Durham E-Theses

Discrete curvatures motivated from Riemannian geometry and optimal transport: Bonnet-Myers-type diameter bounds and rigidity

KAMTUE, SUPANAT

How to cite:

KAMTUE, SUPANAT (2021) *Discrete curvatures motivated from Riemannian geometry and optimal transport: Bonnet-Myers-type diameter bounds and rigidity*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/14124/

Use policy

This work is licensed under a Creative Commons Attribution 3.0 (CC BY)

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

Discrete curvatures motivated from Riemannian geometry and optimal transport

Bonnet-Myers-type diameter bounds and rigidity

Supanat Kamtue

A thesis submitted for the degree of
Doctor of Philosophy

Department of Mathematical Sciences
Durham University
United Kingdom
July 2021

Discrete curvatures motivated from Riemannian geometry and optimal transport

Bonnet-Myers-type diameter bounds and rigidity

Supanat Kamtue

Abstract

This thesis gives an overview of three notions of Ricci curvature for discrete spaces, including Ollivier Ricci curvature (motivated from optimal transport), Bakry-Émery curvature (from Bochner's formula in Riemannian geometry) and Erbar-Maas entropic Ricci curvature (from optimal transport). The first part of the thesis provides background knowledge in optimal transport theory and Riemannian geometry which is essential to the understanding of generalized Ricci curvatures for metric measure spaces and the mentioned Ricci curvatures for graphs.

For each of the three discrete curvature notions, discussed in their respective part of the thesis, we provide the definition of the curvature and use hypercubes as an example for the curvature calculation. We study various curvature results with an emphasis on upper bounds of diameter and lower bounds of the spectral gap for graphs with positive lower bound on the Ricci curvature. These results can be regarded as discrete analogues of the Bonnet-Myers theorem and the Lichnerowicz theorem in Riemannian geometry. In addition, we deeply investigate into the rigidity results (analogous to Cheng's rigidity) in attempt to classify all graphs which yield the sharp diameter bound in the sense of Ollivier Ricci curvature and Bakry-Émery curvature.

Preface

Discrete curvatures in this thesis refer to discrete versions of Ricci curvature (which is a geometric concept on manifolds) for discrete spaces such as graphs and discrete Markov chains. Unlike on manifolds, there is no unified notion of curvature on graphs, and each of discrete curvature notions is inspired by different geometric aspects of Ricci curvature. In particular, we focus on three discrete curvature notions, namely,

1. *Ollivier Ricci curvature*, which is defined via the contraction property of L^1-transportation distance,

2. *Bakry-Émery curvature*, which is defined via Bochner's formula in Riemannian geometry and the Laplace operator,

3. *Erbar-Maas entropic Ricci curvature*, which is defined via the convexity of the entropy functional along L^2-transportation geodesics.

There are other discrete curvature notions that are beyond the scope of the thesis, for example, combinatorial curvature on tessellated surfaces, Forman-Ricci curvature on simplicial complexes, and sectional-type curvature for Alexandrov spaces.

This thesis is divided into four main parts, where each of Parts II III and IV is dedicated to one of the three discrete curvature notions mentioned above. Part I provides background knowledge in optimal transport theory and Riemannian geometry that is helpful for understanding the discrete curvatures. Outlined below are topics which are covered in each part.

Part I begins with an introduction to optimal transport theory (Kantorovich's problem and its duality) with an emphasis towards geometries of the L^2-Wasserstein space of probability measures. Otto's formal calculus [Ott01] is used as a framework to view the Wasserstein space as a "Riemannian manifold" and to describe, for example, the transportation distance in fluid dynamics viewpoint (Benamou-Brenier formula [BB00]) and the heat flow by a gradient flow structure (Jordan-Kinderlehrer-Otto theorem [JKO98]). We also discuss three famous approaches by (i) Sturm and Lott-Villani [Stu06, LV09], (ii) Bakry-Émery [BE84], and (iii) Ollivier [Oll09] to generalize Ricci curvature for metric measure spaces. These approaches give rise to the discrete curvatures under discussion. A central theme of the thesis is about the Bonnet-Myers diameter bound theorem [Mye41] and its rigidity (known as Cheng's rigidity [Che75]). They are classical theorems in Riemannian geometry and their analogous results will be re-appearing in the context of each discrete curvature.

In Part II about Ollivier Ricci curvature on graphs, we revisit an optimal transport in the setting of graphs, which prompt us to the curvature's definition. We discuss various properties of Ollivier Ricci curvature, including Cartesian products, a lower bound on the spectral gap (Lichnerowicz-type theorem) and a Bonnet-Myers-type upper bound on diameter. The key result in this part is my own contribution to the rigidity result in collaboration with David Cushing, Jack Koolen, Shiping Liu, Florentin Münch and my supervisor Norbert Peyerimhoff [CKK+20], including the crucial concept of *transport geodesics* and the classification of all self-centered Bonnet-Myers sharp graphs. The presentation in this thesis deviates somewhat from [CKK+20] by providing an alternative proof of a preliminary rigidity result using only transport geodesics and no properties of the Laplacian. It is also a natural question whether the assumption of self-centeredness can be removed from our classification result. Some partial progress (removal of self-centeredness in diameter three) is provided in [Kam20a].

Part III features Bakry-Émery curvature on weighted graphs and our joint research result [CKLP21] on the curvature reformulation as a minimal eigenvalue of a local matrix. Moreover, we follow the Bakry-Émery theory to provide, in the discrete setting, an equivalent characterization of Bakry-Émery curvature via a gradient estimate of the heat semigroup. In parallel to the previous part, we discuss relevant properties regarding Cartesian products and a lower bound on the spectral gap. We then survey some key ideas of research results by Liu, Münch and Peyerimhoff [LMP18, LMP17] on a Bonnet-Myers-type diameter bound and its rigidity with respect to Bakry-Émery curvature.

In Part IV, we present Erbar and Maas' work [EM12] to define, on discrete Markov chains, a discrete notion of Ricci curvature in the spirit of Sturm and Lott-Villani, called the entropic Ricci curvature. This approach requires a modification of the 2-Wasserstein distance which is motivated from Otto's calculus (in particular, Benamou-Brenier formula). Similarly to Bakry-Émery curvature, we also provide an equivalent reformulation of the entropic Ricci curvature in terms of a Bochner-type formula and a gradient estimate. These tools are crucial to derive a Bonnet-Myers type diameter upper bound for a Markov chain with positive entropic Ricci curvature. This result is a generalization of my paper [Kam20b], which deals with the special case of simple random walks.

Explicit computation of all three discrete curvatures is given for hypercubes as they represent an important family of graphs which possess positive curvature, a product structure, and (potentially) sharp diameter bounds with respect to all discrete curvatures. Each of the three parts finishes with an outlook chapter discussing some related open questions.

Readers who are particularly interested in one of these discrete curvatures can start reading from its respective part without prior knowledge from Part I. During the discussion in discrete curvatures, there usually are pointers which reconnect us to related topics in the manifold case in Part I. Readers with differential geometry and Riemannian geometry background may find the interplay between Ricci curvature and optimal transport interesting, in particular since the Lott-Sturm-Villani synthetic Ricci curvature notion allows to extend various fundamental curvature results to the more general setting of metric measure spaces, which appear naturally via Gromov-Hausdorff convergence of Riemannian manifolds.

In summary, besides its research results, this thesis serves also as a survey on topics which are motivated from Riemannian geometry and optimal transport with special emphasis on Bonnet-Myers-type diameter bounds and the corresponding rigidity results.

Supanat Kamtue
July 2021

Declaration

This thesis is the result of the author's own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. No substantial part of this thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at Durham University or any other institution except as declared in the Preface and specified in the text.

Copyright ©2021 Supanat Kamtue

"The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged."

Acknowledgements

The following list contains my acknowledgement to people in various categories ranging from ones with direct impact to my research projects to ones who contribute to a supporting environment around me. All people in this list are by no means listed by preference, and my appreciation toward them are not comparable (with an exception for the first person I am about to mention).

The first and the most important person whom I like to thank is my PhD supervisor Norbert Peyerimhoff, who undoubtedly has the most direct impact to this work. He gives many insightful suggestion to the structure and contents of this thesis. His supervision is far beyond my dissertation, PhD research or even mathematical knowledge. He has been a very helpful and reliable guidance to my academic life, and his character has inspired me to become an eager mathematician.

The next two persons are David Cushing and Shiping Liu, who are my favorite colleagues and coauthors of many papers in our joint research projects. David always comes up with questions which often become starting points for new projects. Among our research group, Shiping is the person who brought to our attention the trending topic on discrete curvatures. Shiping's knowledge in Riemannian geometry and curvature is always dependable to our projects. I would also like to thank him and his wife, Yuli, and his university (University of Science and Technology of China, Hefei) for their generous hospitality for several of Norbert's and my academic visits.

I would also like to thank all of my coauthors during my PhD study. This list includes Florentin Münch, Jack Koolen, Riikka Kangaslampi, Leyna Watson-May, and Pakawut Jiradilok.

Next I like to thank Durham's math department, including very efficient staff in the office. I would like to thank many people there for building up an encouraging atmosphere for learning. The first group is my office mates and other fellows in the postgraduate seminar group: Irene Pasquinelli, Samuël Borza, John Blackman, Oliver Singh, Matt Northey, Clare Wallace, Laura Midgley, Wanchalerm Promduang, Nawapon Nakharutai, Rob Little, and Daniel Ballesteros. The second group is people in Optimal transport study group including David Bourne, Wilhelm Klingenberg, Djoko Wirosoetisno, Mikaella Iacobelli, Fernando Galaz-Garcia, Alpár Mészáros, an occasional guest speaker, Alessio Figalli.

Next I extend my thanks to all mathematicians whom I had a chance to discuss related mathematics to. Especially, I thank Yong Lin and Bobo Hua for various discussion on analysis on graphs. I thank Nicolas Juillet for a question on the contraction property of random walks on discrete Heisenberg group (which gives a motivation for my upcoming research paper). I also thank Lorenzo Portinale, a colleague whom I met at Lake Como school and Oberwolfach seminar and also works on the entropic Ricci curvature.

I also express my formal gratitude to Thai's Institute for the Promotion of Teaching Science and Technology for the financial supports from my BSc program in the United States to my PhD program in the United Kingdom.

Last but not least, I would like to thanks my mom, dad and grandpa who have supported me from the beginning of my mathematics journey since I was in an elementary school.

Contents

Part I

Curvature notions on metric measure spaces

Chapter 1

Introduction to optimal transport

Let us start with a brief introduction to optimal transport in a general setting of metric measure spaces, where the topics include Monge's problem, Kantorovich's problem, the duality, and the Wasserstein space. In the later section, we will restrict to the setting of Riemannian manifolds, where geometric structures of the Wasserstein space are more enriched. For the contents given in this chapter, we mostly refer to a reader-friendly note by Ambrosio and Gigli [AG13], and occasionally from a book by Santambrogio [San15] and a book by Villani [Vil09].

1.1 Setting of metric measure spaces

Let (X, d) and (Y, d') be Polish spaces (i.e., complete and separable metric spaces), equipped with σ-Borel algebra. Define $\mathscr{P}(X)$ and $\mathscr{P}(Y)$ to be the sets of all Borel probability measures on X and on Y, respectively.

For a given Borel map $T : X \to Y$ and a measure $\mu \in \mathscr{P}(X)$, the ***pushforward*** of μ by T is the measure denoted by $T_{\#}\mu \in \mathscr{P}(Y)$ such that

$$T_{\#}\mu(B) = \mu(T_{-1}(B)),$$

for all Borel subsets $B \subset Y$. Equivalently, the pushforward satisfies the following change of variable formula:

$$\int_Y f(y) dT_{\#}\mu(y) = \int_X f(T(x)) d\mu(x),$$

for all Borel functions $f : X \to Y$.

The original optimal transport problem proposed by Monge in [Mon81] is defined as follows.

Definition 1.1. Given a Borel cost function $c : X \times Y \to \mathbb{R}_{\geq 0}$. Let $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$. ***Monge's optimal transport problem*** is the following minimization problem:

$$\inf_T \left\{ \operatorname{cost}(T) := \int_X c(x, T(x)) d\mu(x) \,\middle|\, T_\# \mu = \nu \right\}. \tag{MP}$$

Any map $T : X \to Y$ such that $T_\# \mu = \nu$ is called a ***transport map*** from μ to ν. A transport map T which minimizes (MP) is called an ***optimal transport map***.

Intuitively, a map T transports the collection of mass distributed as in μ to mass distributed as in ν by moving mass from location x to location $T(x)$ (without splitting mass) with the transport cost given by $c(x, T(x))$ per unit mass. Monge's problem is to find the minimum total cost of all such transport maps T. One limitation of Monge's problem is about the possibility of non-existence of transport maps. For instance, in a situation that $\mu = \delta_x$ is a Dirac measure at x but ν is not a Dirac measure (e.g. $\nu = \frac{1}{2}\delta_y + \frac{1}{2}\delta_{y'}$ with $y \neq y'$), no transport map exists since the splitting of mass at x is not allowed.

In [Kan42], Kantorovich considers a more general optimal transport problem which allows the mass to be split at the source.

Definition 1.2. Given a Borel cost function $c : X \times Y \to \mathbb{R}_{\geq 0}$. Let $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$. A ***transport plan*** from μ to ν is a Borel probability measure $\Gamma \in \mathscr{P}(X \times Y)$ that satisfies the following ***marginal constraints***:

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

for all Borel $A \subseteq X$ and $B \subseteq Y$. Equivalently, the marginal constraints can be expressed by

$$(\pi_1)_\# \Gamma = \mu \text{ and } (\pi_2)_\# \Gamma = \nu,$$

where π_i denotes the ith-projection: $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$. Furthermore, we denote by $\Pi(\mu, \nu)$, the set of all transport plans from μ to ν.

Kantorovich's optimal transport problem (also known as the Monge-Kantorovich problem) is the following minimization problem:

$$\inf_\Gamma \left\{ \operatorname{cost}(\Gamma) := \int_{X \times Y} c(x, y) d\Gamma(x, y) \,\middle|\, \Gamma \in \Pi(\mu, \nu) \right\}. \tag{KP}$$

Any such plan Γ which minimizes (KP) is called an ***optimal transport plan***.

Intuitively, $\Gamma(x, y)$ represents the amount of mass that is moved from x to y. In contrast to Monge's problem, there always exists a transport plan for Kantorovich's problem (as a trivial one is given by $\Gamma = \mu \times \nu$).

It is noteworthy that the support of a transport plan lies in the product of the marginal supports.

Definition 1.3. The support of a measure $\mu \in \mathscr{P}(X)$ is defined as the set

$$\mathrm{supp}(\mu) := \{\, x \in X \mid \mu(U) > 0 \quad \forall \text{ open neighborhood } U \text{ of } x \,\} \subset X.$$

Proposition 1.4. *Let $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$, and let $\Gamma \in \prod(\mu, \nu)$. Then*

$$\mathrm{supp}(\Gamma) \subset \mathrm{supp}(\mu) \times \mathrm{supp}(\nu).$$

Proof. Assume that $(x, y) \in \mathrm{supp}(\Gamma)$. Then the marginal constraints implies that, for any open $x \in A \subset X$ and $y \in B \subset Y$, we have $\mu(A) = \Gamma(A \times X) > 0$ and $\nu(B) = \Gamma(X \times B) > 0$, which means $x \in \mathrm{supp}(\mu)$ and $y \in \mathrm{supp}(\nu)$. □

The connection between transport maps and transport plans can be described as in the following proposition.

Proposition 1.5. *Given $\mu \in \mathscr{P}(X)$ and two Borel maps $F : X \to X$ and $G : X \to Y$. Let the map $(F, G) : X \to X \times Y$ be defined as $x \mapsto (F(x), G(x))$. Then $\Gamma := (F, G)_{\#}\mu$ is a transport plan from $F_{\#}\mu$ to $G_{\#}\mu$, and its cost is given by*

$$\mathrm{cost}(\Gamma) = \int_X c(F(x), G(x)) d\mu(x). \tag{1.1}$$

Proof. For $A \subset X$ and $B \subset Y$, we have

$$\Gamma(A \times B) = \mu((F, G)^{-1}(A \times B)) = \mu(F^{-1}(A) \cap G^{-1}(B)).$$

In particular, Γ satisfies the marginal constraints $\Gamma(A \times X) = \mu(F^{-1}(A)) = F_{\#}\mu(A)$ and $\Gamma(X \times A) = \mu(G^{-1}(B)) = G_{\#}\mu(A)$.

The change of variable formula then gives

$$\mathrm{cost}(\Gamma) = \int_{X \times Y} c(x, y) d(F, G)_{\#}\mu(x, y) = \int_X c((F, G)x) d\mu(x)$$

as desired. □

An immediate consequence from Proposition 1.5 is that, for any Borel map $T : X \to Y$, the pushforward $\Gamma := (\mathrm{id}, T)_{\#}\mu$ is a transport plan from μ to $T_{\#}\mu$ with $\mathrm{cost}(\Gamma) = \mathrm{cost}(T)$. This means the problem (KP) can be considered as a relaxation to the problem (MP), and the infimum in (KP) is less than or equal to the infimum in (MP):

$$\inf_{\Gamma \in \Pi(\mu,\nu)} \mathrm{cost}(\Gamma) \leq \inf_{T_{\#}\mu = \nu} \mathrm{cost}(T).$$

When (KP) is regarded as a linear programming problem in Γ, there occurs naturally a dual problem.

Definition 1.6. Given a Borel cost function $c : X \times Y \to \mathbb{R}_{\geq 0}$. Let $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$. Consider all pairs of functions $\phi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$ which satisfy

$$\phi(x) + \psi(y) \leq c(x, y) \qquad \forall x \in X, \forall y \in Y$$

or written shortly as $\phi \oplus \psi \leq c$.

Kantorovich's dual problem is the following maximization problem:

$$\sup_{\phi,\psi} \left\{ J_{\mu,\nu}(\phi, \psi) := \int_X \phi(x) d\mu(x) + \int_Y \psi(y) d\nu(y) \,\middle|\, \phi \oplus \psi \leq c \right\}. \tag{DP1}$$

For further discussion on the dual problem, let us recall some relevant concepts of c-transform and c-concavity.

Definition 1.7. Given a cost function $c : X \times Y \to \mathbb{R}_{\geq 0}$. Consider functions $\phi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$. The ***c-transform*** of ϕ is a function $\phi^c : Y \to \mathbb{R}$ defined as

$$\phi^c(y) := \inf_{x \in X} c(x, y) - \phi(x) \quad \forall y \in Y.$$

A function $\psi : Y \to \mathbb{R}$ is called ***c-concave*** if $\psi = \phi^c$ for some $\phi : X \to \mathbb{R}$.

Similarly, the c-transform of ψ is a function $\psi^c : X \to \mathbb{R}$ defined as

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

A function $\phi : X \to \mathbb{R}$ is called c-concave if $\phi = \psi^c$ for some $\psi : Y \to \mathbb{R}$. Moreover, we denote by $\mathrm{Lip}^c(X)$, the set of all c-concave functions on X.

Here are some basic properties of c-transform and c-concave functions.

Proposition 1.8. *Let $\phi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$. Then* *(i)* $\phi \oplus \phi^c \leq c$, *and* $\varphi \leq \phi^c$ *for all* $\varphi : Y \to \mathbb{R}$ *such that* $\phi \oplus \varphi \leq c$.

(ii) $\phi^{ccc} = \phi^c$, *and similarly* $\psi^{ccc} = \psi^c$.

(iii) $\phi^{cc} = \phi$ *if and only if* $\phi \in \mathrm{Lip}^c(X)$.

Proof. (i) follows directly from the definition. To see (ii), we observe

$$\begin{aligned}\phi^{ccc}(y) &= \inf_x c(x,y) - \big(\inf_{\tilde{y}} c(x,\tilde{y}) - (\inf_{\tilde{x}} c(\tilde{x},\tilde{y}) - \phi(\tilde{x}))\big) \\ &= \inf_x \sup_{\tilde{y}} \inf_{\tilde{x}} \underbrace{c(x,y) - c(x,\tilde{y}) + c(\tilde{x},\tilde{y}) - \phi(\tilde{x})}_{=:\, \Phi(x,y,\tilde{x},\tilde{y})}.\end{aligned}$$

Thus

$$\begin{aligned}\phi^{ccc}(y) &\geq \inf_x \inf_{\tilde{x}} \Phi(x,y,\tilde{x},y) = \phi^c(y), \\ \phi^{ccc}(y) &\leq \inf_x \sup_{\tilde{y}} \Phi(x,y,x,\tilde{y}) = \phi^c(y),\end{aligned}$$

and therefore $\phi^{ccc} = \phi^c$. (iii) then follows from (ii). □

From the above proposition, one can see that

$$\sup_{\phi \oplus \psi \leq c} J_{\mu,\nu}(\phi,\psi) = \sup_{\phi \in \mathrm{Lip}^c(X)} J_{\mu,\nu}(\phi,\phi^c).$$

Definition 1.9. Kantorovich's dual problem can be reformulated as the maximization problem:

$$\sup_{\phi} \left\{ J_{\mu,\nu}(\phi,\phi^c) := \int_X \phi(x)d\mu(x) + \int_Y \phi^c(y)d\nu(y) \,\middle|\, \phi \in \mathrm{Lip}^c(X) \right\}. \tag{DP2}$$

Any $\phi \in \mathrm{Lip}^c(X)$ that maximizes (DP2) is called an ***optimal Kantorovich potential*** with respect to μ and ν.

The relation between Kantorovich's problem (KP) and its dual problem (DP2) is given by the so-called Kantorovich duality. Moreover, minimizers of (KP) and maximizers of (DP2) exist under a certain condition on the cost function c, as given below. We refer to [AG13, Theorem 1.17] for more details and proof.

Theorem 1.10 (Kantorovich duality)**.** *Let* $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ *and let* $c : X \times Y \to \mathbb{R}_{\geq 0}$ *be a continuous cost function such that*

$$c(x,y) \leq a(x) + b(y), \tag{1.2}$$

for some $a \in L^1(\mu)$ and $b \in L^1(\nu)$. Then Kantorovich's problem and its dual problem satisfy the following ***Kantorovich duality****:*

$$\inf_{\Gamma \in \Pi(\mu,\nu)} \mathrm{cost}(\Gamma) = \sup_{\phi \in \mathrm{Lip}^c(X)} J_{\mu,\nu}(\phi, \phi^c). \tag{1.3}$$

Furthermore, there always exist an optimal transport plan Γ and an optimal Kantorovich potential ϕ.

Remark 1.11. 1. The assumption $c \leq a \oplus b$ with $a \in L^1(\mu)$ and $b \in L^1(\nu)$ guarantees that any transport plan $\Gamma \in \Pi(\mu, \nu)$ has a finite cost:

$$\int_{X \times Y} c(x,y) d\Gamma(x,y) \leq \int_X a(x) d\mu(x) + \int_Y b(y) d\nu(y) < \infty.$$

Without this assumption, the duality (1.3) still holds true in general by allowing infinite values for both sides.

2. If $X = Y$, and the cost function is given by $c = d$, the metric of X, c-concave functions are exactly 1-Lipschitz functions, and the c-transform of $\phi \in \mathrm{Lip}^c(X)$ is given by $\phi^c = -\phi$.

Theorem 1.12 (complementary slackness)**.** *Assume that Γ is an optimal transport plan and ϕ is an optimal Kantorovich potential from μ to ν. Then*

$$\phi(x) + \phi^c(y) = c(x,y) \quad \forall (x,y) \in \mathrm{supp}(\Gamma). \tag{1.4}$$

Proof. Since Γ and ϕ are a minimizer and a maximizer in the duality, we have

$$\begin{aligned} 0 &= \mathrm{cost}(\Gamma) - J(\phi, \phi^c) \\ &= \int_{X \times Y} c(x,y) d\Gamma(x,y) - \int_X \phi(x) d\mu(x) + \int_Y \phi^c(y) d\nu(y) \\ &= \int_{X \times Y} \left(c(x,y) - \phi(x) - \phi^c(y) \right) d\Gamma(x,y), \end{aligned}$$

due to the marginals $d\mu(x) = \int_{y \in Y} d\Gamma(x,y)$ and $d\nu(y) = \int_{x \in X} d\Gamma(x,y)$. Since $c(x,y) - \phi(x) - \phi^c(y) \geq 0$, the equality (1.4) must holds Γ-almost everywhere, and by continuity of c, it holds everywhere on $\mathrm{supp}(\Gamma)$. For more details, we refer to [AG13, Remarks 1.15 and 1.18]. □

From now on, we always consider a situation where $X = Y$ and the cost function c is the power of the distance function d, that is, $c = d^r$ for some fixed $r \in$

$[1,\infty)$. The minimum value of Kantorovich's problem (KP) defines the so-called ***L^r-Wasserstein distance*** between two measures as

$$W_r(\mu,\nu) := \left(\inf_{\Gamma \in \prod(\mu,\nu)} \int_{X\times X} d(x,y)^r d\Gamma(x,y) \right)^{1/r}. \tag{1.5}$$

It is important to remark that W_r defines a metric on $\mathscr{P}_r(X) \subset \mathscr{P}(X)$, the space of all Borel probability measures with finite r-moments, that is,

$$\mathscr{P}_r(X) := \left\{ \mu \in \mathscr{P}(X) \,\middle|\, \int_X d(x_0,x)^r d\mu(x) < \infty \text{ for some } x_0 \in X \right\}.$$

Proposition 1.13. *For $r \in [1,\infty)$, W_r is a metric on $\mathscr{P}_r(X)$.*

A proof to the above proposition can be found in e.g. [AG13, Theorem 2.2] in the case of W_2, but it can be generalized to any W_r for $r \geq 1$. The idea to prove triangle inequality is to apply the gluing lemma to compose transport plans and then apply Minkowski's inequality.

The finite r-moment restriction guarantees that $W_r(\delta_{x_0},\mu) < \infty$ for some (and hence for all) x_0, and therefore $W_r(\mu,\nu) < \infty$ for all $\mu,\nu \in \mathscr{P}_r(X)$ due to the triangle inequality.

Definition 1.14 (Wasserstein space)**.** For $r \in [1,\infty)$, the ***L^r-Wasserstein space*** is the space of probability measures $\mathscr{P}_r(X)$ endowed with the L^r-Wasserstein metric W_r.

The following proposition states some basic topological properties of the Wasserstein space $(\mathscr{P}_r(X), W_r)$; for details see [Vil09, Theorem 6.18 and Remark 6.19].

Proposition 1.15 (Topological properties of $\mathscr{P}_r(X)$)**.** *Let (X,d) be a complete and separable metric space and $r \in [0,\infty)$. Then $(\mathscr{P}_r(X), W_r)$ is also a complete and separable metric space. Moreover, if X is compact, then $\mathscr{P}_r(X)$ is also compact. However, if X is only locally compact, then $\mathscr{P}_r(X)$ is not locally compact.*

Remark 1.16. Note that there is a trivial embedding from X into $\mathscr{P}_r(X)$ given by delta distributions. The embedding is isometric, i.e., $W_r(\delta_x,\delta_y) = d(x,y)$ because transport maps from δ_x to δ_y are those maps T such that $T(x) = y$.

Another important property of the Wasserstein space is about geodesic. Let us recall the definition of geodesics and geodesic spaces.

Definition 1.17 (geodesics in metric spaces)**.** Let (X,d) be a metric space. A curve $\gamma : I \to X$ (for some interval $I \subset \mathbb{R}$) is called a ***constant speed geodesic***

if there is a constant $v \geq 0$ such that $d(\gamma_s, \gamma_t) = v|t-s|$ for all $s, t \in I$. Here we write $\gamma_t := \gamma(t)$.

In particular, a curve $\gamma : [0,1] \to X$ is a constant speed geodesic if for all $s, t \in [0,1]$,

$$d(\gamma_s, \gamma_t) = |t-s| d(\gamma_0, \gamma_1). \tag{1.6}$$

Remark 1.18. Equivalently, one can weaken the condition (1.6) by replacing it with $d(\gamma_s, \gamma_t) \leq |t-s| d(\gamma_0, \gamma_1)$. For $0 \leq s \leq t \leq 1$, the triangle inequality gives

$$d(\gamma_0, \gamma_1) \leq d(\gamma_0, \gamma_s) + d(\gamma_s, \gamma_t) + d(\gamma_t, \gamma_1) \leq (s + (t-s) + (1-s)) d(\gamma_0, \gamma_1),$$

which makes these inequalities hold with equality, and $d(\gamma_s, \gamma_t) = |t-s| d(\gamma_0, \gamma_1)$ is recovered.

Definition 1.19 (geodesic spaces)**.** A metric space (X, d) is a ***geodesic space*** if for every $x, y \in X$, there exists a constant speed geodesic joining x and y, namely $\gamma : [0,1] \to X$ such that $\gamma_0 = x$ and $\gamma_1 = y$ satisfying (1.6).

If there exists a geodesic $\gamma : [0,1] \to X$ from x to y in X, then the curve $t \mapsto \delta_{\gamma(t)}$ is indeed a geodesic in $\mathscr{P}_r(X)$ from δ_x to δ_y. In fact, if X is a geodesic space, then $\mathscr{P}_r(X)$ is also a geodesic space (see e.g. [AG13, Theorem 2.10] for the proof).

Theorem 1.20. *Suppose that (X, d) is a geodesic space. Then the Wasserstein space $(\mathscr{P}_2(X), W_2)$ is also a geodesic space.*

On the other hand, when X is a discrete space (which means X it is not a geodesic space), the Wasserstein space $(\mathscr{P}_2(X), W_2)$ is not a geodesic space (see Chapter 9). This fact leads to Erbar-Maas' approach to define Ricci curvature notion on discrete spaces, which is later discussed in Part IV.

Although it is not a focus of this work, we would like to remark another property of the Wasserstein space related to a lower curvature bound in the sense of Alexandrov. We refer to [Oht14, Section 8.2] for the following theorem and to [BBI01] for general knowledge on Alexandrov spaces.

Definition 1.21. A geodesic space (X, d) is an ***Alexandrov space of curvature*** $\geq k$ if for every geodesic triangle in X is "thicker" than the triangle of the same side length in $\mathbb{M}^2(k)$, the space of constant sectional curvature k. More precisely, the space X must satisfy the following property. Let $\triangle xyz$ be any triangle in X and let $\triangle x'y'z'$ be the corresponding triangle in $\mathbb{M}^2(k)$, that is, $d(x, y) = \mathrm{d}(x', y')$ and so on. For any point p on the edge xy and the corresponding point p' on $x'y'$ (that is, $d(x, p) = \mathrm{d}(x', p')$), it must satisfy $d(z, p) \geq \mathrm{d}(z', p')$.

Note that in particular, $\mathbb{M}^2(1)$ is the sphere of radius one, $\mathbb{M}^2(0)$ is the Euclidean plane, and $\mathbb{M}^2(-1)$ is the real hyperbolic plane.

Theorem 1.22. *A compact geodesic space (X, d) is an Alexandrov space of nonnegative curvature if and only if the Wasserstein space $(\mathscr{P}_2(X), W_2)$ is an Alexandrov space of nonnegative curvature. Moreover, if (X, d) is not an Alexandrov space of nonnegatively curvature, then $(\mathscr{P}_2(X), W_2)$ is not an Alexandrov space of any curvature lower bound.*

1.2 Setting of Riemannian manifolds

In this section, we focus on geometric properties of the L^2-Wasserstein space $(\mathscr{P}_2(M), W_2)$ in the case that $(M, \langle\cdot,\cdot\rangle)$ is a connected complete Riemannian manifold. The understanding these properties, especially McCann's theorem, will prepare us for later discussion about the synthetic Ricci curvature notion. For a reference on geometric properties of $(\mathscr{P}_2(M), W_2)$, we suggest the work by Lott [Lot08].

One of the most remarkable results is about the existence and uniqueness of the optimal transport map on a Riemannian manifold, and it also gives the characterization of such map in terms of a Kantorovich potential. This result is proved by Brenier [Bre87] for Euclidean space $\mathbb{R}^n$, and it is generalized by McCann for Riemannian manifolds [McC01]. These results play an important role in the discussion of Otto's calculus in Chapter 2.

Theorem 1.23 (Brenier's theorem)**.** *Let $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^n)$ with μ absolutely continuous with respect to Lebesgue measure. Then there exists a unique optimal transport plan from μ to ν, and this plan is induced by a transport map $T : \mathbb{R}^n \to \mathbb{R}^n$ given by*

$$T(x) := x - \operatorname{grad}\varphi(x),$$

where $\varphi = \frac{1}{2}\phi$ for some optimal Kantorovich potential ϕ.

Here in the case of Euclidean space, the gradient $\operatorname{grad}\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is given at any point $x = (x_1, ..., x_n) \in \mathbb{R}^n$ by $\operatorname{grad}\varphi(x) = \left(\frac{\partial\varphi}{\partial x_1}(x), ..., \frac{\partial\varphi}{\partial x_n}(x)\right)$. The definition of the gradient in a more general setting of manifolds is discussed in Section 5.1.

Theorem 1.24 (McCann's theorem)**.** *Let M be a connected and closed manifold. Let $\mu, \nu \in \mathscr{P}_2(M)$ such that $\mu \ll \mathrm{vol}$. Then there exists a unique optimal transport plan from μ to ν, and this plan is induced by a transport map $T : M \to M$ given by*

$$T(x) := \exp_x(-\operatorname{grad}\varphi(x)),$$

where $\varphi = \frac{1}{2}\phi$ *for some optimal Kantorovich potential* ϕ*.*

We note that the factors $\frac{1}{2}$ appear in the above statements (but not in the original statements) because the original ones use the cost function $c(x, y) = \frac{1}{2}d^2(x, y)$. Here we only provide the proof for Brenier's theorem to give an intuition how the quadratic cost function play a central role in the argument. For the proof of McCann's theorem, we refer to [McC01, Theorem 8 and 9].

Proof of Brenier's theorem. In view of Theorem 1.10, there exists an optimal transport plan $\pi \in \Pi(\mu, \nu)$ and an optimal Kantorovich potential φ. Let $(x_0, y_0) \in \operatorname{supp}(\pi)$. By complementary slackness theorem, the map $x \mapsto \phi(x) - c(x, y_0)$ assumes the maximum value of $-\phi^c(y_0)$ at $x = x_0$. This implies that the gradient of this map vanishes at x_0:

$$\operatorname{grad}(\phi(\cdot) - c(\cdot, y_0))(x_0) = 0.$$

With the cost function $c(x, y) = h(x - y)$, for some strictly convex function h, we have

$$\operatorname{grad}\phi(x_0) = \operatorname{grad}h(x_0 - y_0).$$

Moreover, $\operatorname{grad}h$ is invertible, so

$$x_0 - y_0 = (\operatorname{grad}h)^{-1}(\operatorname{grad}\phi(x_0)).$$

In the particular case of the quadratic cost $c(x, y) = h(x-y) = |x-y|^2$, its gradient satisfies $\operatorname{grad}h(x) = 2x$ and its inverse is simply $(\operatorname{grad}h)^{-1}(x) = \frac{1}{2}x$. Therefore,

$$y_0 = x_0 - \frac{1}{2}\operatorname{grad}\phi(x_0).$$

□

Corollary 1.25. *Let* M *be a connected and closed manifold and let* $\mu, \nu \in \mathscr{P}_2(M)$ *such that* $\mu \ll \mathrm{vol}$*. Let* $T : M \to M$ *be the unique optimal transport map from* μ *to* ν *given as in McCann's theorem by* $T(x) = \exp_x(-\operatorname{grad}\phi(x))$ *for some function* ϕ*. For* $t \in [0, 1]$*, define the map* $T_t : M \to M$ *by*

$$T_t(x) := \exp_x(-t\operatorname{grad}\phi(x)),$$

and define $\mu_t := (T_t)_\#\mu \in \mathscr{P}_2(M)$ *(so* $\mu_0 = \mu$ *and* $\mu_1 = \nu$*). Then the following two statements hold.*

(a) $\operatorname{supp}(\mu_t) \subset Z_t(\operatorname{supp}(\mu), \operatorname{supp}(\nu))$ *where*

$$Z_t(X, Y) := \left\{ z \in M \,\middle|\, \begin{array}{l} d(x, z) = td(x, y) \text{ and } d(z, y) = (1-t)d(x, y) \\ \text{for some } x \in X, y \in Y \end{array} \right\}.$$

(b) $(\mu_t)_{t\in[0,1]}$ *is a constant speed geodesic in* $(\mathscr{P}_2(M), W_2)$ *from* μ *to* ν.

Proof. As discussed in the proof of [McC01, Lemma 7], $t \mapsto T_t(x)$ is a minimal geodesic for all x where ϕ is differentiable (that is μ-a.e.). Therefore, for such x, $d(T_s(x), T_t(x)) = |t-s| d(x, T(x))$.

In view of Proposition 1.5, we have for $0 \le s \le t \le 1$,

$$\begin{aligned} W_2(\mu_s, \mu_t)^2 &\le \int_M d^2(T_s(x), T_t(x)) d\mu(x) \\ &= \int_M (t-s)^2 d^2(x, T(x)) d\mu(x) = (t-s)^2 W_2(\mu, \nu)^2. \end{aligned}$$

Therefore, $(\mu_t)_{t\in[0,1]}$ is a constant speed geodesic in $(\mathscr{P}_2(M), W_2)$ from μ to ν due to Remark 1.18. □

Furthermore, the absolute continuity of $\mu = \mu_0$ (with respect to vol) implies that $\mu_t = (T_t)_\# \mu$ is also absolutely continuous for all $t \in [0,1)$, say, with the density $\rho_t := d\mu_t / d\,\mathrm{vol}$. The Jacobian determinant $\det DT_t$ satisfies the following ***Monge-Ampère equation***:

$$\rho(x) = \rho_t(T_t(x)) \det DT_t(x). \tag{1.7}$$

This Monge-Ampère equation is essentially the Jacobian equation for change of variables; for more details see e.g. [San15, Section 1.7.6].

It is important to remark that the uniqueness of optimal transport plans and maps in L^2-Wasserstein spaces is attributable to the convexity of the quadratic cost function. This is not the case for L^1-Wasserstein spaces; in fact, there often are plenty of transport maps which give the same optimal transport cost. For an illustration, let us visit the simplest example of transportation on $\mathbb{R}$.

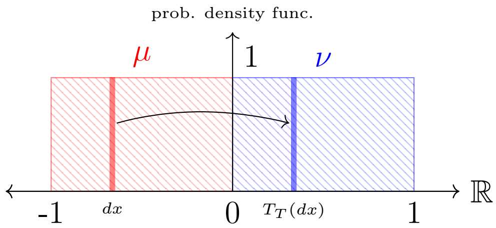
Example 1.26 (linear and quadratic costs in $\mathbb{R}$)**.** Let $\mu = U(-1,0)$ and $\nu = U(0,1)$ be the uniform probability measures on $\mathbb{R}$ whose supports are the real intervals $[-1,0]$ and $[0,1]$. All transport maps $T_\# \mu = \nu$ can have their domain and range restricted to the support of μ and ν, respectively. For a linear cost function $c(x,y) = |x-y|$, the cost of any transport map $T : [-1,0] \to [0,1]$ can be computed as

$$\begin{aligned} \int_{-1}^0 |x - T(x)| d\mu(x) &= \int_{-1}^0 (T(x) - x) d\mu(x) \\ &= \int_0^1 y d\nu(y) - \int_{-1}^0 x d\mu(x) = \frac{1}{2} - \left(\frac{-1}{2}\right) = 1, \end{aligned}$$

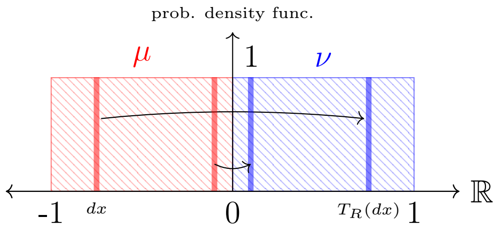
which means all such maps T have the same (and hence optimal) linear cost. On the other hand, the quadratic cost of T satisfies, by Cauchy-Schwarz inequality,

$$\int_{-1}^{0} |x - T(x)|^2 d\mu(x) \geq \frac{(\int_{-1}^{0} |x - T(x)| d\mu(x))^2}{\int_{-1}^{0} 1 d\mu(x)} = 1,$$

and its infimum is attained if and only if $|x - T(x)|$ is constant, i.e., it is the translation map $T_T(x) = x + 1$. In comparison, the reflection map $T_R(x) = -x$ has its quadratic cost equal to $\int_{-1}^{0} |x - (-x)|^2 dx = \frac{4}{3} > 1$. See Figure 1.1 below.



(a) Transport by translation $T_T(x) = x + 1$.



(b) Transport by reflection $T_R(x) = -x$.

Figure 1.1: Comparison of two transport maps between uniform distributions $U(-1, 0)$ and $U(0, 1)$. The translation map has linear and quadratic cost of 1. The reflection map has linear cost of 1 and quadratic cost of $\frac{4}{3}$.

Chapter 2

Otto's calculus

In his booklet [Ott01], Otto develops a framework in which the Wasserstein space is studied via a formal Riemannian calculus (where the word "formal" means that arguments are not completely rigorous, especially due to the concerned Riemannian manifold having infinite dimension). A curve in the Wasserstein space is described by a differential equation called the continuity equation, and the Wasserstein distance is the length of a minimizing curve. In this chapter, we explain Otto's calculus in a formal viewpoint of Riemannian geometry. Otto's calculus will reappear in Part IV of this thesis as an important concept to the definition of Erbar-Maas curvature.

2.1 Wasserstein space as "Riemannian manifold"

For simplicity, we avoid the discussion with specific regularity and often assume smoothness in most of the setting. Throughout this chapter, we let $(M^n, \langle\cdot,\cdot\rangle)$ be a connected complete Riemannian manifold, and we restrict the Wasserstein space of our interest to the space $\mathscr{P}^\infty(M)$ consisting of all smooth positive density functions ρ of an absolutely continuous measure μ with respect to the volume measure (that is, $\rho = d\mu/d\,\mathrm{vol}$):

$$\mathscr{P}^\infty(M) := \left\{ \rho \in C^\infty(M) \,\middle|\, \rho > 0,\ \int_M \rho = 1 \right\},$$

where we identify the distance $W_2(\rho_1, \rho_2) := W_2(\mu_1, \mu_2)$ for $d\mu_i = \rho_i d\,\mathrm{vol}$. The space $\mathscr{P}^\infty(M)$ can be regarded as an ∞-dimensional manifold $\mathcal{M}$, and the tangent space at $\rho \in \mathscr{P}^\infty(M)$ is given by

$$T_\rho \mathscr{P}^\infty(M) = \left\{ (\rho, s) \,\middle|\, s \in C^\infty(M) \text{ with } \int_M s = 0 \right\},$$

where the curve which passes through ρ and represents s is $\rho(t) = \rho_t = \rho + ts \in \mathscr{P}^\infty(M)$ for small enough $|t|$. The key idea in Otto's calculus is to identify this tangent space by another space, namely

$$Tan_\rho \mathscr{P}^\infty(M) = \left\{ (\rho, \operatorname{grad} \phi) \,\middle|\, \phi \in C^\infty(M) \right\},$$

through the identification map $Tan_\rho \mathscr{P}^\infty(M) \to T_\rho \mathscr{P}^\infty(M)$ given by

$$(\rho, \operatorname{grad} \phi) \longmapsto (\rho, s = -\operatorname{div}(\rho \operatorname{grad} \phi)).$$

Here, we refer to Section 5.1 for basic knowledge of differential operators (such as grad, div, and the Laplacian Δ).

To understand this identification map, note that the geodesic ρ_t can be described by $d\mu_t = \rho_t d\operatorname{vol} = (T_t)_\# \rho d\operatorname{vol}$ where T_t is the interpolation of an optimal map given via McCann's theorem by $T_t(x) = \exp_x(t \operatorname{grad} \phi(x))$ for some function ϕ (when ϕ here is chosen to be the negative of the function ϕ in Corollary 1.25). It suffices to justify at $t = 0$ that

$$\frac{\partial}{\partial t}\Big|_{t=0} (T_t)_\# (\rho d\operatorname{vol}) = -\operatorname{div}(\rho \operatorname{grad} \phi) d\operatorname{vol}.$$

For a smooth test function $\varphi \in C^\infty(M)$ and $x \in M$, we have

$$\frac{\partial}{\partial t}\Big|_{t=0} \varphi(T_t(x)) = \langle \operatorname{grad} \varphi(x), \frac{\partial}{\partial t}\Big|_{t=0} T_t(x) \rangle = \langle \operatorname{grad} \varphi(x), \operatorname{grad} \phi(x) \rangle,$$

and therefore

$$\begin{aligned} \frac{\partial}{\partial t}\Big|_{t=0} \int_M \varphi(T_t(x)) \rho(x) &= \int_M \langle \operatorname{grad} \varphi(x), \rho(x) \operatorname{grad} \phi(x) \rangle \\ &= -\int_M \varphi(x) \operatorname{div}(\rho(x) \operatorname{grad} \phi(x)), \end{aligned}$$

as desired. The above identification $s = -\operatorname{div}(\rho \operatorname{grad} \phi))$ can be summarized as the so-called "continuation equation" for probability measures (or probability densities), which is stated in the following proposition.

Proposition 2.1. *Consider any smooth curve $\gamma : [a, b] \to \mathscr{P}^\infty(M)$. Its tangent vector field γ' along the curve γ can be described as $\gamma'(t) = \gamma'_t = (\rho_t, \operatorname{grad} \phi_t)$ which satisfies the following* ***continuity equation****:*

$$\frac{\partial}{\partial t} \rho_t + \operatorname{div}(\rho_t \operatorname{grad} \phi_t) = 0. \tag{2.1}$$

Definition 2.2. Riemannian metric tensor g on $\mathscr{P}^\infty(M)$ is defined as

$$g_\rho(s_1, s_2) = \int_M \rho\langle \operatorname{grad}\phi_1, \operatorname{grad}\phi_2\rangle, \tag{2.2}$$

where $s_i = -\operatorname{div}(\rho \operatorname{grad}\phi_i)$ for $i = 1, 2$.

In principle, with this metric tensor, one can apply Riemannian geometry framework to the Wasserstein spaces and study on e.g. length of curves (which is to be discussed in the next section), Levi-Civita connection, parallel transport, geodesics, as well as Riemann curvature tensor and sectional curvature; for details we refer to the paper by Lott [Lot08]. For example, the following two theorems assert that the nonnegative sectional curvature of $\mathscr{P}^\infty(M)$ is induced from the underlying manifold M.

Theorem 2.3 ([Ott01])**.** *$\mathscr{P}_2(\mathbb{R}^n)$ has nonnegative sectional curvature. Moreover, it is flat if and only if $n = 1$.*

Theorem 2.4 ([Lot08])**.** *If M has nonnegative sectional curvature, then $\mathscr{P}^\infty(M)$ has a nonnegative sectional curvature.*

2.2 Benamou-Brenier formula

In [BB00], Benamou and Brenier introduce an alternative viewpoint to an optimal transportation problem by reintroducing the time-variable and considering it as a "dynamic problem". It is presented in the paper an influential formulation, later known as ***Benamou-Brenier formula***, which is used for the Wasserstein distance computation. The aim of this section is to discuss the Banamou-Brenier formula in the perspective of formal Riemannian geometry and Otto's calculus.

First, let us recall basic facts from Riemannian geometry about the length of curves. Given a Riemannian manifold $(\mathcal{M}, g)$, the length of a curve $\gamma : [0, 1] \to \mathcal{M}$ is $\ell(\gamma) := \int_0^1 |\gamma'(t)|_g dt$, where the norm is induced from Riemannian metric tensor $|v|_g := g(v, v)^{1/2}$. The distance of two points $p, q \in M$ is

$$d(p, q) = \inf_\gamma \ell(\gamma) = \inf_\gamma \int_0^1 |\gamma'(t)|_g dt,$$

where the infimums are taken over all curves $\gamma : [0, 1] \to \mathcal{M}$ with $\gamma(0) = p$ and $\gamma(1) = q$. Such curves γ attaining the infimum are called ***minimal geodesics***, and they are not unique since the length $\ell(\gamma)$ is unchanged under any reparametrization. However, the distance of two points can also be described by a curve which minimizes its kinetic energy $E(\gamma) := \int_0^1 |\gamma'(t)|_g^2 dt$. Such a minimizing curve (if exists) is indeed a minimal geodesic with constant speed.

Proposition 2.5. *Let $(\mathcal{M}, g)$ be a Riemannian manifold. Given $p, q \in \mathcal{M}$, then*

$$d(p,q) = \inf_{\gamma} \int_0^1 |\gamma'(t)|_g dt = \inf_{\gamma} \left(\int_0^1 |\gamma'(t)|_g^2 dt \right)^{\frac{1}{2}}, \tag{2.3}$$

where the infimums are taken over all curves $\gamma : [0,1] \to \mathcal{M}$ with $\gamma(0) = p$ and $\gamma(1) = q$.

Proof. The Cauchy-Schwarz inequality gives

$$\int_0^1 |\gamma'(t)|_g dt \leq \left(\int_0^1 |\gamma'(t)|_g^2 dt \right)^{\frac{1}{2}} \left(\int_0^1 1 dt \right)^{\frac{1}{2}} = \left(\int_0^1 |\gamma'(t)|_g^2 dt \right)^{\frac{1}{2}},$$

with the equality holds when $|\gamma'(t)|_g$ is constant. Moreover, since the length of curves is unchanged under a reparametrization, the infimum $\inf_\gamma \int_0^1 |\gamma'(t)|_g dt$ can be restricted to those curves γ with constant speed. In the case of constant speed γ, the above inequality holds with equality, which in turn yields the equality in (2.3). □

In view of Otto's calculus, one can apply the above description of distance via minimizing energy to the Wasserstein space $\mathscr{P}^\infty(M)$ in order to recover the Wasserstein distance W_2. A precise statement and its rigorous proof (without formal Riemannian calculus arguments) can be found in original paper by Benamou and Brenier [BB00].

Theorem 2.6 (Benamou-Brenier formula)**.** *For $\rho_0, \rho_1 \in \mathscr{P}^\infty(M)$,*

$$W_2(\rho_0, \rho_1) = \inf_{(\rho_t, \phi_t) \in CE(\rho_0, \rho_1)} \left\{ \int_0^1 \int_M \rho_t(x) |\operatorname{grad} \phi_t(x)|^2 d \operatorname{vol}(x) dt \right\}^{1/2}, \tag{2.4}$$

where the infimum is taken over $CE(\rho_0, \rho_1)$ the set of all curves $(\rho_t)_{t\in[0,1]} \in \mathscr{P}^\infty(M)$ and $(\psi_t)_{t\in[0,1]} \in C^\infty(M)$ which satisfy the following continuity equation

$$\frac{\partial}{\partial t} \rho_t + \operatorname{div}(\rho_t \operatorname{grad} \psi_t) = 0 \qquad \forall t \in [0,1], \tag{2.5}$$

with boundary condition $\rho_{t=0} = \rho_0$ and $\rho_{t=1} = \rho_1$.

2.3 Heat flow as gradient flow of entropy

In his original paper [Ott01], Otto uses a gradient flow structure to describe the heat equation. Let us first recall the definition of gradient flow for an abstract Riemannian manifold $(\mathcal{M}, g)$.

Definition 2.7 (gradient flow)**.** Let $(\mathcal{M}, g)$ be a Riemannian manifold, and $E : \mathcal{M} \to \mathbb{R}$ be a function(al) on $\mathcal{M}$. For a fixed time-interval $I \subset \mathbb{R}$, a curve $\rho : I \to \mathcal{M}$ (written as $\rho_t := \rho(t)$), is called an ***integral curve of gradient flow of*** E if the following equation holds true

$$\frac{\partial}{\partial t}\rho_t = -(\operatorname{grad} E)\rho_t, \tag{2.6}$$

or in other words,

$$g_{\rho_t}(\frac{\partial}{\partial t}\rho_t, s) = -DE(\rho_t)(s), \qquad \forall s \in T_{\rho(t)}\mathcal{M} \tag{2.7}$$

Definition 2.8 (entropy)**.** For a complete Riemannian manifold $(M^n, \langle\cdot,\cdot\rangle)$, an entropy functional (with respect to volume measure) $\mathcal{E}_{\text{vol}} : \mathscr{P}(M) \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$\mathcal{E}_{\text{vol}}(\mu) := \begin{cases} \int_M \rho \log \rho & \text{if } \mu \ll \text{vol with } \rho = \frac{d\mu}{d\,\text{vol}}, \\ +\infty & \text{otherwise.} \end{cases}$$

For simplicity in this chapter, we define the entropy for the space of densities $\mathscr{P}^\infty(M)$ as

$$\mathcal{E}_{\text{vol}}(\rho) := \int_M \rho \log \rho$$

for all $\rho \in \mathscr{P}^\infty(M)$.

Definition 2.9 (heat flow)**.** For a fixed time-interval $I \subset \mathbb{R}$, a curve $u : I \to \mathcal{M}$, is called an ***integral curve of heat flow*** if it is a solution of the heat equation: $\frac{\partial}{\partial t}u_t = \Delta u_t$, where $\Delta := \operatorname{div}(\operatorname{grad})$.

We refer to Section 5.5 for details about the heat equation. Now we give a formal proof of the following theorem that the gradient flow of the entropy is in fact the heat flow.

Theorem 2.10. *An integral curve of gradient flow of the entropy functional* $\mathcal{E}_{\text{vol}}$ *on the Riemannian manifold* $(\mathcal{M}, g) = (\mathscr{P}^\infty(M), W_2)$ *is a solution to the heat equation* $\frac{\partial}{\partial t}\rho_t = \Delta\rho_t$.

