WASSERSTEIN-LIKE METRICS ON GRAPHS

VIDUSHI MITTAL*
UNIVERSITY OF CALIFORNIA, SANTA BARBARA
ADVISED BY: PHD CANDIDATE CLAIRE MURPHY AND PROFESSOR KATY CRAIG

May 9, 2025

Abstract. In the space of probability measures on $\mathbb{R}^d$ equipped with the 2-Wasserstein distance, heat flow is the gradient flow of the entropy functional. In this senior thesis, we will closely follow "Gradient flows of the entropy for finite Markov chains" by Jan Maas [3], in which Maas develops an analogue of the 2-Wasserstein distance on the space of probability measures on finite weighted graphs, describing how a new metric allows us to show that heat flow is the gradient flow of the entropy in the graph setting. We will focus primarily on the case of a two-node graph, where we will solve for more explicit formulas. We will first motivate our problem by introducing gradient flows on both Euclidean space and the space of probability measures. From there, we will give an introduction to Markov chains and their properties, which will provide the analogue of heat flow in the graph setting. We will then give the definition of a new metric $\mathcal{W}$ on the space of probability measures on a two-node graph, and prove the isometry of this metric to a subset of $\mathbb{R}$. Next, our isometry helps us develop intuition for geodesics and convexity on our metric space. We will finish by returning to our original motivation, showing how our new metric allows us to prove that heat flow is the gradient flow of the entropy.

1 Introduction to Gradient Flows

Let us first introduce gradient flows in Euclidean space, before generalizing this notion to any metric space. This will help us develop an idea of what a gradient flow might be with respect to the metric we will introduce in Section 3.

DEFINITION 1.1 (Gradient Flow). *Given a function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}$, which is smooth, and a point $x_0 \in \mathbb{R}^n$, a gradient flow is defined as a curve $x(t)$ with starting point $x(0) = x_0$, that moves by following the path that makes $\mathcal{F}$ decrease as steeply as possible, that is, it is the unique solution to the Cauchy problem:*

$$\begin{cases} x'(t) = -\nabla \mathcal{F}(x(t)) \\ x(0) = x_0. \end{cases}$$

When generalizing this to a metric space (X, d), we lack a clear definition of gradient flow and instead turn to a limiting process to define gradient flow. Intuitively, we begin at a point $x_0 \in X$, and at this point, we determine the direction of steepest descent using our metric d, move a small step in this direction, and repeat from there. As the size of the steps goes to zero, this converges to the gradient flow.

Let us introduce a well-known example, the space of probability measures on $\mathbb{R}^d$ with finite second moment. In other words, our space is given by:

$$\mathcal{P}_2(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^2 d\mu < \infty \right\}.$$

We can equip this space with a metric called the 2-Wasserstein distance, $\mathcal{W}_2$. In this metric space, we have a well-defined notion of gradient with respect to $\mathcal{W}_2$, denoted by $\nabla_{\mathcal{W}_2}$.

Given a functional $\mathcal{F}$ with "nice" properties, the JKO scheme [4] describes a process that converges to gradient flow in the space $(\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)$. Let us briefly describe this process more precisely. Given a functional $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, we consider some step size $\tau > 0$. We begin with initial value $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$. Now, given any μ_k^τ,

*VIDUSHIMITTAL15@GMAIL.COM

the next step of the path is given by

$$\mu_{k+1}^{\tau} = \underset{\mu \in \mathcal{P}_2(\mathbb{R}^d)}{\operatorname{argmin}} \mathcal{F}(\mu) + \frac{\mathcal{W}_2^2(\mu, \mu_k^{\tau})}{2\tau}.$$

We call the gradient flow of $\mathcal{F}$ the limit as $\tau \to 0$ of the interpolated paths given by each step size τ. Due to special properties of the Wasserstein metric, we are able to consider gradient flows hand in hand with certain PDEs: the solutions of certain PDEs are the gradient flows of a given functional and vice versa.

For readers familiar with geodesics, we introduce the following definition:

DEFINITION 1.2 (λ-Geodesically Convex in $(\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)$). *A functional $\mathcal{F}$ is λ-geodesically convex, if, given a constant speed geodesic $\mu(t)$ between points $\mu(0)$ and $\mu(1)$, there exists $\lambda > 0$ such that*

$$F(\mu(t)) \leq (1-t)\mathcal{F}(\mu(0)) + t\mathcal{F}(\mu(1)) - \lambda \frac{t(1-t)}{2} \mathcal{W}_2^2(\mu(0), \mu(1))$$

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

Say we have a functional $\mathcal{F}$. Let

$$\rho^* = \underset{\rho \in \mathcal{P}_2(\mathbb{R}^d)}{\operatorname{argmin}} \mathcal{F}(\rho).$$

Say we have gradient flow $\rho(t)$ such that $\rho(0) = \rho_0(\cdot)$. If $\mathcal{F}$ is λ-geodesically convex for $\lambda > 0$, then we have that

$$\mathcal{W}_2(\rho(t), \rho^*) \leq e^{-\lambda t} \mathcal{W}_2(\rho_0, \rho^*).$$

Intuitively, λ-convexity tells us the gradient flow approaches the minimizer quickly [4]. We will postpone our discussion of geodesics for Section 6, but the above will motivate the results of this coming section.

To introduce a major result, Jordan, Kinderlehrer, and Otto [2] showed that, given a gradient flow $\mu(t)$ for a point μ_0, if we define $\nabla_{\mathcal{W}_2} \mathcal{F}(\rho) = -\nabla \cdot \left(\left(\nabla \frac{\partial \mathcal{F}}{\partial \rho} \right) \rho \right)$, we know that

$$\begin{cases} \partial_t \mu(t) = -\nabla_{\mathcal{W}_2} \mathcal{F}(\mu(t)) \\ \mu(0) = \mu_0. \end{cases}$$

Notice that this is identical to the Cauchy problem above!

