Note on Curvature of Cayley Graphs

Anthony Hong

January 18, 2024

Contents

Abstract

The note first makes sense of curvature on geometric objects beyond classical Riemannian manifolds, specifically on Markov chains in metric spaces and on combinatorial graphs. We also examine compatibility of new definitions with classical ones and applicability of new definitions. In particular, algorithmic aspects of computation of curvature of Cayley graph of certain groups will be presented as the main course at the end.

1 Introduction

From the counter-intuitive observation of roundness of seemingly flat Earth to the revolutionary discovery of warping of space time, the captivating concept of curvature leads to various exciting mathematical and physical phenomena.

Curvature is a fundamental characteristic of geometric objects that plays a crucial role in encoding both local geometric information and global topological ambience. It applies to control theory, diffusion process, and shape analysis, to name a few. But can this profound notion of curvature extend to the discrete spaces encountered in various scientific disciplines? Since real-life information is stored in computers via discontinuous data structures, a discrete version of curvature helps researchers handle data in large amount and

recover a pseudo-curvature description to infer the real-life situations. For instance, one may expect that curvature under such definitions converge to the classical ones when discrete models tend to be more and more the same as the Riemannian manifold in terms of thickening density of edges, increasing number of complex cells, or asympototic behavior of random graphs.

The challenge lies in bridging the gap between the smooth world of Riemannian manifolds, where curvature as a measurement of sensitivity is naturally built upon various infinitestimal notions, and the discrete realm of graphs and networks, where some forms of finiteness are often imposed and differences replace differentials. Accordingly, handful tools around the two shift from topological and analytic ones for the former to linear-algebraic and group-theoretic ones for the latter. However, we find inspiration in the fact that certain concepts from Riemannian manifolds still have analogs in discrete spaces, like Laplacians and heat kernels, which motivate us to define discrete versions of curvature that preserve the essential properties.

This note begins with a short survey of ways to generalize curvature, provides some basic languages in probability theory for later use, and applies specifically the Ollivier-Ricci type to Cayley graph with computational examples.

1.1 A Survey on Generalizations of Curvature

The study of curvature traces back to Gauss's *Theorema Egregium* (Outstanding Theorem). He concerned surfaces in $\mathbb{R}^3$, where tangent spaces are naturally identified with the same Euclidean space and thus involve no issues in covariant derivative, a notion of differentiation on Riemannian manifold (In fact, manifolds generalize the surfaces and Euclidean spaces to an extenet that analysis can still find its place). We refer to lecture note [24] for a review of how these ideas originated and discuss some successful attempts of generalization under its framimg of them by three main types, combinatorial curvature, Ollivier-Ricci curvature, and Gromov's δ-hyperbolicity, although there are also others like Steinerberger's potential-theoretic definition [29] that do not quite fall into any of these three.

Combinatorial curvature

This approach is perhaps the most intuitive one, as the manifolds are discretized in an obvious way, namely by their cellular structures. We provide two variants below.

The global result Gauss-Bonnet theorem which relates curvature with combinatorial data, namely the Euler characteristic χ, inspires researchers, among them [13] and [4], to give what [24] (Definition 2.1) summarizes as a curvature notion on combinatorial representation $G = (V, E, F)$ of tessellation $\mathcal{T}$ of a surface S for a vertex x:

$$K(x) = 2\pi - \sum_{f : x \in v(f)} \frac{|e(f)| - 2}{|e(f)|}\pi$$

where $v(f)$ and $e(f)$ stand for the vertex set and edge set of the Euclidean polygon $f \in F$ with interior angles $\frac{|e(f)|-2}{|e(f)|}\pi$.

Forman in [10] sees the potential that both a Riemannian manifold and a graph have Laplacians defined on them. In fact, Bochner-Weitzenböck formula tells us that we can recover information about Ricci curvature from the Laplacian. This inspires Forman to work backwards using combinatorial Laplace operator acting on p-chains to define Ricci curvature for a large class of topological spaces, weighted CW-complexes, including graphs $G = (V, E)$ regarded as 1-skeletons (vertices as 0-cells and edges as 1-cells). [15] writes this for an edge e of graph G as

$$F(e) = w_e \left(\frac{w_{v_1}}{w_e} + \frac{w_{v_2}}{w_e} - \sum_{e_{v_1} \sim e, e_{v_2} \sim e} \left[\frac{w_{v_1}}{\sqrt{w_{e_{v_1}} w_e}} + \frac{w_{v_2}}{\sqrt{w_{e_{v_2}} w_e}} \right] \right)$$

where $e = (v_1\, v_2) \in E$; w_e is the weight of e; w_{v_1} and w_{v_2} denote the weights associated with v_1 and v_2; and $e_{v_1} \sim e$ and $e_{v_2} \sim e$ denote the set of edges incident to v_1 and v_2.

Ollivier-Ricci curvature

This branch started with Bakry and Émery's celebrated paper [2], followed by for instance [28], continued with Sturm [30] and Lott and Villani [19]'s generalization of Ricci curvature on metric measure spaces as geodesic spaces by use of optimal transportation theory and analysis of entropy functional on space of probability measures equipped with W_2 distance, culminated at a far more applicatively rich definition of Ricci curvature for Markov chains on metric spaces by Ollivier [23] using also optimal transport but in W_1 distance. As he claimed, Ollivier-Ricci curvature is "very easy to implement on concrete examples." Lin, Lu, and Yau obtained a modified Ollivier-Ricci curvature for graphs in particular [18]. For Cayley graphs, there are also variants like conjugation curvature introduced by [3] for a computational ease.

Since this note will present their constructions in details, we don't give mathematical expression in this short survey. Basically, they observe that small balls move closer to each other along geodesics of a positively Ricci-curved space and try to generalize curvature notion by finding alternatives of concept of balls in other settings. One recent work [14] is of interest in relating the (generalized) Forman type and Ollivier-Ricci type. Book [21] provides a comprehensive compilation in the combinatorial type and Ollivier-Ricci type.

Gromov's δ-hyperbolicity

Gromov's idea of δ-hyperbolocity comes from the accustomed experience that triangles are shaped differently in spaces of different curvature. Figure 1 shows the typical three triangles on Riemann surfaces of classification theorem.

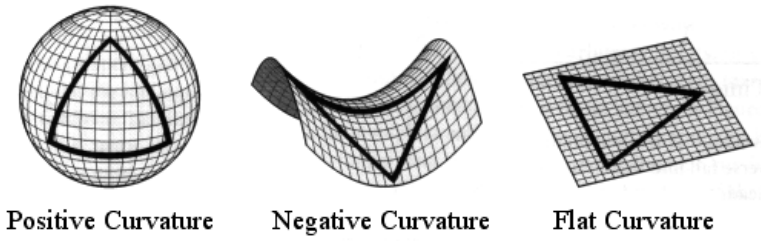

Figure 1: Triangles of three characteristic surfaces of different curvatures.*

A geodesic triangle $\triangle$ (i.e., its three sides are geodesics) is called δ-thin if each side of the triangle is contained in the δ-tubes of the other two sides. A geodesic space X (i.e., each pair of points in X have a minimum-attainable path connecting them) is called δ-hyperbolic if there is some constant $\delta \geq 0$ such that any geodesic triangle $\triangle$ is δ-thin. Figure 2 is an illustration of a δ-thin triangle.

We seek a discretization of the above characterization of hyperbolicity: this only needs us to render graphs also as geodesic metric spaces. Two approaches are considered in [24], and we are not going to enter into details. What's more interesting is that there is also a development for Cayley graphs in particular that amounts to be called geometric group theory. In short, geometric group theory analyzes group actions on geometric spaces and geometric properties of finitely-generated groups by equipping them with word metric, or the length metric of their Cayley graphs.

1.2 Preliminaries in Probability

We refer to [8] for a basic setup in probability.

Let $(E, \mathcal{E})$ be a measurable space, $(F, \mathcal{F}, \nu)$ be a measure space, and $f : F \to E$ be a measurable function. We call $f_{\#}\nu := \nu \circ f^{-1} : \mathcal{E} \to [0, \infty]$ the **image of** ν **under** f. It is easy to verify $\mu = f_{\#}\nu$ is indeed a measure

*Figure taken from Harrison Hartle's blog.
†Figure taken from Wikimedia Commons.

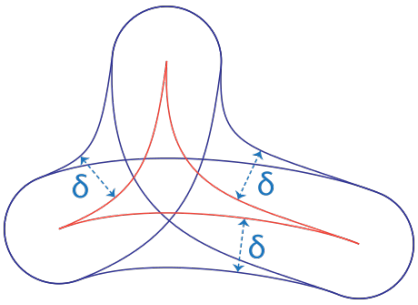

Figure 2: δ-thin triangle illustration[†]

and that for any $g \in \mathcal{E}_+$ (i.e. g is non-negative $\mathcal{E}$-measurbale), one has $\int_E g(y)\,\mu(dy) = \int_F (g \circ f)(x)\,\nu(dx)$.

We observe the fact $\mu \ll \nu \Rightarrow f_\# \mu \ll f_\# \nu$ for measurable function f and the following composition rule:

$$(f \circ g)_\# \mu = f_\#(g_\# \mu) \tag{1}$$

Another measure built from construction is given by the following. Let $(E, \mathcal{E}, \nu)$ be a measure space and $f \in \mathcal{E}_+$. Define $\mu : \mathcal{E} \to [0,\infty]; \mu(A) = \nu(f\mathbb{1}_A) = \int_A f(x)\,\nu(dx)$. One deduces from monotone convergence theorem that $(E, \mathcal{E}, \mu)$ is now a new measure space. μ is called the **indefinite integral of f with respect to ν**. It is easy to check that for any $g \in \mathcal{E}_+$, one has $\int_E g(y)\,\mu(dy) = \int_E f(x)g(x)\,\nu(dx)$.

Recall that for measures μ and ν on a measurable space $(E, \mathcal{E})$, μ is said to be **absolutely continuous with respect to** ν ($\mu \ll \nu$) if $\forall A \in \mathcal{E} : \nu(A) = 0 \Rightarrow \mu(A) = 0$. If μ is the indefinite integral of f with respect to ν in the above sense then we see $\mu \ll \nu$. The Radon-Nikodym theorem gives the converse:

Theorem 1.1 (Radon-Nikodym Theorem)**.** *Suppose ν is a σ-finite measure and $\mu \ll \nu$. Then, there exists $f \in \mathcal{E}_+$ such that*

$$\int_E g(y)\,\mu(dy) = \int_E f(x)g(x)\,\nu(dx) \quad g \in \mathcal{E}_+$$

*The function f is unique in the following sense: if f' is another function with the above property then $f = f'$ $\nu - a.e.$. Besides, f is denoted as $d\mu/d\nu$, called the **Radon-Nikodym derivative** of μ with respect to ν.*

We thus sometimes use $\mu(dy) = f(x)\nu(dx)$ to define measure μ for known $f \in \mathcal{E}_+$ and ν.

