(F_1(x),\, x))} \right)^{n-1} + t \log \left(\frac{\mathrm{sn}_K(t d(x,\, F_1(x)))}{\mathrm{sn}_K(d(x,\, F_1(x)))} \right)^{n-1} \\ \geq\ & (n-1) \left[(1-t) \log \mathrm{sn}_K((1-t) d^x) + t \log \mathrm{sn}_K(t d^x) - \log \mathrm{sn}_K(d^x) \right] \\ \geq\ & \frac{(n-1)K}{2} t(1-t) d^2(x,\, F_1(x)) \\ \geq\ & \frac{\tilde{K}}{2} t(1-t) d^2(x,\, F_1(x))\end{aligned}$$

where $d^x := d(x,\, F_1(x))$ and the last inequality followed from the fact that

$$\begin{aligned}(1-t) \log \mathrm{sn}_K((1-t)r) + t \log \mathrm{sn}_K(tr) - \log \mathrm{sn}_K(r) - \frac{K}{2} t(1-t) r^2 \\ = (1-t)\lambda((1-t)r) + t\lambda(tr) - \lambda(r) \geq 0\end{aligned}$$

for all $t \in [0,\, 1]$ and $r \in \big(0,\, \frac{\pi}{\sqrt{K}}\big)$ (no generality is lost due to the Bonnet-Myers theorem and the fact that $\sinh r \geq 0$ for $r \geq 0$), where $\lambda(r) = \log \mathrm{sn}_K(r) + \frac{K}{6}r^2$, as we showed earlier. Thus

$$\begin{aligned}(1-t)\,\mathrm{Ent}(\mu_0) + t\,\mathrm{Ent}(\mu_1) - \mathrm{Ent}(\mu_t) &= \int_M \log \mathrm{J}_t(x) - t\log \mathrm{J}_1(x) d\mu_0(x)\\ &\geq \int_M \frac{\tilde{K}}{2} t(1-t) d^2(x,\, F_1(x)) d\mu_0(x)\\ &= \frac{\tilde{K}}{2} t(1-t) d^2_{W_2}(\mu_0,\, \mu_1)\end{aligned}$$

hence the $\tilde{K}$-convexity of $\mathrm{Ent}(\cdot)$ is proved.

Now, let's attack the converse statement. For this purpose suppose, on the contrary, that there exist $p \in M$, a unit vector $e_1 \in T_pM$ and some $\varepsilon > 0$ such that $\mathrm{Ric}_p(e_1, e_1) \leq \tilde{K} - \varepsilon$. Complete e_1 to an orthonormal basis of T_pM, $\{e_1,\, \ldots,\, e_n\}$, such that

$$R(e_i,\, e_1)e_1 = k_i e_i$$

where $k_i = \sec(e_1,\, e_i)$, $i = 1,\, \ldots,\, n$. Thus

$$\sum_{i=1}^n k_i = \mathrm{Ric}_p(e_1,\, e_1) \leq \tilde{K} - \varepsilon.$$

Consider a large normal ball around p and take $\delta,\, r > 0$ so that it includes $A_0 := B_\delta(\exp_p(-re_1))$ and $A_1 := B_\delta(\exp_p(re_1))$. Define n numbers:

$$\delta_i := \delta\left[1 + \frac{r^2}{2}\Big(k_i + \frac{\varepsilon}{2n}\Big)\right],\ i = 1,\, \ldots,\, n$$

and choose small enough r and δ so that the ellipsoid

$$A_{1/2} := \exp_p\left(\left\{v \in T_pM : \sum_{i=1}^n \left(\frac{v_i}{\delta_i}\right)^2 \leq 1\right\}\right)$$