As a key example, the solution to the heat equation can be viewed as a gradient flow. In other words, if we have the heat equation

$$\begin{cases} \partial_t \mu = \Delta \mu \\ \mu(0, \cdot) = \mu_0(\cdot), \end{cases}$$

then we can construct the energy functional $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, given by

$$\mathcal{F}(\mu) = \begin{cases} \int_{\mathbb{R}^d} \log \left(\frac{d\mu}{dx}(x) \right) d\mu(x) & \text{if } \mu \ll \lambda, \\ +\infty & \text{otherwise,} \end{cases}$$

where λ denotes Lebesgue measure on $\mathbb{R}^d$. If $\mu(t)$ is the gradient flow of the above functional $\mathcal{F}$, then one can show $\mu(t)$ solves the heat equation. It can also be shown that $\mathcal{F}$ is λ-convex for $\lambda = 0$.

Moving forward, we will be working on the space of probability measures on graphs, rather than Euclidean space. Here, we desire a similar metric, $\mathcal{W}$, and functional, $\mathcal{F}$, to satisfy

$$\partial_t \mu(t) = -\nabla_{\mathcal{W}} \mathcal{F}(\mu(t))$$

for any solution of the heat equation.

2 Introduction to Markov Chains

Intuitively, we can think of a Markov jump process as a set of states in a state space, $X = \{x_1, x_2, ...x_n\}$, along with a transition matrix Q. Q encodes the probability of jumping to state x_j, given we are at state x_i. Say we are at state x_i at current time, and wish to jump to state x_j. After a time with exponential distribution $exp(q_{ij})$, if the process has not jumped to any other state, we will jump to state x_j. The rigorous definition of Q is as follows:

DEFINITION 2.1 (Transition Matrix, Q). *$q_{ij} \geq 0$ describes the rate of transition from state x_i to state x_j. We let $q_{ii} = -\sum_{j\neq i} q_{ij}$, so that, letting $Q = [q_{i,j}]$, the sum of each row of matrix Q is equal to 0.*

Q has the following properties:

- Q has a "memoryless property" meaning that each jump is independent of previous jumps, or how long we have been in the chain. It is solely dependent on which state we are in at the current time.
- If $q_{ij} = 0$ for some i, j, then the probability of jumping from state x_i to x_j is zero.

Knowing the transition rate matrix Q, with some calculation we are able to obtain the probability matrix $H(t)$[1].

DEFINITION 2.2 (Probability Matrix, $H(t)$). *Entry p_{ij} of the matrix $H(t)$ gives us the probability that we will be at state x_j after time t, assuming that we began at state x_i at time 0.*

$H(t)$ has the following properties[1]:

- For each time t, $H(t)$ is a real valued matrix dependent on t.
- $\{H(t) : t \geq 0\}$ is called the transition semigroup.
- $H(t) = e^{tQ} = \sum_{k=0}^{\infty} \frac{t^k}{k!} Q^k$.
- If, for all i, j, there exists $t \geq 0$ such that $H(t)_{ij} > 0$, then it is possible to jump to any state from any other state, and we say the matrix $K := Q + I$ is *irreducible*.

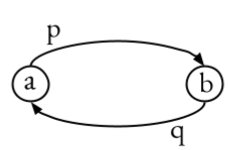
Note here that irreducibility of K ensures the existence of a unique *steady state* π, a row vector with entries corresponding to states in the state space. We can think of a steady state as the "limit" of the Markov jump process as $t \to \infty$. In other words, our steady state, π, is itself a probability measure on the state space, so its entries sum to 1, and it encodes the probability that the process is at each state x_i as time diverges to infinity. We note here as well that steady states are left eigenvectors of $K = Q + I$.

For the context of our problem, say we are given a finite directed complete weighted graph G, with nodes $S = \{s_1, \dots, s_n\}$, and an adjacency matrix $Q = [q_{ij}]$, where $q_{ij} \geq 0$ refers to the edge weight from s_i to s_j. It is possible to frame this weighted graph as a continuous time Markov jump process.

Intuitively, the collection of nodes of our graph can be thought of as "states" of a Markov chain. At any given time, we exist in a certain state, or node. We'll refer to these states by their indices, namely, when considering behavior jumping from a certain state i to j. Fundamentally, we consider ourselves "jumping" from state to state, and we wish to consider the probabilities of jumping from state to state. Our matrix Q of edge weights encodes these probabilities.

3 Defining $\mathcal{W}$ on the Two-Point Space, $\mathcal{P}(Q^1)$

Let us consider the two-point space, a graph with two-nodes $\{a, b\}$, where the edge weight from a to b is given by $p \in (0, 1]$, and the edge weight from b to a is given by $q \in (0, 1]$. We can visualize this as:



We can view this as a continuous time Markov jump process with state space $Q^1 := \{a, b\}$. To obtain transition matrix Q, we know $q_{1,2} = p$ and $q_{2,1} = q$. Since our rows must sum to 0, we know $q_{1,1} = -p$ and $q_{2,2} = -q$. Thus we have the corresponding transition matrix:

$$Q = \begin{bmatrix} -p & p \\ q & -q \end{bmatrix} \text{ and } K = Q + I = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$

From here, our time semigroup is given by $H(t) = \{e^{tQ} : t \geq 0\}$. With some elementary calculation, we obtain that:

$$H(t) = e^{tQ} = \frac{1}{p+q}\left(\begin{bmatrix} q & p \\ q & p \end{bmatrix} + e^{-(p+q)t}\begin{bmatrix} p & -p \\ -q & q \end{bmatrix}\right).$$