Let $(\Omega, \mathcal{D}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ be a measurable space. A map $X : \Omega \to E$ is called a **random variable** in $(E, \mathcal{E})$ if it is measurable relative to $\mathcal{D}$ and $\mathcal{E}$, i.e., $X^{-1}(A) = \{X \in A\} = \{\omega \in \Omega : X(\omega) \in A\}$ is an event for every A in $\mathcal{E}$. A **distribution** of X is the probability measure $\mu = X_\# \mathbb{P}$ on $(E, \mathcal{E})$ as the image of $\mathbb{P}$ under X, i.e., $A \in \mathcal{E} : \mu(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}\{X \in A\}$. Notice that the knowledge of how μ acts on a π-system (a collection of subsets of E that is closed under intersection) that generates $\mathcal{E}$ characterizes μ (see [8]).

Example 1.2 (CDF)**.** When $E = \dot{\mathbb{R}} = [-\infty, \infty]$ and $\mathcal{E} = \mathcal{B}_E$, the Borel σ-algebra of E, the intervals $[-\infty, x]$ with x in $\mathbb{R}$ form a convenient π-system. It is thus enough to specify $\mu = X_\# \mathbb{P} = \mathbb{P} \circ X^{-1}$ by evulating function $c(x) = \mu[-\infty, x] = \mathbb{P}\{X \le x\}$ for each $x \in \mathbb{R}$. And $c : \mathbb{R} \to [0,1]$ is called the **cumulative distribution function (CDF)** of X. CDF is extensively used in elementary probability texts to avoid a measure-theoretic approach. $\Diamond$

Another example of specifying distribution atomically is shown below and will be mainly used for markov chains on countable set for subsequent sections.

Example 1.3 ($(E, \mathcal{E}) = (I, \mathcal{I})$)**.** When the space E takes values is now any countable set I, and the σ-algebra $\mathcal{E}$ is the power set $\mathcal{I} = \wp(I)$, we use the pi-system $i \in I$ to specify the measure $\lambda = \mathbb{P} \circ X^{-1}$:

$$\lambda_i := \lambda(i) = \mathbb{P}(X = i) = \mathbb{P}\{\omega \in \Omega : X(\omega) = i\}$$

Since λ is a measure on $(I, \mathcal{I})$, we see $1 = \lambda(I) = \lambda\left(\bigcup_{i \in I} i\right) = \sum_{i \in I} \lambda_i$. $\Diamond$

1.3 Couplings and Optimal Transportation

[32] and [27] have pedagogical introductions to the theory of optimal transport, the solution to the Monge-Kantorovich problem. Roughly speaking, an optimal transport is the cost-minimizing plan the specifies the transferrence of each unit of mass from one place where the masses are distributed to another where they are distributed differently. We give necessary definitions directly, and later on we shall see the narrative above is clear in discrete settings.

We assume that the spaces below are **Polish space** (i.e., a separable complete metric space).

Definition 1.4 (Coupling; see [32])**.** A **coupling** of two probability measures μ and ν on the measurable spaces $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ respectively is any probability measure π on the product measurable space $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ (where $\mathcal{X} \otimes \mathcal{Y}$ is the smallest σ-algebra containing $\mathcal{X} \times \mathcal{Y}$) whose marginals are μ and ν, i.e.,

$$\pi(A \times Y) = \mu(A), \quad \pi(X \times B) = \nu(B) \tag{2}$$

for all measurable subsets A of X and B of Y. Note that if we let $\mathrm{Pr}_i, i = 1, 2$, be the projections from $X \times Y$ to index spaces we have $A \times Y = \mathrm{Pr}_1^{-1}(A)$ and $X \times B = \mathrm{Pr}_2^{-1}(B)$. Hence, condition (2) is equivalent to

$$\mu = \pi \circ \mathrm{Pr}_1^{-1} = \mathrm{Pr}_{1\#}\pi, \quad \nu = \pi \circ \mathrm{Pr}_2^{-1} = \mathrm{Pr}_{2\#}\pi \tag{3}$$

We denote all such couplings of μ and ν as $\Pi(\mu, \nu)$. ♦

The Kantorovich problem is to find the minimizer π that achieves

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c(x, y)\, d\pi(x, y)$$

where $c(x, y)$ is some cost function, telling us how much it costs to transport one unit of mass from location x to location y. It is shown that if one puts some regularity conditions onto the spaces and functions, the infimum can be actaully attained (see [32] Theorem 1.3). Informally, $d\pi(x, y)$ measures the amount of mass transferred from x to y, and the integration of c over $d\pi$ represents the total cost of a particular transferrence plan π. Inspired by the above result, we consider a scenario in which those regularity conditions are satisfied.

Definition 1.5 (Wasserstein Distance)**.** Let (X, d) be a metric space and μ, ν be two probability measures on X. The **Wasserstein distance** between μ and ν is defined as

$$W(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d(x, y)\, d\pi(x, y). \quad ♦$$

Remark 1.6. One can consult the reference [1], in particular chapter 7, to see that the above definition indeed gives a distance on the space of probability measures $\mathcal{P}(X)$: symmetry is obvious; positivity is checked because $W(\mu, \nu) = 0$ implies $\exists \pi \in \Pi(\mu, \nu)$ s.t. $\int d\, d\pi = 0$ and thus $x = y$ π-a.e., implying that

$$\pi(\text{Diagonal}) > 0,\ \pi(X^2 - \text{Diagonal}) = 0 \Rightarrow \pi = \mathrm{Dup}_{\#}\mu$$

where Dup is the duplicate $x \mapsto (x, x)$, which by composition rule (1) gives $\nu = \mathrm{Pr}_{2\#}\pi = \mathrm{id}_{\#}\mu = \mu$; triangle inequality requires a bit more terminology, namely the composition of transferrence plans (see [1] Remark 5.3.3). However, the triangle inequality is intuitive in our mass-moving narration. Plan A moves masses from place (dist.) μ to place (dist.) ν and plan B from ν to λ, both minimizing the costs on their own, while the optimal plan C minizmies the cost of moving masses from μ to final destination λ, without first going to some intermediate place (dist.) ν.

Example 1.7 (Optimal transport on graph)**.** We note that a graph $G = (V, E)$ (assumed to be simple, connected, and with countable vertex set V) can be realized as a metric space V equipped with metric d, which is the number of edges of the shortest path connecting two vertices. Now probability distributions are specified in manner of Example 1.3, and optimal transport is much simpler in this discrete setting: just as μ

and ν are determined by $\mu(x), \nu(x)$ for $x \in V$, we see that the coupling $\pi \in \Pi(\mu, \nu)$ is a map $\pi : \mathcal{V} \times \mathcal{V} \to [0, 1]$ determined by its value on each $(x, y) \in V \times V$ and it satisfies the marginal conditions, where $\mathcal{V}$ is the power set as in Example 1.3. It is clear that by countable additivity $\pi(A \times V) = \mu(A) \forall A \in \mathcal{V} \Leftrightarrow \pi(x \times V) = \mu(x) \forall x \in V$ with the other for ν analogously given. Wasserstein distance can be defined accordingly. We thus reformulate the definition of optimal transport on graph to be consistent with Lin-Lu-Yau's [18], Peyerimhoff's [24], and Mizukai and Sako's [20] as following. $\Diamond$

Definition 1.8 (Transport Plan and Wasserstein Distance on Graph)**.** Let $G = (V, E)$ be a simple connected graph and $\mu, \nu \in \mathcal{P}(V)$ be two probability measures. The **transport plan** from μ to ν is a map $\pi : V \times V \to [0, 1]$ such that

$$v \in V : \mu(v) = \sum_{w \in V} \pi(v, w) \quad \text{and} \quad w \in V : \nu(w) = \sum_{v \in V} \pi(v, w)$$

The **(total) cost function** of π is

$$\text{cost}(\pi) = \sum_{v,w \in V} d(v, w)\pi(v, w)$$

and we call $d(v, w)\pi(v, w)$ **transport cost from** v **to** w. We denote the set of all such transport plans from μ to ν by $\Pi(\mu, \nu)$. We define the **Wasserstein distance** between the two measures μ and ν in $\mathcal{P}(V)$ as

$$W(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \text{cost}(\pi)$$

A transport plan $\pi \in \Pi$ is called **optimal** if we have $W(\mu, \nu) = \text{cost}(\pi)$. $\blacklozenge$

An example computing optimal transport in the above sense is postponed to section 3.3 (see example 3.6).

2 Markov Chain

2.1 Markov Chains on Countable Set

A **stochastic process** is a collection of E-valued random variables $\{X_t : \Omega \to E\}_{t \in T}$ with **state space** $(E, \mathcal{E})$ and **parameter set** T. For each $\omega \in \Omega$, let $X(\omega)$ denote the function $T \to E; t \mapsto X_t(\omega)$; then $X(\omega)$ is an element of E^T, the collection of all functions from T to E. We may regard the stochastic process $(X_t)_{t \in T}$ as a random variable X that takes value in the product space $(E^T, \mathcal{E}^T)$, since the map $X : \Omega \to E^T; \omega \mapsto X(\omega)$ is measurable relative to $\mathcal{D}$ and $\mathcal{E}^T$. When $T = \mathbb{N} = \{0, 1, \cdots\}$ and $(E, \mathcal{E}) = (I, \mathcal{I})$, we come to the definition of the first major class of stochastic process in this note, adapted from [22]:

Definition 2.1 (Discrete-Time Markov Chain)**.** We say $(X_n)_{n \in \mathbb{N}}$ is a **discrete-time Markov chain** with **initial distribution** λ and **transition matrix** $P = (p_{ij})_{i,j \in \mathbb{N}}$, or **Markov**$(\lambda, P)$ for short, if

(i) X_0 has distribution λ, i.e., $\forall i_0 \in I : \mathbb{P}(X_0 = i_0) = \lambda_{i_0}$

(ii) for $n \geq 0$, conditional on $X_n = i$, X_{n+1} has distribution $(p_{ij})_{j \in I}$, i.e., $\mathbb{P}(X_{n+1} = j | X_n = i) = p_{ij}$;

(iii) for $n \geq 0$, $(X_{n+1} = j | X_n = i)$ is independent of $X_0, \cdots, X_{n-1}$. $\blacklozenge$

We leave it as an exercise by induction to show that (iii) is equivalent of saying

$$\forall i_0, \cdots, i_{n+1} \in I : \mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \cdots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n)$$

Continuous-time Markov chain $(X_t)_{t \geq 0}$ is a stochastic process with t belonging to an uncountable parameter set $T = [0, \infty)$ and random variables $X_t : \Omega \to I$, where I is still a countable set. Formal treatment of this requires us to understand Q-matrices (see [22] chapter 2 and 3). We continue our study of the discrete version. Here are two properties of it.

Proposition 2.2.