Proof. Let ρ_t be the integral curve of gradient flow of $\mathcal{E}_{\text{vol}}$. For a fixed $t \in I$ and fixed $s \in T_{\rho_t}\mathcal{M}$, we first compute $D\mathcal{E}_{\text{vol}}(\rho_t)(s)$ through a curve $c : (-\varepsilon, \varepsilon) \to \mathcal{M}$ such that $c(0) = \rho_t$ and $c'(0) = s$. We have

$$D\mathcal{E}_{\text{vol}}(\rho_t)(s) = \frac{d}{d\tau}\Big|_{\tau=0} \mathcal{E}_{\text{vol}}(c(\tau)) = \int_M (1 + \log c(0))c'(0) = \int_M (1 + \log \rho_t)s.$$

With the identification $s = -\operatorname{div}(\rho_t \operatorname{grad}\phi)$, we have

$$\begin{aligned} g_{\rho_t}(\frac{\partial}{\partial t}\rho_t, s) = -D\mathcal{E}_{\mathrm{vol}}(\rho_t)(s) &= \int_M (1+\log\rho_t)\operatorname{div}(\rho_t \operatorname{grad}\phi) \\ &= -\int_M \rho_t \langle \operatorname{grad}(1+\log\rho_t), \operatorname{grad}\phi\rangle \\ &= -\int_M \rho_t \langle \operatorname{grad}(\log\rho_t), \operatorname{grad}\phi\rangle. \end{aligned}$$

Since the above equality holds for all $s \in T_{\rho_t}\mathcal{M}$, in view of (2.2) we deduce that

$$\frac{\partial}{\partial t}\rho_t = \operatorname{div}(\rho_t \operatorname{grad}(\log\rho_t)) = \operatorname{div}(\operatorname{grad}(\rho_t)) = \Delta(\rho_t).$$

□

2.4 Jordan-Kinderlehrer-Otto theorem

Theorem 2.10 was proved in a rigorous manner by Jordan-Kinderlehrer-Otto [JKO98] (and hence referred to as ***JKO theorem***), where they consider the following discrete-time minimizing scheme. Fix a functional $E : \mathcal{M} \to \mathbb{R}$ and fix $\rho_0 \in \mathcal{M}$. For a given time step size $h > 0$, let $\rho^{(0)} = \rho_0$ and

$$\rho^{(k)} = \operatorname{argmin}_{\rho\in\mathcal{M}} \left(\frac{1}{2h} W_2(\rho^{(k-1)}, \rho) + E(\rho) \right). \tag{2.8}$$

The theorem states that the step functions $\rho_h(t) := \rho^{(k)}$ for $t \in [kh, (k+1)h)$ converges as $h \to 0$ to the solution of the heat equation $\frac{\partial}{\partial t}\rho = \Delta\rho$ with $\rho_{t=0} = \rho_0$.

Here we give a heuristic argument why this minimizing scheme is indeed the discretization of the integral curve. We explain this in a general setting of $(\mathcal{M}, g)$ with the distance function d induced from g and a functional $F : \mathcal{M} \to \mathbb{R}$.

For $x \in \mathcal{M}$, consider $f_h(y) := \frac{1}{2h} d_x(y)^2 + F(y)$, where $d_x(\cdot)$ is the distance from x. Let x_h solve $\min_y f_h(y)$. As $h \to 0$, we know that $x_h \to x$. Now consider a geodesic curve $c : (-\varepsilon, \varepsilon) \to \mathcal{M}$ such that $c(0) = x$ and $c(h) = x_h$. We would like to justify that

$$c'(0) = -\operatorname{grad} F(x).$$

Consider this curve c via polar coordinates: $c(t) = \exp_x(r(t)v(t))$ where $r(t) \in \mathbb{R}_{\geq 0}$ and $v(t) \in T_xM$ is a unit tangent vector. Consider another curve $\gamma(s) := \exp_x(s \cdot v(t))$, so $\operatorname{grad} d_x(c(t)) = \gamma'(r(t)) = D\exp_x(r(t)v(t))(v(t))$.

The assumption that $c(h) = x_h$ minimizes $f_h(\cdot)$ means

$$0 = \operatorname{grad} f_h(y)|_{y=c(h)} = \frac{1}{h} d_x(c(h)) \cdot \operatorname{grad} d_x(c(h)) + \operatorname{grad} F(c(h)).$$

Passing $h \to 0$ and note that $d_x(c(t)) = r(t)$, we have

$$\begin{aligned} -\operatorname{grad} F(x) &= \lim_{h \downarrow 0} \frac{1}{h} r(h) \operatorname{grad} d_x(c(h)) \\ &= \lim_{h \downarrow 0} \frac{r(h)}{h} D \exp_x(r(h)v(h))(v(h)) \\ &= r'(0) \underbrace{D \exp_x(0_x)}_{=\mathrm{Id}}(v(0)) = r'(0)v(0). \end{aligned}$$

On the other hand, $c'(0) = D \exp_x(0_x) \left(r'(0)v(0) + r(0)v'(0) \right) = r'(0)v(0)$, and thus $c'(0) = -\operatorname{grad} F(x)$ as desired.

Chapter 3

Ricci curvature on metric measure spaces

This chapter features a celebrated result from two independent papers by Lott-Villani [LV09] and Sturm [Stu06] that a weighted Riemannian manifold M has Ricci curvature bounded below by $K \in \mathbb{R}$ if and only if the entropy functional on the Wasserstein space $(\mathscr{P}_2(M), W_2)$ is "K-convex" along W_2-geodesics. The latter part of this equivalence does not require differential structure, and hence it allows them to define on metric measure spaces a synthetic notion of Ricci curvature, which is often referred to as ***Lott-Sturm-Villani (LSV) curvature***.

Here we only discuss this main result in the case of nonweighted Riemannian manifold, which was proved earlier by von Renesse and Sturm [vRS05]. The case of weighted manifolds is postponed to Chapter 6.

3.1 Ricci lower curvature bound via displacement convexity

Let us start with the definitions of the relative entropy and the displacement convexity on metric measure spaces.

Definition 3.1 (relative entropy)**.** Let (X, d, m) be a metric measure space with a reference measure $m \in \mathscr{P}(X)$. The ***relative entropy*** (with respect to m) is defined on $\mathscr{P}(X)$ as

$$\mathcal{E}_m(\mu) := \begin{cases} \int_X \rho \log \rho dm = \int_X \log \rho d\mu & \text{if } \mu \ll m \text{ with } \rho = \frac{d\mu}{dm}, \\ +\infty & \text{otherwise.} \end{cases} \tag{3.1}$$

Definition 3.2 (displacement convexity)**.** Let $K \in \mathbb{R}$ and let (X, d) be a metric space. A functional $E : \mathscr{P}_2(X) \to X$ is said to be ***displacement K-convex*** if for any $\mu_0, \mu_1 \in \mathscr{P}_2(X)$, there exists a constant speed geodesic $(\mu_t)_{t\in[0,1]}$ in $(\mathscr{P}_2(X), W_2)$ from μ_0, μ_1 satisfies

$$E(\mu_t) \le (1-t)E(\mu_0) + tE(\mu_1) - \frac{1}{2}Kt(1-t)W_2(\mu_0, \mu_1)^2. \tag{3.2}$$

The word "displacement" here indicates that the curve $(\mu_t)_{t\in[0,1]}$ in consideration is a geodesic with respect to the W_2-metric, and it is not the linear interpolation $\mu_t = (1-t)\mu_0 + t\mu_1$. Let us also remark that for the above definition to be valid, the Wasserstein space $(\mathscr{P}_2(X), W_2)$ must be a geodesic space. In view of Theorem 1.20, it is sensible to assume (X, d) to be a geodesic space as well.

Now we are ready to state the main result which characterizes the Ricci lower bound via the displacement convexity of the entropy functional.

Theorem 3.3 (main theorem, [vRS05])**.** *Given a smooth complete Riemannian manifold (M, g) and any $K \in \mathbb{R}$, the following properties are equivalent:*

(i) $\mathrm{Ric}(M) \ge K$*, that is,* $\mathrm{Ric}_x(v, v) \ge K|v|^2$ *for all* $x \in M$*,* $v \in T_xM$*.*

(ii) The entropy $\mathcal{E}_{\mathrm{vol}}(\cdot)$ *is displacement* K*-convex on* $\mathscr{P}_2(M)$*.*

Here and henceforth, we use $\mathrm{Ric}(M)$ as the convention for the greatest lower bound of the Ricci curvature on M:

$$\mathrm{Ric}(M) := \inf\left\{\frac{\mathrm{Ric}_x(v, v)}{|v|^2} : x \in M, v \in T_xM\right\}.$$

The main theorem gives rise to the following Ricci curvature notion by Lott-Sturm-Villani on metric measure spaces.

Definition 3.4 (Lott-Sturm-Villani)**.** A metric measure space (X, d, m) is said to have ***Ricci curvature bound below by*** $K \in \mathbb{R}$, denoted as $\mathrm{Ric}_\infty(X) \ge K$, if the relative entropy $\mathcal{E}_m(\cdot)$ is displacement K-convex, that is, for any $\mu_0, \mu_1 \in \mathscr{P}_2(X)$, there exists a constant speed geodesic $(\mu_t)_{t\in[0,1]}$ in $\mathscr{P}_2(X)$ from μ_0 to μ_1 such that

$$\mathcal{E}_m(\mu_t) \le (1-t)\mathcal{E}_m(\mu_0) + t\mathcal{E}_m(\mu_1) - \frac{1}{2}Kt(1-t)W_2(\mu_0, \mu_1)^2 \tag{3.3}$$

holds true for all $t \in [0, 1]$.

Remark 3.5 ($CD(K,N)$ spaces). Such a metric measure space (X,d,m) is usually referred to as a $CD(K,\infty)$ space, where CD stands for ***curvature-dimension***. A more general convention of $CD(K,N)$, where $N\in(0,\infty]$, means K is a lower bound of Ricci curvature, and N is "an upper bound of the dimension" (in case X is a manifold). The definition of $CD(K,N)$ spaces where the $N<\infty$ requires the Rényi entropy $S_N(\mu) := -\int_X \rho^{1-1/N} dm$ to satisfy a more complicated K-displacement-convex inequality than (3.2). However, the case $N<\infty$ is not our focus, and we refer readers to a concise summary by Ohta [Oht14, Section 7.4.2] and references therein.

We come back to a proof sketch of the main theorem, which we follow from [Vil09]. Here we only prove the forward implication, and we ignore all regularity concerns. For readers' interest, we will present in Chapter 8 different ideas of proof from the original papers [vRS05] and Cordero-Erausquin, McCann, and Schmuckenschläger's [CEMS01] (the latter of which proves in the case of $\mathrm{Ric}(M)\geq 0$).

Proof sketch of Theorem 3.3. First suppose that $\mu_0,\mu_1 \ll \mathrm{vol}$ otherwise the right-hand-side of (3.3) is infinity. By Theorem 1.24, there exists a unique geodesic $(\mu_t)_{t\in[0,1]}$ in $\mathscr{P}_2(M)$ connecting μ_0 and μ_1, namely $\mu_t = (T_t)_\#\mu_0$ where $T_t(x) = \exp_x(-t\,\mathrm{grad}\,\phi(x))$ for some $\phi: M\to\mathbb{R}$. From the viewpoint that $t\mapsto T_t(x)$ is a minimal constant speed geodesic (see Corollary 1.25), we may denote this geodesic by $\gamma_x:[0,1]\to M$, that is,

$$\gamma_x(t) := T_t(x).$$

For all t, $\mu_t \ll \mathrm{vol}$, so it can be written as $d\mu_t(x) = \rho_t(x)d\,\mathrm{vol}(x)$ with a finite probability density $\rho_t(x)$. Moreover, $J_x(t) := DT_t(x)$ satisfies the Monge-Ampère equation (1.7):

$$\rho_0(x) = \rho_t(F_t(x))\det J_x(t). \tag{3.4}$$

The pushforward relation and Monge-Ampère equation give

$$\begin{aligned}\mathcal{E}_{\mathrm{vol}}(\mu_t) = \int_M \log\rho_t(x)d\mu_t(x) &= \int_M \log\left(\frac{\rho_0(x)}{\det J_x(t)}\right)d\mu_0(x)\\ &= \mathcal{E}_{\mathrm{vol}}(\mu_0) - \int_M \log(\det J_x(t))d\mu_0(x).\end{aligned} \tag{3.5}$$

To derive the second derivatives of $\mathcal{E}_{\mathrm{vol}}(\mu_t)$, we compute that of the term $\log(\det J_x(t))$ for a fixed x as follows:

$$\begin{aligned}\frac{d}{dt}\log(\det J_x(t)) &= \frac{1}{\det J_x(t)}\cdot\det J_x(t)\,\mathrm{tr}\left(J_x'(t)J_x^{-1}(t)\right)\\ &= \mathrm{tr}(J_x'(t)J_x^{-1}(t)),\end{aligned}$$

and

$$\begin{aligned}\frac{d^2}{dt^2}\log(\det J_x(t)) &= \operatorname{tr}\left(J_x''(t)J_x^{-1}(t) - (J_x'(t)J_x^{-1}(t))^2\right)\\ &\leq \operatorname{tr}\left(J_x''(t)J_x^{-1}(t)\right). \qquad (3.6)\end{aligned}$$

Remark that the derivative of an exponential map is a matrix consisting of Jacobi fields, that is, for any $u, v \in T_xM$, by considering the geodesic variation $F(s,t) := \exp_x(su+tv)$, we infer that $\frac{\partial}{\partial s}|_{s=0}F(s,t) = D\exp_x(tv)(u)$ is a Jacobi field. Here, it means for any fixed $x \in M$, $J_x(t) = DT_t(x)$ is a matrix whose entries are Jacobi fields along $\gamma_x(t)$, so it satisfies the following Jacobi equation:

$$J_x''(t) + A_x(t)J_x(t) = 0,$$

where the map $A_x(t) : T_{\gamma_x(t)}M \to T_{\gamma_x(t)}M$ is given by

$$A_x(t)(v) := R(\gamma_x'(t), v)\gamma_x'(t),$$

and R is Riemann curvature tensor. Moreover, the trace of $A_t(x)$ can be expressed in terms of Ricci curvature:

$$-\operatorname{tr}\left(J_x''(t)J_x^{-1}(t)\right) = \operatorname{tr}(A_x(t)) = \operatorname{Ric}_{\gamma_x(t)}(\gamma_x'(t), \gamma_x'(t)). \qquad (3.7)$$

The Ricci curvature assumption $\operatorname{Ric}(M) \geq K$ implies $\operatorname{Ric}_{\gamma_x(t)}(\gamma_x'(t), \gamma_x'(t)) \geq K|\gamma_x'(t)|$. Furthermore, since the curve $\gamma_x(t) = T_t(x)$ is a minimal constant speed geodesic, we have $|\gamma_x'(t)| = d(x, T_1(x))$ for all $t \in [0,1]$. It then follows from (3.5), (3.6) and (3.7) that

$$\frac{d^2}{dt^2}\mathcal{E}_{\mathrm{vol}}(\mu_t) \geq \int_M K d(x, T_1(x))d\mu_0(x) = KW_2(\mu_0, \mu_1)^2.$$

Let $\tilde{E}(t) := \mathcal{E}_{\mathrm{vol}}(\mu_t) + \frac{1}{2}Kt(1-t)W_2(\mu_0, \nu_1)^2$. Note that $\tilde{E}(0) = \mathcal{E}_{\mathrm{vol}}(\mu_0)$, $\tilde{E}(1) = \mathcal{E}_{\mathrm{vol}}(\mu_1)$ and

$$\frac{d^2}{dt^2}\tilde{E}(t) = \frac{d^2}{dt^2}\mathcal{E}_{\mathrm{vol}}(\mu_t) - KW_2(\mu_0, \mu_1)^2 \geq 0.$$

Therefore $\tilde{E}(t)$ is a convex function and it satisfies $\tilde{E}(t) \leq (1-t)\tilde{E}(0) + t\tilde{E}(1)$, which means

$$\mathcal{E}_{\mathrm{vol}}(\mu_t) \leq (1-t)\mathcal{E}_{\mathrm{vol}}(\mu_0) + t\mathcal{E}_{\mathrm{vol}}(\mu_1) - \frac{1}{2}Kt(1-t)W_2(\mu_0, \mu_1)^2.$$

□

Example 3.6 (Euclidean space $\mathbb{R}^n$)**.** For all $n \in \mathbb{N}$, the Euclidean space $\mathbb{R}^n$ has zero (sectional and Ricci) curvature everywhere, and it indeed satisfies $\mathrm{Ric}(\mathbb{R}^n) \geq 0$. To see that the relative entropy $\mathcal{E}_{\mathrm{vol}}$ (with respect to the Lebesgue/volume measure) is displacement 0-convex, one can argue similarly to the proof given above. From Brenier's theorem (and Corollary 1.25), a constant speed W_2-geodesic $(\mu_t)_{t\in[0,1]}$ can be described as $\mu_t = (T_t)_\# \mu_0$, where T_t is the interpolation of the optimal transport map given in an explicit form of $T_t(x) = x - t\nabla\phi(x)$, for some ϕ. The Jacobian matrix is $J_x(t) = DT_t(x) = \mathrm{Id} - t \operatorname{Hess}\phi(x)$, so its second derivative in t vanishes (or alternatively, one may see $J_x''(t) = 0$ from the Jacobi equation with zero Riemann curvature tensor). It is then implies by (3.6) that $\log(\det J_x(t))$ is concave in t, and by (3.5) that $\mathcal{E}_{\mathrm{vol}}(\mu_t)$ is convex, i.e., $\mathcal{E}_{\mathrm{vol}}(\mu_t) \leq (1-t)\mathcal{E}_{\mathrm{vol}}(\mu_0) + t\mathcal{E}_{\mathrm{vol}}(\mu_1)$.

One consequence of the displacement convexity is to provide an alternative proof of the following variation of Brunn-Minkowski inequality on $\mathbb{R}^n$ (see also [Vil09, Theorem 18.5]).

Theorem 3.7 (Brunn-Minkowski)**.** *Let* vol *be the Lebesgue measure of* $\mathbb{R}^n$*. Then for compact subsets* $A_0, A_1 \subset \mathbb{R}^n$*, the following inequality holds*

$$\mathrm{vol}((1-t)A_0 + tA_1) \geq \mathrm{vol}(A_0)^{1-t}\,\mathrm{vol}(A_1)^t.$$

Proof. For each $\in \{0,1\}$, let μ_i be the uniform probability distribution whose support is A_i, that is, $\mu_i(E) = \frac{\mathrm{vol}(E\cap A_i)}{\mathrm{vol}(A_i)}$ for any measurable set $E \subset \mathbb{R}^n$. Its density function satisfies $\rho_i := \frac{d\mu_i}{d\,\mathrm{vol}} = \frac{1}{\mathrm{vol}(A_i)}\mathbb{1}_{A_i}$, and the entropy can be computed as

$$\mathcal{E}_{\mathrm{vol}}(\mu_i) = \int_{\mathbb{R}^n} \log\rho_i d\mu_i = \int_{A_i} \log\left(\frac{1}{\mathrm{vol}(A_i)}\right) d\mu_0 = -\log(\mathrm{vol}(A_i)).$$

Let $\mu_t = (T_t)_\#\mu_0$ be the constant speed W_2-geodesic from μ_0 to μ_1, obtained from Corollary 1.25. For $t \in (0,1)$, let $A_t := \mathrm{supp}(\mu_t)$. Then Jensen's inequality applied to the convex function $s \mapsto s\log s$ yields

$$\begin{aligned}
\mathcal{E}_{\mathrm{vol}}(\mu_t) &= \int_{\mathbb{R}^n} \rho_t \log\rho_t d\,\mathrm{vol} \\
&= \mathrm{vol}(A_t)\int_{A_t} \rho_t\log\rho_t \frac{d\,\mathrm{vol}}{\mathrm{vol}(A_t)} \\
&\geq \mathrm{vol}(A_t)\left(\int_{A_t}\rho_t\frac{d\,\mathrm{vol}}{\mathrm{vol}(A_t)}\right)\log\left(\int_{A_t}\rho_t\frac{d\,\mathrm{vol}}{\mathrm{vol}(A_t)}\right) \\
&= \mathrm{vol}(A_t)\cdot\frac{1}{\mathrm{vol}(A_t)}\cdot\log(\frac{1}{\mathrm{vol}(A_t)}) = -\log(\mathrm{vol}(A_t)).
\end{aligned}$$

Since $\mathcal{E}_{\mathrm{vol}}$ is displacement 0-convex, we have

$$\begin{aligned}-\log(\mathrm{vol}(A_t)) \leq \mathcal{E}_{\mathrm{vol}}(\mu_t) &\leq (1-t)\mathcal{E}_{\mathrm{vol}}(\mu_0) + t\mathcal{E}_{\mathrm{vol}}(\mu_1)\\ &\leq -(1-t)\log(\mathrm{vol}(A_0)) - t\log(\mathrm{vol}(A_1)),\end{aligned}$$

which implies $\mathrm{vol}(A_t) \geq \mathrm{vol}(A_0)^{1-t}\,\mathrm{vol}(A_1)^t$.

Finally we note that $\mathrm{vol}((1-t)A_0 + tA_1) \geq \mathrm{vol}(A_t)$; this is due to that fact that $\mathrm{supp}(A_t)$ lies inside $((1-t)A_0+tA_1)$ as $\mu_t = (T_t)_{\#}\mu_0$ where $t \mapsto T_t(x)$ is a constant speed geodesic for all x. □

Chapter 4

Bonnet-Myers theorem and diameter rigidity

The Bonnet-Myers diameter bound theorem [Mye41] is a classical theorem in Riemannian geometry. It states that a connected complete manifold (M^n, g) with Ricci curvature bounded from below by $K > 0$ is a compact manifold, and its diameter is bounded from above by the diameter of the round sphere of dimension n with Ricci curvature K. Furthermore, there is a rigidity theorem, known as Cheng's rigidity [Che75], which asserts that the diameter upper bound is attained if and only if M^n is isometric to an n-dimensional sphere.

The Bonnet-Myers diameter bound theorem is the main theme of this thesis, and its analogous result will be reoccurring in the context of every curvature notion for discrete spaces.

Theorem 4.1 (Bonnet-Myers)**.** *Let (M^n, g) be a connected and complete Riemannian manifold. Suppose there exists $r > 0$ such that the Ricci curvature satisfies*

$$\mathrm{Ric}_x(v) \geq \frac{n-1}{r^2} > 0$$

for all $x \in M$ and unit vector $v \in T_xM$. Then M is compact and its diameter is bounded from above by $\mathrm{diam}(M) \leq \pi r$*.*

Equivalently, the diameter bound can be written as

$$\mathrm{diam}(M) \leq \pi\sqrt{\frac{n-1}{K}},$$

given that $K := \mathrm{Ric}(M) > 0$. Here we provide a standard proof of this theorem via the second variation formula of the energy (see e.g. the book by do Carmo [dC92]).

Proof. Let $x, y \in M$ be two arbitrary points, and let $a := d(x, y)$. By Hopf-Rinow theorem, the completeness of M implies that there exists a minimal constant-speed geodesic $c : [0, a] \to M$ such that $c(0) = x$, $c(a) = y$, $|c'(t)| = 1$ for all $t \in [0, a]$. It suffices to prove that $a \leq \pi r$. The compactness of M would then follow from M being complete and bounded .

First, we construct proper variations of c as follows. Choose unit vectors $e_i \in T_x M$ such that $\{e_1, e_2, ..., e_{n-1}, c'(0)\}$ forms an orthonormal basis of $T_x M$. For each $1 \leq i \leq n-1$, let V_i be a parallel vector field along c such that $V_i(0) = e_i$. Note that

$$\frac{d}{dt}\langle V_i(t), V_j(t)\rangle = \langle \frac{D}{dt}V_i(t), V_j(t)\rangle + \langle V_i(t), \frac{D}{dt}V_j(t)\rangle = 0$$

since $\frac{D}{dt}V_i(t) = \frac{D}{dt}V_j(t) = 0$ from being parallel. It means $\langle V_i(t), V_j(t)\rangle$ is constant and $\langle V_i(t), V_j(t)\rangle = \langle e_i, e_j\rangle = \delta_{ij}$.

For each i, define $X_i(t) := \sin\left(\frac{\pi t}{a}\right) V_i(t)$, and let $F_i : (-\varepsilon, \varepsilon) \times [0, a] \to M$ be a variation of c whose variational vector field is X_i, that is

$$F_i(0, t) = c(t) \qquad \text{and} \qquad \frac{\partial}{\partial s}F_i(s, t) = X_i(t).$$

Since $X_i(0) = X_i(a) = 0$, it means that for every $s \in (-\varepsilon, \varepsilon)$ the curve $F_i(s, \cdot)$ has the same endpoints as the curve c, that is, F_i is a proper variation of c.

The energy for the curve $F_i(s, \cdot)$ is defined by

$$E_i(s) := \frac{1}{2}\int_0^a \left|\frac{\partial}{\partial t}F_i(s, t)\right|^2 dt.$$

The second variation formula of energy gives

$$\begin{aligned} E_i''(0) &= \int_0^a \left|\frac{D}{dt}X_i(t)\right|^2 - \Big\langle X_i(t), R\Big(c'(t), X_i(t)\Big)c'(t)\Big\rangle dt \\ &= \int_0^a \left|\frac{\pi}{a}\cos\left(\frac{\pi t}{a}\right)V_i(t)\right|^2 - \sin^2\left(\frac{\pi t}{a}\right)\Big\langle V_i(t), R\Big(c'(t), V_i(t)\Big)c'(t)\Big\rangle dt \\ &= \int_0^a \frac{\pi^2}{a^2}\cos^2\left(\frac{\pi t}{a}\right) - \sin^2\left(\frac{\pi t}{a}\right)K(c'(t), V_i(t))dt \qquad (4.1) \end{aligned}$$

where $K(c'(t), V_i(t))$ is sectional curvature of the two-dimensional plane spanned by $c'(t)$ and $V_i(t)$.

Summing (4.1) over i and recalling $\sum_{i=1}^{n-1} K(c'(t), V_i(t)) = \mathrm{Ric}(c'(t)) \geq \frac{n-1}{r^2}$, we obtain

$$\sum_{i=1}^{n-1} E_i''(0) \leq (n-1) \int_0^a \frac{\pi^2}{a^2} \cos^2\left(\frac{\pi t}{a}\right) - \frac{1}{r^2} \sin^2\left(\frac{\pi t}{a}\right) dt$$

$$= (n-1)\left(\frac{\pi^2}{a^2} - \frac{1}{r^2}\right)\frac{a}{2} \tag{4.2}$$

Since c is a minimal constant speed geodesic, $E_i(0) = \inf_{s\in(-\varepsilon,\varepsilon)} E_i(s)$ and thus $E_i''(0) \geq 0$ for all i. Equation (4.2) then implies that $a \leq \pi r$ as desired. □

Theorem 4.2 (Cheng's rigidity [Che75])**.** *A connected and complete Riemannian manifold M with $\mathrm{Ric}(M) \geq K > 0$ and diameter $\mathrm{diam}(M) = \sqrt{\frac{n-1}{K}}$ is isometric to an n-dimensional round sphere S_r^n (where $\mathrm{Ric}(S_r^n) = \frac{n-1}{r^2}$ and $\mathrm{diam}(S_r^n) = \pi r$).*

Remark 4.3. It is crucial to emphasize the necessity of the assumption of the positive lower Ricci curvature bound $\mathrm{Ric}(M) \geq K > 0$ in the Bonnet-Myers theorem. One cannot relax this assumption to "everywhere positive Ricci curvature". As a counterexample, we consider an elliptic paraboloid $x_3 = x_1^2 + x_2^2$. This paraboloid has everywhere positive Gaussian curvature $K = \frac{4}{(1+4x_1^2+4x_2^2)^2}$; however, the infimum of curvature is zero so the Bonnet-Myers theorem is not applicable, and the paraboloid is indeed noncompact.

In [Oht07a], Ohta provides a generalized Bonnet-Myers diameter bound for metric measure spaces, where the lower bound of Ricci curvature is replaced by the Lott-Sturm-Villani curvature-dimension condition $CD(K, N)$ (see a brief discussion of $CD(K, N)$ in Remark 3.5). In his sequential paper [Oht07b], he also provides a necessary condition for the maximal diameter bound (which can be thought of as a weak version of Cheng's rigidity). For a summary of both results, we refer to [Oht14].

Theorem 4.4 (generalized Bonnet-Myers)**.** *Suppose that a metric measure space (X, d, m) satisfies $CD(K, N)$ for some $K > 0$ and $N \in (1, \infty)$. Then*

(i) $\mathrm{diam}(X) \leq \pi\sqrt{(N-1)/K}$.

(ii) Each $x \in X$ has at most one point of distance $\pi\sqrt{(N-1)/K}$ from x.

Theorem 4.5 (maximal diameter)**.** *Suppose (X, d, m) satisfies $CD(K, N)$ for some $K > 0$ and $N \in (1, \infty)$, and $\mathrm{diam}(X) = \pi\sqrt{(N-1)/K}$. Let $x_N, x_S \in X$*

be two points such that $d(x_N, x_S) = \mathrm{diam}(X)$. *Then for all* $z \in X$, *we have* $d(x_N, z) + d(z, x_S) = \mathrm{diam}(X)$. *In particular, there exists a minimal geodesic from* x_N *to* x_S *passing through* z.

If X with the above maximal diameter is assumed to be non-branching (meaning there is no pairs of minimal geodesic $\gamma_1, \gamma_2 : [0, \ell] \to X$ such that $\gamma_1(0) = \gamma_2(0)$ and $\gamma_1(t) = \gamma_2(t)$ for some $t > 0$ but $\gamma_1(\ell) = \gamma_2(\ell)$), then the minimal geodesic from x_N to x_S passing through z is unique. It implies further that X is the spherical suspension of some topological measure space. This maximal diameter and spherical suspension result is not discussed here in details, but we would like to mention that an analogous result will appear later in the context of Ollivier Ricci curvature on graphs.

Chapter 5

Analytic consequences of Ricci curvature: Bakry-Émery theory

In this chapter, we first provide background knowledge about differential operators (gradient, divergence, Laplacian and Hessian) which is necessary in later discussions including Bochner's formula, Laplacian eigenvalues and Lichnerowicz's theorem, and Heat equation and heat kernel. These mentioned topics are building blocks for the Bakry-Émery Theory [BE84] which essentially gives rise to another Ricci curvature notion, known as ***Bakry-Émery curvature*** and related to an important concept about the gradient estimate of the heat flow.

5.1 Differential operators

Let $(M^n, \langle\cdot,\cdot\rangle)$ be a smooth connected complete Riemannian manifold, and let $C^\infty(M)$ and $\mathfrak{X}(M)$ denote the spaces of all smooth real functions and smooth vector fields on M, respectively.

Definition 5.1 (Gradient, divergence and Laplacian).

The ***gradient operator*** $\mathrm{grad} : C^\infty(M) \to \mathfrak{X}(M)$ is defined for any $f \in C^\infty(M)$ uniquely via the following inner product:

$$\langle \mathrm{grad}\, f, X\rangle := Xf \qquad \forall X \in \mathfrak{X}(M).$$

The ***divergence operator*** $\mathrm{div} : \mathfrak{X}(M) \to C^\infty(M)$ is defined for any $X \in \mathfrak{X}(M)$ by

$$(\mathrm{div}\, X)(x) := \mathrm{tr}_{T_xM}(v \mapsto \nabla_v X) \qquad \forall x \in M,$$

where the mapping is from the tangent space T_xM onto itself and ∇ denotes the Levi-Civita connection.

The ***Laplacian operator*** $\Delta : C^\infty(M) \to C^\infty(M)$ is defined to be the composition $\Delta := \operatorname{div}(\operatorname{grad})$.

The product rule for gradient, divergence and Laplacian is given in the following proposition.

Proposition 5.2 (Product rule)**.** *Let $f, h \in C^\infty(M)$ and $X \in \mathfrak{X}(M)$. Then*

(a) $\operatorname{grad}(fh) = f \operatorname{grad} h + h \operatorname{grad} f$

(b) $\operatorname{div}(fX) = \langle \operatorname{grad} f, X\rangle + f \operatorname{div} X$

(c) $\Delta(fh) = 2\langle \operatorname{grad} f, \operatorname{grad} h\rangle + f\Delta h + h\Delta f$

Proof. The product rule for gradient and divergence is induced from the product rule of directional derivative and of Levi-Civita connection, respectively. For Laplacian, we have

$$\begin{aligned}\Delta(fh) &= \operatorname{div}(\operatorname{grad}(fh))\\ &= \operatorname{div}(f \operatorname{grad} h) + \operatorname{div}(h \operatorname{grad} f)\\ &= \langle \operatorname{grad} f, \operatorname{grad} h\rangle + f\operatorname{div}(\operatorname{grad} h) + \langle \operatorname{grad} h, \operatorname{grad} f\rangle + h\operatorname{div}(\operatorname{grad} f)\\ &= 2\langle \operatorname{grad} f, \operatorname{grad} h\rangle + f\Delta h + h\Delta f.\end{aligned}$$

□

Divergence is often referred to as the negative adjoint of gradient as it can be seen from the divergence theorem on a closed manifold M that

$$0 = \int_M \operatorname{div}(fX) d\operatorname{vol} = \int_M \langle \operatorname{grad} f, X\rangle d\operatorname{vol} + \int_M f \operatorname{div} X d\operatorname{vol} \tag{5.1}$$

Definition 5.3 (Hessian)**.** The Hessian tensor of $f \in C^\infty(M)$ is a bilinear form defined as

$$\operatorname{Hess} f(X, Y) := \langle \nabla_X \operatorname{grad} f, Y\rangle$$

for any $X, Y \in \mathfrak{X}(M)$.

Recall the Riemannian property: $X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$. Hessian can then be reformulated as

$$\operatorname{Hess} f(X, Y) = \langle \nabla_X \operatorname{grad} f, Y\rangle = X(Yf) - \langle \operatorname{grad} f, \nabla_X Y\rangle.$$

A fundamental property of the Hessian is the symmetry property.

Proposition 5.4. $\operatorname{Hess} f(X,Y) = \operatorname{Hess} f(Y,X)$

Proof. Recall the Riemannian property: $X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$. We derive

$$\begin{aligned}\operatorname{Hess}(f)(X,Y) &= \langle \nabla_X \operatorname{grad} f, Y\rangle \\ &= X\langle \operatorname{grad} f, Y\rangle - \langle \operatorname{grad} f, \nabla_X Y\rangle \\ &= X(Yf) - \langle \operatorname{grad} f, \nabla_X Y\rangle,\end{aligned}$$

and similarly, we have

$$\operatorname{Hess}(f)(Y,X) = Y(Xf) - \langle \operatorname{grad} f, \nabla_Y X\rangle,$$

Since $\operatorname{Hess} f(X,Y) = X(Yf) - \langle \operatorname{grad} f, \nabla_X Y\rangle$ and similarly $\operatorname{Hess} f(Y,X) = Y(Xf) - \langle \operatorname{grad} f, \nabla_Y X\rangle$, we have

$$\begin{aligned}\operatorname{Hess} f(X,Y) - \operatorname{Hess} f(Y,X) &= X(Yf) - Y(Xf) - \langle \operatorname{grad} f, \nabla_X Y - \nabla_Y X\rangle \\ &= [X,Y](f) - \langle \operatorname{grad} f, [X,Y]\rangle \\ &= [X,Y](f) - [X,Y](f) \\ &= 0\end{aligned}$$

due to the torsion-freeness of ∇: $\nabla_X Y - \nabla_Y X = [X,Y]$, the Lie bracket of vector fields. □

The Hessian tensor can also be represented by a matrix $A = [a_{ij}]$ w.r.t. an arbitrary orthonormal frame $\{E_i\}_{i=1}^n$, that is,

$$a_{ij} = \operatorname{Hess}(f)(E_i, E_j),$$

and the norm $\|\operatorname{Hess} f\|$ is defined as in Hilbert-Schmidt norm:

$$\|\operatorname{Hess} f\| := \sqrt{\operatorname{tr}(AA^\top)} = \sqrt{\sum_{i,j} a_{ij}^2},$$

which is independent of the choice of orthonormal frame E_i's.

Proposition 5.5. *The following two relations hold between Hessian and Laplacian.*

(a) $\operatorname{tr}(\operatorname{Hess} f) = \Delta f$

(b) $\|\operatorname{Hess} f\|^2 \geq \frac{1}{n}(\Delta f)^2$ *where n is the dimension of M.*

Proof. Part (a) follows directly from definitions:

$$\operatorname{tr}(\operatorname{Hess} f) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \langle \nabla_{E_i} \operatorname{grad} f, E_i \rangle = \operatorname{div}(\operatorname{grad} f) = \Delta f,$$

and the part (b) follows by applying Cauchy-Schwarz's inequality:

$$\| \operatorname{Hess} f \|^2 = \sum_{i,j} a_{ij}^2 \geq \sum_{i=1}^{n} a_{ii}^2 \geq \frac{1}{n} \left(\sum_{i=1}^{n} a_{ii} \right)^2 = \frac{1}{n} (\Delta f)^2.$$

□

5.2 Bochner's formula

Bochner's formula is an identity which connects differential operators to Ricci curvature. This formula is a fundamental motivation for the Bakry-Émery curvature notion. In this section, we give the statement of Bochner's formula and its proof, which can also be found in standard Riemannian geometry books, e.g. [GHL90, Proposition 4.15] and [Jos08, Theorem 4.5.1].

Theorem 5.6 (Bochner's formula)**.** *Let (M^n, g) be a complete Riemannian manifold. Then for any function $f \in C^\infty(M)$, the equation*

$$\frac{1}{2} \Delta |\operatorname{grad} f|^2 = \| \operatorname{Hess} f \|^2 + \langle \operatorname{grad} \Delta f, \operatorname{grad} f \rangle + \operatorname{Ric}(\operatorname{grad} f) \tag{5.2}$$

holds pointwise on M.

Here we use the notation $\operatorname{Ric}(v) = \operatorname{Ric}(v, v)$.

Proof. Let $x \in M$ and $\{e_i\}_{i=1}^n$ an orthonormal basis of $T_x M$ consider a local orthonormal frame $\{E_i\}_{i=1}^n$ such that the basis $\{e_i\}_{i=1}^n$ of $T_x M$ defines a Riemannian normal coordinate system at x. More precisely, for neighborhood $U \subset M$ containing x and $V \subset \mathbb{R}^n$ containing 0, the coordinate map $\phi : U \to V$ has its inverse

$$\phi^{-1}(x_1, ..., x_n) = \exp_x(\sum_i x_i e_i)$$

which maps straight lines through origin to geodesics through x. For $y \in U$, $\frac{\partial}{\partial x_i}|_y$ is an element of $T_y M$, and $\frac{\partial}{\partial x_i}|_x = e_i$. In particular, when evaluating at x, we have $\langle e_i, e_j \rangle = \delta_{ij}$ and $\nabla_{e_i} E_j = 0$ for all i and j (and hence $\nabla_u E_j = 0$ for all $u \in T_x M$ by linearity).

By definition of Hessian and its relation to Laplacian, we have the following evaluation,

$$\begin{aligned}\frac{1}{2}\Delta|\operatorname{grad} f|^2 &= \frac{1}{2}\operatorname{tr}(\operatorname{Hess}\|\operatorname{grad} f\|^2) = \frac{1}{2}\sum_{i=1}^{n}\operatorname{Hess}\|\operatorname{grad} f\|^2(E_i, E_i)\\ &= \frac{1}{2}\sum_{i=1}^{n} E_i(E_i\|\operatorname{grad} f\|^2) - \langle \operatorname{grad}\|\operatorname{grad} f\|^2, \nabla_{E_i}E_i\rangle, \end{aligned} \tag{5.3}$$

where the last inner product vanishes at x since $\nabla_{e_i}E_i = 0$. Moreover, by the definition and symmetry of Hessian, we have

$$\begin{aligned}\frac{1}{2}E_i(\|\operatorname{grad} f\|^2) &= \langle \nabla_{E_i}\operatorname{grad} f, \operatorname{grad} f\rangle = \operatorname{Hess} f(E_i, \operatorname{grad} f)\\ &= \operatorname{Hess} f(\operatorname{grad} f, E_i) = \langle \nabla_{\operatorname{grad} f}\operatorname{grad} f, E_i\rangle. \end{aligned} \tag{5.4}$$

We then substitute (5.4) into (5.3) and derive it at x in terms of Riemann curvature tensor:

$$\begin{aligned}\frac{1}{2}\Delta|\operatorname{grad} f|^2(x) &= \sum_{i=1}^{n} e_i\langle \nabla_{\operatorname{grad} f}\operatorname{grad} f, E_i\rangle\\ &= \sum_{i=1}^{n}\langle \nabla_{e_i}\nabla_{\operatorname{grad} f}\operatorname{grad} f, e_i\rangle + \langle \nabla_{\operatorname{grad} f}\operatorname{grad} f(x), \underbrace{\nabla_{e_i}E_i}_{=0}\rangle\\ &= \sum_{i=1}^{n}\langle R(e_i, \operatorname{grad} f(x))\operatorname{grad} f(x), e_i\rangle + \qquad (5.5)\\ &\quad \sum_{i=1}^{n}\langle \nabla_{\operatorname{grad} f(x)}\nabla_{E_i}\operatorname{grad} f, e_i\rangle + \qquad (5.6)\\ &\quad \sum_{i=1}^{n}\langle \nabla_{[E_i, \operatorname{grad} f](x)}\operatorname{grad} f, e_i\rangle. \qquad (5.7)\end{aligned}$$

The term in (5.5) is $\operatorname{Ric}(\operatorname{grad} f(x))$ by definition of Ricci curvature. We will finish the proof by showing that the terms in (5.6) and (5.7) are equal to $\langle \operatorname{grad}\Delta f(x), \operatorname{grad} f(x)\rangle$ and $\|\operatorname{Hess} f\|^2(x)$, respectively. The terms in (5.6) can be computed as

$$\begin{aligned}\sum_{i=1}^{n}\langle \nabla_{\operatorname{grad} f(x)}\nabla_{E_i}\operatorname{grad} f, e_i\rangle &= \sum_{i=1}^{n}(\operatorname{grad} f(x)\langle \nabla_{E_i}\operatorname{grad} f, E_i\rangle\\ &\quad - \langle \nabla_{e_i}\operatorname{grad} f, \underbrace{\nabla_{\operatorname{grad} f(x)}E_i}_{=0}\rangle)\\ &= \operatorname{grad} f(x)(\operatorname{tr}\ \operatorname{Hess} f)\end{aligned}$$

$$= \operatorname{grad} f(x)(\Delta f) = \langle \operatorname{grad} \Delta f(x), \operatorname{grad} f(x) \rangle.$$

Due to torsion-freeness $[E_i, \operatorname{grad} f](x) = \nabla_{E_i} \operatorname{grad} f(x) - \underbrace{\nabla_{\operatorname{grad} f(x)} E_i}_{=0}$, we can compute the term in line (5.7) as

$$\begin{aligned}
\sum_{i=1}^{n} \langle \nabla_{[E_i, \operatorname{grad} f](x)} \operatorname{grad} f, e_i \rangle &= \sum_{i=1}^{n} \operatorname{Hess} f([E_i, \operatorname{grad} f](x), e_i) \\
&= \sum_{i=1}^{n} \operatorname{Hess} f(e_i, \nabla_{e_i} \operatorname{grad} f) \\
&= \sum_{i=1}^{n} \langle \nabla_{e_i} \operatorname{grad} f, \nabla_{e_i} \operatorname{grad} f \rangle = \| \operatorname{Hess} f \|^2(x).
\end{aligned}$$

□

5.3 Eigenvalues of Laplacian and Lichnerowicz spectral gap theorem

In this section, we give a brief overview of the Laplacian eigenvalues of manifolds and the Lichnerowicz theorem which relates the smallest eigenvalue to the lower bound of the Ricci curvature. For more details, we refer to Chavel's book [Cha84].

Suppose for simplicity that (M^n, g) is a closed and connected Riemannian manifold. The ***eigenvalues*** of the Laplacian operator on M are all the real numbers λ such that there exists a nontrivial solution $f \in C^2(M)$ to the following eigenvalue problem

$$\Delta f + \lambda f = 0$$

For each eigenvalue λ, a corresponding solution f is called an ***eigenfunction***. The ***eigenspace*** of λ is the vector space of all eigenfunctions with respect to λ.

It is well-known that the set of these eigenvalues (with their multiplicities) consists of a sequence

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots, \text{where } \lambda_n \to \infty \text{ as } n \to \infty,$$

and each eigenspace has a finite dimension (which is the multiplicity of the corresponding λ). The trivial eigenvalue λ_0 corresponds to the constant function f. The smallest nonzero eigenvalue λ_1 is known as the spectral gap of the Laplacian.

Remark 5.7. Here we list some facts in case M is not closed and connected.

1. If M is not connected, the multiplicity of the trivial eigenvalue equals the number of connected components of M.

2. If M is compact but not closed (i.e., $\partial M \neq \emptyset$), one may consider different eigenvalue problems. For example, the Neumann problem comes with the boundary condition $\hat{\mathrm{n}}(f) = 0$ on ∂M (where $\hat{\mathrm{n}}$ is the unit normal vector). On the other hand, the Dirichlet problem requires $f = 0$ on ∂M, which means constant functions are not eigenfunctions.

3. If M is non-compact, then the Laplacian may not have discrete spectrum. For example, the spectrum of the Euclidean space $\mathbb{R}^n$ is $[0, \infty)$, and the spectrum of the hyperbolic space $\mathbb{H}^n$ is $[\frac{(n-1)^2}{4}, \infty)$.

The Lichnerowicz spectral gap theorem [Lic58] asserts that the smallest nonzero eigenvalue λ_1 can be estimated from below in terms of the positive lower bound of Ricci curvature. The proof of the Lichnerowicz theorem is a straightforward implication of the Bochner's formula. In a general case where M has nonempty boundary, we refer to a proof in, e.g. [GHL90, Theorem 4.70].

Theorem 5.8 (Lichnerowicz)**.** *Let (M^n, g) be a closed Riemannian manifold. Suppose that $K := \mathrm{Ric}(M) > 0$. Then the smallest nonzero Laplacian eigenvalue satisfies*

$$\lambda_1 \geq \frac{n}{n-1} K.$$

Proof. Consider an eigenfunction f satisfying $\Delta f + \lambda f = 0$. Without loss of generality, we assume that $\int_M f^2 = 1$. Bochner's formula (5.2) gives

$$\begin{aligned}
\frac{1}{2}\Delta |\operatorname{grad} f|^2 &= \|\operatorname{Hess} f\|^2 + \langle \operatorname{grad}(-\lambda f), \operatorname{grad} f\rangle + \mathrm{Ric}(\operatorname{grad} f) \\
&= \|\operatorname{Hess} f\|^2 - \lambda |\operatorname{grad} f|^2 + \mathrm{Ric}(\operatorname{grad} f) \\
&\geq \frac{1}{n}(\Delta f)^2 - \lambda |\operatorname{grad} f|^2 + K |\operatorname{grad} f|^2 \\
&= \frac{\lambda^2}{n} f^2 + (K - \lambda) |\operatorname{grad} f|^2 \\
&= \frac{\lambda^2}{n} f^2 + (K - \lambda)(\frac{1}{2}\Delta f^2 + \lambda f^2),
\end{aligned}$$

where the inequality is due to $\|\operatorname{Hess} f\|^2 \geq \frac{1}{n}(\Delta f)^2$ and the lower curvature bound assumption, and the last equality is the product rule : $\Delta f^2 = 2|\operatorname{grad} f|^2 + 2f\Delta f$.

Integrating the above inequality over M and using the fact from the divergence theorem that $\int_M \Delta|\operatorname{grad} f|^2 = 0 = \int_M \Delta f^2$ by , we derive

$$0 \geq \frac{\lambda^2}{n} + (K-\lambda)\lambda.$$

In particular, for $\lambda_1 > 0$, we conclude $\lambda_1 \geq \frac{n}{n-1}K$ as desired. □

Furthermore, like in Cheng's rigidity to the Bonnet-Myers diamter bound, there is also a rigidity theorem for the Lichnerowicz spectral bound, known as Obata's rigidity [Oba62] which confirms that the bound is sharp if and only if the manifold is isometric to a round sphere.

Theorem 5.9 (Obata's rigidity, [Oba62])**.** *Let $(M^n, \langle\cdot,\cdot\rangle)$ be a compact Riemannian manifold with positive lower Ricci curvature bound $K := Ric(M) > 0$. Then $\lambda_1(M) = \frac{n}{n-1}K$ if and only if M is isometric to the sphere S^n_r.*

5.4 Curvature-dimension inequality and Γ-calculus

Due to the Bochner's formula and $\|\operatorname{Hess} f\|^2 \geq \frac{1}{n}(\Delta f)^2$, the pointwise lower Ricci curvature bound $\operatorname{Ric}_x \geq K$ immediately implies the following ***curvature-dimension inequality*** in the sense of Bakry-Émery [BE84]:

$$\frac{1}{2}\Delta|\operatorname{grad} f|^2(x) - \langle \operatorname{grad}\Delta f(x), \operatorname{grad} f(x)\rangle \geq \frac{1}{n}(\Delta f(x))^2 + K|\operatorname{grad} f(x)|^2. \quad (5.8)$$

Bakry and Émery also introduce the following Γ-calculus, consisting of the first iteration Γ and the second iteration Γ_2, which helps to reformulate the curvature-dimension inequality.

Definition 5.10 (Γ and Γ_2)**.** For $f, g \in C^\infty(M)$, define bilinear operators

$$2\Gamma(f,g) := \Delta(fg) - f\cdot\Delta g - \Delta f\cdot g, \quad (5.9)$$
$$2\Gamma_2(f,g) := \Delta\Gamma(f,g) - \Gamma(f,\Delta g) - \Gamma(\Delta f, g), \quad (5.10)$$

and for shortened notation, we write $\Gamma f := \Gamma(f,f)$ and $\Gamma_2 f := \Gamma_2(f,f)$.

By the product rule, one can realize that

$$\Gamma(f,g) = \langle \operatorname{grad} f, \operatorname{grad} g\rangle, \text{ and } \Gamma f = |\operatorname{grad} f|^2. \quad (5.11)$$

The inequality (5.8) can then be rewritten as $\Gamma_2 f(x) \geq \frac{1}{n}(\Delta f(x))^2 + K\Gamma f(x)$, which motivates the following definition of Bakry-Émery curvature notion.

Definition 5.11. Let (M^n, g) be a Riemannian manifold and let $K \in \mathbb{R}$ and $N \in [0, \infty]$. A point $x \in M$ is said to satisfy $BE(K, N)$ if the following inequality holds true:

$$\Gamma_2 f(x) \geq \frac{1}{N}(\Delta f(x))^2 + K\Gamma f(x) \qquad \forall f \in C^\infty(M). \tag{5.12}$$

Here K is an lower bound of the Ricci curvature and N is an upper bound for the dimension of M. In the case that $N = \infty$, the $BE(K, \infty)$ inequality at x is read as

$$\Gamma_2 f(x) \geq K\Gamma f(x). \tag{5.13}$$

In fact, the following result follows immediately from Bochner's formula.

Theorem 5.12. *Let (M^n, g) be a Riemannian manifold and $x \in M$ such that $\mathrm{Ric}_x(v) \geq K|v|^2$ for all $v \in T_xM$. Then x satisfies $BE(K, N)$ for all $N \geq n$.*

5.5 Heat kernel and heat semigroup operator on manifolds

In this section, we briefly overview the heat kernel p_t and the heat semigroup operator P_t. They are main ingredients to another characterization of $BE(K, n)$ in terms of gradient estimate, which is the topic of the next section. Basic and useful properties of p_t and P_t will be provided here without proofs, and we refer to Grigoryan's book [Gri09] for details.

We consider the following ***Cauchy problem*** for the heat equation on a complete Riemannian manifold. Given a bounded continuous function $f \in C_b(M)$, the problem is to solve $u : [0, \infty) \times M \to \mathbb{R}$ (which is continuous, C^1 in $t \in (0, \infty)$ and C^2 in $x \in M$) such that

$$\begin{cases} \frac{\partial}{\partial t} u &= \Delta u \qquad \text{on } (0, \infty) \times M, \\ u(0, x) &= f(x). \end{cases} \tag{5.14}$$

Definition 5.13 (Fundamental solution)**.** A function $p : (0, \infty) \times M \times M \to \mathbb{R}$ is called a ***fundamental solution*** of the heat equation, if for every $f \in C_b(M)$, the function

$$u(t, x) = \begin{cases} \int_M p(t, x, y) f(y) d\,\mathrm{vol}(y) & \text{for } t > 0, \\ f(x) & \text{for } t = 0, \end{cases} \tag{5.15}$$

is a solution to Cauchy problem (5.14) with initial data f.

The existence of solutions u is proved via the following Dodziuk's construction [Dod83] (see also Chavel's book [Cha84]). First, we construct an exhaustion of M by regular domains $\{D_\ell\}_{\ell=1}^{\infty}$ (relatively compact and open subsets of M with smooth boundaries) so that $\bar{D}_\ell \subset D_{\ell+1}$ and $\bigcup_\ell D_\ell = M$. Then for each ℓ, consider $p_\ell : (0,\infty) \times D_\ell \times D_\ell \to \mathbb{R}^+$ to be the ***Dirichlet heat kernel***, that is, p_ℓ is the unique fundamental solution of the following Dirichlet problem:

$$\begin{cases} \frac{\partial}{\partial t} u & = \Delta u \qquad \text{on } (0,\infty) \times D_\ell, \\ u(t,x) & = 0 \qquad \text{on } (0,\infty) \times \partial D_\ell, \\ u(0,x) & = f(x). \end{cases} \tag{5.16}$$

Two important properties of p_ℓ are that for all $t > 0$, $\int_{D_\ell} p_\ell(t,x,y) d\,\mathrm{vol}(y) \le 1$ for all $x \in D_\ell$ and $p_\ell(t,x,y) \le p_{\ell+1}(t,x,y)$ for all $x, y \in D_\ell$ (which is a consequence of parabolic maximum principle); see [Dod83, Lemma 3.3]. It follows that the limit

$$p(t,x,y) := \lim_{\ell \to \infty} p_\ell(t,x,y)$$

exists, and the convergence is locally uniform due to Dini's theorem. This p is indeed the smallest positive fundamental solution of the heat equation on M. We call this p the ***heat kernel*** of M. Furthermore, the heat semigroup operator is defined via the kernel p as follows.

Definition 5.14 (heat semigroup operator)**.** The ***heat semigroup*** $\{P_t\}_{t \ge 0}$ is the family of operators $P_t : C_b(M) \to C_b(M)$ defined as $P_0 = \mathrm{Id}$ and for $t > 0$,

$$P_t f(x) := \int_M p(t,x,y) f(y) d\,\mathrm{vol}(y),$$

for all $f \in C_b(M)$ and $x \in M$.

