

## GEODESICS OF MINIMAL LENGTH IN THE SET OF PROBABILITY MEASURES ON GRAPHS

WILFRID GANGBO, WUCHEN LI AND CHENCHEN MOU\*

**Abstract.** We endow the set of probability measures on a weighted graph with a Monge–Kantorovich metric induced by a function defined on the set of edges. The graph is assumed to have  $n$  vertices and so the boundary of the probability simplex is an affine  $(n - 2)$ -chain. Characterizing the geodesics of minimal length which may intersect the boundary is a challenge we overcome even when the endpoints of the geodesics do not share the same connected components. It is our hope that this work will be a preamble to the theory of mean field games on graphs.

**Mathematics Subject Classification.** 49K35, 49Q20, 60J27.

Received June 6, 2018. Accepted September 24, 2018.

### 1. INTRODUCTION

The past two decades have witnessed an increasing number of studies on geodesics of minimal length on the set of probability measures on manifolds and Hilbert spaces [2, 24] (*cf.* for applications [1, 4, 5, 15–18]). In these cases, the geodesics are characterized by Hamilton–Jacobi equations which appear through a duality argument [2, 11–14, 24]. The story is different when Hilbert spaces or manifolds are replaced by discrete states which are not length spaces. In modeling and numerics [7–9], one faces the issue of dealing with geodesics of minimal length on the set of probability measures on graphs, the probability simplexes. Here, the graph represents either a discrete sample space or the discretization of a continuous domain. In these applications, an essential question is to characterize these geodesics. This is not such a so difficult task when the geodesic stays in the interior of probability simplexes. However, a challenge arise when one needs to go beyond understanding the differential structures in the interior and push the study to the boundary. To our knowledge, the literature on the problem is meager. This paper introduces some new ideas, which allow to characterize the geodesics without excluding the possibility that the complement of the endpoints touches the boundary.

Let  $G = (V, E, \omega)$  denote an undirected graph of vertices  $V = \{1, \dots, n\}$  and edges  $E$ , with a weighted metric  $\omega = (\omega_{ij})$ . It is given by an  $n$  by  $n$  symmetric matrix with non-negative entries  $\omega_{ij}$  such that  $\omega_{ij} > 0$  if  $(i, j) \in E$ . For simplicity, assume that the graph is connected, simple, with no self-loops or multiple edges. Let  $\mathcal{P}(G)$  denote

---

*Keywords and phrases:* Optimal transport on simplexes, manifold with boundary, Geodesic, Hamilton–Jacobi equations on graphs.

Department of Mathematics, UCLA, Los Angeles, CA 90095, USA.

\* Corresponding author: [muchenchen@math.ucla.edu](mailto:muchenchen@math.ucla.edu)

the probability simplex

$$\left\{ \rho \in [0, 1]^n \mid \sum_{i=1}^n \rho_i = 1 \right\},$$

the set of probability measures on  $V$  and let  $\mathcal{P}_0(G) := \mathcal{P}(G) \cap (0, 1)^n$  denote the interior of  $\mathcal{P}(G)$ . Any symmetric function  $g : [0, \infty)^2 \rightarrow [0, \infty)$  induces an equivalence relation on  $S^{n \times n}$ , the set of  $n$  by  $n$  skew-symmetric matrix: for  $\rho \in \mathcal{P}(G)$ ,  $v, \tilde{v} \in S^{n \times n}$  are equivalent if

$$(v_{ij} - \tilde{v}_{ij})g_{ij}(\rho) = 0 \quad \forall (i, j) \in E \quad (1.1)$$

holds. The quotient space  $\mathbb{H}_\rho$  is endowed with a metric tensor  $(g_{ij}(\rho))_{ij} = (g(\rho_i, \rho_j))_{ij}$  which yields the inner product

$$(v, \tilde{v})_\rho := \frac{1}{2} \sum_{(i,j) \in E} v_{ij} \tilde{v}_{ij} g_{ij}(\rho) \quad \forall v, \tilde{v} \in S^{n \times n}.$$

The function  $g$  is used to produce the underlying Hamiltonian  $\mathcal{H}_g : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  given by

$$\mathcal{H}_g(\rho, \phi) = \frac{1}{4} \sum_{(i,j) \in E} \omega_{ij} g_{ij}(\rho) (\phi_i - \phi_j)^2. \quad (1.2)$$

We define the minimal action needed to connect  $\rho^0 \in \mathcal{P}(G)$  to  $\rho^1 \in \mathcal{P}(G)$  to be

$$\frac{1}{2} \mathcal{W}_g^2(\rho^0, \rho^1) := \inf \left\{ \int_0^1 \mathcal{H}_g(\rho, \phi) dt \mid \dot{\rho} = \nabla_\phi \mathcal{H}_g(\rho, \phi), \rho(0) = \rho^0, \rho(1) = \rho^1 \right\}. \quad (1.3)$$

Making appropriate assumptions on  $g$ , this infimum will be shown to coincide with

$$\inf_{(\rho, v)} \left\{ \int_0^1 (v, v)_\rho dt \mid \dot{\rho} + \operatorname{div}_\rho(v) = 0, \rho(0) = \rho^0, \rho(1) = \rho^1 \right\} \quad (1.4)$$

and  $\mathcal{W}_g$  will be shown to be a metric on  $\mathcal{P}(G)$ . Note that for a well-chosen sequence  $(\phi_k)_k \subset \mathbb{R}^n$  whose norm tends to  $\infty$ , we may have that  $(\mathcal{H}_g(\rho, \phi_k))_k$  is identically null and so,  $\mathcal{H}_g(\rho, \cdot)$  is not coercive. This makes it a harder task to use direct methods of the calculus of variations to assert existence of a minimizer in (1.3). To circumvent this obstacle, we instead use the equivalent formulation (1.4) and resort to identifying a dual to (1.4).

Any minimizer  $(\rho, \phi)$  of (1.3) such that  $\rho, \phi \in W^{1,2}(0, 1; \mathbb{R}^n)$  and the range of  $\rho$  does not intersect the boundary of  $\mathcal{P}(G)$  satisfies Hamilton's equations

$$\dot{\rho} = \nabla_\phi \mathcal{H}_g(\rho, \phi), \quad \dot{\phi} = -\nabla_\rho \mathcal{H}_g(\rho, \phi). \quad (1.5)$$

This is a Hamiltonian system which reads off

$$\dot{\rho}_i + \sum_{j \in N(i)} \sqrt{\omega_{ij}} (\phi_j - \phi_i) g(\rho_i, \rho_j) = 0, \quad \dot{\phi}_i + \frac{1}{2} \sum_{j \in N(i)} \omega_{ij} \partial_1 g(\rho_i, \rho_j) (\phi_i - \phi_j)^2 = 0. \quad (1.6)$$

Here,  $N(i) := \{j \in V \mid (i, j) \in E\}$  denotes the neighborhood of a vertex  $i \in V$  and  $\partial_1 g$  denotes the partial derivative of  $g$  with respect to its first variable.

Hardly enough, even if  $\rho^0, \rho^1$  are chosen in the interior of  $\mathcal{P}(G)$ , a minimizer  $(\rho, \phi)$  of (1.3) may be such that the range of  $\rho$  intersects the boundary of  $\mathcal{P}(G)$ , unless (cf. [20])

$$C_g := \int_0^1 \frac{dr}{\sqrt{g(r, 1-r)}} = \infty. \quad (1.7)$$

The condition (1.7) precisely forces  $\mathcal{W}_g$  to assume infinite values and so, it cannot be a metric on the whole set  $\mathcal{P}(G)$ . When  $C_g < \infty$ , we endeavor to identify the appropriate substitute of (1.6), by characterizing minimizers of (1.4), even when  $\rho((0, 1))$  intersects the boundary of  $\mathcal{P}(G)$ .

We define a Poincaré function  $\gamma_P : \mathcal{P}(G) \rightarrow \mathbb{R}$ . It is a concave function which is strictly positive in the interior of  $\mathcal{P}(G)$  but may remain positive on a subset of the boundary of  $\mathcal{P}(G)$ . When  $\gamma_P(\rho^0), \gamma_P(\rho^1) > 0$  we show that  $(\rho, \nu)$  minimizes (1.4) if and only if there exists  $\lambda \in \text{BV}_{\text{loc}}(0, 1; \mathbb{R}^n)$  such that  $-\dot{\lambda}_i$  is a Borel regular measure,

$$g_{ij}(\rho) \left[ v_{ij} - \sqrt{\omega_{ij}}(\lambda_i - \lambda_j) \right] = 0 \quad \forall (i, j) \in E,$$

$$0 = H(\dot{\lambda}^{\text{abs}}, \nabla_G \lambda) = (\dot{\lambda}^{\text{abs}}, \rho) + \frac{1}{2} \|\nabla_G \lambda\|_{\rho}^2 \quad \mathcal{L}^1 \text{ a.e.} \quad (1.8)$$

and

$$0 = H_0\left(\frac{d\dot{\lambda}^{\text{sing}}}{d\nu}\right) = \left(\frac{d\dot{\lambda}^{\text{sing}}}{d\nu}, \rho\right) \quad \nu \text{ a.e.} \quad (1.9)$$

Here,  $\dot{\lambda}^{\text{abs}}$  is the absolutely continuous part of  $\dot{\lambda}$ ,  $\dot{\lambda}^{\text{sing}}$  the singular part of  $\dot{\lambda}$ , and  $\nu$  is any non-negative measure such that  $\nu$  and  $\mathcal{L}^1|_{(0,1)}$  are mutually singular and  $|\dot{\lambda}^{\text{sing}}| \ll \nu$ . We have defined

$$H(a, b) = \sup_{\rho \in \mathcal{P}(G)} \left\{ (a, \rho) + \frac{1}{2} \|b\|_{\rho}^2 \right\} \quad \forall (a, b) \in \mathbb{R}^n \times S^{n \times n} \quad (1.10)$$

and set

$$H_0(a) = \max_{1 \leq i \leq n} a_i \quad \forall a \in \mathbb{R}^n.$$

There is a relation between  $H_0$  and the recession function of  $H$  since

$$\lim_{l \rightarrow \infty} \frac{H(la, lb)}{l} = \begin{cases} H_0(a) & \text{if } b = 0 \\ \infty & \text{if } b \neq 0. \end{cases}$$

What seems surprising at a first glance is that, even when  $\dot{\lambda}$  has no singular part, the expression in (1.8) is still not linear in  $\dot{\lambda}$ . This means the geodesics of minimal length are characterized by Hamilton–Jacobi equations in the form  $0 = H(\dot{\lambda}^{\text{abs}}, \nabla_G \lambda)$ , with a non-linear dependence in  $\dot{\lambda}^{\text{abs}}$ . This is in contrast to what happens in the continuum setting, where geodesics of minimal length are characterized by Hamilton–Jacobi equations in the form  $\partial_t u + H_*(\nabla u) = 0$ , hence linear in  $\partial_t u$ . We pause here to draw the attention of the reader to [23] which proposes a class of Hamilton–Jacobi equations which can hardly be compared with (1.8) and (1.9).

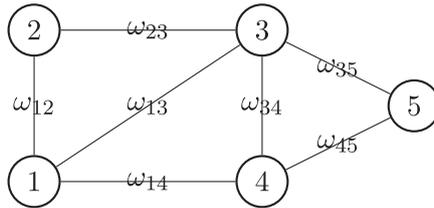


FIGURE 1. An illustration example of a weighted graph. Here  $V = \{1, 2, 3, 4, 5\}$ ,  $E = \{(1, 2), (1, 3), (1, 4), (2, 3), (3, 4), (3, 5), (4, 5)\}$ .

A comparison between our work and the work [20] by J. Maas becomes at this point unavoidable. There, the author considers an irreducible Markov kernel  $(K_{ij})_{ij}$  with a finite right invariant measure. Our hypothesis that  $(\omega_{ij})_{ij} := (K_{ij}\pi_i)_{ij}$  is symmetric, is equivalent to the requirement in [20] that  $K$  is reversible. When  $C_g = \infty$  [20] gave a remarkable characterization of the pairs for which  $\mathcal{W}_g(\rho^0, \rho^1) < \infty$ . The necessary and sufficient condition is that both  $\rho^0$  and  $\rho^1$  must have the same  $g$ -connected components (see Sect. 2 for the definition of  $g$ -connected components). As a consequence, if  $(\rho, v)$  is a minimizer in (1.4) then the  $g$ -connected components of  $\rho(t)$  are independent of  $t \in (0, 1)$  and they coincide with those of  $\rho^0$ . The search of paths of minimal actions in (1.4) reduces then to a finite collection of searches of paths of minimal actions which are known to be entirely contained in the interior of simplexes. In this case, the Euler–Lagrange equations are obtained by standard arguments.

In this paper, we assume that  $C_g < \infty$  and so, our study of geodesics of minimal norms complements that in [20]. We further assume that  $g$  is concave, 1-homogeneous, positive in  $(0, \infty)^2$  and  $C^\infty$  in this open set. These assumptions, while facilitating our study, still encompass a large number of metrics, useful in applications. The class of functions  $g$  we choose are motivated by studies [7–9, 20, 22], which recently appeared in the literature. The main purpose of the paper is to give a necessary and sufficient condition for being a geodesic in the set of probability measures on a graph endowed with the metric  $\mathcal{W}_g$ , without ruling out paths which touch the boundary.

The study in this paper will be little more than a disappointment if the set of geodesics starting and ending in the interior of  $\mathcal{P}(G)$  would never intersect the boundary of  $\mathcal{P}(G)$ . Unlike the study in [20], Proposition 3.11 supports the fact that when  $C_g < \infty$  then the set of such geodesics is not empty. Another feature of the condition  $C_g < \infty$  is that if  $\rho : [0, 1] \rightarrow \mathcal{P}(G) \setminus \mathcal{P}_0(G)$  is a geodesic of minimal length, then the  $g$ -connected components of  $\rho(t)$  need not be time independent (*cf.* Prop. 3.8), unlike the case when  $C_g = \infty$  [20]. One could combine Propositions 3.8 and 3.11 to construct more intricate geodesics which intersect the boundary of  $\mathcal{P}(G)$ .

The paper is organized as follows. In Section 2, we introduce the notation used in the paper. Section 3 contains preliminary remarks. For instance there, we comment on the sufficient condition for  $\mathcal{W}_g$  to assume only finite values. In Section 4, we show the existence of geodesics of minimal norms. Sections 5 and 6 contain ingredients we later use in Section 7 to characterize the geodesics of minimal path through a dual formulation.

## 2. NOTATION

We denote the one-dimensional Lebesgue measure by  $\mathcal{L}^1$  and denote the set of skew-symmetric  $n \times n$  matrices as  $S^{n \times n}$ . Let  $G = (V, E, \omega)$  denote an undirected graph of vertices  $V = \{1, \dots, n\}$  and edges  $E$ , with a weighted metric  $\omega = (\omega_{ij})$  given by an  $n$  by  $n$  symmetric matrix with non-negative entries  $\omega_{ij}$  and such that  $\omega_{ij} > 0$  if  $(i, j) \in E$ . An example is provided in Figure 1.

For simplicity, assume that the graph is connected and is simple, with no self-loops or multiple edges.

**Functions on a graph.** It is customary to identify a function  $\phi : V \rightarrow \mathbb{R}$  with a vector  $\phi = (\phi_i)_{i=1}^n \subset \mathbb{R}^n$ . We use the standard *inner product* on  $\mathbb{R}^n$ :

$$(\phi, \tilde{\phi}) := \sum_{i=1}^n \phi_i \tilde{\phi}_i, \quad \forall \phi, \tilde{\phi} \in \mathbb{R}^n.$$

**Vector fields and gradient operator.** A vector field  $m$  on  $G$  is a *skew-symmetric matrix* on the edges set  $E$ , denoted by  $m$ :

$$m := (m_{ij})_{(i,j) \in E}, \quad \text{with } m_{ij} = -m_{ji}.$$

Special elements of  $S^{n \times n}$  are the so-called potential vector fields which are *discrete gradients* of functions  $\phi$  on  $V$ , denoted  $\nabla_G \phi$  and defined as

$$\nabla_G \phi := \sqrt{\omega_{ij}}(\phi_i - \phi_j)_{(i,j) \in E}.$$

**The range and kernel of the gradient operator.** We denote by  $R(\nabla_G)$  the range of  $\nabla_G$  and by  $\mathbf{1} \in \mathbb{R}^n$  the vector whose entries are all equal to 1. Since  $G$  is connected, the kernel of  $\nabla_G$  is the one-dimensional space spanned by  $\mathbf{1}$ . The orthogonal in  $\mathbb{R}^n$  of the latter space is  $\ker(\nabla_G)^\perp$ , the set of  $h \in \mathbb{R}^n$  such that  $\sum_{i=1}^n h_i = 0$ .

**G-Divergence of vector field.** The *divergence* operator associates to any vector field  $m$  on  $G$  a function on  $V$  defined by

$$\nabla_G \cdot (m) = \text{div}_G(m) := \left( \sum_{j \in N(i)} \sqrt{\omega_{ij}} m_{ji} \right)_{i=1}^n.$$

**Set of probability measures and its boundary.** We identify  $\mathcal{P}(G)$ , the set of probability measures on  $V$ , with a simplex as follows

$$\mathcal{P}(G) = \left\{ \rho = (\rho_i)_{i=1}^n \subset [0, 1]^n \mid \sum_{i=1}^n \rho_i = 1 \right\}.$$

Let  $\mathcal{P}_0(G) := \mathcal{P}(G) \cap (0, 1)^n$  denote the interior of  $\mathcal{P}(G)$ . The boundary of  $\mathcal{P}(G)$  is  $\mathcal{P}(G) \setminus \mathcal{P}_0(G)$ .