(i) $(X_n)_{0\le n\le N}$ *is Markov*(λ, P) *if and only if*

$$\forall i_0, \cdots, i_N \in I : \mathbb{P}(X_0 = i_0, \cdots, X_1 = i_1, \cdots, X_N = i_N) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N}$$

(ii) ***(Markov property)****: Let* $(X_n)_{n\ge 0}$ *be Markov*(λ, P)*. Then, conditional on* $X_m = i$*,* $(X_{n+m})_{n\ge 0}$ *is Markov*(δ_i, P) *and is independent of the random variables* $X_0, \cdots, X_m$*, where* $\delta_i = (\delta_{ij} : j \in I)$ *is the* ***unit mass*** *at* i *and*

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

We regard distributions and measures λ as row vectors whose components are indexed by I, just as P is a matrix whose entries are indexed by $I \times I$. We extend the matrix multiplication and multiplicaiton of matrix by row vector to the general sense in the obvious way, defining a new measure λP and a new matrix P^2 by

$$(\lambda P)_j = \sum_{i\in I} \lambda_i p_i j, \quad (P^2)_{ij} = \sum_{k\in I} p_{ik} p_{kj}.$$

We define P^m similarly for any n. We set $P^0 = I$ where $(I)_{ij} = \delta_{ij}$. We write $p_{ij}^{(n)} = (P^n)_{ij}$ for the (i, j) entry in P^n. In the case where $\lambda_i > 0$ we shall write $\mathbb{P}_i(A)$ for the conditional probability $\mathbb{P}(A|X_0 = i)$. By the Markov property at time $m = 0$, under $\mathbb{P}_i$, $(X_n)_{n\ge 0}$ is Markov(δ_i, P). So the behaviour of $(X_n)_{n\ge 0}$ under $\mathbb{P}_i$ does not depend on λ.

Theorem 2.3 ([22] Theorem 1.1.3)**.** *Let* $(X_n)_{n\ge 0}$ *be Markov*(λ, P)*. Then, for all* $n, m \ge 0$*,*

(i) $\mathbb{P}(X_n = j) = (\lambda P^n)_j$*;*

(ii) $\mathbb{P}_i(X_n = j) = \mathbb{P}(X_{n+m} = j | X_m = i) = p_{ij}^{(n)}$

Proof. (i) By Proposition 2.2 (i)

$$\begin{aligned} \mathbb{P}(X_n = j) &= \sum_{i_0\in I} \cdots \sum_{i_{n-1}\in I} \mathbb{P}(X_0 = i_0, \cdots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_0\in I} \cdots \sum_{i_{n-1}\in I} \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} j} = (\lambda P^n)_j \end{aligned}$$

(ii) By the Markov property, conditional on $X_m = i$, $(X_{m+n})_{n\ge 0}$ is Markov(δ_i, P); then take $\lambda = \delta_i$ in (i). □

In light of this theorem we call $p_{ij}^{(n)}$ the **n-step transition probability from** i **to** j. The following examples illustrates how to calculate it.

Example 2.4 ([22] Example 1.1.4)**.** Suppose that whether it rains tomorrow depends on previous weather conditions only through whether it is raining today. Suppose further that if it is raining today, then it won't rain tomorrow with probability α, and if it is not raining today, then it will rain tomorrow with probability β. If we say that the system is in state 0 when it rains and state 1 when it does not, then the preceding system is a two-state Markov chain having transition probability matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

and is represented by Figure 3.

We exploit the relation $P^{n+1} = P^n P$ to write

$$p_{00}^{(n+1)} = (P^{n+1})_{00} = (P^n P)_{00} = \sum_{k=0,1} p_{0k}^{(n)} p_{k0} = p_{00}^{(n)} p_{00} + p_{01}^{(n)} p_{10} = p_{00}^{(n)}(1-\alpha) + p_{01}^{(n)} \beta$$

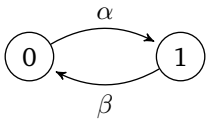

Figure 3: A two-state Markov chain

We also know that $p_{00}^n + p_{01}^{(n)} \overset{\text{Thm 2.3 (ii)}}{=\!=\!=} \mathbb{P}_0(X_n = 0 \text{ or } 2) = 1$, so by eliminating $p_{01}^{(n)}$ we get a recurrence relation for $p_{00}^{(n)}$:

$$p_{00}^{(n+1)} = p_{00}^{(n)}(1-\alpha-\beta)+\beta, \quad p_{00}^{(0)} = 1.$$

This has a unique solution:

$$p_{00}^{(n)} = \begin{cases} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n, & \text{for } \alpha+\beta > 0, \\ 1, & \text{for } \alpha+\beta = 0. \end{cases} \qquad \lozenge$$

As illustrated in the scenario above, we percieve $X = (X_n)_{n\geq 0}$ as a kind of "random walker" on a graph with $|I|$ vertices ($|I| = 2$ in this case). The example only specifies the P matrix, but a complete discrete-time Markov chain also includes an intial distirbution by which the random variable X_0 assigns the position i_0 where the walker is born. X_1 assigns i_1 where the walker goes from i_0, $\cdots$, X_{n+1} assigns i_{n+1} where the walker goes from i_n. p_{ij} is the probability that X will go to position j given X is currently at position i, but notice that X's choice is regardless of the time or step when X is at. Namely,

$$\forall i_0, \cdots, i_{n+1} \in I : \mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \cdots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) = p_{ij}$$

A **random walk** is a process for traversing a graph where at every step we follow an outgoing edge chosen uniformly at random. As a comparison, a Markov chain is similar except the outgoing edge is chosen according to an arbitrary fixed distribution. That is, the distribution is not necessarily a uniform distribution.

Example 2.5 ([12])**.** We consider a "random walker" in a very small town consisting of four streets, and four street-corners v_1, v_2, v_3 and v_4 arranged as in Figure 4.

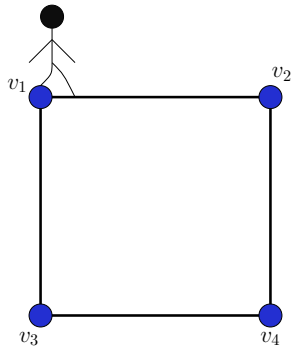

Figure 4: A random walker in a very small town

At time 0, the random walker stands in corner v_1. At time 1, he flips a fair coin and moves immediately to v_2 or v_4 according to whether the coin comes up heads or tails. At time 2, he flips the coin again to decide which of the two adjacent corners to move to, with the decision rule that if the coin comes up head, then he moves one step clockwise, while if it is tail, he moves one step counterclockwise. Then iterate this at times $3, 4, \cdots$

For each n, let X_n denote the index of the street-corner at which the walker stands at time n. Hence, $(X_0, X_1, \cdots)$ is a random process taking values in $\{1, 2, 3, 4\}$. Since the walker starts at time 0 in v_1, we have

$\mathbb{P}(X_0 = 1) = 1$. Next, he will move to v_2 or v_4 with probability $\frac{1}{2}$ each, so that $\mathbb{P}(X_1 = 2) = \mathbb{P}(X_1 = 2) = \frac{1}{2}$. To compute the distribution of X_n for $n \geq 2$ requires a little more thought. Because of the current coin-flipping mechanism for deciding where to go next at time $n+1$ is independent of all past coin-flippings, we get the same conditional probabilities if we condition on the full history of the process up to time n, i.e.,

$$\mathbb{P}(X_{n+1} = v_1 | X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1}, X_n = v_2) = \frac{1}{2}$$

$$\mathbb{P}(X_{n+1} = v_3 | X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1}, X_n = v_2) = \frac{1}{2}$$

for any choice of $i_0, \cdots, i_{n-1}$. Hence, this random walk is a Markov chain with finite state space $\{1, 2, 3, 4\}$ and transition matrix

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

with initial distribution $\lambda = (1, 0, 0, 0)$. Notice that every row sums to 1, and that's simply because every row is a conditional probability distribution. $\Diamond$

2.2 Markov Chains on Metric Space

We now consider Markov chain on a more general space: the state space $(E, \mathcal{E})$ is now $(X, \mathcal{B}_X)$ instead of $(I, \mathcal{I})$ where X is a Polish space with metric d and $\mathcal{B}_X$ is the Borel σ-algebra of X.

Definition 2.6 (Transition Kernel)**.** Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be two measurable spaces, then a **transition kernel** K is a map $X \times \mathcal{Y} \to [0, 1]$ for which

(i) fix $x \in X$, then the following map is a probability measure on $(Y, \mathcal{Y})$.

$$\begin{aligned} K_x : \mathcal{Y} &\to [0, 1] \\ A &\mapsto K(x, A) \end{aligned}$$

(ii) fix $A \in \mathcal{Y}$, then the following map is a $\mathcal{X}$-measurable function.

$$\begin{aligned} K^A : X &\to [0, 1] \\ x &\mapsto K(x, A) \end{aligned}$$

We say K is a **Markov kernel** if in addition $(X, \mathcal{X}) = (Y, \mathcal{Y})$. $\blacklozenge$

Notice that when $(X, \mathcal{X}) = (I, \mathcal{I})$, the Markov kernel K is specified by the numbers $K(x, \{y\})$ and can be regarded as a matrix. In fact, for discrete-time markov chain, the i-th row

$$(\, p_{i1}\, , \cdots, \, p_{ij}\, , \cdots), \text{ or } (\, \mathbb{P}(X_{n+1} = 1 | X_n = i), \cdots, \mathbb{P}(X_{n+1} = j | X_n = i), \cdots)\,,$$

specifies a conditional probability distribution $\mathbb{P}(\,\cdot\,|X_n = i)$ at point i as an image measure of the conditioned X_{n+1}, and thus a probability measure on $(I, \mathcal{I})$. Check [6] Example 1.3 for more details.

Theorem 2.7 ([8] I. Thm 6.3)**.** *Let K be a transition kernel from $(E, \mathcal{E})$ into $(F, \mathcal{F})$. Then*

(i) $x \in E : Kf(x) = \int_F f(y) K(x, dy)$ *defines a function* $Kf : \mathcal{F}_+ \to \mathcal{E}_+$ *for every function* $f \in \mathcal{F}_+$*;*

(ii) $B \in \mathcal{F} : \mu K(B) = \int_E K(x, B) \mu(dx)$ *defines a measure* μK *on* $(F, \mathcal{F})$ *for each measure* μ *on* $(E, \mathcal{E})$*;*

(iii) for every measure μ on $(E, \mathcal{E})$ and function f in $\mathcal{F}_+$ one has

$$(\mu K)f = \mu(Kf) = \int_E \mu(dx) \int_F f(y)K(x, dy).$$

A formal definition of Markov chain on metric space requires a bit more reading about filtration and adapted process which one can find in [6]. We do not intend to present them there, becauese for now it suffices to use kernel to understand Ollivier-Ricci curvature and the see that a kernel gives rise to a Markov chain.

2.3 Markov Chains on Riemannian Manifold

Riemannian manifold is a nice space where we can do generalized analysis. Markov chains can be put onto Riemannian manifold in the same sense as the previous subsection, because Riemannian manifold is metrizable with length metric. Before we show this realization we first introduce some basic notions. We use the same notationas as [17] and refer to it for details.