It means the function $u(t,x) := P_t f(x)$ solves the Cauchy problem (5.14), or in short,

$$\frac{\partial}{\partial t} P_t = \Delta P_t.$$

Moreover, $P_t f(x)$ is C^∞ in $(0,\infty) \times M$ with the limit $\lim_{t \to 0} P_t f(x) = f(x)$ locally uniformly in $x \in M$. It also satisfies

$$\inf f \le P_t f(x) \le \sup f$$

(see [Gri09, Theorem 7.16]). Basic properties of the heat kernel p_t and the heat semigroup are summarized in the following two propositions.

Proposition 5.15. *The heat kernel* $p : (0, \infty) \times M \times M \to \mathbb{R}$ *satisfies the following properties.*

1. *symmetry:* $p(t, x, y) = p(t, y, x)$.
2. *semigroup identity:* $p(s + t, x, y) = \int_M p(s, x, z)p(t, z, y)d\,\mathrm{vol}(z)$.
3. *density-like:* $p(t, x, y) > 0$ *and*

$$\int_M p(t, x, y)d\,\mathrm{vol}(y) \leq 1.$$

Proposition 5.16. *The heat semigroup operator* $P_t : C_b(M) \to C_b(M)$ *satisfies the following properties.*

1. $P_t \circ P_s = P_{s+t}$.
2. $\Delta P_t = P_t \Delta$.
3. $f(x) \geq 0$ *for all* x *implies* $P_t f(x) \geq 0$ *for all* x.
4. $\|P_t(f)\|_\infty \leq \|f\|_\infty$.

The uniqueness of a bounded solution to the Cauchy problem (5.14) requires a certain condition on the manifold M called ***stochastic completeness***, which roughly says that the heat is preserved within the system.

Definition 5.17 (Stochastic completeness)**.** A Riemannian manifold (M, g) is said to be ***stochastically complete*** if the heat kernel p satisfies

$$\int_M p(t, x, y)d\,\mathrm{vol}(y) = 1,$$

for all $t > 0$ and $x \in M$.

Theorem 5.18 ([Gri09, Theorem 8.18])**.** *A Riemannian manifold* (M, g) *is stochastically complete if and only if the Cauchy problem* (5.14) *has exactly one bounded solution* u.

A criterion for the stochastic completeness by the volume test is provided in [Gri09, Theorem 11.8].

Theorem 5.19 ([Gri09]). *Let (M, g) be a complete connected Riemannian manifold. If for some $x_0 \in M$,*

$$\int_0^\infty \frac{r}{\log \operatorname{vol}(B_r(x_0))} dr = \infty, \tag{5.17}$$

then M is stochastically complete. Here $B_r(x_0)$ denotes the geodesic ball of radius r centered at x_0.

The condition (5.17) holds in particular if $\operatorname{vol}(B_r(x)) \le \exp(Cr^2)$ for all r large enough. For example, the Euclidean space $\mathbb{R}^n$ with vanishing curvature and the hyperbolic space $\mathbb{H}^n$ with constant negative curvature equal to -1 are stochastically complete. More generally, it is proved in [Yau76] that any complete manifold with Ricci curvature bounded from below is stochastically complete.

Remark 5.20. In [Gri09], the author considers the Cauchy problem (5.14) first for $L^2(M)$ before the case $C_b(M)$ is covered. In the case of a (geodesically) complete Riemannian manifold (M, g), the Laplacian on functions with compact support is known to be essentially self-adjoint (see e.g., [Gaf54, Che73, Kar84]) and Spectral Theory can be employed to construct the self-adjoint heat semigroup operators $P_t : L^2(M) \to L^2(M)$, namely $P_t = e^{t\Delta}$. In the L^2-Cauchy problem, the solution is a unique: $u = P_t f$ with these uniquely determined operators $P_t : L^2(M) \to L^2(M)$, which are bounded, nonnegative, commute with Δ and map $L^2(M)$ to $C^\infty(M)$ for $t > 0$. They are integral operators having smooth kernels which agree with the minimal heat kernel $p(t, x, y)$ from above.

5.6 Gradient estimate

An important result from Bakry and Émery [BE84] is the characterization of $BE(K, \infty)$ in terms of gradient estimate, which we present here in the following two variations.

Definition 5.21 (gradient estimates). A point $x \in M$ is said to satisfy the ***gradient estimate*** (or the **Γ*-gradient estimate***, respectively) if for all $f \in C_c^\infty(M)$ and $t \in [0, \infty)$,

$$|\operatorname{grad} P_t f|(x) \le \mathrm{e}^{-Kt} P_t |\operatorname{grad} f|(x), \tag{GE}$$

or

$$\Gamma(P_t f)(x) \le \mathrm{e}^{-2Kt} P_t(\Gamma f)(x), \tag{GE-Γ}$$

respectively. Moreover, we say that M satisfies gradient estimate (or Γ-gradient estimate) if such condition holds for all $x \in M$.

The condition (GE-Γ) can be considered as a weaker version of (GE) since

$$(P_t|\operatorname{grad} f|)^2 \le P_t(|\operatorname{grad} f|^2) = P_t(\Gamma f).$$

In fact, it is among the most famous results in the original work of Bakry and Émery that these gradient estimates are equivalent to the curvature-dimension condition $BE(K,\infty)$ on manifolds.

Theorem 5.22. *Let (M,g) be a complete Riemannian manifold. Let $x \in M$ and $K \in \mathbb{R}$. Then the following properties are equivalent.*

1. *$BE(K,\infty)$ at x: $\Gamma_2 f(x) \ge K\Gamma f(x)$ for all f.*
2. *Γ-gradient estimate at x: $\Gamma(P_t f)(x) \le \mathrm{e}^{-2Kt}\, P_t(\Gamma f)(x)$ for all f.*
3. *gradient estimate at x: $\sqrt{\Gamma(P_t f)}(x) \le \mathrm{e}^{-Kt}\, P_t(\sqrt{\Gamma f})(x)$ for all f.*

The proof of this theorem, provided below, relies on the well-known semigroup interpolation argument. The equivalence (1) ⇔ (2) will make an appearance again when we discuss the Bakry-Émery curvature on graphs and the Erbar-Maas entropic Ricci curvature. Although the equivalence (1) ⇔ (3) looks similar, it indeed requires an additional diffusion property of Δ on manifolds (referred to [Bak97, pp. 4]) as we shall see in the proof.

Proof. (1) ⇒ (2): Fix $t \in [0,\infty)$ and $f \in C_c^\infty(M)$. Define a function $F : [0,t] \to \mathbb{R}$ as

$$F(s) := e^{-2Ks} P_s(\Gamma(P_{t-s}f))(x).$$

Every function here is evaluated at x and we omit it from the writing. Moreover, let $g = P_{t-s}f$. The derivative of F is computed as:

$$\begin{aligned} F'(s) &= e^{-2Ks}\Big(-2KP_s(\Gamma g) + (\frac{\partial}{\partial s}P_s)(\Gamma g) + P_s(\frac{\partial}{\partial s}\Gamma g)\Big) \\ &= e^{-2Ks}P_s\Big(-2K\Gamma P_{t-s}f + \Delta(\Gamma g) + \frac{\partial}{\partial s}\Gamma g\Big), \end{aligned}$$

where in the second line we recall that $\frac{\partial}{\partial s}P_s = \Delta P_s = P_s\Delta$. Next, the term $\frac{\partial}{\partial s}\Gamma g$ is equal to

$$\begin{aligned} \frac{\partial}{\partial s}\Gamma(P_{t-s}f, P_{t-s}f) = 2\Gamma\Big(\frac{\partial}{\partial s}P_{t-s}f, P_{t-s}f\Big) &= -2\Gamma(\Delta P_{t-s}f, P_{t-s}f) \\ &= 2\Gamma_2 g - \Delta(\Gamma g). \end{aligned} \tag{5.18}$$

Substituting (5.18) into the calculation of $F'(s)$ yields

$$F'(s) = 2e^{-2Ks}P_s\Big(\Gamma_2 g - K\Gamma g\Big). \tag{5.19}$$

Under the assumption $BE(K,\infty)$ at x, we have $F'(s) \geq 0$ for all $s \in [0,t]$. Thus $F(0) \leq F(t)$ which yields (GE-Γ).

(2) ⇐ (1): Suppose x satisfies (GE-Γ). Differentiating (GE-Γ) near $t=0$ gives

$$\begin{aligned}
0 &\leq \lim_{t\to 0}\frac{1}{t}\big(e^{-2Kt}P_t(\Gamma f) - \Gamma(P_t f)\big) = \frac{\partial}{\partial t}\Big|_{t=0}\big(e^{-2Kt}P_t(\Gamma f) - \Gamma(P_t f)\big) \\
&= \Delta(\Gamma f) - 2K\Gamma f - \frac{\partial}{\partial t}\Big|_{t=0}\Gamma(P_t f, P_t f) \\
&= \Delta(\Gamma f) - 2K\Gamma f + \Big[2\Gamma_2(P_t f) - \Delta(\Gamma(P_t f))\Big]_{t=0} = 2\Gamma_2 f - 2K\Gamma f.
\end{aligned}$$

Thus x satisfies $BE(K,\infty)$ as desired.

Similarly, the proof of (1) ⇒ (3) starts by considering the function

$$\tilde{F}(s) := e^{-Ks}P_s(\sqrt{\Gamma P_{t-s} f})(x).$$

The derivative $\tilde{F}(s)$ (with $g := P_{t-s}$) gives

$$\begin{aligned}
\tilde{F}'(s) &= e^{-Ks}\Big(-KP_s(\sqrt{\Gamma g}) + (\frac{\partial}{\partial s}P_s)(\sqrt{\Gamma g}) + P_s(\frac{\partial}{\partial s}\sqrt{\Gamma g})\Big) \\
&= e^{-Ks}P_s\Big(-K\sqrt{\Gamma g} + \Delta\sqrt{\Gamma g} + \frac{1}{2\sqrt{\Gamma g}}\cdot\frac{\partial}{\partial s}\Gamma g\Big) \\
&\overset{(5.18)}{=} e^{-Ks}P_s\Big[\frac{1}{2\sqrt{\Gamma g}}\Big(2\Gamma_2 g - 2K\Gamma g + 2\sqrt{\Gamma g}\Delta\sqrt{\Gamma g} - \Delta(\Gamma g)\Big)\Big] \\
&= e^{-Ks}P_s\Big[\frac{1}{\sqrt{\Gamma g}}\Big(\Gamma_2 g - K\Gamma g - \Gamma(\sqrt{\Gamma g})\Big)\Big].
\end{aligned}$$

It was shown in [Bak97, Lemmas 2.4 and 2.5] that $BE(K,\infty)$ together with the diffusion property gives

$$\Gamma_2 g - K\Gamma g - \Gamma(\sqrt{\Gamma g}) \geq 0.$$

(This inequality is a self-improvement of $\Gamma_2 - K\Gamma \geq 0$, and it is proved by employing the coordinate-system to formulate Δ, Γ, Γ_2 and Hessian, which we do not cover here). Thus $\tilde{F}'(s) \geq 0$, yielding (GE). Lastly, (3) ⇒ (1) can be viewed as (GE) ⇒ (GE-Γ) ⇒ $BE(K,\infty)$. □

Finally, we conclude the equivalent characterizations of the global lower Ricci curvature bound in spirit of Bakry-Émery.

Theorem 5.23. *Let (M^n, g) be a complete Riemannian manifold of dimension n and let $K \in \mathbb{R}$. Then the following conditions on M are equivalent:*

(i) $\mathrm{Ric}(M) \geq K$*,*

(ii) $BE(K, n)$*,*

(iii) $BE(K, \infty)$*,*

(iv) gradient estimate: $\sqrt{\Gamma(P_t f)} \leq \mathrm{e}^{-Kt} P_t(\sqrt{\Gamma f})$ *for all* f*.*

(v) Γ*-gradient estimate:* $\Gamma(P_t f) \leq \mathrm{e}^{-2Kt} P_t(\Gamma f)$ *for all* f*.*

Proof. The implication (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) follows from the Bochner's formula, the curvature-dimension inequality, and the work of Γ-calculus. The implication (iii) $\Rightarrow$ (iv) $\Rightarrow$ (v) follows from Theorem 5.22. The implication (v) $\Rightarrow$ (i) is due to [vRS05, Theorem 1.3]. □

Chapter 6

Bakry-Émery manifolds

A ***Bakry-Émery manifold***, also known as ***weighted manifold***, is a smooth Riemannian manifold $(M, \langle , \rangle)$ together with a reference measure $d\nu(x) = e^{-V(x)} d\,\mathrm{vol}(x)$ for a fixed function $V \in C^2(M)$.

The divergence operator is given by $\mathrm{div}_V(X) := \mathrm{div}\, X - \langle \mathrm{grad}\, V, X\rangle$ and it is the negative adjoint of gradient (with respect to ν), that is,

$$\begin{aligned}\int_M \langle \mathrm{grad}\, f, X\rangle d\nu &= \int_M \langle \mathrm{grad}\, f, e^{-V} X\rangle d\,\mathrm{vol} = -\int_M f\, \mathrm{div}(e^{-V} X) d\,\mathrm{vol} \\ &= -\int_M f(e^{-V}\,\mathrm{div}\, X + \langle \mathrm{grad}\, e^{-V}, X\rangle) d\,\mathrm{vol} \\ &= -\int_M f(e^{-V}\,\mathrm{div}\, X - e^{-V}\langle \mathrm{grad}\, V, X\rangle) d\,\mathrm{vol} \\ &= -\int_M f(\underbrace{\mathrm{div}\, X - \langle \mathrm{grad}\, V, X\rangle}_{=:\mathrm{div}_V(X)}) d\nu.\end{aligned}$$

The Laplacian Δ_V is then given by

$$\Delta_V f := \mathrm{div}_V(\mathrm{grad}\, f) = \Delta f - \langle \mathrm{grad}\, V, \mathrm{grad}\, f\rangle,$$

and the generalized Ricci tensor (with ∞ dimension) is defined to be

$$\mathrm{Ric}_{V,\infty} := \mathrm{Ric} + \mathrm{Hess}\, V.$$

We provide two justifications of this generalized Ricci tensor. The first one is the weighted Bochner's formula (c.f. [OS14, Formula (1.2)] and [Vil09, Chapter 14]). The second one is the generalized result of the main theorem (Theorem 3.3) on weighted manifolds as presented in [LV09] and [Stu06].

Theorem 6.1 (weighted Bochner's formula)**.** *Let (M^n, g) be a complete Riemannian manifold. Then*

$$\frac{1}{2}\Delta_V |\operatorname{grad} f|^2 = \|\operatorname{Hess} f\|^2 + \langle \operatorname{grad} \Delta_V f, \operatorname{grad} f\rangle + \operatorname{Ric}_{V,\infty}(\operatorname{grad} f) \tag{6.1}$$

holds pointwise for any function $f \in C^\infty(M)$.

Proof. Subtracting (6.1) from the original Bochner's formula (5.2), it is left to show that

$$\frac{1}{2}\langle \operatorname{grad} V, \operatorname{grad} |\operatorname{grad} f|^2\rangle = \langle \operatorname{grad}\langle \operatorname{grad} V, \operatorname{grad} f\rangle, \operatorname{grad} f\rangle - \operatorname{Hess} V(\operatorname{grad} f, \operatorname{grad} f).$$

By the definition of the gradient and the Hessian and the Riemannian property, we have

$$\begin{aligned}\frac{1}{2}\langle \operatorname{grad} V, \operatorname{grad} |\operatorname{grad} f|^2 &= \frac{1}{2}\operatorname{grad} V\langle \operatorname{grad} f, \operatorname{grad} f\rangle\\ &= \langle \nabla_{\operatorname{grad} V} \operatorname{grad} f, \operatorname{grad} f\rangle\\ &= \operatorname{Hess} f(\operatorname{grad} V, \operatorname{grad} f).\end{aligned}$$

Similarly,

$$\begin{aligned}\langle \operatorname{grad}\langle \operatorname{grad} V, \operatorname{grad} f\rangle, \operatorname{grad} f\rangle &= \operatorname{grad} f\langle \operatorname{grad} V, \operatorname{grad} f\rangle\\ &= \langle \nabla_{\operatorname{grad} f} \operatorname{grad} V, \operatorname{grad} f\rangle + \langle \operatorname{grad} V, \nabla_{\operatorname{grad} f} \operatorname{grad} f\rangle\\ &= \operatorname{Hess} V(\operatorname{grad} f, \operatorname{grad} f) + \operatorname{Hess} f(\operatorname{grad} f, \operatorname{grad} V).\end{aligned}$$

The desired identity then follows from the symmetry of the Hessian. □

Theorem 6.2. *For any smooth complete Riemannian manifold M and any $K \in \mathbb{R}$, the following properties are equivalent:*

(i) $\operatorname{Ric}_{V,\infty}(M) \geq K$*, that is* $\operatorname{Ric}_x(v) + \operatorname{Hess}_x V(v,v) \geq K|v|^2$ *for all $x \in M$, $v \in T_xM$.*

(ii) The relative entropy $\mathcal{E}_\nu(\cdot)$ with respect to $d\nu(x) = e^{-V(x)}d\operatorname{vol}(x)$ is displacement K-convex on $\mathscr{P}_2(M)$.

Proof. The proof here is similar and uses the same setup as the one given in Theorem 3.3. Again, we only prove that $(i) \Rightarrow (ii)$. Suppose that $\mu_0, \mu_1 \ll \operatorname{vol}$. Then there exists a unique geodesic $(\mu_t)_{t\in[0,1]}$ in $\mathscr{P}_2(M)$ from μ_0 to μ_1 given by $\mu_t = (T_t)_\#\mu_0$. For each fixed x, we consider the curve $\gamma_x : [0,1] \to M$ given by $\gamma_x(t) := T_t(x)$. We also write $J_x(t) := DT_t(x)$.

For all t, $\mu_t \ll \mathrm{vol}$, so we can write

$$d\mu_t(x) = \rho_t(x) d\,\mathrm{vol}(x) = \rho_t(x) e^{V(x)} d\nu(x).$$

The relative entropy can be written as

$$\begin{aligned}
\mathcal{E}_\nu(\mu_t) &= \int_M \log\left(\rho_t(x) e^{V(x)}\right) d\mu_t(x) \\
&= \mathcal{E}_{\mathrm{vol}}(\mu_t) + \int_M V(x) d\mu_t(x) \\
&= \mathcal{E}_{\mathrm{vol}}(\mu_0) - \int_M \log(\det J_x(t)) d\mu_0(x) + \int_M V(x) d\mu_t(x) \\
&= \mathcal{E}_{\mathrm{vol}}(\mu_0) + \int_M \left(-\log(\det J_x(t)) + V(\gamma_x(t))\right) d\mu_0(x),
\end{aligned}$$

where the third equality comes from (3.5) and the last equality is the change of variables formula for pushforward $\mu_t = (T_t)_\# \mu_0$ together with the identification $T_t(x) = \gamma_x(t)$. To compute the second derivative of $\mathcal{E}_\nu(\mu_t)$, we need the derivative of the two terms in the above integrand. First, recall from (3.6) and (3.7) that

$$-\frac{d^2}{dt^2} \log(\det J_x(t)) \geq -\operatorname{tr}(J_x''(t) J_x^{-1}(t)) = \mathrm{Ric}_{\gamma_x(t)}(\gamma_x'(t), \gamma_x'(t)). \tag{6.2}$$

For the second term $V(\gamma_x(t))$, its second derivative can then be expressed in terms of Hessian as

$$\begin{aligned}
\frac{d^2}{dt^2} V(\gamma_x(t)) &= \frac{d}{dt} \langle \operatorname{grad} V(\gamma_x(t)), \gamma_x'(t) \rangle \\
&= \langle \nabla_{\gamma_x'(t)} (\operatorname{grad} V), \gamma_x'(t) \rangle + \langle \operatorname{grad} V(\gamma_x(t)), \frac{D}{dt} \gamma_x'(t) \rangle \\
&= \operatorname{Hess} V(\gamma_x'(t), \gamma_x'(t)),
\end{aligned} \tag{6.3}$$

where the latter inner-product vanishes because $\frac{D}{dt}\gamma_x'(t) = 0$ as γ_x is a geodesic.

Now the assumption $\mathrm{Ric}_{V,\infty}(M) \geq K$ can be applied to equations (6.2) and (6.3), yielding

$$\begin{aligned}
\frac{d^2}{dt^2}\left(-\log(\det J_x(t)) + V(\gamma_x(t))\right) &\geq \mathrm{Ric}_{\gamma_x(t)}(\gamma_x'(t), \gamma_x'(t)) + \operatorname{Hess} V(\gamma_x'(t), \gamma_x'(t)) \\
&\geq K |\gamma_x'(t)|^2 = K d(x, T_1(x))^2,
\end{aligned}$$

where the last equality $|\gamma_x'(t)| = d(x, T_1(x))$ is, again, due to the fact that γ_x is a minimal constant speed geodesic whose length is $d(x, T_1(x))$. Finally, we obtain

$$\frac{d^2}{dt^2} \mathcal{E}_\nu(\mu_t) = \int_M \frac{d^2}{dt^2} \left(-\log(\det J_x(t)) + V(\gamma_x(t))\right) d\mu_0(x)$$

$$\geq \int_M Kd(x, T_1(x))^2 d\mu_0(x) = KW_2(\mu_0, \mu_1)^2,$$

since T_1 is optimal transport map from μ_0 to μ_1. and we can derive (by using the same trick as in the proof of Theorem 3.3) that

$$\mathcal{E}_\nu(\mu_t) \leq (1-t)\mathcal{E}_\nu(\mu_0) + t\mathcal{E}_\nu(\mu_1) - \frac{1}{2}Kt(1-t)W_2(\mu_0, \mu_1)^2.$$

□

Chapter 7

Ollivier's coarse Ricci curvature

Another synthetic notion of Ricci curvature on metric measure spaces was proposed by Ollivier in [Oll09], which is called ***coarse Ricci curvature***. While the curvature notion by Lott-Sturm-Villani is based on the displacement convexity of the entropy functional on the L^2-Wasserstein space, Ollivier's coarse Ricci is defined via the contraction property of L^1-Wasserstein metric.

The definition of Ollivier's is motivated from the following remarkable observation by von Renesse and Sturm (see [vRS05, Theorem 1.5 (xii)]. Let (M^n, g) be a Riemannian manifold with lower curvature bound $\mathrm{Ric}(M) \geq K$, and let $m_{r,x}$ denote a uniform probability measure (with respect to the volume measure) on a geodesic ball of radius r centered at $x \in M$. The L^1-Wasserstein distance $W_1(m_{r,x}, m_{r,y})$ satisfies the asymptotic estimate:

$$W_1(m_{r,x}, m_{r,y}) \leq \left(1 - \frac{K}{2(n+2)} r^2 + o(r^2)\right) \cdot d(x, y).$$

In particular, if $K > 0$ then this Wasserstein distance is shorter than the distance between the centers. A refined statement by Ollivier [Oll09] of this observation will be discussed in the upcoming section.

For now, we shall provide the definition Ollivier's notion of coarse Ricci curvature for a metric measure space. Here we work in the setting that (X, d) is a Polish space with σ-Borel algebra, and $\mathscr{P}_1(X)$ denotes the set of all Borel probability measures with finite 1-moments.

Definition 7.1 (random walk)**.** A ***random walk*** m on X is a family of probability measures $m_x(\cdot) \in \mathscr{P}_1(X)$ for each $x \in X$ such that the measure m_x depends measurably on $x \in X$, that is, for any Borel set $A \subset X$ and $c \in \mathbb{R}$, the set $\{x \in X : m_x(A) < c\}$ is Borel.

Definition 7.2 (coarse Ricci curvature)**.** Let (X, d, m) be a Polish space with a random walk m. Given two different points $x, y \in X$, the ***coarse Ricci curvature*** along xy is defined as

$$\kappa(x,y) := 1 - \frac{W_1(m_x, m_y)}{d(x,y)}. \tag{7.1}$$

7.1 Distance between two balls

Now we explain two important results from [Oll09] which give justification to the terminology of coarse Ricci curvature defined as in (7.1).

Proposition 7.3 ([Oll09, Proposition 6])**.** *Let (M^n, g) be a smooth connected and complete Riemannian manifold and let $x \in M$. Let $\delta, \varepsilon > 0$ small enough, and let $v, w \in T_xM$ be unit tangent vectors at x. Let $y := \exp_x(\delta v)$ and let $w' \in T_yM$ be the tangent vector at y obtained by the parallel transport of w along the geodesic* $\exp_x(tv)$*. Moreover, let $x' := \exp_x(\varepsilon w)$ and $y' := \exp_y(\varepsilon w')$. Then*

$$d(x', y') \le \delta\Big(1 - \varepsilon^2 K(v,w) + O(\varepsilon^3 + \varepsilon^2\delta)\Big), \tag{7.2}$$

where $K(v, w)$ is the sectional curvature of the two-dimensional plane spanned by $\{v, w\}$.

Example 7.4 ([Oll09, Example 7])**.** Let (M^n, g) be a smooth connected and complete Riemannian manifold. For some small $r > 0$, let the Markov chain m^r be defined as

$$m_x^r(A) := \frac{\mathrm{vol}(A \cap B_r(x))}{\mathrm{vol}\, B_r(x)} \qquad \forall \text{ Borel } A \subset M$$

(that is m_x^r represents the uniform probability measure on the geodesic ball $B_r(x)$). Let $x \in M$ and let $v \in T_xM$ be a unit tangent vector. Let y be a point on the geodesic issuing from v, with $d(x, y)$ small enough. Then the L^1-Wasserstein distance between m_x^r and m_y^r satisfies the following estimate:

$$W_1(m_x^r, m_y^r) = d(x,y)\left(1 - \frac{r^2\,\mathrm{Ric}(v,v)}{2(n+2)} + O(r^3 + r^2 d(x,y))\right), \tag{7.3}$$

or equivalently, the coarse Ricci curvature satisfies

$$\kappa(x,y) = \frac{r^2\,\mathrm{Ric}(v,v)}{2(n+2)} + O(r^3 + r^2 d(x,y)).$$

The idea to prove the statement in Example 7.4 is to average (7.2) over all w in the tangent space T_xM and over $\varepsilon \in (0, r]$, and then derive the $\leq$ inequality of (7.3). The converse inequality is proved by employing the Kantorovich duality. We omit the proof of this example which can be found in [Oll09, Section 8], and we only provide Ollivier's proof of the above proposition.

Proof of Proposition 7.3. Let $c : [0, \delta] \to M$ be the geodesic $c(s) := \exp_x(sv)$, and let $v_s := c'(s) \in T_{c(s)}M$. For each s, let $w_s \in T_{c(s)}M$ obtained by the parallel transport of w along c. Thus $\langle v_s, w_s\rangle$ is constant in $s \in [0, \delta]$. Moreover, $v_0 = v$, $w_0 = w$, and $w_\delta = w'$. Consider $F : [0, \delta] \times [0, \varepsilon] \to M$ a geodesic variation defined by

$$F(s, t) := c_s(t) := \exp_{c(s)}(tw_s).$$

For each fixed s_0, let J_{s_0} be the variational vector field associated to F of the geodesic c_{s_0}, that is,

$$\frac{\partial}{\partial s}\bigg|_{s=s_0} F(s, t) = J_{s_0}(t).$$

Hence J_{s_0} is a Jacobi field along c_{s_0} and satisfies the Jacobi equation

$$J''_{s_0}(t) + R(c'_{s_0}(t), J_{s_0}(t))c'_{s_0}(t) = 0$$

where $J'_{s_0}(t) = \frac{D}{dt} J_{s_0}(t) = \nabla_{c'_{s_0}(t)} J_{s_0}$ and $J''_{s_0}(t) = \frac{D^2}{dt^2} J_{s_0}(t) = \nabla_{c'_{s_0}(t)}(\nabla_{c'_{s_0}(t)} J_{s_0})$.

Let $\gamma : [0, \delta] \to M$ be the curve $\gamma(s) := c_s(\varepsilon) = \exp_{c(s)}(\varepsilon w_s)$ from $\gamma(0) = x'$ to $\gamma(\delta) = y'$. We aim to compute the length of γ, which in turn gives an upper bound for the distance $d(x', y')$. We have

$$\gamma'(s_0) = \frac{d}{ds}\bigg|_{s_0} F(s, \varepsilon) = J_{s_0}(\varepsilon).$$

The Taylor's expansion of $f(t) = \|J_{s_0}(t)\|^2$ is given by

$$\|\gamma'(s_0)\|^2 = f(\varepsilon) = f(0) + \varepsilon f'(0) + \frac{\varepsilon^2}{2} f''(0) + O(\varepsilon^3),$$

where $f(0)$, $f'(0)$, and $f''(0)$ are computed as

1. $f(0) = \|J_{s_0}(0)\|^2 = \|c'(s_0)\|^2 = 1.$

2. $f'(0) = \frac{d}{dt}\Big|_{t=0} \langle J_{s_0}(t), J_{s_0}(t)\rangle = 2\langle J'_{s_0}(0), J_{s_0}(t)\rangle = 0$ because

$$J'_{s_0}(0) = \frac{D}{dt}\bigg|_{t=0} J_{s_0}(t) = \frac{D}{dt}\bigg|_{t=0} \frac{d}{ds}\bigg|_{s_0} F(s, t)$$

$$= \frac{D}{ds}\bigg|_{s_0} \frac{d}{dt}\bigg|_{t=0} F(s,t) \qquad \text{(symmetry lemma)}$$
$$= \frac{D}{ds}\bigg|_{s_0} w_s = 0 \qquad (w_s \text{ is parallel along } c).$$

3. $\frac{1}{2}f''(0) = \|J'_{s_0}(0)\|^2 + \langle J_{s_0}(0), J''_{s_0}(0)\rangle$
$= 0 - \langle J_{s_0}(0), R(c'_{s_0}(0), J_{s_0}(0))c'_{s_0}(0)\rangle$ (Jacobi equation)
$= -\langle v_{s_0}, R(w_{s_0}, v_{s_0})w_{s_0}\rangle$
$= -\langle R(w,v)w, v\rangle + O(\delta) \qquad \forall s_0 \in [0,\delta]$,
where the last equality holds true by a linear approximation of a continuously differentiable function $A(s) := \langle v_s, R(w_s, v_s)w_s\rangle$ around $s = 0$.

Thus

$$\begin{aligned}
\|\gamma'(s_0)\|^2 = f(\varepsilon) &= f(0) + \varepsilon f'(0) + \frac{\varepsilon^2}{2} f''(0) + O(\varepsilon^3) \\
&= 1 - \varepsilon^2 \langle R(w,v)w, v\rangle + O(\varepsilon^3 + \varepsilon^2\delta) \\
&\leq 1 - \varepsilon^2 K(v,w) + O(\varepsilon^3 + \varepsilon^2\delta),
\end{aligned}$$

for all $s_0 \in [0,\delta]$. Finally, we obtain the desired upper bound of $d(x', y')$ via the length of γ:

$$d(x', y') \leq \int_0^\delta \|\gamma'(s)\|^2 ds \leq \delta(1 - \varepsilon^2 K(v,w) + O(\varepsilon^3 + \varepsilon^2\delta)).$$

□

7.2 W_1-contraction property of random walks

Given a probability measure $\mu \in \mathscr{P}_1(X)$, the image of μ by a random walk m is a measure $\mu * m$ defined via the following convolution with m:

$$\mu * m := \int_{x \in X} d\mu(x) m_x.$$

One can also see from the definition that $\delta_x * m = m_x$. Moreover, a measure obtaind after n-step random walk m starting from μ is given by the iteration $\mu * m^{*n} := (\mu * m^{*n-1}) * m$. A ***stationary distribution*** (or an ***invariant distribution***) is a probability measure π such that $\pi * m = \pi$.

Next we present the so-called Bubley-Dyer theorem [BD97] about W_1 contraction property. Here the result is reformulated in the context of spaces with coarse Ricci curvature bound, [Oll09, Proposition 20].

Theorem 7.5. *Let (X, d, m) be a metric space with a random walk m. Let $\kappa \in \mathbb{R}$. Then we have $\kappa(x, y) \geq \kappa$ for all $x, y \in X$ if and only if for any two probability distributions μ, ν one has*

$$W_1(\mu * m, \nu * m) \leq (1 - \kappa) W_1(\mu, \nu).$$

Moreover, $\mu * m \in \mathscr{P}_1(X)$.

An important consequence of the above theorem is the exponential convergence of multi-step random walk.

Corollary 7.6. *Assume $\kappa(x, y) \geq \kappa$ for all $x, y \in X$. For $\mu, \nu \in \mathscr{P}_1(X)$,*

$$W_1(\mu * m^{*n}, \nu * m^{*n}) \leq (1 - \kappa)^n W_1(\mu, \nu). \tag{7.4}$$

Consequently, if $\kappa > 0$, then the random walk has a unique stationary distribution $\mu_0 \in \mathscr{P}_1(X)$. In fact, for all $\mu \in \mathscr{P}_1(X)$, the n-step random walk $\mu * m^{*n}$ converges in W_1 exponentially to μ_0, i.e.,

$$W_1(\mu * m^{*n}, \mu_0) \leq (1 - \kappa)^n W_1(\mu, \mu_0) \tag{7.5}$$

We would like to remark, in comparison to the W_1-contraction property of random walks for Ollivier's coarse Ricci curvature, the following continous-time W_r-contraction of the heat semigroup for Ricci curvature on Riemannian manifolds; see [vRS05, Corollary 1.4] and also [DS08, Formulae (2.3) and (2.4)] for the proof which uses the fact that heat flow is the gradient flow of entropy.

Theorem 7.7. *Let (M, g) be a Riemannian manifold with $\mathrm{Ric}(M) \geq K$. Let $f d\,\mathrm{vol}, g d\,\mathrm{vol} \in \mathscr{P}(M)$ be two absolutely continuous measures with density function $f, g \in C^\infty(M)$. Then for any $r \in [1, \infty]$,*

$$W_r(P_t f d\,\mathrm{vol}, P_t g d\,\mathrm{vol}) \leq e^{-Kt} W_r(f d\,\mathrm{vol}, g d\,\mathrm{vol}), \tag{7.6}$$

where P_t denotes the heat semigroup operator.

This exponential convergence of random walks in case of positive coarse Ricci curvature plays a role of the spectral gap. The following proposition from [Oll09, Proposition 30] provides an analogous result to Lichnerowicz theorem (c.f. Theorem 5.8) in the sense of coarse Ricci curvature.

Definition 7.8 (averaging operator)**.** An ***averaging operator*** M is defined to be

$$Mf(x) := \int_{y\in X} f(y)dm_x(y),$$

and the associated random walk Laplacian is $\Delta_M := \mathrm{Id} - M$.

Proposition 7.9 (spectral gap for coarse Ricci curvature)**.** *Let (X, d, m) be a metric space with random walk m, with invariant distribution π. Suppose that the coarse Ricci curvature of X is at least $\kappa > 0$ and that $\sigma < \infty$. Suppose that φ is reversible, or that X is finite. Then the first nonzero eigenvalue λ_1 of the Laplacian Δ_M satisfies*

$$\lambda_1(\Delta_M) \geq \kappa.$$

Chapter 8

Original proof of Ricci curvature via displacement convexity of entropy

The goal of this chapter is to explain the original proofs by von Renesse and Sturm [vRS05] and Cordero-Erausquin, McCann and Schmuckenschläger [CEMS01] that the lower Ricci curvature bound can be characterized via the displacement convexity on the entropy functional.

To simplify matters, we assume here that all objects are smooth and we ignore conjugate points and cut loci. Technical details can be found in the original papers.

8.1 Differential of exponential maps, Hessian of squared distance and Jacobi fields

Let us start with a technical result about the differential of exponential maps and the Hessian of the squared distance function. This result and its corollary about a particular Jacobi field will help to clarify some proof ideas from [CEMS01] in the later section.

Theorem 8.1. *Let M be a complete manifold with no conjugate points and let $u \in T_xM$ be a unit vector. Let $x_s := \exp_x(su)$ and let $Z : [0,1] \to TM$ be a vector field along the geodesic x_s. We introduce the curve z_s in M given by*

$$z_s := \exp_{x_s}(Z(s)),$$

and denote $z := z_0 = \exp_x(Z(0))$. Then

$$\dot{z}_{s=0} = D\exp_x(Z(0))\left(\mathrm{Hess}_x \frac{d_z^2}{2}(u) + \frac{D}{ds}\Big|_{s=0} Z\right).$$

Here we recall that $d_p(\cdot) := d(p, \cdot)$ is the distance function from a point $p \in M$. Moreover, instead of viewing the Hessian $\mathrm{Hess}_x f$ as symmetric bilinear form (c.f. Definition 5.3), we can also consider it as a self-adjoint linear operator $\mathrm{Hess}_x f : T_xM \to T_xM$ defined via

$$\langle \mathrm{Hess}_x f(v), w\rangle = \mathrm{Hess}\, f(v, w),$$

for all $v, w \in T_xM$.

Proof. We consider $z_s = \exp(Z(s))$ for the exponential map $\exp : TM \to M$ on tangent bundle, whose derivative at $w \in T_xM$ is the map $D\exp(w) : T_wTM \to T_wM$. Let us recall the canonical identification $T_wTM \cong T_xM \times T_xM$ for any $w \in T_xM$, which is explained as follows.

Any $\xi \in T_wTM$ can be represented by $\xi = \dot{C}(0) = \frac{d}{ds}|_{s=0}\, C(s)$ where $C : [0, 1] \to TM$ is a curve in TM such that $C(0) = w$. Let $c : [0, 1] \to M$ be the footpoint curve of C, that is, $C(s) \in T_{c(s)}M$ and $c(0) = x$. In other words, $C : [0, 1] \to TM$ is a vector field along c. Then the identification is given by

$$\begin{aligned} T_wTM \ni \xi = \frac{d}{ds}\Big|_{s=0} C(s) &= \left(\frac{d}{ds}\Big|_{s=0} c(s), \frac{D}{ds}\Big|_{s=0} C(s)\right) \\ &= \left(\dot{c}(0), \frac{D}{ds}\Big|_{s=0} C(s)\right) \in T_xM \times T_xM, \end{aligned} \tag{8.1}$$

where $\dot{c}(0) \in T_xM$ and $\frac{D}{ds}|_{s=0}C(s) \in T_xM$ are called ***horizontal*** and ***vertical*** components of $\xi = \dot{C}(0) \in T_wTM$, respectively.

Now we differentiate the curve $s \mapsto z_s = \exp(Z(s))$ at $s = 0$ and use the chain rule to obtain

$$T_zM \ni \frac{d}{ds}\Big|_{s=0} z_s = D\exp(Z(0)) \left(\frac{d}{ds}\Big|_{s=0} Z(s)\right) \tag{8.2}$$

The trick is to write $Z(s) = \tilde{Z}(s) + W(s)$ where $\tilde{Z}(s) = -\,\mathrm{grad}\, \frac{d_z^2}{2}(x_s)$, and note that the curve given by $s \mapsto \tilde{z}_s := \exp \tilde{Z}(s)$ is indeed a point curve since

$$\tilde{z}_s = \exp \tilde{Z}(s) = \exp_{x_s}\left(-\,\mathrm{grad}\, \frac{d_z^2}{2}(x_s)\right) = \exp_{x_s}\left(-d_z(x_s)\,\mathrm{grad}\, d_z(x_s)\right) = z.$$

It follows that

$$T_zM \ni 0_z = \frac{d}{ds}\Big|_{s=0} \tilde{z}_s = D\exp(\tilde{Z}(0)) \left(\frac{d}{ds}\Big|_{s=0} \tilde{Z}(s)\right). \tag{8.3}$$

Moreover, the fact that $\tilde{z}_0 = z_0$ implies $Z(0) = \tilde{Z}(0)$ since exponential maps are bijective. Both $\frac{d}{ds}|_{s=0}\, Z(s)$ and $\frac{d}{ds}|_{s=0}\, \tilde{Z}(s)$ have the same horizontal component, namely $\dot{x}_{s=0}$, so the subtraction (8.3) from (8.2) yields only the vertical component:

$$\begin{aligned}
\dot{z}_{s=0} &= D\exp(Z(0))\left(\frac{d}{ds}\Big|_{s=0} Z(s) - \frac{d}{ds}\Big|_{s=0} \tilde{Z}(s)\right) \\
&= D\exp(Z(0))\left(\dot{x}_{s=0} - \dot{x}_{s=0}, \frac{D}{ds}\Big|_{s=0}(Z(s) - \tilde{Z}(s))\right) \\
&= D\exp(Z(0))\left(0_x, \frac{D}{ds}\Big|_{s=0}(Z(s) - \tilde{Z}(s))\right) \\
&= D\exp_x(Z(0))\left(\frac{D}{ds}\Big|_{s=0}(Z(s) - \tilde{Z}(s))\right) \\
&= D\exp_x(Z(0))\left(\frac{D}{ds}\Big|_{s=0} Z(s) + \operatorname{Hess}_x \frac{d_z^2}{2}(u)\right).
\end{aligned}$$

□

The above theorem has the following consequence.

Corollary 8.2. *Let $u \in T_xM$ be a unit vector and $c(s) = \exp_x(su)$. Let $v \in T_xM$ be arbitrary and let $\gamma : [0,1] \to M$ denote the geodesic $\gamma(x) := \exp_x(tv)$ with $y = \gamma(1)$. Consider a geodesic variation given by*

$$\gamma_s(t) := \exp_{c(s)}\left(-t \operatorname{grad} \frac{d_y^2}{2}(c(s))\right),$$

where $d_y(\cdot) := d(y, \cdot)$ denotes the distance function from y. Then γ_s is a geodesic variation of γ, that is, $\gamma_0(t) = \gamma(t)$. Moreover, define

$$V(t) := \frac{\partial}{\partial s}\Big|_{s=0} \gamma_s(t)$$

to be the corresponding variational vector field, which is a Jacobi field along γ. Then we have

$$V(t) = D\exp_x(tv)\left(\operatorname{Hess}_x \frac{d_{\gamma(t)}^2}{2}(u) - t \operatorname{Hess}_x \frac{d_y^2}{2}(u)\right).$$

Proof. First, it is straightforward to check that $\gamma_0(t) = \gamma(t)$:

$$\gamma_0(t) = \exp_x\left(-t \operatorname{grad} \frac{d_y^2}{2}(x)\right) = \exp_x\left(-t d_y(x) \operatorname{grad} d_y(x)\right) = \gamma(t),$$

since $\operatorname{grad} d_y(x) = -\frac{\gamma'(0)}{d_y(x)}$.

Next, we aim to apply Theorem 8.1, so we start with the same setup by letting $x_s := \exp_x(su) = c(s)$. For a fixed t, we choose a vector field $Z : [0,1] \to TM$ along the geodesic $c(s)$ to be

$$Z(s) := -t \operatorname{grad} \frac{d_y^2}{2}(c(s)).$$

Moreover, let $z_s := \exp_{x_s}(Z(s)) = \gamma_s(t)$ and $z := z_0 = \gamma(t)$. We note that $c(0) = x$ and $c'(0) = u$, so we derive

$$\frac{D}{ds}\Big|_{s=0} Z = -t\nabla_{c'(0)} \operatorname{grad} \frac{d_y^2}{2} = -t \operatorname{Hess}_x \frac{d_y^2}{2}(u). \tag{8.4}$$

Applying Theorem 8.1, we obtain

$$\begin{aligned} V(t) = \frac{\partial}{\partial s}\Big|_{s=0} \gamma_s(t) = \dot{z}_{s=0} \ &= \ D\exp_x(Z(0))\left(\operatorname{Hess}_x \frac{d_z^2}{2}(u) + \frac{D}{ds}\Big|_{s=0} Z\right) \\ &\overset{(8.4)}{=} D\exp_x(Z(0))\left(\operatorname{Hess}_x \frac{d_z^2}{2}(u) - t\operatorname{Hess}_x \frac{d_y^2}{2}(u)\right). \end{aligned}$$

□

8.2 Volume distortion coefficients and Jacobi fields

In this section, we survey two important results in [CEMS01]. The first one describes the volume distortion coefficient in terms of the differential of exponential maps and the Hessian of squared distance function. The second one then uses these coefficients to give an estimation of Jacobian determinant of the interpolating map.

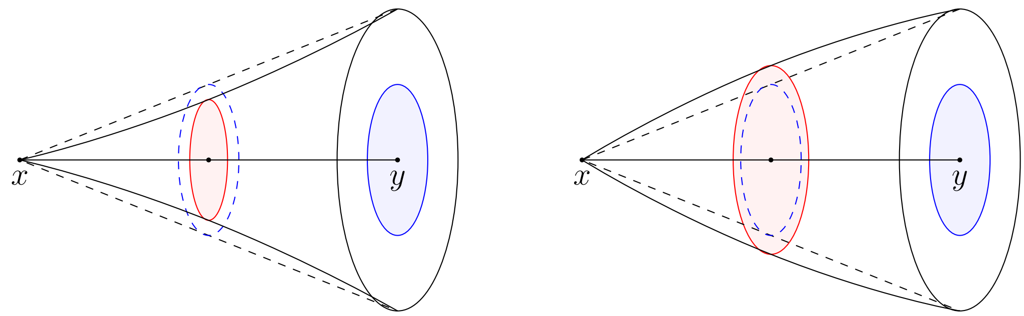
Throughout this section, let M be a smooth connected complete Riemannian manifold of dimension n.

Definition 8.3. For $x \in M$, $Y \subset M$ and $t \in [0,1]$, define the following locus of points

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

Let $B_r(y) \subset M$ be the open ball of radius r centered at $y \in M$. The ***volume distortion coefficient*** between x and y is defined as

$$v_t(x,y) := \lim_{r\to 0} \frac{\operatorname{vol}[Z_t(x,B_r(y))]}{\operatorname{vol}[B_{tr}(y)]} > 0. \tag{8.5}$$



(a) In negatively curved space, the coefficient is less than 1.

(b) In positively curved space, the coefficient is greater than 1.

Figure 8.1: Distortion due to curvature effects: the volume distortion coefficient is the ratio of the volume in red to the volume of in blue.

In the case that $\mathrm{Ric}(M) \geq K$, $K \in \mathbb{R}$, Bishop's comparison theorem [BC64] gives the following estimate of the volume distortion coefficient :

$$v_t(x, y) \geq \left(\frac{\mathtt{s}_K(td(x, y))}{\mathtt{s}_K(d(x, y))} \right)^{n-1}, \tag{8.6}$$

where $\mathtt{s}_K$ is defined by

$$\mathtt{s}_K(r) := \begin{cases} \dfrac{\sin\left(\sqrt{\frac{K}{n-1}} \cdot r\right)}{\sqrt{\frac{K}{n-1}} \cdot r} & \text{if K>0,} \\ 1 & \text{if K=0,} \\ \dfrac{\sinh\left(\sqrt{\frac{-K}{n-1}} \cdot r\right)}{\sqrt{\frac{-K}{n-1}} \cdot r} & \text{if K<0.} \end{cases}$$

Intuitively, the term $\mathtt{s}_K(r)^{n-1}$ is proportional to the area of the sphere of radius r in the n-dimensional space form of constant sectional curvature $K/(n-1)$ (and consequently Ricci curvature K).

Next we present the first result about volume distortion coefficients in terms of the determinants of certain Jacobi matrices.

Definition 8.4. Let $v \in T_x M$ and let $\gamma : [0, 1] \to M$ be the geodesic given by $\gamma(0) = x$ and $\gamma(t) := \exp_x(tv)$. A ***Jacobi matrix*** along γ is a map $A(t) : T_x M \to T_{\gamma(t)} M$ such that $A(t)u$ is a Jacobi field along γ for all $u \in T_x M$.

Remark 8.5. The Jacobi matrix $A(t)$ is uniquely determined by $A(0)$ and $A(1)$ if there are no conjugate points along γ. A well-known example of Jacobi matrices is $A(t) := tD\exp_x(tv)$ (see e.g. [GHL90]).

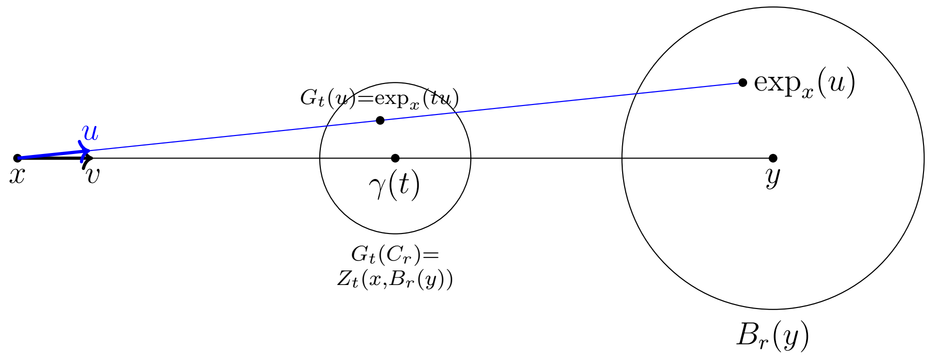


Figure 8.2: The diffeomorphic map $G_t : C_r \to Z_t(x, B_r(y))$.

Lemma 8.6 (Lemma 2.1 of [CEMS01])**.** *Fix $x, y \in M$ and let $\gamma(t) := \exp_x(tv)$ be the minimal geodesic joining $x = \gamma(0)$ and $y = \gamma(1)$. For $t \in [0,1]$, define $Y(t) := D\exp_x(tv)$ and $H(t) := \mathrm{Hess}_x \frac{d^2_{\gamma(t)}}{2}$. Then the volume distortion coefficient satisfies*

$$v_t(x,y) = \frac{\det Y(t)}{\det Y(1)} = \det Y(t)Y(1)^{-1} > 0, \tag{8.7}$$

and for $t \neq 1$,

$$v_{1-t}(y,x) = \det \frac{Y(t)\big(H(t) - tH(1)\big)}{1-t}. \tag{8.8}$$

Proof. For each $t \in [0,1]$, consider a map $G_t : T_xM \to M$ given by $G_t(u) = \exp_x(tu)$. Its differential at v is $DG_t(v) = tD\exp_x(tv) = tY(t)$. For small $r > 0$, we introduce the set $C_r := \{u \in T_xM \mid \exp_x(u) \in B_r(y)\}$ and note that G_t maps from C_r diffeomorphically to $Z_t(x, B_r(y))$, as illustrated in Figure 8.2.

The Jacobian transformation yields

$$\mathrm{vol}[Z_t(x, B_r(y))] = \int_{G_t(C_r)} 1 = \int_{u \in C_r} |\det DG_t(u)| = \mathrm{vol}(C_r) \cdot |\det DG_t(u')|$$

for some $u' \in C_r$ by the mean value theorem. As $r \to 0$, u' converges v, and we obtain

$$t^n \det Y(t) = \det(tY(t)) = \det(DG_t(v)) = \pm \lim_{r\to 0} \frac{\mathrm{vol}[Z_t(x, B_r(y))]}{\mathrm{vol}[C_r]},$$

where the sign is positive because $\det Y(t) \neq 0$ for all $t \in [0,1]$ and $Y(0) = \mathrm{Id}$. Since $Z_1(x, B_r(y)) = B_r(y)$, it follows that

$$\frac{\det Y(t)}{\det Y(1)} = \lim_{r\to 0} \frac{\mathrm{vol}[Z_t(x, B_r(y))]}{t^n \mathrm{vol}[B_r(y)]} = \lim_{r\to 0} \frac{\mathrm{vol}[Z_t(x, B_r(y))]}{\mathrm{vol}[B_{tr}(y)]} = v_t(x,y), \tag{8.9}$$

which proves the first half of the theorem.

Let $P_t^\gamma : T_xM \to T_{\gamma(t)}M$ be the parallel transport along γ (and note that $\det P_t^\gamma = 1$). Consider the unique Jacobi matrix $A(t) : T_xM \to T_{\gamma(t)}M$ along γ such that $A(0) = 0$ and $A(1) = P_1^\gamma$.

We note from Remark 8.5 that $tY(t) = tD\exp_x(tv) : T_xM \to T_{\gamma(t)}M$ is a Jacobi matrix along γ and that $Y(1)^{-1}P_1^\gamma \in \mathrm{End}(T_xM)$. Hence, the following map $A_\gamma(t) : T_xM \to T_{\gamma(t)}M$ defined as

$$A_\gamma(t) := tY(t)Y(1)^{-1}P_1^\gamma$$

is a Jacobi matrix along γ with $A_\gamma(0) = 0$ and $A_\gamma(1) = P_1^\gamma$.

In particular, it follows from (8.9) that, for $t \neq 0$,

$$v_t(x,y) = \frac{\det Y(t)}{\det Y(1)} = \det \frac{A_\gamma(t)}{t}. \tag{8.10}$$

Next, consider the geodesic $\gamma^- : [0,1] \to M$ from y to x given by $\gamma^-(t) := \gamma(1-t)$. We obtain similarly to (8.10) that, for $t \neq 1$,

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

where A_{γ^-} is a Jacobi matrix along γ^- with $A_{\gamma^-}(0) = 0$ and $A_{\gamma^-}(1) = (P_1^\gamma)^{-1}$.

Furthermore, define another Jacobi matrix $B(t) : T_xM \to T_{\gamma(t)}M$ along γ as $B(t) := A_{\gamma^-}(1-t)P_1^\gamma$. We have $B(0) = \mathrm{Id}$, $B(1) = 0$ and

$$v_{1-t}(y,x) = \det \frac{A_{\gamma^-}(1-t)}{1-t} = \det \frac{B(t)}{1-t}. \tag{8.11}$$

On the other hand, we define $\tilde{B}(t) := Y(t)(H(t) - tH(1)) : T_xM \to T_{\gamma(t)}M$ and recall that for all $u \in T_xM$,

$$\tilde{B}(t)(u) = D\exp_x(tv)\left(\mathrm{Hess}_x \frac{d_{\gamma(t)^2}}{2}(u) - t\,\mathrm{Hess}_x \frac{d_{y^2}}{2}(u)\right)$$

is a Jacobi field along γ due to Corollary 8.2. Thus $\tilde{B}(t)$ is a Jacobi matrix along γ, and it has boundary values $\tilde{B}(0) = \mathrm{Id}$ and $\tilde{B}(1) = 0$. Under the assumption of no conjugate points, we conclude that $B(t)$ coincides with $\tilde{B}(t)$, and (8.11) yields the desired inequality. □

To prepare for upcoming results, we start with the setup and notation given as follows. Let $\phi : M \to \mathbb{R}$ be a smooth function. For $t \in [0,1]$, define

$$\begin{aligned} F_t(x) &:= \exp_x(-t \operatorname{grad} \phi(x)), \\ Y(t) &:= D\exp_x(-t \operatorname{grad} \phi(x)), \\ H(t) &:= \operatorname{Hess}_x \frac{d^2_{F_t(x)}}{2}, \\ J_t(x) &:= \det DF_t(x). \end{aligned} \tag{8.12}$$

This function F_t plays a role of the optimal transport map as in McCann's theorem (Theorem 1.24).