**The set  $\mathcal{C}(\rho^0, \rho^1)$  of paths connections probability measures.** Given  $\rho^0, \rho^1 \in \mathcal{P}(G)$ , we denote as  $\mathcal{C}(\rho^0, \rho^1)$  the set of pairs  $(\rho, m)$  such that

$$\rho \in H^1(0, 1; \mathcal{P}(G)), \quad m \in L^2(0, 1; S^{n \times n}), \quad (\rho(0), \rho(1)) = (\rho^0, \rho^1)$$

and

$$\dot{\rho}_i + \sum_{j \in N(i)} \sqrt{\omega_{ij}} m_{ji} = 0, \quad \text{in the weak sense on } (0, 1). \quad (2.1)$$

Throughout this paper  $g : [0, \infty) \times [0, \infty) \rightarrow \mathbb{R}_+$  satisfies the following assumptions:

(H-i)  $g$  is continuous on  $[0, \infty) \times [0, \infty)$  and is of class  $C^\infty$  on  $(0, \infty) \times (0, \infty)$ ;

(H-ii)  $g(r, s) = g(s, r)$  for any  $s, r \in \mathbb{R}_+$ ;

- (H-iii)  $g(r, s) > 0$  for any  $r, s \in (0, \infty)$ ;  
(H-iv)  $g(\lambda r, \lambda s) = \lambda g(r, s)$  for any  $\lambda, s, r \in (0, \infty)$ ;  
(H-v)  $g$  is concave.

We extend  $g$  by setting its value to be  $-\infty$  outside  $[0, \infty)^2$ , to obtain a function on  $\mathbb{R}^2$  which we still denote  $g$ . Observe that the extension is concave and upper semicontinuous. We define

$$g_{ij}(\rho) = g(\rho_i, \rho_j) \quad \forall \rho \in \mathbb{R}^n, \quad \forall i, j \in V.$$

**A constant depending solely of  $g$ .** Since  $g$  is continuous on the compact set  $[0, 1]^2$ ,

$$\epsilon_0(g) := \sup_{r, s > 0} g\left(\frac{r}{r+s}, \frac{s}{r+s}\right)$$

is a finite number.

**The Hilbert spaces  $\mathbb{H}_\rho$ .** If  $\rho \in \mathcal{P}(G)$ , we say that  $v, \tilde{v} \in S^{n \times n}$  are  $\rho$ -equivalent if (1.1) holds, which means  $v_{ij} = \tilde{v}_{ij}$  whenever  $g_{ij}(\rho) > 0$ . We denote by  $\mathbb{H}_\rho$  the set of classes of equivalence. This is a Hilbert space when endowed with the discrete *inner product* and the *discrete norm*

$$(v, \tilde{v})_\rho := \frac{1}{2} \sum_{(i,j) \in E} v_{ij} \tilde{v}_{ij} g_{ij}(\rho), \quad \|v\|_\rho = \sqrt{(v, v)_\rho} \quad \forall v, \tilde{v} \in S^{n \times n}.$$

Here, the coefficient  $1/2$  accounts for the fact that whenever  $(i, j) \in E$  then  $(j, i) \in E$ . Similarly, if  $m, \tilde{m} \in S^{n \times n}$  we set

$$(m, \tilde{m}) := \frac{1}{2} \sum_{(i,j) \in E} m_{ij} \tilde{m}_{ij}, \quad \|m\|^2 = (m, m).$$

**The tangent spaces and the projection operator  $\pi_\rho$ .** We denote as  $T_\rho \mathcal{P}(G)$  the closure of the range of  $\nabla_G$  in  $\mathbb{H}_\rho$ . We refer to  $T_\rho \mathcal{P}(G)$  as the tangent space to  $\mathcal{P}(G)$  at  $\rho$ . Using an elementary Hilbert space projection argument, given  $v \in \mathbb{H}_\rho$  there exists a unique  $\pi_\rho(v) \in T_\rho \mathcal{P}(G)$  that minimizes  $\|v - \cdot\|_\rho$  over  $T_\rho \mathcal{P}(G)$ . It is characterized by the property

$$(v - \pi_\rho(v), w)_\rho = 0 \quad \forall w \in T_\rho \mathcal{P}(G). \quad (2.2)$$

**The divergence operator.** The operator  $\nabla_G : \mathbb{R}^n \rightarrow \mathbb{H}_\rho$  admits an adjoint  $-\text{div}_\rho : \mathbb{H}_\rho \rightarrow \mathbb{R}^n$  given by

$$\text{div}_\rho(v) = \left( \sum_{j \in N(i)} \sqrt{\omega_{ij}} v_{ji} g_{ij}(\rho) \right)_{i=1}^n \quad \forall v \in S^{n \times n}.$$

We call  $\text{div}_\rho$  the divergence operator. Note the integration by parts formula:

$$(\nabla_G \phi, v)_\rho = -(\phi, \text{div}_\rho(v)). \quad (2.3)$$

Let  $\mathcal{H}_g$  be the Hamiltonian defined in (1.2). Observe that

$$\text{div}_\rho(\nabla_G \phi) = -\nabla_\phi \mathcal{H}_g(\rho, \phi). \quad (2.4)$$

**The Monge–Kantorovich metric on  $G$ .** The square of 2-Monge–Kantorovich metric which measures the square distance between  $\rho^0 \in \mathcal{P}(G)$  and  $\rho^1 \in \mathcal{P}(G)$  is

$$\mathcal{W}_g^2(\rho^0, \rho^1) := \inf_{(\rho, v)} \left\{ \int_0^1 (v, v)_\rho dt \mid \dot{\rho} + \operatorname{div}_\rho(v) = 0, \rho(0) = \rho^0, \rho(1) = \rho^1 \right\}. \quad (2.5)$$

Here, the infimum is performed over the set of pairs  $(\rho, v)$  such that  $\rho \in H^1(0, 1; \mathbb{R}^n)$ ,  $v : [0, 1] \rightarrow S^{n \times n}$  is measurable.

**Connected components.** Let  $\rho \in \mathcal{P}(G)$ . We say that  $i, j \in V$  are  $g$ -connected if there are integers  $i_1, i_2, \dots, i_k \in V$  such that  $i_1 = i$ ,  $i_k = j$ ,  $(i_l, i_{l+1}) \in E$  for  $l = 1, \dots, k-1$  and

$$g_{i_1 i_2}(\rho) \cdots g_{i_{k-1} i_k}(\rho) > 0.$$

The largest  $g$ -connected set containing  $i$  is called the  $g$ -connected component of  $i$ . The  $g$ -connected components of  $\rho$  form a partition of a subset of  $V$ .

**Poincaré functions on graphs.** We define the Poincaré function  $\gamma_P$  on  $G$  as

$$\gamma_P(\rho) = \inf_{\beta} \left\{ \frac{1}{2} \sum_{(i,j) \in E} g_{ij}(\rho) \omega_{ij} (\beta_i - \beta_j)^2 \mid \sum_{i=1}^n \beta_i = 0, \sum_{i=1}^n \beta_i^2 = 1 \right\} \quad \forall \rho \in \mathcal{P}(G).$$

**Action.** Consider the lower semicontinuous convex function  $f : \mathbb{R}^2 \rightarrow [0, \infty]$  defined as

$$f(t, s) = \begin{cases} \frac{s^2}{t} & \text{if } t > 0 \\ 0 & \text{if } s = t = 0 \\ \infty & \text{otherwise.} \end{cases} \quad (2.6)$$

Observe that if  $t \geq 0$  and  $\mu \in \mathbb{R}$  then

$$2\mu s < f(t, s) + \mu^2 t \quad (2.7)$$

unless  $t\mu = s$  in which case equality holds.

For  $\rho \in \mathbb{R}^n$  and  $m \in S^{n \times n}$ , we define

$$F(\rho, m) = \frac{1}{2} \sum_{(i,j) \in E} f(g_{ij}(\rho), m_{ij}).$$

If  $\rho \in L^2(0, 1; \mathbb{R}^n)$  and  $m \in L^2(0, 1; S^{n \times n})$  we define the action

$$\mathcal{A}(\rho, m) = \frac{1}{2} \int_0^1 F(\rho, m) dt.$$

Let  $H : \mathbb{R}^n \times S^{n \times n} \rightarrow \mathbb{R}$  denote the Hamiltonian defined as in (1.10).

In the remainder of the paper, unless the contrary is explicitly stated, we assume that

$$C_g := \int_0^1 \frac{dr}{\sqrt{g(r, 1-r)}} < +\infty. \quad (2.8)$$

**Example 2.1.** Examples satisfying (H-i)–(H-v) and (2.8) include  $g(r, s) = \frac{r+s}{2}$ . Other examples (which appeared in [20]) are

$$g(r, s) = \int_0^1 r^{1-t} s^t dt = \begin{cases} \frac{r-s}{\log r - \log s} & \text{if } r \neq s \\ 0 & \text{if } r = 0 \text{ or } s = 0 \\ r & \text{if } r = s, \end{cases}$$

and

$$g(r, s) = \begin{cases} 0 & \text{if } r = 0 \text{ or } s = 0 \\ \frac{1}{\frac{1}{r} + \frac{1}{s}} & \text{otherwise.} \end{cases}$$

### 3. PRELIMINARIES

Throughout the section, we use the same notation as in Section 2 and assume (2.8) holds. In the section, we showed that any  $\rho^0, \rho^1 \in \mathcal{P}(G)$  can be connected by a feasible path if  $C_g < \infty$ . Moreover, we construct a path such that the end points are in the interior of  $\mathcal{P}(G)$  while the range intersects the boundary of  $\mathcal{P}(G)$ .

**Lemma 3.1.** *The Poincaré function  $\gamma_P : \mathcal{P}(G) \rightarrow \mathbb{R}$  is concave.*

*Proof.* Note  $\gamma_P$  is obtained by taking the infimum of concave functions of  $\rho$ . □

**Lemma 3.2.** *If  $\rho \in \mathcal{P}(G)$ ,  $\lambda \in \mathbb{R}^n$ ,  $na = \sum_{j=1}^n \lambda_j$  and  $\tilde{\lambda}_i := \lambda_i - a$  then*

$$\|\nabla_G \tilde{\lambda}\|_\rho^2 \geq \gamma_P(\rho) \|\tilde{\lambda}\|^2.$$

*Proof.* Set

$$\beta_i := \frac{\tilde{\lambda}_i}{\sqrt{\sum_{j=1}^n \tilde{\lambda}_j^2}}.$$

Then

$$\sum_{i=1}^n \beta_i = \frac{\sum_{i=1}^n \tilde{\lambda}_i}{\sqrt{\sum_{j=1}^n \tilde{\lambda}_j^2}} = 0 \text{ and } \sum_{i=1}^n \beta_i^2 = \frac{\sum_{i=1}^n \tilde{\lambda}_i^2}{\sum_{j=1}^n \tilde{\lambda}_j^2} = 1.$$

The desired inequality follows from the definition of  $\gamma_P$ . □

**Proposition 3.3.** *Suppose  $\rho \in \mathcal{P}(G)$  has only one  $g$ -connected component which is the whole set  $V$ . Then the range of  $\nabla_G$  is a closed subset of  $\mathbb{H}_\rho$  and so, it is  $T_\rho \mathcal{P}(G)$ .*

*Proof.* Suppose  $(\phi^k)_k \subset \mathbb{R}^n$  is such that  $(\nabla_G \phi^k)_k$  converges to  $v$  in  $\mathbb{H}_\rho$ . We are to show that  $v \in R(\nabla_G)$ . For any  $e \in V \setminus \{1\}$ , there exist  $e_1, \dots, e_l \in V$  such that  $e_1 = 1$ ,  $e_l = e$ ,  $(e_j, e_{j+1}) \in E$  and  $g_{e_j e_{j+1}}(\rho) > 0$  for any  $j \in \{1, \dots, l-1\}$ . We have for any  $j \in \{1, \dots, l-1\}$

$$\lim_{k \rightarrow \infty} \left( \sqrt{\omega_{e_j e_{j+1}}} (\phi_{e_j}^k - \phi_{e_{j+1}}^k) - v_{e_j e_{j+1}} \right)^2 g_{e_j e_{j+1}}(\rho) = 0. \quad (3.1)$$

Replacing  $\varphi_{e_j}^k$  by  $\varphi_{e_j}^k - \varphi_1^k$  if necessary, we may assume without loss of generality that  $\varphi_1^k = 0$ .

Setting  $j = 1$  in (3.1), we obtain that  $(\phi_{e_2}^k)_k$  converges to  $\frac{-v_{e_1 e_2}}{\sqrt{\omega_{e_1 e_2}}}$ . Setting  $\phi_{e_1} = 0$ ,  $\phi_{e_2} = \frac{-v_{e_1 e_2}}{\sqrt{\omega_{e_1 e_2}}}$  we inductively obtain

$$\sqrt{\omega_{e_j e_{j+1}}} \phi_{e_{j+1}} := \lim_{k \rightarrow \infty} \sqrt{\omega_{e_j e_{j+1}}} \phi_{e_{j+1}}^k = \lim_{k \rightarrow \infty} \sqrt{\omega_{e_j e_{j+1}}} \phi_{e_j}^k - v_{e_j e_{j+1}} = \sqrt{\omega_{e_j e_{j+1}}} \phi_{e_j} - v_{e_j e_{j+1}}$$

for any  $j \in \{2, \dots, l-1\}$ . This is sufficient to verify  $v = \nabla_G \phi$ .  $\square$

**Lemma 3.4.** *Assume that  $\rho \in \mathcal{P}(G)$ . Then  $\rho$  has only one  $g$ -connected component which is the whole set  $V$  iff  $\gamma_P(\rho) > 0$ .*

*Proof.* Assume that  $\rho$  has only one  $g$ -connected component which is the whole set  $V$ . Then for any  $e \in V \setminus \{1\}$  there exist  $e_1, \dots, e_l \in V$  such that  $e_1 = 1$ ,  $e_l = e$ ,  $(e_j, e_{j+1}) \in E$  and  $g_{e_j e_{j+1}}(\rho) > 0$  for any  $j \in \{1, \dots, l-1\}$ . Suppose that  $\gamma_P(\rho) = 0$ , i.e. there exists  $\beta \in \mathbb{R}^n$  such that  $\sum_{i=1}^n \beta_i = 0$ ,  $\sum_{i=1}^n \beta_i^2 = 1$  and

$$\sum_{(i,j) \in E} g_{ij}(\rho) \omega_{ij} (\beta_i - \beta_j)^2 = 0. \quad (3.2)$$

(3.2) implies that

$$0 = \sum_{(i,j) \in E} g_{ij}(\rho) \omega_{ij} (\beta_i - \beta_j)^2 \geq \sum_{j=1}^{l-1} g_{e_j e_{j+1}}(\rho) \omega_{e_j e_{j+1}} (\beta_{e_j} - \beta_{e_{j+1}})^2 = 0$$

and, thus,  $\beta_{e_1} = \beta_{e_2} = \dots = \beta_{e_l}$ . Since  $e$  is arbitrary, we have  $\beta_1 = \beta_2 = \dots = \beta_n = 0$ . This is in contradiction with the fact that  $\sum_{i=1}^n \beta_i^2 = 1$ .

Suppose that  $\gamma_P(\rho) > 0$ . We want to prove that  $\rho$  has only one  $g$ -connected component which is the whole set  $V$ . If not, there exist  $i_1, j_1 \in V$  such that  $i_1$  and  $j_1$  are not in the same  $g$ -connected component. Let  $V_{i_1} = \{i_1, i_2, \dots, i_k\}$  and  $V_{j_1} = \{j_1, j_2, \dots, j_{\tilde{k}}\}$  be the  $g$ -connected components of  $i_1$  and  $j_1$ , respectively. Set  $\beta_i = 0$  whenever  $i \in V \setminus (V_1 \cup V_2)$ . Otherwise, set

$$\beta_{i_1} = \beta_{i_2} = \dots = \beta_{i_k} = \sqrt{\frac{\tilde{k}}{k(k+\tilde{k})}} \text{ and } \beta_{j_1} = \beta_{j_2} = \dots = \beta_{j_{\tilde{k}}} = -\sqrt{\frac{k}{\tilde{k}(k+\tilde{k})}}.$$

Then we have  $\sum_{i=1}^n \beta_i = 0$ ,  $\sum_{i=1}^n \beta_i^2 = 1$  and (3.2) holds. This is at variance with the fact that  $\gamma_P(\rho) > 0$ .  $\square$

**Lemma 3.5.** *We assert the following*

- (i) *The function  $F$  is convex and lower semicontinuous on  $\mathcal{P}(G) \times S^{n \times n}$ .*
- (ii) *Suppose  $m, b \in S^{n \times n}$  and  $\rho \in \mathcal{P}(G)$  are such that  $m_{ij} = 0$  whenever  $g_{ij}(\rho) = 0$ . Then*

$$F(\rho, m) + \|b\|_\rho^2 > 2(m, b),$$

*unless  $m_{ij} = g_{ij}(\rho) b_{ij}$  for all  $(i, j) \in E$ , in which case equality holds.*

*Proof.* (i) Since  $g$  is concave,  $f$  is convex and  $f(\cdot, s)$  is monotone non-increasing,  $(\rho, m) \rightarrow f(g_{ij}(\rho), m_{ij})$  is convex and so, the function  $F$  is convex. One checks that  $(\rho, m) \rightarrow f(g_{ij}(\rho), m_{ij})$  is lower semicontinuous. Thus,  $F$  is a convex and lower semicontinuous as a sum of convex, lower semicontinuous functions.