Since $p, q \neq 0$, for any state in Q^1, we can reach any other state in Q^1, and thus K is irreducible. Since steady states are left eigenvectors of K, we obtain steady state π as follows:

$$\begin{bmatrix} \pi(a) & \pi(b) \end{bmatrix}\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} = \begin{bmatrix} \pi(a) - p\pi(a) + q\pi(b) & p\pi(a) + \pi(b) - q\pi(b) \end{bmatrix} = \begin{bmatrix} \pi(a) & \pi(b) \end{bmatrix}$$

$$\implies \pi(a) = \frac{q}{p+q}, \pi(b) = \frac{p}{p+q}.$$

Now any probability measure on $\{a, b\}$ must be of the following form, for some $\beta \in [-1, 1]$:

$$\mu^\beta = \frac{1}{2}((1-\beta)\delta_a + (1+\beta)\delta_b).$$

For any measure μ^β, we desire a corresponding probability density function $\rho^\beta : Q^1 \to \mathbb{R}$ with respect to π. For any subset $A \subseteq Q^1$, ρ^β should satisfy the following:

$$\mu^\beta(A) = \rho^\beta(a)\pi(a)\delta_A(a) + \rho^\beta(b)\pi(b)\delta_A(b).$$

To explicitly solve for ρ^β in terms of μ^β:

$$\rho^\beta = \begin{bmatrix} \rho^\beta(a) & \rho^\beta(b) \end{bmatrix} = \begin{bmatrix} \frac{\mu^\beta(a)}{\pi(a)} & \frac{\mu^\beta(b)}{\pi(b)} \end{bmatrix} = \begin{bmatrix} \frac{(p+q)(1-\beta)}{2q} & \frac{(p+q)(1+\beta)}{2p} \end{bmatrix}.$$

Our goal is to construct a metric comparing ρ^β for various values of $\beta \in [-1, 1]$. In order to construct our metric, we choose a "nice" function $\theta : [0, \infty) \times [0, \infty) \to [0, \infty)$ that satisfies the following:

(i) θ is continuous on $[0, \infty) \times [0, \infty)$
(ii) θ is C^∞ on $(0, \infty) \times (0, \infty)$
(iii) $\theta(s, t) = \theta(t, s)$ for $s, t \geq 0$
(iv) $\theta(s, t) > 0$ for $s, t > 0$

Then, let

$$\hat{\rho}(\beta) := \theta(\rho^\beta(a), \rho^\beta(b)).$$

We introduce our metric $\mathcal{W}$:

Definition 3.1. *For $\alpha, \beta \in [-1, 1]$, we define:*

$$\mathcal{W}(\rho^\alpha, \rho^\beta) = \frac{1}{2}\sqrt{\frac{1}{p} + \frac{1}{q}}\left|\int_\alpha^\beta \frac{1}{\sqrt{\hat{\rho}(r)}} dr\right|.$$

Given a function θ, we now have an explicit definition for a metric $\mathcal{W}$ on the space $\mathcal{P}(Q^1)$. Moving forward, we will show the isometry of this metric to a subset of the real line, allowing us to explore geodesics, convexity, and gradient flows.

4 The Isometry Induced by $\mathcal{W}$

Let us define

$$\varphi(\beta) := \frac{1}{2}\sqrt{\frac{1}{p}+\frac{1}{q}}\int_0^\beta \frac{1}{\sqrt{\hat{\rho}(r)}}dr.$$

Notice first that φ is continuous and strictly increasing, by assumptions placed on θ. For $\alpha, \beta \in [-1, 1]$, we have that:

$$\begin{aligned}
|\varphi(\alpha)-\varphi(\beta)| &= \left|\frac{1}{2}\sqrt{\frac{1}{p}+\frac{1}{q}}\int_0^\alpha \frac{1}{\sqrt{\hat{\rho}(r)}}dr - \frac{1}{2}\sqrt{\frac{1}{p}+\frac{1}{q}}\int_0^\beta \frac{1}{\sqrt{\hat{\rho}(r)}}dr\right| \\
&= \frac{1}{2}\sqrt{\frac{1}{p}+\frac{1}{q}}\left|\int_\alpha^\beta \frac{1}{\sqrt{\hat{\rho}(r)}}dr\right| \\
&= \mathcal{W}(\rho^\alpha, \rho^\beta).
\end{aligned}$$

We define

$$(-1, 1)_* = \{\beta \in [-1, 1] : |\varphi(\beta)| < \infty\}.$$

Consider

$$\mathcal{P}_1(\mathcal{Q}^1) := \{\rho^\beta \in \mathcal{P}(\mathcal{Q}^1) : \beta \in (-1, 1)_*\}.$$

To see that $\mathcal{W}$ is a metric on $\mathcal{P}_1(\mathcal{Q}^1)$, note that the following hold:

(i) $\mathcal{W}(\rho^\alpha, \rho^\beta) = 0 \iff |\varphi(\alpha) - \varphi(\beta)| = 0 \iff \varphi(\alpha) = \varphi(\beta) \iff \alpha = \beta$

(ii) $\mathcal{W}(\rho^\alpha, \rho^\beta) = |\varphi(\alpha) - \varphi(\beta)| = |\varphi(\beta) - \varphi(\alpha)| = \mathcal{W}(\rho^\beta, \rho^\alpha)$

(iii) $\mathcal{W}(\rho^\alpha, \rho^\gamma) = |\varphi(\alpha) - \varphi(\gamma)| \le |\varphi(\alpha) - \varphi(\beta)| + |\varphi(\beta) - \varphi(\gamma)| = \mathcal{W}(\rho^\alpha, \rho^\beta) + \mathcal{W}(\rho^\beta, \rho^\gamma)$.