Let V be a vector space. We use $T^k(V^*) = \bigotimes_{i=1}^k V^* \cong \left\{ \bigtimes_{i=1}^k V \xrightarrow{multilinear} \mathbb{R} \right\}$ to denote the set of all covaraint k-tensors on V. More conveniently, we denote the mixed types as $T^{(k,l)}(V) = V^{\otimes k} \otimes V^{*\otimes l}$. The set of all symmetric and alternating covaraint k-tensors is denoted as $\Sigma^k(V^*)$ and $\Lambda^k(V^*)$ respectively. We can also bundle them up with V being tangent space T_pM of manifold M (or T_p^*M for contravariant case) by taking disjoint union. Any one of them is called a tensor bundle. A section of a tensor bundle is a tensor field on M. Let (M, g) be a Riemannian manifold, where the Riemannian metric g is a smooth symmetric covariant 2-tensor field on M (i.e.,$g_p \in \Sigma^2(T_p^*M)$, or $g_p(v, w) = g_p(w, v)$, $\forall v, w \in T_pM$). Riemannian metric thus associates each tangent space T_pM an inner product $\langle \cdot, \cdot \rangle_p$. For each point p in one of the smooth charts $(U, \varphi = (x^i))$, we note that T_pM has basis $\{\frac{\partial}{\partial x^1}|_p, \cdots, \frac{\partial}{\partial x^n}|_p\}$ and T_p^*M has basis $\{dx^1|_p, \cdots, dx^n|_p\}$, so g_p can be written as

$$g_p = \sum_{1 \le i,j \le n} g_{ij}(p) dx^i|_p \otimes dx^j|_p$$

where

$$g_{ij}(p) = g_p\left(\left.\frac{\partial}{\partial x^i}\right|_p, \left.\frac{\partial}{\partial x^j}\right|_p \right)$$

and we can write g as a function taking p as follows

$$g = \sum_{1 \le i,j \le n} g_{ij} dx^i \otimes dx^j$$

where $g_{ij} : p \mapsto (g_p)_{ij}$ becomes a smooth function, and $(g_{ij})_{n\times n}$ is a symmetric positive definite matrix of smooth functions g_{ij}. We often write it in Einstein summation convention, $g = g_{ij} dx^i \otimes dx^j$, and furthermore, $g = g_{ij} dx^i dx^j$ by abbreviation of symmetrization. One can use other bases (orthonormal frames) and duals to write the above expansions. The first example is Euclidean metric $\bar{g} = \delta_{ij} dx^i dx^j = \sum (dx^i)^2$ on the manifold $\mathbb{R}^n$. Every smooth manifold with or without boundary admits a Riemannian metric by a standard partition-of-unity argument. The length of an admissible (piecewise smooth) curve $\gamma : [a, b] \to M$ is defined as

$$L_g\gamma = \int_a^b |\gamma'(t)|_g \, dt$$

The integrand is bounded and continuous everywhere on $[a, b]$ except for possibly finitely many points where γ is not smooth, so the integral is well defined. Now for each pair of points $p, q \in M$, we define the **Riemannian distance from** p **to** q, denoted by $d_g(p, q)$, to be the infimum of the lengths of all admissible curves from p to q:

$$d_g(p, q) = \inf \left\{ L_g(\gamma) \middle| \gamma : [a, b] \xrightarrow{p.w.C^\infty} M; \gamma(a) = p; \gamma(b) = q \right\}$$

By path-connectedness of the connected manifold and compactness of $[a, b]$, it can be shown that any two points of M can be joined by an admissible curve, guaranteeing that $d_g(p, q)$ is a well defined nonnegative real number for each $p, q \in M$. We shall assume our Riemannian manifold is connected from now on. We comes to the theorem promised at the beginning:

Theorem 2.8 (Riemannian Manifolds as Metric Spaces; see [16] Thm 13.29)**.**
Let (M, g) be a connected Riemannian manifold. With the Riemannian distance function, M is a metric space whose metric topology is the same as the original manifold topology.

It is worth-noting that if the connected Riemannian manifold (M, g) as a metric space is complete then it is a Polish space because a manifold is by definition second countable and thus separable. A well-known theorem named after Hopf and his student Rinow claims that metric completeness is equivalent of geodesic completeness.

Now, (M, g) is a metric space. We can also give the space a Riemannian measure from which one gets other measures on M. While abstract integration of function is defined from measure by approximating arbitrary measurable function step by step, we work backwards defining Riemannian measure from integration by generalizing local definition to global version. That's because we already have Lebesgue and Riemann's integration on U open in $\mathbb{R}^n$, which can be "pushforward" to the smooth chart homeomorphic to U. We then get a definition of integration of n-form compactly supported on M again thanks to partition of unity. This gives rise to the integration of **Riemannian volume form** ω_g and functions on M. A local formula of ω_g is $\omega_g = \sqrt{\det(g_{ij})}dx^1 \wedge \cdots \wedge dx^n$. Now suppose $f : M \to \mathbb{R}^n$ is compactly supported. Then $f\omega_g$ is a compactly supported n-form and the integral $\int_M f\omega_g$ is well define. We also use dV_g to denoted ω_g. When $f = 1$, we have $\text{vol}(M) := \int_M dV_g$, called **volumne** of M. Thus, $d$vol is also used. In general, oriented or not, we have Riemannian density μ_g (its relationship with ω_g is $\mu_g = |\omega_g|$ and $\int_M f\mu_g = \int_M f\omega_g$ when M is oriented). The density then gives a measure. Due to [16] Corollary 12.28 we can get a local formula for this:

Definition 2.9. Let (ψ_α) be a partition of unity subordinate to an atlas $(U_\alpha, \varphi_\alpha = (x^i_\alpha))$ of M, then we set

$$\text{vol}(A) = \sum_\alpha \int_{\varphi_\alpha(A \cap U_\alpha)} \left(\psi_\alpha \sqrt{\det(G_\alpha)}\right) \circ \varphi_\alpha^{-1} dx^1_\alpha \cdots dx^n_\alpha$$

as the **Riemannian measure** of A whenever each integral exists in the Lebesgue sense. ♦

Remark 2.10.

(i) It is standard to check independency of this definition upon the choices of the atlas and partition of 1.

(ii) In the case the sum diverges, we say that the volume of A is infinite.

As usual, for any $1 \leq p < 1$ one can define the L_p norm on C_c^∞ by $\|f\|_{L^p} := \left(\int_M |f|^p \, d\text{vol}\right)^{\frac{1}{p}}$. The completion of C_c^∞ under L^p norm is denoted as $L^p(M)$. One also defines $L^\infty(M)$ analogous to that in measure theory. For $p = 2$, there is a natural inner product making it a Hilbert space, which is largely considered in spectral geometry.

Equipped with suitable measure on (M, g), we now see an example of Markov chain on M.

Example 2.11 ([23] Example 4 and Example 7)**.** For a metric measure space (mm-space) (X, d, μ), one can define an ε-random walk for some $\varepsilon > 0$ if we assume measure of balls $\mu(B_\varepsilon(x))$ to be finite and Supp$\mu = X$: the random walker jumps impartially in the ball $B_\varepsilon(x)$, with probability proportional to μ:

$$m_x = \frac{\mu|_{B_\varepsilon(x)}}{\mu(B_\varepsilon(x))} \tag{4}$$

Let (M^n, g) be a connected complete Riemannian manifold (note: we often use upper index on manifold M to indicate the dimension). With Riemannian measure vol and length metric d_g, we can define a similar

random walk m^ε for the mm-space (M, d_g, vol) via the Markov kernel

$$m_x^\varepsilon(dy) = \begin{cases} \frac{1}{\text{vol}(B_\varepsilon(x))}\text{vol}(dy), & \text{if } y \in B_\varepsilon(x), \\ 0, & \text{if } y \notin B_\varepsilon(x). \end{cases}$$

Recalling section 1, we then have measure m_x^ε as an indefinite integral of $\frac{\mathbb{1}_{B_\varepsilon(x)}}{\text{vol}(B_\varepsilon(x))}$ with respect to measure vol. Namely,

$$m_x^\varepsilon(A) = \int_A \frac{\mathbb{1}_{B_\varepsilon(x)}}{\text{vol}(B_\varepsilon(x))}\,\text{vol}(dy) = \frac{\text{vol}(B_\varepsilon(x) \cap A)}{\text{vol}(B_\varepsilon(x))}, \quad A \in \mathcal{B}(M) \tag{5}$$

This is exactly the same as equation (4). $\Diamond$

3 Curvature

We now come to the core of this note. Equipped with classical settings in Riemannian manifold in subsection 3.1, we seek a discrete version of curvature of Ollivier-Ricci type as we emphasized in the introduction.

3.1 Connection and Curvature on Riemannian Manifold

Let (M, g) be a Riemannian manifold and ∇ be the unique metric connection on TM (i.e., $\forall X, Y, Z \in \mathfrak{X}(M) : \nabla_X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$) that is also symmetric (i.e., $\forall X, Y \in \mathfrak{X}(M), \nabla_X Y - \nabla_Y X = [X, Y]$), called **Levi-Civita connection**. There are other formulae of ∇ suitable for different cases and various properties of ∇ (see [17]).

A **covariant derivative of C^∞ vector field V along** γ is a map $D_t : \mathfrak{X}(\gamma) \to \mathfrak{X}(\gamma)$ that satisfies linearity on $\mathbb{R}$, product rule with $f \in C^\infty(I)$, and $(D_tV)(t) = \nabla_{\gamma'(t)}\tilde{V}$ for an extension $\tilde{V}$ of V. The last rule connects the connection on TM with that along a curve and needs careful interpretation. A useful formula is

$$(D_tV)(t) = \left(\dot{V}^k(t) + \dot{\gamma}^i(t)V^j(t)\Gamma_{ij}^k(\gamma(t))\right)\partial_k|_{\gamma(t)}$$

A smooth curve γ is called a **geodesic** if it keeps a zero acceleration, $D_t\gamma' = 0$. Utilizing the above formula, one gets the geodesic equation and solves it by considering corresponding curve in the local trivialization of TM (see [9] p.62), given initial position $\gamma(0)$ and velocity $\dot{\gamma}(0)$.

A smooth vector field V along a smooth curve γ is said to be **parallel along** γ (with respect to ∇) if $D_tV = 0$. Thus a geodesic can be characterized as a curve whose velocity vector field is parallel along the curve. Visually, each v_p is "parallel" with each other. With similar trick in solving ODEs, we see a vector $v \in T_{\gamma(t_0)}M$ uniquely defines a **parallel transport** V such that $V(t_0) = v$. This allows us to correspond a tangent vector on a point to a tangent vector on another point by parallelism.

Let γ_v denote the unique maximal geodesic (i.e., the domain of the geodesic cannot be extended any further) issuing from point $p = \gamma(0)$ with velocity $v = \gamma'(0)$. Define

$$\mathcal{D} = \coprod_{p\in M} \mathcal{D}_p = \coprod_{p\in M} \{v_p \in T_pM : [0, 1] \subseteq \text{Dom}(\gamma_v)\} \subseteq TM$$

and the **exponential map** $\exp : \mathcal{D} \to M;\ v \mapsto \gamma_v(1)$ with its restriction $\exp_p := \exp|_{\mathcal{D}_p}$.

The **Riemann curvature endomorphism** is the map

$$\begin{aligned} R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) &\to \mathfrak{X}(M) \\ (X, Y, Z) &\mapsto R(X, Y)Z := \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z \end{aligned}$$

which defines a $(1, 3)$-tensor field, still denoted as R, by tensor characterization lemma ([17] Lemma B.6). In local coordinates, R is written as $R = R_{ijk}{}^l dx^i \otimes dx^j \otimes dx^k \otimes \partial_l$, where the coefficients $R_{ijk}{}^l$ are defined

by $R(\partial_i, \partial_j)\partial_k = R_{ijk}{}^{l}\partial_l$. We also define the **Riemann curvature tensor** as the covaraint 4-tensor field $Rm = R^\flat$ obtained from the $(1,3)$-tensor field R by lowering the last index. Its action on vector fields is given by $Rm(X,Y,Z,W) = \langle R(X,Y)Z, W\rangle$ and in coordinates $Rm = R_{ijkl}dx^i \otimes dx^j \otimes dx^k \otimes dx^l$ where $R_{ijkl} = g_{lm}R_{ijk}{}^{m}$.