(ii) is a direct consequence of (2.7).  $\square$

**Lemma 3.6.** *Let  $\mathcal{H}_g$  be as in (1.2) and let  $\partial_i g$  denote derivative of  $g$  with respect to the  $i$ -th variable for  $i = 1, 2$ .*

(i) If  $r, s > 0$  then

$$\partial_1 g(r, s) = \partial_2 g(s, r) \quad \text{and} \quad (\nabla g(r, s), (r, s)) = g(r, s).$$

(ii) For any  $\rho \in [0, \infty)^n$  and  $\phi \in \mathbb{R}^n$  we have

$$(\nabla_\phi \mathcal{H}_g(\rho, \phi), \phi) = 2\mathcal{H}_g(\rho, \phi).$$

(iii) For any  $\rho \in (0, \infty)^n$  and  $\phi \in \mathbb{R}^n$  we have

$$(\nabla_\rho \mathcal{H}_g(\rho, \phi), \rho) = \mathcal{H}_g(\rho, \phi).$$

*Proof.* Recall that  $g$  has been extended to an upper semicontinuous on  $\mathbb{R}^2$ , which we still denote as  $g$ .

(i) Since  $g(r, s) = g(s, r)$ , differentiating, we obtain the first identity in (i). Let  $G^*$  denote the Legendre transform of the convex, degree 1-homogeneous function  $G = -g$ . If  $\alpha, \beta \in \mathbb{R}$  then  $G^*(\alpha, \beta) = \infty$ , unless  $\alpha\bar{r} + \beta\bar{s} \leq G(\bar{r}, \bar{s})$  for all  $\bar{r}, \bar{s} \in \mathbb{R}$ , in which case  $G^*(\alpha, \beta) = 0$ . If  $r, s > 0$  then  $G$  is differentiable at  $(r, s)$  and so, setting  $(\alpha, \beta) = \nabla G(r, s)$  we have

$$(\nabla G(r, s), (r, s)) = \alpha r + \beta s = G(r, s) + G^*(\alpha, \beta) = G(r, s).$$

This completes the verification of (i).

(ii) We have

$$\frac{\partial \mathcal{H}_g}{\partial \phi_i}(\rho, \phi) = \sum_{j \in N(i)} \omega_{ij} g_{ij}(\rho)(\phi_i - \phi_j).$$

We use this to verify that (ii) holds.

(iii) We use the first identity in (i) to infer

$$\frac{\partial \mathcal{H}_g}{\partial \rho_i}(\rho, \phi) = \frac{1}{2} \sum_{j \in N(i)} \omega_{ij} \partial_1 g(\rho_i, \rho_j)(\phi_i - \phi_j)^2.$$

This, together with the second identity in (i) complete the verification of (iii).  $\square$

**Proposition 3.7.** *For any  $\rho^0, \rho^1 \in \mathcal{P}(G)$  there is a path  $(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)$  such that  $\mathcal{A}(\rho, m) < +\infty$ . In other words, we have a feasible path in  $\mathcal{C}(\rho^0, \rho^1)$ .*

*Proof.* Let  $C_g$  be as defined in (2.8). Changing variables, we infer

$$C_g = \sqrt{2} \int_0^1 \frac{dr}{\sqrt{g(1+r, 1-r)}}.$$

This new formulation of  $C_g$  allows us to attribute this Proposition to [20] even if one may think our setting is a variant of his or the other way around. For completeness we lay down the main arguments supporting our statement. Before we start, we outline the strategy of the proof: when  $C_g < \infty$ , we first prove that, if the cardinality of  $V$  is 2, there exists a finite energy path connecting any  $\rho^0, \rho^1 \in \mathcal{P}(G)$ . Then we extend the result to any connected graph by moving mass between two vertices in each step.

Let  $G_2 = (V_2, E_2, \omega_2)$  be a graph of vertices  $V_2 = \{1, 2\}$ , edges  $E_2 = \{(1, 2), (2, 1)\}$ , endow with the weight  $\omega_{12} = \omega_{21} > 0$ . Let  $\rho^0, \rho^1 \in \mathcal{P}(G_2)$ . To avoid trivialities, assume  $\rho_1^0 \neq \rho_1^1$ . Without loss of generality, assume  $\rho_1^0 < \rho_1^1$ . Note the strictly increasing function

$$\tau \rightarrow G(\tau) := \int_0^\tau \frac{dr}{\sqrt{g(r, 1-r)}}$$

has an inverse function  $G^{-1}$  which is differentiable. Set

$$C = G(\rho_1^1) - G(\rho_1^0), \quad \rho_1(t) = G^{-1}(G(\rho_1^0) + Ct), \quad \rho_2(t) = 1 - \rho_1(t).$$

Define  $m_{21} = -m_{12}$  through the identity

$$\sqrt{\omega_{12}m_{21}} = -\dot{\rho}_1.$$

Observe that the path  $t \rightarrow (\rho_1(t), 1 - \rho_1(t))$  connects  $(\rho_1^0, \rho_2^0)$  to  $(\rho_1^1, \rho_2^1)$  and

$$\omega_{12}m_{12}^2 = \omega_{12}m_{21}^2 = \left(\frac{d\rho_1}{dt}\right)^2 = C^2 g(\rho_1, 1 - \rho_1) \in L^1(0, 1).$$

By definition  $(\rho, m)$  satisfies (2.1). Check that  $2\omega_{12}\mathcal{F}(\rho, m) = C^2$ . This covers the case  $n = 2$ .

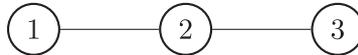
When  $n > 2$ , if  $\rho \in \mathcal{P}(G)$  is such that there is a feasible path in  $\mathcal{C}(\rho^0, \rho)$  and a feasible path in  $\mathcal{C}(\rho^1, \rho)$  then by concatenation, there is a feasible path in  $\mathcal{C}(\rho^0, \rho^1)$ . This means we may assume without loss of generality that  $\rho^1 = (0, \dots, 0, 1)$ . Let

$$V[\rho] := \{i \in \{1, \dots, n-1\} \mid \rho_i > 0\}.$$

If  $V[\rho^0] = \emptyset$ , then  $\rho^0 = \rho^1$  and so,  $(\rho, m) \equiv (\rho^0, 0)$  is a feasible path in  $\mathcal{C}(\rho^0, \rho^1)$ . Assuming that  $V[\rho^0] \neq \emptyset$ , one iteratively constructs a finite sequence  $\tilde{\rho}^0, \dots, \tilde{\rho}^{l_0}$  in  $\mathcal{P}(G)$  satisfying the following properties:

- (i)  $\tilde{\rho}^0 = \rho^0$  and  $\tilde{\rho}^{l_0} = \rho^1$ ;
- (ii) the cardinality of  $V[\tilde{\rho}^l]$  is strictly smaller than that of  $V[\tilde{\rho}^{l-1}]$  whenever  $l \leq l_0$ ;
- (iii) there is a feasible path in  $\mathcal{C}(\tilde{\rho}^{l-1}, \tilde{\rho}^l)$ . □

The next example features a geodesic path which connects two probabilities  $\rho^0, \rho^1 \in \mathcal{P}(G)$  which have distinct  $g$ -connected components.



In this example, we define  $g$  as follows:

$$g(r, s) = \begin{cases} \frac{r+s}{2} & \text{if } r, s \geq 0 \\ -\infty & \text{otherwise.} \end{cases} \quad (3.3)$$

Define

$$Q := \{(r_1, r_3) \in (0, 1)^2 \mid r_1 + r_3 < 1\}. \quad (3.4)$$

**Proposition 3.8.** *Let  $V = \{1, 2, 3\}$ ,  $E = \{(1, 2), (2, 3)\}$  and let  $\omega$  denote a  $2 \times 2$  symmetric matrix such that  $\omega_{12} = \omega_{23} = 1$ . Let  $G$  denote the weighted graph  $(V, E, \omega)$ . Let  $g$  be as in (3.3). Let  $\rho^0 = (0, 0, 1)$  and  $\rho^1 =$*

$(0, 1/2, 1/2)$  so that  $\rho^0$  and  $\rho^1$  lie on the boundary of  $\mathcal{P}(G)$ . Observe  $\rho^0$  has two  $g$ -connected components which are  $\{1\}$  and  $\{2, 3\}$ , and  $\rho^1$  has only one  $g$ -connected component which is  $\{1, 2, 3\}$ . We claim that any geodesic of minimal norm  $\rho : [0, 1] \rightarrow \mathcal{P}(G)$ , connecting  $\rho^0$  to  $\rho^1$  lies in the boundary of  $\mathcal{P}(G)$ . Furthermore, the  $g$ -connected components of  $\rho(t)$  are not constant in  $t$ .

*Proof.* By Theorem 4.5 there is  $(\rho, m)$  that minimizes  $\mathcal{A}$  over  $\mathcal{C}(\rho^0, \rho^1)$ .

In order to show the range of  $\rho$  lies in the boundary of  $\mathcal{P}(G)$ , it suffices to show that  $\rho_1(t) \equiv 0$ . To achieve that goal, it suffices to show that for any  $(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)$  such that  $\rho_1 \not\equiv 0$ , we can construct  $(\bar{\rho}, \bar{m}) \in \mathcal{C}(\rho^0, \rho^1)$  such that  $\bar{\rho}_1 \equiv 0$  and  $\mathcal{A}(\bar{\rho}, \bar{m}) < \mathcal{A}(\rho, m)$ . Let then assume  $(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)$  is such that  $\rho_1 \not\equiv 0$ . We have

$$\dot{\rho}_1 + m_{21} = 0, \quad \dot{\rho}_2 + m_{12} + m_{32} = 0, \quad \dot{\rho}_3 + m_{23} = 0. \quad (3.5)$$

Set

$$(\bar{\rho}_1, \bar{\rho}_2, \bar{\rho}_3) := (0, \rho_1 + \rho_2, \rho_3), \quad (\bar{m}_{12}, \bar{m}_{23}) := (0, m_{23}), \quad (\bar{m}_{21}, \bar{m}_{32}) := (0, m_{32}).$$

Note

$$\bar{\rho}(0) = \rho^0, \quad \bar{\rho}(1) = \rho^1 \quad \text{and} \quad \dot{\bar{\rho}}_1 + \bar{m}_{21} = 0.$$

We use (3.5) to infer

$$\dot{\bar{\rho}}_2 + \bar{m}_{12} + \bar{m}_{32} = \dot{\rho}_1 + \dot{\rho}_2 + \bar{m}_{32} = -m_{21} - m_{12} - m_{32} + m_{32} = 0.$$

Similarly,

$$\dot{\bar{\rho}}_3 + \bar{m}_{32} = \dot{\rho}_3 + m_{32} = 0.$$

Thus, we verified that  $(\bar{\rho}, \bar{m}) \in \mathcal{C}(\rho^0, \rho^1)$  and  $\bar{\rho}_1 \equiv 0$ . Note we cannot have  $m_{12} \equiv 0$  otherwise, we would have  $\dot{\rho}_1 \equiv 0$  which would imply  $\rho_1(t) = \rho(0) = 0$ . Therefore

$$2\mathcal{A}(\rho, m) = \int_0^1 \left( \frac{m_{12}^2}{g_{12}(\rho)} + \frac{m_{23}^2}{g_{23}(\rho)} \right) dt > \int_0^1 \frac{m_{23}^2}{g_{23}(\rho)} dt = \int_0^1 2 \frac{m_{23}^2}{\rho_2 + \rho_3} dt. \quad (3.6)$$

We have

$$2\mathcal{A}(\bar{\rho}, \bar{m}) = \int_0^1 \frac{\bar{m}_{23}^2}{g_{23}(\bar{\rho})} dt = \int_0^1 2 \frac{m_{23}^2}{\rho_1 + \rho_2 + \rho_3} dt \leq \int_0^1 2 \frac{m_{23}^2}{\rho_2 + \rho_3} dt. \quad (3.7)$$

By (3.6) and (3.7),  $\mathcal{A}(\rho, m) > \mathcal{A}(\bar{\rho}, \bar{m})$ . We conclude the proof of the proposition thanks to the observation that since  $\rho^0$  and  $\rho^1$  do not have the same  $g$ -connected components, the  $g$ -connected components of  $\rho^*(t)$  cannot be constant in  $t$ .  $\square$

**Remark 3.9.** Let  $G = (V, E, \omega)$  denote the weighted graph in Proposition 3.8 and let  $g$  denote the function used there. Suppose  $\rho^0, \rho^1 \in \mathcal{P}(G)$ ,  $(\rho, m)$  minimizes  $\mathcal{A}$  over  $\mathcal{C}(\rho^0, \rho^1)$ , and the range of  $\rho$  is entirely contained in the interior of  $\mathcal{P}(G)$ . Then using (3.5) we have

$$\mathcal{A}(\rho, m) = \int_0^1 \left( \frac{(\dot{\rho}_1)^2}{\rho_1 + \rho_2} + \frac{(\dot{\rho}_3)^2}{\rho_3 + \rho_2} \right) dt = \int_0^1 \mathcal{L}_0((\rho_1, \rho_3), (\dot{\rho}_1, \dot{\rho}_3)) dt =: \mathcal{A}_0(\rho_1, \rho_3)$$

where

$$\mathcal{L}_0(q, u) := \frac{u_1^2}{1 - q_3} + \frac{u_3^2}{1 - q_1}, \quad q = (q_1, q_3), \quad u = (u_1, u_3).$$

From  $\rho_1$  and  $\rho_3$  we recover  $\rho_2 = 1 - (\rho_1 + \rho_3)$ . We have that  $(\rho_1, \rho_3)$  minimizes  $\mathcal{A}_0$  over the set of  $(\rho_1, \rho_3) : [0, 1] \rightarrow Q$  where  $Q$  is given by (3.4).

**Proposition 3.10.** *Let  $\mathcal{H}_g$  be as in (1.2) and  $H$  be as in (1.10). Let  $\rho^0, \rho^1 \in \mathcal{P}(G)$  be such that  $(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)$ . Assume  $\lambda \in H^1(0, 1; \mathbb{R}^n)$  is such that  $H(\dot{\lambda}, \nabla_G \lambda) \leq 0$  almost everywhere.*

(i) *We have*

$$(\lambda(1), \rho^1) - (\lambda(0), \rho^0) \leq \mathcal{A}(\rho, m).$$

(ii) *Equality holds in (i) if and only if*

$$m_{ij} = g_{ij}(\rho)(\nabla_G \lambda)_{ij} \quad \forall (i, j) \in E, \quad H(\dot{\lambda}, \nabla_G \lambda) = (\rho, \dot{\lambda}) + \frac{1}{2} \|\nabla_G \lambda\|_\rho^2 = 0 \quad \text{a.e.}$$

(iii) *If the range of  $\rho$  is almost everywhere contained in  $(0, \infty)^n$  and  $(\rho, \lambda)$  satisfies almost everywhere the Hamiltonian system*

$$\begin{cases} \dot{\rho} &= \nabla_\phi \mathcal{H}_g(\rho, \lambda) \\ \dot{\lambda} &= -\nabla_\rho \mathcal{H}_g(\rho, \lambda), \end{cases} \quad (3.8)$$

*then equality holds in (i) and so,  $(\rho, m)$  minimizes  $\mathcal{A}$  over  $\mathcal{C}(\rho^0, \rho^1)$  where  $m_{ij} = g_{ij}(\rho)(\nabla_G \lambda)_{ij}$  for any  $(i, j) \in E$ .*

*Proof.* (i) We have

$$(\lambda(1), \rho^1) - (\lambda(0), \rho^0) = \int_0^1 ((\dot{\rho}, \lambda) + (\rho, \dot{\lambda})) dt = \int_0^1 (-(\operatorname{div}_G(m), \lambda) + (\rho, \dot{\lambda})) dt.$$

Integrating by parts and then using Lemma 3.5 (ii) in the subsequent identity, we conclude

$$\begin{aligned} (\lambda(1), \rho^1) - (\lambda(0), \rho^0) &= \int_0^1 \left( (m, \nabla_G \lambda) + (\rho, \dot{\lambda}) \right) dt \\ &\leq \int_0^1 \left( \frac{1}{2} F(\rho, m) + \frac{1}{2} \|\nabla_G \lambda\|_\rho^2 + (\rho, \dot{\lambda}) \right) dt \end{aligned} \quad (3.9)$$

$$\leq \int_0^1 \left( \frac{1}{2} F(\rho, m) + H(\dot{\lambda}, \nabla_G \lambda) \right) dt \leq \mathcal{A}(\rho, m). \quad (3.10)$$

This, verifies (i).

(ii) Note that equality holds in (i) if and only if equality hold in (3.9) and (3.10). Using Lemma 3.5 (ii), we conclude the proof of (ii).

(iii) Assume  $(\rho, \lambda)$  satisfies almost everywhere the Hamiltonian system (3.8). We use it and then use Lemma 3.6 to obtain

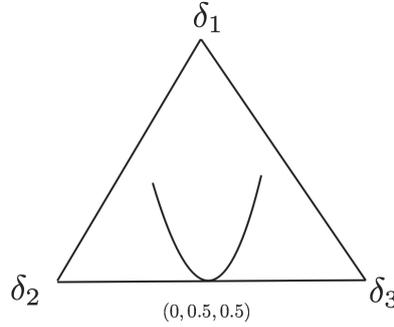
$$0 = (\rho, \dot{\lambda} + \nabla_{\rho} \mathcal{H}_g(\rho, \lambda)) = (\rho, \dot{\lambda}) + (\rho, \nabla_{\rho} \mathcal{H}_g(\rho, \lambda)) = (\rho, \dot{\lambda}) + \frac{1}{2} \|\nabla_G \lambda\|_{\rho}^2 \leq H(\dot{\lambda}, \nabla_G \lambda) \leq 0.$$

Thus,

$$0 = (\rho, \dot{\lambda}) + \frac{1}{2} \|\nabla_G \lambda\|_{\rho}^2 \leq H(\dot{\lambda}, \nabla_G \lambda).$$

Setting  $m_{ij} = g_{ij}(\rho)(\nabla_G \lambda)_{ij}$  for any  $(i, j) \in E$  we use (ii) to conclude the proof of (iii).  $\square$

We next construct a geodesic path which touches the boundary of probability simplex, even if the initial and terminal probability measures lie in the interior, see the following figure. In Proposition 3.11, we shall use the same weighted graph and the same metric as in Proposition 3.8.