will also be included in the large normal ball. Now since the geodesic that connects the centers of A_0 and A_1 passes through p we have that $\gamma\big(\frac{1}{2}\big) \in A_{1/2}$ for each minimizing geodesic $\gamma : [0,1] \to M$ with $\gamma(0) \in A_0$, $\gamma(1) \in A_1$. In order to check that, parallel translate the basis $\{e_1,\, \ldots,\, e_n\}$ along the geodesic joining the centers of A_0, A_1 and Taylor expand γ at p. Let $\mu_0,\, \mu_1,\, \nu$ be the normalized uniform distributions in A_0, A_1 and $A_{1/2}$ respectively. From the first property we proved for entropy and a local calculation on normal coordinates we have

$$
\begin{aligned}
\mathrm{Ent}(\mu_0) &= -\log \mathrm{vol}_{\mathrm{g}}(A_0) = -\log \omega_n - n\log\delta + O(\delta^2)\\
\mathrm{Ent}(\mu_1) &= -\log \mathrm{vol}_{\mathrm{g}}(A_1) = -\log \omega_n - n\log\delta + O(\delta^2)\\
\mathrm{Ent}(\nu) &= -\log \mathrm{vol}_{\mathrm{g}}(A_{1/2}) = -\log \omega_n - \sum_{i=1}^{n}\log\delta_i + O(\delta^2)
\end{aligned}
$$

where $\omega_n = \mathrm{Vol}_{\mathbb{R}^n}(B_1(0))$. From the definition of δ_i one has

$$
\begin{aligned}
\sum_{i=1}^{n}\log\delta_i &= \sum_{i=1}^{n}\left\{\log\delta + \log\left(1+\frac{r^2}{2}\Big(k_i+\frac{\varepsilon}{2n}\Big)\right)\right\}\\
&\leq \sum_{i=1}^{n}\left\{\log\delta + \frac{r^2}{2}\Big(k_i+\frac{\varepsilon}{2n}\Big)\right\}\\
&= n\log\delta + \frac{r^2}{2}\left(\sum_{i=1}^{n}k_i+\frac{\varepsilon}{2}\right)\\
&\leq n\log\delta + \frac{r^2}{2}\Big(\tilde{K}-\frac{\varepsilon}{2}\Big)
\end{aligned}
$$

hence

$$
\mathrm{Ent}(\nu) \geq -\log\omega_n - n\log\delta - \frac{r^2}{2}\Big(\tilde{K}-\frac{\varepsilon}{2}\Big) + O(\delta^2).
$$

Since the optimal mass transport from μ_0 to μ_1 (with respect to d_{W_2}) is along geodesics of M, we must have that $\mathrm{supp}(\mu_{1/2}) \subseteq A_{1/2}$ and the second property we proved for entropy yields

$$
\mathrm{Ent}(\mu_{1/2}) \geq \mathrm{Ent}(\nu).
$$

Combining the above we have:

$$
\begin{aligned}
&\mathrm{Ent}(\mu_{1/2}) - \frac{1}{2}\mathrm{Ent}(\mu_0) - \frac{1}{2}\mathrm{Ent}(\mu_1)\\
\geq& -\frac{r^2}{2}\Big(\tilde{K}-\frac{\varepsilon}{2}\Big) + O(\delta^2)\\
=& -\frac{r^2}{2}\tilde{K} + \frac{\varepsilon}{4}r^2 + O(\delta^2)\\
=& -\frac{r^2}{2}\tilde{K} + \frac{\varepsilon}{4}r^2 - \frac{\delta^2}{2}\tilde{K} + O(\delta^2)
\end{aligned}
$$

since $\frac{\delta^2}{2}\tilde{K} = O(\delta^2)$ and $O(\delta^2)+O(\delta^2) = O(\delta^2)$. Choosing $\delta << r$ small enough such that

$$
O(\delta^2) > -\frac{\varepsilon}{8}r^2 \quad \& \quad r > \frac{8|\tilde{K}|}{\varepsilon}\delta
$$

we get

$$\begin{aligned}
&\operatorname{Ent}(\mu_{1/2}) - \frac{1}{2}\operatorname{Ent}(\mu_0) - \frac{1}{2}\operatorname{Ent}(\mu_1)\\
\geq &-\frac{r^2}{2}\tilde{K} + \frac{\varepsilon}{4}r^2 - \frac{\delta^2}{2}\tilde{K} + O(\delta^2)\\
> &-\frac{r^2}{2}\tilde{K} + \frac{\varepsilon}{8}r^2 - \frac{\delta^2}{2}\tilde{K}\\
> &-\frac{r^2}{2}\tilde{K} + |\tilde{K}|r\delta - \frac{\delta^2}{2}\tilde{K}
\end{aligned}$$