The **Ricci curvature** Ric is the covaraint 2-tensor field defined as the trace of R on its first and last indices. The components of Rc are usually denoted as R_{ij}, so that

$$R_{ij} = (Rc)_{ij} := (\text{tr}R)_{ij} = R_{kij}{}^{k} = g^{km}R_{kijm}$$

The **scalar curvature** is the function S defined as the trace of Rc:

$$S := \text{tr}Rc = R_i^i = g^{ij}R_{ij}$$

The following example may help us have a look of more variants of curvature on M.

Example 3.1 ([23] Proposition 6)**.** Let (M, g) be a complete Riemannian manifold with length metric d_g and $v, w \in U_pM$. Let $\varepsilon, \delta > 0$. Denote $q = \exp_p(\delta v) = \gamma_v(\delta)$, and we parallel-transport w at p to w' at q along geodesic $\exp_x(tv) = \gamma_v(t)$. Then,

$$d_g(\exp_p(\varepsilon w), \exp_q(\varepsilon w')) = d_g(\gamma_w(\varepsilon), \gamma_{w'}(\varepsilon)) = \delta\left(1 - \frac{\varepsilon^2}{2}K(v,w) + O(\varepsilon^3 + \varepsilon^2\delta)\right)$$

as $(\varepsilon, \delta) \to 0$. Here $K(v,w)$ is the sectional curvature of the 2-plane $\Pi = (v,w) \subset T_pM$, i.e., the Gaussian curvature of the surface S_Π at p with induced metric ($S_\Pi := \exp_p(\Pi \cap \mathcal{V})$ for any $\mathcal{V} \subset T_pM$ that is an neighborhood of zero on which $\exp_p$ is a diffeomorphism). $\lozenge$

Remark 3.2. The proof is given in [23] Section 8. Gaussian curvaute (and mean curvature) is built upon principal curvatures, or the eigenvalues of the shape operator s, a self-adjoint linear operator on finite dimensional T_pM. These curvatures enjoy concrete geometric interpretations and requires some reading in Riemannian submanifold.

3.2 Ollivier-Ricci Curvature

Here is a visualization of geometric meaing of Ricci curvature by which Ollivier used to develop his definition. We can imagine two walkers starting from x and y in the space X, and they are having a running match. Their lines should be parallel and straight enough to be fair and easy to compare. This means the two walkers are heading with the same direction and walk on the straight line starting with that direction. Mathematically speaking, we have a tangent vector v at x and another v' on y parallelly transported from v. They are also on the geodesics with initial positions and velocities specified accordingly. Now, if they are closer to each other as time goes, it indicates that the curvature is positive (think about starting points on equator of the sphere and longitudes as geodesics). This measurement of dilation allows us to consider the following generalization.

Definition 3.3 (Coarse Ricci Curvature; see [23] Definition 3)**.** Let (X, d) be a metric space with a random walk $m = (m_x(\cdot))_{x\in X}$. Let $x, y \in X$ be two distinct points. The coarse Ricci curvature of (X, d, m) along (xy) is

$$\kappa(x,y) = 1 - \frac{W(m_x, m_y)}{d(x,y)}$$

We see some examples illustrating the use of optimal transportation and coarse ricci curvature.

Example 3.4 (follows Example 2.11)**.** Two resutls connecting the old and new definitions of Ricci curvature are given:

- [26] Theorem 3 (xii) shows that the normalized Riemannian uniform distribution on balls

$$m_x^\varepsilon(A) = \frac{\text{vol}(B_\varepsilon(x) \cap A)}{\text{vol}(B_\varepsilon(x))}, \quad A \in \mathcal{B}(M)$$

in Example 2.11 satisfies the asympototic estimate

$$\frac{W(m_x^\varepsilon, m_y^\varepsilon)}{d_g(x,y)} \leq 1 - \frac{K}{2(n+2)}\varepsilon^2 + o(\varepsilon^2)$$

if and only if $Rc \geq K$ for some $K \in \mathbb{R}$, where $Rc \geq K$ is a shorthand for $Rc \geq Kg$ (recall that both Rc and g are covariant 2-tensor fields). Equivalently, $Rc(v,v) \geq K$ for each unit tangent vector $v \in U_pM$.

- [23] Section 8 proves that for a unit tangent vector v at point p and a point y on the maximal geodesic γ_v (guaranteed by for example [17] Corollary 4.28), one has

$$\kappa(x,y) = \frac{\varepsilon^2 Rc(v,v)}{2(n+2)} + O\left(\varepsilon^2 + \varepsilon^2 d_g(x,y)\right)$$

with suffciently small $d_g(x,y)$. Perhaps a sketch of proof will be better, because it involves many of the concepts in Riemannian geometry we defined in section 3.1 $\lozenge$

Now we comes to Lin, Lu, and Yau's modified version of Ollivier's definition on grpah. [18] They considered a particular Markov chain on the graph, tuned with a parameter, and obtained the curvature on graph by limiting the parameter.

3.3 Lin-Lu-Yau Curvature

As in Example 1.7, we let $G = (V, E)$ be a simple connected graph with countable vertex set V and edge set E, equipped with metric d counting edges in geodesic. Consider Definition 3.3 and Example 2.11 again and see the following correspondences:

- ε-ball $B_\varepsilon(x) = \{y \in X : d_g(x,y) \leq \varepsilon\} \longleftrightarrow B(x) = B_1(x) = \{y \in V : d(x,y) \leq 1\}$. This is the same as $N(x) = \Gamma(x) \cup \{x\}$ where $\Gamma(x) = \{y \in V : (xy) \in E\}$ in [18]'s notation.
- random walk on Riemannian manifold (or a metric space) $m = (m_x(\cdot))_{x \in X} \longleftrightarrow \alpha$-parameterized random walk on graph $m = (m_x^\alpha(\cdot))_{x \in V}$, where the probability measures m_x^α in the sense of Example 1.3 are defined by [18] below.

$$v \in V : \ m_x^\alpha(v) = \begin{cases} \alpha & \text{if } v = x, \\ \frac{1-\alpha}{\deg_x} & \text{if } v \in \Gamma(x), \\ 0 & \text{otherwise.} \end{cases}$$

where $\deg_x$ is the degree of x, $\deg_x = |\Gamma(x)|$. Since α is the probability that the random walker will stay at vertex x, we call α the **idleness parameter**. This is shown in Figure 5.

Then we have the following notion of curvature.

Definition 3.5 (Lin-Lu-Yau curvature)**.** For any $x, y \in V$, the α**-curvature** κ_α is given by

$$\kappa_\alpha(x,y) = 1 - \frac{W(m_x^\alpha, m_y^\alpha)}{d(x,y)}$$

where W is the Wasserstein distance in Definition 1.8. Due to [18] Lemma 2.1 on concavity of $\kappa_\alpha(x,y)$ with respect to α and fixed x, y , we arrive at the **Lin-Lu-Yau curvature**

$$\kappa(x,y) = \lim_{\alpha \to 1} \frac{\kappa_\alpha(x,y)}{1-\alpha}. \quad \blacklozenge$$

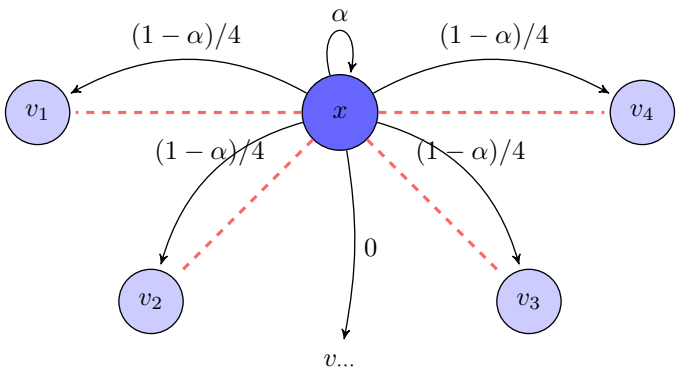

Figure 5: An example of a probability measure on a 4-degree vertex x with idleness parameter α. Red dashed lines stand for the edges. Black arrows stand for the Markov chain diagram.

Two remarks are given: (1) If we require our graph $G = (V, E)$ to be a connected locally finite graph (i.e. each $B(x)$ is a finite set), then the vertex set V is automatically a countable set. (2) Note that when the random walker is completely idle, namely $\alpha = 1$, the meausres m_x^α and m_y^α become Dirac measures δ_x and δ_y, and

$$W(m_x^1, m_y^1) = W(\delta_x, \delta_y) = d(x, y)$$

via the transport plan $\pi(v, w) = \delta_x(v)\delta_y(w)$, i.e., directly sending 1 on x to 1 on y. Thus, $\kappa_1(x, y) = 0$ for any $x, y \in V$.

Example 3.6 (Complete graphs K_n)**.** We continue using the 4-degree vertex example in figure 5 for an illustration: a graph K_n with $n = 5$ is shown in Figure 6. Due to the symmetry, it suffices to analyze any (xy). In fact, any pair of vertices are connected. As we shall see immediately, this strong connectivity reduces our cost of transportation significantly.

The left graph shows the distribution of masses for vertex x and the right graph shows the distirbution of masses for vertex y. To redistribute the masses, one observes that only masses on x and y are different, so intuitively we don't bother planning to move the masses on v_2, v_3, v_4. The remaining is just to move $\frac{5\alpha-1}{4}$ $(= \alpha - \frac{1-\alpha}{4})$ from x to y. This plan is then specified by

$$\begin{aligned}
&\pi(x, x) = \frac{1-\alpha}{4}, \pi(x, y) = \frac{5\alpha-1}{4}, \pi(x, \text{any other vertex}) = 0 \\
&\pi(y, y) = \frac{1-\alpha}{4}, \pi(y, \text{any vertex}) = 0 \\
&\pi(v_i, v_i) = \frac{1-\alpha}{4}, \pi(v_i, \text{any other vertex}) = 0 \quad \forall i = 2, 3, 4
\end{aligned}$$

and the total cost is

$$\text{cost}(\pi) = \frac{5\alpha-1}{4} \cdot 1 + \frac{1-\alpha}{4} \cdot 0 + \frac{1-\alpha}{4} \cdot 0 + \frac{1-\alpha}{4} \cdot 0 = \frac{5\alpha-1}{4} \tag{6}$$

It is intuitive that this plan is the optimal, but a rigorous argument needs a duality trick that will be discussed in the next subsection. We thus compute the curvature

$$\kappa_\alpha(x, y) = 1 - \frac{\text{cost}(\pi)}{d(x, y)} = \frac{5(1-\alpha)}{4}$$

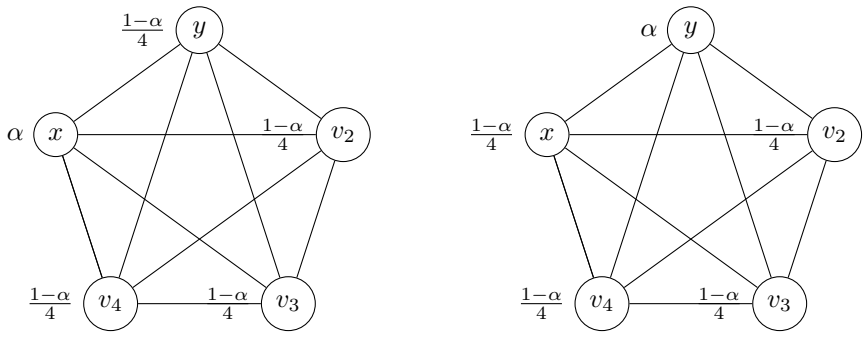

Figure 6: Complete graphs with probability distribution labelled.