**Proposition 3.11.** *Let  $G = (V, E, \omega)$  denote the weighted graph in Proposition 3.8 and let  $g$  denote the function used there. Let  $\mathcal{H}_g$  be as in (1.2) and  $H$  be as in (1.10). There exist  $\rho^0, \rho^1$  in the interior of  $\mathcal{P}(G)$  and there is a geodesic of minimal norm  $\rho : [0, 1] \rightarrow \mathcal{P}(G)$ , connecting  $\rho^0$  to  $\rho^1$  which intersects the boundary of  $\mathcal{P}(G)$ .*

*Proof.* The comments in Remark 3.9 led us to the following considerations which rely on the Lagrangian  $\mathcal{L}_0$  introduced there.

Set  $q := (q_1, q_3)$  and choose  $\delta \in (0, 0.1]$  such that the system of differential equations

$$\frac{d}{dt} \nabla_u \mathcal{L}_0(q, \dot{q}) = \nabla_q \mathcal{L}_0(q, \dot{q}) \quad \text{on } (-\delta, \delta), \quad (3.11)$$

together with the initial conditions

$$q(0) = (0, 0.5), \quad \dot{q}(0) = (0, 1) \quad (3.12)$$

has a unique solution. We have the conserved quantity

$$\frac{\dot{q}_1^2}{1 - q_3} + \frac{\dot{q}_3^2}{1 - q_1} = 1.$$

For  $\delta$  small enough, we have

$$|q_1| \leq 0.085, \quad 0.48 \leq q_3 \leq 0.62 \quad \text{on } [-\delta, \delta]. \quad (3.13)$$

By (3.11)

$$\ddot{q}_1 = -\frac{\dot{q}_1 \dot{q}_3}{1 - q_3} + \frac{1}{2} \frac{\dot{q}_3^2 (1 - q_3)}{(1 - q_1)^2} \quad \text{and} \quad \ddot{q}_3 = -\frac{\dot{q}_1 \dot{q}_3}{1 - q_1} + \frac{1}{2} \frac{\dot{q}_1^2 (1 - q_1)}{(1 - q_3)^2}. \quad (3.14)$$

We use (3.13) in (3.14) to obtain a constant  $C_1 > 0$  independent of  $\delta \in (0, 0.1)$  such that

$$|\dot{q}_1| + |\ddot{q}_1| + |\dot{q}_3| + |\ddot{q}_3| \leq C_1 \quad \text{on} \quad [-\delta, \delta]. \quad (3.15)$$

Differentiating the expressions in (3.14), we obtain explicit expressions of  $d^3 q_1/dt^3$  and  $d^3 q_3/dt^3$  in terms of  $q_1, q_3, \dot{q}_1, \dot{q}_3, \ddot{q}_1, \ddot{q}_3$ . We use the identities in (3.15) in these expressions to obtain a constant  $C > 0$  such that

$$|\dot{q}_1| + |\ddot{q}_1| + \left| \frac{d^3 q_1}{dt^3} \right| + |\dot{q}_3| + |\ddot{q}_3| + \left| \frac{d^3 q_3}{dt^3} \right| \leq C \quad \text{on} \quad [-\delta, \delta]. \quad (3.16)$$

By (3.12) and (3.14),  $\ddot{q}_1(0) = 0.25$ . This, together with (3.16) implies

$$\ddot{q}_1 \geq 0.25 - Ct \quad \forall t \in [-\delta, \delta].$$

Thus, choosing  $\delta_1$  strictly between 0 and  $\min\{\delta, 0.15C^{-1}\}$  we have

$$\ddot{q}_1 \geq 0.1 \quad \text{on} \quad [-\delta_1, \delta_1].$$

Since  $\dot{q}_1(0) = 0$  then for any  $t \in [-\delta_1, \delta_1]$ ,

$$q_1(t) \geq \frac{0.1t^2}{2} \quad \text{on} \quad [-\delta_1, \delta_1].$$

This, together with (3.13) yields

$$0.05t^2 \leq q_1(t) \leq 0.085, \quad 0.48 \leq q_3(t) \leq 0.62, \quad \text{and} \quad q_1(t) + q_3(t) \leq 0.705, \quad \forall t \in [-\delta_1, \delta_1]. \quad (3.17)$$

Setting

$$q_2 := 1 - q_1 - q_3 \quad \text{and} \quad \tilde{q} := (q_1, q_2, q_3).$$

(3.17) implies

$$\tilde{q} \in C^3([-\delta_1, \delta_1], \mathbb{R}^n), \quad \tilde{q}([-\delta_1, \delta_1] \setminus \{0\}) \subset \mathcal{P}_0(G), \quad \tilde{q}(0) \in \mathcal{P}(G) \setminus \mathcal{P}_0(G). \quad (3.18)$$

Define

$$l_1(t) = -\int_0^t \frac{\dot{q}_1^2}{(1 - q_3)^2} ds, \quad l_2(t) = l_1(t) - \frac{2\dot{q}_1}{1 - q_3}, \quad l_3(t) := l_2(t) + \frac{2\dot{q}_3}{1 - q_1}. \quad (3.19)$$

Observe

$$l_2 - l_1 = \frac{-2\dot{q}_1}{1 - q_3}, \quad l_3 - l_2 = \frac{2\dot{q}_3}{1 - q_1}. \quad (3.20)$$

Differentiating the expressions in (3.19) and using (3.20) we obtain

$$\dot{l}_1 = -\frac{1}{4}(l_1 - l_2)^2, \quad \dot{l}_2 = -\frac{1}{4}(l_1 - l_2)^2 - \frac{1}{4}(l_3 - l_2)^2, \quad \dot{l}_3 = -\frac{1}{4}(l_3 - l_2)^2.$$

Thus,

$$\dot{l} = -\nabla_\rho \mathcal{H}_g(\tilde{q}, l). \quad (3.21)$$

We have

$$\partial_{\phi_1} \mathcal{H}_g(\tilde{q}, l) = g(q_1, q_2)(l_1 - l_2) = \frac{q_1 + q_2}{2}(l_1 - l_2) = \frac{1 - q_3}{2}(l_1 - l_2).$$

This, combined with the first identity in (3.20) yields,

$$\partial_{\phi_1} \mathcal{H}_g(\tilde{q}, l) = \dot{q}_1.$$

Analogously, computing  $\partial_{\phi_2} \mathcal{H}_g$  and  $\partial_{\phi_3} \mathcal{H}_g$  and using (3.20) we obtain

$$\dot{q} = \nabla_\phi \mathcal{H}_g(\tilde{q}, l). \quad (3.22)$$

Set

$$\rho(s) := \tilde{q}(2\delta_1 s - \delta_1), \quad \lambda(s) := 2\delta_1 l(2\delta_1 s - \delta_1), \quad m_{ij} := g_{ij}(\rho)(\nabla_G \lambda)_{ij} \quad \forall (i, j) \in E.$$

By (3.18)

$$\rho \in C^3([0, 1], \mathbb{R}^n), \quad \rho\left([0, 1] \setminus \left\{\frac{1}{2}\right\}\right) \subset \mathcal{P}_0(G), \quad \rho\left(\frac{1}{2}\right) \in \mathcal{P}(G) \setminus \mathcal{P}_0(G). \quad (3.23)$$

By (3.22) and (3.21) we have

$$\dot{\rho} = \nabla_\phi \mathcal{H}_g(\rho, \lambda), \quad \dot{\lambda} = -\nabla_\rho \mathcal{H}_g(\rho, \lambda). \quad (3.24)$$

The latter identity implies

$$H(\dot{\lambda}, \nabla_G \lambda) = 0. \quad (3.25)$$

Combining (3.24) and (3.25), using Proposition 3.10, we obtain that  $(\rho, m)$  minimizes  $\mathcal{A}$  over  $\mathcal{C}(\rho(0), \rho(1))$ . We learn from (3.23) that the end points of  $\rho$  are in the interior of  $\mathcal{P}(G)$  while the range of  $\rho$  intersects the boundary of  $\mathcal{P}(G)$ .  $\square$

#### 4. EXISTENCE OF A MINIMIZER

In this section, we use the same notation as in Section 2 and assume (2.8) holds. The main purpose of the section is to prove (2.5) admits a minimizer. We start with the following Lemma.

**Lemma 4.1.** For any  $E_0 \subset E$  and  $(\rho, m) \in \mathbb{R}_+^n \times S^{n \times n}$

$$\epsilon_0(g) \sum_{k=1}^n \rho_k \sum_{(i,j) \in E_0} f(g_{ij}(\rho), m_{ij}) \geq \sum_{(i,j) \in E_0} m_{ij}^2.$$

*Proof.* Let  $(\rho, m) \in \mathbb{R}_+^n \times S^{n \times n}$ . Observe that to prove the lemma we only need to take into account  $(i, j) \in E_0$  such that  $m_{ij} \neq 0$ . In that case, we may only account for  $(i, j) \in E_0$  such that  $g_{ij}(\rho) > 0$ . We then have

$$g_{ij}(\rho) = g(\rho_i, \rho_j) \leq \epsilon_0(g)(\rho_i + \rho_j) \leq \epsilon_0(g) \sum_{k=1}^n \rho_k.$$

The desired inequality follows since  $\epsilon_0(g)f(g_{ij}(\rho), m_{ij}) \sum_{k=1}^n \rho_k \geq m_{ij}^2$ .  $\square$

**Remark 4.2.** Let  $h \in \ker(\nabla_G)^\perp$ , and let  $\rho^0, \rho^1 \in \mathcal{P}(G)$ .

- (i) If  $\phi, \tilde{\phi} \in \mathbb{R}^n$  are such that  $\nabla_G \phi = \nabla_G \tilde{\phi}$ , since  $G$  is connected we obtain  $a := \phi_i - \tilde{\phi}_i$  is independent of  $i \in V$  and so,

$$(\phi - \tilde{\phi}, h) = a \sum_{i=1}^n h_i = 0.$$

Hence, on  $R(\nabla_G)$ , the linear operator defined by  $L(\nabla_G \phi) := (h, \phi)$  is well defined. Since  $R(\nabla_G)$  is finite-dimensional,  $L$  is continuous and so, it admits a unique linear extension  $L_\rho : T_\rho \mathcal{P}(G) \rightarrow \mathbb{R}$ , which is in turn continuous.

- (ii) By the Riesz representation theorem, there exists a unique  $l_\rho(h) \in T_\rho \mathcal{P}(G)$  such that

$$L_\rho = (\cdot, l_\rho(h))_\rho.$$

By the fact that  $L_\rho(\nabla \phi) = (\phi, h)$  for every  $\phi \in \mathbb{R}^n$ , we have  $h + \operatorname{div}_\rho(l_\rho(h)) = 0$ .

Set

$$\mathcal{E}[\rho, v](h) := \frac{1}{2} \|v\|_\rho^2 - (v, l_\rho(h))_\rho \quad \forall v \in S^{n \times n}.$$

**Proposition 4.3.** Let  $m \in S^{n \times n}$ , let  $\rho \in \mathcal{P}(G)$  and let  $h \in \ker(\nabla_G)^\perp$ .

- (i) Then  $\nabla_G \cdot (m) \in \ker(\nabla_G)^\perp$ .  
 (ii) Then  $l_\rho(h)$  is the unique minimizer of  $\mathcal{E}[\rho, \cdot](h)$  over  $T_\rho \mathcal{P}(G)$  and over  $\mathbb{H}_\rho$ .  
 (iii) If  $w \in T_\rho \mathcal{P}(G)$  and  $h = -\operatorname{div}_\rho(w)$  then  $w = l_\rho(h)$ .  
 (iv) If  $h = -\nabla_G \cdot (m)$  then  $F(\rho, m) \geq \|l_\rho(h)\|_\rho^2$ .

*Proof.* (i) We use the fact that  $m_{ij} + m_{ji} = 0$  for any  $(i, j) \in E$  and that  $(i, j) \in E$  if  $(j, i) \in E$  to obtain (i).

(ii) Let  $v \in S^{n \times n}$  be such that  $v \neq \pi_\rho(v)$ . We have  $\|v\|_\rho > \|\pi_\rho(v)\|_\rho$  and by the characterization of  $\pi_\rho(v)$  in (2.2), we have  $(v, l_\rho(h))_\rho = (\pi_\rho(v), l_\rho(h))_\rho$ . Hence,

$$\mathcal{E}[\rho, v](h) > \frac{1}{2} \|\pi_\rho(v)\|_\rho^2 - (\pi_\rho(v), l_\rho(h))_\rho = \frac{1}{2} \|\pi_\rho(v) - l_\rho(h)\|_\rho^2 - \frac{1}{2} \|l_\rho(h)\|_\rho^2.$$

If  $\pi_\rho(v) \neq l_\rho(h)$ , we conclude that

$$\mathcal{E}[\rho, v](h) > -\frac{1}{2}\|l_\rho(h)\|_\rho^2 = \mathcal{E}[\rho, l_\rho(h)](h).$$

This concludes the proof of (ii).

(iii) Since

$$(\nabla_G \phi, l_\rho(h))_\rho = (\phi, h) = (\nabla_G \phi, w)_\rho \quad \forall \phi \in \mathbb{R}^n,$$

we have

$$(v, l_\rho(h) - w)_\rho = 0 \quad \forall v \in T_\rho \mathcal{P}(G),$$

which proves (iii).

(iv) Assume  $h = -\nabla_G \cdot (m)$  and set  $l_\rho(h) = w$ . If  $F(\rho, m) = \infty$  there is nothing to prove. Assume that  $F(\rho, m) < \infty$  so that  $m_{ij} = 0$  whenever  $g_{ij}(\rho) = 0$ . There is a unique vector field  $v$  such that  $g_{ij}(\rho)v_{ij} = m_{ij}$  and  $v_{ij} = 0$  whenever  $g_{ij}(\rho) = 0$ . We have

$$F(\rho, m) = \sum_{g_{ij}(\rho) > 0} \frac{m_{ij}^2}{2g_{ij}(\rho)} = \frac{1}{2} \sum_{g_{ij}(\rho) > 0} g_{ij}(\rho)v_{ij}^2 = \frac{1}{2} \sum_{(i,j) \in E} g_{ij}(\rho)v_{ij}^2 = \|v\|_\rho^2 \geq \|\pi_\rho(v)\|_\rho^2. \quad (4.1)$$

Since

$$h = -\nabla_G \cdot (m) = -\operatorname{div}_\rho(v) = -\operatorname{div}_\rho(\pi_\rho(v)),$$

(iii) implies  $\pi_\rho(v) = w$ . This, together with (4.1) proves (iv).  $\square$

**Remark 4.4.** The following remarks are needed in the paper.

(i) If  $\rho \in L^2(0, 1; \mathbb{R}^n)$  and  $m \in L^2(0, 1; S^{n \times n})$  are such that  $\mathcal{A}(\rho, m) < \infty$  then  $f(g_{ij}(\rho), m_{ij}) \in L^1(0, 1)$  and for any  $(i, j) \in E$ ,

$$\mathcal{L}^1\left(\left\{t \in (0, 1) \mid g_{ij}(\rho(t)) = 0, m_{ij}(t) \neq 0\right\}\right) = 0. \quad (4.2)$$

(ii)  $\mathcal{A}$  is non-negative and lower semicontinuous on  $L^2(0, 1; \mathbb{R}^n) \times L^2(0, 1; S^{n \times n})$  for the weak convergence.

*Proof.* We skip the proof of (i). Since  $F$  is a non-negative convex, lower semicontinuous function, by the standard theory of the calculus of variations ( cf. e.g. [10]) we obtain (ii).  $\square$

**Theorem 4.5.** Assume  $\rho^0, \rho^1 \in \mathcal{P}(G)$ .

(i) There exists  $(v^*, \rho^*, m^*)$  such that  $(v^*, \rho^*)$  is a minimizer in (2.5) and  $(\rho^*, m^*)$  minimizes  $\mathcal{A}$  over  $\mathcal{C}(\rho^0, \rho^1)$ .  
(ii) Furthermore,

$$2 \inf_{\mathcal{C}(\rho^0, \rho^1)} \mathcal{A} = \mathcal{W}_g^2(\rho^0, \rho^1) = \int_0^1 \|v^*\|_{\rho^*}^2 dt = 2\mathcal{A}(\rho^*, m^*).$$

(iii) We have

$$F(\rho^*, m^*)(t) = F(\rho^*, m^*)(0) \quad \text{a.e. on } (0, 1).$$

*Proof.* By Proposition 3.7, there is a positive number  $i_0$  and a path  $(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)$  such that  $\mathcal{A}(\rho, m) \leq i_0$ . By Lemma 4.1,

$$\int_0^1 |m(t)|^2 dt \leq 2\epsilon_0(g)i_0.$$

Using the differential equation linking  $\dot{\rho}$  to  $m$  we conclude that for a constant  $C$  depending only on  $i_0, \epsilon_0(g), w$  and  $n$ , we have

$$\|\dot{\rho}\|_{L^2(0,1)}^2 \leq C.$$

Increasing the value of  $C$  is necessary, we use the Poincaré–Wintiger inequality to obtain

$$\|\rho\|_{H^1(0,1)}^2 \leq C.$$

As a consequence, the intersection of  $\mathcal{C}(\rho^0, \rho^1)$  with any sub-level subsection of  $\mathcal{A}$  is precompact set of  $H^1(0, 1; \mathbb{R}^n) \times L^2(0, 1; S^{n \times n})$  for the weak topology. By Remark 4.4 (ii),  $\mathcal{A}$  is weakly lower semicontinuous and so, it achieves its minimum at some  $(\rho^*, m^*) \in \mathcal{C}(\rho^0, \rho^1)$ .