We have shown that for all $\alpha, \beta \in (-1, 1)_*$, $\mathcal{W}(\rho^\alpha, \rho^\beta) = |\varphi(\alpha) - \varphi(\beta)|$. Let us first define

$$I = \{\varphi(\beta) : \beta \in (-1, 1)_*\}.$$

Then, defining $J : \mathcal{P}_1(\mathcal{Q}^1) \to I$, where $\rho^\beta \mapsto \varphi(\beta)$, it is clear that J preserves distance from $(\mathcal{P}_1(\mathcal{Q}^1), \mathcal{W})$ to $(I, |\cdot|)$, and thus J is an isometry. In other words, rather than considering each ρ^β as a probability density function, we may instead think of the scalar value $J(\rho^\beta) = \varphi(\beta)$, which will allow us to simplify our calculations later on.

To summarize our isometries and variables, each measure on $\mathcal{P}_1(\mathcal{Q}^1)$ is uniquely defined by a value $\beta \in [-1, 1]$. We have used the above isometry to show that $(\mathcal{P}_1(\mathcal{Q}^1), \mathcal{W})$ and $(I, |\cdot|)$ are isomorphic, and thus, using the interval I to only consider the values of β where $|\varphi(\beta)| < \infty$, we can apply J to any measure ρ^β, and instead observe nice properties of $\varphi(\beta)$.

5 Example: θ is the Logarithmic Mean

As an example, let us consider the case where $\theta(s,t) = \int_0^1 s^{1-r}t^r dr$. If we have $\beta \in [-1, 1]$,

$$\begin{aligned}
\hat{\rho}(\beta) &= \theta\left(\rho^\beta(a), \rho^\beta(b)\right) \\
&= \theta\left(\frac{p+q}{q}\frac{1-\beta}{2}, \frac{p+q}{p}\frac{1+\beta}{2}\right) \\
&= \int_0^1 \left(\frac{p+q}{2q}(1-\beta)\right)^{1-r}\left(\frac{p+q}{2p}(1+\beta)\right)^r dr
\end{aligned}$$

$$= \frac{p+q}{2pq} \frac{q(1+\beta) - p(1-\beta)}{\ln q(1+\beta) - \ln p(1-\beta)}.$$

Notice first that $\hat{\rho}(-1) = \hat{\rho}(1) = 0$. Furthermore,

$$\begin{aligned}
(-1,1)_* &= \{\beta \in [-1,1] : |\varphi(\beta)| < \infty\} \\
&= \{\beta \in [-1,1] : \left| \frac{1}{2}\sqrt{\frac{1}{p} + \frac{1}{q}} \int_0^\beta \frac{1}{\sqrt{\hat{\rho}(r)}} \right| dr < \infty\} \\
&= \{\beta \in [-1,1] : \left| \int_0^\beta \frac{1}{\sqrt{\hat{\rho}(r)}} \right| dr < \infty\}.
\end{aligned}$$

We see that in this example, $(-1,1)_* = [-1,1]$, so $I = [\varphi(-1), \varphi(1)]$ is compact. Let us explicitly compute $\mathcal{W}(\rho^\alpha, \rho^\beta)$ for $\alpha \le \beta$.

$$\begin{aligned}
\mathcal{W}(\rho^\alpha, \rho^\beta) &= \frac{1}{2}\sqrt{\frac{1}{p} + \frac{1}{q}} \left| \int_\alpha^\beta \frac{1}{\sqrt{\hat{\rho}(r)}} dr \right| \\
&= \frac{1}{2}\sqrt{\frac{1}{p} + \frac{1}{q}} \left(\sqrt{\frac{2pq}{p+q}} \right) \int_\alpha^\beta \sqrt{\frac{\ln q(1+r) - \ln p(1-r)}{q(1+r) - p(1-r)}} dr \\
&= \frac{1}{\sqrt{2}} \int_\alpha^\beta \sqrt{\frac{\ln q(1+r) - \ln p(1-r)}{q(1+r) - p(1-r)}} dr.
\end{aligned}$$

In the case where $p = q$, we can simplify this further. Observe:

$$\begin{aligned}
\hat{\rho}(\beta) &= \frac{p+q}{2pq} \frac{q(1+\beta) - p(1-\beta)}{\ln q(1+\beta) - \ln p(1-\beta)} \\
&= \frac{2p}{2p^2} \frac{p(1+\beta) - p(1-\beta)}{\ln p(1+\beta) - \ln p(1-\beta)} \\
&= \frac{2\beta}{\ln(1+\beta) - \ln(1-\beta)} \\
&= \frac{\beta}{\operatorname{arctanh}(\beta)}.
\end{aligned}$$

Thus, in the case $p = q$, we have:

$$\begin{aligned}
\mathcal{W}(\rho^\alpha, \rho^\beta) &= \frac{1}{2}\sqrt{\frac{1}{p} + \frac{1}{q}} \left| \int_\alpha^\beta \frac{1}{\sqrt{\hat{\rho}(r)}} dr \right| \\
&= \frac{1}{\sqrt{2p}} \int_\alpha^\beta \sqrt{\frac{\operatorname{arctanh}(r)}{r}} dr.
\end{aligned}$$

6 $\mathcal{W}$-Geodesics

A geodesic from point a to point b takes the shortest path possible from a to b with respect to the given metric. Furthermore, a constant speed geodesic is a geodesic that moves at a constant pace, in other words:

Definition 6.1 (Constant Speed Geodesic). *A constant speed geodesic in a metric space (X, d) is a curve $u : [0,1] \to X$ such that*

$$d(u(s), u(t)) = |t - s| d(u(0), u(1))$$

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

We can consider Euclidean space as an example. Given two points $a, b \in \mathbb{R}^d$, the shortest path from a to b is a straight line. Furthermore, if we want to maintain constant speed along our geodesic, it must be of the form $u(t) = (1-t)a + tb$.