Lemma 8.7. *For all $t \in [0,1]$, $H(t) - tH(1)$ and $H(1) - \operatorname{Hess}_x \phi$ are positive semidefinite.*

This lemma is described in more details in [CEMS01, Lemma 2.3 and Proposition 4.1(a)]

Proof. For any given $u \in T_xM$, let $x_s := \exp_x(su)$. By expressing the Hessian as the second derivative of the function evaluated along the representing curve, we have

$$\langle (H(t) - tH(1))u, u\rangle = \frac{d^2}{ds^2}\Big|_{s=0} \left(\frac{d^2_{F_t(x)}}{2} - t\frac{d^2_{F_1(x)}}{2} \right)(x_s).$$

To check that this term is nonnegative, it suffices to show that $\left(\frac{d^2_{F_t(x)}}{2} - t\frac{d^2_{F_1(x)}}{2} \right)(\cdot)$ is minimum at x, that is,

$$d^2(F_t(x), m) - td^2(F_1(x), m) \geq d^2(F_t(x), x) - td^2(F_1(x), x).$$

holds true for all $m \in M$.

For shortened notations, we denote $\ell_m := d(F_1(x), m)$ and $\ell := d(x, F_1(x))$, and we recall that $d(F_{t_1}(x), F_{t_2}(x)) = |t_1 - t_2|\ell$ for all t_1, t_2 since $F_\cdot(x)$ is a constant speed geodesic. We apply the triangle inequality and derive

$$\begin{aligned} d^2(F_t(x), m) - td^2(F_1(x), m) &\geq \big(d(F_1(x), m) - d(F_t(x), F_1(x))\big)^2 - td^2(F_1(x), m) \\ &= (1-t)\left(\ell_m^2 + (1-t)\ell^2 - 2\ell_m\ell\right) \\ &= (1-t)\left((\ell_m - \ell)^2 - t^2\ell^2\right) \\ &\geq -t(1-t)\ell^2 \end{aligned}$$

$$= d^2(F_t(x), x) - td^2(F_1(x), x),$$

as desired. Similarly, to see that $H(1) - \mathrm{Hess}_\phi$ is positive semidefinite, we need to show that $\left(\frac{d^2_{F_1(x)}}{2} - \phi\right)(\cdot)$ is locally minimum at x, that is, to show

$$\mathrm{grad}\, \frac{d^2_{F_t(x)}}{2}(x) - \mathrm{grad}\, \phi(x) = 0.$$

This holds true as we recall $F_1(x) = \exp_x(-\mathrm{grad}\, \phi(x))$ and deduce that

$$\mathrm{grad}\, \frac{d^2_{F_1(x)}}{2}(x) = d_{F_1(x)}(x) \cdot \mathrm{grad}\, d_{F_1(x)}(x) = |\,\mathrm{grad}\, \phi(x)| \cdot \frac{\mathrm{grad}\, \phi(x)}{|\,\mathrm{grad}\, \phi(x)|} = \mathrm{grad}\, \phi(x).$$

□

Now we are ready for the second result about an estimate of the Jacobian determinant in terms of volume distortion coefficients.

Lemma 8.8 (Lemma 6.1 of [CEMS01])**.** *Given the same setup as in* (8.12)*, The Jacobian determinant* $J_t(x) := \det DF_t(x)$ *can be identified as*

$$J_t(x) = \det Y(t)\big(H(t) - t\, \mathrm{Hess}_x\, \phi\big). \tag{8.13}$$

Moreover, it satisfies

$$J_t(x)^{\frac{1}{n}} \geq (1-t)v_{1-t}(F_1(x), x)^{\frac{1}{n}} + tv_t(x, F_1(x))^{\frac{1}{n}} J_1(x)^{\frac{1}{n}} \tag{8.14}$$

Proof of Lemma 8.8. First, we aim to apply Theorem 8.1 to verify the following identity:

$$DF_t(x) = Y(t)\big(H(t) - t\, \mathrm{Hess}_x\, \phi\big).$$

For any $u \in T_xM$, we follow most setup from Theorem 8.1: let $x_s := \exp_x(su)$ and choose $Z : [0,1] \to TM$ to be

$$Z(s) := -t\, \mathrm{grad}\, \phi(x_s) \in T_{x_s}M,$$

and let $z_s := \exp_{x_s}(Z(s))$ with $z := z_0 = \exp_x(-t\, \mathrm{grad}\, \phi(x)) = F_t(x)$.

As $x_0 = 0$ and $\dot{x}_{s=0} = u$, we have

$$DF_t(x)(u) = \frac{d}{ds}\Big|_{s=0} F_t(x_s) = \frac{d}{ds} \exp_{x_s} Z(s) = \dot{z}_{s=0}.$$

Applying Theorem 8.1 then yields

$$DF_t(x)(u) = \dot{z}_{s=0} = D\exp_x(Z(0)) \left(\mathrm{Hess}_x\, \frac{d_z^2}{2}(u) + \frac{D}{ds}\Big|_{s=0} Z\right)$$

$$= Y(t)\left(\mathrm{Hess}_x \frac{d^2_{F_t(x)}}{2}(u) - t\frac{D}{ds}\Big|_{s=0} \mathrm{grad}\,\phi(x_s)\right)$$
$$= Y(t)\left(H(t)(u) - t\,\mathrm{Hess}_x\,\phi(u)\right),$$

as desired, and (8.13) follows immediately.

To see (8.14), we write for $t \in [0,1)$,

$$J_t(x) = \det Y(t) \det\left[(1-t)\frac{H(t)-tH(1)}{1-t} + t\big(H(1) - \mathrm{Hess}_x\,\phi\big)\right],$$

and we recall from Lemma 8.7 that both matrices $H(t)-tH(1)$ and $H(1)-\mathrm{Hess}_x\,\phi$ are positive semidefinite.

Using the fact that $\det^{\frac{1}{n}}(\cdot)$ is a concave function over $n\times n$ positive semidefinite real symmetric matrices, and the fact that $\det Y(t) > 0$, we have

$$\begin{aligned} J_t(x)^{\frac{1}{n}} &\geq \det^{\frac{1}{n}} Y(t)\cdot\left[(1-t)\mathrm{det}^{\frac{1}{n}}\frac{H(t)-tH(1)}{1-t} + t\mathrm{det}^{\frac{1}{n}}\big(H(1)-\mathrm{Hess}_x\,\phi\big)\right] \\ &= (1-t)\mathrm{det}^{\frac{1}{n}}\frac{Y(t)(H(t)-tH(1))}{1-t} + t\mathrm{det}^{\frac{1}{n}}\big(Y(t)Y^{-1}(1)\big)\cdot J_1(x)^{\frac{1}{n}} \\ &= (1-t)v_{1-t}(F_1(x),x)^{\frac{1}{n}} + tv_t(x,F_1(x))^{\frac{1}{n}}J_1(x)^{\frac{1}{n}}, \end{aligned}$$

where the last equality is due to (8.7) and (8.8). □

8.3 Curvature equivalence on manifolds (revisited)

Now we are properly prepared to follow and understand the proof ideas of the main theorem (c.f. Theorem 3.3) as originally given in [CEMS01].

Theorem 8.9 (Theorem 1.1 of [vRS05])**.** *For any smooth complete Riemannian manifold M and any $K \in \mathbb{R}$, the following properties are equivalent:*

1. $\mathrm{Ric}(M) \geq K$, *that is* $\mathrm{Ric}_x(v) \geq K|v|^2$ *for all* $x \in M$, $v \in T_xM$.

2. *The entropy* $\mathcal{E}_{\mathrm{vol}}(\cdot)$ *is displacement* K*-convex on* $\mathscr{P}^2(M)$.

Original Proof of (i) ⇒ (ii). Assume that μ_0 and μ_1 are absolutely continuous with respect to volume measure. By McCann's theorem (Theorem 1.24), there exists a unique geodesic $(\mu_t)_{t\in[0,1]}$ in $\mathscr{P}^2(M)$ connecting μ_0 and μ_1, namely $\mu_t =$

$(F_t)_\# \mu_0$ where $F_t(x) = \exp_x(-t \operatorname{grad} \phi(x))$ with some function $\phi : M \to \mathbb{R}$. Here we recall that the pushforward $\mu_t = (F_t)_\# \mu_0$ means

$$\int_M b(y) d\mu_t(y) = \int_M b(F_t(x)) d\mu_0(x)$$

for all Borel functions $b : M \to \mathbb{R}$. Moreover, μ_t is absolutely continuous and can be written as $d\mu_t(x) = \rho_t(x) d\operatorname{vol}(x)$ with finite probability density $\rho_t(x)$. The Jacobian determinant $J_t(x) := \det DF_t(x)$ satisfies the Monge-Ampère equation (1.7):

$$\rho_0(x) = \rho_t(F_t(x)) J_t(x).$$

Applying the pushforward relation and the Monge-Ampère equation, we can derive the entropy term as

$$\begin{aligned}
\mathcal{E}_{\mathrm{vol}}(\mu_t) = \int_M \log(\rho_t(y)) d\mu_t(y) &= \int_M \log(\rho_t(F_t(x))) d\mu_0(x) \\
&= \int_M \log\left(\frac{\rho_0(x)}{J_t(x)}\right) d\mu_0(x) \\
&= \mathcal{E}_{\mathrm{vol}}(\mu_0) - \int_M \log J_t(x) d\mu_0(x).
\end{aligned}$$

Next we derive a lower estimate for the convex combination of entropy terms, namely

$$\mathcal{E}(t) := -\mathcal{E}_{\mathrm{vol}}(\mu_t) + (1-t)\mathcal{E}_{\mathrm{vol}}(\mu_0) + t\mathcal{E}_{\mathrm{vol}}(\mu_1),$$

by applying Lemma 8.8 and Bishop's comparison theorem as follows.

$$\begin{aligned}
\mathcal{E}(t) &= \int_M \log J_t(x) d\mu_0(x) - t \int_M \log J_1(x) d\mu_0(x) \\
&\overset{(8.14)}{\geq} n \int_M \log[(1-t) v_{1-t}(F_1(x), x)^{\frac{1}{n}} + t v_t(x, F_1(x))^{\frac{1}{n}} J_1(x)^{\frac{1}{n}}] d\mu_0(x) \\
&\quad - t \int_M \log J_1(x) d\mu_0(x) \\
&\geq \int_M \left[(1-t)\log v_{1-t}(F_1(x), x) + t \log\big(v_t(x, F_1(x)) J_1(x)\big)\right] d\mu_0(x)
\end{aligned}$$

$$
\begin{aligned}
& - t \int_M \log J_1(x) d\mu_0(x) \qquad\qquad \text{(using concavity of log)} \\
&= \int_M \left[(1-t)\log v_{1-t}(F_1(x), x) + t \log v_t(x, F_1(x)) \right] d\mu_0(x) \\
&\overset{(8.6)}{\geq} (n-1) \int_M [(1-t)\log \mathbf{s}((1-t)d(x,F_1(x))) + t\log \mathbf{s}(td(x,F_1(x))) \\
&\quad - \log \mathbf{s}(d(x,F_1(x)))] d\mu_0(x).
\end{aligned}
$$

In order conclude that $\mathcal{E}_{\mathrm{vol}}(\cdot)$ is displacement K-convex, we need to prove that

$$
\mathcal{E}(t) \geq \frac{K}{2} t(1-t) W_2^2(\mu_0, \mu_1) = \frac{K}{2} t(1-t) \int_M d^2(x, F_1(x)) d\mu_0(x).
$$

Compared to the lower estimate of $\mathcal{E}(t)$ derived earlier, we are left to show in general that for all $r > 0$,

$$
(1-t)\log \mathbf{s}((1-t)r) + t \log \mathbf{s}_K(tr) - \log \mathbf{s}_K(r) \geq \frac{t(1-t)}{2} \frac{K}{n-1} r^2,
$$

which can be rewritten as

$$
(1-t)\lambda((1-t)r) + t\lambda(tr) - \lambda(r) \geq 0,
$$

where $\lambda(r) := \log \mathbf{s}_K(r) + \frac{1}{6}\frac{K}{n-1} r^2$.

It suffices to show that $\lambda'(r) \leq 0$ because the monotonicity of λ would then imply $\lambda((1-t)r) \geq \lambda(r)$ and $\lambda(tr) \geq \lambda(r)$ and hence the desired inequality.

With the substitution $\sqrt{\frac{|K|}{n-1}} r = \mathbf{r}$, we obtain

$$
\lambda\left(\frac{n-1}{|K|}\mathbf{r}\right) = \begin{cases} \log \frac{\sin \mathbf{r}}{\mathbf{r}} + \frac{1}{6}\mathbf{r}^2 & \text{if } K > 0, \\ \log \frac{\sinh \mathbf{r}}{\mathbf{r}} - \frac{1}{6}\mathbf{r}^2 & \text{if } K < 0. \end{cases} \tag{8.15}
$$

To verify that its derivative is nonpositive, we need to check that

$$
\mathbf{r}\cos\mathbf{r} - \sin\mathbf{r} + \frac{1}{3}\mathbf{r}^2 \sin\mathbf{r} \quad \leq 0 \tag{8.16}
$$

in case $K > 0$, and that

$$
\mathbf{r}\cosh\mathbf{r} - \sinh\mathbf{r} + \frac{1}{3}\mathbf{r}^2 \sinh\mathbf{r} \quad \leq 0. \tag{8.17}
$$

For $K < 0$, by differentiating of (8.17) and dividing by $\frac{\mathbf{r}}{3}$, we need $\mathbf{r}\cos\mathbf{r} - \sin\mathbf{r} + \frac{1}{3}\mathbf{r}^2\sin\mathbf{r} \leq 0$. Again by differentiation, this follows from $-\mathbf{r}\sinh\mathbf{r} \leq 0$, which is also true for all $\mathbf{r} \leq 0$.

Similarly, in case $K > 0$, to check (8.16) requires that $-\mathbf{r}\sin\mathbf{r} \leq 0$. This is true when $\mathbf{r} \in [0, \pi]$, and such restriction is allowed due to Bonnet-Myers theorem. $\square$

Chapter 9

Overview of curvature notions on discrete spaces

In this chapter, we discuss brief motivations which give rise to the three notions of Ricci curvature on discrete spaces, namely, Ollivier Ricci curvature, Bakry-Émery curvature, and Erbar-Maas entropic Ricci curvature. Each of the following three parts of this thesis is focusing one of these curvature notions.

9.1 Setting of discrete spaces

Let us start with the standard convention for three different settings of discrete spaces in which these curvature notions will be discussed.

First is a ***(combinatorial) graph***, denoted by $G = (V, E)$, where V is the set of vertices (or nodes), and E is the set of undirected edges. Two vertices $x, y \in V$ are said to be adjacent (or neighbors) if there is an edge between x and y, that is, $\{x, y\} \in E$, and we denote this edge or this adjacency by $x \sim y$ (or $y \sim x$). For $x \in V$, the ***vertex degree*** of x, denoted by $\deg(x)$, is the number of neighbors of x. All graphs G are assumed to be simple (i.e., no loops nor multiple edges), connected and locally finite (i.e., every vertex has finite degree). The graph distance function $d : V \times V \to \mathbb{Z}_{\geq 0}$ is defined for a pair of vertices $x, y \in V$ to be the length (i.e., the number of edges) of a shortest path (also called a ***geodesic***) between x and y. The ***diameter*** of G is given by $\mathrm{diam}(G) := \sup_{x,y \in V} d(x, y) \in \mathbb{N} \cup \{\infty\}$. For $r \in \mathbb{N}$, the r***-sphere*** and the r***-ball*** centered at a vertex x consists of all vertices whose distance from x is equal to, and less than or equal to r, respectively:

$$S_r(x) := \{\, y \in V \mid d(x, y) = r \,\},$$

$$B_r(x) := \{ y \in V \mid d(x,y) \le r \}.$$

In particular, $S_1(x)$ is the set of all neighbors of x. An ***interval*** $[x,y]$ is the set of all vertices lying on some geodesic(s) between x and y, that is,

$$[x,y] = \{ z \in V \mid d(x,z) + d(z,y) = d(x,y) \}. \tag{9.1}$$

Second is a weighted graph, denoted by $G = (V, \mu, w)$, consisting of a vertex measure $\mu : V \to \mathbb{R}_+$ and a symmetric edge-weight function $w : V \times V \to \mathbb{R}_{\ge 0}$, where we write $u_x = \mu(x)$ and $w_{xy} := w(x,y)$. Two vertices x, y are adjacent if and only if $w(x,y) > 0$. Similarly to the case of graphs, all weighted graphs are also assumed to have no loops (that is, $w_{xx} = 0$ for all $x \in V$) and to be connected and locally finite (that is, for each $x \in V$, there exists finitely many $y \in V$ such that w_{xy} is nonzero). The definition of graph distance, geodesics, diameter, spheres and balls is carried from its underlying non-weighted graph. Furthermore, we define the transition rate from x to y as $p_{xy} := \frac{w_{xy}}{\mu_x}$ and define the weighted vertex degree as $\mathrm{Deg}(x) := \frac{1}{\mu_x} \sum_{y \in V} w_{xy} = \sum_{y \in V} p_{xy}$.

Third is a ***discrete Markov chain*** (X, Q, π), consists of a finite set of vertices X and a Markov kernel $Q : X \times X \to \mathbb{R}_{\ge 0}$ satisfying $\sum_{y \in X} Q(x,y) = 1$. We always assume Q to be irreducible and reversible (see further discussion in Chapter 20), and consequently there exists a unique stationary probability measure (also known as a steady state) $\pi : X \to \mathbb{R}_+$ satisfying $\sum_{x \in X} \pi(x) = 1$, $\sum_{x \in X} Q(x,y)\pi(x) = \pi(y)$ and the detailed balance equation $Q(x,y)\pi(x) = Q(y,x)\pi(y)$. The ***vertex*** π***-degree*** of x is defined as $\mathrm{Deg}_\pi(x) := \frac{1}{\pi(x)} \sum_{y \in S_1(x)} \pi(y)$.

Furthermore, a Markov chain (X, Q, π) can be regarded as a weighted graph $G = (X, w, \mu)$ by viewing $w_{xy} = Q(x,y)\pi(x) = w_{yx}$ as a symmetric edge-weight, and $\mu_x = \pi(x)$ as a vertex measure. The transition rate is then given by $p_{xy} := \frac{w_{xy}}{\mu_x} = Q(x,y)$, which is the transitional probability to go from x to y. Here we allow $Q(x,x)$ to possibly be nonzero (which means the Markov chain can have self-loops), but whether $Q(x,x) = 0$ is required or not has no effect on the associated Laplacian $\Delta f := \sum_y (f(y) - f(x)) Q(x,y)$ and the entropic Ricci curvature.

9.2 Lin-Lu-Yau's modification of Ollivier Ricci curvature for graphs

The notion of coarse Ricci curvature $\kappa(x,y) = 1 - \frac{W_1(\mathfrak{m}_x, \mathfrak{m}_y)}{d(x,y)}$, which was introduced by Ollivier, have been studied in the context of combinatorial graphs $G = (V, E)$ by Lin, Lu and Yau [LLY11]. They consider, for each vertex $x \in V$, the probability measure $\mathfrak{m}_x = \mu_x^\alpha$ to be obtained by a one-step lazy simple random walk from a vertex x with the probability $\alpha \in (0,1]$ to stay idle. More precisely, $\mu_x^\alpha(x) = \alpha$ and $\mu_x^\alpha(v) = (1-\alpha)/\deg(x)$ if v is a neighbor of x. Then the α-Ollivier Ricci curvature between two different vertices $x, y \in V$ is defined as

$$\kappa_\alpha(x,y) := 1 - \frac{W_1(\mu_x^\alpha, \mu_y^\alpha)}{d(x,y)}.$$

This curvature notion is the central focus of Part II.

9.3 Bakry-Émery curvature on graphs

Recall from Section 5.4 the following Bakry-Émery curvature-dimension condition on a Riemannian manifold $(M^n, \langle\cdot,\cdot\rangle)$. Given $K \in \mathbb{R}$ and $N \in [0,\infty]$, a point $x \in M$ satisfies $BE(K,N)$ if the following inequality holds for all $f \in C^\infty(M)$,

$$\Gamma_2 f(x) \geq \frac{1}{N}(\Delta f(x))^2 + K\Gamma f(x), \tag{9.2}$$

where

$$\begin{aligned} 2\Gamma(f,g) &:= \Delta(fg) - f\Delta g - g\Delta f, \\ 2\Gamma_2(f,g) &:= \Delta(\Gamma(f,g)) - \Gamma(f,\Delta g) - \Gamma(g,\Delta f), \end{aligned}$$

and $\Gamma f := \Gamma(f,f)$ and $\Gamma_2 f := \Gamma_2(f,f)$. The fact that the condition (9.2) and two iterations Γ and Γ_2 are defined via Δ allows one to generalize the Bakry-Émery curvature-dimension condition $BE(K,N)$ for any spaces admitting the Laplace operator. Early works by Elworthy [Elw91], by Schmuckenschlaäger [Sch99], and by Lin and Yau [LY10] study this Bakry-Émery curvature-dimension condition on weighted graphs $G = (V, w, \mu)$ with the graph Laplacian Δ defined as

$$\Delta f(x) := \frac{1}{\mu_x} \sum_{y \in V} w_{xy}.$$

In particular, a vertex $x \in V$ is said to satisfy $BE(K,N)$ if (9.2) holds for all functions $f : V \to \mathbb{R}$. Furthermore, the Bakry-Émery curvature at x, denoted as $\mathcal{K}(G,x;N)$, is the maximum K such that $BE(K,N)$ is satisfied at x. This curvature notion is the central focus of Part III.

9.4 Erbar-Maas' approach to adapt LSV curvature for discrete Markov chains

Recall the definition by Lott-Sturm-Villani (see Definition 3.4) which defines the lower Ricci curvature bound for a metric measure space (X, d, m) by the displacement convexity of the entropy functional. This definition requires firstly that there exists a geodesic connecting any two given probability measures; in other words, the Wasserstein space $\mathscr{P}_2(X)$ needs to be a geodesic space, which is the case for when X is geodesic space (see Theorem 1.20). On the other hand, when X is a discrete space, there is no constant speed geodesic joining any two distinct probability measures.

Proposition 9.1. *Let (X, d) be a discrete metric space. Then there is no constant speed geodesic $(\mu_t)_{t\in[0,1]}$ in the Wasserstein space $(\mathscr{P}_2(X), W_2)$, except trivial geodesics where μ_t is constant on t.*

Proof. Assume that $(\mu_t)_{t\in[0,1]}$ is a constant speed geodesic. Fix an arbitrary point $x \in X$. Since X is discrete, there exist $\delta > 0$ such that $\delta < d(x, y)$ for all $y \in X\backslash\{x\}$. For any transport plan $\pi \in \prod(\mu_s, \mu_t)$, its marginal constraints imply that $\pi(x, x) \leq \mu_t(x)$ and $\pi(x, x) \leq \mu_s(x)$, and that

$$\sum_{y\in X\backslash\{x\}} \pi(x, y) = \mu_s(x) - \pi(x, x) \geq \max\{\mu_s(x) - \mu_t(x),\ 0\};$$
$$\sum_{y\in X\backslash\{x\}} \pi(y, x) = \mu_t(x) - \pi(x, x) \geq \max\{\mu_t(x) - \mu_s(x),\ 0\}.$$

The quadratic cost of π then satisfies

$$\sum_{X\times X} d^2(u, v)\pi(u, v) > \delta^2 \sum_{y\in X\backslash\{x\}} (\pi(x, y) + \pi(y, x)) \geq \delta^2|\mu_s(x) - \mu_t(x)|.$$

The constant speed property of $(\mu_t)_{t\in[0,1]}$ then gives

$$|s - t|W_2(\mu_0, \mu_1) = W_2(\mu_s, \mu_t) \geq \delta\sqrt{|\mu_s(x) - \mu_t(x)|},$$

which implies $|\mu_s(x)-\mu_t(x)| \leq L|s-t|^2$ for all $s, t \in [0, 1]$, where $L := W_2(\mu_0, \mu_1)^2/\delta$ is a constant. By fixing s, one has the limit $\lim_{t\to s} \left|\frac{\mu_t(x)-\mu_s(x)}{t-s}\right| \leq \lim_{t\to s} L|s - t| = 0$, which means $t \mapsto \mu_t(x)$ is differentiable at $t = s$ and its derivative is zero. It follows that $\mu_t(x)$ is constant on t for every $x \in X$. □

There were two approaches to circumvent the lack of constant speed geodesics in the Wasserstein space $(\mathscr{P}_2(X), W_2)$ for discrete spaces X. The first approach by Bonciocat and Sturm [BS09] introduces "approximated points" along "rough geodesics". The second approach by Erbar and Maas [EM12], modifies W_2-metric into a new transport metric, called $\mathcal{W}$ by applying the concept of Otto's calculus and Benamou-Brenier formula in such a way that $(\mathscr{P}_2(X), \mathcal{W})$ is a geodesic space. This approach by Erbar and Maas gives rise the entropic Ricci curvature on discrete Markov chains, which is the central focus of Part IV.

Part II

Ollivier Ricci curvature on graphs

Chapter 10

W_1 on graphs

In this chapter, we briefly recall basic concepts of optimal transport, including the Kantorovich cost-minimizing transportation problem and its duality. These concepts were already discussed in Section 1.1, but we would like to present them here in the context of L^1-Wasserstein distance W_1 on graphs. Readers are encouraged to consult the book by Peyré and Cuturi [PC19, Chapter 6].

10.1 Cost-minimizing problem and dual problem

As usual, let $G = (V, E)$ be a simple, connected and locally finite graph and let

$$\mathscr{P}(V) := \left\{ \mu : V \to [0,1] \;\middle|\; \sum_{x \in V} \mu(x) = 1,\ |\operatorname{supp}(\mu)| < \infty \right\}$$

be the set of finitely supported probability measures on G. For $\mu, \nu \in \mathscr{P}(V)$, the 1-Wasserstein distance W_1 is defined via the following ***Kantorovich optimal transportation problem***:

$$W_1(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \left\{ \operatorname{cost}(\pi) := \sum_{x,y \in V} d(x,y)\pi(x,y) \right\}, \tag{10.1}$$

where the infimum runs over the set of all transport plans from μ to ν,

$$\Pi(\mu,\nu) := \left\{ \pi : V \times V \to [0,1] \;\middle|\; \begin{array}{ll} \sum_{y \in V} \pi(x,y) = \mu(x) & \forall x \in V \\ \sum_{x \in V} \pi(x,y) = \nu(y) & \forall y \in V \end{array} \right\}. \tag{10.2}$$

Note that the marginal constraints in (10.2) implies that $\mathrm{supp}(\pi) \subset \mathrm{supp}(\mu) \times \mathrm{supp}(\nu)$, and consequently

$$\sum_{y \in \mathrm{supp}(\nu)} \pi(x,y) = \mu(x) \qquad \text{and} \qquad \sum_{x \in \mathrm{supp}(\mu)} \pi(x,y) = \nu(y).$$

Any minimizer π to the problem (10.1) is called an ***optimal transport plan*** from μ to ν, and we denote by $\Pi_{opt}(\mu,\nu)$ the set of all such optimal transport plans. On the other hand, the 1-Wasserstein distance can also be reformulated via the following ***Kantorovich dual problem***:

$$W_1(\mu,\nu) = \sup_{\phi \in \mathrm{Lip}_1(V)} \sum_{x \in V} \phi(x)(\mu(x) - \nu(x)), \tag{10.3}$$

where $\mathrm{Lip}_1(V) := \{ \phi : V \to \mathbb{R} \mid \phi(x) - \phi(y) \le d(x,y) \ \forall x,y \}$ is the set of all 1-Lipschitz functions on V. Any maximizer ϕ to the problem (10.3) is called an ***optimal Kantorovich potential***.

The Existence of optimal plans and optimal Kantorovich potentials is asserted from a general optimal transport theory on metric spaces (see Theorem 1.10 and Remark 1.11). Alternatively, one can regard (10.1) as a standard finite-dimensional linear optimization problem with a variable $\pi = (\pi(x,y))_{\mathrm{supp}(\mu)\times\mathrm{supp}(\nu)} \ge 0$, whose minimizer always exists.

An important aspect of the Kantorovich duality is the relation between optimal transport plans and optimal Kantorocih potential, which can be described by the following complementary slackness theorem.

Theorem 10.1 (Complementary Slackness)**.** *Let $G = (V,E)$ be a graph. Suppose that π is an optimal transport plan and ϕ is an optimal Kantorovich potential from μ to ν in $\mathscr{P}(V)$. Then*

$$\phi(x) - \phi(y) = d(x,y) \quad \forall (x,y) \in \mathrm{supp}(\pi).$$

10.2 Triangle inequality and transport geodesics

It holds in general that for every $r \in [1,\infty)$ and a metric measure space (X,d), the Wasserstein distance W_r defines a metric on the finite-r-moment probability measure space $\mathscr{P}_r(X)$ (see Proposition 1.13). Here we decide to present a proof in our particular case of W_1 on graphs. Minkowski's inequality for L^1 spaces is simply the triangle-inequality, and the gluing lemma shares the same concept as the concatenation of transport plans, which we will introduce as follows.

Proposition 10.2. *Let $G = (V, E)$ be a graph. Then W_1 defines a metric on the space of probability measures $\mathscr{P}(V)$.*

Proposition 10.3 (concatenation of transport plans)**.** *Let $G = (V, E)$ be a graph and $\mu, \nu, \rho \in \mathscr{P}(V)$ be three probability measures. For any transport plans $\pi_1 \in \prod(\mu, \nu)$ and $\pi_2 \in \Pi(\nu, \rho)$, the **concatenation** of plans π_1 and π_2, denoted as $\pi_2 \circ \pi_1$, is defined to be the function $\pi_3 : V \times V \to [0, 1]$ such that*

$$\pi_3(x, z) := \sum_{y \in \text{supp}(\nu)} \frac{\pi_1(x, y)\pi_2(y, z)}{\nu(y)} \qquad \forall x, z \in V. \tag{10.4}$$

Then π_3 defines a transport plan from μ to ρ.

Moreover, if π_1, π_2 are given by transport maps $T_1, T_2 : V \to V$, then the concatenation $\pi_3 = \pi_2 \circ \pi_1$ is also given by a composite map $T_3 = T_2 \circ T_1$.

Proof of Proposition 10.3. We need to verify the marginal constraints for $\pi_3 \in \Pi(\mu, \rho)$. First, for all $x \in V$, we have

$$\begin{aligned}\sum_z \pi_3(x, z) &= \sum_z \sum_{y \in \text{supp}(\nu)} \frac{\pi_1(x, y)\pi_2(y, z)}{\nu(y)} \\ &= \sum_{y \in \text{supp}(\nu)} \frac{\pi_1(x, y)}{\nu(y)} \cdot \nu(y) = \sum_{y \in \text{supp}(\nu)} \pi_1(x, y) = \mu(x),\end{aligned}$$

where we apply the marginals $\sum_z \pi_2(y, z) = \nu(y)$ and $\sum_y \pi_1(x, y) = \mu(x)$. Similarly, it can be shown that $\sum_x \pi_3(x, z) = \rho(z)$, and hence $\pi_3 \in \Pi(\mu, \rho)$ as desired.

Moreover, if transport maps $T_1, T_2 : V \to V$ induces the plans π_1, π_2, namely,

$$\pi_1(x, y) = \begin{cases} \mu(x) & \text{if } y = T_1(x), \\ 0 & \text{otherwise} \end{cases}; \ \pi_2(y, z) = \begin{cases} \nu(y) & \text{if } z = T_2(y), \\ 0 & \text{otherwise} \end{cases}.$$

Then $\pi_3(x, z)$ defined via (10.4) is nonzero if and only if $\pi_1(x, y) > 0$ and $\pi_2(y, z) > 0$, which occurs exactly when $y = T_1(x)$ and $z = T_2(y)$. In this case, $\sum_z \pi_3(x, z) = \pi_3(x, T_2(T_1(x))) = \mu(x)$. Therefore,

$$\pi_3(x, z) = \begin{cases} \mu(x) & \text{if } z = T_2 \circ T_1(x), \\ 0 & \text{otherwise} \end{cases},$$

which means π_3 is induced from the composition $T_2 \circ T_1 : V \to V$. □

Now we are ready to prove that W_1 satisfies the three metric axioms.

Proof of Proposition 10.2. The symmetry property $W_1(\mu,\nu) = W_1(\nu,\mu)$ is obvious, since for every transport plan $\pi \in \Pi(\mu,\nu)$, the reverse plan $\pi^{(-1)} \in \Pi(\nu,\mu)$ defined by $\pi^{(-1)}(x,y) := \pi(y,x)$ has the same cost as π. The identity property $W_1(\mu,\nu) = 0 \ \Rightarrow \mu = \nu$ can also be proved easily by a contrapositive argument. We are left to check the triangle inequality, namely

$$W_1(\mu,\nu) + W_1(\nu,\rho) \geq W_1(\mu,\rho) \qquad \forall \mu,\nu,\rho \in \mathscr{P}(V).$$

Assume $\pi_1 \in \Pi(\mu,\nu)$ and $\pi_2 \in \Pi(\nu,\rho)$, and consider the concatenation $\pi_3 := \pi_2 \circ \pi_1 \in \Pi(\mu,\rho)$. As one would expect, the cost of π_3 is no more than that of π_1 and π_2 combined:

$$\begin{aligned}
\operatorname{cost}(\pi_3) &= \sum_{x,z} d(x,z)\pi_3(x,z) \\
&= \sum_{x,z} \sum_{y\in\operatorname{supp}(\nu)} d(x,z) \cdot \frac{\pi_1(x,y)\pi_2(y,z)}{\nu(y)} \\
&\overset{\triangle}{\leq} \sum_{x,z} \sum_{y\in\operatorname{supp}(\nu)} \Big(d(x,y) + d(y,z)\Big) \frac{\pi_1(x,y)\pi_2(y,z)}{\nu(y)} \\
&= \sum_{x} \sum_{y\in\operatorname{supp}(\nu)} d(x,y)\pi_1(x,y) + \sum_{z} \sum_{y\in\operatorname{supp}(\nu)} d(y,z)\pi_2(y,z) \\
&= \sum_{x,y} d(x,y)\pi_1(x,y) + \sum_{y,z} d(y,z)\pi_2(y,z) = \operatorname{cost}(\pi_1) + \operatorname{cost}(\pi_2).
\end{aligned} \tag{10.5}$$

By choosing π_1 and π_2 to be optimal transport plans, we conclude that

$$W_1(\mu,\rho) \leq \operatorname{cost}(\pi_3) \leq W_1(\mu,\nu) + W_1(\nu,\rho). \tag{10.6}$$

□

Remark 10.4. We may refer to the ***Wasserstein space*** $(\mathscr{P}(V), W_1)$ as the space of probability measures $\mathscr{P}(V)$ equipped with the Wasserstein metric W_1. Moreover, the underlying metric space (V,d) is embedded isometrically into the Wasserstein space $(\mathscr{P}(V), W_1)$ by $x \mapsto \delta_x$ since $W_1(\delta_x,\delta_y) = d(x,y)$ (compared to Definition 1.14 and Remark 1.16).

Corollary 10.5. *Given $\mu,\nu,\rho \in \mathscr{P}(V)$, suppose*

$$W_1(\mu,\rho) \leq W_1(\mu,\nu) + W_1(\nu,\rho).$$

Then for any optimal transport plans $\pi_1 \in \prod_{opt}(\mu,\nu)$ and $\pi_2 \in \prod_{opt}(\nu,\rho)$, we have

1. *the concatenation $\pi_3 = \pi_2 \circ \pi_1$ is optimal, and*

2. $d(x, y)+d(y, z) = d(x, z)$ *for all pairs* $(x, y) \in \operatorname{supp}(\pi_1)$ *and* $(y, z) \in \operatorname{supp}(\pi_2)$.

Proof of Corollary 10.5. The fact that $W_1(\mu, \rho) = W_1(\mu, \nu)+W_1(\nu, \rho)$ implies that the inequalities (10.6) hold with equality, which means π_3 is optimal. Moreover, the triangle inequality (10.5) holds with equality, so we must have $d(x, z) = d(x, y) + d(y, z)$ whenever $\pi_1(x, y) > 0$ and $\pi_2(y, z) > 0$. □

Moreover, we imitate the convention of ***intervals***, introduced as in (9.1), for Wasserstein space $(\mathscr{P}(V), W_1)$ and say that an interval

$$[\mu, \rho] := \{\nu \in \mathscr{P}(V)|\ W_1(\mu, \nu) + W_1(\nu, \rho) = W_1(\mu, \rho)\}$$

consists of all measures ν lying on a W_1-geodesic from μ to ρ. Then one can reformulate the above corollary and extend the result inductively as follows.

Corollary 10.6. *Suppose that $\mu_1, \mu_2, ..., \mu_n \in \mathscr{P}(V)$ lie orderly on a W_1-geodesic, that is,*

$$\sum_{i=1}^{n-1} W_1(\mu_i, \mu_{i+1}) = W_1(\mu_1, \mu_n).$$

For all $1 \le i \le n-1$, let $\pi_i \in \prod_{opt}(\mu_i, \mu_{i+1})$ be an optimal transport plan. Then the concatenation of the supports of $\pi_1, \pi_2, ..., \pi_n$ forms a geodesic on V, that is,

$$\sum_{i=1}^{n-1} d(x_i, x_{i+1}) = d(x_1, x_n)$$

for all pairs $(x_i, x_{i+1}) \in \operatorname{supp}(\pi_i)$.

Definition 10.7 (transport geodesics)**.** Suppose that $\mu_1, \mu_2, ..., \mu_n \in \mathscr{P}(V)$ lie orderly on a W_1-geodesic. Then any such sequence of vertices $x_1, x_2, ..., x_n$ (which may contain repetitions among consecutive vertices) from Corollary 10.6 is called a ***transport geodesic*** associated to the sequence $\mu_1, \mu_2, ..., \mu_n$.

Chapter 11

Ollivier Ricci curvature (ORC) with Lin-Lu-Yau modification

11.1 Relevant results about Ollivier Ricci curvature

Lin, Lu and Yau introduce in [LLY11] a convention for Ollivier Ricci curvature on a combinatorial graph by considering a "small ball" centered at a vertex x to be obtained after a one-step lazy simple random walk with idleness/laziness α from x.

Definition 11.1 ([LLY11])**.** Let $G = (V, E)$ be a locally finite graph. Let $\alpha \in [0, 1]$ be an ***idleness*** parameter. For any vertex $x \in V$, the probability measure $\mu_x^\alpha : V \to [0, 1]$ is given by

$$\mu_x^\alpha(v) := \begin{cases} \alpha & \text{if } v = x, \\ \frac{1-\alpha}{\deg(x)} & \text{if } v \in S_1(x), \\ 0 & \text{otherwise.} \end{cases}$$

The α***-Ollivier Ricci curvature*** between two different vertices $x, y \in V$ is given by

$$\kappa_\alpha(x, y) := 1 - \frac{W_1(\mu_x^\alpha, \mu_y^\alpha)}{d(x, y)},$$

and the ***Lin-Lu-Yau curvature*** is defined to be

$$\kappa(x, y) := \lim_{\alpha \to 1} \frac{\kappa_\alpha(x, y)}{1 - \alpha}.$$

The following lemma provides a useful fact that Ollivier Ricci curvature between two vertices with great distance apart can be bounded from below by the curvature between some pair of vertices which are closer.

Lemma 11.2. *Suppose that $y \in [x, z]$. Then for all $\alpha \in [0, 1]$, we have*

$$\kappa_\alpha(x, z) \geq \min\{\kappa_\alpha(x, y), \kappa_\alpha(y, z)\}.$$

Proof. Fix any $\alpha \in [0, 1]$, and let $K_\alpha := \min\{\kappa_\alpha(x, y), \kappa_\alpha(y, z)\}$. The triangle inequality $W_1(\mu_x^\alpha, \mu_z^\alpha) \leq W_1(\mu_x^\alpha, \mu_y^\alpha) + W_1(\mu_y^\alpha, \mu_z^\alpha)$ implies that

$$\begin{aligned}(1 - \kappa_\alpha(x, z))d(x, z) &\leq (1 - \kappa_\alpha(x, y))d(x, y) + (1 - \kappa_\alpha(y, z))d(y, z) \\ &\leq (1 - K_\alpha)(d(x, y) + d(y, z)) \\ &= (1 - K_\alpha)d(x, z),\end{aligned}$$

due to $y \in [x, z]$. We can then conclude that $\kappa_\alpha(x, z) \geq K_\alpha$ as desired. □

An immediate consequence of the above lemma is that the infimum of curvature among all pairs of different vertices is equal to the infimum of curvature between any two adjacent vertices. Therefore, when we discuss the lower bound of curvature on graphs, it makes sense to consider Ollivier Ricci curvature restrictively on edges of graphs.

Corollary 11.3. *For all $\alpha \in [0, 1]$, we have*

$$\inf_{x \neq y} \kappa_\alpha(x, y) = \inf_{x \sim y} \kappa_\alpha(x, y).$$

Example 11.4 (hypercube)**.** The hypercube $\mathcal{Q}^n$ can be viewed as the graph whose vertices are elements of $\{0, 1\}^n$, and two vertices are adjacent if and only if their Hamming distance (i.e., the number of coordinates at which the corresponding values are different) is one. Consider a vertex $x = (0, 0, ..., 0) = \bar{0}$ with n neighbors represented by $\{\, \mathrm{e}_i \mid 1 \leq i \leq n \,\}$, where e_i is the standard unit vector with a 1 in the i-th coordinate and 0's elsewhere. Let $y = \mathrm{e}_1$, and we aim to compute $\kappa_\alpha(x, y) = 1 - W_1(\mu_x^\alpha, \mu_y^\alpha)$, where the probability measures $\mu_x^\alpha, \mu_y^\alpha$ are

$$\mu_x^\alpha(z) = \begin{cases} \alpha, & \text{if } z = \bar{0}, \\ \frac{1-\alpha}{n}, & \text{if } z = \mathrm{e}_1, \\ \frac{1-\alpha}{n}, & \text{if } z = \mathrm{e}_i \\ & \text{for } 2 \leq i \leq n, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and } \mu_y^\alpha(z) = \begin{cases} \alpha, & \text{if } z = \mathrm{e}_1, \\ \frac{1-\alpha}{n}, & \text{if } z = \bar{0}, \\ \frac{1-\alpha}{n}, & \text{if } z = \mathrm{e}_1 + \mathrm{e}_i \\ & \text{for } 2 \leq i \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

Denote $m := \min\{\alpha, \frac{1-\alpha}{n}\} \in \mathbb{R}_{\geq 0}$. One may be convinced that to transport optimally from μ_x^α to μ_y^α, the masses can stay put at $\bar{0}$ and e_1 for at most m units each. The other masses of $1-2m$ units, collectively, need to move with the distance one (between $\bar{0}$ and e_1, and from e_i to $\mathrm{e}_1+\mathrm{e}_i$ for $i \geq 2$), so the total cost is $1-2m$. Explicitly, the corresponding transport plan $\pi \in \Pi(\mu_x^\alpha, \mu_y^\alpha)$ is given by $\pi(\bar{0},\bar{0}) = \pi(\mathrm{e}_1,\mathrm{e}_1) = m$, $\pi(\bar{0},\mathrm{e}_1) = \alpha - m$, $\pi(\mathrm{e}_1,\bar{0}) = \frac{1-\alpha}{n} - m$ and $\pi(\mathrm{e}_i,\mathrm{e}_i) = \frac{1-\alpha}{n}$ for all $2 \leq i \leq n$. This gives the total cost of π to be

$$\mathrm{cost}(\pi) = (\alpha - m) + \left(\frac{1-\alpha}{n} - m\right) + (n-1)\left(\frac{1-\alpha}{n}\right) = 1 - 2m,$$

which means $W_1(\mu_x^\alpha, \mu_y^\alpha) \leq 1-2m$.

To verify that $W_1(\mu_x^\alpha, \mu_y^\alpha) \geq 1-2m$, we need to employ the Kantorovich Duality. We construct a 1-Lipschitz function $\phi : V \to \mathbb{R}$ as follows:

$$\big(\phi(\bar{0}), \phi(\mathrm{e}_1)\big) = \begin{cases} (0,1) & \text{if } \alpha \in [0, \frac{1}{n+1}], \\ (1,0) & \text{if } \alpha \in [\frac{1}{n+1}, 1], \end{cases}$$

and $\phi(\mathrm{e}_i) = 1$ and $\phi(\mathrm{e}_1 + \mathrm{e}_i) = 0$ for all $2 \leq i \leq n$. So far, ϕ satisfies the 1-Lipschitz condition among vertices in $\mathrm{supp}(\mu_x^\alpha) \cup \mathrm{supp}(\mu_y^\alpha)$, and it can be 1-Lipschitz extended to all other vertices. Then we can deduce from the dual problem that

$$\begin{aligned} W_1(\mu_x^\alpha, \mu_y^\alpha) &\geq \sum_{z \in V} \phi(z)(\mu_x^\alpha(z) - \mu_y^\alpha(z)) \\ &= \frac{1-\alpha}{n}(n-1) + (\phi(\bar{0}) - \phi(\mathrm{e}_1))\left(\alpha - \frac{1-\alpha}{n}\right) \\ &= \begin{cases} 1 - 2\alpha & \text{if } \alpha \in [0, \frac{1}{n+1}], \\ 1 - \frac{2(1-\alpha)}{n} & \text{if } \alpha \in [\frac{1}{n+1}, 1] \end{cases} = 1 - 2m. \end{aligned}$$

Therefore, $W_1(\mu_x^\alpha, \mu_y^\alpha) = 1-2m$. In conclusion, the α-Ollivier Ricci curvature on an edge of the hypercube $\mathcal{Q}^n$ is given by

$$\kappa_\alpha(x,y) = 1 - W_1(\mu_x^\alpha, \mu_y^\alpha) = 2m = \begin{cases} 2\alpha & \text{if } \alpha \in [0, \frac{1}{n+1}], \\ \frac{2(1-\alpha)}{n} & \text{if } \alpha \in [\frac{1}{n+1}, 1], \end{cases}$$

and the Lin-Lu-Yau curvature is $\kappa(x,y) = \lim_{\alpha \to 1} \frac{\kappa_\alpha(x,y)}{1-\alpha} = \frac{2}{n}$.

Note that the non-lazy curvature κ_0 of a hypercube is zero. However, hypercubes (as the discrete counterpart of round spheres) are expected to have positive curvature, so the curvature with a nonzero idleness parameter α may serve as a better representation of curvature. Common choices of α are $\frac{1}{2}$, or $\frac{1}{d+1}$ in case of a d-regular graph. For readers' interest, general curvature notions for hypercubes are also discussed by Ollivier and Villani in [OV12].

In fact, Bourne et al. [BCL+18, Theorem 1.1 and Remark 5.4] discovers that for any graph, the α-Ollivier Ricci curvature on edges with large enough α is a linear function on α, and hence it only differs from the Lin-Lu-Yau curvature by a multiplication factor. This result is extended further for the curvature defined on any pair of different vertices; see [CK19, Corollary 3.4].

Theorem 11.5 (α-ORC and Lin-Lu-Yau)**.** *Let $G = (V, E)$ be a graph. Consider two different vertices $x, y \in V$ with vertex degrees* $\deg(x) = d_x$ *and* $\deg(y) = d_y$*.*

1. *If $x \sim y$, then $\kappa(x, y) = \frac{\kappa_\alpha(x,y)}{1-\alpha}$ for all $\alpha \in [\frac{1}{\max(d_x,d_y)+1}, 1)$. In particular, if $d_x = d_y = d$,*
$$\kappa(x, y) = 2\kappa_{\frac{1}{2}}(x, y) = \frac{d+1}{d}\kappa_{\frac{1}{d+1}}(x, y).$$

2. *In general, for any $x \neq y$, we have $\kappa(x, y) = \frac{\kappa_\alpha(x,y)}{1-\alpha}$ for all $\alpha \in [\frac{1}{2}, 1)$. In particular,*
$$\kappa(x, y) = 2\kappa_{\frac{1}{2}}(x, y). \tag{11.1}$$

From this relation between κ_α and κ, we see that Lemma 11.2 and Corollary 11.3 also hold with κ instead of κ_α.

Another important result is the computation of the curvature for the Cartesian product of two regular graphs. We refer [LLY11, Theorem 3.1] for a proof, and [BCL+18] for a generalized result on κ_α, and [CK19, Theorem 6.2] for an extended result to nonadjacent vertices.

Theorem 11.6 (Cartesian product)**.** *Let $G = (V_G, E_G)$ be a d_G-regular graph and $H = (V_H, E_H)$ be a d_H-regular graph. Let $x_1, x_2 \in V_G$ with $x_1 \sim x_2$ and $y_1, y_2 \in V_H$ with $y_1 \sim y_2$. Then*

$$\kappa^{G\times H}((x_1, y_1), (x_2, y_1)) = \frac{d_G}{d_G + d_H}\kappa^G(x_1, x_2),$$
$$\kappa^{G\times H}((x_1, y_1), (x_1, y_2)) = \frac{d_H}{d_G + d_H}\kappa^H(y_1, y_2).$$

Remark 11.7. By viewing the hypercube $\mathcal{Q}^n$ as the Cartesian products of n copies of the complete graph K_2, one may apply Theorem 11.6 inductively and deduce that the curvature of $\mathcal{Q}^n$ is $\kappa(x, y) = \frac{2}{n}$ for all $x \sim y$.

11.2 Normalized Laplacian and Lichnerowicz theorem

The ***normalized graph Laplacian*** Δ is defined on a graph $G = (V, E)$ as

$$\Delta f(x) := \frac{1}{\deg(x)} \sum_{z \in N(x)} (f(z) - f(x)) \qquad \forall f \in \mathbb{R}^V, \forall x \in V.$$

In case $G = (V, E)$ is a finite and connected graph, the Laplacian eigenvalues $\lambda \in \mathbb{R}$ solving $\Delta f + \lambda f = 0$ are known to be nonnegative real numbers and can be arranged in an increasing order with multiplicities: $0 = \lambda_0 < \lambda_1 \leq ... \leq \lambda_{n-1}$, where $n = |V|$. The trivial eigenvalue $\lambda_0 = 0$ corresponds to the constant eigenfunction, and the second smallest eigenvalue λ_1 is strictly positive as we assume the connectivity of the graph G. The next theorem is a discrete analogue to the Lichnerowicz spectral gap theorem on manifolds (c.f. Theorem 5.8 and Proposition 7.9). Here we follow a proof from [LLY11, Theorem 4.2].

Theorem 11.8 (Lichnerowicz for ORC)**.** *Let $G = (V, E)$ be a connected graph. Assume that $K := \inf_{x \neq y} \kappa(x, y) > 0$. Then the smallest nonzero eigenvalue satisfies*

$$\lambda_1 \geq K.$$

Proof. First, we note that G must be a finite graph due to the Bonnet-Myer diameter bound result (see Theorem 12.2). Fix $\alpha \in [\frac{1}{2}, 1)$, and recall from Theorem 11.5 that $\kappa(x, y) = \frac{\kappa_\alpha(x,y)}{1-\alpha}$ for all $x \neq y$. We define an average operator $M_\alpha : \mathbb{R}^V \to \mathbb{R}^V$ as $M_\alpha f(x) := \sum_{z \in V} \mu_x^\alpha(z) f(z)$, which is simplified to

$$M_\alpha f(x) = \alpha f(x) + \sum_{z \in S_1(x)} \frac{1-\alpha}{\deg(x)} f(z) = f(x) + (1-\alpha)\Delta f(x).$$

Let f_1 be an eigenfunction satisfying $\Delta f_1 + \lambda_1 f_1 = 0$. Then we have

$$M_\alpha f_1(x) = (1 - (1-\alpha)\lambda_1) f_1(x). \tag{11.2}$$

Consider the following Lipschitz constant of f_1, namely

$$\ell := \max_{x,y \in V} \frac{|f_1(x) - f_1(y)|}{d(x, y)},$$

and suppose that the maximum is attained at (x_1, y_1). Note that $\ell \neq 0$ since f_1 is not constant. We then have

$$\left| M_\alpha \left(\frac{f_1}{\ell}\right)(x_1) - M_\alpha \left(\frac{f_1}{\ell}\right)(y_1) \right| = |1-(1-\alpha)\lambda_1|\, d(x_1, y_1). \tag{11.3}$$

On the other hand, since $\frac{f_1}{\ell}$ is 1-Lipschitz, the Kantorovich duality yields

$$\begin{aligned} \left| M_\alpha \left(\frac{f_1}{\ell}\right)(x_1) - M_\alpha \left(\frac{f_1}{\ell}\right)(y_1) \right| &= \left| \sum_{z\in V} \frac{f(z)}{\ell} \big(\mu_{x_1}^\alpha(z) - \mu_{y_1}^\alpha(z)\big) \right| \\ &\leq W_1(\mu_{x_1}^\alpha, \mu_{y_1}^\alpha) \\ &= \big(1-\kappa_\alpha(x_1,y_1)\big) d(x,y) \end{aligned} \tag{11.4}$$

Comparing (11.4) to (11.3) gives $\lambda_1 \geq \frac{\kappa_\alpha(x_1,y_1)}{1-\alpha} = \kappa(x_1, y_1) \geq K$ as desired. □

Remark 11.9. For a hypercube $\mathcal{Q}^n$, the eigenvalues of the adjacency matrix $A_{\mathcal{Q}^n}$ are $(n-2i)$ with multipicities $\binom{n}{i}$ for all $0 \leq i \leq n$. The smallest nonzero Laplacian eigenvalue (corresponging to $n-2$ of $A_{\mathcal{Q}^n}$) is $\lambda_1 = \frac{2}{n}$, and hence the Lichnerowicz spectral bound is sharp. In fact, the class of all ***Lichnerowicz sharp*** graphs are much larger and not yet fully classified. For readers' interest, some partial classification result is provided in [CKK+20, Section 6].