It is easy to see that the general complete graph K_n has $\kappa_\alpha(x,y) = \frac{n(1-\alpha)}{n-1}$. In particular $\alpha = 0$ gives the curvature $\frac{n}{n-1}$, which is [18] example 1. In fact, this example hints us to consider some symmetries to put on the graph to ease computation. It turns out that Cayley grpah is one of the way to realize vertex-transitivity, a property among the various symmetries. $\lozenge$

3.4 Computational Aspects of Discrete Curvature

Definition 1.8 can be made into the following suitable form for programming:

Let $G = (V, E)$ be a graph, assumed to be locally finite but need not to be finite. We can label V as $\{1, 2, 3, \cdots\}$. Let the cost matrix be the distance matrix $\mathcal{D}$, that is, the cost of transportation from vertex i to vertex j is equal to $d(i, j)$:

$$\mathcal{D} = \begin{pmatrix} 0 & d(1,2) & \cdots & d(1,n) & \cdots \\ d(2,1) & 0 & \cdots & d(2,n) & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ d(n,1) & d(n,2) & \cdots & 0 & \vdots \\ \vdots & \vdots & \cdots & \vdots & \ddots \end{pmatrix}$$

Let $x, y \in V$. The measures m_x and m_y on V are determined by two vectors $(m_x(i))_{i\in\mathbb{N}}, (m_y(j))_{i\in\mathbb{N}}$. Notice that m_x and m_y are only supported on $B(x)$ and $B(y)$ with $|B(x)| = k$ and $|B(y)| = l$ (because the graph is locally finite), so they have $k-1$ and $l-1$ neighborhoods respectively. In case of a Cayley graph with a finite symmetric set, which is a regular graph, we have $k = l$. Then $B(x)$ and $B(y)$ are labelled as subsequence of V, i.e., $(x_i) = \{x_1, x_2, \cdots, x_k\}$ and $(y_j) = \{y_1, y_2, \cdots, y_l\}$, where we can assume $x_1 = x$ and $y_1 = y$. There is no need to store the zeros on the non-neighborhoods, and we define

$$\mathbf{x} = (m_x(x_i))_{x_i \in B(x)}, \mathbf{y} = (m_y(y_j))_{y_j \in B(y)}$$

The transport plan $\pi : V \times V \to [0, 1]$ corresponds each pair of $x_i \in B(x)$ and $y_j \in B(y)$ with a value in $[0, 1]$, so it can be regarded as a $k \times l$-matrix $\mathbf{P}$ where $\mathbf{P}_{ij} = \pi(x_i, y_j)$. Then the set of all plans $\Pi(\mu, \nu)$ is

$$\Pi(m_x, m_y) = \Pi(\mathbf{x}, \mathbf{y}) = \left\{\mathbf{P} \in M_{k\times l}(\mathbb{R}) : \mathbf{P}\mathbb{1}_l = \mathbf{x}, \mathbf{P}^T\mathbb{1}_k = \mathbf{y}\right\}$$

where $\mathbb{1}_l$ is the column vector with l ones, $\mathbb{1}_k$ with k ones. We let $\mathbf{D}$ be the submatrix of $\mathcal{D}$ with rows $x_i \in B(x)$ and columns $y_j \in B(y)$. Then the Kantorovich problem, i.e., finding the Wasserstein distance, is presented as

$$W(\mathbf{x}, \mathbf{y}) = \min_{\mathbf{P}\in\Pi(\mathbf{x},\mathbf{y})} \langle \mathbf{D}, \mathbf{P}\rangle_F = \min_{\mathbf{P}\in\Pi(\mathbf{x},\mathbf{y})} \sum_{i,j} \mathbf{D}_{ij}\mathbf{P}_{ij} \tag{7}$$

where $\langle \mathbf{D}, \mathbf{P} \rangle_F$ is the Frobenius norm of the matrix $\mathbf{D}$ and $\mathbf{P}$. Problem (7) is a linear optimization problem with $k \times l$ decision variables and $k + l$ equality constraints. In particular, Problem (7) can be seen as an uncapacitated network flow problem (NFP) on a directed complete $k \times l$ bipartite graph $K_{k,l}$, that is, the vertices in $B(x)$ are not adjacent to each other in $K_{k,l}$ but is adjacent to every other nodes in $K_{k,l}$, and the same is for vertices in $B(y)$.

The NFP aims to minimize the total cost of flow $\langle \mathbf{D}, \mathbf{P} \rangle_F$. The decision variables $\mathbf{P}_{ij} = \pi(x_i, y_j)$ are called **flow variables** as they represent the amount of flow of sands through each edge of the graph $K_{k,l}$ from a vertex x_i in $B(x)$ to a vertex y_j in $B(y)$. Each $d(x_i, y_j)$ is now the cost of a unit of flow from a vertex x_i to a vertex y_j. Note that each node (vertex) has a numerical information b_i representating the supply (if $b_i > 0$) or demand (if $b_i < 0$) of sands. Nodes supplying sands will be called **sources**; nodes demanding sands will be called **sinks**. See Figure 7 for such a bipartite representation of Kantorovich problem.

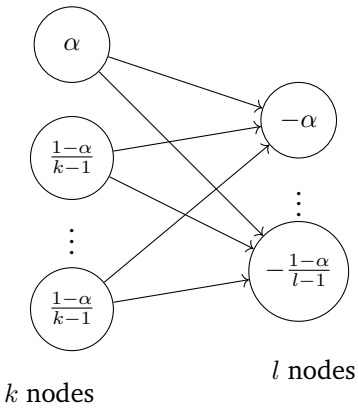

Figure 7: Bipartite-graphical representation of Kantorovich problem

We call the above graph $G' = (\mathcal{N}, \mathcal{A})$ the **derived graph** for Problem (7) associated to vertices x and y, where

$$\mathcal{N} = \{v_1(= x_1 = x), \cdots, v_k(= x_k), v_{k+1}(= y_1 = y), \cdots, v_{k+l}(= y_l)\}$$

and

$$\mathcal{A} = \{(v_1, v_{k+1}), \cdots, (v_1, v_{k+l}); (v_2, v_{k+1}), \cdots, (v_2, v_{k+l}); \cdots; (v_k, v_{k+1}), \cdots, (v_k, v_{k+l})\}$$

or simply

$$\mathcal{A} = \{(1, k+1), \cdots, (1, k+l); (2, k+1), \cdots, (2, k+l); \cdots; (k, k+1), \cdots, (k, k+l)\}$$

To write problem (7) in standard form of linear optimization and specifically NFP, we need to vectorize the matrix-valued problem, following notations used in [7] Section 7.2. We align the decision variables $\mathbf{P}_{ij}$ in the matrix $\mathbf{P}$ with the ordering such that the **vector of flows f** is the following column vector

$$\mathbf{f} = [P_{11}, \cdots, P_{1l}; P_{21}, \cdots, P_{2l}; \cdots; P_{k1}, \cdots, P_{kl}]^T.$$

Using indices in $\mathcal{A}$, we have

$$\mathbf{f} = \left[f_{v_1,v_{k+1}}, \cdots, f_{v_1,v_{k+l}}; f_{v_2,v_{k+1}}, \cdots, f_{v_2,v_{k+l}}; \cdots; f_{v_k,v_{k+1}}, \cdots, f_{v_k,v_{k+l}}\right]^T.$$

or simply

$$\mathbf{f} = [f_{1,k+1}, \cdots, f_{1,k+l}; f_{2,k+1}, \cdots, f_{2,k+l}; \cdots; f_{k,k+1}, \cdots, f_{k,k+l}]^T.$$

Do the same for the matrix $\mathbf{D}$ to get a column vector

$$\begin{aligned}
\mathbf{c} &= [D_{11}, \cdots, D_{1l}; D_{21}, \cdots, D_{2l}; \cdots; D_{k1}, \cdots, D_{kl}]^T \\
&= [d(v_1, v_{k+1}), \cdots, d(v_1, v_{k+l}); d(v_2, v_{k+1}), \cdots, d(v_2, v_{k+l}); \cdots; d(v_k, v_{k+1}), \cdots, d(v_k, v_{k+l})]^T \\
&= [d(x_1, y_1), \cdots, d(x_1, y_l); d(x_2, y_1), \cdots, d(x_2, y_l); \cdots; d(x_k, y_1), \cdots, d(x_k, y_l)]^T
\end{aligned}$$

Consider the $(k+l)\times kl$-dimensional **incidence matrix** $\mathbf{A}$ of G', where each entry $\mathbf{A}_{ij}$ is associated with the i-th node and j-th edge of the derived graph, a directed graph, and is defined as

$$\mathbf{A}_{ij}=\begin{cases}1, \text{ if } i \text{ is the start node of the } j \text{ th edge}\\ -1, \text{ if } i \text{ is the end node of the } j \text{ th edge}\\ 0, \text{ otherwise.}\end{cases}$$

It is easy to see that j-th edge is $\left(\lceil\frac{j}{l}\rceil, k+j-l\lfloor\frac{j}{l}\rfloor\right)$ in $\mathcal{A}$, so we know exactly how the incidence matrix $\mathbf{A}$ looks like by only numerical inputs: $\mathbf{A}_{ij}$ is 1 iff $i=\lceil\frac{j}{l}\rceil$ and is -1 iff $i=k+j-l\lfloor\frac{j}{l}\rfloor$. Conversely, the edge $(i,j)\in\mathcal{A}$ is the $l(i-1)+j-k$-th edge. The matrix has the following appearence

$$\mathbf{A}=\begin{array}{c} \\ v_1\\ v_2\\ \vdots\\ v_k\\ v_{k+1}\\ v_{k+2}\\ \vdots\\ v_{k+l}\end{array}\begin{array}{c}\begin{array}{cccccccccc}(1,k+1) & (1,k+2) & \cdots & (1,k+l); & \cdots & ;(k,k+1) & (k,k+1) & \cdots & (k,k+l)\end{array}\\ \begin{bmatrix}1 & 1 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & \cdots & 1 & 1 & \cdots & 1\\ -1 & 0 & \cdots & 0 & \cdots & -1 & 0 & \cdots & 0\\ 0 & -1 & \cdots & 0 & \cdots & 0 & -1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & -1 & \cdots & 0 & 0 & \cdots & -1\end{bmatrix}\end{array}$$

$$=\begin{array}{c} \\ v_1\cdots v_k\\ v_{k+1}\cdots v_{k+l}\end{array}\begin{array}{c}\begin{array}{cccc}(1,k+1)\cdots(1,k+l) & (2,k+1)\cdots(2,k+l) & \cdots & (k,k+1)\cdots(k,k+l)\end{array}\\ \begin{bmatrix}\mathbf{1}_1 & \mathbf{1}_2 & \cdots & \mathbf{1}_k\\ -\mathbf{I}_l & -\mathbf{I}_l & \cdots & -\mathbf{I}_l\end{bmatrix}\end{array}$$

where $\mathbf{1}_i$ is the $k\times l$ matirx with i-th row all ones and all other entries zeros. The sparsity of matrix A and its block form can facilitate implementation of optimization algorithms.