Generalizing to other manifolds, we can intuitively consider constant speed geodesics as the shortest path between two points that travels at constant speed.The existence and uniqueness of constant speed geodesics will later aid us in showing the convexity of the functional $\mathcal{F}$ along these geodesics. The convexity of the functional proves interesting when considering curvature properties of a manifold, and also helps ensure fast convergence to a minimizer of a functional, $\mathcal{F}$.

In our context, a constant speed geodesic is a curve $\mu : [0,1] \to \mathcal{P}_1(Q^1)$ such that

$$\mathcal{W}(\mu(s), \mu(t)) = |t-s|\mathcal{W}(\mu(0), \mu(1))$$

for all $s, t \in [0,1]$. Notice that for any $\mu(t) \in \mathcal{P}(Q^1)$, we have that $\mu(t) = \rho^\beta$ for some $\beta \in [-1,1]$. Thus, rather than seeking a function $\mu : [0,1] \to \mathcal{P}(Q^1)$, we instead seek $\gamma : [0,1] \to [-1,1]$ so that $\mu(t) = \rho^{\gamma(t)}$.

THEOREM 6.2. *Let* $\rho^\alpha, \rho^\beta \in \mathcal{P}_1(Q^1)$ *such that* $\alpha, \beta \in (-1,1)_*$. *There exists a unique constant speed geodesic* $\rho^{\gamma(t)} : [0,1] \to \mathcal{P}(Q^1)$ *so that* $\rho^{\gamma(0)} = \rho^\alpha$ *and* $\rho^{\gamma(1)} = \rho^\beta$. *Furthermore,* $\gamma \in C^1([0,1],[-1,1])$ *and, for* $t \in [0,1]$, γ *satisfies*

$$\gamma'(t) = 2(sgn(\beta-\alpha))\mathcal{W}(\rho^\alpha, \rho^\beta)\sqrt{\frac{pq}{p+q}\hat{\rho}(\gamma(t))}.$$

Proof. Fix $\alpha, \beta \in (-1,1)_*$. Consider the ordinary differential equation:

$$\begin{cases} \gamma'(t) = 2(sgn(\beta-\alpha))\mathcal{W}(\rho^\alpha, \rho^\beta)\sqrt{\frac{pq}{p+q}\hat{\rho}(\gamma(t))} \\ \gamma(0) = \alpha. \end{cases}$$

Let us consider $f(x) = 2(sgn(\beta-\alpha))\mathcal{W}(\rho^\alpha, \rho^\beta)\sqrt{\frac{pq}{p+q}\hat{\rho}(x)}$. Since $f(x)$ and $f'(x)$ are continuous, there exists a unique function $\gamma \in C^1([0,1],[-1,1])$ such that γ is the solution to the above ordinary differential equation. Now, we see that:

$$\begin{aligned} \mathcal{W}(\rho^{\gamma(t)}, \rho^{\gamma(s)}) &= |\varphi(\gamma(t)) - \varphi(\gamma(s))| \\ &= \left| \int_s^t \varphi'(\gamma(r))\gamma'(r)dr \right| \\ &= \left| \int_s^t \frac{1}{2}\sqrt{\frac{1}{p}+\frac{1}{q}}\frac{1}{\sqrt{\hat{\rho}(\gamma(r))}}\gamma'(r)dr \right| \\ &= \left| \int_s^t \frac{1}{2}\sqrt{\frac{1}{p}+\frac{1}{q}}\frac{1}{\sqrt{\hat{\rho}(\gamma(r))}}2(sgn(\beta-\alpha))\mathcal{W}(\rho^\alpha, \rho^\beta)\sqrt{\frac{pq}{p+q}\hat{\rho}(\gamma(r))}dr \right| \\ &= \mathcal{W}(\rho^\alpha, \rho^\beta)|(t-s)|. \end{aligned}$$

Thus $t \mapsto \rho^{\gamma(t)}$ is a constant speed geodesic. Next we show that $\gamma(1) = \beta$. To begin, notice that:

$$\begin{aligned} \mathcal{W}(\rho^{\gamma(0)}, \rho^{\gamma(1)}) &= |\varphi(\gamma(0)) - \varphi(\gamma(1))| \\ &= |\varphi(\alpha) - \varphi(\gamma(1))| && \gamma(0) = \alpha \\ &= \mathcal{W}(\rho^\alpha, \rho^\beta)|0-1| && \text{Definition of constant speed geodesic} \\ &= |\varphi(\beta) - \varphi(\alpha)|. \end{aligned}$$

Since $\gamma(0) = \alpha$ by assumption, we now have

$$|\varphi(\gamma(1)) - \varphi(\alpha)| = |\varphi(\beta) - \varphi(\alpha)|.$$

Since $\varphi(\beta)$ is strictly increasing for all $\beta \in (-1,1)_*$, this tells us $\varphi(\gamma(1)) = \varphi(\beta)$, allowing us to conclude $\gamma(1) = \beta$. Thus, $t \mapsto \rho^{\gamma(t)}$ is a constant speed geodesic.