Chapter 12

Bonnet-Myers diameter bound and Maximal diameter theorem

In this chapter, we present a Bonnet-Myers type diameter bound theorem for Ollivier Ricci curvature. The diameter bound result appears in Ollivier's original work [Oll09, Proposition 23] as well as in the paper by Lin-Lu-Yau [LLY11, Theorem 4.1], and its proof features a simple application of the triangle inequality. In the second half of this chapter, we introduce the concept of the maximal diameter, which will prepare us for the discussion of the rigidity result in the next chapter.

All results in this chapter are stated in terms of the Lin-Lu-Yau curvature κ, but most proofs use arguments based on κ_α due to the relation $\kappa_\alpha = (1-\alpha)\kappa$ for $\alpha \in [\frac{1}{2}, 1]$.

Lemma 12.1. *Let $G = (V, E)$ be a graph, and let $x, y \in V$ with $x \neq y$. Then*

$$\kappa(x, y) \leq \frac{2}{d(x, y)}.$$

Proof. We aim to prove that $\kappa_\alpha(x, y) \leq \frac{2(1-\alpha)}{d(x,y)}$ for any α.

Let $\ell := d(x, y)$ and denote a geodesic in G from x to y by $x_0 \sim x_1 \sim \cdots \sim x_\ell$, where $x_0 = x$ and $x_\ell = y$.

The triangle inequality for the metric W_1 gives

$$\begin{aligned} \ell = W_1(\delta_x, \delta_y) &\leq W_1(\delta_{x_0}, \mu_{x_0}^\alpha) + W_1(\mu_{x_0}^\alpha, \mu_{x_\ell}^\alpha) + W_1(\mu_{x_\ell}^\alpha, \delta_{x_\ell}) \\ &= (1-\alpha) + \ell(1 - \kappa_\alpha(x, y)) + (1-\alpha), \end{aligned} \tag{12.1}$$

where $W_1(\delta_v, \mu_v^\alpha) = 1 - \alpha$ for any $v \in V$ since the total mass of $1-\alpha$ unit is transported from v to its neighbors). □

Theorem 12.2 (Bonnet-Myers)**.** *If a locally finite graph $G = (V, E)$ has a positive α-Ollivier Ricci lower curvature bound for some $\alpha \in [0, 1)$, namely $K_\alpha := \inf_{x\sim y} \kappa_\alpha(x, y) > 0$, then the diameter of G is bounded above by*

$$\operatorname{diam}(G) \leq \frac{2(1-\alpha)}{K_\alpha}, \tag{12.2}$$

and hence G is finite.

Similarly, if G has a positive Lin-Lu-Yau lower curvature bound $K := \inf_{x\sim y} \kappa(x, y) > 0$, then

$$\operatorname{diam}(G) \leq \frac{2}{K}. \tag{12.3}$$

The short version of the proof follows from Lemma 12.1 by choosing $x, y \in V$ such that $d(x, y) = \operatorname{diam}(G)$ and using Corollary 11.3. However for later purposes involving transport geodesics, we provide another proof which is direct and does not use the Lemma 12.1.

Proof. For arbitrary $x, y \in V$, let $L := d(x, y)$ and denote a geodesic in G from x to y by $x_0 \sim x_1 \sim \cdots \sim x_L$, where $x_0 = x$ and $x_L = y$.

The triangle inequality gives

$$\begin{aligned} L = W_1(\delta_x, \delta_y) &\leq W_1(\delta_{x_0}, \mu^\alpha_{x_0}) + \sum_{i=0}^{L-1} W_1(\mu^\alpha_{x_i}, \mu^\alpha_{x_{i+1}}) + W_1(\mu^\alpha_{x_L}, \delta_{x_L}) \\ &\leq (1-\alpha) + L(1-K_\alpha) + (1-\alpha), \end{aligned} \tag{12.4}$$

where $W_1(\mu^\alpha_{x_i}, \mu^\alpha_{x_{i+1}}) = 1 - \kappa_\alpha(x_i, x_{i+1}) \leq 1 - K_\alpha$ for all i. Rearranging (12.4) yields $L \leq \frac{2(1-\alpha)}{K_\alpha}$, which means the diameter of G is no more than $\frac{2(1-\alpha)}{K_\alpha}$. A finite diameter bound for a locally finite graph G then implies that G is also finite.

The second statement of the theorem follows immediately due to the relation $\kappa = \frac{\kappa_\alpha}{1-\alpha}$ for $\alpha \in [\frac{1}{2}, 1)$. □

Definition 12.3. A graph $G = (V, E)$ is said to be **Bonnet-Myers sharp** if (12.3) holds with equality, that is,

$$\inf_{x\sim y} \kappa(x, y) = \frac{2}{\operatorname{diam}(G)}.$$

Equivalently, G is Bonnet-Myers sharp iff (12.2) holds with equality for some (and for all) $\alpha \in [\frac{1}{2}, 1)$, that is,

$$\inf_{x\sim y} \kappa_\alpha(x, y) = \frac{2(1-\alpha)}{\operatorname{diam}(G)}.$$

Moreover, a graph is called (D, L)-**Bonnet-Myers sharp** if it is D-regular and has the diameter L.

Hypercubes are examples of Bonnet-Myers sharp graphs. A complete classification of Bonnet-Myers sharp graphs is open, but we present in a full classification under the additional assumption of self-centeredness later in Chapter 13. Hypercubes were also examples of Lichnerowicz sharp graphs. In fact, regular Bonnet-Myers sharp graphs are Lichnerowicz sharp; see [CKK+20, Theorem 1.5].

An important characteristic of Bonnet-Myers sharp graphs is related to W_1-geodesics between the Dirac measures of the two opposite poles as explained in the following theorem.

Theorem 12.4. *Let $G = (V, E)$ be a Bonnet-Myers sharp graph and let $x, y \in V$ with $d(x, y) = \operatorname{diam}(G) =: L$. Assume that a sequence of vertices $x = v_0, v_1, v_2, ...,$ $v_{n-1}, v_n = y$ lie orderly on a geodesic in G. Then for any $\alpha \in [\frac{1}{2}, 1)$,*

- *(a) the sequence of probability measures $\delta_x, \mu_x^\alpha, \mu_{v_1}^\alpha, ..., \mu_{v_{n-1}}^\alpha, \mu_y^\alpha, \delta_y \in \mathscr{P}(V)$ also lie orderly on a W_1-geodesic, and*
- *(b) $\kappa_\alpha(v_i, v_j) = \frac{2(1-\alpha)}{L}$ and $\kappa(v_i, v_j) = \frac{2}{L}$ for all $v_i \neq v_j$.*

Additionally, if G is D-regular, then (a) and (b) also hold with $\alpha = \frac{1}{D+1}$.

Proof. Being Bonnet-Myers sharp means $\inf_{u \sim v} \kappa_\alpha(u, v) = \frac{2(1-\alpha)}{L}$ for all $\alpha \in [\frac{1}{2}, 1]$. In view of Corollary 11.3, it implies that for all $u, v \in V$ with $u \neq v$,

$$\frac{2(1-\alpha)}{L} \leq \kappa_\alpha(u, v) = 1 - \frac{W_1(\mu_u^\alpha, \mu_v^\alpha)}{d(u, v)},$$

that is, $W_1(\mu_u^\alpha, \mu_v^\alpha) \leq \frac{L-2(1-\alpha)}{L} \cdot d(u, v)$.

Recall also that $W_1(\delta_x, \mu_x^\alpha) = 1 - \alpha = W_1(\mu_y^\alpha, \delta_y)$. Therefore, by the triangle inequality on $(\mathscr{P}(V), W_1)$, we have

$$\begin{aligned} L = W_1(\delta_x, \delta_y) &\overset{\Delta}{\leq} W_1(\delta_x, \mu_x^\alpha) + \sum_{i=0}^{n-1} W_1(\mu_{v_i}^\alpha, \mu_{v_{i+1}}^\alpha) + W_1(\mu_y^\alpha, \delta_y) \\ &\leq 2(1-\alpha) + \frac{L - 2(1-\alpha)}{L} \sum_{i=0}^{n-1} d(v_i, v_{i+1}) = L, \end{aligned} \tag{12.5}$$

where the last equality comes from the assumption that $\sum_{i=0}^{n-1} d(v_i, v_{i+1}) = d(x, y) = L$. The equality of (12.5) implies that both involved inequalities must hold with equality. The first one means $\delta_x, \mu_x^\alpha, \mu_{v_1}^\alpha, ..., \mu_{v_{n-1}}^\alpha, \mu_y^\alpha, \delta_y$ lie orderly on a W_1-geodesic, which yields (a). The second one implies that $W_1(\mu_{v_i}^\alpha, \mu_{v_j}^\alpha) = \frac{L-2(1-\alpha)}{L} \cdot d(v_i, v_j)$ for any $v_i \neq v_j$. This means $\kappa_\alpha(v_i, v_j) = \frac{2(1-\alpha)}{L}$ and $\kappa(v_i, v_j) = \frac{2}{L}$, which yields (b). Additionally, if G is D-regular, then $\inf_{u\sim v} \kappa_\alpha(u, v) = \frac{2(1-\alpha)}{L}$ also holds for $\alpha = \frac{1}{D+1}$ due to Theorem 11.5. The statements (a) and (b) are therefore proved with the same arguments. □

Our approach to understand more the property of Bonnet-Myers sharp graphs is to apply the concept of transport geodesics (Corollary 10.6 and Definition 10.7) to the W_1-geodesics obtained from Theorem 12.4. This approach, which led us to our classification result, will be discuss in more details in the next chapter. Let us finish this chapter by presenting the following simple application of the transport geodesic concept: Bonnet-Myers sharp graphs must satisfy the "spherical suspension" property in the sense that every vertex must lie on a geodesic from any pair of two opposite poles (compared to Theorem 4.5). This property has already been proved by Matsumoto in his Master thesis [Mat10] written in Japanese, and it was proved alternatively with a Laplacian method in [CKK+20, Theorem 5.5] in the special case of regular graphs. Similar result was proved in the setting of directed graphs in [OSY20, Theorem 1.2].

Theorem 12.5 (maximal diameter)**.** *Let $G = (V, E)$ be a Bonnet-Myers sharp graph, i.e., $\inf_{x\sim y} \kappa(x, y) = \frac{2}{\mathrm{diam}(G)}$. Then for any $x, y \in V$ with $d(x, y) = \mathrm{diam}(G)$, we have $[x, y] = V$.*

Proof. By connectivity of G, it suffices to show that if a vertex v lies in the interval $[x, y]$ then any neighbor of v also lies in $[x, y]$.

Fix any $\alpha \in [\frac{1}{2}, 1)$. Suppose $v \in [x, y]$, that is, x, v, y lie orderly on a geodesic. Theorem 12.4 implies in particular that $\delta_x, \mu_v^\alpha, \delta_y$ lie orderly on a W_1-geodesic. Consider optimal transport plans $\pi_1 \in \Pi_{opt}(\delta_x, \mu_v^\alpha)$ and $\pi_2 \in \Pi_{opt}(\mu_v^\alpha, \delta_y)$.

For any neighbor of v, namely z (so $z \in \mathrm{supp}(\mu_v^\alpha)$), we know that $\pi_1(x, z) > 0$ and $\pi_2(z, y) > 0$. Corollary 10.6 then implies that x, z, y lie on a geodesic, i.e., $z \in [x, y]$ as desired. □

Chapter 13

Rigidity of Bonnet-Myers sharp graphs

In this chapter, we aim to present the main result in [CKK+20] to find all possible Bonnet-Myers sharp graphs, i.e., satisfying $\mathrm{diam}(G) = \frac{2}{\inf_{x\sim y}\kappa(x,y)}$ (under the additional assumption of self-centeredness).

We restrict our interest to Bonnet-Myers sharp graphs which are regular graphs (i.e., every vertex has the same degree). By considering a (D, L)-Bonnet-Myers sharp graph $G = (V, E)$, where $D = \deg(G)$ and $L := \mathrm{diam}(G)$, the most natural choice of the idleness parameter is $\alpha = \frac{1}{D+1}$, which gives the curvature on an edge $x \sim y$ as

$$\kappa(x, y) = \frac{D+1}{D}\kappa_\alpha(x, y) = \frac{D+1}{D}(1 - W_1(\mu_x, \mu_y)),$$

where here and henceforth $\mu_x := \mu_x^{\frac{1}{D+1}}$ is the probability measure uniformly distributed in the one-ball $B_1(x)$, that is,

$$\mu_x(z) = \begin{cases} \frac{1}{D+1}, & \text{if } z \in B_1(x), \\ 0 & \text{otherwise.} \end{cases}$$

Remark 13.1. Due to the fact that all individual masses in μ_x and μ_y are equal to $\frac{1}{D+1}$, there always exists an optimal transport plan $\pi \in \Pi_{opt}(\mu_x, \mu_y)$ without splitting masses. In other words, π is induced by a bijective ***optimal transport map*** $T : B_1(x) \to B_1(y)$, where for each $v \in B_1(x)$, the image $T(v)$ is the corresponding vertex in $B_1(y)$ such that $\pi(v, T(v)) = \frac{1}{D+1}$ (and $\pi(v, z) = 0$ for all vertices z other than $T(v)$). Moreover, we can assume without loss of generality that we do not need to move masses from origin that already lie in the destination; in other words, $T(v) = v$ for all $v \in B_1(x) \cap B_1(y)$ (see arguments in, e.g., [BCL+18, Lemma 4.1]).

13.1 Transport geodesics of a regular Bonnet-Myers sharp graph

Definition 13.2. Consider a (D, L)-Bonnet-Myers sharp graph $G = (V, E)$ and assume a full-length geodesic of G, namely,

$$g : x_0 \sim x_1 \sim x_2 \sim \cdots \sim x_L.$$

For every $1 \leq j \leq L$, consider an optimal transport map T_j from $\mu_{x_{j-1}}$ to μ_{x_j} satisfying the following properties:

(P1) $T_j : B_1(x_{j-1}) \to B_1(x_j)$ is a bijection.

(P2) $T_j(v) = v$ if and only if $v \in B_1(x_{j-1}) \cap B_1(x_j)$.

Such a map T_j is called a ***good*** optimal transport map. The existence of good maps T_j is explained in Remark 13.1 (but these maps are not necessarily unique a priori).

Moreover, define maps $T^j : B_1(x_0) \to B_1(x_j)$ as the composition

$$T^j := T_j \circ \cdots \circ T_1 \quad \forall 1 \leq j \leq L,$$

and T^0 is the identity map on $B_1(x_0)$ by convention. These maps T^j are also bijections. For each $z \in B_1(x_0)$, we define $z(0) := z$ and for $1 \leq j \leq L$,

$$z(j) := T^j(z) := T_j \circ \cdots \circ T_1(z) \in B_1(x_j).$$

Note in particular that $x_0(0) = x_0(1) = x_0$ by the condition (P2). One may define transport geodesics along the core geodesic g as follows.

Proposition 13.3. *Let $G = (V, E)$ be a (D, L)-Bonnet-Myers sharp graph with a full-length geodesic g and maps T_j and T^j defined as in Definition 13.2. Then for every $z \in B_1(x_0)$, the vertices $x_0, z(0), z(1), z(2), ..., z(L), x_L$ lie orderly on a geodesic, that is,*

$$L = d(x_0, x_L) = d(x_0, z(0)) + d(z(0), z(1)) + d(z(1), z(2)) + ... + d(z(L-1), z(L)) + d(z(L), x_L).$$

*This sequence of vertices $x_0, z(0), z(1), z(2), ..., z(L), x_L$ (which contains some repetitions among consecutive vertices) is called a **transport geodesic** along geodesic g.*

Proof of Proposition 13.3. We apply Theorem 12.4 in the case of D-regular Bonnet-Myers sharp graph and $\alpha = \frac{1}{D+1}$, and deduce that the sequence of probability measures $\delta_{x_0}, \mu_{x_0}, \mu_{x_1}, ..., \mu_{x_{L-1}}, \mu_{x_L}, \delta_{x_L} \in \mathscr{P}(V)$ lie orderly on a W_1-geodesic. Note also that along this sequence of probability measures, the mass of $\frac{1}{D+1}$ unit travels in the following chain of optimal transport plan/maps:

$$x_0 \mapsto z(0) \overset{T_1}{\mapsto} z(1) \overset{T_2}{\mapsto} z(2) \overset{T_3}{\mapsto} ... \overset{T_L}{\mapsto} z(L) \mapsto x_L.$$

In view of Theorem 10.6 and Definition 10.7, the sequence of vertices x_0, $z(0)$, $z(1)$, $z(2)$, ..., $z(L)$, x_L is then a transport geodesic associated to the probability measures δ_{x_0}, μ_{x_0}, μ_{x_1}, ..., $\mu_{x_{L-1}}$, μ_{x_L}, δ_{x_L}. □

Corollary 13.4. *For any $z \in B_1(x_0)$, we have $d(z, T_1(z)) \leq 1$.*

Proof. Proposition 13.3 asserts in particular that $x_0, z(0), z(1)$ lie on a geodesic; in other words, $d(x_0, T_1(z)) = d(x, z) + d(z, T_1(z))$ for all $z \in B_1(x_0)$. In the case of $z = x_0$, then $T_1(x_0) = x_0$, so $d(x_0, T_1(x_0)) = 0$ as desired. Otherwise, if $z \in S_1(x_0)$, then $d(z, T_1(z)) = d(x_0, T_1(z)) - 1$. Recall further that $T_1(z) \in B_1(x_1) \subset B_2(x_0)$, so $d(x_0, T_1(z)) \leq 2$. Therefore, $d(z, T_1(z)) \leq 1$ as desired. □

13.2 Transport geodesics of a Self-centered Bonnet-Myers sharp graph

In this section, we assume that our (D, L)-Bonnet-Myers sharp graph $G = (V, E)$ has the extra condition that G is ***self-centered***: every vertex is a pole, i.e., for every $x \in V$, there exists $y \in V$ such that $d(x, y) = L$. Such a vertex y is called an ***antipole*** of x.

As a consequence of G being self-centered, the following proposition asserts that any vertex $v \in B_1(x_{j-1})$ must be adjacent or equal to the image $T_j(v)$ for any $1 \leq j \leq L$.

For a shortened notion, we henceforth write $u \simeq v$ to represent that u is adjacent or equal to v.

Proposition 13.5. *Let $G = (V, E)$ be a self-centered (D, L)-Bonnet-Myers sharp with a full-length geodesic g and maps $T_j : B_1(x_{j-1}) \to B_1(x_j)$ defined as in Definition 13.2. For any $1 \leq j \leq L$ and any $v \in B_1(x_{j-1})$, we have $v \simeq T_j(v)$. In particular,*

(P2) *$v = T_j(v)$ if $v \in B_1(x_{j-1}) \cap B_1(x_j)$.*

(P3) $v \sim T_j(v)$ *if* $v \in B_1(x_{j-1}) \setminus B_1(x_j)$.

Proof. Fix $1 \le j \le L$. By treating x_{j-1} as a pole together with an antipole y (such that $d(x_{j-1}, y) = L$), it is implied by the Maximal diameter theorem that $[x_{j-1}, y] = V$. Now since $x_j \in [x_{j-1}, y]$, one can construct a full-length geodesic from x_{j-1} to y passing through x_j, namely, $g' : x_{j-1} \sim x_j \sim ... \sim y$.

Then Corollary 13.4 (with roles of x_0, z, T_1 replaced by x_{j-1}, v, T_j, respectively) asserts that $d(v, T_j(v)) \le 1$, which means $v \simeq T_j(v)$. Since the property (P2) was assumed, the property (P3) follows immediately. □

In words, (P1)-(P3) tell us that the bijection $T_j : B_1(x_{j-1}) \to B_1(x_j)$ is based on ***triangles and a perfect matching***: that is, T_j fixes every vertex z such that $x_{j-1}x_jz$ forms a triangle (including x_{j-1} and x_j), and T_j forms a perfect matching the other points in its domain and range (i.e., $B_1(x_{j-1}) \setminus B_1(x_j)$ and $B_1(x_j) \setminus B_1(x_{j-1})$). This property was proved differently via Laplacian arguments in [CKK+20, Theorem 5.7 and Corollary 5.10].

The combination of Propositions 13.3 and 13.5 immediately yields the following result about the transport geodesics along g of a self-centered Bonnet-Myers sharp graph.

Proposition 13.6. *Let $G = (V, E)$ be a self-centered (D, L)-Bonnet-Myers sharp. Given a full-length geodesic g and maps T_j and T^j (for $1 \le j \le L$) defined as above. Then every $z \in B_1(x_0)$ induces a transport geodesic along g, namely,*

$$x_0 \simeq z(0) \simeq z(1) \simeq z(2) \simeq \cdots \simeq z(L) \simeq x_L,$$

and we denote this geodesic by g_z.

Remark 13.7. The definition of a transport geodesic g_z depends on a full-length geodesic g and sets of transport maps $\{T_j\}_{j=1}^L$. Each T_j is a priori not uniquely defined, since the definition of T_j is based on triangles and a perfect matching, the latter of which is not necessarily unique. We will see later (cf. Remark 13.13) that in fact the maps $\{T_j\}_{j=1}^L$ are already uniquely determined by g in the case of self-centered Bonnet-Myers sharp graphs.

13.3 Antipoles of intervals in a self-centered Bonnet-Myers sharp graph

We still assume that our graph $G = (V, E)$ is a self-centered (D, L)-Bonnet-Myers sharp graph. Henceforth we will use the following notation related to intervals:

Given an interval $[x,y] \subset V$ in G and a vertex $z \in [x,y]$, we call a vertex $\overline{z} \in [x,y]$ an ***antipole of*** z ***w.r.t.*** $[x,y]$ if $d(x,y) = d(z,\overline{z})$. Note that antipoles were already introduced for graphs and this definition simply means that z and $\overline{z}$ are antipoles of the induced subgraph of $[x,y]$. We now focus on identifying antipoles w.r.t. intervals via the method of transport geodesics.

Theorem 13.8. *Let $G = (V,E)$ be a self-centered (D,L)-Bonnet-Myers sharp with a full-length geodesic $g: \quad x_0 \sim x_1 \sim \cdots \sim x_L$. Then for any $2 \le k \le L$, x_1 has a unique antipole w.r.t. the interval $[x_0,x_k]$, which we will then denote as $\mathrm{ant}_{[x_0,x_k]}(x_1)$. In fact, we show that*

$$\mathrm{ant}_{[x_0,x_k]}(x_1) = x_0(k) = T^k(x_0) \in B_1(x_k)$$

for any fixed $\{T_j, T^j\}_{j=1}^L$ defined in Definition 13.2.

Proof. First, fix a set of transport maps $\{T_j, T^j\}_{j=1}^L$ associated to g. Suppose that there exists $z \in [x_0,x_k]$ which is an antipole of x_1 w.r.t. $[x_0,x_k]$, that is $z \in [x_0,x_k]$ and $d(x_1,z) = d(x_0,x_k) = k$. Since $x_1 \sim x_0$ and $d(x_1,z) = k$, we have $d(x_0,z) \ge k-1$. Since $z \in [x_0,x_k]$ and $z \ne x_k$, we must have $d(x_0,z) = k-1$ and $d(z,x_k) = 1$. Since $z \in B_1(x_k)$, there is a unique $a \in B_1(x_0)$ such that $a(k) = z$, that is $a = (T^k)^{-1}(z)$ (because T^k is a bijective map). By Proposition 13.6, we obtain a geodesic

$$\begin{array}{c} x_0 \simeq a(0) \simeq a(1) \simeq \ldots \simeq a(k) \\ \hphantom{x_0 \simeq a(0) \simeq a(1) \simeq \ldots \simeq{}} \parallel \\ \hphantom{x_0 \simeq a(0) \simeq a(1) \simeq \ldots \simeq{}} z \end{array}$$

as a part of the geodesic g_a. Therefore it satisfies

$$k - 1 = d(x_0,z) = d(x_0,a(0)) + d(a(0),a(1)) + d(a(1),z). \tag{13.1}$$

On the other hand, since $d(x_1,z) = k$ and $a(1) \in B_1(x_1)$, the triangle inequality gives $d(a(1),z) \ge k-1$. Equation (13.1) then implies $x_0 = a(0) = a(1)$, which means $a = x_0$. Thus $z = a(k) = x_0(k)$. So far we have shown that, for every $2 \le k \le L$, $x_0(k)$ is the only candidate for an antipole of x_1 w.r.t. $[x_0,x_k]$. It remains to show that $x_0(k)$ is in fact the antipole of x_1 w.r.t $[x_0,x_k]$.

In particular, when $k = L$, the antipole of x_1 w.r.t. $[x_0,x_L] = V$ exists by the assumption that G is self-centered. Denote this antipole by $\overline{x}_1$ (which obviously differs from x_L). By the previous argument, $x_0(L)$ must be $\overline{x}_1$.

Consider the transport geodesic g_{x_0} and recall that $x_0 = x_0(0) = x_0(1)$ and $x_0(L) = \overline{x}_1 \ne x_L$:

$$g_{x_0}: \quad x_0 = x_0(0) = x_0(1) \simeq x_0(2) \simeq \cdots \simeq \underset{\overline{x}_1}{\underset{\parallel}{x_0(L)}} \sim x_L.$$

Computing the length of g_{x_0} yields the inequality

$$L = d(x_0, x_L) = \sum_{j=1}^{L-1} d(x_0(j), x_0(j+1)) + d(x_0(L), x_L) \leq (L-1) + 1 = L,$$

which holds with equality. It means $x_0(j) \sim x_0(j+1)$ for all $1 \leq j \leq L-1$.

By removing x_L from the geodesic g_{x_0}, we obtain a new geodesic $x_0 \sim x_0(2) \sim \cdots \sim x_0(L)$ of length $L-1$. Next, by extending this geodesic by $x_1 \sim x_0$, we obtain a new sequence of vertices, namely,

Next we consider a new sequence of vertices obtained by removing x_L and extending $x_1 \sim x_0$ to the left of the geodesic g_{x_0}, namely,

$$g' : \quad x_1 \sim \underbrace{x_0 \sim x_0(2) \sim \cdots \sim x_0(L)}_{\text{geodesic of length } L-1}.$$

Together with the assumption that $d(x_1, x_0(L)) = L$, we know g' is also a geodesic. Consequently, we can read off from the geodesic g' that for every $k \in \{2, ..., L\}$, we have

1. $d(x_1, x_0(k)) = k$, and
2. $x_0(k) \in [x_0, x_k]$, because $k = d(x_0, x_k) \leq d(x_0, x_0(k)) + d(x_0(k), x_k) \leq (k-1) + 1$.

Therefore, $x_0(k)$ is the unique antipole of x_1 w.r.t. $[x_0, x_k]$ as desired. □

Let us first discuss an immediate consequence of Theorem 13.8. Note that the theorem implies that there is a well-defined antipole map

$$\text{ant}_{[x,y]} : [x, y] \cap B_1(x) \to [x, y] \cap B_1(y).$$

The existence and uniqueness of antipoles for neighbours of x w.r.t. $[x, y]$ implies the following result.

Corollary 13.9. *Let $G = (V, E)$ be a self-centered Bonnet-Myers sharp graph, $x, y \in V$ be two different vertices. Then the antipole map*

$$\text{ant}_{[x,y]} : [x, y] \cap B_1(x) \to [x, y] \cap B_1(y)$$

is bijective and, consequently,

$$|[x, y] \cap B_1(x)| = |[x, y] \cap B_1(y)|.$$

Remark 13.10. Let $x, y \in V$ be two different vertices and $x' \in [x, y] \cap B_1(x)$ with its antipole $y' = \mathrm{ant}_{[x,y]}(x')$. Observe that then $x, y \in [x', y']$ and $y = \mathrm{ant}_{[x',y']}(x)$.

Another immediate consequence of Theorem 13.8 is the following corollary about the intervals of length two.

Corollary 13.11. *Let $G = (V, E)$ be a self-centered Bonnet-Myers sharp graph. Then for any $x, y \in V$ such that $d(x, y) = 2$, the induced subgraph of $[x, y]$ is a* ***cocktail party graph****, that is, for any $z \in [x, y]$, there exists a unique $z' \in [x, y]$ such that $d(z, z') = 2$.*

Proof. Let $x, y \in V$ with $d(x, y) = 2$ and let $z \in [x, y]$. If $z = x$ (or resp. $z = y$), it follows immediately from the definition of intervals that such vertex $z' \in [x, y]$ with $d(z, z') = 2$ must be $z' = y$ (or resp. $z' = x$). Now we consider the case of $z \notin \{x, y\}$. Let $x = x_0$ has an antipole x_L, and let $z = x_1$ and $y = x_2$. Now we have $d(x_0, x_L) = L, d(x_0, x_1) = d(x_1, x_2) = 1$. By the maximal diameter theorem, $x_2 \in [x_0, x_L]$, that is $d(x_2, x_L) = L - 2$. Thus x_0, x_1, x_2, x_L lie orderly on some geodesic g of length L, namely,

$$\begin{array}{rccccccc} g: & x_0 & \sim & x_1 & \sim & x_2 & \sim \cdots \sim & x_L. \\ & \parallel & & \parallel & & \parallel & & \\ & x & & z & & y & & \end{array}$$

Applying Theorem 13.8 with $k = 2$, we conclude that there is a unique $z' \in [x, y]$ satisfying $d(z, z') = 2$ as desired. □

Remark 13.12. The fact that all intervals of length two are cocktail party graphs allows us to naturally introduce a ***switching map***, defined as follows. Consider a pair $x, y \in V$ with $d(x, y) = 2$. We define $N_{xy} := S_1(x) \cap S_1(y)$ to be the set of all common neighbors of x and y. Then the switching map $\sigma_{xy} : N_{xy} \to N_{xy}$ is defined by $\sigma_{xy}(z) := \mathrm{ant}_{[x,y]}(z)$ and satisfies $\sigma_{xy}^2 = \mathrm{Id}_{N_{xy}}$.

Remark 13.13. Recall from Remark 13.7 that for general Bonnet-Myers sharp graphs the perfect matchings defining the maps T_j are not necessarily unique. However, under the additional condition of self-centeredness, the fact that all intervals of length two are cocktail party graphs implies the uniqueness of these perfect matchings and the associated transport maps T_j. Therefore, the definition of a transport geodesic g_z depends only on the geodesic g.

In particular, the transport geodesic g_{x_0} containing all antipoles of x_1 w.r.t. increasing intervals $[x_0, x_k]$ (see Theorem 13.8) can be also understood as been generated via the following recursive process of switching maps, as illustrated in Figure 13.1:

$$x_0(2) = \sigma_{x_0 x_2}(x_1),$$

$$x_0(3) = \sigma_{x_0(2)x_3}(x_2),$$
$$\vdots$$
$$x_0(k) = \sigma_{x_0(k-1)x_k}(x_{k-1}),$$
$$\vdots$$
$$x_0(L) = \sigma_{x_0(L-1)x_L}(x_{L-1}).$$

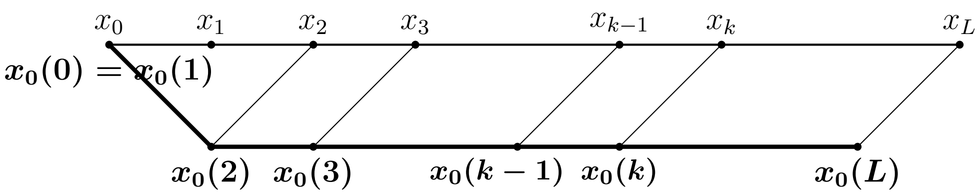


Figure 13.1: Transport geodesic g_{x_0} (along g) shown in **bold**

13.4 Self-centered Bonnet-Myers sharp graphs are strongly spherical (strategy)

In this section and the next section, we aim to prove a rigidity result that all self-centered Bonnet-Myers sharp graphs are strongly spherical. Strongly spherical graphs are a special class of graphs which generalize the hypercubes via the combinatorial property about intervals. A complete classification of strongly spherical graphs is given by Koolen, Moulton and Stevanović [KMS04]. This leads to the main result in [CKK+20] about the classification of all self-centered Bonnet-Myers sharp graphs (see Theorem 13.26).

Let us first give the definitions of relevant combinatorial properties of graphs. For a finite connected graph $G = (V, E)$,
• G is ***self-centered*** if for every $x \in V$ there exists $\overline{x} \in V$ such that $d(x, \overline{x}) = \mathrm{diam}(G)$. The vertex $\overline{x}$ is then called an ***antipole*** of x.
• G is ***antipodal*** if for every $x \in V$ there exists $\overline{x} \in V$ such that $[x, \overline{x}] = V$. The vertex $\overline{x}$ is then called an ***antipode*** of x.
• G is ***strongly spherical*** if G is antipodal, and the induced subgraph of every interval of G is antipodal.

Remark 13.14. It is important to notice the distinction between the notions “antipole” and “antipode”. Here are basic facts about antipodes for any graph $G = (V, E)$:

- Antipodes are also antipoles: Let $\overline{x}$ be an antipode of x in G, that is, $[x, \overline{x}] = V$. We choose arbitrary $y, z \in V$ such that $d(y, z) = \mathrm{diam}(G)$. Then we have by $y, z \in [x, \overline{x}]$ and the triangle inequality

$$\begin{aligned} \mathrm{diam}(G) \geq d(x, \overline{x}) &= \frac{1}{2}(d(x, y) + d(y, \overline{x})) + \frac{1}{2}(d(x, z) + d(z, \overline{x})) \\ &= \frac{1}{2}(d(y, x) + d(x, z)) + \frac{1}{2}(d(y, \overline{x}) + d(\overline{x}, z)) \geq d(y, z) = \mathrm{diam}(G). \end{aligned}$$

- Antipodes are necessarily unique: Assume $\overline{x}_1$ and $\overline{x}_2$ are antipodes of x. Then $\overline{x}_2$ lies on a geodesic from x to $\overline{x}_1$. Since $d(x, \overline{x}_1) = d(x, \overline{x}_2)$, this implies $\overline{x}_1 = \overline{x}_2$.

Remark 13.15. The maximal diameter theorem (Theorem 12.5) asserts that every antipole of a Bonnet-Myers sharp graph G must be an antipode. With an additional assumption that G is self-centered, it implies that G is antipodal. We want to show that self-centered BM-sharps are not only antipodal but also "strongly spherical", which mean all intervals are also antipodal.

The following four steps provides an outline of our proof that a self-centered Bonnet-Myers sharp graph $G = (V, E)$ is strongly spherical, i.e., not only that G itself is antipodal but every interval $[x, y]$ of G is also antipodal.

Step 1: Let $x' \in [x, y] \cap S_1(x)$ with antipole $y' = \mathrm{ant}_{[x,y]}(x')$. We prove for every $z \in [x, y] \cap B_1(x)$ that $z \in [x', y']$ (see Theorem 13.16).

Step 2: Let $x' \in [x, y] \cap S_1(x)$ with antipole $y' = \mathrm{ant}_{[x,y]}(x')$. We prove for every $z \in [x, y]$ that $z \in [x', y']$ (see Theorem 13.20).

Step 3: Let $x' \in [x, y] \cap S_1(x)$ with antipole $y' = \mathrm{ant}_{[x,y]}(x')$. We prove that $[x, y] = [x', y']$ (see Corollary 13.21).

Step 4: Let $x' \in [x, y]$. We prove that there exists $y' \in [x, y]$ such that $[x, y] = [x', y']$ (see Theorem 13.22).

13.5 Self-centered Bonnet-Myers sharp graphs are strongly spherical (proof)

Let us now prove the statements in **Steps 1-4** in this logical order. Recall that the existence of antipoles of vertices in $[x, y] \cap B_1(x)$ w.r.t. $[x, y]$ is guaranteed by Corollary 13.9.

Theorem 13.16. *Let $G = (V, E)$ be self-centered Bonnet-Myers sharp. Let $x, y \in V$ be two different vertices, and consider any $x' \in [x, y] \cap S_1(x)$ with its antipole $y' = \mathrm{ant}_{[x,y]}(x')$. Then every $z \in [x, y] \cap B_1(y)$ satisfies $z \in [x', y']$.*

We start with the set-up and introduce particular sets A, A_1, A_2, Z, Z_1, Z_2 and a function F which will be important for the proof of the above theorem.

Let $k = d(x, y)$. We re-label the vertices as $x = x_0$ and $y = x_k$ and $x' = x_1$ and $y' = \overline{x}_1$, as illustrated in Figure 13.2. Keep in mind that $\overline{x}_1 = \mathrm{ant}_{[x_0,x_k]}(x_1)$ and $x_0 \sim x_1$ and $x_k \sim \overline{x}_1$.

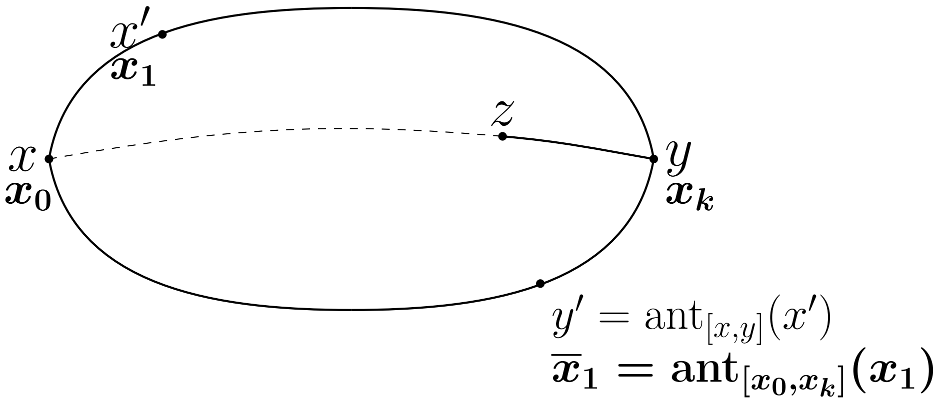


Figure 13.2: The interval $[x, y]$ with the re-labelled vertices in **bold**, and $z \in [x_0, x_k] \cap B_1(x_k)$.

We define the following sets

$$A := [x_0, x_k] \cap B_1(x_0), \qquad Z := [x_0, x_k] \cap B_1(x_k),$$
$$A_1 := A \cap S_1(x_1) \setminus \{x_0\}, \qquad Z_1 := Z \cap S_1(\overline{x}_1) \setminus \{x_k\},$$
$$A_2 := A \cap S_2(x_1), \qquad Z_2 := Z \cap S_2(\overline{x}_1).$$

Note that the sets A and Z can be partitioned into

$$A = \{x_0, x_1\} \sqcup A_1 \sqcup A_2 \qquad \text{and} \qquad Z = \{x_k, \overline{x}_1\} \sqcup Z_1 \sqcup Z_2.$$

Now fix an arbitrary full-length geodesic g from x_0 to x_L (the antipole of x_0) which passes through x_1 and x_k (this can be done since $x_1 \in [x_0, x_k]$ and $x_k \in [x_0, x_L]$ due to the maximal diameter theorem: Theorem 12.5), namely

$$g: \quad x_0 \sim x_1 \sim x_2 \sim \cdots \sim x_k \sim x_{k+1} \sim \cdots \sim x_L.$$

Consider the transport map $T^k : B_1(x_0) \to B_1(x_k)$ introduced in Subsection 13.2. Recall that T^k is bijective. Then define a function $F : Z \to A$ to be $F(z) :=$

$(T^k)^{-1}(z)$ for all $z \in Z \subset B_1(x_k)$. Lemma 13.17 below guarantees that $F(Z) \subseteq A$, hence F is well-defined.

In order to conclude Theorem 13.16, we need to prove that $\forall z \in Z: \ z \in [x_1, \overline{x}_1]$, which is divided into Lemma 13.18 (dealing with the case $z \in Z_2 \sqcup \{x_k, \overline{x_1}\}$) and Lemma 13.19 (dealing with the case $z \in Z_1$).

Lemma 13.17. *$F(Z) \subseteq A$ and $F : Z \to A$ is bijective.*

Lemma 13.18. *$F(Z_2 \sqcup \{x_k, \overline{x}_1\}) = A_2 \sqcup \{x_0, x_1\}$ and $\forall z \in Z_2 \sqcup \{x_k, \overline{x}_1\}: \ z \in [x_1, \overline{x}_1]$.*

Lemma 13.19. *$F(Z_1) = A_1$ and $\forall z \in Z_1: \ z \in [x_1, \overline{x}_1]$.*

Now we will prove the above three lemmas in order, and then conclude Theorem 13.16.

Proof of Lemma 13.17. First we show that $F(z) \in A$ for all $z \in Z$. Let $a := F(z) \in B_1(x_0)$, that is $a = a(0)$ and $z = a(k)$. By Proposition 13.6, we know that

$$x_0 \simeq a(0) \simeq a(1) \simeq ... \simeq a(k)$$

is a geodesic. Moreover, since $a(k) = z \in [x_0, x_k]$, this geodesic can be extended to another geodesic γ, namely

$$\gamma : \quad x_0 \simeq \underset{\underset{a}{\|}}{a(0)} \simeq a(1) \simeq ... \simeq \underset{\underset{z}{\|}}{a(k)} \simeq x_k. \tag{13.2}$$

Therefore, a must lie in the interval $[x_0, x_k]$, which means $a \in A$ and we have $F(Z) \subseteq A$.

Next, note that the function $F : Z \to A$, which is a restriction of $(T^k)^{-1}$, must be injective (because T^k is bijective). Note also that $|A| = |Z|$ because of Corollary 13.9. Therefore, F must be bijective. □

Proof of Lemma 13.18. A main feature of the following proof is to show $A_2 \sqcup \{x_0, x_1\} \subseteq F(Z_2 \sqcup \{x_k, \overline{x}_1\})$. For that reason we start with an element $a \in A_2 \sqcup \{x_0, x_1\}$. Then there exists a uniqe $z \in Z$ with $F(z) = a$. Consequently, $z = a(k)$ and $z \in [x_0, x_k]$. Consider the following two cases.

Case $a = x_0$: From Theorem 13.8, we have $a = x_0 = (T^k)^{-1}(\overline{x}_1) = F(\overline{x}_1)$, so $z = \overline{x}_1$ and $z \in [x_1, \overline{x}_1]$.

Case $a \in A_2 \sqcup \{x_1\}$: As in the proof of Lemma 13.17, we have the following geodesic γ of length k (referred to the one in (13.2)):

$$\gamma: \quad x_0 \sim \underset{\underset{a}{\parallel}}{a(0)} \simeq a(1) \simeq ... \simeq \underset{\underset{z}{\parallel}}{a(k)} \simeq x_k.$$

From this geodesic γ and an observation that

$$d(x_0, a(1)) = \begin{cases} 1, & \text{if } a = x_1 \text{ (and therefore also } a(1) = x_1), \\ 2, & \text{if } a \in A_2 \text{ (and therefore } a(1) \neq a(0)), \end{cases}$$

we conclude

$$d(a(1), x_k) = \begin{cases} k-1, & \text{if } a = x_1, \\ k-2, & \text{if } a \in A_2. \end{cases} \tag{13.3}$$

Now we extend the geodesic

$$a(1) \simeq ... \simeq \underset{\underset{z}{\parallel}}{a(k)} \simeq x_k$$

to

$$x_1 \simeq a(1) \simeq ... \simeq \underset{\underset{z}{\parallel}}{a(k)} \simeq x_k \sim \overline{x}_1,$$

which is, again a geodesic because of (13.3) (and recall that $d(x_1, \overline{x}_1) = k$). We can then read off from the above geodesic that $z = x_k$ or $z \in S_2(\overline{x}_1) \cap [x_0, x_k] = Z_2$ and $z \in [x_1, \overline{x}_1]$.

We conclude from both cases that $A_2 \sqcup \{x_0, x_1\} \subseteq F(Z_2 \sqcup \{x_k, \overline{x}_1\})$. Since F is bijective, it follows that $|A_2| \leq |Z_2|$. By switching the roles between x_0 and x_k and between the antipoles x_1 and $\overline{x}_1$ w.r.t. $[x_0, x_k]$, we obtain the opposite inequality $|Z_2| \leq |A_2|$. Therefore, we have $|Z_2| = |A_2|$, and thus $A_2 \sqcup \{x_0, x_1\} = F(Z_2 \sqcup \{x_k, \overline{x}_1\})$, as desired.

Consequently, if we consider any $z \in Z_2 \sqcup \{x_k, \overline{x}_1\}$, then $a \in A_2 \sqcup \{x_0, x_1\}$ falls into one of the above cases, in which we have shown $z \in [x_1, \overline{x}_1]$. □

Proof of Lemma 13.19. Since $F : Z \to A$ is bijective and $F(Z_2 \sqcup \{x_k, \overline{x}_1\}) = A_2 \sqcup \{x_0, x_1\}$, we conclude $F(Z_1) = A_1$.

Moreover, consider $z \in Z_1$. It follows that $z \sim \overline{x}_1$ and $d(x_1, z) \leq k - 1$, because $z \neq \overline{x}_1 = \text{ant}_{[x_0, x_k]}(x_1)$. Therefore

$$d(x_1, z) + d(z, \overline{x}_1) = d(x_1, z) + 1 \leq (k-1) + 1 = k,$$

which means $z \in [x_1, \overline{x}_1]$. □

Proof of Theorem 13.16. Recalling the original set-up and notation, we only need to show that $z \in [x_1, \overline{x}_1]$. This follows immediately from Lemma 13.18 and Lemma 13.19. □

The next theorem generalizes Theorem 13.16 by removing the restriction $z \in B_1(y)$.

Theorem 13.20. *Let $G = (V, E)$ be self-centered Bonnet-Myers sharp. Let $x, y \in V$ be two different vertices, and consider any $x' \in [x, y] \cap S_1(x)$ with its antipole $y' = \mathrm{ant}_{[x,y]}(x')$. Then every $z \in [x, y]$ satisfies $z \in [x', y']$.*

Proof of Theorem 13.20. Let $d_1 = d(x, y)$ and $d_2 = d(z, y)$ (note that $0 \leq d_2 \leq d_1$). We will prove the statement of the theorem by induction on d_1 and d_2.

Base step: For any value of d_1, the cases $d_2 = 0, 1$ are both covered by Theorem 13.16.

Inductive step: Assume that the statement is true for $d_1 = k - 1$ and all d_2, and assume that the statement is true for $d_1 = k$ and $d_2 = j - 1$ for some $2 \leq j \leq k - 1$. Now consider $d(x, y) = k$ and $z \in [x, y] \cap S_j(y)$. Choose an arbitrary $z_1 \in [z, y] \cap S_1(y)$. Hence x, z, z_1, y lies in a geodesic , see Figure 13.3. In particular, $z \in [x, z_1]$.

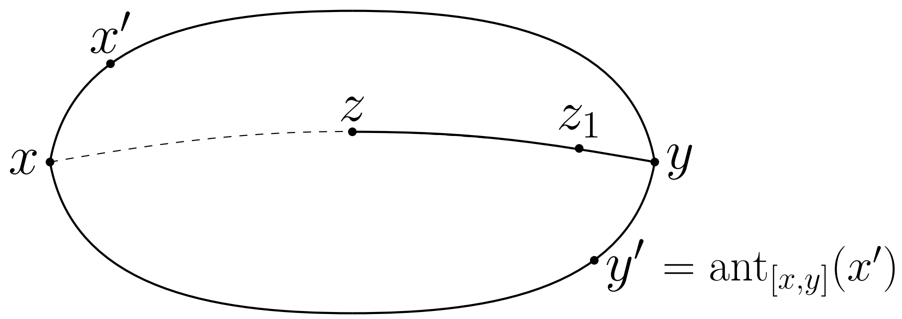


Figure 13.3: The interval $[x, y]$ and $z \in [x, y]$ with $d(z, y) = j \geq 2$ and $z_1 \in [z, y] \cup S_1(y)$.

Now consider the following three cases whether $d(z_1, y')$ is 0, 1, or 2.

Case $z_1 = y'$: It follows immediately that $z \in [x, z_1] = [x, y'] \subseteq [x', y']$ where the last inclusion is due to $x \in [x', y']$.

Case $z_1 \sim y'$: Since $z_1 \in [x, y]$, by Theorem 13.8 there is a unique $a_1 = \mathrm{ant}_{[x,y]}(z_1) \in [x, y]$. Since $a_1 \in [x, y] \cap B_1(x)$ and $z_1 \in [x, y] \cap B_1(y)$, by Theorem 13.16, $z_1, a_1 \in [x', y']$.

The fact that $a_1, z_1 \in [x', y']$ and that $d(a_1, z_1) = d(x, y) = d(x', y')$ altogether implies that a_1 must be the unique antipole $\mathrm{ant}_{[x',y']}(z_1)$ by Corollary 13.9 since $z_1 \sim y'$. This is illustrated in Figure 13.4. By Remark 13.10, it implies that $y' = \mathrm{ant}_{[a_1,z_1]}(x')$.

Observe also that $z \in [x, z_1] \subset [a_1, z_1]$ with $d(z, z_1) = j - 1$.

We are now in a position to apply the induction hypothesis for the interval $[a_1, z_1]$ (instead of $[x, y]$) and $z \in [a_1, z_1]$ with $d(z, z_1) = j - 1$. Note that $d(a_1, z_1) = k$. Note also that $x' \in [a_1, z_1] \cap S_1(a_1)$ and $y' = \mathrm{ant}_{[a_1,z_1]}(x')$. Then the induction hypothesis implies $z \in [x', y']$, finishing this case.

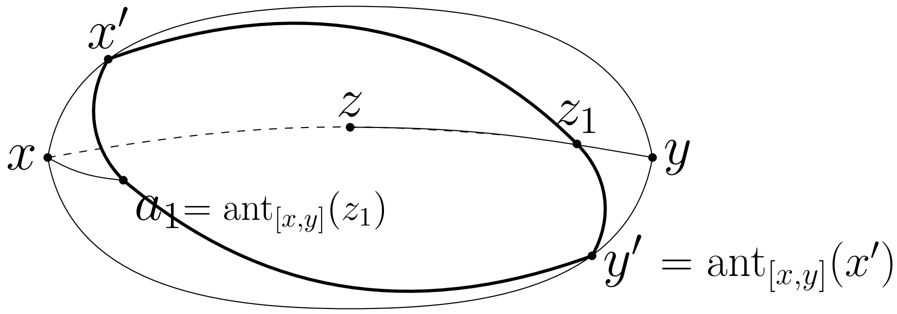


Figure 13.4: Picture for Case $z_1 \sim y'$. The **bold** cycle represents the fact that a_1 and z_1 are antipoles w.r.t. not only $[x, y]$ but also $[x', y']$.

Case $d(z_1, y') = 2$: Since $z_1 \in [x, y] \cap B_1(y)$, by Theorem 13.16, we have $z_1 \in [x', y']$. The condition $d(z_1, y') = 2$ then implies that $d(x', z_1) = d(x', y') - 2 = k - 2$. It follows that

$$d(x, x') + d(x', z_1) + d(z_1, y) = 1 + (k - 2) + 1 = k = d(x, y)$$

which means that x' and z_1 lie on a geodesic from x to y. Let us denote this geodesic by g^*:

$$g^* : \quad x \sim x' \sim \cdots \sim z_1 \sim y.$$

In particular, $x' \in [x, z_1]$. Then $y'' := \mathrm{ant}_{[x,z_1]}(x')$ exists by Corollary 13.9. The situation is illustrated in Figure 13.5.

Next we apply the induction hypothesis for the interval $[x, z_1]$ (instead of $[x, y]$) and $z \in [x, z_1]$ with $d(z, z_1) = j - 1$. Note that $d(x, z_1) = k - 1$. Note also that $x' \in [x, z_1] \cap S_1(x)$ and $y'' = \mathrm{ant}_{[x,z_1]}(x')$. Then the induction hypothesis implies $z \in [x', y'']$.

So far we have that $d(x', z) + d(z, y'') = d(x', y'') = d(x, z_1) = k - 1$. It remains to show that $d(y'', y') = 1$ which would imply

$$k = d(x', y') \leq d(x', z) + d(z, y'') + d(y'', y') = (k - 1) + 1 = k,$$

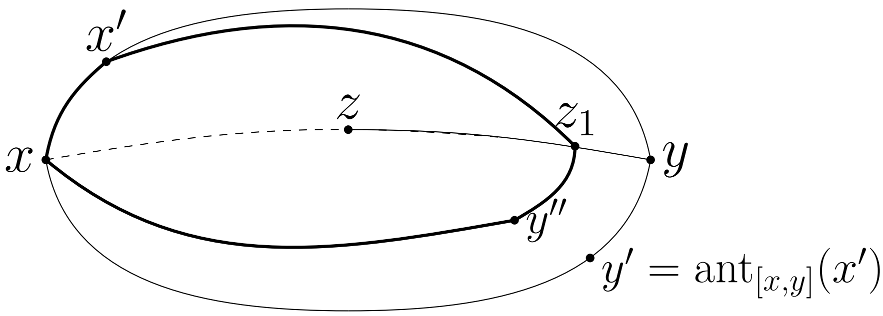


Figure 13.5: Picture for Case $d(z_1, y') = 2$. The **bold** cycle represents $y'' = \mathrm{ant}_{[x,z_1]}(x')$.

that is $z \in [x', y']$, as desired.

To prove $d(y'', y') = 1$ we use transport geodesic techniques. Therefore, we relabel the vertices of the geodesic g^* and extend g^* to a full-length geodesic g in G starting from $x = x_0$ as follows:

$$g: \quad \underset{x_0}{\overset{x}{\parallel}} \sim \underset{x_1}{\overset{x'}{\parallel}} \sim \cdots \sim \underset{x_{k-1}}{\overset{z_1}{\parallel}} \sim \underset{x_k}{\overset{y}{\parallel}} \sim x_{k+1} \sim \cdots \sim x_L,$$

and consider the transport geodesic along g starting at x_0.