Since  $\int_0^1 F(\rho^*(t), m^*(t)) < \infty$ , by Remark 4.4 (i), the set obtained as the union over  $(i, j) \in E$  of the sets  $\{g_{ij}(\rho^*) = 0\} \cap \{m_{ij}^* \neq 0\}$  is of null measure. Thus, the functions  $v_{ij}^* : (0, 1) \rightarrow \mathbb{R}$  defined as

$$v_{ij}^*(t) = \begin{cases} \frac{m_{ij}^*(t)}{g_{ij}(\rho^*(t))} & \text{if } g_{ij}(\rho^*(t)) > 0 \\ 0 & \text{if } g_{ij}(\rho^*(t)) = 0 \end{cases} \quad (4.3)$$

are measurable and satisfies  $m_{ij}^* = g_{ij}(\rho^*)v_{ij}^*$ . Let  $(\rho, v)$  be an admissible path in (2.5). This means we are assuming that  $(\rho(0), \rho(1)) = (\rho^0, \rho^1)$ ,  $\rho \in H^1(0, 1; \mathbb{R}^n)$ ,  $v : [0, 1] \rightarrow S^{n \times n}$  is Borel measurable and

$$\int_0^1 \|v\|_{\rho}^2 dt < \infty.$$

Setting  $m_{ij} = g_{ij}(\rho)v_{ij}$  we have

$$\int_0^1 |m|^2 dt = \frac{1}{2} \sum_{(i,j) \in E} \int_0^1 v_{ij}^2 g_{ij}^2(\rho) dt \leq \epsilon_0(g) \sum_{(i,j) \in E} \int_0^1 v_{ij}^2 g_{ij}(\rho) dt = 2\epsilon_0(g) \int_0^1 \|v\|_{\rho}^2 dt.$$

Thus,  $m \in L^2(0, 1; S^{n \times n})$  and so,  $(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)$ . By the definition of  $v^*$

$$\|v^*\|_{\rho^*}^2 = F(\rho^*, m^*). \quad (4.4)$$

By the minimality property of  $(\rho^*, m^*)$ , we have

$$\int_0^1 \|v^*\|_{\rho^*}^2 dt = 2\mathcal{A}(\rho^*, m^*) \leq 2\mathcal{A}(\rho, m) = \int_0^1 \|v\|_{\rho}^2 dt. \quad (4.5)$$

This proves (i) and also (ii).

(iii) Here, we borrow ideas from [6]. Let  $\zeta \in C_c^1(0, 1)$  be arbitrary and set  $S(t) = t + \epsilon\zeta(t)$ . We have  $S(0) = 0$ ,  $S(1) = 1$  and  $\dot{S}(t) = 1 + \epsilon\dot{\zeta}(t) > 1/2$  for  $|\epsilon| \ll 1$ . Thus,  $S : [0, 1] \rightarrow [0, 1]$  is a diffeomorphism. Let  $T := S^{-1}$  and set

$$u(s) = \rho^*(T(s)), \quad w(s) = \dot{T}(s)m^*(T(s)).$$

We have

$$\dot{u} + \nabla_G \cdot (w) = 0, \quad u(0) = \rho^0, \quad u(1) = \rho^1.$$

Thus,  $(u, w) \in \mathcal{C}(\rho^0, \rho^1)$  and so,

$$\int_0^1 F(\rho^*, m^*) dt \leq \int_0^1 F(u, w) ds = \int_0^1 \dot{T}^2 F(\rho^*(T), m^*(T)) ds.$$

We use the fact that  $dt = \dot{T}(s)ds$  and  $\dot{T}(S(t))\dot{S}(t) = 1$  to conclude that

$$\int_0^1 F(\rho^*, m^*) dt \leq \int_0^1 \frac{1}{\dot{S}} F(\rho^*, m^*) dt = \int_0^1 (1 - \epsilon\dot{\zeta} + o(\epsilon)) F(\rho^*, m^*) dt.$$

Since  $\epsilon \rightarrow \int_0^1 (1 - \epsilon\dot{\zeta} + o(\epsilon)) F(\rho^*, m^*) dt$  admits its minimum at 0, we conclude that its derivative there is null, *i.e.*,

$$\int_0^1 \dot{\zeta}(t) F(\rho^*, m^*) dt = 0.$$

This proves that the distributional derivative of  $F(\rho^*, m^*)$  is null and so,  $F(\rho^*, m^*)$  is independent of  $t$ .  $\square$

**Remark 4.6.** Let  $(\rho^*, m^*, v^*)$  as in Theorem 4.5

- (i) We have  $\|v^*\|_{\rho^*}^2 = F(\rho^*, m^*) = \mathcal{W}_g^2(\rho^0, \rho^1)$  and so,  $\|v^*\|_{\rho^*} \in L^\infty(0, 1)$ .
- (ii) We have  $\rho^* \in W^{1, \infty}(0, 1; \mathbb{R}^n)$ ,  $m^* \in L^\infty(0, 1; S^{n \times n})$  and

$$|m_{ij}^*| \leq \mathcal{W}_g(\rho^0, \rho^1) \sqrt{\max_{[0, 1]^2} g}.$$

*Proof.* (i) is a direct consequence of Theorem 4.5 (iii) and (4.4).

(ii) The same argument which led to the definition of  $v^*$  in (4.3) can be used to conclude that if  $(i, j) \in E$  then either  $m_{ij}^*(t) = 0$  or

$$\frac{(m_{ij}^*)^2}{g_{ij}(\rho^*)}(t) \leq F(\rho^*, m^*)(t) = \mathcal{W}_g^2(\rho^0, \rho^1).$$

In any of these two cases,

$$(m_{ij}^*)^2 \leq \mathcal{W}_g^2(\rho^0, \rho^1) g_{ij}(\rho^*) \leq \mathcal{W}_g^2(\rho^0, \rho^1) \max_{[0, 1]^2} g,$$

which yields the desired inequality in (ii). This shows  $m^* \in L^\infty(0, 1; S^{n \times n})$ , which together with the identity  $\dot{\rho} = -\nabla_G \cdot m^*$ , shows  $\dot{\rho} \in L^\infty(0, 1; \mathbb{R}^n)$ .  $\square$

**Theorem 4.7.**  $\mathcal{W}_g$  is a metric on  $\mathcal{P}(G)$ .

*Proof.* The symmetry of  $\mathcal{W}_g$  is a direct consequence of the definition of  $\mathcal{W}_g$ .

It is obvious that  $\mathcal{W}_g(\rho_0, \rho_1) = 0$  if  $\rho^0 = \rho^1 \in \mathcal{P}(G)$ . We claim that  $\mathcal{W}_g(\rho^0, \rho^1) = 0$  implies  $\rho^0 = \rho^1$ . Let  $(\rho^*, m^*, v^*)$  be as in Theorem 4.5 and let  $q = (q_1, \dots, q_n)$  be an arbitrary vector in  $\mathbb{R}^n$ . Using Hölder's inequality, there exists a constant  $C > 0$  such that

$$\begin{aligned} \left| \sum_{i=1}^n q_i (\rho_i^0 - \rho_i^1) \right| &= \left| \int_0^1 (q, \rho^*) dt \right| \\ &= \left| \int_0^1 (\nabla_G q, v^*)_{\rho^*} dt \right| \\ &\leq \left( \int_0^1 (\nabla_G q, \nabla_G q)_{\rho^*} dt \right)^{\frac{1}{2}} \mathcal{W}_g(\rho^0, \rho^1) \\ &\leq C \mathcal{W}_g(\rho^0, \rho^1) \max_{i \in V} |q_i|. \end{aligned}$$

Therefore,  $\sum_{i=1}^n |\rho_i^0 - \rho_i^1| \leq C \mathcal{W}_g(\rho^0, \rho^1)$ .

At the end, the triangle inequality follows from Remark 4.6.  $\square$

## 5. DUALITY IN A SMOOTH SETTING

Throughout this section, we further assume that  $g$  satisfies (2.8). The main purpose of the section is to find the Euler–Lagrange equations satisfied by geodesics of minimal action connecting  $\rho^0, \rho^1 \in \mathcal{P}(G)$ . We will express these Euler–Lagrange equations in terms of the Hamiltonian  $H$  defined in (1.10). It is convenient to set

$$A := \left\{ \rho \in L^2(0, 1; \mathbb{R}^n) \mid \sum_{i=1}^n \rho_i = 1, \rho_i \geq 0 \forall i = 1, \dots, n \right\} \times L^2(0, 1; S^{n \times n}), \quad B := H^1(0, 1; \mathbb{R}^n).$$

For  $l \in (0, \infty)$  we set

$$A^l := \left\{ \rho \in L^2(0, 1; \mathbb{R}^n) \mid \sum_{i=1}^n \rho_i \leq l, \rho_i \geq 0 \forall i = 1, \dots, n \right\} \times L^2(0, 1; S^{n \times n}), \quad (5.1)$$

and

$$B_l := \{ \lambda \in B \mid \|\lambda\|_{H^1(0, 1; \mathbb{R}^n)} \leq l \}.$$

We plan to prove the duality property

$$\min_{(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)} \{ \mathcal{A}(\rho, m) \} = \sup_{\lambda \in B} \left\{ (\lambda(1), \rho^1) - (\lambda(0), \rho^0) \mid H(\dot{\lambda}, \nabla_G \lambda) = 0 \right\}. \quad (5.2)$$

As we will show, this reduces to a minimax identity for

$$\mathcal{L}(\rho, m, \lambda) := (\lambda(1), \rho^1) - (\lambda(0), \rho^0) + \mathcal{A}(\rho, m) - \int_0^1 \left( (\dot{\lambda}, \rho) + (m, \nabla_G \lambda) \right) dt.$$

**Proposition 5.1.** *For any  $l > 0$ ,  $e \geq 1$  and  $A_* \in \{A, A^e\}$  we have*

$$\inf_{(\rho, m) \in A_*} \sup_{\lambda \in B_l} \mathcal{L}(\rho, m, \lambda) = \sup_{\lambda \in B_l} \inf_{(\rho, m) \in A_*} \mathcal{L}(\rho, m, \lambda).$$

*Proof.* By Lemma 3.5(i),  $\mathcal{L}$  is convex in  $(\rho, m)$ . By inspection,  $\mathcal{L}$  is linear in  $\lambda$  and so it is concave. Let  $(\bar{\rho}, \bar{m}) \in A_*$ ,  $\bar{\lambda} \in B_l$  and  $C \in \mathbb{R}$  be arbitrary. To show the proposition, according to the standard minimax theorem (*cf. e.g.* Proposition 2.5 [21] on page 16), it suffices to show the following properties:  $B_l$  is a convex and compact set for the weak topology (which is obviously the case),  $A_*$  is a convex topological space (which is obviously the case for the weak topology),  $\{\lambda \in B_l \mid \mathcal{L}(\bar{\rho}, \bar{m}, \lambda) \geq C\}$  is a closed convex set in  $B_l$  and  $\{(\rho, m) \in A_* \mid \mathcal{L}(\rho, m, \bar{\lambda}) \leq C\}$  is a closed convex set in  $A_*$ .

When  $\mathcal{A}(\bar{\rho}, \bar{m}) = \infty$  then  $\mathcal{L}(\bar{\rho}, \bar{m}, \cdot) \equiv \infty$  and so,  $\{\lambda \in B_l \mid \mathcal{L}(\bar{\rho}, \bar{m}, \lambda) \geq C\} = B_l$  is a closed convex set. When  $\mathcal{A}(\bar{\rho}, \bar{m}) < \infty$ ,  $\mathcal{L}(\bar{\rho}, \bar{m}, \cdot)$  is linear and so,  $\{\lambda \in B_l \mid \mathcal{L}(\bar{\rho}, \bar{m}, \lambda) \geq C\}$  is convex. We use the fact that bounded subsets of  $H^1(0, 1)$  are compact in  $C[0, 1]$ , that  $\bar{m} \in L^2$  and  $\bar{\rho} \in L^\infty$  to conclude that  $\mathcal{L}(\bar{\rho}, \bar{m}, \cdot)$  is a continuous function on  $B_l$  and so,  $\{\lambda \in B_l \mid \mathcal{L}(\bar{\rho}, \bar{m}, \lambda) \geq C\}$  is a closed subset of  $B_l$ .

By the fact that  $F$  is a convex function and  $A_*$  is a convex set,  $\{(\rho, m) \in A_* \mid \mathcal{L}(\rho, m, \bar{\lambda}) \leq C\}$  is convex subset of  $A_*$ . One part of  $\mathcal{L}(\cdot, \cdot, \bar{\lambda})$  is a linear functional and by Remark 4.4 (ii), the other part is weakly lower semicontinuous. Thus,  $\mathcal{L}(\cdot, \cdot, \bar{\lambda})$  is itself weakly lower semicontinuous and so,  $\{(\rho, m) \in A_* \mid \mathcal{L}(\rho, m, \bar{\lambda}) \leq C\}$  is a closed subset of  $A_*$ . This, concludes the proof of the Proposition.  $\square$

Since  $\mathcal{L}$  is linear with respect to  $\lambda$ , we have

$$\sup_{B_l} \mathcal{L}(\rho, m, \lambda) = \mathcal{A}(\rho, m) + l\mathcal{E}(\rho, m) \quad (5.3)$$

where

$$\mathcal{E}(\rho, m) = \sup_{\lambda \in B_1} (\lambda(1), \rho^1) - (\lambda(0), \rho^0) - \int_0^1 \left( (\dot{\lambda}, \rho) + (m, \nabla_G \lambda) \right) dt.$$

Observe that

$$\mathcal{E}(\rho, m) = \begin{cases} 0 & \text{if } (\rho, m) \in \mathcal{C}(\rho^0, \rho^1) \\ > 0 & \text{otherwise.} \end{cases} \quad (5.4)$$

**Remark 5.2.** Let  $l > 0$ ,  $e \geq 1$  and  $A_* \in \{A, A^e\}$ . By Theorem 4.5,  $\mathcal{A}$  achieves its minimum over  $\mathcal{C}(\rho^0, \rho^1)$  at some  $(\rho^*, m^*)$ . By (5.4), we obtain that  $\mathcal{E}(\rho^*, m^*) = 0$  and thus the infimum of  $\mathcal{A} + l\mathcal{E}$  over  $A_*$  is between 0 and  $\mathcal{A}(\rho^*, m^*)$ .

**Lemma 5.3.** *Let  $e \geq 1$  and let  $A_* \in \{A, A^e\}$ . The following hold.*

- (i)  $\mathcal{E}$  is convex and weakly lower semi-continuous on  $L^2(0, 1; \mathbb{R}^n) \times L^2(0, 1; S^{n \times n})$ .
- (ii) For any  $l > 0$ , there exists  $(\rho^{*,l}, m^{*,l})$  which minimizes  $\mathcal{A} + l\mathcal{E}$  over  $A_*$ .
- (iii) The set  $\{(\rho^{*,l}, m^{*,l}) \mid l > 0\}$  is pre-compact in  $A_*$ .

*Proof.* (i) As a supremum of continuous linear functionals,  $\mathcal{E}$  is convex and weakly lower semi-continuous.

(ii) By Remark 5.2,  $\mathcal{A} + l\mathcal{E}$  is not identically  $\infty$  over  $A_*$ . Since  $\mathcal{E} \geq 0$ , any sub-level subset of  $\mathcal{A} + l\mathcal{E}$  is a sub-level subset of  $\mathcal{A}$ . By Lemma 4.1, the sub-level sets of  $\mathcal{A}$  are contained in a bounded subset of  $L^2(0, 1; \mathbb{R}^n) \times L^2(0, 1; S^{n \times n})$  and so, they are pre-compact. Thus, the sub-level subsets of  $\mathcal{A} + l\mathcal{E}$  are pre-compact. By Remark 4.4  $\mathcal{A}$  is weakly lower semi-continuous on  $L^2(0, 1; \mathbb{R}^n) \times L^2(0, 1; S^{n \times n})$  and by (i)  $\mathcal{E}$  is weakly lower semi-continuous on that same set. That is all we need to prove that  $\mathcal{A} + l\mathcal{E}$  achieves its minimum over the closed set  $A_*$ .