Finally, we want to show that $\rho^{\gamma(t)}$ is the *unique* constant speed geodesic. Fix $\alpha, \beta \in (-1,1)_*$. We have already established the existence of a constant speed geodesic $\gamma(t)$ such that $\rho^{\gamma(0)} = \rho^\alpha$ and $\rho^{\gamma(1)} = \rho^\beta$. Say there exists another such geodesic, $\rho^{\sigma(t)}$ such that $\rho^{\sigma(0)} = \rho^\alpha$ and $\rho^{\sigma(1)} = \rho^\beta$. Fix $t \in (0,1)$. We have that:

$$\begin{aligned} |\varphi(\gamma(t)) - \varphi(\alpha)| &= \mathcal{W}(\rho^{\gamma(t)}, \rho^\alpha) \\ &= \mathcal{W}(\rho^{\gamma(t)}, \rho^{\gamma(0)}) \\ &= t \cdot \mathcal{W}(\rho^\alpha, \rho^\beta). \end{aligned}$$

Similarly,

$$\begin{aligned} |\varphi(\sigma(t)) - \varphi(\alpha)| &= \mathcal{W}(\rho^{\sigma(t)}, \rho^\alpha) \\ &= \mathcal{W}(\rho^{\sigma(t)}, \rho^{\sigma(0)}) \\ &= t \cdot \mathcal{W}(\rho^\alpha, \rho^\beta). \end{aligned}$$

This tells us $\gamma(t) = \sigma(t)$ for all $t \in (0,1)$, so $\gamma(t)$ is unique. □

Let us impose an additional condition on our function θ, now assuming the existence of $f \in C([0,\infty),\mathbb{R}) \cap C^\infty((0,\infty),\mathbb{R})$ such that $f''(t) > 0$ for all $t > 0$ and

$$\theta(s,t) = \frac{s - t}{f'(s) - f'(t)}$$

for all $s, t > 0$ where $s \neq t$.

Furthermore, let us define $\mathcal{F} : \mathcal{P}_1(Q^1) \to \mathbb{R}$ by

$$\begin{aligned} \mathcal{F}(\rho) &= \sum_{x \in Q_1} f(\rho(x))\pi(x) \\ &= \frac{q}{p+q} f(\rho(a)) + \frac{p}{p+q} f(\rho(b)). \end{aligned}$$

To consider the convexity of the above functional $\mathcal{F}$, we define the function $K : (-1,1) \to \mathbb{R}$ by

$$K(\beta) := \frac{p+q}{2} + \frac{1}{2}\hat{\rho}(\beta)(qf''(\rho^\beta(b)) + pf''(\rho^\beta(a)))$$

and we let

$$\kappa := inf\{K(\beta) : \beta \in (-1,1)\}.$$

Since, by assumption, $f'' > 0$, we know that $\kappa \geq \frac{p+q}{2}$.

THEOREM 6.3 (Convexity along Geodesics). *Let κ be defined as above. Then, the functional $\mathcal{F}$ is κ-convex along geodesics. More explicitly, let $\rho^\alpha, \rho^\beta \in \mathcal{P}_1(Q^1)$, and let $\rho^{\gamma(t)}$ be the unique constant speed geodesic satisfying $\rho^\gamma(0) = \rho^\alpha$ and $\rho^\gamma(1) = \rho^\beta$. Then, for all $t \in [0,1]$, the following inequality holds:*

$$\mathcal{F}(\rho^{\gamma(t)}) \leq (1-t)\mathcal{F}(\rho^\alpha) + t\mathcal{F}(\rho^\beta) - \frac{\kappa}{2}t(1-t)\mathcal{W}^2(\rho^\alpha, \rho^\beta).$$

Proof. Let $\alpha, \beta \in (-1,1)_*$. Let $\rho^{\gamma(t)}$ be the unique constant speed geodesic satisfying $\rho^\gamma(0) = \rho^\alpha$ and $\rho^\gamma(1) = \rho^\beta$.

Let us assume WLOG that $\alpha \leq \beta$, and for ease of notation, let $w = \mathcal{W}(\rho^\alpha, \rho^\beta)$. We want to show

$$\mathcal{F}(\rho^{\gamma(t)}) \leq (1-t)\mathcal{F}(\rho^\alpha) + t\mathcal{F}(\rho^\beta) - \frac{\kappa}{2}t(1-t)w^2.$$

Let us define $\zeta(t) = \mathcal{F}(\rho^{\gamma(t)})$, and it will suffice to show

$$\zeta''(t) \geq w^2\kappa.$$

Applying our definition of $\mathcal{F}$ from above, we know that

$$\begin{aligned}\zeta'(t) &= \frac{q}{p+q}f'(\rho^{\gamma(t)}(a))\left(\frac{p+q}{q}\cdot\frac{-1}{2}\right)\gamma'(t) + \frac{p}{p+q}f'(\rho^{\gamma(t)}(b))\left(\frac{p+q}{p}\cdot\frac{1}{2}\right)\gamma'(t)\\ &= \frac{1}{2}\gamma'(t)(f'(\rho^{\gamma(t)}(b) - f'(\rho^{\gamma(t)}(a)).\end{aligned}$$

Now, we can apply definition of $\gamma'(t)$ from Theorem 6.2. Substituting this in the above equation, and applying our extra assumption of θ, we see that

$$\begin{aligned}\zeta'(t) &= \frac{1}{2}(2)\mathcal{W}(\rho^\alpha, \rho^\beta)\sqrt{\frac{pq}{p+q}\hat{\rho}(\gamma(t))}(f'(\rho^{\gamma(t)}(b) - f'(\rho^{\gamma(t)}(a)))\\ &= w\sqrt{\frac{pq}{p+q}\cdot\frac{\rho^{\gamma(t)}(b) - \rho^{\gamma(t)}(a)}{f'(\rho^{\gamma(t)}(b) - f'(\rho^{\gamma(t)}(a))}}(f'(\rho^{\gamma(t)}(b) - f'(\rho^{\gamma(t)}(a)))\\ &= w\sqrt{\frac{pq}{p+q}}\sqrt{(\rho^{\gamma(t)}(b) - \rho^{\gamma(t)}(a))(f'(\rho^{\gamma(t)}(b) - f'(\rho^{\gamma(t)}(a)))}.\end{aligned}$$

By differentiating once more, and yet again substituting $\gamma'(t)$, we obtain

$$\zeta''(t) = w^2K(\gamma(t)) \geq w^2\kappa.$$

And thus we have shown κ-convexity along constant speed geodesics $\rho^{\gamma(t)}$. As a result, we know gradient flows of $\mathcal{F}$ converge to minimizers exponentially quickly. □

7 Gradient Flows with respect to $\mathcal{W}$

In Euclidean space, Brownian motion may be thought of in the following way: Let $X_h(t)$ be a random walk on a lattice with edge length h. As $h \to 0$, our random walk converges to Brownian motion. A heat flow $\rho(t, x)$ may be thought of as the probability density function describing the location of a particle moving according to Brownian motion at time t. This idea can be extended to a more general manifold. The shape of the manifold determines the heat flow, i.e, the probability that a point is in a certain place at a certain time.