Theorem 13.8 guarantees that $x_0(m) = \mathrm{ant}_{[x_0,x_m]}(x_1)$ for all $2 \leq m \leq L$. In particular, we have $y'' = \mathrm{ant}_{[x_0,x_{k-1}]}(x_1) = x_0(k-1)$ and $y' = \mathrm{ant}_{[x_0,x_k]}(x_1) = x_0(k)$. Therefore, $y'' = x_0(k-1)$ and $y' = x_0(k)$ must be adjacent vertices (as illustrated in Figure 13.6), thus completing the proof.

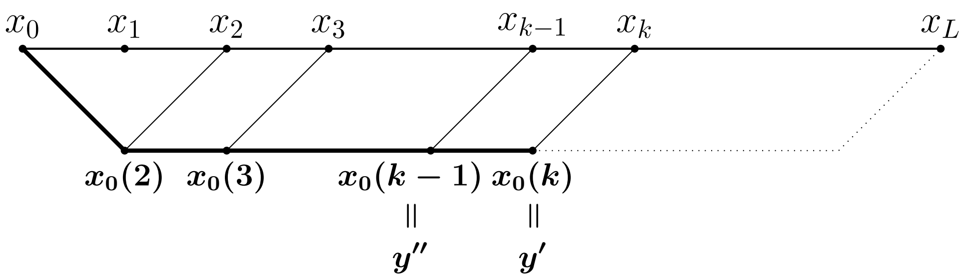


Figure 13.6: Transport geodesic g_{x_0} and the antipoles of x_1 w.r.t. increasing intervals $[x_0, x_m]$.

□

An immediate but important consequence of the above theorem is the following corollary.

Corollary 13.21. *Let $G = (V, E)$ be self-centered Bonnet-Myers sharp. Let $x, y \in V$ be two different vertices, and consider any $x' \in [x, y] \cap S_1(x)$ with its antipole $y' = \mathrm{ant}_{[x,y]}(x')$. Then $[x', y'] = [x, y]$.*

Proof of Corollary 13.21. Theorem 13.20 can be rephrased as $[x, y] \subseteq [x', y']$. Since $y = \mathrm{ant}_{[x',y']}(x)$ by Remark 13.10, we can interchange the roles of x, y and x', y' to obtain the opposite inclusion $[x', y'] \subseteq [x, y]$. Therefore $[x', y'] = [x, y]$, as desired. □

Now, we are ready to conclude the ultimate result of this section by applying Corollary 13.21 inductively.

Theorem 13.22. *Let $G = (V, E)$ be self-centered Bonnet-Myers sharp. Then for two different vertices $x, y \in V$, the induced subgraph of the interval $[x, y]$ is antipodal. Therefore G is strongly spherical.*

Proof of Theorem 13.22. Let $x' \in [x, y]$. We will prove by induction on the distance $d(x, x')$ that there exists a vertex $y' \in [x, y]$ such that $[x, y] = [x', y']$.

<u>Base step:</u> The case $d(x, x') = 0$ is trivial and $d(x, x') = 1$ is covered by Corollary 13.21.

<u>Inductive step:</u> We assume the statement of the Theorem is true for all $d(x, x') \leq m - 1$ with $2 \leq m \leq \mathrm{diam}(G)$. Let $x' \in [x, y]$ with $d(x, x') = m$. We choose a vertex $x_1 \in [x, y] \cap S_1(x)$ such that x_1, x' lie on a geodesic from x to y. By the induction hypothesis, there exists $y_1 \in [x, y]$ such that

$$[x, y] = [x_1, y_1].$$

Since $d(x_1, x') = m - 1$, the induction hypothesis again implies the existence of $y' \in [x_1, y_1] = [x, y]$ such that

$$[x_1, y_1] = [x', y'] = [x, y].$$

This finishes the proof. □

13.6 Classification of Self-centered Bonnet-Myers sharp graphs

Since we have shown that all regular self-centered Bonnet-Myers sharp graphs must be strongly spherical (Theorem 13.22), the task to find all such graphs is reduced to checking through the list of all the strongly spherical graphs, which is classified by Koolen, Moulton and Stevanović [KMS04] as follows.

Theorem 13.23 ([KMS04]). *Strongly spherical graphs are precisely all the Cartesian products of the following graphs:*

1. *hypercubes Q^n, $n \geq 1$;*
2. *cocktail party graphs $CP(n)$, $n \geq 3$;*
3. *the Johnson graphs $J(2n, n)$, $n \geq 3$;*
4. *even-dimensional demi-cubes $Q^{2n}_{(2)}$, $n \geq 3$;*
5. *the Gosset graph.*

For readers' convenience, a brief description of these graphs is given below.

1. The hypercube Q^n has 2^n vertices indexed by $\{0,1\}^n$ and edges between them if Hamming distance equals one.
2. The cocktail party graph $CP(n)$ is obtained by removal of a perfect matching from the complete graph K_{2n};
3. the Johnson graph $J(2n, n)$ has vertices corresponding to n-element subsets of $\{1, 2, \dots, 2n\}$ and edges between them if they overlap in $n-1$ elements.
4. The even-dimensional demi-cube $Q^{2n}_{(2)}$ is one of the two isomorphic connected components of the vertex set $\{0,1\}^{2n}$ and edges between them if Hamming distance equals two.
5. The Gosset graph with 56 vertices: the vertices are in one-to-one correspondence with the edges $\{i,j\}$ and $\{i,j\}'$ of two disjoint copies of K_8, respectively, and $\{i,j\} \sim \{k,l\}$ if $|\{i,j\} \cap \{k,l\}| = 1$ and $\{i,j\} \sim \{k,l\}'$ if $\{i,j\} \cap \{k,l\} = \emptyset$.

These graphs (with n prescribed as in Theorem 13.23) are non-isomorphic to one another, and they are irreducible (with the exception of the hypercubes Q^n, $n \geq 2$). Moreover, they are vertex- and edges-transitive and self-centered. Their specific information including the curvature on an edge (see [CKK+20] for calculation details) is listed in the following table:

One can see that $\kappa(x,y) = \frac{2}{L}$ in all examples above, confirming that they are all (D, L)-Bonnet-Myers sharp. Interestingly, unlike in the case of strongly regular graphs, Bonnet-Myers sharpness is not necessarily preserved under taking any Cartesian products. In order for the product to be Bonnet-Myers sharp, it requires an extra condition that each component has the same ratio of degree and diameter, as explained in the following theorem.

G	$\lvert V\rvert$	$D=\deg(G)$	$L=\mathrm{diam}(G)$	$\kappa(x,y)$
Q^n	2^n	n	n	$\frac{2}{n}$
$CP(n)$	$2n$	$2n-2$	2	1
$J(2n,n)$	$\binom{2n}{n}$	n^2	n	$\frac{2}{n}$
$Q^{2n}_{(2)}$	2^{2n-1}	$2n^2-n$	n	$\frac{2}{n}$
Gosset	56	27	3	$\frac{2}{3}$

Table 13.1: Examples of (D,L)-Bonnet-Myers sharp graphs

Theorem 13.24 ([CKK+20]). *Let $\{G_i=(V_i,E_i)\}_{i=1}^N$ be a family of regular graphs where G_i has valency D_i for each i. Let L_i be the diameter of G_i. Let $G=G_1\times\cdots\times G_N$. The following are equivalent:*

1. *G is Bonnet-Myers sharp.*
2. *Each G_i is Bonnet-Myers sharp and $\frac{D_1}{L_1}=\cdots=\frac{D_N}{L_N}$.*

Proof. We may assume without loss of generality that $N=2$ (and the case for $N>2$ can be argued inductively). Recall from Theorem 11.6 that the curvature of the Cartesian product is given by

$$\begin{aligned}\inf_{\substack{u,v\in V(G)\\ u\sim v}}\kappa^G(u,v)&=\min\left\{\inf_{\substack{x_1,x_2\in V_1\\ x_1\sim x_2\\ y\in V_2}}\kappa^G((x_1,y),(x_2,y)),\inf_{\substack{y_1,y_2\in V_2\\ y_1\sim y_2\\ x\in V_1}}\kappa^G((x,y_1),(x,y_2))\right\}\\ &=\min\left\{\inf_{\substack{x_1,x_2\in V_1\\ x_1\sim x_2}}\frac{D_1}{D_1+D_2}\kappa^{G_1}(x_1,x_2),\inf_{\substack{y_1,y_2\in V_2\\ y_1\sim y_2}}\frac{D_2}{D_1+D_2}\kappa^{G_2}(y_1,y_2)\right\}.\end{aligned}\tag{13.4}$$

First we prove that (i) implies (ii). Since G is Bonnet-Myers sharp, togther with the Bonnet-Myers diameter bound on G_1, we have

$$\frac{2}{L_1+L_2}=\inf_{\substack{u,v\in V(G)\\ u\sim v}}\kappa^G(u,v)\le\inf_{\substack{x_1,x_2\in V_1\\ x_1\sim x_2}}\frac{D_1}{D_1+D_2}\kappa^{G_1}(x_1,x_2)\le\frac{D_1}{D_1+D_2}\cdot\frac{2}{L_1},$$

which is equivalent to $\frac{D_2}{L_2}\le\frac{D_1}{L_1}$.

On the other hand, the Bonnet-Myers diameter bound on G_2 gives

$$\frac{2}{L_1+L_2}=\inf_{\substack{u,v\in V(G)\\ u\sim v}}\kappa^G(u,v)\le\inf_{\substack{y_1,y_2\in V_2\\ y_1\sim y_2}}\frac{D_2}{D_1+D_2}\kappa^{G_2}(y_1,y_2)\le\frac{D_2}{D_1+D_2}\cdot\frac{2}{L_2},$$

which is equivalent to $\frac{D_1}{L_1} \leq \frac{D_2}{L_2}$.

Therefore, we can conclude that $\frac{D_1}{L_1} = \frac{D_2}{L_2}$, and all the inequalities above are sharp, that is G_1 and G_2 are Bonnet-Myers sharp as well.

To prove (ii) implies (i): we simply plug into (13.4)

$$\inf_{\substack{x_1,x_2\in V_1\\ x_1\sim x_2}} \kappa^{G_1}(x_1,x_2) = \frac{2}{L_1} \quad \text{and} \quad \inf_{\substack{y_1,y_2\in V_2\\ y_1\sim y_2}} \kappa^{G_2}(y_1,y_2) = \frac{2}{L_2}$$

and use the assumption that $\frac{D_1}{L_1} = \frac{D_2}{L_2}$. As a result, we obtain $\inf_{\substack{u,v\in V(G)\\ u\sim v}} \kappa^G(u,v) = \frac{2}{L_1+L_2}$. □

Remark 13.25. In contrast to the necessary and sufficient condition for Bonnet-Myers sharpness in Theorem 13.24, much less is required for the Cartesian product $G = G_1 \times G_2$ to be Lichnerowicz sharp. In fact, G is Lichnerowicz sharp already if G_1 is Lichenerowicz sharp and G_2 is an arbitrary graph with its curvature lower bound large enough, as explained in the following argument.

Let $\lambda_1^{G_1}, \lambda_1^{G_2}, \lambda_1^{G}$ be the smallest positive eigenvalues of the Laplacians on G_1, G_2, G. We have

$$\begin{aligned}\inf_{\substack{u,v\in V(G)\\ u\sim v}} \kappa^G(u,v) &= \min\left\{\inf_{\substack{x_1,x_2\in V_1\\ x_1\sim x_2}} \frac{D_1}{D_1+D_2}\kappa^{G_1}(x_1,x_2), \inf_{\substack{y_1,y_2\in V_2\\ y_1\sim y_2}} \frac{D_2}{D_1+D_2}\kappa^{G_2}(y_1,y_2)\right\} \\ &\leq \min\left\{\frac{D_1}{D_1+D_2}\lambda_1^{G_1}, \frac{D_2}{D_1+D_2}\lambda_2^{G_2}\right\} = \lambda_1^G, \end{aligned} \tag{13.5}$$

where the inequality comes from Lichnerowicz' Theorem on each graph G_i: $\inf \kappa^{G_i} \leq \lambda_1^{G_i}$. In order to obtain the equality in (13.5), a sufficient condition is $\inf \kappa^{G_1} = \lambda_1^{G_1}$ (i.e. G_1 is Lichnerowicz sharp) and $\frac{D_1}{D_2} \inf \kappa^{G_1} \leq \inf \kappa^{G_2}$.

Finally, we can combine Theorems 13.22, 13.23 and 13.24 to conclude the complete classification of regular self-centered Bonnet-Myers sharp graphs.

Theorem 13.26 (main result, [CKK+20]). *Regular self-centered Bonnet-Myers sharp graphs are precisely the Cartesian products $G = G_1 \times G_2 \times \cdots \times G_N$, where each factor G_i is any of the following graphs:*

1. *hypercubes Q^n, $n \geq 1$;*
2. *cocktail party graphs $CP(n)$, $n \geq 3$;*

3. *the Johnson graphs* $J(2n, n)$, $n \geq 3$;

4. *even-dimensional demi-cubes* $Q^{2n}_{(2)}$, $n \geq 3$;

5. *the Gosset graph,*

and satisfies the condition $\frac{D_1}{L_1} = \frac{D_2}{L_2} = \cdots = \frac{D_N}{L_N}$.

Chapter 14

Outlook for Ollivier Ricci curvature

Let us conclude Part II with some interesting questions, especially about Cheng's and Obata's rigidity for Ollivier Ricci curvature.

1. While the main classification result of Bonnet-Myers sharp graphs (Theorem 13.26) assumes the graphs in consideration to be regular and self-centered, it is a natural question to ask if we could drop either or both of these assumptions. From an optimistic viewpoint, we conjecture that all regular Bonnet-Myers sharp graphs must be self-centered. This conjecture holds in the extremal cases where the graph diameter is equal to two (in this case, the graph has to be a Cocktail party graph) or to the vertex degree (in this case, the graph has to be a hypercube); for the proof see [CKK+20]. Moreover, in [Kam20a] we confirm this conjecture for all such graphs of diameter three.

2. It is shown in [CKK+20] that Bonnet-Myers sharp graphs are Lichnerowicz sharp, and the class of all Lichnerowicz sharp graphs is much larger. The paper also presents some partial results about the classification of Lichnerowicz sharp graphs which are distance-regular with a certain value of the second largest adjacency eigenvalue. It would be interesting to find further classification results of Lichnerowicz sharp graphs.

3. It was brought to our attention by Paul Terwilliger that the classification found for the Bonnet-Myers sharp graphs agrees with the class of graphs related hypermetric spaces mentioned in [TD87]. This agreement is likely not to be coincidental, and there could potentially be some connection between these two classes.

Part III

Bakry-Émery curvature on graphs

Chapter 15

Discrete Bakry-Émery theory

The application of the Bakry and Émery theory to define the Ricci curvature notion for discrete spaces was initiated and promoted by, e.g., Elworthy [Elw91], Schmuckenschläger [Sch99] and Lin-Yau [LY10]. In this chapter, we revisit the Γ-calculus and the definition Bakry-Émery curvature which were previously given in Chapter 5 but are now discussed in the context of weighted graphs.

15.1 Setting of weighted graphs and Laplacians

Let $G = (V, w, \mu)$ be a weighted graph consisting of a countable vertex set V, a vertex measure $\mu : V \to \mathbb{R}^+$, and an edge-weight function $w : V \times V \to \mathbb{R}_{\geq 0}$ which is a symmetric function with $w_{xx} = 0$ for all $x \in V$. Two vertices $x, y \in V$ are adjacent (denoted as $x \sim y$) if and only if $w_{xy} > 0$. The graph G is assumed to be connected, and locally finite, that is, each vertex has only finitely many neighbors. For $r \in \mathbb{N}$, the sphere (resp. ball) of radius r centered at $x \in V$, denoted by $S_r(x)$ (resp. $B_r(x)$), is the set of all vertices whose minimum number of edges from x is equal to (resp. less than or equal to) r. In particular, $S_1(x)$ contains all neighbors of x.

Furthermore, let $p_{xy} := \frac{w_{xy}}{\mu_x}$ be the ***transition rate*** from x to y. Let $\mathrm{Deg}(x) := \frac{1}{\mu_x} \sum_{y \in V} w_{xy} = \sum_{y \in V} p_{xy}$ be the ***vertex degree*** of x, and mostly we assume maximal degree to be finite, i.e., $\mathrm{Deg}_{\max} := \sup_{x \in V} \mathrm{Deg}(x) < \infty$. In the special case of $\mathrm{Deg}(x) = 1$ for all $x \in V$, the terms p_{xy} can be understood as transition probabilities of a reversible Markov chain (which is the topic of discussion in Chapter 20). Another special situation is a ***non-weighted*** (or combinatorial) graph $G = (V, E)$ where E is the set of edges (without loops and multiple edges), that is, $\mu \equiv 1$ and $w_{xy} = 1$ iff x is adjacent to y, and $w_{xy} = 0$ otherwise. In this case, $\mathrm{Deg}(x) = \sum_{y \in V} w_{xy}$ equals the number of neighbors of x, which is the usual definition of the vertex degree.

The Laplacian $\Delta : \mathbb{R}^V \to \mathbb{R}^V$ (where $\mathbb{R}^V$ is the vector space of all functions $f : V \to \mathbb{R}$) is a linear operator on V given by

$$\begin{aligned}\Delta f(x) := \frac{1}{\mu_x} \sum_{y \in V} w_{xy}(f(y) - f(x)) &= \sum_{y \in V} p_{xy}(f(y) - f(x)) \\ &= \sum_{y \in S_1(x)} p_{xy}(f(y) - f(x)),\end{aligned}$$

where the summation can be restricted to only those $y \in S_1(x)$ since by definition $x \sim y$ if and only if $w_{xy} > 0$. The Laplacian associated to non-weighted graphs is also known as the ***non-normalized Laplacian***.

15.2 Γ-calculus on graphs

The Laplacian Δ gives rise to the symmetric bilinear forms Γ and Γ_2, namely,

$$\begin{aligned} 2\Gamma(f,g) &:= \Delta(fg) - f\Delta g - g\Delta f, \qquad (15.1)\\ 2\Gamma_2(f,g) &:= \Delta(\Gamma(f,g)) - \Gamma(f,\Delta g) - \Gamma(g,\Delta f),\end{aligned}$$

for all $f, g \in \mathbb{R}^V$. Moreover, we write $\Gamma f := \Gamma(f,f)$ and $\Gamma_2 f := \Gamma_2(f,f)$.

In words, Γ represents how much the Laplacian is far away from the usual chain rule for derivative $D(fg) = fDg + gDf$, and Γ_2 is an iteration from Γ where the product of two functions fg are replaced by $\Gamma(f,g)$.

Some basic properties of Δ and Γ on graphs are given in the following proposition.

Proposition 15.1. *Let $G = (V, w, \mu)$ be a weighted graph. For all $f, g \in \mathbb{R}^V$ and $x \in V$, we have*

(L1) $2\Gamma(f,g)(x) = \sum_{y \in S_1(x)} p_{xy}\big(f(y)-f(x)\big)\big(g(y)-g(x)\big)$*, and particularly,* $2\Gamma f(x) = \sum_{y \in S_1(x)} p_{xy}\big(f(y) - f(x)\big)^2$*;*

(L2) $(\Delta f(x))^2 \leq 2\,\mathrm{Deg}(x)\Gamma f(x)$*;*

(L3) If G is finite, then $\langle \Delta f, g\rangle = \langle f, \Delta g\rangle$,
where the norm is denoted by $\langle f, g\rangle := \sum_{x \in V} \mu_x f(x) g(x)$*. In particular,* $\langle \Delta f, 1\rangle = \sum_{x \in V} \mu_x \Delta f(x) = 0$.

Remark 15.2. The properties (L1) and (L3) are discrete analogues of properties in Riemannian manifolds. With an additional notion of *discrete gradient* given by $\nabla f(x,y) := f(y) - f(x)$, the property (L1) can be written as $2\Gamma(f,g) = \langle \nabla f, \nabla g\rangle$ and $2\Gamma f = |\nabla f|^2$, which is a discrete analogue to (5.11) (but notice the factor 2 discrepancy). The property (L3) means Δ is self-adjoint on L^2-norm, and its consequence $\sum_{x\in V}\mu_x \Delta f(x) = 0$ is a discrete version of the divergence theorem for closed manifolds. The property (L2), on the other hand, is a specific property in the discrete setting of graphs and does not have a counterpart in Riemannian manifolds (as one should not expect a bound like $|\Delta f| \le C|\operatorname{grad} f|^2$ since the Laplacian is a second-order differential operator, while gradient is only a first-order differential operator).

Proof of Proposition 15.1. (L1) A straightforward calculation gives

$$\begin{aligned}
2\Gamma(f,g)(x) &= \Delta(fg)(x) - f(x)\cdot\Delta g(x) - \Delta f(x)\cdot g(x)\\
&= \sum_{y\in S_1(x)} p_{xy}\big[f(y)g(y) - f(x)g(x)\\
&\qquad\qquad - f(x)\big(g(y)-g(x)\big) - g(x)\big(f(y)-f(x)\big)\big]\\
&= \sum_{y\in S_1(x)} p_{xy}\big(f(y)-f(x)\big)\big(g(y)-g(x)\big),
\end{aligned}$$

where the identity $2\Gamma f(x) = \sum_{y\in S_1(x)} p_{xy}\big(f(y)-f(x)\big)^2$ follows immediately.

(L2) We apply Cauchy-Schwarz inequality to (L1) and obtain

$$\begin{aligned}
(\Delta f(x))^2 &= \left(\sum_{y\in S_1(x)} p_{xy}\big(f(y)-f(x)\big)\right)^2\\
&\le \left(\sum_{y\in S_1(x)} p_{xy}\big(f(y)-f(x)\big)^2\right)\cdot \sum_{y\in S_1(x)} p_{xy} \overset{(L1)}{=} 2\operatorname{Deg}(x)\Gamma f(x).
\end{aligned}$$

(L3) Suppose G is finite. Then

$$\begin{aligned}
\sum_{x\in V}\mu_x\Delta f(x)g(x) &= \sum_x\sum_y w_{xy}\big(f(y)-f(x)\big)g(x)\\
&= \sum_x\sum_y w_{xy}f(y)g(x) - \sum_x\sum_y w_{xy}f(x)g(x)\\
&= \sum_y\sum_x w_{xy}f(x)g(y) - \sum_x\sum_y w_{xy}f(x)g(x)
\end{aligned}$$

$$= \sum_{x \in V} \mu_x f(x) \Delta g(x).$$

□

15.3 Bakry-Émery curvature in local matrix form

The previous section has prepared us with the discrete version of bilinear operators Γ and Γ_2. Now we are ready to present in the context of weighted graphs, the definition of Bakry-Émery curvature, which uses Γ and Γ_2 as the key ingredients (cf. Section 5.4).

Definition 15.3 (Bakry-Émery curvature)**.** Let $G = (V, w, \mu)$ be a locally finite weighted graph. Let $K \in \mathbb{R}$ and $N \in (0, \infty]$. We say that a vertex $x \in V$ satisfies $BE(K, N)$ if the following Bakry-Émery's ***curvature-dimension inequality*** holds for all $f \in \mathbb{R}^V$:

$$\Gamma_2 f(x) \geq K \Gamma f(x) + \frac{1}{N} (\Delta f(x))^2. \tag{15.2}$$

The ***Bakry-Émery curvature*** at a vertex x, denoted by $\mathcal{K}(G, x; N)$, is then defined to be the largest K such that x satisfies $BE(K, N)$. Moreover, we say that G satisfies $BE(K, N)$ if all $x \in V$ satisfy $BE(K, N)$.

Most of the times, saying that "a graph G satisfies $BE(K, N)$" is almost the same as $K = \inf_{x \in V} \mathcal{K}(G, x; N)$, except that in the former statement K does not need to be maximum possible one. We may use both of them interchangeably to express the global lower bound of Bakry-Émery curvature.

Here, the parameter N acts as an upper bound for "dimension" of the weighted graph, and K acts as a "Ricci lower curvature bound" at x. There is no unified notions of dimension and Ricci curvature for graphs, so this convention is only to be compared with Theorem 5.12 for manifolds.

Remark 15.4. Both $\Delta f(x)$ and $\Gamma(f, g)(x)$ only involves the information within $B_1(x)$ (i.e., neighbors of x including x itself). The iterated $\Gamma_2(f, g)$ involves the information within $B_2(x)$ (i.e., all vertices with distance ≤ 2 from x). Therefore, the test functions for (15.2) can be restricted to those $f : B_2(x) \to \mathbb{R}$.

Remark 15.5. If x satisfies $BE(K, N)$ then x also satisfies $BE(K', N')$ for all $K' \leq K$ and all $N' \geq N$ (including $N' = \infty$) because $\Gamma f(x) \geq 0$ for all f. Equivalently, the curvature at x is given by

$$\mathcal{K}(G, x; N) = \inf_f \frac{\Gamma_2 f(x) - \frac{1}{N} (\Delta f(x))^2}{\Gamma f(x)},$$

where the infimum is taken over all functions $f \in \mathbb{R}^V$ which are non-constant on $B_1(x)$ and therefore $\Gamma f(x) > 0$.

Alternatively, it is easier to view the involved operators in (15.2) by their matrix representations. The linear operator $\Delta(\cdot)(x)$ and the bilinear forms $\Gamma(\cdot,\cdot)(x), \Gamma_2(\cdot,\cdot)(x)$ can be represented by a local vector $\Delta(x)$ and local matrices $\Gamma(x), \Gamma_2(x)$ as

$$\begin{aligned} \Delta f(x) &= \Delta(x)^\top \vec{f}, \\ \Gamma(f,g)(x) &= \vec{f}^\top \Gamma(x) \vec{g}, \\ \Gamma_2(f,g)(x) &= \vec{f}^\top \Gamma_2(x) \vec{g}. \end{aligned}$$

These vector and matrices are "local" in the sense that its entries are only indexed by vertices near x instead the whole graph. More precisely, in the first two equations above, $\vec{f}$ and $\vec{g}$ are vector representations indexed by vertices in $B_1(x)$ as $\vec{f} = \begin{pmatrix} f(x) & f(y_1) & \cdots & f(y_m) \end{pmatrix}^\top$, where $S_1(x) = \{y_1, ..., y_m\}$. In the third equation, the vectors are indexed by vertices in $B_2(x)$ as

$$\vec{f} = \begin{pmatrix} f(x) & f(y_1) & \cdots & f(y_m) & f(z_1) & \cdots & f(z_n) \end{pmatrix}^\top,$$

where $S_2(x) = \{z_1, ..., z_n\}$.

The Bakry-Émery curvature $\mathcal{K}(G, x; N)$ can then be reformulated as the solution to the following semidefinite programming:

$$\begin{aligned} &\text{maximize } \ K \\ &\text{subject to } \ \Gamma_2(x) - \frac{1}{N}\Delta(x)\Delta(x)^\top - K\Gamma(x) \succeq 0. \end{aligned} \tag{P}$$

Here $M \succeq 0$ (resp. $M \succ 0$) means M is positive semidefinite (resp. positive definite). Note also that $\Gamma_2(x) - \frac{1}{N}\Delta(x)\Delta(x)^\top - K\Gamma(x)$ is a symmetric metric due to the symmetric bilinear forms Γ and Γ_2 at x. The above computing method has been studied by Schmuckenschläger [Sch99], and later on in e.g. [CLP20].

Chapter 16

Curvature as a minimal eigenvalue problem

The goal of this chapter is to present one of the main results from [CKLP21] which gives a reformulation of Bakry-Émery curvature $\mathcal{K}(G, x; N)$ (previously given by the semidifinite problem (P)) as a smallest eigenvalue problem by employing the Schur complement of a square block matrix M_{22} in $M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$, namely $M/M_{22} := M_{11} - M_{12}M_{22}^{-1}M_{21}$, applied to the matrix

$$\Gamma_2(x)_{\hat{1}} = \begin{pmatrix} \Gamma_2(x)_{S_1,S_1} & \Gamma_2(x)_{S_1,S_2} \\ \Gamma_2(x)_{S_2,S_1} & \Gamma_2(x)_{S_2,S_2} \end{pmatrix}.$$

Here, the matrix $\Gamma_2(x)_{\hat{1}}$ refers to the principle submatrix of $\Gamma_2(x)$ obtained by removing its first row and column corresponding to the vertex x. The matrix $\Gamma_2(x)_{S_i,S_j}$ refers to the submatrix of $\Gamma_2(x)$ whose rows and columns are indexed by the vertices of the combinatorial spheres $S_i(x)$ and $S_j(x)$.

We use the notation $Q(x) := \Gamma_2(x)_{\hat{1}}/\Gamma_2(x)_{S_2,S_2}$ for simplicity, and define

$$\begin{aligned} A_\infty(x) &:= 2\,\mathrm{diag}(\mathbf{v}_0(x))^{-1}Q(x)\,\mathrm{diag}(\mathbf{v}_0(x))^{-1}, \\ A_N(x) &:= A_\infty(x) - \frac{2}{N}\mathbf{v}_0(x)\mathbf{v}_0(x)^\top, \end{aligned} \tag{16.1}$$

where $\mathbf{v}_0(x) := (\sqrt{p_{xy_1}}\ \sqrt{p_{xy_2}}\ \dots\ \sqrt{p_{xy_m}})^\top$ with $S_1(x) = \{y_1, y_2, \dots, y_m\}$ labelling the neighbours of x. Note that the matrices $Q(x), A_\infty(x), A_N(x)$ are all symmetric matrices, and that $A_N(x)$ is a rank one perturbation of $A_\infty(x)$. The formulation of Bakry-Émery curvature can be summarized as follows.

Theorem 16.1. *Let $G = (V, w, \mu)$ be a weighted graph. For $x \in V$ and $N \in (0, \infty]$, the Bakry-Émery curvature $\mathcal{K}(G, x; N)$ is the smallest eigenvalue of the symmetrix matrix $A_N(x)$ defined in (16.1), that is,*

$$\mathcal{K}(G, x; N) = \lambda_{\min}(A_N(x)).$$

16.1 Curvature matrix A_∞

This section is dedicated to the proof from [CKLP21] of Theorem 16.1 that the Bakry-Émery curvature at a vertex x, namely

$$\mathcal{K}(G, x; N) = \arg\max_K \left\{ \Gamma_2(x) - \frac{1}{N}\Delta(x)\Delta(x)^\top - K\Gamma(x) \succeq 0 \right\}$$

is equal to the smallest eigenvalue of the so-called ***curvature matrix*** $A_\infty(x)$.

First, let us recall that $\mathcal{K}(G, x; N)$ is a local concept and uniquely determined by the structure of the two-ball $B_2(x)$. In particular, the symmetric matrix $\Gamma_2(x)$ is of size $|B_2(x)|$, and the symmetric matrices $\Delta(x)\Delta(x)^\top$ and $\Gamma(x)$ are of sizes $|B_1(x)|$ (and trivially extended by zeros to matrices of sizes $|B_2(x)|$).

Schmuckenschläger [Sch99] observes that the size of these matrices can be reduced by one: since $\Gamma_2(f), \Gamma(f), \Delta f$ all vanish for constant functions f, the curvature-dimension inequality $BE(K, N)$ remains valid after shifting f by an additive constant. It is therefore sufficient to verify (15.2) for all functions $f : V \to \mathbb{R}$ with $f(x) = 0$. This observation allows us remove from these matrices the row and column corresponding to the vertex x, and we are able to reformulate the problem (P) as

$$\begin{aligned} &\text{maximize } \ K && (P') \\ &\text{subject to } \ M_{K,N}(x) := \left(\Gamma_2(x) - \frac{1}{N}\Delta(x)\Delta(x)^\top - K\Gamma(x)\right)_{S_1 \cup S_2, S_1 \cup S_2} \succeq 0. \end{aligned}$$

Next we recall the concept of the Schur complement, which allows us to further reduce the size of the involved symmetric matrices in (P').

Lemma 16.2. *Consider a real symmetric matrix $M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$, where M_{11} and M_{22} are square submatrices, and assume that $M_{22} \succ 0$. The* ***Schur complement*** *M/M_{22} is defined as*

$$M/M_{22} := M_{11} - M_{12}M_{22}^{-1}M_{21}. \tag{16.2}$$

Then $M/M_{22} \succeq 0$ if and only if $M \succeq 0$.

The proof of this lemma can be found in, e.g., [G+10, Proposition 2.1] or [CLP20, Proposition 5.13]. We aim to apply this lemma for the symmetric matrix $M_{K,N}(x)$ given in (P'). Since $\Delta(x)$ and $\Gamma(x)$ have zero entries in the $S_2(x)$-structure, it means the matrix $M_{K,N}(x)$ has the following block structure:

$$M_{K,N}(x) = \begin{pmatrix} \Gamma_2(x)_{S_1,S_1} - \frac{1}{N}\Delta(x)_{S_1}\Delta(x)_{S_1}^\top - K\Gamma(x)_{S_1,S_1} & \Gamma_2(x)_{S_1,S_2} \\ \Gamma_2(x)_{S_2,S_1} & \Gamma_2(x)_{S_2,S_1} \end{pmatrix}.$$

By folding $M_{K,N}(x)$ into the upper left block, the Schur complement is given by

$$\begin{aligned} &M_{K,N}(x)/\Gamma_2(x)_{S_2,S_2} \\ &= \Gamma_2(x)_{S_1,S_1} - \frac{1}{N}\Delta(x)_{S_1}\Delta(x)_{S_1}^\top - K\Gamma(x)_{S_1,S_1} - \Gamma_2(x)_{S_1,S_2}\Gamma_2(x)_{S_2,S_2}^{-1}\Gamma_2(x)_{S_2,S_1} \\ &= Q(x) - \frac{1}{N}\Delta(x)_{S_1}\Delta(x)_{S_1}^\top - K\Gamma(x)_{S_1,S_1}, \end{aligned}$$

where $Q(x) := \begin{pmatrix} \Gamma_2(x)_{S_1,S_1} & \Gamma_2(x)_{S_1,S_2} \\ \Gamma_2(x)_{S_2,S_1} & \Gamma_2(x)_{S_2,S_2} \end{pmatrix} / \Gamma_2(x)_{S_2,S_2}$.

The importance of $\Gamma_2(x)_{S_1,S_1}$ for a lower curvature bound was already mentioned in Schmuckenschläger [Sch99, pp. 194–195] where he used the notation A_{II}. It is then implied by Lemma 16.2 that

$$\mathcal{K}(G,x;N) = \arg\max_K \left\{ Q(x) - \frac{1}{N}\Delta(x)_{S_1}\Delta(x)_{S_1}^\top - K\Gamma(x)_{S_1,S_1} \succeq 0 \right\}. \quad (16.3)$$

One can see from Proposition 15.1(L1) that $\Gamma(x)_{S_1,S_1} = \frac{1}{2}\operatorname{diag}(\Delta(x)_{S_1})$ and $\Delta(x)_{S_1} = (p_{xy_1}\ p_{xy_2}\ \dots\ p_{xy_m})^\top$, where $S_1(x) = \{y_1, y_2, \dots, y_m\}$.

Denote the vector $\mathbf{v}_0 := \mathbf{v}_0(x) = (\sqrt{p_{xy_1}}\ \sqrt{p_{xy_2}}\ \dots\ \sqrt{p_{xy_m}})^\top$. The maximizer K in (16.3) does not change under the multiplication by $\operatorname{diag}(\mathbf{v}_0)^{-1} \succ 0$ both from left and right sides, that is,

$$\mathcal{K}_{G,x}(N) = \arg\max_K \left\{ \operatorname{diag}(\mathbf{v}_0)^{-1} Q(x) \operatorname{diag}(\mathbf{v}_0)^{-1} - \frac{1}{N}\mathbf{v}_0\mathbf{v}_0^\top - \frac{K}{2}\mathrm{Id} \succeq 0 \right\}. \quad (16.4)$$

We finally deduce that

$$\begin{aligned} \mathcal{K}(G,x;N) &= \lambda_{\min}\left(2\operatorname{diag}(\mathbf{v}_0)^{-1}Q(x)\operatorname{diag}(\mathbf{v}_0)^{-1} - \frac{2}{N}\mathbf{v}_0\mathbf{v}_0^\top\right) \\ &= \lambda_{\min}\left(A_\infty(x) - \frac{2}{N}\mathbf{v}_0\mathbf{v}_0^\top\right) = \lambda_{\min}\left(A_N(x)\right), \end{aligned}$$

as desired.

Results in later sections and chapters are sometimes presented in terms of Bakry-Émery curvature at dimension $N = \infty$. For convenience, we use a shortened notation for the curvature as defined below.

Definition 16.3 (curvature at $N = \infty$)**.** When a weighted graph $G = (V, w, \mu)$ is fixed and unambiguous, the Bakry-Émery curvature at $x \in V$ is given by

$$\begin{aligned}\mathcal{K}_\infty(x) := \mathcal{K}(G, x; \infty) &= \max\{\, K \in \mathbb{R} \mid \Gamma_2 f(x) \geq K\Gamma f(x) \quad \forall f \in \mathbb{R}^V \,\} \\ &= \lambda_{\min}\left(A_\infty(x)\right).\end{aligned}$$

Remark 16.4. Both the Bakry-Émery curvature $\mathcal{K}_\infty(x)$ and the curvature matrix $A_\infty(x)$ are completely determined by the weighted structure of the ***incomplete two-ball*** around x, namely $B_2^{\text{inc}}(x)$, which is obtained from the induced subgraph of $B_2(x)$ by removing all edges connecting vertices within $S_2(x)$. The explicit expression of $A_\infty(x)$ is provided in [CKLP21, Appendix A].

16.2 Curvature of Cartesian products and graph modifications

One important application of the curvature matrix A_∞ is the curvature computation of the Cartesian products. Let us first provide the definition of (weighted) Cartesian product for weighted graphs.

Definition 16.5 (weighted Cartesian product)**.** Given two weighted graphs G, G' and two fixed positive numbers $\alpha, \beta \in \mathbb{R}^+$, the weighted Cartesian product $G \times_{\alpha,\beta} G'$ is defined with the following weight function and vertex measure: for $x, y \in G$ and $x', y' \in G'$,

$$\begin{aligned} w_{(x,x')(y,x')} &:= \alpha w_{xy}\mu_{x'}, \\ w_{(x,x')(x,y')} &:= \beta w_{x'y'}\mu_x, \\ \mu_{(x,x')} &:= \mu_x\mu_{x'}. \end{aligned}$$

The parameters α and β serve two purposes.

1. In the case of non-weighted graphs G and G' (i.e., $\mu \equiv 1$ and $w \in \{0, 1\}$), the choice of $\alpha = \beta = 1$ gives the usual Cartesian product graph $G \times G'$.

2. In the case of G and G' representing Markov chains (i.e., when $\sum_y w_{xy} = \mu_x$ and $\sum_{y'} w_{x'y'} = \mu_{x'}$), the choice of $\alpha + \beta = 1$ gives the weighted product $G \times_{\alpha,\beta} G$ which represents the random walk with probability α and β following horizontal and vertical edges, respectively.

It was shown in [CKLP21] that the curvature matrix of the Cartesian product of two graphs is simply the direct sum of the curvature matrices of each graph.

The proof of this result, which we omit here, relies on the explicit formula for the curvature matrix.

Theorem 16.6. *The curvature matrix of the product $G \times_{\alpha,\beta} G'$ is the weighted direct sum of the curvature matrices G and G':*

$$A_\infty^{G\times_{\alpha,\beta}G'}((x,x')) = \alpha A_\infty^G(x) \oplus \beta A_\infty^{G'}(x').$$

As an immediate consequence, the curvature of the product is the minimum of the curvature in each factor.

Corollary 16.7. *The curvature of the weighted product $H := G \times_{\alpha,\beta} G'$ is*

$$\mathcal{K}_\infty^H((x,x')) = \min\{\alpha\mathcal{K}_\infty^G(x), \beta\mathcal{K}_\infty^{G'}(x')\}.$$

Example 16.8 (hypercubes)**.** Consider the weighted graph K_2 consisting of two points x and y with the Laplacian given by

$$\Delta f(x) = p_{xy}(f(y) - f(x)) \text{ and } \Delta f(y) = p_{yx}(f(x) - f(y)),$$

where $p_{xy} = \frac{w_{xy}}{\mu_x}$ and $p_{yx} = \frac{w_{yx}}{\mu_y} = \frac{w_{xy}}{\mu_y}$. Explicit expressions of Γ and Γ_2 can be derived as

$$\begin{aligned} 2\Gamma(f,g)(x) &= p_{xy}(f(y) - f(x))(g(y) - g(x)),\\ 2\Gamma(f,g)(y) &= p_{yx}(f(y) - f(x))(g(y) - g(x)), \end{aligned}$$

and

$$\begin{aligned} 2\Gamma_2(f,g)(x) &= \Delta\Gamma(f,g)(x) - \Gamma(f,\Delta g)(x) - \Gamma(\Delta f, g)(x)\\ &= p_{xy}\big(\Gamma(f,g)(y) - \Gamma(f,g)(x)\big) - \frac{1}{2}p_{xy}(f(y) - f(x))(\Delta g(y) - \Delta g(x))\\ &\quad - \frac{1}{2}p_{xy}(\Delta f(y) - \Delta f(x))(g(y) - g(x))\\ &= \big(\frac{1}{2}p_{xy}(p_{yx} - p_{xy}) + 2\cdot\frac{1}{2}p_{xy}(p_{yx} + p_{xy})\big)(f(y) - f(x))(g(y) - g(x))\\ &= \big(\frac{3}{2}p_{xy}p_{yx} + \frac{1}{2}p_{xy}^2\big)(f(y) - f(x))(g(y) - g(x)). \end{aligned}$$

In particular, $\Gamma_2 f(x) = \big(\frac{3}{2}p_{yx} + \frac{1}{2}p_{xy}\big)\Gamma f(x)$ and thus the curvature is $\mathcal{K}_\infty(x) = \frac{3}{2}p_{yx} + \frac{1}{2}p_{xy}$.

Alternatively, the matrix forms $\Gamma(x)$ and $\Gamma_2(x)$ are given by

$$2\Gamma(x) = p_{xy}\begin{pmatrix} 1 & -1\\ -1 & 1\end{pmatrix}, \quad 2\Gamma_2(x) = \big(\frac{3}{2}p_{xy}p_{yx} + \frac{1}{2}p_{xy}^2\big)\begin{pmatrix} 1 & -1\\ -1 & 1\end{pmatrix}.$$

The Schur complement term is simply $Q(x) = \Gamma_2(x)_{S_1,S_1}$ since $S_2(x)$ is empty. The curvature matrix $A_\infty(x)$ (of size $|S_1(x)| = 1$) then equals

$$\begin{aligned} A_\infty(x) &= 2\,\mathrm{diag}(\mathbf{v}_0)^{-1} Q(x)\,\mathrm{diag}(\mathbf{v}_0)^{-1} \\ &= 2(\sqrt{p_{xy}})^{-1}\big(\frac{3}{4}p_{xy}p_{yx} + \frac{1}{4}p_{xy}^2\big)(\sqrt{p_{xy}})^{-1} = \big(\frac{3}{2}p_{yx} + \frac{1}{2}p_{xy}\big), \end{aligned}$$

and of course, $\mathcal{K}_\infty(x) = \lambda_{\min}(A_\infty) = \frac{3}{2}p_{yx} + \frac{1}{2}p_{xy}$.

In the particular case of the non-weighted K_2 (so $p_{xy} = p_{yx} = 1$), its curvature matrix is $A_\infty^{K_2}(x) = (2)$. By viewing the n-dimensional hypercube Q^n as the Cartesian product $(K_2)^n$, its curvature matrix and Bakry-Émery curvature are given by

$$A_\infty^{Q^n}(x) = \begin{pmatrix} 2 & & \\ & \ddots & \\ & & 2 \end{pmatrix}_{n\times n}, \text{ and } \mathcal{K}_\infty^{Q^n}(x) = 2.$$

We would also like to present without proof another application of the relation between $\mathcal{K}_\infty(x)$ and $A_\infty(x)$ to derive a certain graph modification result in [CKLP21].

Theorem 16.9. *Let $G = (V, w, \mu)$ be a weighted graph and fix a vertex $x \in V$. Assume that $p^-(y) = p_{yx}$ is independent of $y \in S_1(x)$.*[1] *Consider a modified weighted graph $\widetilde{G}$ obtained from G by one of the following operations:*

(O1) Increase the edge-weight between a fixed pair $y, y' \in S_1(x)$ with $y \neq y'$ by $\tilde{w}_{yy'} = w_{yy'} + C_1$ for any constant $C_1 > 0$.

(O2) Delete a vertex $z_0 \in S_2(x)$ and remove all of its incident edges, i.e., $\tilde{w}_{yz_0} = 0$ for all $y \in S_1(x)$. Increase the edge-weight between all pairs $y, y' \in S_1(x)$ with $y \neq y'$ by

$$\tilde{w}_{yy'} = w_{yy'} + C_2 w_{yz_0} w_{z_0 y'} \tag{16.5}$$

with any constant $C_2 \geq \dfrac{p^-(y)}{\mu_x p^{(2)}_{xz_0}}$.

Then $\mathcal{K}(\widetilde{G}, x; N) \geq \mathcal{K}(G, x; N)$ for all $N \in (0, \infty]$.

[1] A vertex x satisfying this property is called ***S_1-in regular***.

Chapter 17

Discrete Laplacian and heat semigroup operator

In this chapter, we present results which go in parallel with Sections 5.3 and 5.5 about the Laplacian, the spectral gap and the heat semigroup operator.

As usual, let $G = (V, w, \mu)$ be a connected and locally finite weighted graph with its graph Laplacian Δ given by

$$\Delta f(x) = \sum_{y \in S_1(x)} p_{xy}(f(y) - f(x))$$

for all $f \in \mathbb{R}^V$ and all $x \in V$. The eigenvalues-eigenvectors of Δ are the solutions $(\lambda, f) \in \mathbb{R} \times \mathbb{R}^V$ to the equation $\Delta f + \lambda f = 0$ on G. The eigenspace of each λ is the vector space consisting of all eigenfunctions with respect to λ. If G is finite, its Laplacian eigenvalues are discrete: $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{|V|-1}$ including their multiplicities. If G is infinite, the Laplacian is self-adjoint operator (under suitable boundary conditions), and its spectrum is contained in $[0, \infty)$. For example, if G is a non-weighted d-regular infinite tree, the spectrum of the non-normalized Laplacian is given by

$$\sigma(-\Delta) = \left[d - 2\sqrt{d-1}, d + 2\sqrt{d-1}\right].$$

Since the d-regular tree is the universal covering of any connected d-regular graph, this interval agrees precisely with the interval required for the nontrivial eigenvalues (excluding 0, and $2d$ in the bipartite case) of d-regular Ramanujan graphs.

Apart from the trivial $\lambda_0 = 0$ which corresponds to the constant eigenfunction f, the smallest nonzero eigenvalue λ_1 is the Laplacian spectral gap (and it is closely related to difference between the two largest eigenvalues of the graph's adjacency matrix).

17.1 Lichnerowicz spectral gap theorem

In the smooth setting of Riemannian manifolds, we already obtain Lichnerowicz's bound on the smallest nonzero eigenvalue by taking an integral of Bochner's formula and applying the divergence theorem (see Section 5.3). Here we present an analogous result on weighted graphs (see also, e.g. [BCLL17, Theorem 2.1]).

Theorem 17.1 (Lichnerowicz for $BE(K,N)$)**.** *Fix $N \in [0,\infty)$ and let $G = (V,w,\mu)$ be a weighted graph with* $\mathrm{Deg}_{\max} < \infty$ *and $K := \mathcal{K}(G,x;N) > 0$. Then the smallest nonzero Laplacian eigenvalue satisfies*

$$\lambda_1 \geq \frac{N}{N-1}K.$$

Similarly, for $N = \infty$, we have $\lambda_1 \geq K$.

Proof. First, we remark that G is finite due to the Bonnet-Myers diameter bound (see Corollary 18.2), since G satisfies $BE(K,N)$ (and hence $BE(K,\infty)$) with $K > 0$.

Suppose f is an eigenfunction with respect to λ_1, so $\Delta f = -\lambda f$. Using $BE(K,N)$ condition and the divergence theorem $\langle \Delta g, 1\rangle = 1$ for all g (see Proposition 15.1(L3)), we deduce that

$$\begin{aligned}
0 &\leq \langle \Gamma_2 f - K\Gamma f - \frac{1}{N}(\Delta f)^2, 1\rangle \\
&\leq \langle \frac{1}{2}\Delta\Gamma f - \Gamma(f,\Delta f) - K\Gamma f - \frac{\lambda^2}{N} f^2, 1\rangle \\
&= \langle \lambda\Gamma f - K\Gamma f - \frac{\lambda^2}{N} f^2, 1\rangle \\
&= \langle (\lambda - K)(\frac{1}{2}\Delta(f^2) - f\Delta f) - \frac{\lambda^2}{N} f^2, 1\rangle \\
&= \langle (\lambda - K)(\lambda f^2) - \frac{\lambda^2}{N} f^2, 1\rangle \\
&= \left(\frac{N-1}{N}\lambda^2 - K\lambda\right)\langle f^2, 1\rangle.
\end{aligned}$$

Since $\lambda_1 > 0$, we conclude that $\lambda_1 \geq \frac{N}{N-1}K$ as desired.

□

17.2 Heat semigroup and gradient estimate

In this section, we introduce and discuss properties of the heat semigroup operators P_t and the heat kernel p_t on weighted graphs. They are very important in the analysis of Bakry-Émery curvature, especially the semigroup characterization via gradient estimate. The definitions and properties of the heat semigroup operators and the heat kernel on graphs are in a similar fashion to those on manifolds, which we discussed in Section 5.5. We refer to the arXiv version of Wojciechowski's PhD thesis [Woj08] for technical details in the case of graphs, including e.g. Dodzuik's construction of the heat kernel and the stochastic completeness. At the same time, we remark that our convention for the graph Laplacian Δ is weighted and has the opposite sign to the one in [Woj08], which is the combinatorial Laplacian $\Delta f = \sum_{y \in S_1(x)}(f(x) - f(y))$.

Let $G = (V, w, \mu)$ be a locally finite and connected weighted graph. We denote by $C_b(V) := \{ f : V \to \mathbb{R} \mid \|f\|_\infty < \infty \}$ the space of all bounded functions on V.

Definition 17.2. The ***heat kernel*** $p : [0, \infty) \times M \times M \to [0, 1]$ (written as $p_t(x, y) = p(t, x, y)$) is the smallest nonnegative ***fundamental solution of the heat equation***, that is, for every $f \in C_b(V)$, the function

$$u(t, x) = \sum_{y \in V} \mu_y p_t(x, y) f(y) \tag{17.1}$$

is a solution to the Cauchy problem

$$\begin{cases} \frac{\partial}{\partial t} u & = \Delta u, \\ u(0, \cdot) & = f. \end{cases} \tag{17.2}$$

Equivalently, the heat kernel is the smallest nonnegative function p which solves the heat equation

$$\frac{\partial}{\partial t} p_t(x, y) = \Delta p_t(x, y),$$

where the Laplacian Δ is taken with respect to either x or y, and p satisfies the initial condition

$$p_0(x, y) = \delta_x(y).$$

Furthermore, the ***heat semigroup operator*** $P_t : C_b(V) \to C_b(V)$, $t \geq 0$ is given by

$$P_t f(x) := \sum_{y \in V} \mu_y p_t(x, y) f(y),$$

for all $f \in C_b(V)$ and $x \in M$.

By definition, it means the function $u(t,x) = P_t f(x)$ solves the Cauchy problem (17.2), or in short, we write

$$\frac{\partial}{\partial t} P_t = \Delta P_t.$$

Fundamental and very useful properties of p_t and P_t are summarized in the following proposition.

Proposition 17.3. *P_t and p_t satisfy the following properties:*

(H1) $p_t(x,y) = p_t(y,x)$.

(H2) $p_t(x,y) > 0$ *for all* x, y *and* $t > 0$.

(H3) $\sum_{y\in V} \mu_y p_t(x,y) \leq 1$.

(H4) $p_{s+t}(x,y) = \sum_{z\in V} \mu_z p_s(x,z) p_t(z,y)$.

(H5) *If* $f(x) \geq 0$ *for all* x, *then* $P_t f(x) \geq 0$ *for all* x *and* $t \geq 0$. *Additionally, if* f *is not identically zero, then* $P_t f(x) > 0$ *all* x *and* $t > 0$.

(H6) $\|P_t f\|_\infty \leq \|f\|_\infty$ *for all* $f \in C_b(V)$.

(H7) $P_t \circ P_s = P_{s+t}$ *for all* $s, t \geq 0$.

(H8) $\Delta P_t = P_t \Delta$.

Proof. Properties (H1)-(H4) are due to [Woj08, Theorems 2.1.2 and 2.1.5]. (H5), (H6) and (H7) follow from (H2), (H3) and (H4), respectively, via $P_t f(x) = \sum_{y\in V} \mu_y p_t(x,y) f(y)$. Lastly, (H8) follows from the proof in [Woj08, Theorem 2.2.1]. □

The next theorem gives a characterization of the $BE(K,\infty)$ in terms of (Γ-type) gradient estimate of the heat semigroup P_t. This result should be compared to Theorem 5.23, and its proof also employs the same technique of the semigroup interpolation. We refer to a paper by Lin and Liu [LL15] for details about this characterization which also includes $BE(K,N)$ for finite N.

Theorem 17.4 (Semigroup characterization of $BE(K,\infty)$)**.** *Let* $G = (V, w, \mu)$ *be a weighted graph. A vertex* x *satisfies* $BE(K,\infty)$ *if and only if the* Γ***-gradient estimate:***

$$\Gamma(P_t f)(x) \leq e^{-2Kt} P_t(\Gamma f)(x) \tag{17.3}$$

holds true for all $f \in C_b(V)$ *and* $t \geq 0$.