We observe that for small enough $\delta << r << 1$ such that we're working on normal coordinates we have

$$0 < 2r - 2\delta \leq d_{W_2}(\mu_0,\, \mu_1) \leq 2r + 2\delta$$

since we have probability measures. Now, we distinguish two cases:

- If $\tilde{K} \geq 0$ we have

$$\begin{aligned}
&-\frac{r^2}{2}\tilde{K} + |\tilde{K}|r\delta - \frac{\delta^2}{2}\tilde{K}\\
= &-\frac{\tilde{K}}{8}(2r - 2\delta)^2\\
\geq &-\frac{\tilde{K}}{8}d^2_{W_2}(\mu_0,\, \mu_1)
\end{aligned}$$

  since $-\frac{\tilde{K}}{8} \leq 0$.

- If $\tilde{K} < 0$ we have

$$\begin{aligned}
&-\frac{r^2}{2}\tilde{K} + |\tilde{K}|r\delta - \frac{\delta^2}{2}\tilde{K}\\
= &-\frac{\tilde{K}}{8}(2r + 2\delta)^2\\
\geq &-\frac{\tilde{K}}{8}d^2_{W_2}(\mu_0,\, \mu_1)
\end{aligned}$$

  since $-\frac{\tilde{K}}{8} > 0$.

Thus

$$\operatorname{Ent}(\mu_{1/2}) > \frac{1}{2}\operatorname{Ent}(\mu_0) + \frac{1}{2}\operatorname{Ent}(\mu_1) - \frac{\tilde{K}}{8}d^2_{W_2}(\mu_0,\, \mu_1)$$

which is a contradiction, since entropy is $\tilde{K}$-convex. $\square$

Bibliography

[1] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften, Vol. 338. Springer-Verlag Berlin Heidelberg, 2008.

[2] C. Villani. *Topics in Optimal Transportation Theory*. Graduate Studies in Mathematics, Vol. 58. American Mathematical Society, 2003.

[3] J. Lee. *Introduction to Riemannian manifolds*. Graduate Texts in Mathematics, Vol. 176. Springer International Publishing, 2018.

[4] N. Dunford; J.T. Schwartz. *Linear operators, Vol. 1*. Wiley-Interscience, 1958.

[5] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Progress in Nonlinear Differential Equations and Their Applications, 87. Birkhäuser, 2015.

[6] L. C. Evans; R. F. Gariepy. *Measure Theory and Fine Properties of Functions, Revised Edition*. Textbooks in Mathematics. Chapman and Hall/CRC, 2015.

[7] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Pure and applied mathematics. New York: Wiley, 1999.

[8] D. Burago; Yu. Burago & S. Ivanov. *A course in metric geometry*. American Mathematical Society, Providence, RI, 2001.

[9] I. Chavel. *Riemannian Geometry: A Modern Introduction*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2006.

[10] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[11] P. Petersen. *Riemannian Geometry*. Graduate Texts in Mathematics, Vol. 171. Springer-Verlag New York, 2006.

[12] M. do Carmo. *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser Basel, 1992.

[13] Qi S. Zhang. *Sobolev Inequalities, Heat Kernels under Ricci Flow, and the Poincaré Conjecture*. Mathematics: Theory & Applications. CRC Press, 2010.

[14] K.T. Sturm. "On the geometry of metric measure spaces. I". In: *Acta Math.* 196 (2006), pp. 65–131. DOI: 10.1007/s11511-006-0002-8.