Intuitively, the shape of the manifold governs the movement of the particle. In our context, a continuous time Markov jump process on the graph, the Markov kernel plays the role of the manifolds above; it is what governs the movement of a particle moving randomly on the graph, determining where it is most likely for the particle to travel. If we desire a sense of heat flow that mimics a probability density function of the location of a particle at a certain time, it is natural to turn to the continuous time semigroup – which gives the probability a particle is at a state j, given it began at time $t = 0$ at state i.

Recall that we defined our probability matrix $H(t) = e^{tQ}$. In this time dependent matrix, the ij-th entry of matrix $H(t)$ gives the probability that we are at state j at time t given that we began at state i at time $t = 0$. While each entry of $H(t)$ gives a single probability with respect to two nodes, we desire a "path" through the space of probability measures. By multiplying $H(t)$ on the right by ρ^β, we obtain a probability measure in $\mathcal{P}_1(Q^1)$, and we will use this to define our notion of heat flow. Since $H(t)\rho^\beta$ is a probability measure, we know there exists $\beta_t \in [-1,1]$ such that $\rho^{\beta_t} = H(t)\rho^{\beta}$. With some calculation, we can explicitly solve to obtain

$$\beta_t = \frac{p-q}{p+q}(1-e^{-(p+q)t}) + \beta e^{-(p+q)t}.$$

With the above context in mind, for some $\beta \in [-1,1]$, we define

$$\mu : \mathbb{R} \to \mathcal{P}(\mathcal{Q}^1) \text{ defined by } t \mapsto \rho^{\beta_t} = H(t)\rho^{\beta}$$

to be the heat flow trajectory, starting at ρ^{β} at time $t=0$. Now, we introduce a major result:

THEOREM 7.1 (Heat flow is the gradient flow of the entropy). *For $\beta \in [-1,1]$, our function $\mu(t)$, as defined above, is a gradient flow trajectory of the functional $\mathcal{F}$ in the Riemannian manifold $(\mathcal{P}_1(\mathcal{Q}^1), \mathcal{W})$*

Proof. To show the above, we desire to show that

$$\frac{d}{dt}\mu(t) = -\nabla_{\mathcal{W}}\mathcal{F}(\mu(t)).$$

Intuitively, our isometry from section 4 allows us to view the above from a different perspective. Since gradient flow is uniquely determined by the metric, our isometry provides a one to one correspondence between gradient flows in $(\mathcal{P}_1(\mathcal{Q}^1), \mathcal{W})$ and gradient flows in $(I, |\cdot|)$. Though this is outside the scope of this thesis, [4] provides a good resource to explore this further. Rather than showing the above directly, we define $\tilde{\mathcal{F}} = \mathcal{F} \circ J^{-1}$, and we will now equivalently show:

$$\frac{d}{dt}\varphi(\beta_t) = -(\tilde{\mathcal{F}}(\varphi(\beta_t)).$$

Let us define $\ell(\beta) = \rho^{\beta}(a)$ and $r(\beta) = \rho^{\beta}(b)$, as well as $c_{pq} = \frac{1}{2}\sqrt{\frac{1}{p} + \frac{1}{q}}$. We calculate

$$\ell'(\beta) = \frac{d}{d\beta}\left(\frac{p+q}{q}\frac{1-\beta}{2}\right) = -\frac{p+q}{2q}$$

and

$$r'(\beta) = \frac{d}{d\beta}\left(\frac{p+q}{p}\frac{1+\beta}{2}\right) = \frac{p+q}{2p}.$$

Recall also that

$$\hat{\rho}(\beta) = \theta(\rho^{\beta}(a), \rho^{\beta}(b)) = \theta(\rho^{\beta}(b), \rho^{\beta}(a)) = \frac{\rho^{\beta}(b) - \rho^{\beta}(a)}{f'(\rho^{\beta}(b)) - f'(\rho^{\beta}(a))} = \frac{r(\beta) - \ell(\beta)}{f'(r(\beta)) - f'(\ell(\beta))}$$

and

$$\varphi(\beta) = c_{pq}\int_0^{\beta} \frac{1}{\sqrt{\hat{\rho}(r)}}dr,$$

which allows us to calculate

$$\frac{d}{d\beta}\varphi(\beta) = c_{pq}\sqrt{\frac{f'(r(\beta)) - f'(\ell(\beta))}{r(\beta) - \ell(\beta)}}.$$

Since $f, r, \ell,$ and φ are continuously differentiable, for all $\beta \in I$, we must have that $\tilde{\mathcal{F}}$ is continuously differentiable on I as well. Now, we know that

$$\frac{d}{d\beta}\tilde{\mathcal{F}}(\varphi(\beta)) = \tilde{\mathcal{F}}'(\varphi(\beta)) \cdot \varphi'(\beta).$$

Since we have already calculated $\varphi'(\beta)$, we can calculate $\frac{d}{d\beta}\tilde{\mathcal{F}}(\varphi(\beta))$ so that we may find $\tilde{\mathcal{F}}'(\varphi(\beta))$. First,

$$\begin{aligned}\tilde{\mathcal{F}}(\varphi(\beta)) &= \tilde{\mathcal{F}}(J(\rho^\beta)) = \mathcal{F}(\rho^\beta)\\ &= \frac{q}{p+q} f(\ell(\beta)) + \frac{p}{p+q} f(r(\beta)).\end{aligned}$$