Proof. ($\Rightarrow$): Fix $t \geq 0$ and let $F : [0,t] \to \mathbb{R}$ be a smooth function given by

$$F(s) := e^{-2Ks} P_s(\Gamma P_{t-s} f)(x).$$

We must show that $F(0) \leq F(t)$, so and suffices to prove $F'(s) \geq 0$ on $[0,t]$. Differentiation F and using the product rule and the chain rule, we obtain

$$F'(s) = e^{-2Ks}\Big[-2KP_s(\Gamma P_{t-s}f)(x) + (\frac{\partial}{\partial s}P_s)(\Gamma P_{t-s}f)(x) + P_s(\frac{\partial}{\partial s}\Gamma(P_{t-s}f))(x)\Big]$$

With the relation $\frac{\partial}{\partial s}P_s = \Delta P_s = P_s \Delta$ being replaced into the second summand, we obtain $F'(s) = e^{-2Ks}P_s Hf(x)$ where H denotes the operator

$$Hf := -2K\Gamma P_{t-s}f + \Delta(\Gamma P_{t-s}f) + \frac{\partial}{\partial s}\Gamma(P_{t-s}f).$$

Moreover, the last summand of H is equal to

$$\begin{aligned}
\frac{\partial}{\partial s}\Gamma(P_{t-s}f) &= \frac{\partial}{\partial s}\Gamma(P_{t-s}f, P_{t-s}f) \\
&= 2\Gamma\Big(\frac{\partial}{\partial s}(P_{t-s}f), P_{t-s}f\Big) \\
&= -2\Gamma(\Delta P_{t-s}f, P_{t-s}f) \\
&= 2\Gamma_2(P_{t-s}f) - \Delta(\Gamma P_{t-s}f).
\end{aligned}$$

Therefore $Hf(x) = -2K\Gamma(P_{t-s}f)(x) + 2\Gamma_2(P_{t-s}f)(x) \geq 0$, since x satisfies $BE(K,\infty)$. Proposition 17.3(H5) then implies that $P_s Hf \geq 0$, which gives $F'(s) \geq 0$ as desired.

($\Leftarrow$): We differentiate (17.3) at $t = 0$ and obtain

$$\begin{aligned}
0 &\leq \lim_{t\to 0}\frac{1}{t}\big(e^{-2Kt}P_t(\Gamma f) - \Gamma(P_t f)\big) \\
&= \frac{\partial}{\partial t}\Big|_{t=0}\big(e^{-2Kt}P_t(\Gamma f) - \Gamma(P_t f)\big) \\
&= \Delta(\Gamma f) - 2K\Gamma f - \frac{\partial}{\partial t}\Big|_{t=0}\Gamma(P_t f, P_t f) \\
&= \Delta(\Gamma f) - 2K\Gamma f + \Big[2\Gamma_2(P_t f) - \Delta(\Gamma(P_t f))\Big]_{t=0} \\
&= 2\Gamma_2 f - 2K\Gamma f,
\end{aligned}$$

and therefore x satisfies $BE(K,\infty)$ □

In [LMP17, Theorem 3.4], the authors provide equivalent conditions where the Γ-gradient estimate (17.3) is sharp. This result, which is stated and proved below, will be helpful when we discuss the rigidity of diameter bound in an upcoming chapter.

Theorem 17.5. *Let $G = (V, w, \mu)$ be a weighted graph with $\mathrm{Deg}_{\max} < \infty$ that satisfies $BE(K, \infty)$ with $K > 0$. Let $f \in C_b(V)$. Then the following are equivalent.*

(i) *There exists some $x_0 \in V$ such that $\Gamma P_t f(x_0) = e^{-2Kt} P_t \Gamma f(x_0)$ for all $t \geq 0$.*

(ii) $\Gamma P_t f = e^{-2Kt} P_t \Gamma f$ *for all* $t \geq 0$.

(iii) $\Gamma_2 f = K \Gamma f$.

(iv) *$f = \varphi + c$ for a constant $c \in \mathbb{R}$ and an eigenfunction φ such that $\Delta\varphi + K\varphi = 0$.*

Moreover, if any of the above holds, then

(a) *Γf is constant.*

Proof. Firstly, G is finite due to the Bonnet-Myers diameter bound (Corollary 18.2). We will prove this equivalence in the following order:

$$(ii) \Rightarrow (i) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (ii).$$

$(ii) \Rightarrow (i)$ is trivial.

$(i) \Rightarrow (iii)$: Let $F_x(s) := e^{-2Ks} P_s(\Gamma P_{t-s} f)(x)$ for $s \in [0, t]$, and recall that $F_x'(s) = 2e^{-2Ks} P_s(\Gamma_2(P_{t-s} f) - K\Gamma(P_{t-s} f))(x)$. We know from (i) that

$$0 = F_{x_0}(t) - F_{x_0}(0) = \int_0^t 2e^{-2Ks} P_s(\Gamma_2(P_{t-s} f) - K\Gamma(P_{t-s} f))(x_0) ds.$$

Due to $BE(K, \infty)$, the above integrand is nonnegative, and hence it must be zero for all s. Particularly at $s = t$, we have $P_t(\Gamma_2 f - K\Gamma f)(x_0) = 0$ which holds for all $t > 0$. Proposition 17.3(H5) then asserts that $\Gamma_2 f - K\Gamma f$ is identically zero (otherwise $P_t(\Gamma_2 f - K\Gamma f)(x_0) > 0$), so (iii) is true.

Next, suppose for the sake of contradiction that (ii) is false, that is, $\Gamma P_t f < e^{-2Kt} P_t \Gamma f < 0$ at some $x_1 \in V$ and $t > 0$. Then

$$0 < F_{x_1}(t) - F_{x_1}(0) = \int_0^t 2e^{-2Ks} P_s(\Gamma_2(P_{t-s} f) - K\Gamma(P_{t-s} f))(x_1) ds,$$

which implies $\Gamma_2(P_{t-s} f) - K\Gamma(P_{t-s} f)$ is not identically zero. Proposition 17.3(H5) asserts that $P_s(\Gamma_2 P_{t-s} - K\Gamma P_{t-s}) > 0$ for all $x \in V$ and all $0 < s < t$, which in turn gives $\Gamma P_t f - e^{-2Kt} P_t \Gamma f < 0$ for all $x \in V$, contradicting to (1).

$(iii) \Rightarrow (iv)$: Using Proposition 15.1(L3), we derive

$$\begin{aligned}\langle \Gamma_2 f, 1\rangle = \langle -\Gamma(f, \Delta f), 1\rangle &= \langle \frac{1}{2} f\Delta(\Delta f) + \frac{1}{2}(\Delta f)^2, 1\rangle \\ &= \frac{1}{2}\langle \Delta(\Delta f), f\rangle + \frac{1}{2}\langle \Delta f, \Delta f\rangle \\ &= \langle \Delta f, \Delta f\rangle.\end{aligned}$$

On the other hand, $\Gamma_2 f = K\Gamma f$ gives

$$\langle \Gamma_2 f, 1\rangle = K\langle \Gamma f, 1\rangle = -K\langle f\Delta f, 1\rangle = -K\langle f, \Delta f\rangle,$$

and therefore we have $-K\langle f, \Delta f\rangle = \langle \Delta f, \Delta f\rangle$.

Next, we spectrally decompose f into $f = \sum_{i=0}^{|V|-1} \alpha_1 f_i$, where $\alpha_i \in \mathbb{R}$ and f_i is an eigenfunction corresponding to the Laplacian eigenvalue λ_i (listed as $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{|V|-1}$), and $\{f_i\}_{i=0}^{|V|-1}$ forms an orthonormal basis of $\mathbb{R}^V$, that is, $\langle f_i, f_j\rangle = \delta_{ij}$. This decomposition gives

$$K\sum_i \lambda_i \alpha_i^2 = -K\langle f, \Delta f\rangle = \langle \Delta f, \Delta f\rangle = \sum_i \lambda_i^2 \alpha_i^2,$$

which means $\sum_i \lambda_i(\lambda_i - K)\alpha_i^2 = 0$. Recall that $\lambda_0 = 0$ and from Lichnerowicz theorem that $\lambda_i \geq K$ for all $i \geq 1$. It implies that $\alpha_i = 0$ for all i such that $\lambda_i \notin \{0, K\}$, and hence we can write $f = c + \varphi$ with some constant $c \in \mathbb{R}$ and $\Delta\varphi + K\varphi = 0$.

Next, we prove $(iv) \Rightarrow (a)$, which will help proving $(iv) \Rightarrow (ii)$ later. We know from $f = \varphi + c$ that $\Gamma f = \Gamma\varphi$. The condition $BE(K, N)$ applied for φ gives

$$K\Gamma\varphi \leq \Gamma_2\varphi = \frac{1}{2}\Delta((\Gamma\varphi) - \Gamma(\varphi, \Delta\varphi) = \frac{1}{2}\Delta(\Gamma\varphi) + K\Gamma\varphi,$$

and thus $\Delta(\Gamma\varphi) \geq 0$. Due to the divergence theorem $\langle \Delta(\Gamma\varphi), 1\rangle = 0$, it must be that $\Delta(\Gamma\varphi) = 0$, and hence $\Gamma\varphi$ is constant as desired.

$(iv) \Rightarrow (ii)$: Fix any $x \in V$ and let $H(t) := P_t\varphi(x) - e^{-Kt}\varphi(x)$. Note that $H(0) = 0$ and $H'(t) = \Delta P_t\varphi(x) + Ke^{-Kt}\varphi(x)$.

By applying $P_t = e^{t\Delta} = \sum_{n=0}^{\infty} \frac{t^n\Delta^n}{n!}$ to φ, we have

$$P_t\varphi = \sum_{n=0}^{\infty} \frac{t^n(-K)^n\varphi}{n!} = e^{-Kt}\varphi.$$

Using $f = \varphi + c$, we finally derive

$$\Gamma P_t f = \Gamma P_t\varphi = \Gamma e^{-Kt}\varphi = e^{-2Kt}\Gamma\varphi = e^{-2Kt}\Gamma f \overset{(a)}{=} e^{-2Kt}P_t\Gamma f.$$

□

Chapter 18

Bonnet-Myers diameter bound and rigidity of hypercube

In this chapter, we survey for a Bonnet-Myers diameter bound result in the context of Bakry-Émery curvature, and its correspoding rigidity result from the two papers by Liu, Münch and Peyerimhoff [LMP18, LMP17].

18.1 Diameter bound for $BE(K, \infty)$

For the sake of simplicity, the results in this survey are presented only in the case of $N = \infty$. The arguments for curvature at a finite dimension N are carried out in a more complicated fashion and can be found in the original paper [LMP18].

Theorem 18.1 ([LMP18, Theorem 2.1])**.** *Let $G = (V, w, \mu)$ be a weighted graph with the positive lower curvature bound $K := \inf_{x \in V} \mathcal{K}_\infty(x) > 0$ and $\mathrm{Deg}_{\max} < \infty$. Then for all $x_0, y_0 \in V$, we have*

$$d(x_0, y_0) \leq \frac{\sqrt{2\,\mathrm{Deg}(x_0)} + \sqrt{2\,\mathrm{Deg}(y_0)}}{K}. \tag{18.1}$$

Proof. Let $L := d(x_0, y_0)$ and define a function $f(x) := \max\{L - d(x_0, x), 0\}$, which is bounded and 1-Lipschitz with $f(x_0) = L$ and $f(y_0) = 0$. Moreover,

$$\Gamma f(x) = \frac{1}{2} \sum_y p_{xy} (f(y) - f(x))^2 \leq \frac{1}{2} \sum_y p_{xy} = \frac{\mathrm{Deg}(x)}{2},$$

which means $\|\Gamma f\|_\infty \leq \frac{\mathrm{Deg}_{\max}}{2}$.

The triangle inequality gives, for all $t \geq 0$,

$$\begin{aligned} L &= |f(x_0) - f(y_0)| \\ &\overset{\triangle}{\leq} |f(x_0) - P_t f(x_0)| + |P_t f(x_0) - P_t f(y_0)| + |P_t f(y_0) - f(y_0)|. \end{aligned} \tag{18.2}$$

For any $x \in V$, the term $|f(x_0) - P_t f(x_0)|$ is bounded from above by

$$|f(x) - P_t f(x)| \leq \int_0^t \left| \frac{\partial}{\partial s} P_s f(x) \right| ds = \int_0^t \left| \Delta P_s f(x) \right| ds, \tag{18.3}$$

due to the fundamental theorem of calculus. Moreover,

$$\begin{aligned} (\Delta P_s f(x))^2 \leq 2\operatorname{Deg}(x)\Gamma(P_s f)(x) &\leq 2\operatorname{Deg}(x) e^{-2Ks} P_s(\Gamma f)(x) \\ &\leq 2\operatorname{Deg}(x) e^{-2Ks} \|\Gamma f\|_\infty \leq \operatorname{Deg}(x)^2 e^{-2Ks}, \end{aligned} \tag{18.4}$$

due to Proposition 15.1(L2), Theorem 17.4, and Proposition 17.3(H6).

Combining (18.3) and (18.4) yields for all $x \in V$,

$$|f(x) - P_t f(x)| \leq \operatorname{Deg}(x) \int_0^t e^{-Ks} ds = \frac{\operatorname{Deg}(x)}{K}(1 - e^{-Kt}) \leq \frac{\operatorname{Deg}(x)}{K}.$$

Thus $|f(x_0) - P_t f(x_0)| \leq \frac{\operatorname{Deg}(x_0)}{K}$ and $|P_t f(y_0) - f(y_0)| \leq \frac{\operatorname{Deg}(y_0)}{K}$. We also know from (18.4) that $\Gamma P_t f \to 0$ as $t \to \infty$, which implies that $|P_t f(y) - P_t f(x)| \to 0$ for all $x \sim y$. Due to connectivity of G, the triangle inequality can be applied to deduce that $|P_t f(y) - P_t f(x)| \to 0$ also for all $x, y \in V$. Finally, we can conclude from (18.2) that $L \leq \frac{\operatorname{Deg}(x_0) + \operatorname{Deg}(y_0)}{K}$ as desired. □

As an immediate sequence of the above theorem, we obtain the following Bonnet-Myers-type diameter bound result.

Corollary 18.2 (Bonnet-Myers for $BE(K, \infty)$)**.** *Let $G = (V, w, \mu)$ be a weighted graph with the positive lower curvature bound $K := \inf_{x \in V} \mathcal{K}_\infty(x) > 0$ and $\operatorname{Deg}_{\max} < \infty$. Then G is finite and*

$$\operatorname{diam}(G) \leq \frac{2\operatorname{Deg}_{\max}}{K}. \tag{18.5}$$

We remark from Example 16.8 that in the non-weighted case, the curvature of n-dimensional hypercube $\mathcal{Q}^n$ is $\mathcal{K}_\infty^{\mathcal{Q}^n} = 2$, and $\operatorname{Deg}_{\max} = \operatorname{diam}(\mathcal{Q}^n) = n$. Thus, the diameter bound (18.5) is sharp for any hypercube. In fact, we will discuss in the next section that hypercubes are the only nonweighted graphs for which (18.5) is sharp.

Example 18.3 (antitrees)**.** Now we would like to make a comparison to Remark 4.3 about a paraboloid being a noncompact space with positive curvature everywhere but the infimum of the curvature is zero (and thus the classical Bonnet-Myers theorem does not apply). The family of graphs that has the same analogy are known as ***antitrees***, whose name and general definition come from a paper by Keller, Lenz and Wojciechowski [KLW13]. For an infinite sequence of positive integers $(a_k)_{k\in\mathbb{N}}$, an antitree with respect to (a_k) is defined to be a simple and connected graph with the vertex set

$$V = \bigsqcup_k V_k, \quad |V_k| = a_k,$$

and satisfies the following two properties.

(AT1) Any root vertex $x \in V_1$ is connected to all vertices in V_2 and no vertex in V_k, $k \geq 3$.

(AT2) Any vertex $x \in V_k$, $k \geq 2$, is connected to all vertices in V_{k-1} and V_{k+1} and no vertex in V_l, $|l - k| \geq 2$.

In [CLMP20], the authors provide explicit formulae to calculate both the Ollivier Ricci curvature and (non-weighted) Bakry-Émery curvature for antitrees which satisfy additional properties that (a_k) is an increasing sequence, and any $x \in V_k$ is connected to all other vertices in V_k. Such a unique antitree is denoted by $\mathcal{AT}((a_k))$. In the particular case that $a_k = 1 + (k-1)t$ for a fixed $t \in \mathbb{N}$, the infinite antitree $\mathcal{AT}((a_k))$ has positive curvature everywhere and the infimum of the curvature is zero (in both contexts of Bakry-Émery curvature and Ollivier Ricci curvature); see [CLMP20, Theorems 2.1.4 and 2.1.6]. Thus these antitrees serve to be an example of infinite graphs with positive curvature everywhere but a Bonnet-Myers-type diameter bound theorem is not applicable.

18.2 Rigidity of hypercubes

Continuing from the Bonnet-Myers-type diameter bound result in the previous section, we now present the main result in [LMP17] about the rigidity in the special case of non-weighted graphs; see Cheng's rigidity on manifolds (Theorem 4.2) for comparison. Readers are welcome to consult the original paper for more technical details when dealing with weighted graphs.

As usual, our non-weighted graph $G = (V, E)$ here is connected and locally finite. Moreover, we assume that G has finite maximal vertex degree, $\mathrm{Deg}_{\max} < \infty$.

Theorem 18.4. *Let $G = (V, E)$ be a graph with $K := \inf_{x \in V} \mathcal{K}_\infty(x) > 0$. Then G satisfies the sharp Bonnet-Myers diameter bound, i.e.,*

$$\operatorname{diam}(G) = \frac{2\operatorname{Deg}_{\max}}{K}$$

if and only if G is a hypercube.

We summarize the authors' ideas to prove the forward implication of the above theorem into the following three steps (see Theorems 18.5, 18.7 and 18.8). Afterwards, we present only the proof of Theorem 18.5 to demonstrate how to utilize the sharpness of inequalities involved in the derivation of the distance bound (18.1).

Theorem 18.5. *Let $G = (V, E)$ be a graph with $K := \inf_{x \in V} \mathcal{K}_\infty(x) > 0$, and let $D := \operatorname{Deg}_{\max}$. Suppose G satisfies the sharp Bonnet-Myers diameter bound, $\operatorname{diam}(G) = \frac{2D}{K}$. Let $x_0 \in V$ be a pole, that is, there exists $y_0 \in V$ with $d(x_0, y_0) = \operatorname{diam}(G)$. Then $K = 2$ and G has the **hypercube shell structure** $HSS(D, 1, x_0)$, which means*

(i) G is D-regular and bipartite, and

(ii) the in-degree $d^-_{x_0}(x) := |\{v \in S_1(x) \mid d(x_0, v) = d(x_0, x) - 1\}|$ is equal to $d(x_0, x)$.

The authors of [LMP17] prove further that a graph $G = (V, E)$ with $K := \inf_{x \in V} \mathcal{K}_\infty(x) = 2$ must satisfy two combinatorial properties called the small two-sphere property (SSP) and the non-clustering property (NCP).

Definition 18.6. Let $G = (V, E)$ be a D-regular graph. We say a vertex $x \in V$ satisfy the ***small two-sphere property*** (SSP) if $|S_2(x)| \leq \binom{D}{2}$. We say x satisfies the ***non-clustering property*** (NCP) if the following implication holds true: if $d^-_x(z) = 2$ for all $z \in S_2(x)$, then for all different $y_1, y_2 \in S_1(x)$ there is at most one $z \in S_2(x)$ that $y_1 \sim z \sim y_2$. Moreover, G satisfies (SSP) (or (NCP)) if such property is satisfied for all $x \in V$.

Theorem 18.7. *Let $G = (V, E)$ be a D-regular bipartite graph. Suppose that a vertex $x \in V$ satisfies $\mathcal{K}_\infty(x) \geq 2$. Then x satisfies both (SSP) and (NCP). In particular, if $K := \inf_{x \in V} \mathcal{K}_\infty(x) = 2$, then G satisfies (SSP) and (NCP).*

The authors then conclude that (SSP), (NCP) and $\mathrm{HSS}(D, 1, x_0)$ for some $x_0 \in V$ altogether imply that G must be a D-dimensional hypercube.

Theorem 18.8. *Let $G = (V, E)$ be a graph with $HSS(D, 1, x_0)$ at some $x_0 \in V$. Suppose that G satisfies both (SSP) and (NCP). Then G is isomorphic to the D-dimensional hypercube Q^D.*

Proof of Theorem 18.5. We remark G is finite due to the Bonnet-Myers diameter bound theorem. Let $f_0 \in \mathbb{R}^V$ be given by $f_0(x) := d(x_0, x)$ for all $x \in V$. We follow closely the arguments in the proof of Theorem 18.1 that

$$\begin{aligned}
\operatorname{diam}(G) = f_0(y_0) - f_0(x_0) &\le \int_0^\infty |\Delta P_t f_0(x_0)| + |\Delta P_t f_0(y_0)| dt \\
&\le \sqrt{2D} \int_0^\infty \sqrt{\Gamma P_t f_0(x_0)} + \sqrt{\Gamma P_t f_0(y_0)} dt \qquad (18.6)\\
&\le \sqrt{2D} \int_0^\infty e^{-Kt}(\sqrt{P_t \Gamma f_0(x_0)} + \sqrt{P_t \Gamma f_0(y_0)}) dt \qquad (18.7)\\
&\le \sqrt{2D} \int_0^\infty 2e^{-Kt} \sqrt{\|\Gamma f_0\|_\infty} dt \\
&\le \sqrt{2D} \int_0^\infty 2e^{-Kt} \sqrt{\frac{D}{2}} dt = \frac{2D}{K}. \qquad (18.8)
\end{aligned}$$

The assumption of sharpness $\operatorname{diam}(G) = \frac{2D}{K}$ implies that all the involved inequalities must hold with equality. The equality of (18.6) implies $\operatorname{Deg}(x_0) = D$, and the equality of (18.7) implies that $\Gamma P_t f_0(x_0) = e^{-2Kt} P_t \Gamma f_0(x_0)$ for all $t \ge 0$. Theorem 17.5 asserts that Γf_0 is constant, and particularly for all $x \in V$ we have

$$2\Gamma f_0(x) = 2\Gamma f_0(x_0) = \sum_{y \in S_1(x_0)} (d(x_0, y) - d(x_0, x_0))^2 = D.$$

It follows that

$$D = 2\Gamma f_0(x) = \sum_{y \in S_1(x)} (d(x_0, y) - d(x_0, x))^2 \le \sum_{y \in S_1(x)} 1 = \operatorname{Deg}(x) \le D.$$

The sharpness of the above inequalities implies that $\operatorname{Deg}(x) = D$ for all x (i.e., G is D-regular) and that $|d(x_0, y) - d(x_0, x)| = 1$ for all $y \sim x$. This means there is no spherical edge ($x \sim y$ such that $d(x_0, y) = d(x_0, x)$), so G is indeed a bipartite graph (between vertices in $\bigcup_{i \text{ even}} S_i(x_0)$ and $\bigcup_{i \text{ odd}} S_i(x_0)$). Observe the Laplacian Δf_0 in terms of degrees:

$$\Delta f_0(x) = \sum_{y \in S_1(x)} (d(x_0, y) - d(x_0, x)) = d^+_{x_0}(x) - d^-_{x_0}(x) = D - 2d^-_{x_0}(x), \qquad (18.9)$$

where the last equality comes from the fact that $d^+_{x_0}(x) + d^-_{x_0}(x) = D$ since G is D-regular and $d^0_{x_0}(x) = 0$ (i.e., there is no spherical edge).

Recall again from Theorem 17.5 that $f_0 = \varphi + c$ where $\Delta\varphi + K\varphi = 0$, so

$$\Delta f_0 = \Delta\varphi = -K\varphi = -Kf_0 + cK.$$

To find the values of c and K, we evaluate at x_0 and at its neighbor x_1. At x_0, we have $f_0(x_0) = 0$ and $\Delta f_0(x_0) = D$, so the substitution into the above equation yields $cK = D$, which means $\Delta f_0 = -K f_0 + D$. At any $x_1 \in S_1(x_0)$, we have $f_0(x_1) = 1$ and from (18.9) that $\Delta f_0(x_1) = D - 2$, so the substitution yields $K = 2$. Therefore, $\Delta f_0 = -2f_0 + D$, and by comparing this equation to (18.9) we can conclude that $d^-_{x_0}(x) = f_0(x) = d(x_0, x)$ as desired. □

Chapter 19

Outlook for Bakry-Émery curvature

In this final chapter of Part III, we collect some interesting questions about Bakry-Émery curvature on non-weighted graphs and its connection to Ollivier Ricci curvature.

1. The Bonnet-Myers-type diameter bounds with respect to Ollivier Ricci curvature and to Bakry-Émery curvature are similar in comparison:

$$\operatorname{diam}(G) \leq \frac{2}{\inf_{x,y} \kappa(x,y)}, \tag{19.1}$$

$$\operatorname{diam}(G) \leq \frac{2}{\inf_{x} \mathcal{K}_\infty(x)}. \tag{19.2}$$

   However, we have learned that their corresponding rigidity results are different: (19.2) is only sharp for hypercubes whereas (19.1) is sharp for a larger class of graphs, including all self-centered Bonnet-Myers sharp graphs as listed in Theorem 13.26.

   In [CKK+20, Section 10.3], we pose an interesting conjecture that for all regular graphs $G = (V, E)$, the Bakry-Émery curvature satisfies the following inequality

$$\inf_{x \in V} \mathcal{K}_\infty(x) \leq \frac{1}{\deg(G)} + \frac{1}{\operatorname{diam}(G)}, \tag{19.3}$$

   which is an improvement of (19.2) in the case of $\deg(G) > \operatorname{diam}(G)$. In fact, it is shown that the new inequality (19.3) is sharp for all self-centered Bonnet-Myers sharp graphs.

2. A broader question about the relation between Ollivier Ricci curvature and Bakry-Émery curvature is to understand graphs which have positive (or non-

negative) curvature in both contexts. Prominent examples are abelian Cayley graphs. Some partial results in this direction can be found in a paper by Ralli [Ral17] and our joint paper [CKK+21]. In the latter paper, we also draw connections of Ollivier Ricci and Bakry-Émery to various versions of Ricci-flatness, a curvature notion which was originally introduced by Chung and Yau in [CY96].

Part IV

Erbar-Maas entropic Ricci curvature on Markov chains

Chapter 20

Setting of discrete Markov chains

20.1 Irreducible and reversible Markov chains and graph structure

We start with a finite Markov chain (X, Q), where X is a finite set, and $Q : X \times X \to [0,1]$ is a Markov kernel, where $Q(x,y)$ serves as the transition probability from x to y and it satisfies $\sum_{y \in X} Q(x,y) = 1$ for all $x \in V$. The kernel Q can be represented by a square matrix of size $|X|$ with nonnegative entries whose sum in each row equals one. The kernel Q induces a directed graph on X by assigning an directed edge $x \to y$ whenever $Q(x,y) > 0$.

Throughout this thesis, the Markov kernel Q is always assumed to be ***irreducible*** and ***reversible***, which we briefly discuss for readers' convenience.

Definition 20.1 (irreducibility)**.** Let A be an $n \times n$ square matrix with nonnegative entries. The matrix A is ***irreducible*** if, for any $k, l \leq n$, there exists $m \in \mathbb{N}$ such that $A^m_{k,l} > 0$.

The notion of irreducibility translates naturally for a Markov kernel. A Markov chain is irreducible means its induced directed graph is ***strongly connected***, i.e., it has a directed path from any vertex to any vertex.

Definition 20.2 (aperiodicity)**.** Let A be an irreducible $n \times n$ matrix A. Each index $1 \leq k \leq n$ has a ***period*** $p \in \mathbb{N}$ to be the greatest common divisor of all $m \in \mathbb{N}$ such that $A^m_{k,k} > 0$. The irreducibility asserts that all indexes share the same period p (and hence p is called the period of A). The matrix A is called ***aperiodic*** if $p = 1$.

Theorem 20.3 (Perron-Frobenius)**.** *Let A be an irreducible square matrix with all entries nonnegative. Then the spectral radius $\sigma(A) > 0$ is a real (left- and right-) eigenvalue $r > 0$ with multiplicity of one. The (left-) eigenvector corresponding to r has all entries positive. If A is additionally aperiodic, then all other eigenvalues λ has modulus strictly less than r.*

In the case of aperiodicity, the infimum $\inf\{\sigma(A) - |\lambda| : \lambda \text{ is eigenvalue}\}$ is strictly positive and called the spectral gap. In the particular case of a Markov kernel Q, the left-spectral radius equals one due to $\sum_{y \in X} Q(x, y) = 1$. Consequently, there exists a unique stationary probability measure π on X satisfying $\pi(x) > 0$ for all x, $\sum_{x \in X} \pi(x) = 1$ and $\sum_{x \in X} Q(x, y)\pi(x) = \pi(y)$ (or written compactly as $\pi^\top Q = \pi^\top$). The spectral gap in fact describes the rate of convergence (i.e., the mixing rate) from any initial probability distribution to the limiting distribution, which is π.

Definition 20.4 (reversibility)**.** An irreducible Markov kernel $Q : X \times X \to \mathbb{R}_{\geq 0}$ (with the unique stationary probability measure π) is ***reversible*** if it satisfies the following detailed balance equation

$$Q(x, y)\pi(x) = Q(y, x)\pi(y) \quad \forall x, y \in X. \tag{20.1}$$

Henceforth, the kernel Q, assumed to be irreducible and reversible, always comes with a unique stationary probability measure π, which is treated as a reference measure.

One may define the symmetric function $w : X \times X \to \mathbb{R}_{\geq 0}$ by $w_{xy} := Q(x, y)\pi(x) = w_{yx}$, which can be regarded as the symmetric weight function for the induced weighted graph $G = (X, w, \pi)$ (compared to the setting of weighted graphs in Section 15.1). This induced weighted graph is no longer a directed graph, since there exists an edge $x \sim y$ if and only if any of the following equivalences holds true: $Q(x, y) > 0 \Leftrightarrow w_{xy} > 0 \Leftrightarrow Q(y, x) > 0$. Moreover, we define the ***vertex π-degree*** to be $\mathrm{Deg}_\pi(x) := \frac{1}{\pi(x)} \sum_{y \in V} \pi(y)$. Note that this degree is not the same as the vertex degree for weighted graphs: $\mathrm{Deg}(x) := \frac{1}{\pi(x)} \sum_{y \in S_1(x)} w_{xy} = \sum_{y \in S_1(x)} Q(x, y) = 1$ for all x.

A special case is a simple random walk (without laziness) on a finite d-regular connected graph $G = (V, E)$ which is given by $X = V$, $w_{xy} = \frac{1}{d}$ iff $x \sim y$, and $\pi(x) = \frac{1}{|V|}$ for $x \in V$. In this case, we have $\mathrm{Deg}_\pi(x) = d$, the usual vertex degree.

20.2 Probability space and discrete differential operators

The space of our interest $\mathscr{P}(X) \subset \mathbb{R}^X$ is the space of all probability densities (with respect to the reference measure π), which is given by

$$\mathscr{P}(X) := \left\{ \rho : X \to \mathbb{R}_{\geq 0} \,\middle|\, \sum_{x \in X} \rho(x)\pi(x) = 1 \right\}.$$

Furthermore, a function $V : X \times X \to \mathbb{R}$ is called a ***discrete vector field*** if it satisfies

$$V(x,y) = \begin{cases} -V(y,x) & \text{for } x \sim y, \\ 0 & \text{for } x \not\sim y. \end{cases}$$

We denote the space of all discrete vector fields by $\mathfrak{X}(X) \subset \mathbb{R}^{X \times X}$.

Next we give the explicit formulae of the operators ***discrete gradient*** $\nabla : \mathbb{R}^X \to \mathfrak{X}(X)$, ***discrete divergence*** $\operatorname{div} : \mathfrak{X}(X) \to \mathbb{R}^X$, and ***discrete Laplacian*** $\Delta := \operatorname{div} \circ \nabla$ as well as the first Γ-iteration, namely $2\Gamma(f,g) := \Delta(fg) - f\Delta g - g\Delta f$.

Definition 20.5. For all $f, g \in \mathbb{R}^X$ and $V \in \mathfrak{X}(X)$,

$$\begin{aligned}
\nabla f(x,y) &:= \begin{cases} f(y) - f(x) & \text{for } x \sim y, \\ 0 & \text{for } x \not\sim y, \end{cases} \\
\operatorname{div} V(x) &:= \frac{1}{2} \sum_{y \in X} (V(x,y) - V(y,x)) Q(x,y) = \sum_{y \in X} V(x,y) Q(x,y), \\
\Delta f(x) &:= \sum_{y \in X} (f(y) - f(x)) Q(x,y), \\
2\Gamma(f,g)(x) &:= \sum_{y \in X} (f(y) - f(x))(g(y) - g(x)) Q(x,y).
\end{aligned}$$

The second Γ-iteration is defined by $2\Gamma_2(f,g) := \Delta\Gamma(f,g) - \Gamma(f,\Delta g) - \Gamma(g,\Delta f)$. For shortened notations, we also write $\Gamma f(x) := \Gamma(f,f)(x)$ and $\Gamma_2 f(x) := \Gamma_2(f,f)(x)$.

All these operators in the setting of Markov chains are the same ones as in the setting of weighted graphs where we discuss the Bakry-Émery curvature. Furthermore, inner-products for functions and for vector fields are defined as follows.

Definition 20.6. For $f_i \in \mathbb{R}^X$ and $V_i \in \mathfrak{X}(X)$,

$$\begin{aligned} \langle f_1, f_2 \rangle &:= \sum_{x \in X} f_1(x) f_2(x); \\ \langle f_1, f_2 \rangle_\pi &:= \sum_{x \in X} f_1(x) f_2(x) \pi(x); \\ \langle V_1, V_2 \rangle_\pi &:= \frac{1}{2} \sum_{x,y \in X} V_1(x,y) V_2(x,y) Q(x,y) \pi(x). \end{aligned}$$

The following proposition asserts that, with respect to the inner products $\langle \cdot, \cdot \rangle_\pi$, the divergence is the negative adjoint of the gradient (compared to (5.1) in the smooth setting of manifolds).

Proposition 20.7. *For all $f \in \mathbb{R}^X$ and $V \in \mathfrak{X}(X)$, we have $\langle \nabla f, V \rangle_\pi = -\langle f, \operatorname{div} V \rangle_\pi$.*

Proof. A straightforward calculation gives

$$\begin{aligned} \langle f, \operatorname{div} V \rangle_\pi &= \sum_{x \in X} f(x) \pi(x) \left(\frac{1}{2} \sum_{y \in X} (V(x,y) - V(y,x)) Q(x,y) \right) \\ &= \frac{1}{2} \sum_{x,y \in X} f(x) V(x,y) Q(x,y) \pi(x) - \frac{1}{2} \sum_{x,y \in X} f(x) V(y,x) Q(x,y) \pi(x) \\ &= \frac{1}{2} \sum_{x,y \in X} f(x) V(x,y) Q(x,y) \pi(x) - \frac{1}{2} \sum_{y,x \in X} f(y) V(x,y) Q(x,y) \pi(y) \\ &= \frac{1}{2} \sum_{x,y \in X} (f(x) - f(y)) V(x,y) Q(x,y) \pi(x) = -\langle \nabla f, V \rangle_\pi. \end{aligned}$$

where the third inequality is due to the interchange of x and y, together with the detailed balance equation $Q(y,x)\pi(y) = Q(x,y)\pi(x)$. □

While it is natural to multiply a function ρ to a vector field V on manifolds, such multiplication is not allowed in the discrete setting since a discrete vector field V is defined on the product space $X \times X$. In order to make them compatible, we transform $\rho \in \mathbb{R}^X_{\geq 0}$ into $\hat{\rho} \in \mathbb{R}^{X \times X}_{\geq 0}$ via the logarithmic mean, explained as follows

Definition 20.8. [Maa11, EM12] The function $\theta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is assumed to be the logarithmic mean, that is,

$$\theta(s,t) = \int_0^1 s^{1-\alpha} t^a d\alpha = \frac{t - s}{\log t - \log s},$$

where the convention in extremal cases is $\theta(0,t) = \theta(t,0) = 0$ and $\theta(t,t) = t$.

For $\rho \in \mathscr{P}(X)$, the average $\hat{\rho} : X \times X \to \mathbb{R}_{\geq 0}$ is defined as

$$\hat{\rho}(x,y) := \theta(\rho(x), \rho(y)).$$

Another inner-product for vector fields is defined with respect to $\rho \in \mathscr{P}(X)$ as

$$\langle V_1, V_2 \rangle_\rho := \langle \hat{\rho} V_1, V_2 \rangle_\pi = \frac{1}{2} \sum_{x,y \in X} V_1(x,y) V_2(x,y) \hat{\rho}(x,y) Q(x,y) \pi(x). \tag{20.2}$$

We remark that the symmetric property $\hat{\rho}(x,y) = \hat{\rho}(y,x)$ asserts that the multiplication $\hat{\rho}V$ is skew-symmetric and remains to be a vector field.

In [Maa11, EM12], the authors allow θ to be any suitable mean which satisfies certain properties: θ is continuous on $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ and C^∞ on $(0,\infty) \times (0,\infty)$; $\theta(s,t) = \theta(t,s)$; $\theta(s,t) > 0$ for $s,t > 0$; $\theta(r,t) \geq \theta(s,t)$ for $r \geq s$; $\theta(\lambda s, \lambda t) = \lambda\theta(s,t)$ and $\theta(1,1) = 1$.

The importance of the choice of θ being logarithmic mean (which is a crucial ingredient for Erbar-Maas Ricci curvature notion later) is to re-establish the chain rule "$\mathrm{grad}(\log \rho) = \frac{1}{\rho}\,\mathrm{grad}\,\rho$" that appears in the continuous setting:

$$\hat{\rho}(x,y) = \frac{\rho(y) - \rho(x)}{\log \rho(y) - \log \rho(x)} = \frac{\nabla \rho(x,y)}{\nabla \log \rho(x,y)}. \tag{20.3}$$

Another interesting choice of θ is the arithmetic mean, which relates Erbar-Maas curvature to Bakry-Émery curvature (see Proposition 22.6).

Chapter 21

Otto's calculus for discrete spaces

In this chapter, we revisit the arguments of Otto's calculus (see Chapter 2) about the Riemannian structure of the probability space $\mathscr{P}(X)$. However, since now X is a finite set, arguments certainly become rigorous for a finite-dimensional manifold $\mathscr{P}(X)$ (in comparison to Otto's formal calculus for an infinite-dimensional $\mathscr{P}(M)$).

21.1 Identification of tangent spaces of $\mathscr{P}(X)$

First, we regard as a manifold $\mathcal{M}$, the subspace of strictly positive probability densities

$$\mathcal{M} := \mathscr{P}_*(X) := \left\{ \rho : X \to \mathbb{R}^+ \middle| \sum_{x \in X} \rho(x)\pi(x) = 1 \right\} \subset \mathscr{P}(X).$$

Note that $\mathcal{M}$ is an open and convex subspace of the affine hyperplane $\mathcal{V} := \left\{ \rho \in \mathbb{R}^X \middle| \sum_{x \in X} \rho(x)\pi(x) = 1 \right\}$, so it is indeed a smooth manifold of dimension $|X| - 1$.

Its tangent space at $\rho \in \mathcal{M}$ can be naturally realized as

$$T_\rho \mathcal{M} = \left\{ (\rho, s) \text{ where } s \in \mathbb{R}^X \middle| \sum_{x \in X} s(x)\pi(x) = 0 \right\},$$

where the curve in $\mathcal{M}$ which passes through the point ρ in the direction s (in the small neighborhood of ρ) is given by $\rho(t) = \rho_t = \rho + ts \in \mathcal{M}$ for small enough $|t|$.

The main goal of this chapter is to follow the ideas from Otto's calculus and identify this tangent space $T_\rho\mathcal{M}$ with the following space

$$Tan_\rho\mathcal{M} := \left\{(\rho, \nabla\psi) \text{ where } \psi \in \mathbb{R}^X\right\}$$

through an identification map $\mathcal{T}(\rho) : Tan_\rho\mathcal{M} \xrightarrow{\sim} T_\rho\mathcal{M}$ given by

$$(\rho, \nabla\psi) \longmapsto (\rho, s = -\operatorname{div}(\hat{\rho}\nabla\psi)).$$

Theorem 21.1. *The map $\mathcal{T}(\rho) : Tan_\rho\mathcal{M} \to T_\rho\mathcal{M}$ is well-defined, and it is a linear isomorphism.*

Proof. Here the footpoint $\rho \in \mathcal{M}$ is fixed and omitted from calculation, that is, we only write

$$Tan_\rho\mathcal{M} = \left\{\nabla\psi : \psi \in \mathbb{R}^X\right\}, \text{ and } T_\rho\mathcal{M} = \{s \in \mathbb{R}^X : \sum_{x\in X} s(x)\pi(x) = 0\}.$$

First we check that the image $-\operatorname{div}(\hat{\rho}\nabla\psi)$ lies in $Tan_\rho\mathcal{M}$. This can be seen as a consequence of the divergence theorem

$$-\sum_{x\in X} \operatorname{div}(\hat{\rho}\nabla\psi)\pi(x) = -\langle 1, \operatorname{div}(\hat{\rho}\nabla\psi)\rangle_\pi = \langle \nabla 1, \hat{\rho}\nabla\psi\rangle_\pi = 0.$$

Now consider the following commutative diagram

$$\mathbb{R}^X \xrightarrow{\ \nabla\ } Tan_\rho\mathcal{M} \xrightarrow{\mathcal{T}(\rho) = -\operatorname{div}(\hat{\rho}\ \cdot)} T_\rho\mathcal{M} \hookrightarrow \mathbb{R}^X$$
$$\mathbb{R}^X \xrightarrow{\ \tilde{A}(\rho)\ } \mathbb{R}^X$$

where the map $\psi \mapsto \nabla\psi$ is a surjection on $Tan_\rho\mathcal{M}$ by construction, and $T_\rho\mathcal{M} \subset \mathbb{R}^X$ is a canonical embedding. The map $\tilde{A}(\rho) : \mathbb{R}^X \to \mathbb{R}^X$ is a linear operator given by $\tilde{A}(\rho)\psi := -\operatorname{div}(\hat{\rho}\nabla\psi)$. It satisfies

$$\begin{aligned}
-\operatorname{div}(\hat{\rho}\nabla\psi)(x) &= -\sum_{y\in X} \hat{\rho}(x,y)(\psi(y) - \psi(x))Q(x,y) \\
&= \left(\sum_{y\neq x} \hat{\rho}(x,y)Q(x,y)\right)\psi(x) - \sum_{y\neq x} \hat{\rho}(x,y)Q(x,y)\psi(y),
\end{aligned}$$

so its matrix representation is

$$\tilde{A}(\rho)_{x,y} = \begin{cases} \sum_{z\neq x} \hat{\rho}(x,z)Q(x,z) & \text{if } x = y, \\ -\hat{\rho}(x,y)Q(x,y)\psi(y) & \text{if } x \neq y. \end{cases}$$

Note that the matrix $\tilde{A}(\rho)$ has zero sum in each row, which corresponds to the fact that constant functions ϕ are in its kernel. We claim that the kernel of $\tilde{A}(\rho)$ has only one dimension and contains only the constant functions. Consider a vector $a = (a_y) \in \mathbb{R}^X$ in the kernel of $\tilde{A}(\rho)$ and denote $a_{\max} := \max_{y\in X}(a_y)$. Pick any $x_0 \in X$ such that $a_{x_0} = a_{\max}$. Since $-\tilde{A}(\rho)_{x_0,y}$ is nonnegative for all $y \neq x_0$ and vanishes for all $y \not\sim_Q x_0$, the relation $\sum_{y\in X} \tilde{A}(\rho)_{x_0,y} a_y = 0$ implies that

$$\tilde{A}(\rho)_{x_0,x_0} a_{x_0} = \sum_{\substack{y\neq x_0 \\ y\sim_Q x_0}} (-\tilde{A}(\rho))_{x_0,y} a_y \leq a_{x_0} \sum_{\substack{y\neq x_0 \\ y\sim_Q x_0}} (-\tilde{A}(\rho))_{x_0,y} = \tilde{A}(\rho)_{x_0,x_0} a_{x_0},$$

where the last equality is due to the zero sum on x_0-row. Since this inequality holds with equality, we can infer that $a_y = a_{x_0} = a_{\max}$ for all $y \sim_Q x_0$. Then we may pick any $x_1 \in \{y : y \sim_Q x_0\}$ (so that $a_{x_1} = a_{\max}$) and argue similarly that $a_y = a_{x_1} = a_{\max}$ for all $y \sim_Q x_1$. Since Q is irreducible and hence the induced graph is connected, we may repeat this process until we can conclude that $a_v = a_{\max}$ for all $v \in X$, that is, the vector $a \in \ker \tilde{A}(\rho)$ must be a constant vector as claimed.

To see that $\mathcal{T}(\rho) = -\operatorname{div}(\hat{\rho}\,\cdot)$ is injective on $Tan_\rho M$, assume that $-\operatorname{div}(\hat{\rho}\,\nabla\psi) = 0 = \tilde{A}(\rho)\psi$. Then $\psi \in \ker \tilde{A}(\rho)$ must be constant, and hence $\nabla\psi = 0$. Lastly, we observe that dimensions of $Tan_\rho\mathcal{M}$ and of $T_\rho\mathcal{M}$ coincide:

$$\dim Tan_\rho\mathcal{M} = \dim \mathbb{R}^X - \dim\ker\nabla = |X| - 1 = \dim T_\rho\mathcal{M},$$

since $T_\rho\mathcal{M}$ is defined via a nontrivial linear condition. Therefore, $\mathcal{T}(\rho) : Tan_\rho\mathcal{M} \to T_\rho\mathcal{M}$ is indeed an isomorphism. □

So far we have identified $s \in T_\rho\mathcal{M}$ with $\nabla\psi \in Tan_\rho\mathcal{M}$ through the relation $s = -\operatorname{div}(\hat{\rho}\nabla\psi)$. As a consequence, for a smooth curve $\gamma : [a,b] \to \mathcal{M}$, the tangent vector field γ' along the curve γ, which is normally described by $\gamma'(t) = \gamma'_t = (\rho_t, s_t) \in T_{\rho_t}\mathcal{M}$, can be identified with $(\rho_t, \nabla\psi_t) \in Tan_{\rho_t}\mathcal{M}$ where $\frac{d}{dt}\rho_t = s_t = -\operatorname{div}(\hat{\rho}_t\nabla\psi_t)$. In other words, a smooth curve in $\mathcal{M}$ can be represented by $(\rho_t, \psi_t)_{t\in[a,b]}$ which satisfies the following ***discrete continuity equation***:

$$\frac{d}{dt}\rho_t + \operatorname{div}(\hat{\rho}_t\nabla\psi_t) = 0. \tag{21.1}$$

21.2 Induced metric from metric tensor

The smooth manifold $\mathcal{M} = \mathscr{P}_*(X) := \left\{\rho : X \to \mathbb{R}^+ \,\middle|\, \sum_{x\in X} \rho(x)\pi(x) = 1\right\}$ can be equipped by the metric tensor $g_\rho : T_\rho\mathcal{M} \times T_\rho\mathcal{M} \to \mathbb{R}$ given by

$$g_\rho(s_1, s_2) := \langle\nabla\psi_1, \nabla\psi_2\rangle_\rho$$

$$=: \langle A(\rho)\psi_1, \psi_2\rangle$$

where $s_i = -\operatorname{div}(\hat{\rho}\nabla\psi_i)$ for $i \in \{1,2\}$, and the linear operator A is defined uniquely via Riesz representation theorem for the symmetric bilinear form $(\psi_1, \psi_2) \mapsto \langle \nabla\psi_1, \nabla\psi_2\rangle_\rho$ as follows.

Definition 21.2 (operator $A(\rho)$)**.** For a fixed $\rho \in \mathscr{P}_*(X)$, a linear operator $A(\rho) : \mathbb{R}^X \to \mathbb{R}^X$ is uniquely defined by

$$\langle A(\rho)\psi_1, \psi_2\rangle := \langle \nabla\psi_1, \nabla\psi_2\rangle_\rho,$$

for all $\psi_1, \psi_2 \in \mathbb{R}^X$. Equivalently, $A(\rho) = -\pi \operatorname{div}(\hat{\rho}\nabla\cdot)$, which can be seen via the following relation:

$$\langle \nabla\psi_1, \nabla\psi_2\rangle_\rho = \langle \hat{\rho}\nabla\psi_1, \nabla\psi_2\rangle_\pi = \langle -\operatorname{div}(\hat{\rho}\nabla\psi_1), \psi_2\rangle_\pi = \langle -\pi \operatorname{div}(\hat{\rho}\nabla\psi_1), \psi_2\rangle.$$

With this metric tensor g, the length of a smooth curve $\gamma : [a,b] \to \mathcal{M}$ with the tangent vector field $\gamma'(t) = \gamma_t' = (\rho_t, s_t)$ is given by

$$\ell(\gamma) := \int_a^b g_{\rho_t}(s_t, s_t)^{\frac{1}{2}} dt = \int_a^b \langle \nabla\psi_t, \nabla\psi_t\rangle_\rho^{\frac{1}{2}} dt,$$

and its energy is given by

$$E(\gamma) := \int_a^b g_{\rho_t}(s_t, s_t) dt = \int_a^b \langle \nabla\psi_t, \nabla\psi_t\rangle_{\rho_t} dt.$$

In view of Proposition 2.5, the metric $\mathcal{W}$ on $\mathscr{P}_*(X)$ induced by g can be characterized by the infimum of length of curves or by the infimum of energy. More precisely, the distance between any given $\rho_0, \rho_1 \in \mathcal{M} = \mathscr{P}_*(X)$ can be written as

$$\mathcal{W}(\rho_0, \rho_1) = \inf_\gamma \ell(\gamma) = \inf_\gamma E(\gamma)^{1/2},$$

which the infimum is taken over all curves $\gamma : [0,1] \to \mathcal{M}$ whose endpoints are ρ_0 and ρ_1. This metric $\mathcal{W}$ was introduced and studied in [Maa11, EM12], and it can be formulated as follows.

Definition 21.3 (Discrete transport metric)**.** The Riemannian metric tensor g at $\rho \in \mathcal{M} = \mathscr{P}_*(X)$ is defined as

$$g_\rho(s_1, s_2) := \langle \nabla\psi_1, \nabla\psi_2\rangle_\rho, \tag{21.2}$$

where $s_i = -\operatorname{div}(\hat{\rho}\nabla\psi_i)$ for $i \in \{1,2\}$. The metric $\mathcal{W}$ induced by g is given, for all pairs $\rho_0, \rho_1 \in \mathscr{P}_*(X)$, as

$$\mathcal{W}(\rho_0, \rho_1) := \inf \left\{ \int_0^1 \|\nabla\psi_t\|_{\rho_t}^2 dt \;\middle|\; (\rho_t, \psi_t) \in CE(\rho_0, \rho_1) \right\}^{\frac{1}{2}},$$

where $CE(\rho_0, \rho_1)$ consists of all smooth curves $(\rho_t)_{t\in[0,1]}$ in $\mathscr{P}_*(X)$ together with $(\psi_t)_{t\in[0,1]}$ in $\mathbb{R}^X$ which satisfies the discrete continuity equation (21.1): $\frac{d}{dt}\rho_t + \operatorname{div}(\hat{\rho}_t \nabla \psi_t) = 0$ with $\rho_{t=0} = \rho_0$ and $\rho_{t=1} = \rho_1$.

In [Maa11, Theorem 3.31], Maas formulates the following geodesic equation, which refers to the concept of co-geodesic flows. Here we discuss an alternative idea to prove this formula by applying the first variation formula to a constant speed geodesic from ρ_0 to ρ_1 which minimizes the energy: $\mathcal{W}(\rho_0, \rho_1) = \inf_\gamma E(\gamma)$.

Theorem 21.4. *Given any $\rho_0, \rho_1 \in \mathscr{P}_*(X)$, a curve $(\rho_t, \psi_t)_{t\in[0,1]} \in CE(\rho_0, \rho_1)$ is a geodesic from ρ_0 to ρ_1 if it satisfies (in addition to the discrete continuity equation) the following* ***discrete geodesic equation****:*

$$\frac{d}{dt}\psi_t(x) + \frac{1}{2}\sum_{y\in X}(\psi_t(y) - \psi_t(x))^2 \partial_1\theta(\rho_t(x), \rho_t(y))Q(x, y) = 0 \quad \forall x \in X, \tag{21.3}$$

where $\partial_1\theta$ is the partial derivative of θ with respect to the first coordinate.

Partial proof. Here we will only prove that the function

$$\Phi_t(x) := \frac{d}{dt}\psi_t(x) + \frac{1}{2}\sum_{y\in X}(\psi_t(y) - \psi_t(x))^2 \partial_1\theta(\rho_t(x), \rho_t(y))Q(x, y)$$

is a constant function on $x \in X$.

Consider a proper variation of a curve $(\rho_t)_{t\in[0,1]}$, namely $\rho_t^s := \rho_t + s\mathfrak{s}_t \in \mathscr{P}_*(X)$ for $s \in (-\varepsilon, \varepsilon)$ (for some fixed $\varepsilon > 0$) and $t \in [0, 1]$ with $\mathfrak{s}_t \in T_{\rho_t}\mathscr{P}_*(X)$. Here, being a proper variation means that $\rho_0^s = \rho_0$ and $\rho_1^s = \rho_1$ for all s, that is, $\mathfrak{s}_0 = \mathfrak{s}_1 = 0$. Moreover, for each s, we represent the curve $(\rho_t^s)_{t\in[0,1]}$ by (ρ_t^s, ψ_t^s) which satisfies the continuity equation $\frac{d}{dt}\rho_t^s + \operatorname{div}(\hat{\rho_t^s}\nabla\psi_t^s) = 0$, that is,

$$\frac{d}{dt}\rho_t + s\frac{d}{dt}\mathfrak{s}_t = -\operatorname{div}(\hat{\rho_t^s}\nabla\psi_t^s). \tag{21.4}$$

At $s = 0$, the continuity equation is simply $\frac{d}{dt}\rho_t = -\operatorname{div}(\hat{\rho}_t\nabla\psi_t)$, where the convention $\psi_t := \psi_t^0$ is used here. Taking derivative of (21.4) at $s = 0$ yields

$$\frac{d}{dt}\mathfrak{s}_t = -\operatorname{div}(\hat{\rho}_t\nabla(\partial_s\psi_t^s)) - \operatorname{div}((\partial_s\hat{\rho_t^s})\nabla\psi_t). \tag{21.5}$$

(Here and henceforth in the proof, we write in short $\partial_s := \frac{\partial}{\partial s}|_{s=0}$.)