(iii) By Remark 5.2, for any  $l > 0$ ,

$$\mathcal{A}(\rho^{*,l}, m^{*,l}) \leq \mathcal{A}(\rho^*, m^*).$$

We apply again Lemma 4.1 to conclude that  $\{m^{*,l} \mid l > 0\}$  is bounded in  $L^2$ . This is sufficient to verify (iii).  $\square$

**Lemma 5.4.** *Let  $e \geq 1$  and let  $A_* \in \{A^e, A\}$ . We have*

$$\inf_{(\rho, m) \in A_*} \sup_{\lambda \in B} \mathcal{L}(\rho, m, \lambda) = \sup_{\lambda \in B} \inf_{(\rho, m) \in A_*} \mathcal{L}(\rho, m, \lambda).$$

*Proof.* We use (5.3) and the first identity in (5.4) to obtain for any  $l \geq 1$ ,

$$\mathcal{A}(\rho^*, m^*) = \sup_{\lambda \in B_l} \mathcal{L}(\rho^*, m^*, \lambda) \geq \inf_{(\rho, m) \in A_*} \sup_{\lambda \in B_l} \mathcal{L}(\rho, m, \lambda).$$

This, together with Proposition 5.1 implies

$$\mathcal{A}(\rho^*, m^*) \geq \sup_{\lambda \in B_l} \inf_{(\rho, m) \in A_*} \mathcal{L}(\rho, m, \lambda). \quad (5.5)$$

This means

$$\mathcal{A}(\rho^*, m^*) \geq \mathcal{A}(\rho^{*,l}, m^{*,l}) + l\mathcal{E}(\rho^{*,l}, m^{*,l}) \geq \mathcal{A}(\rho^{*,l}, m^{*,l}). \quad (5.6)$$

By Lemma 5.3 (iii), the second inequality in (5.6) yields that the set  $\{(\rho^{*,l}, m^{*,l}) \mid l \geq 1\}$  is pre-compact in  $A_*$  and so, it admits a point of accumulation  $(\rho^\infty, m^\infty)$ . Since Lemma 5.3 (i) ensures that  $\mathcal{E}$  is weakly lower semi-continuous, we may divide the expression in (5.6) by  $l$  and then let  $l$  tend to  $\infty$  in the subsequent inequality and use the fact that  $\mathcal{E}$  is non-negative, to obtain  $\mathcal{E}(\rho^\infty, m^\infty) = 0$ . Thanks to (5.4) we obtain that

$$(\rho^\infty, m^\infty) \in \mathcal{C}(\rho^0, \rho^1).$$

By the minimality property of  $(\rho^*, m^*)$  obtained in Theorem 4.5, we have

$$\mathcal{A}(\rho^*, m^*) \leq \mathcal{A}(\rho^\infty, m^\infty). \quad (5.7)$$

We let  $l$  tend to  $\infty$  in (5.6) and use the lower semicontinuity property of  $\mathcal{A}$  given in Remark 4.4 to reverse the inequality in (5.7). In conclusion,

$$\mathcal{A}(\rho^*, m^*) = \mathcal{A}(\rho^\infty, m^\infty).$$

and

$$\varliminf_{l \rightarrow +\infty} l\mathcal{E}(\rho^{*,l}, m^{*,l}) = 0.$$

To summarize, we have proven that

$$\mathcal{A}(\rho^\infty, m^\infty) \leq \varliminf_{l \rightarrow +\infty} \mathcal{A}(\rho^{*,l}, m^{*,l}) + l\mathcal{E}(\rho^{*,l}, m^{*,l}) = \varliminf_{l \rightarrow +\infty} \inf_{(\rho, m) \in A_*} \sup_{\lambda \in B_l} \mathcal{L}(\rho, m, \lambda).$$

We first apply the duality identity in Proposition 5.1 to interchange  $\inf_{A_*} \sup_{B_l}$  and  $\sup_{B_l} \inf_{A_*}$ . Then we use the fact that Since  $B_l \subset B$  to conclude that

$$\mathcal{A}(\rho^\infty, m^\infty) \leq \varliminf_{l \rightarrow +\infty} \sup_{\lambda \in B_l} \inf_{(\rho, m) \in A_*} \mathcal{L}(\rho, m, \lambda) \leq \sup_{\lambda \in B} \inf_{(\rho, m) \in A_*} \mathcal{L}(\rho, m, \lambda). \quad (5.8)$$

But as  $\mathcal{L}(\rho^\infty, m^\infty, \cdot) \equiv \mathcal{A}(\rho^\infty, m^\infty)$  we infer

$$\mathcal{A}(\rho^\infty, m^\infty) = \sup_{\lambda \in B} \mathcal{L}(\rho^\infty, m^\infty, \lambda) \geq \inf_{(\rho, m) \in A_*} \sup_{\lambda \in B} \mathcal{L}(\rho, m, \lambda).$$

This, together with (5.8) yields,

$$\inf_{(\rho, m) \in A_*} \sup_{\lambda \in B} \mathcal{L}(\rho, m, \lambda) \leq \sup_{\lambda \in B} \inf_{(\rho, m) \in A_*} \mathcal{L}(\rho, m, \lambda).$$

The reverse inequality  $\sup_B \inf_{A_*} \leq \inf_{A_*} \sup_B$  being always true, we conclude the proof of the lemma.  $\square$

Recall  $A^l$  is defined earlier in (5.1). Set

$$A^\infty := \left\{ \rho \in L^2(0, 1; \mathbb{R}^n) \mid \rho_i \geq 0 \ \forall i = 1, \dots, n \right\} \times L^2(0, 1; S^{n \times n}).$$

**Lemma 5.5.** *Let  $\lambda \in B$ .*

(i) *For any  $l > 0$*

$$\inf_{(\rho, m) \in A^l} \mathcal{L}(\rho, m, \lambda) = (\lambda(1), \rho^1) - (\lambda(0), \rho^0) - l \int_0^1 (H(\dot{\lambda}, \nabla_G \lambda))_+ dt.$$

(ii)

$$\inf_{(\rho, m) \in A^\infty} \mathcal{L}(\rho, m, \lambda) = \begin{cases} (\lambda(1), \rho^1) - (\lambda(0), \rho^0) & \text{if } (H(\dot{\lambda}, \nabla_G \lambda))_+ \leq 0 \text{ a.e. in } (0, 1) \\ -\infty & \text{otherwise.} \end{cases}$$

*Proof.* Expressing the inf in terms of  $-\sup$  and using (2.7) we have

$$\inf_{m \in L^2} \int_0^1 \frac{1}{2} F(\rho, m) - (\dot{\lambda}, \rho) - (m, \nabla_G \lambda) dt = - \int_0^1 \left( (\dot{\lambda}, \rho) + \frac{1}{2} \|\nabla_G \lambda\|_\rho^2 \right) dt. \quad (5.9)$$

Since  $g$  is 1-homogeneous

$$\int_0^1 \sup_{\sum_{i=1}^n \rho_i \leq l} \left( (\dot{\lambda}, \rho) + \frac{1}{2} \|\nabla_G \lambda\|_\rho^2 \right) dt = l \int_0^1 (H(\dot{\lambda}, \nabla_G \lambda))_+ dt.$$

This, together with (5.9), proves (i). We let  $l$  tend to  $\infty$  to verify (ii).  $\square$

**Lemma 5.6.** *Let  $\lambda \in H^1(0, 1; \mathbb{R}^n)$  and  $\alpha \in H^1(0, 1)$  and set  $\bar{\lambda}_i = \lambda_i + \alpha$ . Then*

$$H(\dot{\bar{\lambda}}, \nabla_G \bar{\lambda}) = H(\dot{\lambda}, \nabla_G \lambda) + \dot{\alpha}.$$

*Proof.* Observe that for any  $\rho \in \mathcal{P}(G)$  we have

$$(\dot{\bar{\lambda}}, \rho) = (\dot{\lambda}, \rho) + \dot{\alpha} \quad \text{and} \quad \nabla_G \bar{\lambda} = \nabla_G \lambda. \quad (5.10)$$

Since

$$H(\dot{\bar{\lambda}}, \nabla_G \bar{\lambda}) = \sup_{\rho \in \mathcal{P}(G)} \left\{ (\dot{\bar{\lambda}}, \rho) + \frac{1}{2} \|\nabla_G \bar{\lambda}\|_{\rho}^2 \right\},$$

we use (5.10) to conclude the proof.  $\square$

**Proposition 5.7.** *Given  $\lambda \in H^1(0, 1; \mathbb{R}^n)$ , there is  $\bar{\lambda} \in H^1(0, 1; \mathbb{R}^n)$  such that  $H(\dot{\bar{\lambda}}, \nabla_G \bar{\lambda})_+ \equiv 0$  and*

$$\inf_{(\rho, m) \in A^1} \mathcal{L}(\rho, m, \lambda) = \inf_{(\rho, m) \in A^1} \mathcal{L}(\rho, m, \bar{\lambda}).$$

*Proof.* Let

$$O := \{t \in (0, 1) \mid H(\dot{\lambda}, \nabla_G \lambda) > 0\}.$$

If  $\mathcal{L}^1(O) = 0$ , we are done by letting  $\bar{\lambda} = \lambda$ . Assume that  $\mathcal{L}^1(O) > 0$ . Set

$$\alpha(t) = - \int_0^t \chi_O(s) H(\dot{\lambda}(s), \nabla_G \lambda(s)) ds.$$

Since (H-i) holds and  $\lambda \in H^1(0, 1; \mathbb{R}^n)$ , we have

$$|\dot{\alpha}| = |\chi_O H(\dot{\lambda}, \nabla_G \lambda)| \leq \left| \sup_{\rho \in \mathcal{P}(G)} \left\{ (\dot{\lambda}, \rho) + \frac{1}{2} \|\nabla_G \lambda\|_{\rho}^2 \right\} \right| \leq |\dot{\lambda}| + \max_{[0, 1]^n} \frac{g}{2} \|\nabla_G \lambda\|^2 \in L^2(0, 1).$$

Thus,  $\alpha \in H^1(0, 1)$ . Set  $\bar{\lambda}_i = \lambda_i + \alpha$ . By Lemma 5.6, we obtain

$$H(\dot{\bar{\lambda}}, \nabla_G \bar{\lambda}) = H(\dot{\lambda}, \nabla_G \lambda) + \dot{\alpha} = (1 - \chi_O) H(\dot{\lambda}, \nabla_G \lambda) \leq 0.$$

Hence,  $H(\dot{\bar{\lambda}}, \nabla_G \bar{\lambda})_+ \equiv 0$ , which verifies the first claim of the proposition.

By Lemma 5.5 and (i), we infer,

$$\inf_{(\rho, m) \in A^1} \mathcal{L}(\rho, m, \bar{\lambda}) = (\lambda(1), \rho^1) - (\lambda(0), \rho^0) + \alpha(1) - \alpha(0). \quad (5.11)$$

Since

$$\alpha(1) - \alpha(0) = - \int_0^1 \left( H(\dot{\lambda}, \nabla_G \lambda) \right)_+ dt, \quad (5.12)$$

we use Lemma 5.5 again and combine (5.11) and (5.12) to verify (ii).  $\square$

**Proposition 5.8.** *Suppose  $\lambda \in H^1(0, 1; \mathbb{R}^n)$  and  $H(\dot{\lambda}, \nabla_G \lambda) \leq 0$ . Then there exists  $\bar{\lambda} \in H^1(0, 1; \mathbb{R}^n)$  such that  $H(\dot{\bar{\lambda}}, \nabla_G \bar{\lambda}) = 0$  and*

$$(\bar{\lambda}(1), \rho^1) - (\bar{\lambda}(0), \rho^0) \geq (\lambda(1), \rho^1) - (\lambda(0), \rho^0).$$

*Proof.* Let  $O := \{H(\dot{\lambda}, \nabla_G \lambda) < 0\}$  and to avoid trivialities, assume  $\mathcal{L}^1(O) > 0$ . Set

$$\alpha(t) = - \int_0^t \chi_O H(\dot{\lambda}, \nabla_G \lambda) ds$$

and  $\bar{\lambda}_i = \lambda_i + \alpha$ . As done in the proof of Proposition 5.7, we have  $\alpha \in H^1(0, 1)$  and by Lemma 5.6,

$$H(\dot{\bar{\lambda}}, \nabla_G \bar{\lambda}) = (1 - \chi_O)H(\dot{\lambda}, \nabla_G \lambda) = 0.$$

Finally,

$$(\bar{\lambda}(1), \rho^1) - (\bar{\lambda}(0), \rho^0) = (\lambda(1), \rho^1) - (\lambda(0), \rho^0) - \int_0^1 H(\dot{\lambda}, \nabla_G \lambda) dt \geq (\lambda(1), \rho^1) - (\lambda(0), \rho^0).$$

□

**Corollary 5.9.** *We have*

$$\sup_{\lambda \in B} \inf_{(\rho, m) \in A^\infty} \mathcal{L}(\rho, m, \lambda) = \inf_{(\rho, m) \in A^\infty} \sup_{\lambda \in B} \mathcal{L}(\rho, m, \lambda). \quad (5.13)$$

*Proof.* Since  $A^1 \subset A^\infty$ ,

$$\sup_{\lambda \in B} \inf_{(\rho, m) \in A^\infty} \mathcal{L}(\rho, m, \lambda) \leq \sup_{\lambda \in B} \inf_{(\rho, m) \in A^1} \mathcal{L}(\rho, m, \lambda).$$

Define

$$\mathcal{L}^*(\lambda) = \inf_{(\rho, m) \in A^1} \mathcal{L}(\rho, m, \lambda).$$

Assume that  $\{\lambda_n\}_n \subset B$  is a maximizing sequence such that

$$\sup_{\lambda \in B} \mathcal{L}^*(\lambda) = \lim_{n \rightarrow +\infty} \mathcal{L}^*(\lambda_n). \quad (5.14)$$

By Proposition 5.7, there exists  $\{\bar{\lambda}_n\}_n \subset B$  such that

$$\mathcal{L}^*(\bar{\lambda}_n) = \mathcal{L}^*(\lambda_n) \quad \text{and} \quad \left( H(\dot{\bar{\lambda}}_n, \nabla_G \bar{\lambda}_n) \right)_+ = 0. \quad (5.15)$$

Combining (5.14) and (5.15) and using Lemma 5.5 we obtain

$$\begin{aligned} \sup_{\lambda \in B} \inf_{(\rho, m) \in A^1} \mathcal{L}(\rho, m, \lambda) &= \lim_{n \rightarrow \infty} \mathcal{L}^*(\lambda_n) \\ &= \lim_{n \rightarrow \infty} (\bar{\lambda}_n(1), \rho^1) - (\bar{\lambda}_n(0), \rho^0) \\ &\leq \sup_{\lambda \in B} \left\{ (\lambda(1), \rho^1) - (\lambda(0), \rho^0) \mid H(\dot{\lambda}, \nabla_G \lambda) \leq 0 \right\} \\ &= \sup_{\lambda \in B} \inf_{(\rho, m) \in A^\infty} \mathcal{L}(\rho, m, \lambda). \end{aligned}$$

Since  $A^1 \subset A^\infty$ , using the duality result in Lemma 5.4, we have proven that

$$\sup_{\lambda \in B} \inf_{(\rho, m) \in A^\infty} \mathcal{L}(\rho, m, \lambda) = \sup_{\lambda \in B} \inf_{(\rho, m) \in A^1} \mathcal{L}(\rho, m, \lambda) = \inf_{(\rho, m) \in A^1} \sup_{\lambda \in B} \mathcal{L}(\rho, m, \lambda).$$

We exploit once more the fact that  $A^1 \subset A^\infty$  to infer

$$\sup_{\lambda \in B} \inf_{(\rho, m) \in A^\infty} \mathcal{L}(\rho, m, \lambda) \geq \inf_{(\rho, m) \in A^\infty} \sup_{\lambda \in B} \mathcal{L}(\rho, m, \lambda).$$

Since the reverse inequality always holds, we conclude the proof of the Corollary.  $\square$

We now state the main result of this section.

**Theorem 5.10.** (5.2) holds.

*Proof.* Define

$$I_{\mathcal{C}(\rho^0, \rho^1)} = \begin{cases} 0 & \text{if } (\rho, m) \in \mathcal{C}(\rho^0, \rho^1) \\ \infty & \text{if } (\rho, m) \in A^\infty \setminus \mathcal{C}(\rho^0, \rho^1). \end{cases}$$

By (5.3) and (5.4), for any  $(\rho, m) \in A^\infty$ , we have

$$\sup_{\lambda \in B} \mathcal{L}(\rho, m, \lambda) = \mathcal{A}(\rho, m) + I_{\mathcal{C}(\rho^0, \rho^1)}(\rho, m).$$

Thus,

$$\inf_{(\rho, m) \in A^\infty} \sup_{\lambda \in B} \mathcal{L}(\rho, m, \lambda) = \inf_{(\rho, m) \in A^\infty} \left\{ \mathcal{A}(\rho, m) + I_{\mathcal{C}(\rho^0, \rho^1)}(\rho, m) \right\}. \quad (5.16)$$

Since  $\mathcal{C}(\rho^0, \rho^1) \subset A^\infty$ , exploiting (5.16), we infer

$$\min_{(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)} \mathcal{A}(\rho, m) = \inf_{(\rho, m) \in A^\infty} \left\{ \mathcal{A}(\rho, m) + I_{\mathcal{C}(\rho^0, \rho^1)}(\rho, m) \right\} = \inf_{(\rho, m) \in A^\infty} \sup_{\lambda \in B} \mathcal{L}(\rho, m, \lambda). \quad (5.17)$$

We first use Corollary 5.9 in (5.17) to conclude that

$$\min_{(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)} \mathcal{A}(\rho, m) = \sup_{\lambda \in B} \inf_{(\rho, m) \in A^\infty} \mathcal{L}(\rho, m, \lambda).$$

We reach the desired conclusion by noting that in light of Lemma 5.5 and Proposition 5.8, the right hand side of this last identity is nothing but the supremum in (5.2).  $\square$

## 6. INGREDIENTS FOR DUALITY IN THE NON-SMOOTH CASE

Throughout this section, we further assume  $g$  satisfies (2.8). This section is a preliminary toward a necessary and sufficient condition for a path to be a geodesic  $(\mathcal{P}(G), \mathcal{W}_g)$ . We will make use of  $H_0$ , the restriction of the recession function of  $H$  (cf. e.g. [3]) to  $\mathbb{R}^n \times \{0\}$ :

$$H_0(a) = \sup_{\rho \in \mathcal{P}(G)} (a, \rho) = \max_{1 \leq i \leq n} a_i.$$

**Lemma 6.1.** *Assume  $\nu$  is a non-negative Borel regular measure on  $(0, 1)$  such that  $\nu$  and  $\mathcal{L}^1|_{(0,1)}$  are mutually singular. Let  $m \in L^2(0, 1; S^{n \times n})$ , let  $\beta : (0, 1) \rightarrow \mathbb{R}^n$  be a Borel map (defined  $\mathcal{L}^1$  a.e.) and let  $\bar{g} : (0, 1) \rightarrow \mathbb{R}^n$  be a Borel map (defined  $\nu$  a.e.). Then the following assertions are equivalent*

(i)

$$\begin{cases} H(\beta, m) \leq 0 & \mathcal{L}^1 \text{ a.e. in } (0, 1) \\ H_0(\bar{g}) \leq 0 & \nu \text{ a.e. in } (0, 1). \end{cases} \quad (6.1)$$

(ii) *For every non-negative function  $\varphi \in C([0, 1])$ , we have  $s[\varphi] \leq 0$  if we set*

$$s[\varphi] := \sup_{\rho} \left\{ \int_0^1 \left( (\rho, \beta) dt + (\rho, \bar{g}\nu(dt)) + \frac{1}{2} \|m\|_{\rho}^2 dt \right) \varphi(t) \mid \rho \in \mathcal{C}_G \right\}.$$

Here,  $\mathcal{C}_G$  is the set of Borel maps of  $(0, 1)$  into  $\mathcal{P}(G)$ .