Now, we can differentiate, and see that

$$\begin{aligned}\frac{d}{d\beta}\tilde{\mathcal{F}}(\varphi(\beta)) &= \frac{q}{p+q} f'(\ell(\beta))\ell'(\beta) + \frac{p}{p+q} f'(r(\beta))r'(\beta)\\ &= \frac{q}{p+q}\left(-\frac{p+q}{2q}\right) f'(\ell(\beta)) + \frac{p}{p+q}\left(\frac{p+q}{2p}\right) f'(r(\beta))\\ &= \frac{1}{2}(f'(r(\beta)) - f'(\ell(\beta))).\end{aligned}$$

Finally, we can put this together to obtain

$$\begin{aligned}\tilde{\mathcal{F}}'(\varphi(\beta)) &= \frac{\frac{d}{d\beta}\tilde{\mathcal{F}}(\varphi(\beta))}{\varphi'(\beta)}\\ &= \frac{f'(r(\beta)) - f'(\ell(\beta))}{2\varphi'(\beta)}\\ &= \frac{f'(r(\beta)) - f'(\ell(\beta))}{2c_{pq}\sqrt{\frac{f'(r(\beta)) - f'(\ell(\beta))}{r(\beta)-\ell(\beta)}}}\\ &= \frac{1}{2c_{pq}}\sqrt{(r(\beta) - \ell(\beta))(f'(r(\beta)) - f'(\ell(\beta))}.\end{aligned}$$

Note that this above formula holds for β_t, i.e.

$$\tilde{\mathcal{F}}'(\varphi(\beta_t)) = \frac{1}{2c_{pq}}\sqrt{(r(\beta_t) - \ell(\beta_t))(f'(r(\beta_t)) - f'(\ell(\beta_t))} \qquad (1).$$

Now recall that

$$\beta_t = \frac{p-q}{p+q}(1 - e^{-(p+q)t}) + \beta e^{-(p+q)t}.$$

Then we have that

$$\frac{d}{dt}\beta_t = p(1-\beta_t) - q(1+\beta_t).$$

We can use this to compute $\frac{d}{dt}\varphi(\beta_t)$, since we know that

$$\begin{aligned}\frac{d}{dt}\varphi(\beta_t) &= \frac{d}{d\beta}\varphi(\beta_t)\cdot\frac{d}{dt}\beta_t\\ &= c_{pq}\sqrt{\frac{f'(r(\beta_t)) - f'(\ell(\beta_t))}{r(\beta_t) - \ell(\beta_t)}}\cdot(p(1-\beta_t) - q(1+\beta_t))\\ &= c_{pq}\sqrt{\frac{f'(r(\beta_t)) - f'(\ell(\beta_t))}{r(\beta_t) - \ell(\beta_t)}}\left(-2\left(\frac{pq}{p+q}\right)\left(\frac{p+q}{p}\frac{1+\beta_t}{2} - \frac{p+q}{q}\frac{1-\beta_t}{2}\right)\right)\\ &= c_{pq}\sqrt{\frac{f'(r(\beta_t)) - f'(\ell(\beta_t))}{r(\beta_t) - \ell(\beta_t)}}\left(-\frac{2}{c_{pq}^2}(\rho^{\beta_t}(b) - \rho^{\beta_t}(a))\right)\end{aligned}$$

$$= c_{pq}\sqrt{\frac{f'(r(\beta_t)) - f'(\ell(\beta_t))}{r(\beta_t) - \ell(\beta_t)}}\left(-\frac{1}{c_{pq}^2}(r(\beta_t) - \ell(\beta_t))\right)$$
$$= -\frac{1}{2c_{pq}}\sqrt{(r(\beta_t) - \ell(\beta_t))(f'(r(\beta_t)) - f'(\ell(\beta_t))}.$$

By equation (1), we see that $\frac{d}{dt}\varphi(\beta_t) = -\tilde{\mathcal{F}}'(\varphi(\beta_t))$. ∎

Conclusion: Extending to the general space

In this thesis, we have worked solely in the space of probability measures on two-node graphs, $\mathcal{P}(\mathcal{Q}^1)$, introducing a metric $\mathcal{W}$, examining its properties, before introducing and proving a major result, that heat flow is the gradient flow of the entropy. Moving forward, readers can refer to [3] to see how Maas generalizes these ideas to the more general space of graphs, using the base ideas developed in the two-node example.

Acknowledgments

I'd like to sincerely thank Claire Murphy for her patience, time, and care throughout this process. Having her as a mentor and friend has been the most valuable part of this project, her patience and enthusiasm have helped me understand more than I ever thought possible. This thesis truly would not be the same without her. Thank you to Professor Katy Craig for always encouraging me, helping me find my path in mathematics, and being the best role model I could ask for. I would also like to thank my mother, Kamna Mittal, for supporting me throughout my time at UC Santa Barbara, and my best friend, Greta, for keeping me laughing, even through the stressful days.

REFERENCES

[1] M. Aldridge, *Math2750: Introduction to Markov Processes*. https://mpaldridge.github.io/math2750/, 2021. Accessed: 2025-05-04.

[2] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker–Planck equation*, SIAM Journal on Mathematical Analysis, 29 (1998), pp. 1–17.

[3] J. Maas, *Gradient flows of the entropy for finite Markov chains*, Journal of Functional Analysis, 261 (2011), pp. 2250–2292.

[4] F. Santambrogio, {*Euclidean, Metric, and Wasserstein*} *Gradient Flows: an overview*, Bulletin of Mathematical Sciences, 7 (2017), pp. 87–154.