For each s, the energy of the curve $(\rho_t^s)_{t\in[0,1]}$ is given by

$$E(\rho_\cdot^s) = \int_0^1 \langle\nabla\psi_t^s, \nabla\psi_t^s\rangle_{\rho_t^s} dt = \int_0^1 \langle -\operatorname{div}(\hat{\rho_t^s}\nabla\psi_t^s), \psi_t^s\rangle_\pi dt$$

$$= \int_0^1 \langle \frac{d}{dt}\rho_t + s\frac{d}{dt}\mathfrak{s}_t, \psi_t^s \rangle_\pi dt.$$

Its derivative at $s = 0$ can be computed as

$$\begin{aligned} \frac{d}{ds}|_{s=0} E(\rho_\cdot^s) &= \int_0^1 \langle \frac{d}{dt}\mathfrak{s}_t, \psi_t \rangle_\pi + \langle \frac{d}{dt}\rho_t, \partial_s \psi_t^s \rangle_\pi dt \\ &= \int_0^1 \langle \frac{d}{dt}\mathfrak{s}_t, \psi_t \rangle_\pi + \langle -\operatorname{div}(\hat{\rho}_t \nabla \psi_t), \partial_s \psi_t^s \rangle_\pi dt. \end{aligned} \tag{21.6}$$

The inner product $\langle -\operatorname{div}(\hat{\rho}_t \nabla \psi_t), \partial_s \psi_t^s \rangle_\pi$ can be computed using (21.5) as follows:

$$\begin{aligned} \langle -\operatorname{div}(\hat{\rho}_t \nabla \psi_t), \partial_s \psi_t^s \rangle_\pi &= \langle \hat{\rho}_t \nabla \psi_t, \nabla(\partial_s \psi_t^s) \rangle_\pi = \langle \nabla \psi_t, \hat{\rho}_t \nabla(\partial_s \psi_t^s) \rangle_\pi \\ &= \langle \psi_t, -\operatorname{div}(\hat{\rho}_t \nabla(\partial_s \psi_t^s)) \rangle_\pi \\ &= \langle \psi_t, \frac{d}{dt}\mathfrak{s}_t + \operatorname{div}((\partial_s \hat{\rho_t^s}) \nabla \psi_t) \rangle_\pi \\ &= \langle \psi_t, \frac{d}{dt}\mathfrak{s}_t \rangle_\pi - \langle \nabla \psi_t, (\partial_s \hat{\rho_t^s}) \nabla \psi_t \rangle_\pi. \end{aligned} \tag{21.7}$$

Next, we plug (21.7) into (21.6) and use integral by parts to obtain

$$\begin{aligned} \frac{d}{ds}|_{s=0} E(\rho_\cdot^s) &= \int_0^1 2\langle \frac{d}{dt}\mathfrak{s}_t, \psi_t \rangle_\pi - \langle \nabla \psi_t, (\partial_s \hat{\rho_t^s}) \nabla \psi_t \rangle_\pi dt \\ &= 2\langle \mathfrak{s}_1, \psi_1 \rangle_\pi - 2\langle \mathfrak{s}_0, \psi_0 \rangle_\pi - \int_0^1 2\langle \mathfrak{s}_t, \frac{d}{dt}\psi_t \rangle_\pi + \langle \nabla \psi_t, (\partial_s \hat{\rho_t^s}) \nabla \psi_t \rangle_\pi dt \\ &= -\int_0^1 2\langle \mathfrak{s}_t, \frac{d}{dt}\psi_t \rangle_\pi + \langle \nabla \psi_t, (\partial_s \hat{\rho_t^s}) \nabla \psi_t \rangle_\pi dt, \end{aligned} \tag{21.8}$$

where $\langle \mathfrak{s}_0, \psi_0 \rangle_\pi = \langle \mathfrak{s}_1, \psi_1 \rangle_\pi = 0$ since $\mathfrak{s}_0 = \mathfrak{s}_1 = 0$.

We recall that $\partial_s \rho_t^s = \partial_s(\rho_t + s\mathfrak{s}_t) = \mathfrak{s}_t$, and hence

$$\partial_s \hat{\rho_t^s}(x, y) = \partial_s \theta(\rho_t^s(x), \rho_t^s(y)) = \partial_1 \theta(\rho_t(x), \rho_t(y))\mathfrak{s}_t(x) + \partial_2 \theta(\rho_t(x), \rho_t(y))\mathfrak{s}_t(y).$$

Then

$$\begin{aligned} &\langle \nabla \psi_t, (\partial_s \hat{\rho_t^s}) \nabla \psi_t \rangle_\pi \\ &= \frac{1}{2} \sum_{x,y \in X} (\psi_t(y) - \psi_t(x))^2 (\partial_1 \theta(\rho_t(x), \rho_t(y))\mathfrak{s}_t(x) + \partial_2 \theta(\rho_t(x), \rho_t(y))\mathfrak{s}_t(y)) Q(x, y)\pi(x) \\ &= \sum_{x,y \in X} (\psi_t(y) - \psi_t(x))^2 \partial_1 \theta(\rho_t(x), \rho_t(y))\mathfrak{s}_t(x) Q(x, y)\pi(x). \end{aligned} \tag{21.9}$$

Plugging (21.9) into (21.8) then gives

$$
\begin{aligned}
&\frac{d}{ds}|_{s=0}E(\rho^s_\cdot)\\
&= -\int_0^1 \sum_{x\in X} \mathfrak{s}_t(x)\pi(x)\left(\frac{d}{dt}\psi_t(x) + \sum_{y\in X}(\psi_t(y)-\psi_t(x))^2\partial_1\theta(\rho_t(x),\rho_t(y))Q(x,y)\right)dt\\
&= -\int_0^1 \langle \mathfrak{s}_t, \Phi_t\rangle_\pi dt.
\end{aligned}
$$

The original curve $(\rho_t)_{t\in[0,1]}$ being a geodesic means that it minimizes the energy among its proper variation $\rho_t^s = \rho_t + s\mathfrak{s}_t$, that is, $0 = \frac{d}{ds}|_{s=0}E(\rho^s_\cdot) = -\int_0^1 \langle \mathfrak{s}_t, \Phi_t\rangle_\pi dt$. Since $\mathfrak{s}_t$ is chosen arbitrarily, we can conclude that $\Phi_t(x) = c_t$, a constant depending on t. □

The metric $\mathcal{W}$ in Definition 21.3 can be extended to a metric on $\mathscr{P}(X)$, and particularly $\mathcal{W}(\rho_0,\rho_1) < \infty$ for all $\rho_0,\rho_1 \in \mathscr{P}(X)$; see [Maa11, Theorems 3.8 and 3.12]. Moreover, the metric space $(\mathscr{P}(X),\mathcal{W})$ is complete (see [Maa11, Theorems 3.22]). The mentioned metric and completeness results in the the original paper are stated with more technical details to deal with more general choices of θ. In case of logarithmic mean however, things are simplified as $C_\infty := \inf_0^1 \frac{1}{\sqrt{\theta(1-r,1+r)}}dr < \infty$ and $\mathscr{P}_\alpha(X) := \{\rho \mid \mathcal{W}(\rho,\alpha) < \alpha\} = \mathscr{P}(X)$.

21.3 Discrete JKO theorem

The ***entropy functional*** (relative to a reference measure π) is defined for a probability density $\rho \in \mathscr{P}(X)$ by

$$\mathrm{Ent}(\rho) := \sum_{x\in X} \rho(x)\log\rho(x)\pi(x) = \langle \rho, \log\rho\rangle_\pi. \tag{21.10}$$

Maas states and proves in [Maa11, Theorem 4.7 and Corollary 4.8] the discrete analogue of theorem by Jordan-Kinderlehrer-Otto (JKO theorem; see details in Sections 2.3 and 2.4) which asserts that heat flow is gradient flow of the entropy.

Theorem 21.5 (Discrete JKO, [Maa11]). *Let θ be the logarithmic mean defined by $\theta(a,b) := \int_0^1 a^{1-t}b^t dt = \frac{a-b}{\log a - \log b}$. An integral curve $t \mapsto \rho_t$ of gradient flow of the entropy functional* $\mathrm{Ent}(\cdot)$ *(in the Riemannian structure of $(\mathscr{P}(X),\mathcal{W})$) is the solution to the heat equation $\frac{\partial}{\partial t}\rho_t = \Delta\rho_t$.*

Proof. Let ρ_t be the integral curve of gradient flow of $\mathrm{Ent}(\cdot)$. For a fixed $t \in \mathbb{R}$ and $s \in T_{\rho_t}\mathcal{M}$, we first compute $D\,\mathrm{Ent}(\rho_t)(s)$ through a curve $c : (-\varepsilon, \varepsilon) \to \mathscr{P}(X)$ given by $c(\tau) = \rho_t + \tau s$ (so that $c(0) = \rho_t$ and $c'(0) = s$). We have

$$\begin{aligned} D\,\mathrm{Ent}(\rho_t)(s) = \frac{d}{d\tau}|_{\tau=0}\,\mathrm{Ent}(c(\tau)) &= \frac{d}{d\tau}|_{\tau=0}\langle c(\tau), \log c(\tau)\rangle_\pi \\ &= [\langle 1 + \log c(\tau), c'(\tau)\rangle_\pi]_{\tau=0} = \langle 1 + \log \rho_t, s\rangle_\pi. \end{aligned}$$

With the identification $s = -\,\mathrm{div}(\hat{\rho}_t \nabla\psi)$, we have

$$\begin{aligned} g_{\rho_t}(\frac{d}{dt}\rho_t, s) &= -D\,\mathrm{Ent}_{\mathrm{vol}}(\rho_t)(s) = -\langle 1 + \log \rho_t, s\rangle_\pi \\ &= \langle 1 + \log \rho_t, \mathrm{div}(\hat{\rho}_t \nabla\psi)\rangle_\pi \\ &= -\langle \nabla \log \rho_t, \hat{\rho}_t \nabla\psi\rangle_\pi \\ &= -\langle \nabla \log \rho_t, \nabla\psi\rangle_{\rho_t} = -g_{\rho_t}(\mathfrak{s}, s), \end{aligned}$$

where $\mathfrak{s} = \mathrm{div}(\hat{\rho}_t \nabla \log \rho_t)$ via the defined metric tensor (21.2).

Since the above equality holds for all s, we can conclude that

$$\frac{d}{dt}\rho_t = -\mathfrak{s} = \mathrm{div}(\hat{\rho}_t \nabla \log \rho_t) = \mathrm{div}(\nabla \rho_t) = \Delta \rho_t.$$

Here $\hat{\rho} = \frac{\nabla \rho}{\nabla \log \rho}$ from (20.3) due to the choice of θ being the logarithmic mean.

□

Chapter 22

Erbar-Maas entropic Ricci curvature

In [EM12], Erbar and Maas introduce for finite Markov chains the following definition of lower Ricci curvature bound in the sense of Lott-Sturm-Villani curvature notion (cf. Definition 3.4).

Definition 22.1 (Entropic Ricci curvature, [EM12])**.** An irreducible and reversible finite Markov chain (X, Q, π) is said to have the ***entropic Ricci curvature bound below by*** $K \in \mathbb{R}$, or written as $\operatorname{Ric}(Q) \geq K$, if the relative entropy $\operatorname{Ent}(\cdot)$ is ***displacement*** K***-convex***, that is, every constant speed geodesic $(\rho_t)_{t\in[0,1]}$ in $(\mathscr{P}(X), \mathcal{W})$ satisfies the following K-convexity inequality:

$$\operatorname{Ent}(\rho_t) \leq (1-t)\operatorname{Ent}(\rho_0) + t\operatorname{Ent}(\rho_1) - K\frac{t(1-t)}{2}\mathcal{W}(\rho_0, \rho_1)^2 \tag{22.1}$$

for all $t \in [0, 1]$.

It is remarked in [EM12] that it suffices to verify (22.1) for only those constant speed geodesics $(\rho_t)_{t\in[0,1]}$ in $\mathscr{P}_*(X)$. Since $\operatorname{Ent}(\cdot)$ is smooth on $\mathscr{P}_*(X)$, this is equivalent to the ***infinitesimal*** K***-convexity***: $\frac{d^2}{dt^2}\operatorname{Ent}(\rho_t) \geq K\mathcal{W}(\rho_0, \rho_1)^2 = K\|\nabla\psi_t\|_{\rho_t}^2$, where the rightmost equality is due to the fact that ρ_t has constant speed. Hence we may reformulate the above definition of the entropic Ricci curvature as follows.

Definition 22.2. Let (X, Q, π) be an irreducible and reversible finite Markov chain. A point $\rho \in \mathscr{P}_*(X)$ is said to satisfy $\operatorname{Ric}_Q(\rho) \geq K$ if the inequality

$$\frac{d^2}{dt^2}\Big|_{t=0}\operatorname{Ent}(\rho_t) \geq K\|\nabla\psi\|_\rho^2 \tag{22.2}$$

holds for all $\psi \in \mathbb{R}^X$ and all $(\rho_t, \psi_t)_{t\in(-\varepsilon,\varepsilon)}$ satisfying (21.1) and (21.3) with $(\rho_0, \psi_0) = (\rho, \psi)$.

Note that the entropic Ricci curvature given in Definition 22.2 is of local-type (i.e., it is defined at every point in $\mathscr{P}_*(X)$), in contrast to the curvature defined globally as in Definition 22.1. In fact, a Markov chain (X, Q, π) satisfies $\mathrm{Ric}(Q) \geq K$ if and only if $\mathrm{Ric}_Q(\rho) \geq K$ holds for all $\rho \in \mathscr{P}_*(X)$.

22.1 Equivalent reformulations of the entropic Ricci curvature via Bochner's formula and heat semigroup

This section features an important result from Erbar and Maas [EM12] about the characterization of the entropic Ricci curvature (originally defined via displacement convexity of the Entropy) in terms of the curvature-dimension inequality derived from Bochner's formula (see Theorem 22.5 below). This inequality resembles the condition $BE(K, \infty)$ for Bakry-Émery curvature, and consequently it has another equivalent interpretation in terms of gradient estimate.

Let us first mention a key component introduced in [EM12], namely $\mathcal{B}(\rho, \psi)$.

Definition 22.3. Let $\hat{\rho}$ and $\hat{\Delta}\rho$ be defined as

$$\begin{aligned} \hat{\rho}(x, y) &:= \theta(\rho(x), \rho(y)), \\ \hat{\Delta}\rho &:= \partial_1\theta(\rho(x), \rho(y))\Delta\rho(x) + \partial_2\theta(\rho(x), \rho(y))\Delta\rho(y), \end{aligned}$$

where $\partial_1\theta$ and $\partial_2\theta$ is the partial derivatives with respect to the first and second coordinate of θ, respectively. For $\rho \in \mathscr{P}_*(X)$ and $\psi \in \mathbb{R}^X$, the term $\mathcal{B}(\rho, \psi) \in \mathbb{R}$ is defined to be

$$\mathcal{B}(\rho, \psi) := \frac{1}{2}\langle \hat{\Delta}\rho \cdot \nabla\psi, \nabla\psi\rangle_\pi - \langle \hat{\rho} \cdot \nabla\psi, \nabla\Delta\psi\rangle_\pi. \tag{22.3}$$

If compared to the Bochner's formula (5.2), $\mathcal{B}(\rho, \psi)$ vaguely represents the terms $\frac{1}{2}\Delta|\operatorname{grad} f|^2 - \langle \operatorname{grad}\Delta f, \operatorname{grad} f\rangle$. In fact, it can be regarded as the following Hessian term (see also [EM12, Proposition 4.3])

Proposition 22.4 ([EM12]). *For all $\rho \in \mathscr{P}_*(X)$ and $\psi \in \mathbb{R}^X$, the following identity*

$$\mathcal{B}(\rho, \psi) = \mathrm{Hess}_{\mathcal{W}}\,\mathrm{Ent}(s, s). \tag{22.4}$$

holds where $s := -\operatorname{div}(\hat{\rho}\nabla\psi) \in T_\rho\mathscr{P}_*(X)$.

Proof. Given any $\rho \in \mathscr{P}_*(X)$ and $\psi \in \mathbb{R}^X$, let $s = -\operatorname{div}(\hat{\rho}\nabla\psi) \in T_\rho\mathscr{P}_*(X)$. Let $(\rho_t)_{t\in(-\varepsilon,\varepsilon)}$ be a geodesic in $(\mathscr{P}_*(X), \mathcal{W})$ such that $\rho_0 = \rho$ and $\dot{\rho}_0 = s$, and recall the fact that the second derivative in time of a functional along a geodesic can be expressed in terms of Hessian, that is,

$$\begin{aligned}\frac{d^2}{dt^2}\operatorname{Ent}(\rho_t) &= \frac{d}{dt} g_{\rho_t}(\operatorname{grad}_{\mathcal{W}}\operatorname{Ent}(\rho_t), \dot\rho_t)\\ &= g_{\rho_t}(\nabla_{\dot\rho_t}(\operatorname{grad}_{\mathcal{W}}\operatorname{Ent}), \dot\rho_t) + g_{\rho_t}(\operatorname{grad}_{\mathcal{W}}\operatorname{Ent}(\rho_t), \underbrace{\frac{D}{dt}\dot\rho_t}_{=0\ \because\ \text{geodesic}})\\ &= \operatorname{Hess}_{\mathcal{W}}\operatorname{Ent}(\dot\rho_t, \dot\rho_t).\end{aligned}$$

Hence it suffices to prove that $\mathcal{B}(\rho_t, \psi_t) = \frac{d^2}{dt^2}\operatorname{Ent}(\rho_t)$. First, take derivative in time of $\operatorname{Ent}(\rho_t) = \langle \rho_t, \log \rho_t\rangle_\pi$ and use the continuity equation (21.1) together with the fact that $\hat\rho_t = \frac{\nabla\rho_t}{\nabla \log \rho_t}$ from (20.3):

$$\begin{aligned}\frac{d}{dt}\operatorname{Ent}(\rho_t) &= \langle 1 + \log \rho_t, \frac{d}{dt}\rho_t\rangle_\pi\\ &= -\langle 1 + \log\rho_t, \operatorname{div}(\hat\rho_t \nabla\psi_t)\rangle_\pi\\ &= \langle \nabla \log \rho_t, \hat\rho_t\nabla\psi_t\rangle_\pi = \langle \nabla\rho_t, \nabla\psi_t\rangle_\pi.\end{aligned}$$

The second derivative of $\operatorname{Ent}(\rho_t)$ is given by

$$\begin{aligned}\frac{d^2}{dt^2}\operatorname{Ent}(\rho_t) &= \langle \nabla(\frac{d}{dt}\rho_t), \nabla\psi_t\rangle_\pi + \langle \nabla\rho_t, \nabla(\frac{d}{dt}\psi_t)\rangle_\pi\\ &= -\langle \frac{d}{dt}\rho_t, \Delta\psi_t\rangle_\pi - \langle \Delta\rho_t, \frac{d}{dt}\psi_t\rangle_\pi,\end{aligned}$$

where the continuity equation (21.1) gives

$$\langle \frac{d}{dt}\rho_t, \Delta\psi_t\rangle_\pi = -\langle \operatorname{div}(\hat\rho_t\nabla\psi_t), \Delta\psi_t\rangle_\pi = \langle \hat\rho_t\nabla\psi_t, \nabla\Delta\psi_t\rangle_\pi,$$

and the geodesic equation (21.3) gives

$$\begin{aligned}\langle \Delta\rho_t, \frac{d}{dt}\psi_t\rangle_\pi &= -\frac{1}{2}\sum_{x,y\in X}\Delta\rho_t(x)(\psi_t(y)-\psi_t(x))^2\partial_1\theta(\rho_t(x),\rho_t(y))Q(x,y)\pi(x)\\ &= -\frac{1}{4}\sum_{x,y\in X}(\psi_t(y)-\psi_t(x))^2\Big(\partial_1\theta(\rho_t(x),\rho_t(y))\Delta\rho_t(x) +\\ &\qquad \partial_2\theta(\rho_t(x),\rho_t(y))\Delta\rho_t(y)\Big)Q(x,y)\pi(x)\\ &= -\frac{1}{2}\langle \hat\Delta\rho\cdot\nabla\psi_t, \nabla\psi_t\rangle_\pi.\end{aligned}$$

Hence $\frac{d^2}{dt^2}\operatorname{Ent}(\rho_t) = -\langle\hat\rho_t\nabla\psi_t, \nabla\Delta\psi_t\rangle_\pi + \frac{1}{2}\langle\hat\Delta\rho\cdot\nabla\psi_t,\nabla\psi_t\rangle_\pi = \mathcal{B}(\rho_t,\psi_t)$ as desired. □

Now we are ready to state and prove the equivalent reformulations of the entropic Ricci curvature. This result is due to [EM12] and Erbar and Fathi [EF18].

Theorem 22.5. *Given an irreducible and reversible finite Markov chain (X, Q, π), the following statements are equivalent.*

1. $\mathrm{Ric}(Q) \geq K$, *i.e.,* (22.1) *holds.*
2. *All* $\rho \in \mathscr{P}_*(X)$ *satisfies* $\mathrm{Ric}_Q(\rho) \geq K$, *i.e.,* (22.2) *holds.*
3. *The following Bochner's inequality holds for all* $\rho \in \mathscr{P}_*(X)$ *and* $\psi \in \mathbb{R}^X$*:*
$$\mathcal{B}(\rho, \psi) \geq K \|\nabla \psi\|_\rho^2, \tag{22.5}$$
where $\mathcal{B}(\rho, \psi)$ *is defined in* (22.3) *with* θ *being the logarithmic mean.*
4. *Global gradient estimate holds for for all* $\rho \in \mathscr{P}(X)$ *and* $\psi \in \mathbb{R}^X$ *and all* $t \geq 0$,
$$\|\nabla P_t \psi\|_\rho^2 \leq e^{-2\kappa t} \|\nabla \psi\|_{P_t \rho}^2,$$
or more explicitly
$$\begin{aligned} &\frac{1}{2} \sum_{u,v} (P_t\psi(v) - P_t\psi(u))^2 \hat{\rho}(u,v) Q(u,v) \pi(u) \\ \leq\, & e^{-2\kappa t} \frac{1}{2} \sum_{u,v} (\psi(v) - \psi(u))^2 \widehat{P_t\rho}(u,v) Q(u,v) \pi(u) \end{aligned} \tag{22.6}$$
where P_t *denotes the heat semigroup operator.*

Proof. The equivalence (1) ⇔ (2) is discussed in the beginning of this chapter. The equivalence (2) ⇔ (3) is a consequence of Proposition 22.4 applied to (22.2). The equivalence (3) ⇔ (4) is due to [EF18, Theorem 3.1], following a similar argument in Theorem 5.22 by differentiating $F(s) := e^{-2Ks} \|\nabla P_{t-s} \psi\|_{P_s \rho}^2$. □

The above equivalence (2) ⇔ (3) requires the fact that θ is the logarithmic mean. We can generalize this curvature notion by defining it directly from Bochner's inequality. More precisely, we define the entropic Ricci curvature at a fixed $\rho \in \mathscr{P}(X)$ as

$$\mathrm{Ric}_{Q,\theta}(\rho) := \sup\{\, K \in \mathbb{R} \mid (22.5) \text{ holds for all } \psi \in \mathbb{R}^X \,\},$$

with a given choice of the mean θ. Another interesting choice of θ is the arithmetic mean (AM), where the resulting curvature notion is closely related to the Bakry-Émery curvature. This observation is mentioned in [Maa17], and we clarify it in the following proposition.

Proposition 22.6. *Let* (X, Q, π) *be a finite irreducible and reversible Markov chain. For any* $x \in X$*, consider the probability density* $\delta_x := \frac{1}{\pi(x)}\mathbf{1}_x \in \mathscr{P}(X)$*. Then* $\mathrm{Ric}_{Q,\mathrm{AM}}(\delta_x) \geq K$ *if and only if* x *satisfies* $BE(K, \infty)$*.*

Proof. The choice of θ being the arithmetic mean implies that $\Delta\hat{\rho}(u,v) = \frac{1}{2}\rho(u) + \frac{1}{2}\rho(v)$. With the choice of $\rho = \delta_x$, we have nontrivial values of $\Delta\hat{\rho}$ and $\hat{\Delta}\rho$ as follows:

$$\begin{aligned}
\hat{\rho}(x,y) = \hat{\rho}(y,x) &= \frac{1}{2\pi(x)}, \\
\Delta\hat{\rho}(x,y) = \Delta\hat{\rho}(y,x) &= -\frac{1}{2\pi(x)} + \frac{Q(y,x)}{2\pi(x)}, \\
\Delta\hat{\rho}(y,z) = \Delta\hat{\rho}(z,y) &= \frac{Q(y,x)}{2\pi(x)} = \frac{Q(x,y)}{2\pi(y)}.
\end{aligned}$$

for all $y \in S_1(x)$ and $z \in S_2(x)$. We must show that the inequality $\mathcal{B}(\rho,\psi) \geq K\|\nabla\psi\|_\rho^2$ is simplified into $\Gamma_2\psi(x) \geq K\Gamma\psi(x)$. First, the term $\langle\hat{\Delta}\rho \cdot \nabla\psi, \nabla\psi\rangle_\pi$ is given by

$$\begin{aligned}
\langle\hat{\Delta}\rho \cdot \nabla\psi, \nabla\psi\rangle_\pi &= \frac{1}{2}\sum_{u,v}(\psi(v)-\psi(u))^2 Q(u,v)\pi(u)\hat{\Delta}\rho(u,v) \\
&= \sum_{y\in S_1(x)} (\psi(y)-\psi(x))^2 Q(x,y)(-\frac{1}{2} + \frac{Q(y,x)}{2}) + \\
&\quad \sum_{y\in S_1(x)} \sum_{z\in S_1(y)\setminus\{x\}} (\psi(z)-\psi(y))^2 Q(y,z)\frac{Q(x,y)}{2} \\
&= \sum_{y\in S_1(x)} (\Gamma\psi(y)-\Gamma\psi(x))Q(x,y) = \Delta(\Gamma\psi)(x). \qquad (22.7)
\end{aligned}$$

The second term $\langle\hat{\rho}\cdot\nabla\psi, \nabla\Delta\psi\rangle_\pi$ is given by

$$\begin{aligned}
\langle\hat{\rho}\cdot\nabla\psi, \nabla\Delta\psi\rangle_\pi &= \frac{1}{2}\sum_{u,v}(\psi(v)-\psi(u))(\Delta\psi(v)-\Delta\psi(u))Q(u,v)\pi(u)\hat{\rho}(u,v) \\
&= \sum_{y\in S_1(x)} (\psi(y)-\psi(x))(\Delta\psi(v)-\Delta\psi(u))\frac{Q(x,y)}{2} \\
&= \Gamma(\psi, \Delta\psi)(x), \qquad (22.8)
\end{aligned}$$

and similarly, $\|\nabla\psi\|_\rho^2 = \langle\hat{\rho}\cdot\nabla\psi, \nabla\psi\rangle_\pi = \Gamma\psi(x)$. Combining (22.7) and (22.8) yields $\mathcal{B}(\rho,\psi) = \frac{1}{2}\Delta(\Gamma\psi)(x) - \Gamma(\psi,\Delta\psi)(x) = \Gamma_2\psi(x)$ as desired. □

22.2 Products of Markov chains and hypercube example

One of the main results in [EM12] which follows from the characterization of entropic curvature by the Bochner's inequalty $\mathcal{B}(\rho,\psi) \geq K\|\nabla\psi\|_\rho^2$ is the curvature estimate for the Cartesian product of two Markov chains.

The following definition of the Cartesian products of Markov chains is almost perfectly aligned with The definition of Cartesian products is previously discussed in the setting of weighted graphs (see Definition 16.5). The products of Markov chains are very similar (with additional exception that they are allowed to be lazy, i.e. $Q(x,x) > 0$). Products here are defined for $n \geq 2$ components.

Definition 22.7. Given n irreducible and reversible finite Markov chains (X_i, Q_i, π_i), $1 \leq i \leq n$ and a fixed $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n) \in (\mathbb{R}_{\geq 0})^n$ with $\sum_{i=1}^n \alpha_i = 1$. Their Cartesian product (X, Q_α, π) is defined on the product space $X = \prod_i X_i$ with the Markov kernel given for $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ in X as

$$Q_\alpha(\mathbf{x},\mathbf{y}) := \begin{cases} \sum_{i=1}^n \alpha_i Q_i(x_i,x_i) & \text{if } \mathbf{x} = \mathbf{y}, \\ \alpha_i Q_i(x_i,y_i) & \text{if } x_i \neq y_i \text{ and } x_j = y_j \forall j \neq i, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, $\pi(\mathbf{x}) := \prod_i \pi_i(x_i)$ is indeed the unique stationary distribution for this kernel.

Theorem 22.8. *[EM12, Theorem 6.2] Assume that* $\mathrm{Ric}(Q_i) \geq K_i$ *for* $1 \leq i \leq n$. *Then*

$$\mathrm{Ric}(Q_\alpha) \geq \min_i \{\alpha_i K_i\}.$$

The proof in the original paper applies the concept of mapping representations for the terms $\mathcal{B}(\rho,\psi)$ and $\|\nabla\psi\|_\rho^2$, which is not covered here in this survey. However, we would like to mention the resemblance between this result and the Cartesian product result for the Bakry-Émery curvature given in Corollary 16.7.

Now we revisit the classical example of the hypercubes and compute the lower curvature bound for (a simple random walk on) the hypercube.

Example 22.9 (hypercubes)**.** We start with the Markov chain of the two-point space $K_2 = \{x,y\}$ with $Q(x,y) = Q(y,x) = 1$ (so it represent the (nonlazy) simple random walk of K_2) and $\pi(x) = \pi(y) = \frac{1}{2}$. Note that $\Delta f(x) = f(y) - f(x) = -\Delta f(y)$ for all $f \in \mathbb{R}^{\{x,y\}}$. To derive $\mathcal{B}(\rho,\psi)$ and $\|\nabla\psi\|_\rho^2$, we first compute $\|\nabla\psi\|_\rho^2 = \frac{1}{2}(\psi(y) - \psi(x))^2 \hat{\rho}(x,y)$, and

$$\langle \hat{\Delta}\rho \cdot \nabla\psi, \nabla\psi \rangle_\pi = \frac{1}{2}(\psi(y) - \psi(x))^2 \hat{\Delta}\rho(x,y)$$

$$\begin{aligned}\langle \hat{\rho} \cdot \nabla\psi, \nabla\Delta\psi\rangle_\pi &= \frac{1}{2}(\psi(y) - \psi(x))(\Delta\psi(y) - \Delta\psi(x))\hat{\rho}(x,y) \\ &= -(\psi(y) - \psi(x))^2\hat{\rho}(x,y).\end{aligned}$$

Thus $\mathcal{B}(\rho,\psi) = \big(\frac{1}{2}\frac{\hat{\Delta}\rho(x,y)}{\hat{\rho}(x,y)} + 2\big)\|\nabla\psi\|_\rho^2$. And hence the optimal bound for the curvature is

$$K = 2 + \inf_\rho \frac{1}{2}\frac{\hat{\Delta}\rho(x,y)}{\hat{\rho}(x,y)}.$$

For shortened notations, we write $\rho(x) = a$, $\rho(y) = b$ and $\log b - \log a = c$. Thus

$$\begin{aligned}\hat{\Delta}\rho(x,y) &= \partial_1\theta(\rho(x),\rho(y))(\rho(y) - \rho(x)) + \partial_2\theta(\rho(x),\rho(y))(\rho(x) - \rho(y)) \\ &= \frac{1}{c^2}\big(-c + \frac{b-a}{a}\big)(b-a) + \frac{1}{c^2}\big(c - \frac{b-a}{b}\big)(a-b) \\ &= \frac{b-a}{c}\big(-2 + \frac{(b-a)}{c}(\frac{1}{a} + \frac{1}{b})\big) \geq 0,\end{aligned}$$

where the last inequality is due to the logarithmic mean being greater or equal to the harmonic mean: $\frac{b-a}{\log b - \log a} \geq \frac{2}{\frac{1}{a}+\frac{1}{b}}$. Moreover, choosing $\rho(x) = \rho(y)$ gives $\hat{\Delta}\rho(x,y) = 0$. Therefore, the entropic Ricci curvature for the simple random walk on K_2 is given by the sharp estimate $\mathrm{Ric}(K_2, srw) \geq 2$.

The simple random walk on the n-dimensional hypercube $\mathcal{Q}^n$ is the Cartesian product of n copies of the simple random walk on K_2 with weights $\alpha_i \equiv \frac{1}{n}$. It then follows immediately that $\mathrm{Ric}(\mathcal{Q}^n, srw) \geq \frac{2}{n}$. In fact, it is further discussed in [EM12, Corollary 7.10] that $\frac{2}{n}$ is the sharp curvature bound.

Chapter 23

Diameter bound for Markov chain with positive curvature

In this chapter, we present a generalization of the main result in [Kam20b] that for a Markov chain (X, Q, π) with positive entropic Ricci curvature $\mathrm{Ric}(Q) \geq K > 0$, the diameter of the underlying graph is bounded from above by

$$\mathrm{diam}(X) \leq \frac{2}{K}\sqrt{\frac{D\log D}{D-1}},$$

where D is the maximal vertex π-degree. The proof is divided into two steps. The first step is to follow the ideas by Erbar and Fathi [EF18] to derive a similar local gradient estimate as follows.

Theorem 23.1 (Local gradient estimate). *If* $\mathrm{Ric}(Q) \geq K$*, then for all* $x \in X$ *and* $\psi \in \mathbb{R}^X$ *and all* $\varepsilon \geq 0$ *we have*

$$\Gamma(P_t\psi)(x)\pi(x) \leq \frac{e^{-2Kt}}{2\theta(1,\varepsilon)}[P_t(\Gamma\psi)(x)\pi(x) + \varepsilon \sum_{y\in N(x)} P_t(\Gamma\psi)(y)\pi(y)]. \tag{23.1}$$

where $2\Gamma f(x) := \sum_y (f(y) - f(x))^2 Q(x,y)$.

Proof. The proof follows ideas from [EF18, Corollary 3.4]. For convenience, we recall the global gradient estimate (22.6) here:

$$\begin{aligned}&\frac{1}{2}\sum_{u,v}(P_t\psi(v) - P_t\psi(u))^2\hat{\rho}(u,v)Q(u,v)\pi(u)\\ &\leq e^{-2Kt}\frac{1}{2}\sum_{u,v}(\psi(v) - \psi(u))^2\widehat{P_t\rho}(u,v)Q(u,v)\pi(u).\end{aligned}$$

We localize (22.6) by choosing $\rho = \mathbf{1}_x + \varepsilon \sum_{y \in S_1(x)} \mathbf{1}_y$ for a fixed $x \in V$ and a parameter $\varepsilon \in [0, \infty)$. Note that we do not require ρ to be a probability density (i.e., $\sum_x \rho(x)\pi(x) = 1$) since (22.6) is homogeneous in ρ.

For our particular choice of ρ, we have $\hat{\rho}(x, y) = \hat{\rho}(y, x) = \theta(1, \varepsilon)$ for all $y \in S_1(x)$. The left-hand-side of (22.6) is bounded from below by

$$L.H.S. \geq \sum_{y \in S_1(x)} (P_t\psi(y) - P_t\psi(x))^2 \theta(1, \varepsilon) Q(x, y)\pi(x) = 2\theta(1, \varepsilon)\Gamma(P_t\psi)(x)\pi(x). \tag{23.2}$$

On the other hand, the right-hand-side of (22.6) is bounded from above by

$$\begin{aligned} R.H.S. \quad &\leq \quad e^{-2Kt} \frac{1}{2} \sum_{u,v} (\psi(v) - \psi(u))^2 \frac{P_t\rho(u) + P_t\rho(v)}{2} Q(u, v)\pi(u) \\ &= \quad e^{-2Kt} \frac{1}{2} \sum_{u,v} (\psi(v) - \psi(u))^2 P_t\rho(u) Q(u, v)\pi(u) \end{aligned} \tag{23.3}$$

due to $\theta(s, t) \leq (s + t)/2$ and the symmetry from interchanging u and v.

We now apply the heat kernel $p_t(\cdot, \cdot)$ given by $P_t g(u) = \sum_z p_t(u, z) g(z)\pi(z)$ for every function g. With our choice of ρ, we obtain

$$P_t\rho(u) = p_t(u, x)\rho(x)\pi(x) + \varepsilon \sum_{y \in S_1(x)} p_t(u, y)\rho(y)\pi(y),$$

which we substitute into (23.3) and use the symmetry of heat kernel: $p_t(u, v) = p_t(v, u)$ to derive

$$\begin{aligned} R.H.S. \quad &\leq \frac{e^{-2Kt}}{2} \Big[\quad \pi(x) \sum_u p_t(u, x)\pi(u) \sum_v (\psi(v) - \psi(u))^2 Q(u, v) + \\ &\qquad \varepsilon \sum_{y \in S_1(x)} \pi(y) \sum_u p_t(u, y)\pi(u) (\sum_v (\psi(v) - \psi(u))^2 Q(u, v)) \Big] \\ &= e^{-2Kt} \Big[\quad \pi(x) P_t(\Gamma\psi)(x) + \varepsilon \sum_{y \in S_1(x)} \pi(y) P_t(\Gamma\psi)(y) \Big]. \end{aligned} \tag{23.4}$$

The desired inequality then follows from combining (23.2) and (23.4). □

We have the following corollary as an immediate consequence of the above theorem.

Corollary 23.2. *If* $\mathrm{Ric}(Q) \geq K$, *then for all* $x \in X$ *and* $\psi \in \mathbb{R}^X$, *we have*

$$\Gamma(P_t\psi)(x) \leq c \cdot e^{-2Kt} \|P_t(\Gamma\psi)\|_\infty \tag{23.5}$$

where $c := \frac{D \log D}{D-1}$ *and* D *is the maximal vertex* π*-degree,*

$$D := \max_{x \in X} \operatorname{Deg}_\pi(x) = \max_{x \in X} \frac{1}{\pi(x)} \sum_{y \in S_1(x)} \pi(y).$$

.

Proof. It is implied by Theorem 23.1 that $\Gamma(P_t\psi)(x) \le c_{\varepsilon,x} \cdot e^{-2Kt} \|P_t(\Gamma\psi)\|_\infty$, where the constant term $c_{\varepsilon,x}$ (depending on ε and x) is given by

$$c_{\varepsilon,x} := \frac{1}{2\theta(1,\varepsilon)} \left(1 + \varepsilon \sum_{y \in S_1(x)} \frac{\pi(y)}{\pi(x)}\right) = \frac{1 + \varepsilon \operatorname{Deg}_\pi(x)}{2\theta(1,\varepsilon)} \le \frac{1 + \varepsilon D}{2\theta(1,\varepsilon)}.$$

In particular when choosing $\varepsilon = \frac{1}{D}$, we have $c_{\varepsilon,x} \le \frac{D \log D}{D-1}$. □

The second step is to prove a Bonnet-Myers-type diameter bound on the underlying graph of (X, Q, π) using the same technique from [LMP18] (see Theorem 18.1 for comparison).

Theorem 23.3 (Diameter bound)**.** *Let* (X, Q, π) *be an irreducible and reversible finite Markov chain with strictly positive entropic Ricci curvature* $\operatorname{Ric}(Q) \ge K > 0$. *Then the diameter of the induced weighted graph* X *is bounded from above by*

$$\operatorname{diam}(X) \le \frac{2}{K} \sqrt{\frac{D \log D}{D-1}}, \tag{23.6}$$

where D *is the maximal vertex* π*-degree,*

$$D := \max_{x \in X} \operatorname{Deg}_\pi(x) = \max_{x \in X} \frac{1}{\pi(x)} \sum_{y \in S_1(x)} \pi(y).$$

Remark 23.4. We previously discuss that the entropic Ricci curvature of the hypercube $\mathcal{Q}^n$ is sharply estimated by $\frac{2}{n}$. Therefore, in view of the hypercube $\mathcal{Q}^n$, the bound (23.6) is not optimal:

$$n = \operatorname{diam}(\mathcal{Q}^n) \le n \sqrt{\frac{n}{n-1} \log n}.$$

Proof of Theorem 23.3. Consider a function $f \in \mathbb{R}^X$ given by $f(x) := d(x, x_0)$ for an arbitrary reference point $x_0 \in X$. Since f is a 1-Lipschitz function, it follows that $2\Gamma f(x) = \sum_{y \in S_1(x)} Q(x,y)(f(y) - f(x))^2 \le 1$ for all $x \in X$, i.e., $\|2\Gamma f\|_\infty \le 1$, which then implies $\|2P_t(\Gamma f)\|_\infty \le \|2\Gamma f\|_\infty \le 1$.

Moreover, the Cauchy-Schwartz inequality and the inequality (23.5) yield

$$\begin{aligned}|\Delta P_t f(x)|^2 &= \left(\sum_{y\in S_1(x)} Q(x,y)(P_t f(y) - P_t f(x))\right)^2 \\ &\le \left(\sum_{y\in S_1(x)} Q(x,y)(P_t f(y) - P_t f(x))^2\right)\left(\sum_{y\in S_1(x)} Q(x,y)\right) \\ &= 2\Gamma(P_t f)(x) \le e^{-2Kt} c\|2P_t(\Gamma f)\|_\infty \le ce^{-2Kt}. \end{aligned} \tag{23.7}$$

From the fundamental theorem of calculus and the definition of P_t, we then obtain

$$|f(x) - P_T f(x)| \le \int_0^T \left|\frac{\partial}{\partial t} P_t f(x)\right| dt = \int_0^T |\Delta P_t f(x)|\, dt \le \int_0^T \sqrt{c}e^{-Kt} dt \le \frac{\sqrt{c}}{K},$$

which holds true for all $T > 0$.

Moreover, (23.7) implies that $|P_t f(y) - P_t f(x)| \to 0$ as $t \to \infty$ for all edges $y \sim x$, and the same result is extended for all pairs of x, y which are not necessarily neighbors.

Passing to the limit $T \to \infty$, we conclude from triangle inequality that

$$\begin{aligned} d(x, x_0) &= |f(x) - f(x_0)| \\ &\le |f(x) - P_T f(x)| + |f(x_0) - P_T f(x_0)| + |P_T f(x) - P_T f(x_0)| \\ &\le \frac{2\sqrt{c}}{K}, \end{aligned}$$

which implies the desired diameter bound. □

Chapter 24

Outlook for Erbar-Maas entropic Ricci curvature

Let us finish this Part IV with further questions about the entropic Ricci curvature on discrete Markov chains.

1. Curvature in general is a local concept. Ollivier Ricci curvature is defined on edges of graphs, while Bakry-Émery is defined on vertices. The entropic Ricci curvature, on the other hand, is defined to be the global lower bound of Ricci curvature for whole spaces. Previously, we mentioned about the possibility to define the entropic Ricci curvature "pointwise on $\mathscr{P}(X)$", namely $\mathrm{Ric}_Q(\rho)$ for each $\rho \in \mathscr{P}(X)$. However, it is unlikely that this local property can be pushed further to "pointwise or locally on X" (since $\mathrm{Ric}_Q(\mu) = +\infty$ for any μ which does not have full support in the case of the logarithmic mean). However, it would be interesting to investigate whether there are measures $\rho_0 \in \mathscr{P}_*(X)$ with
$$\mathrm{Ric}_Q(\rho_0) = \inf_{\rho \in \mathscr{P}(X)} \mathrm{Ric}_Q(\rho),$$
and if they exist, to provide a characterization of them.

2. To derive the Bonnet-Myers type diameter bound in Theorem 23.3, our approach applies the gradient estimate only to those ρ with local structure, namely $\rho = 1_x + \varepsilon \sum_{y \in S_1(x)} 1_y$. This could be a main reason that this diameter bound is likely not optimal. It would be interesting to investigate whether utilizing non-local measures could lead to a better diameter bound, which possibly becomes sharp in the case of a simple random walk on hypercubes.

Bibliography

[AG13] Luigi Ambrosio and Nicola Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013.

[Bak97] D. Bakry. On Sobolev and logarithmic Sobolev inequalities for Markov semigroups. In *New trends in stochastic analysis (Charingworth, 1994)*, pages 43–75. World Sci. Publ., River Edge, NJ, 1997.

[BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[BC64] Richard L. Bishop and Richard J. Crittenden. *Geometry of manifolds*. Pure and Applied Mathematics, Vol. XV. Academic Press, New York-London, 1964.

[BCL+18] D. P. Bourne, D. Cushing, S. Liu, F. Münch, and N. Peyerimhoff. Ollivier-Ricci idleness functions of graphs. *SIAM J. Discrete Math.*, 32(2):1408–1424, 2018.

[BCLL17] F. Bauer, F. Chung, Y. Lin, and Y. Liu. Curvature aspects of graphs. *Proc. Amer. Math. Soc.*, 145(5):2033–2042, 2017.

[BD97] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pages 223–231. IEEE, 1997.

[BE84] Dominique Bakry and Michel Émery. Hypercontractivité de semigroupes de diffusion. *C. R. Acad. Sci. Paris Sér. I Math.*, 299(15):775–778, 1984.

[Bre87] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.

[BS09] Anca-Iuliana Bonciocat and Karl-Theodor Sturm. Mass transportation and rough curvature bounds for discrete spaces. *J. Funct. Anal.*, 256(9):2944–2966, 2009.

[CEMS01] Dario Cordero-Erausquin, Robert J. McCann, and Michael Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146(2):219–257, 2001.

[Cha84] Isaac Chavel. *Eigenvalues in Riemannian geometry*, volume 115 of *Pure and Applied Mathematics*. Academic Press, Inc., Orlando, FL, 1984. Including a chapter by Burton Randol, With an appendix by Jozef Dodziuk.

[Che73] Paul R. Chernoff. Essential self-adjointness of powers of generators of hyperbolic equations. *J. Functional Analysis*, 12:401–414, 1973.

[Che75] Shiu Yuen Cheng. Eigenvalue comparison theorems and its geometric applications. *Math. Z.*, 143(3):289–297, 1975.

[CK19] D. Cushing and S. Kamtue. Long-scale Ollivier Ricci curvature of graphs. *Anal. Geom. Metr. Spaces*, 7(1):22–44, 2019.

[CKK+20] D. Cushing, S. Kamtue, J. Koolen, S. Liu, F. Münch, and N. Peyerimhoff. Rigidity of the Bonnet-Myers inequality for graphs with respect to Ollivier Ricci curvature. *Adv. Math.*, 369:107188, 53, 2020.

[CKK+21] David Cushing, Supanat Kamtue, Riikka Kangaslampi, Shiping Liu, and Norbert Peyerimhoff. Curvatures, graph products and Ricci flatness. *J. Graph Theory*, 96(4):522–553, 2021.

[CKLP21] David Cushing, Supanat Kamtue, Shiping Liu, and Norbert Peyerimhoff. Bakry-Émery curvature on graphs as an eigenvalue problem. *arXiv preprint arXiv:2102.08687*, 2021.

[CLMP20] David Cushing, Shiping Liu, Florentin Münch, and Norbert Peyerimhoff. *Curvature calculations for antitrees*, page 21–54. London Mathematical Society Lecture Note Series. Cambridge University Press, 2020.

[CLP20] David Cushing, Shiping Liu, and Norbert Peyerimhoff. Bakry-Émery curvature functions on graphs. *Canad. J. Math.*, 72(1):89–143, 2020.

[CY96] F. R. K. Chung and S.-T. Yau. Logarithmic Harnack inequalities. *Math. Res. Lett.*, 3(6):793–812, 1996.

[dC92] Manfredo Perdigão do Carmo. *Riemannian geometry.* Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[Dod83] Jozef Dodziuk. Maximum principle for parabolic inequalities and the heat flow on open manifolds. *Indiana Univ. Math. J.*, 32(5):703–716, 1983.

[DS08] Sara Daneri and Giuseppe Savaré. Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.*, 40(3):1104–1122, 2008.

[EF18] Matthias Erbar and Max Fathi. Poincaré, modified logarithmic Sobolev and isoperimetric inequalities for Markov chains with non-negative Ricci curvature. *J. Funct. Anal.*, 274(11):3056–3089, 2018.

[Elw91] K. D. Elworthy. Manifolds and graphs with mostly positive curvatures. In *Stochastic analysis and applications (Lisbon, 1989)*, volume 26 of *Progr. Probab.*, pages 96–110. Birkhäuser Boston, Boston, MA, 1991.

[EM12] Matthias Erbar and Jan Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.

[G$^+$10] Jean Gallier et al. The schur complement and symmetric positive semidefinite (and definite) matrices. *Penn Engineering*, pages 1–12, 2010.

[Gaf54] Matthew P. Gaffney. A special Stokes's theorem for complete Riemannian manifolds. *Ann. of Math. (2)*, 60:140–145, 1954.

[GHL90] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian geometry.* Universitext. Springer-Verlag, Berlin, second edition, 1990.

[Gri09] Alexander Grigor'yan. *Heat kernel and analysis on manifolds*, volume 47 of *AMS/IP Studies in Advanced Mathematics.* American Mathematical Society, Providence, RI; International Press, Boston, MA, 2009.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[Jos08] Jürgen Jost. *Riemannian geometry and geometric analysis*. Universitext. Springer-Verlag, Berlin, fifth edition, 2008.

[Kam20a] Supanat Kamtue. Bonnet-myers sharp graphs of diameter three. *arXiv preprint arXiv:2005.06704*, 2020.

[Kam20b] Supanat Kamtue. A note on a bonnet-myers type diameter bound for graphs with positive entropic ricci curvature. *arXiv preprint arXiv:2003.01160*, 2020.

[Kan42] L Kantorovich. On the transfer of masses: Dokl. *Acad. Nauk. USSR*, 37:7–8, 1942.

[Kar84] Leon Karp. Noncompact Riemannian manifolds with purely continuous spectrum. *Michigan Math. J.*, 31(3):339–347, 1984.

[KLW13] Matthias Keller, Daniel Lenz, and Radosław K. Wojciechowski. Volume growth, spectrum and stochastic completeness of infinite graphs. *Math. Z.*, 274(3-4):905–932, 2013.

[KMS04] J. H. Koolen, V. Moulton, and D. Stevanović. The structure of spherical graphs. *European J. Combin.*, 25(2):299–310, 2004.

[Lic58] André Lichnerowicz. *Géométrie des groupes de transformations*. Travaux et Recherches Mathématiques, III. Dunod, Paris, 1958.

[LL15] Yong Lin and Shuang Liu. Equivalent properties of cd inequality on graph. *arXiv preprint arXiv:1512.02677*, 2015.

[LLY11] Yong Lin, Linyuan Lu, and Shing-Tung Yau. Ricci curvature of graphs. *Tohoku Math. J. (2)*, 63(4):605–627, 2011.

[LMP17] Shiping Liu, Florentin Münch, and Norbert Peyerimhoff. Rigidity properties of the hypercube via Bakry-Emery curvature. *arXiv preprint arXiv:1705.06789*, 2017.

[LMP18] Shiping Liu, Florentin Münch, and Norbert Peyerimhoff. Bakry-Émery curvature and diameter bounds on graphs. *Calc. Var. Partial Differential Equations*, 57(2):Paper No. 67, 9, 2018.

[Lot08] John Lott. Some geometric calculations on Wasserstein space. *Comm. Math. Phys.*, 277(2):423–437, 2008.

[LV09] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[LY10] Yong Lin and Shing-Tung Yau. Ricci curvature and eigenvalue estimate on locally finite graphs. *Math. Res. Lett.*, 17(2):343–356, 2010.

[Maa11] Jan Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.

[Maa17] Jan Maas. Entropic Ricci curvature for discrete spaces. In *Modern approaches to discrete curvature*, volume 2184 of *Lecture Notes in Math.*, pages 159–174. Springer, Cham, 2017.

[Mat10] M. Matsumoto. (Japanese). Master's thesis, Mathematical Institute, Tokyo University, Japan, 2010.

[McC01] Robert J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11(3):589–608, 2001.

[Mon81] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

[Mye41] S. B. Myers. Riemannian manifolds with positive mean curvature. *Duke Math. J.*, 8:401–404, 1941.

[Oba62] Morio Obata. Certain conditions for a Riemannian manifold to be isometric with a sphere. *J. Math. Soc. Japan*, 14:333–340, 1962.

[Oht07a] Shin-ichi Ohta. On the measure contraction property of metric measure spaces. *Comment. Math. Helv.*, 82(4):805–828, 2007.

[Oht07b] Shin-Ichi Ohta. Products, cones, and suspensions of spaces with the measure contraction property. *J. Lond. Math. Soc. (2)*, 76(1):225–236, 2007.

[Oht14] Shin-ichi Ohta. Ricci curvature, entropy, and optimal transport. In *Optimal transportation*, volume 413 of *London Math. Soc. Lecture Note Ser.*, pages 145–199. Cambridge Univ. Press, Cambridge, 2014.

[Oll09] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.

[OS14] Shin-ichi Ohta and Karl-Theodor Sturm. Bochner-Weitzenböck formula and Li-Yau estimates on Finsler manifolds. *Adv. Math.*, 252:429–448, 2014.

[OSY20] Ryunosuke Ozawa, Yohei Sakurai, and Taiki Yamada. Maximal diameter theorem for directed graphs of positive Ricci curvature. *arXiv preprint arXiv:2011.00755*, 2020.

[Ott01] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[OV12] Y. Ollivier and C. Villani. A curved Brunn-Minkowski inequality on the discrete hypercube, or: what is the Ricci curvature of the discrete hypercube? *SIAM J. Discrete Math.*, 26(3):983–996, 2012.

[PC19] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[Ral17] Peter Ralli. Bounds on curvature in regular graphs. *arXiv preprint arXiv:1701.08205*, 2017.

[San15] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[Sch99] Michael Schmuckenschläger. Curvature of nonlocal Markov generators. In *Convex geometric analysis (Berkeley, CA, 1996)*, volume 34 of *Math. Sci. Res. Inst. Publ.*, pages 189–197. Cambridge Univ. Press, Cambridge, 1999.

[Stu06] Karl-Theodor Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[TD87] Paul Terwilliger and Michel Deza. The classification of finite connected hypermetric spaces. *Graphs Combin.*, 3(3):293–298, 1987.

[Vil09] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[vRS05] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

[Woj08] Radoslaw Krzysztof Wojciechowski. *Stochastic completeness of graphs*. PhD thesis, City University of New York, 2008.

[Yau76] Shing Tung Yau. Some function-theoretic properties of complete Riemannian manifold and their applications to geometry. *Indiana Univ. Math. J.*, 25(7):659–670, 1976.