*Proof.* Let  $\varphi \in C([0, 1])$  be non-negative. Since  $\nu$  and  $\mathcal{L}^1|_{(0,1)}$  are mutually singular

$$s[\varphi] = \sup_{\rho} \left\{ \int_0^1 \left( (\rho, \beta) + \frac{1}{2} \|m\|_{\rho}^2 \right) \varphi(t) dt \mid \rho \in \mathcal{C}_G \right\} + \sup_{\rho} \left\{ \int_0^1 (\rho, \bar{g}\nu(dt)) \varphi(t) \mid \rho \in \mathcal{C}_G \right\}.$$

Thus,

$$s[\varphi] = \int_0^1 H(\beta, m) \varphi(t) dt + \int_0^1 H_0(\bar{g}) \varphi(t) \nu(dt).$$

Using again the fact that  $\nu$  and  $\mathcal{L}^1|_{(0,1)}$  are mutually singular, we conclude that  $s[\varphi] \leq 0$  for all non-negative  $\varphi \in C([0, 1])$  if and only if (6.1) holds.  $\square$

**Lemma 6.2.** *Let  $R : [0, 1] \rightarrow [0, 1]$  be a Lipschitz function and let  $h \in L^2(0, 1)$  be a monotone non-decreasing function. We have*

$$h(1)R(1) - h(0)R(0) = \int_0^1 \dot{R}h dt + \int_0^1 R\dot{h}(dt). \quad (6.2)$$

**Remark 6.3.** In Lemma 6.2, observe that  $h \in \text{BV}_{\text{loc}}(0, 1)$  and if  $h(0^+) > -\infty$  then  $h \in \text{BV}(0, 3/4)$  and if  $h(1^-) < +\infty$  then  $h \in \text{BV}(1/4, 1)$ . Recall that when  $h(0^+)$  is finite, it is the trace of  $h$  at 0 and will simply be denoted as  $h(0)$ . Similarly, we denote as  $h(1)$  the trace of  $h$  at 1 when it exists. When  $R(0) = 0$ , even if  $h(0^+) = -\infty$ , we interpret  $R(0)h(0)$  as 0. Similarly if  $R(1) = 0$ , even if  $h(1^-) = \infty$ , in which case, we interpret  $R(1)h(1)$  as 0.

*Proof of Lemma 6.2.* For each natural number  $k$  we define the function  $\varphi_k \in W_0^{1,\infty}(0, 1)$  as

$$\varphi_k(t) := \begin{cases} kt & \text{if } 0 \leq t \leq k^{-1} \\ 1 & \text{if } \frac{1}{k} < t < 1 - k^{-1} \\ k(1-t) & \text{if } 1 - k^{-1} < t \leq 1. \end{cases}$$

Since  $R\varphi_k \in W_0^{1,\infty}(0,1)$  we have

$$0 = \int_0^1 \left( \frac{d(R\varphi_k)}{dt} h dt + R\varphi_k \dot{h}(dt) \right) = \int_0^1 \varphi_k \left( \dot{R} h dt + R \dot{h}(dt) \right) + k \int_0^{\frac{1}{k}} R h dt - k \int_{1-\frac{1}{k}}^1 R h dt. \quad (6.3)$$

We use the dominated convergence theorem and then the monotone convergence theorem (since  $\dot{h}$  is a Borel regular measure) to obtain

$$\lim_{k \rightarrow \infty} \int_0^1 \dot{R} \varphi_k h dt = \int_0^1 \dot{R} h dt \quad \text{and} \quad \lim_{k \rightarrow \infty} \int_0^1 R \varphi_k \dot{h}(dt) = \int_0^1 R \dot{h}(dt). \quad (6.4)$$

Observe that

$$\left| \int_0^{\frac{1}{k}} R h dt - R(0) \int_0^{\frac{1}{k}} h dt \right| \leq \frac{\text{Lip}(R)}{k} \int_0^{\frac{1}{k}} |h| dt, \quad \left| \int_{1-\frac{1}{k}}^1 R h dt - R(1) \int_{1-\frac{1}{k}}^1 h dt \right| \leq \frac{\text{Lip}(R)}{k} \int_{1-\frac{1}{k}}^1 |h| dt.$$

Since, when we use the above interpretation of  $R(0)h(0)$  and  $R(1)h(1)$  we have

$$\lim_{k \rightarrow \infty} R(0)k \int_0^{\frac{1}{k}} h dt = R(0)h(0) \quad \text{and} \quad \lim_{k \rightarrow \infty} R(1)k \int_{1-\frac{1}{k}}^1 h dt = R(1)h(1)$$

we conclude that

$$\lim_{k \rightarrow \infty} k \int_0^{\frac{1}{k}} R h dt = R(0)h(0) \quad \text{and} \quad \lim_{k \rightarrow \infty} k \int_{1-\frac{1}{k}}^1 R h dt = R(1)h(1). \quad (6.5)$$

Combining (6.3–6.5) we verify (6.2).  $\square$

**Definition 6.4.** Let  $\lambda \in \text{BV}_{\text{loc}}(0,1; \mathbb{R}^n)$  such that the distributional derivative  $\dot{\lambda}$  is the sum of an absolutely continuous part  $\dot{\lambda}^{\text{abs}} \mathcal{L}^1$  and a singular part (a Borel regular measure)  $\dot{\lambda}^{\text{sing}}$ . Here,  $\dot{\lambda}^{\text{abs}} : (0,1) \rightarrow (-\infty, 0]^n$  is a Borel function. Choose a non-negative Borel regular measure  $\nu$  such that  $-\dot{\lambda}_i^{\text{sing}} \ll \nu$ , and  $\nu$  and  $\mathcal{L}^1$  are mutually singular. We say that  $\lambda$  belongs to  $\mathcal{B}_*$  if

$$H \left( \dot{\lambda}^{\text{abs}}, \nabla_G \lambda \right) = 0 \quad \mathcal{L}^1 \quad \text{a.e. in } (0,1), \quad (6.6)$$

and

$$\max_{i=1 \leq i \leq n} \left\{ \frac{d\dot{\lambda}_i^{\text{sing}}}{d\nu} \right\} = 0, \quad \nu \quad \text{a.e. in } (0,1). \quad (6.7)$$

**Lemma 6.5.** Let  $\lambda \in L^2(0,1; \mathbb{R}^n)$  be such that  $\lambda_i$  is monotone non-increasing for any  $i \in \{1, \dots, n\}$ . Let  $(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)$  be such that  $\rho$  is Lipschitz,  $\rho_i(0) = 0$  whenever  $\lambda_i \notin \text{BV}(0, 0.75)$ , and  $\rho_i(1) = 0$  whenever  $\lambda_i \notin \text{BV}(0.25, 1)$ . Let  $\nu$  be the one in Definition 6.4. If

$$\begin{cases} H(\dot{\lambda}^{\text{abs}}, \nabla_G \lambda) \leq 0 & \mathcal{L}^1 \quad \text{a.e. in } (0,1) \\ H_0 \left( \frac{d\dot{\lambda}^{\text{sing}}}{d\nu} \right) \leq 0 & \nu \quad \text{a.e. in } (0,1), \end{cases} \quad (6.8)$$

then

$$\mathcal{A}(\rho, m) \geq (\lambda(1), \rho^1) - (\lambda(0), \rho^0).$$

Here, for each  $i \in \{1, \dots, n\}$ , we have interpreted  $\lambda_i(0)\rho_i^0$  and as  $\lambda_i(1)\rho_i^1$  as in Remark 6.2.

*Proof.* To avoid trivialities, we assume that  $\mathcal{A}(\rho, m) < \infty$ , in which case for  $\mathcal{L}^1$  almost every  $t \in (0, 1)$ ,  $m_{ij}(t) = 0$  if  $g_{ij}(\rho(t)) = 0$ . By Lemma 3.5 (ii) we have

$$F(\rho, m) + \|b\|_\rho^2 > 2(m, b) \quad (6.9)$$

unless  $m_{ij} = g_{ij}(\rho)b_{ij}$  for all  $(i, j) \in E$ . We have

$$\begin{aligned} \int_0^1 \frac{1}{2} F(\rho, m) dt &\geq \int_0^1 \left( \frac{1}{2} F(\rho, m) dt + H(\dot{\lambda}^{\text{abs}}, \nabla_G \lambda) dt + H_0 \left( \frac{d\dot{\lambda}^{\text{sing}}}{d\nu} \right) d\nu \right) \\ &\geq \int_0^1 \left( \frac{1}{2} F(\rho, m) dt + (\dot{\lambda}^{\text{abs}}, \rho) dt + \frac{1}{2} \|\nabla_G \lambda\|_\rho^2 dt + \left( \frac{d\dot{\lambda}^{\text{sing}}}{d\nu}, \rho \right) d\nu \right) \\ &= \int_0^1 \left( \frac{1}{2} F(\rho, m) dt + (\dot{\lambda}(dt), \rho) + \frac{1}{2} \|\nabla_G \lambda\|_\rho^2 dt \right). \end{aligned} \quad (6.10)$$

We use Remark 6.2, then the fact that  $(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)$ , to obtain after integrating by parts,

$$\int_0^1 (\dot{\lambda}(dt), \rho) = - \int_0^1 (\lambda, \dot{\rho}) dt + (\lambda(1), \rho^1) - (\lambda(0), \rho^0) = \int_0^1 (\lambda, \nabla_G \cdot m) dt + (\lambda(1), \rho^1) - (\lambda(0), \rho^0).$$

This, together with (6.10) yields

$$\mathcal{A}(\rho, m) \geq \int_0^1 \left( \frac{1}{2} F(\rho, m) - (\nabla_G \lambda, m) dt + \frac{1}{2} \|\nabla_G \lambda\|_\rho^2 \right) dt + (\lambda(1), \rho^1) - (\lambda(0), \rho^0).$$

Thanks to (6.9), we reach the desired conclusions.  $\square$

**Remark 6.6.** Let  $(\rho, m)$  and  $\lambda$  be as in Lemma 6.5.

(i) A necessary and sufficient condition to have  $\mathcal{A}(\rho, m) = (\lambda(1), \rho^1) - (\lambda(0), \rho^0)$  is that

$$(a) \quad m_{ij} = g_{ij}(\rho)(\nabla_G \lambda)_{ij} \quad \mathcal{L}^1 \quad \text{a.e in } (0, 1) \text{ for all } (i, j) \in E. \quad (6.11)$$

$$(b) \quad 0 = H(\dot{\lambda}^{\text{abs}}, \nabla_G \lambda) = (\dot{\lambda}^{\text{abs}}, \rho) + \frac{1}{2} \|\nabla_G \lambda\|_\rho^2 \quad \mathcal{L}^1 \quad \text{a.e. in } (0, 1), \quad (6.12)$$

$$0 = H_0 \left( \frac{d\dot{\lambda}^{\text{sing}}}{d\nu} \right) = \left( \frac{d\dot{\lambda}^{\text{sing}}}{d\nu}, \rho \right) \quad \nu \quad \text{a.e. in } (0, 1). \quad (6.13)$$

(ii) For any  $0 \leq s < t \leq 1$  we have

$$(\lambda(t), \rho(t)) = \min_{p \in \mathcal{P}(G)} \left\{ (\lambda(0), p) + \frac{1}{2(t-s)} \mathcal{W}_g^2(\rho(t), p) \right\}$$

## 7. CHARACTERIZATION OF GEODESICS AND EXTENDED HAMILTON–JACOBI EQUATIONS

Throughout this section, we assume (2.8) hold and characterize the minimizers of (5.2).

**Lemma 7.1.** *Let  $\lambda \in H^1(0, 1; \mathbb{R}^n)$  be such that*

$$H(\dot{\lambda}, \nabla_G \lambda) = 0 \text{ a.e. in } (0, 1). \quad (7.1)$$

(i) *We have  $\dot{\lambda}_i \leq 0$  a.e. in  $(0, 1)$  for any  $i \in V$ .*

(ii) *If we further assume that  $\gamma_P(\rho^0), \gamma_P(\rho^1) > 0$  and  $C$  a constant such that*

$$(\lambda(1), \rho^1) - (\lambda(0), \rho^0) \geq C$$

*then there exists  $C_0$  depending on  $C, \gamma_P(\rho^0), \gamma_P(\rho^1), w, n$  such that  $\|\nabla_G \lambda\|_{L^2(0,1)} \leq C_0$ .*

*Proof.* (i) Since (7.1) holds for  $\lambda$  a.e. on  $(0, 1)$

$$(\dot{\lambda}, \rho) + \frac{1}{2} \|\nabla_G \lambda\|_\rho^2 \leq 0 \text{ for any } \rho \in \mathcal{P}(G). \quad (7.2)$$

For  $i \in \{1, \dots, n\}$  fixed, we set  $\rho_j = \delta_{ij}$  to discover that (i) holds.

(ii) Let

$$\tilde{\rho}(t) := (1-t)\rho^0 + t\rho^1 \in \mathcal{C}(\rho^0, \rho^1). \quad (7.3)$$

Let  $\bar{\rho} := (1/n, \dots, 1/n) \in \mathcal{P}_0(G)$ . By Remark 4.2, there is a unique  $l_{\bar{\rho}}(\rho^1 - \rho^0)$  such that

$$\dot{\tilde{\rho}} = -\operatorname{div}_{\bar{\rho}}(l_{\bar{\rho}}(\rho^1 - \rho^0)) = -g\left(\frac{1}{n}, \frac{1}{n}\right) \operatorname{div}_G(l_{\bar{\rho}}(\rho^1 - \rho^0)) \quad \text{and} \quad l_{\bar{\rho}}(\rho^1 - \rho^0) \in T_{\bar{\rho}}\mathcal{P}(G).$$

Note

$$\tilde{m} := g\left(\frac{1}{n}, \frac{1}{n}\right) l_{\bar{\rho}}(\rho^1 - \rho^0) \in L^2(0, 1; S^{n \times n}) \quad (7.4)$$

satisfies

$$\dot{\tilde{\rho}} + \operatorname{div}_G(\tilde{m}) = 0. \quad (7.5)$$

Using Lemma 3.1, we have

$$\gamma_P(\tilde{\rho}(t)) \geq (1-t)\gamma_P(\rho^0) + t\gamma_P(\rho^1) \geq \min\{\gamma_P(\rho^0), \gamma_P(\rho^1)\} =: \epsilon_1. \quad (7.6)$$

Using (7.5) we obtain

$$C \leq (\lambda(1), \rho^1) - (\lambda(0), \rho^0) = \int_0^1 (\dot{\lambda}, \tilde{\rho}) + (\lambda, \dot{\tilde{\rho}}) dt = \int_0^1 (\dot{\lambda}, \tilde{\rho}) - (\lambda, \operatorname{div}_G(\tilde{m})) dt. \quad (7.7)$$

Setting  $\tilde{\lambda}_i := \lambda_i - 1/n \sum_{j=1}^n \lambda_j$ , using Lemma 3.2, we observe that (7.2), together with (7.7) implies

$$\begin{aligned} C &\leq \int_0^1 \left( -\frac{1}{2} \|\nabla_G \tilde{\lambda}\|_{\tilde{\rho}}^2 - (\tilde{\lambda}, \operatorname{div}_G(\tilde{m})) \right) dt \leq \int_0^1 \left( -\frac{\gamma_P(\tilde{\rho})}{2} \|\tilde{\lambda}\|^2 - (\tilde{\lambda}, \operatorname{div}_G(\tilde{m})) \right) dt \\ &\leq \int_0^1 \left( -\frac{\epsilon_1}{2} \|\tilde{\lambda}\|^2 - (\tilde{\lambda}, \operatorname{div}_G(\tilde{m})) \right) dt. \end{aligned}$$

This reads off

$$\frac{\epsilon_1}{2} \int_0^1 \|\tilde{\lambda}\|^2 dt + \int_0^1 (\tilde{\lambda}, \operatorname{div}_G(\tilde{m})) dt \leq -C.$$

Therefore, there exists a constant  $C_1$  depending only on  $C$ ,  $\epsilon_1$  and  $\int_0^1 \|\operatorname{div}_G(\tilde{m})\|^2 dt$  such that

$$\int_0^1 \|\tilde{\lambda}\|^2 dt \leq C_1.$$

Hence, there exists  $C_0$  depending on  $C$ ,  $\gamma_P(\rho^0)$ ,  $\gamma_P(\rho^1)$ ,  $w$  and  $n$  such that

$$\int_0^1 \|\nabla_G \tilde{\lambda}\|^2 dt \leq C_0.$$

Since  $\nabla_G \tilde{\lambda} = \nabla_G \lambda$  we conclude the proof.  $\square$

**Remark 7.2.** Since  $\rho^0, \rho^1 \in \mathcal{P}(G)$ , there are  $i, j \in \{1, \dots, n\}$  such that  $\rho_i^0, \rho_j^1 > 0$ . The following lemma draws consequences from these facts.