Denoting i-th row vector of $\mathbf{A}$ as $\mathbf{a}_i$ and the sets of all vertices of edges outgoing from v_i (resp. incoming to v_i) as $O(i)$ (resp. $I(i)$), we see that the general constraints in NFP

$$\begin{aligned}i\in[\![k+l]\!]:\quad & b_i+\sum_{j\in I(i)}\mathbf{f}_{ji}=\sum_{j\in O(i)}\mathbf{f}_{ij}\\ \Rightarrow & \mathbf{a}_i\mathbf{f}=\sum_{j\in O(i)}\mathbf{f}_{ij}-\sum_{j\in I(i)}\mathbf{f}_{ji}=b_i=\begin{cases}m_x^\alpha(v_i), & 1\leqslant i\leqslant k\\ -m_y^\alpha(v_i), & k+1\leqslant i\leqslant k+l\end{cases}\end{aligned}\tag{8}$$

read as

$$m_x^\alpha(v_i)=m_x^\alpha(x_i)=b_i=\mathbf{a}_i\mathbf{f}=\sum_{j\in O(i)}\mathbf{f}_{ij}=\sum_{j=k+1}^{k+l}\mathbf{f}_{ij},\quad i=1,\cdots,k$$

$$-m_y^\alpha(v_i)=-m_y^\alpha(y_{i-k})=b_i=\mathbf{a}_i\mathbf{f}=-\sum_{j\in I(i)}\mathbf{f}_{ji}=-\sum_{j=1}^{k}\mathbf{f}_{ji},\quad i=k+1,\cdots,k+l$$

which are exactly the constraints specified by the coupling measure, i.e., $\mathbf{P}\mathbb{1}_l=\mathbf{u},\mathbf{P}^T\mathbb{1}_k=\mathbf{v}$.

Therefore, the Kantorovich problem of finding an optimal transport from probability measure m_x^α to probability measure m_y^α is equivalent to the NFP in the following standard form

$$\begin{aligned}\min\quad & \mathbf{c}^T\mathbf{f}\\ \text{subject to}\quad & \mathbf{A}\mathbf{f}=\mathbf{b}\\ & \mathbf{f}\geq\mathbf{0}\end{aligned}\tag{9}$$

where $\mathbf{f} \geq 0$ is exerted due to nonnegativity of the coupling measure. The dual form is given by

$$\begin{aligned} \max \quad & \mathbf{p}^T \mathbf{b} \\ \text{subject to} \quad & \mathbf{p}^T \mathbf{A} \leq \mathbf{c}^T \end{aligned} \tag{10}$$

where $\mathbf{p} = [p_1, \cdots, p_{k+l}]^T$ is a $(k+l) \times 1$ vector free of constraint. By the structure of incidence matrix A, the constraints $\mathbf{p}^T \mathbf{A} \leq \mathbf{c}^T$ read as $\forall (i,j) \in \mathcal{A} : p_i - p_j \leq d(v_i, v_j) = c_{l(i-1)+j-k}$. The strong duality result (see for example [7] Theorem 4.4) ensures that the optimal value are the same when one of the primal and dual has an optimal solution. Observing that $\mathbf{b} = \begin{bmatrix} \mathbf{x} \\ -\mathbf{y} \end{bmatrix}$, we can split $\mathbf{p}$ as $\begin{bmatrix} \mathbf{h} \\ \mathbf{g} \end{bmatrix}$ to get the following form

$$\max_{(\mathbf{h},\mathbf{g}) \in \mathbf{R}(\mathbf{D})} \langle \mathbf{h}, \mathbf{x} \rangle + \langle \mathbf{g}, -\mathbf{y} \rangle = \max_{(\mathbf{h},\mathbf{g}) \in \mathbf{R}(\mathbf{D})} \langle \mathbf{h}, \mathbf{x} \rangle - \langle \mathbf{g}, \mathbf{y} \rangle$$

where the set of admissible dual variables is

$$\mathbf{R}(\mathbf{D}) = \left\{ \mathbf{p} \in \mathbb{R}^{k+l} : \mathbf{p}^T \mathbf{A} \leq \mathbf{c}^T \right\} = \left\{ (\mathbf{h}, \mathbf{g}) \in \mathbb{R}^k \times \mathbb{R}^l : \forall (i,j) \in \mathcal{A}, h_i - g_j \leq \mathbf{D}_{ij} = d(v_i, v_j) \right\}$$

In fact, regarding $\mathbf{p} = \begin{bmatrix} \mathbf{h} \\ \mathbf{g} \end{bmatrix}$ as a function on the derived graph G', one can show that it induces a 1-Lipschitz function on G. The dual problem above is equivalent to the one formulated in theoretical literatures (e.g. [18],[20]) on optimal transport, i.e.,

$$W(\mu, \nu) = \sup_{f \in \mathrm{Lip}_1(G)} \sum_{x \in V} f(x)(\mu(x) - \nu(x)) \tag{11}$$

Example 3.7 (Continued on Example 3.6)**.** We already established the duality result and now we apply it to the complete graph in Example 3.6 with $n = 5$ and $\alpha = \frac{1}{4}$. Consider the following choice

$$\mathbf{h} = [1,0,0,0,0]^T, \mathbf{g} = [0,1,0,0,0]^T, \mathbf{p} = [1,0,0,0,0,0,1,0,0,0]^T$$

we get $\mathbf{p}^T \mathbf{b} = 1b_1 + 1b_7 = \frac{1}{4} + \left(-\frac{3}{16}\right) = \frac{1}{16} = 0.0625$. Note that $\mathbf{p}$ corresponds to 1-Lipschitz map f sending x to 1 leaving others 0 in problem (11). By weak duality ([7] Theorem 4.3), we get a lower bound $\frac{1}{16}$ of $\mathbf{c}^T \mathbf{f}$ in problem (9) or problem (7):

$$\mathbf{c}^T \mathbf{f} \geq \frac{1}{16}$$

However, by equation (6), we also have the upper bound

$$\mathbf{c}^T \mathbf{f} \leq \frac{\frac{5}{4} - 1}{4} = \frac{1}{16}$$

Therefore, by [7] Corollary 4.2, both problems attain their optimalities and the Wasserstein distance is

$$W(m_x^{1/4}, m_y^{1/4}) = \frac{1}{16}$$

This example summarizes several equivalent formulations of the optimization problem involved in finding the optimal transport and computing Wasserstein distance. $\diamondsuit$

We also observe from the above example that it is generally difficult to tackle the problem by a Trial-and-Error approach, i.e., guessing a transportation plan to obtain an upper bound and guessing a price vector $\mathbf{p}$ or a 1-Lipschitz map to obtain a lower bound with the hope that the two bounds coincide. When the size of the problem is large and complicated and involves many possibilities for guessing, this approach is infeasible. Therefore, we resort to computational programs. Below are three ways to get the same result in Example 3.7. We will discuss algorithmic improvements in the next section.

Example 3.8 (programs for Example 3.6 and Example 3.7)**.** We first build the incidence matrix **A**. Since both primal and dual are linear optimization problems, we can use the most general way simplex method, which is packaged in the function "linprog" of "scipy.optimize" library. The third method imports directly Wasserstein distance function from "ot" library (see POT: Python Optimal Transport).

```
def matrix_A(k, l):
  # Initial empty matrix
  A = np.zeros((k + l, k*l))
  # Fill the upper matrix
  for i in range(k):
      for j in range (l):
          A[i, i*l+j] = 1
  # Fill the lower matrix with block identity matrices
  for i in range(k):
      for j in range (l):
          A[k+j, i*l+j] = -1

  return A

#Solve Kantorovich problem of complete graph with n=5 and alpha=1/4
c = [1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1]
b = [1/4, 3/16, 3/16, 3/16, 3/16, -1/4, -3/16, -3/16, -3/16, -3/16]

# primal
result_primal = linprog(c, A_eq=matrix_A(5,5), b_eq=b, method="simplex")
# output:
# optimal_value = result_primal.fun = 0.0625
# optimal_solution = [0.0625 0.1875 0. 0. 0. 0. 0. 0.1875 0. 0. 0. 0. 0. 0.1875 0.
# 0. 0. 0. 0. 0.1875 0.1875 0. 0. 0. 0.]

#dual
neg_b = [-x for x in b]
result_dual = linprog(neg_b, A_ub=matrix_A(5,5).T, b_ub=c, method="simplex")
# output:
# optimal_value = result_primal.fun = -0.0625 (note that max(f)=-min(-f))
# optimal_solution = [1. 0. 0. 0. 0. 0. 1. 0. 0. 0.]

# ot library
a = [1/4, 3/16, 3/16, 3/16, 3/16] # distribution m^0.25_x
b = [1/4, 3/16, 3/16, 3/16, 3/16] # distribution m^0.25_y
M = [
[1, 0, 1, 1, 1],
[1, 1, 0, 1, 1],
[1, 1, 1, 0, 1],
[1, 1, 1, 1, 0],
[0, 1, 1, 1, 1]
]

Wd = ot.emd2(a, b, M)
# output: 0.0625
```

4 Cayley Graph and its Curvature

We will introduce Cayley graph in the next section and come back to the algorithmic discussion. We then choose certain groups and compute the curvature of their Cayley graphs and explain the patterns theoretically.

Definition 4.1 (Cayley graph)**.** Let G be a group. Let $S \subseteq G$ be a finite subset with $|S| = k$. We assume S generates G (see Remark 4.3 for the reason) It is called a **connection set** if it satisfies the following two properties:

- (exclusion of identity) $1 \notin S$;
- (symmetric property) element s is in S if and only if the inverse s^{-1} is in S (denoted as $S = S^{-1}$).