[15] S. Ohta. "On the measure contraction property of metric measure spaces". In: *Comment. Math. Helv* 82 (2007), pp. 805–828.

[16] Nicola Gigli. "The splitting theorem in non-smooth context". In: (2013).

[17] M.K. von Renesse; K.T. Sturm. "Transport inequalities, gradient estimates, entropy and Ricci curvature". In: *Comm. Pure Appl. Math.* 58 (2004). DOI: 10.1002/cpa.20060.

[18] Y. Brenier. "Décomposition polaire et réarrangement monotone des champs de vecteurs". In: *C.R. Acad. Sci. Paris Sér. I Math.* 305 (1987), pp. 805–808.

[19] Y. Brenier. "Polar factorization and monotone rearrangement of vector-valued functions". In: *Comm. Pure Appl. Math.* 44 (1991), pp. 375–417.

[20] J.-D. Benamou & Y. Brenier. "Weak existence for the semi-geostrophic equations formulated as a coupled Monge-Ampére / transport problem". In: *SIAM J. Appl. Math.* 58 (1998), pp. 1450–1461.

[21] S. Ohta. "Ricci curvature, entropy, and optimal transport". In: *Y. Ollivier, H. Pajot & C. Villani (Eds.), Optimal Transport: Theory and Applications (London Mathematical Society Lecture Note Series)* (2014), pp. 145–200. DOI: 10.1017/CBO9781107297296.008.

[22] J. Lott; C. Villani. "Ricci Curvature for Metric-Measure Spaces via Optimal Transport". In: *Annals of Mathematics, Vol. 169* (2009), pp. 903–991.

[23] Yu. Burago; M. Gromov & G. Perelman. "A.D. Alexandrov spaces with curvatures bounded below". In: *Russian Math. Surveys* ***47*** (1992), pp. 1–58.

[24] A.D. Alexandrov. "Almost everywhere existence of the second differential of a convex function and ome properties of convex surfaces connected with it (in Russian)". In: *Russian Math. Surveys* ***6*** (1939), pp. 3–35.

[25] G. Vitali. "On groups of points and functions of real variables (in Italian)". In: *'Accademia delle Scienze di Torino* ***43*** (1908), pp. 75–92.

[26] V. Bangert. "Analytishe Eigenschaften konvexer Funktionen auf Riemannschen Manigfaltigkeiten". In: *J. Reine Angew. Math.* ***307*** (1979), pp. 309–324.

[27] Y. Otsu; T. Shioya. "The Riemannian structure of Alexandrov spaces". In: *J. Differential Geom.* ***39*** (1994), pp. 629–658.

[28] A. Figalli; C. Villani. "Optimal Transport and Curvature". In: *Nonlinear PDE's and Applications. Lecture Notes in Mathematics, Vol. 2028* (2011), pp. 171–217.

[29] L. Ambrosio; N. Gigli. "A user's guide to optimal transport". In: *Modelling and Optimisation of Flows on Networks. Lecture Notes in Mathematics* Vol. 2062 (2009). DOI: https://doi.org/10.1007/978-3-642-32160-3_1.

[30] R. J. McCann. "Polar factorization of maps on Riemannian manifolds". In: *GAFA, Geom. funct. anal.* Vol. 11 (1999), pp. 589–608. DOI: https://doi.org/10.1007/PL00001679.

[31] D. Cordero-Erausquin; R. J. McCann; M. Schmuckenschläger. "A Riemannian interpolation inequality à la Borell, Brascamp and Lieb". In: *Invent. math* Vol. 146 (2001), pp. 219–257. DOI: https://doi.org/10.1007/s002220100160.

[32] D. Cordero-Erausquin; R. J. McCann; M. Schmuckenschläger. "Prékopa–Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport". In: *Annales de la Faculté des sciences de Toulouse : Mathématiques* Vol. 15 (2006), pp. 613–635.