**Lemma 7.3.** *Assume  $\gamma_P(\rho^0), \gamma_P(\rho^1) > 0$  and  $\lambda$  is as in Lemma 7.1 and satisfies (ii) of that lemma. Then there exists some constant  $C_1$  depending on  $C$ ,  $\gamma_P(\rho^0)$ ,  $\gamma_P(\rho^1)$ ,  $w$ ,  $n$  such that*

- (i) if  $\rho_i^0 \geq \epsilon > 0$  for some  $i \in V$ ,  $\delta \|\dot{\lambda}_i\|_{L^1(0,1-\delta)} \leq C_1 \epsilon^{-1}$  for any  $0 < \delta < 1$ ;
- (ii) if  $\rho_i^1 \geq \epsilon > 0$  for some  $i \in V$ ,  $\delta \|\dot{\lambda}_i\|_{L^1(\delta,1)} \leq C_1 \epsilon^{-1}$  for any  $0 < \delta < 1$ ;
- (iii) if  $\rho_i^0, \rho_i^1 \geq \epsilon > 0$  for some  $i \in V$ ,  $\|\dot{\lambda}_i\|_{L^1(0,1)} \leq C_1 \epsilon^{-1}$ ;
- (iv)  $\|\lambda\|_{L^2(0,1)} \leq C_1$ .

*Proof.* Let  $(\tilde{\rho}(t), \tilde{m}(t))$  be the pair defined in (7.3) and (7.4). We exploit (7.7) to obtain

$$C \leq \int_0^1 (\dot{\lambda}, \tilde{\rho}) dt + \|\nabla_G \lambda\|_{L^2(0,1)} \|\tilde{m}\|_{L^2(0,1)}.$$

We use Lemma 7.1 to obtain a constant  $C_1$  depending on  $C$ ,  $\gamma_P(\rho^0)$ ,  $\gamma_P(\rho^1)$ ,  $w$ ,  $n$  such that

$$-\int_0^1 (\dot{\lambda}, \tilde{\rho}) dt \leq C_1.$$

From one line to another, we may increase the value of  $C_1$  when necessary.

(i) By Lemma 7.1,  $\dot{\lambda}_j \leq 0$  for any  $j \in \{1, \dots, n\}$  and so,  $-\dot{\lambda}_i \tilde{\rho}_i \leq -(\dot{\lambda}, \tilde{\rho})$ . Thus, if  $\rho_i^0 \geq \epsilon > 0$  then  $\tilde{\rho}_i \geq \delta \epsilon$  on  $t \in [0, 1 - \delta]$ . Hence,

$$\int_0^{1-\delta} \delta \epsilon |\dot{\lambda}_i| dt \leq - \int_0^1 \tilde{\rho}_i(t) \dot{\lambda}_i dt \leq - \int_0^1 (\dot{\lambda}, \tilde{\rho}) dt \leq C_1.$$

The proofs of (ii) and (iii) follow the same lines of argument as that of (i).

(iv) Observe that as  $\rho^0, \rho^1 \in \mathcal{P}(G)$  there exist  $i, j \in V$  such that  $n\rho_i^0, n\rho_j^1 \geq 1$ . Adding a constant to  $\lambda_i$  if necessary, without loss of generality we assume that  $\lambda_i(0) = 0$ . By (i), we have

$$|\lambda_i(t)| = \int_0^t |\dot{\lambda}_i| ds \leq C_1 \tag{7.8}$$

for  $t \in (0, 0.75]$ . Since the graph  $G$  is connected and  $\|\nabla_G \lambda\|_{L^2} \leq C_0$ , we obtain, for a bigger constant we still denote as  $C_1$ ,

$$\|\lambda\|_{L^2(0, 0.75)} \leq C_1. \tag{7.9}$$

Therefore, the set  $\mathcal{T}$  of  $t \in [0.25, 0.75]$  such that  $|\lambda| \leq 2C_1$  is a set of positive measure. Using (ii), we have

$$\int_{t_0}^t |\dot{\lambda}_j| ds \leq C_1 \tag{7.10}$$

if  $0.25 \leq t_0 \leq t \leq 1$ . In particular, taking  $t_0 \in \mathcal{T}$  and increasing the value of  $C_1$ , (7.10) implies

$$\|\lambda_j\|_{L^\infty(0.25, 1)} \leq C_1. \tag{7.11}$$

As above, we use again the fact that the graph  $G$  is connected and  $\|\nabla_G \lambda\|_{L^2(0,1)} \leq C_0$  to obtain

$$\|\lambda\|_{L^2(0.25, 1)} \leq C_1. \tag{7.12}$$

This, together with (7.9) proves (iv).  $\square$

**Theorem 7.4.** *Let  $B_*$  be as in Definition 6.4. Assume  $\gamma_P(\rho^0), \gamma_P(\rho^1) > 0$ .*

(i) *We have*

$$\min_{(\rho, m)} \left\{ \mathcal{A}(\rho, m) \mid (\rho, m) \in \mathcal{C}(\rho^0, \rho^1) \right\} = \sup_{\lambda} \left\{ (\lambda(1), \rho^1) - (\lambda(0), \rho^0) \mid \lambda \in B_* \right\}.$$

(ii) *Then there exists  $\lambda^* \in B_*$  such that*

$$(\lambda^*(1), \rho^1) - (\lambda^*(0), \rho^0) = \sup_{\lambda \in B} \left\{ (\lambda(1), \rho^1) - (\lambda(0), \rho^0) \mid H(\dot{\lambda}, \nabla_G \lambda) = 0 \right\}.$$

*Proof.* (i) Since  $B_* \cap B$  is a subset of  $B_*$ , Lemma 6.5 and Theorem 5.10 imply

$$\min_{(\rho, m)} \left\{ \mathcal{A}(\rho, m) \mid (\rho, m) \in \mathcal{C}(\rho^0, \rho^1) \right\} = \sup_{\lambda} \left\{ (\lambda(1), \rho^1) - (\lambda(0), \rho^0) \mid \lambda \in B_* \right\}. \tag{7.13}$$

(ii) Let  $\{\lambda^l\}_l \subset B$  be the maximizing sequence of

$$\sup_{\lambda \in B} \left\{ (\lambda(1), \rho^1) - (\lambda(0), \rho^0) \mid H(\dot{\lambda}, \nabla_G \lambda) = 0 \right\} =: M.$$

Without loss of generality we can assume that

$$M - 1 \leq (\lambda^l(1), \rho^1) - (\lambda^l(0), \rho^0). \quad (7.14)$$

Lemma 7.3 ensures  $\|\lambda^l\|_{L^2(0,1)} \leq C$  for a constant  $C$  independent of  $l$ . Therefore, for any  $i \in V$  and any integer  $k \geq 2$ , there exist  $s_{i,k} \in [0, \frac{1}{k}]$  and  $\tilde{s}_{i,k} \in [1 - \frac{1}{k}, 1]$  such that

$$\max \{ |\lambda_i^l(s_{i,k})|, |\lambda_i^l(\tilde{s}_{i,k})| \} \leq kC.$$

Since by Lemma 7.1,  $\dot{\lambda}_i^l \leq 0$  a.e. in  $(0, 1)$  for any  $i \in V$ , we obtain

$$\left\| \dot{\lambda}_i^l \right\|_{L^1(\frac{1}{k}, 1 - \frac{1}{k})} + \left\| \lambda_i^l \right\|_{L^1(\frac{1}{k}, 1 - \frac{1}{k})} =: \left\| \lambda_i^l \right\|_{\text{BV}(\frac{1}{k}, 1 - \frac{1}{k})} \leq 2Ck + (k - 2)C.$$

Thus, there is an increasing sequence  $(n_l) \subset \mathbb{N}$  and  $\lambda \in \text{BV}_{\text{loc}}(0, 1)$  such that

- (i)  $(\lambda^{n_l})_l$  converges weakly to  $\lambda^*$  in  $L^2(0, 1)$ , in  $\text{BV}_{\text{loc}}(0, 1)$  and strongly in  $L^2_{\text{loc}}(0, 1)$ .
- (ii) for any  $k \in \mathbb{N}$ ,  $\|\lambda_i^* \|_{\text{BV}(\frac{1}{k}, 1 - \frac{1}{k})} \leq 3Ck$  for any  $i \in V$ .

We denote the singular part of  $-\dot{\lambda}_i^*$  as  $-\dot{\lambda}_i^{*\text{sing}}$  and denote the absolutely continuous part as  $-\dot{\lambda}_i^{*\text{abs}} \mathcal{L}^1$ . Let  $I$  be the set of  $i \in \{1, \dots, n\}$  such that  $\rho_i^0 > 0$  and let  $J$  be the set of  $i \in \{1, \dots, n\}$  such that  $\rho_i^1 > 0$ . By Lemma 7.3  $(\lambda_i^{n_l})_l$  is a bounded sequence in  $\text{BV}(0, 0.75)$  for any  $i \in I$  and is a bounded sequence in  $\text{BV}(0.25, 1)$  for any  $i \in J$ . By the convergence of traces of functions of bounded variations, we may assume that

$$\lim_{l \rightarrow +\infty} \lambda_i^{n_l}(0) = \lambda_i^*(0) \quad \text{and} \quad \lim_{l \rightarrow +\infty} \lambda_j^{n_l}(1) = \lambda_j^*(1) \quad \forall (i, j) \in I \times J. \quad (7.15)$$

Let  $\mathcal{C}_G$  be the set of Borel maps of  $(0, 1)$  into  $\mathcal{P}(G)$ . Fix  $\varphi \in C_c^0(0, 1)$  non-negative,  $\rho \in \mathcal{C}_G$  and  $m \in L^2(0, 1; S^{n \times n})$  such that  $\mathcal{A}(\rho, m) < \infty$ . We use Lemma 3.5 (ii) to infer

$$\begin{aligned} 0 &= \lim_{l \rightarrow +\infty} \int_0^1 \varphi(t) H \left( \frac{d\lambda^{n_l}}{dt}, \nabla_G \lambda^{n_l} \right) dt \\ &\geq \lim_{l \rightarrow +\infty} \int_0^1 \varphi(t) \left( (\dot{\lambda}^{n_l}, \rho) - \frac{1}{2} F(\rho, m) + (m, \nabla_G \lambda^{n_l}) \right) dt \\ &= \int_0^1 \varphi(t) \left( (\dot{\lambda}^*, \rho) - \frac{1}{2} F(\rho, m) + (m, \nabla_G \lambda^*) \right) dt. \end{aligned}$$

Setting  $m_{ij} = g_{ij}(\rho)(\nabla_G \lambda^*)_{ij}$  we use Lemma 3.5 (ii) again to conclude that

$$0 \geq \int_0^1 \varphi(t) \left( (\dot{\lambda}^*, \rho) + \frac{1}{2} \|\nabla_G \lambda^*\|_\rho^2 \right) dt.$$

Since  $\varphi$  and  $\rho$  are arbitrary, we use Lemma 6.1 to verify that

$$\begin{cases} H(\dot{\lambda}^{*\text{abs}}, \nabla_G \lambda^*) & \leq 0 & \mathcal{L}^1 & \text{a.e. in } (0,1) \\ H_0\left(\frac{d\dot{\lambda}^{*\text{sing}}}{d\nu}\right) & \leq 0 & \nu & \text{a.e. in } (0,1) \end{cases}$$

where  $\nu$  is a non-negative Borel regular measure such that  $-\dot{\lambda}_i^{\text{sing}} \ll \nu$ , and  $\nu$  and  $\mathcal{L}^1$  are mutually singular. Thanks to (7.15) and since  $\{\lambda^n\}_n \subset B$  is a maximizing sequence, we have

$$M = \lim_{l \rightarrow +\infty} (\lambda^{n_l}(1), \rho^1) - (\lambda^{n_l}(0), \rho^0) = \sum_{j \in J} \lambda_j^*(1) \rho_j^1 - \sum_{i \in I} \lambda_i^*(0) \rho_i^0.$$

In light of the convention in Remark 6.2 we infer

$$M = (\lambda^*(1), \rho^1) - (\lambda^*(0), \rho^0). \quad (7.16)$$

Let  $(\rho^*, m^*) \in H^1(0, 1; \mathbb{R}^n) \times L^2(0, 1; S^{n \times n})$  be a minimizer of (5.2). We combine (7.13) and (7.16) to obtain

$$\mathcal{A}(\rho^*, m^*) = (\lambda^*(1), \rho^1) - (\lambda^*(0), \rho^0).$$

This, together with Remark 6.6 (i) yields (6.6) and (6.7).  $\square$

**Theorem 7.5.** *Assume  $\gamma_P(\rho^0), \gamma_P(\rho^1) > 0$ ,  $(\rho, m) \in \mathcal{C}(\rho^0, \rho^1)$  and  $\mathcal{A}(\rho, m) < \infty$ . A necessary and sufficient condition of  $(\rho, m)$  to minimize  $\mathcal{A}$  over  $\mathcal{C}(\rho^0, \rho^1)$  is that there exists  $\lambda \in \text{BV}_{\text{loc}}(0, 1; \mathbb{R}^n)$  such that  $\nabla_G \lambda$  and the distributional derivative  $\dot{\lambda}$ , which is the sum of an absolutely continuous part  $\dot{\lambda}^{\text{abs}} \mathcal{L}^1$  and a singular part  $\dot{\lambda}^{\text{sing}}$ , satisfy (6.11)-(6.13)*

*Proof.* Suppose  $(\rho, m)$  to minimize  $\mathcal{A}$  over  $\mathcal{C}(\rho^0, \rho^1)$ . By Theorem 7.4, there is  $\lambda \in \text{BV}_{\text{loc}}(0, 1; \mathbb{R}^n)$  such that

$$(\lambda(1), \rho^1) - (\lambda(0), \rho^0) = \mathcal{A}(\rho, m).$$

We use Remark 6.6 (i) to conclude that (6.11)-(6.13) hold.

Conversely, suppose there is  $\lambda \in \text{BV}_{\text{loc}}(0, 1; \mathbb{R}^n)$  such that (6.11)-(6.13) hold. Relying on Remark 6.6 (i), we conclude that  $(\rho, m)$  minimizes  $\mathcal{A}$  over  $\mathcal{C}(\rho^0, \rho^1)$ .  $\square$

*Acknowledgements.* The research of WG was supported by NSF grant DMS-17 00 202 and AFOSR grant FA9550-18-1-0502. WL and CM acknowledge generous support of the AFOSR through grant FA9550-18-1-0502.

## REFERENCES

- [1] L. Ambrosio and W. Gangbo, Hamiltonian ODEs in the Wasserstein space of probability measures. *Commun. Pure Appl. Math.* **61** (2008) 18–53.
- [2] L. Ambrosio, N. Gigli and G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, 2nd edition. *Lectures in Mathematics ETH Zürich*. Birkhäuser Verlag, Basel (2008).
- [3] G. Buttazzo, Semicontinuity, Relaxation and Integral Representation in the Calculus of Variations. Vol. 207 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow (1989).
- [4] P. Cardaliaguet, Notes on Mean-Field Games, Lectures by P.L. Lions. Collège de France (2010).
- [5] P. Cardaliaguet, F. Delarue, J.-M. Lasry and P.-L. Lions, The master equation and the convergence problem in mean field games. Preprint [arXiv:1509.02505](https://arxiv.org/abs/1509.02505) (2015).
- [6] Y. Chen, W. Gangbo, T.T. Georgiou and A. Tannenbaum, On the Matrix Monge-Kantorovich Problem. Preprint [arXiv:1701.02826](https://arxiv.org/abs/1701.02826) [math] (2017).
- [7] S.-N. Chow, W. Huang, Y. Li and H. Zhou, Fokker–Planck equations for a free energy functional or Markov process on a graph. *Arch. Ration. Mech. Anal.* **203** (2012) 969–1008.

- [8] S.-N. Chow, W. Li and H. Zhou, Entropy dissipation of Fokker–Planck equations on graphs. *Discrete Contin. Dyn. Syst.* **38** (2018) 4929–4950.
- [9] S.-N. Chow, W. Li and H. Zhou, A discrete Schrödinger equation via optimal transport on graphs. Preprint [arXiv:1705.07583](https://arxiv.org/abs/1705.07583) [math] (2017).
- [10] I. Ekeland and R. Témam, Convex Analysis and Variational Problems. *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA (1999).
- [11] L.C. Evans and W. Gangbo, Differential equations methods for the Monge–Kantorovich mass transfer problem. *Mem. AMS* **137** (1999) 1–66.
- [12] W. Gangbo, An elementary proof of the polar decomposition of vector-valued functions. *Arch. Ration. Mech. Anal.* **128** (1995) 380–399.
- [13] W. Gangbo and R. McCann, Optimal maps in Monge’s mass transport problem. *C.R. Acad. Sci. Paris* **321** (1995) 1653–1658.
- [14] W. Gangbo and R. McCann, The geometry of optimal transport. *Acta Math.* **177** (1996) 113–161.
- [15] W. Gangbo, T. Nguyen and A. Tudorascu, Hamilton–Jacobi equations in the Wasserstein space. *Meth. Appl. Anal.* **15** (2008) 155–184.
- [16] W. Gangbo and A. Swiech, Existence of a solution to an equation arising from the theory of Mean Field Games. *J. Differ. Equ.* **259** (2015) 6573–6643.
- [17] W. Gangbo and A. Swiech, Metric viscosity solutions of Hamilton–Jacobi equations depending on local slopes. *Calc. Var. Partial Differ. Equ.* **54** (2015) 1183–1218.
- [18] W. Gangbo and A. Tudorascu, Lagrangian dynamics on an infinite-dimensional torus; a Weak KAM theorem. *Adv. Math.* **224** (2010) 260–292.
- [19] J.-M. Lasry and P.-L. Lions, Mean field games. *Jpn. J. Math.* **2** (2007) 229–260.
- [20] J. Maas, Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.* **261** (2011) 2250–2292.
- [21] J.-F. Mertens, S. Sorin and S. Zamir, Repeated games. Vol. 55 of *Econometric Society Monographs*. Cambridge University Press, New York, 2015.
- [22] A. Mielke, A gradient structure for reaction–diffusion systems and for energy-drift-diffusion systems. *Nonlinearity* **24** (2011) 1329–1346.
- [23] Y. Shu, Hamilton–Jacobi equations on graph and applications. *Potential Anal.* **48** (2018) 125–157.
- [24] C. Villani, Topics in optimal transportation. Vol. 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.