Then **Cayley graph** $\mathrm{Cay}(G; S)$ is a graph (V, E) with vertex set V and edge set E defined as follows:

- V contains exactly all elements of G;
- $E = \{(g, gs) | g \in G, s \in S\}$. ♦

Definition 4.2. An **automorphism** of a graph $\Gamma = (V, E)$ is a permutation (a bijection) $\phi : V \to V$ of the vertex set with the following property:

$$(*): \quad \forall v, w \in E,\ (\phi v, \phi w) \in E \iff (v, w) \in E.$$

Note that this definition works well for simple graphs and thus suffices for our discussion due to Remark 4.3. The graph $\Gamma = (V, E)$ is said to be **vertex-transitive** if the automorphism group $\mathrm{Aut}(\Gamma)$ acts transitively on vertex set E, i.e., the action only has one orbit (for any $x, y \in E$, there is some $\phi \in \mathrm{Aut}(\Gamma)$ such that $\phi \cdot x := \phi(x) = y$.) We also use $x \sim y$ to mean $(x, y) \in E$. ♦

Note that the set of all automorphisms $\mathrm{Aut}(\Gamma) := \{\phi \in S_V : (*) \text{ is satisfied}\}$ is a subgroup of S_V in an obvious manner: identity is in $\mathrm{Aut}(\Gamma)$; inverse of ϕ is in $\mathrm{Aut}(\Gamma)$ because $(\phi v, \phi w) \in E \Leftrightarrow (\phi^{-1}(\phi v), \phi^{-1}(\phi w)) = (v, w) \in E$; product is in $\mathrm{Aut}(\Gamma)$ becauese $(v \sim w) \in E \Leftrightarrow (\phi v, \phi w) \in E \Leftrightarrow (\varphi\phi v, \varphi\phi w) \in E$ and each of the three iterates over E making the other twos do so as well.

Remark 4.3 (Cayley graph is simple, regular, connected, and vertex-transitive)**.**

Since $g = (gs)s^{-1}$ for any $s \in S$, we see if there is an edge $(g, gs) \in E$, then there is also an edge $(gs, g) \in E$. They will be considered as a single edge instead. Then the Cayley graph is a simple graph: (1) Thanks to the above remark, there are no multiple edges connecting two vertices, because we also have $gs_1 = gs_2 \Rightarrow s_1 = s_2$; (2) There are no loops, because $gs = g \Rightarrow s = 1$ (but $1 \notin S$ by definition).

We also note that each vertex x in the Cayley graph has k edges connecting to other vertices, so $\mathrm{Cay}(G, S)$ with $|S| = k$ is also a k-regular graph.

Clearly, a Cayley graph is connected iff the connection set S is a set of generators of G, i.e.,

$$G = \langle S \rangle = \bigcap_{S \subseteq H \leq G} H = \{w = x_1^{e_1} x_2^{e_2} \dots x_n^{e_n} | x_i \in S, e_i = \pm 1, n \geq 1\}$$

Then Cayley graph $\Gamma = \mathrm{Cay}(G, S)$ is vertex-transitive because $\forall x, y \in E$ we have $L_{yx^{-1}}(x) = y$ where $L_g : g \mapsto gx$ is the left translation, and $L_g \in \mathrm{Aut(s)}$ (because $L_g(y) \sim L_g(x) \Leftrightarrow gy = (gx)s \Leftrightarrow y = xs \Leftrightarrow y \sim x$).

4.1 Examples of Groups and their Cayley Graphs

Example 4.4 (Grids)**.** The group is $G = (\mathbb{Z}^n, +)$ and set connection set S is

$$S = \left\{ (x_1, x_2, \cdots, x_n) \in \mathbb{Z}^n \middle| \sum_{i=1}^{n} |x_i| = 1 \right\},$$

that is, the set of points with exactly one nonzero component ± 1. In case $n = 1$, the set S degenerates to $\{-1, 1\}$. Thus, $(x, y) \in E$ if $x - y = 1$ or $x - y = -1$. Graph of group $\mathbb{Z}$ is shown in Figure 8.

Figure 8: graph $\mathbb{Z}$

When $n = 2$, the set S becomes $\{(0, -1), (0, 1), (-1, 0), (1, 0)\}$. We then have graph of $\mathbb{Z}^2$ in Figure 9.

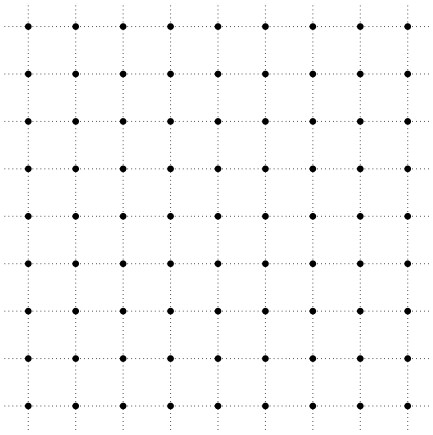

Figure 9: graph $\mathbb{Z}^2$

Example 4.5 (Cycles)**.** Let $G = \mathbb{Z}_n$ where $n > 2$, and let $S = \{\pm 1\}$. That is, each class $\overline{k} = \overline{0}, \overline{1}, \cdots, \overline{n-1}$ has two neighborhoods, $\overline{k-1}$ and $\overline{k+1}$. For example, $\overline{n-1}$ has the neighbors $\overline{n-2}$ and $\overline{n} = \overline{0}$. The graph $(\mathbb{Z}_n, S)$ is called the **n-cycle** and is denoted by C_n.

Example 4.6 (Hypercubes)**.** n-dimensional hypercube Q_n is the one-skeleton of the geometric n-dimensional cubes I^n in $\mathbb{R}^n$. Figure 10 gives Q_1, Q_2, Q_3, and Q_4. Q_n is the Cayley graph of the group $\mathbb{Z}_2^n$, the product

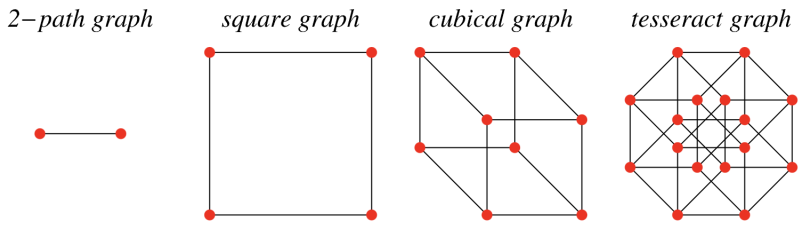

Figure 10: hypercubes taken from Wolfram.

of n copies of the cyclic group of order 2, with the connection set $S = \{e_i = (0, \cdots, 1, \cdots, 0) : i \in [\![n]\!]\}$. For example, in $(\mathbb{Z}_2)^2$, the generating set is $\{(1,0),(0,1)\}$, and in $(\mathbb{Z}_2)^3$, it is $\{(1,0,0),(0,1,0),(0,0,1)\}$, and so on.

4.2 Computation of Curvatures of Cayley Graphs

4.3 Symmetry, Diameter, and Curvature of Cayley Graphs

References

[1] Ambrosio, Luigi, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*, Springer Science+Business Media, 2005.

[2] Bakry, D. and Émery, M. *Diffusions hypercontractives*, Séminaire de Probabilités, XIX (1985). Lecture Notes in Math. 1123 177–206. Springer, Berlin.

[3] Bar-Natan, Assaf, Moon Duchin, and Robert Kropholler. *Conjugation Curvature for Cayley Graphs*, Journal of topology and analysis 14.02 (2022): 439-459.

[4] Baues, Oliver, and Norbert Peyerimhoff. *Geodesics in Non-positively Curved Plane Tessellations*, Advanced in Geometry 6(2) (2006): 243-263.

[5] Beineke, Lowell W., and Robin J. Wilson. *Topics in Algebraic Graph Theory*, Cambridge University Press, 2004.

[6] Benaïm, Michel, and Tobias Hurth. *Markov Chains on Metric Spaces: A Short Course*, Springer Nature Switzerland AG, 2022.

[7] Bertsimas, Dimitris, and John N. Tsitsiklis. *Introduction to Linear Optimization*, Vol. 6. Belmont, MA: Athena scientific, 1997.

[8] Cinlar, Erhan. *Probability and Stochastics*, Springer Science+Business Media, LLC, 2011.

[9] Carmo, Do. *Riemannian Geometry*, Birkhäuser Boston, 1992.

[10] Forman, Robin. *Bochner's Method for Cell Complexes and Combinatorial Ricci Curvature*, Discrete Comput Geom 29 (2003): 323–374.

[11] Grigor'yan, Alexander. *Introduction to Analysis on Graphs*, University Lecture Series (vol. 71), American Mathematical Society, 2018.

[12] Häaggström, Olle. *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, 2002.

[13] Higuchi, Y. *Combinatorial Curvature for Planar Graphs*, J. Graph Theory 38(4) (2001), 220-229.

[14] Jost, Jürgen, and Florentin Münch. *Characterizations of Forman curvature*, arXiv preprint arXiv:2110.04554 (2021).

[15] Sreejith, R. P., et al. *Forman Curvature for Complex Networks*, Journal of Statistical Mechanics: Theory and Experiment 2016.6 (2016): 063206.

[16] Lee, John. M. *Introduction to Smooth Manifolds*, second edition, Springer Science+Business Media New York, 2013.

[17] Lee, John. M. *Introduction to Riemannian Manifolds*, second edition, Springer International Publishing AG, 2018.

[18] Lin, Yong, Linyuan Lu, and Shing-Tung Yau. *Ricci Curvature on Graphs*, Tohoku Math. I. 63 (2011), 605-627, 2010.

[19] Lott, John, and Cédric Villani. *Ricci Curvature for Metric-Measure Spaces via Optimal Transport*, Annals of Mathematics (2009): 903-991.

[20] Mizukai, Iwao, and Akifumi Sako. *Ollivier Ricci Curvature of Cayley Graphs for Dihedral Groups, Generalized Quaternion Groups, and Cyclic Groups*, arXiv preprint arXiv:2210.00860 (2022).

[21] Najman, L, and P. Romon (editors), *Modern Approaches to Discrete Curvature*, Lecture Notes in Mathematics 2184, Springer, Cham, 2017.

[22] Norris, James. *Markov Chains*, Cambridge University Press, 1997.

[23] Ollivier, Yann. *Ricci Curvature of Markov Chains on Metric Spaces*, Journal of Functional Analysis 256.3 (2009): 810-864.

[24] Peyerimhoff, Norbert. *Curvature Notation on Graphs Leeds Summer School*, 2019. https://www.maths.dur.ac.uk/users/norbert.peyerimhoff/peyerimhoff-lecture-notes.pdf.

[25] Peyre, Gabriel, and Marco Cuturi. *Computational Optimal Transport*, arXiv preprint arXiv:1803.00567v4 (2020).

[26] Renesse, M.-K. von, and K.-T.Sturm. *Transport Inequalities, Gradient Estimates, Entropy and Ricci Curvature*, Comm. Pure Appl. Math. 58.7 (2005): 923-940.

[27] Santambrogio, Filippo. *Introduction to Optimal Transport Theory*, arXiv preprint arXiv:1009.3856 (2010).

[28] Schmuckenschläger, Michael. *Curvature of nonlocal Markov generators*, Convex Geometric Analysis (Berkeley, CA, 1996) 34 (1998): 189-197.

[29] Steinerberger, Stefan. *Curvature on Graphs Via Equilibrium Measures*, Journal of Graph Theory 103.3 (2023): 415-436.

[30] Sturm, K.-T. *On the Geometry of Metric Measure Spaces I and II*. Acta Math., 196 (2006), 65–131 and 133-177.

[31] Villani, Cedric. *Optimal Transport, Old and New*, Springer-Verlag Berlin Heidelberg, 2009.

[32] Villani, Cedric. *Topics in Optimal Transportation*, American Mathematical Society, 2003.

[33] Xu, Junming. *Topological Structure and Analysis of Interconnection Networks*, Kluwer Academic Publishers, 2001.