

# Transportation on spheres via an entropy formula

**Gordon Blower**

Department of Mathematics and Statistics,  
Lancaster University,  
Lancaster LA1 4YF (g.blower@lancaster.ac.uk)

13th July 2022

The paper proves transportation inequalities for probability measures on spheres for the Wasserstein metrics with respect to cost functions that are powers of the geodesic distance. Let  $\mu$  be a probability measure on the sphere  $\mathbf{S}^n$  of the form  $d\mu = e^{-U(x)}dx$  where  $dx$  is the rotation invariant probability measure, and  $(n-1)I + \text{Hess } U \geq \kappa_U I$ , where  $\kappa_U > 0$ . Then any probability measure  $\nu$  of finite relative entropy with respect to  $\mu$  satisfies  $\text{Ent}(\nu | \mu) \geq (\kappa_U/2)W_2(\nu, \mu)^2$ . The proof uses an explicit formula for the relative entropy which is also valid on connected and compact  $C^\infty$  smooth Riemannian manifolds without boundary. A variation of this entropy formula gives the Lichnérowicz integral.

*Key words:* Wasserstein metric; curvature; transport; convexity

*2020 Mathematics subject classification* 60E15, 58C35

## 1. Transportation on the sphere

Optimal transportation involves moving unit mass from one probability distribution to another, at minimal cost, where the cost is measured by Wasserstein's distance.

DEFINITION Let  $(M, d)$  be a compact metric space and let  $\mu$  and  $\nu$  be probability measures on  $M$ . Then for  $1 \leq p < \infty$ , Wasserstein's distance from  $\mu$  to  $\nu$  is  $W_p(\nu, \mu)$ , where

$$W_p(\nu, \mu)^p = \inf_{\pi} \left\{ \iint_{M \times M} d(x, y)^p \pi(dx dy) : \pi \in \text{Prob}(M \times M) \right\} \quad (1.1)$$

where the probability measure  $\pi$  has marginals  $\nu$  and  $\mu$ . (See [8], [14].)

Transportation inequalities are results that bound the transportation cost  $W_p(\nu, \mu)^p$  in terms of  $\mu$ ,  $\nu$  and geometrical quantities of  $(M, d)$ . Typically, one chooses  $\mu$  to satisfy special conditions, and then one imposes minimal hypotheses on  $\nu$ . In this section, we consider the case where  $(M, d)$  is the unit sphere  $\mathbf{S}^2$  in  $\mathbf{R}^3$ , and obtain transportation inequalities by vector calculus. In section two, we extend these methods to a connected, compact and  $C^\infty$  smooth Riemannian manifold  $(M, d)$ .

On  $\mathbf{S}^2$ , let  $\theta \in [0, 2\pi)$  be the longitude and  $\phi \in [0, \pi]$  the colatitude, so the area measure is  $dx = \sin \phi d\phi d\theta$ . Let  $ABC$  be a spherical triangle where  $A$  is the North Pole; then by [10]

the Green's function  $G(B, C) = -(4\pi)^{-1} \log(1 - \cos d(B, C))$  may be expressed in terms of longitude and co latitude of  $B$  and  $C$  via the spherical cosine formula. A related cost function is listed in [14], p 972. Given probability measures  $\mu$  and  $\nu$  on  $\mathbf{S}^2$ , we can form

$$G(\mu - \nu)(x) = \int_{\mathbf{S}^2} G(x, y)(\mu(dy) - \nu(dy))$$

with gradient in the  $x$  variable

$$\nabla G(\mu - \nu)(x) = \int_{\mathbf{S}^2} \nabla_x G(x, y)(\mu(dy) - \nu(dy)).$$

PROPOSITION 1.1. *Let  $\mu$  and  $\nu$  be nonatomic probability measures on  $\mathbf{S}^2$ . Then*

$$W_1(\mu, \nu) \leq \int_{\mathbf{S}^2} \|\nabla G(\mu - \nu)(x)\| dx. \quad (1.2)$$

*Proof.* The Green's function is chosen so that  $\nabla \cdot \nabla G(B, C) = \delta_B(C) - 1/(4\pi)$  in the sense of distributions. Given non-atomic probability measures  $\mu$  and  $\nu$  on  $\mathbf{S}^2$ , their difference  $\mu - \nu$  is orthogonal to the constants on  $\mathbf{S}^2$ , so for a 1-Lipschitz function  $\varphi : \mathbf{S}^2 \rightarrow \mathbf{R}$ , we have

$$\begin{aligned} \int_{\mathbf{S}^2} \varphi(x)(\mu(dx) - \nu(dx)) &= \int_{\mathbf{S}^2} \varphi(x) \nabla \cdot \nabla G(\mu - \nu)(x) dx \\ &= - \int_{\mathbf{S}^2} \nabla \varphi(x) \cdot \nabla G(\mu - \nu)(x) dx \end{aligned} \quad (1.3)$$

so by Kantorovich's duality theorem [8], the Wasserstein transportation distance is bounded by

$$W_1(\mu, \nu) \leq \int_{\mathbf{S}^2} \|\nabla G(\mu - \nu)(x)\| dx. \quad (1.4)$$

DEFINITION Suppose that  $\mu$  is a probability measure and  $\nu$  is a probability measure that is absolutely continuous with respect to  $\mu$ , so  $d\nu = v d\mu$  for some probability density function  $v \in L^1(\mu)$ . Then the relative entropy of  $\nu$  with respect to  $\mu$  is

$$\text{Ent}(\nu | \mu) = \int_{\mathbf{S}^2} \log v(y) \nu(dy), \quad (1.5)$$

where  $0 \leq \text{Ent}(\nu | \mu) \leq \infty$  by Jensen's inequality.

At  $x \in \mathbf{S}^2$ , we have tangent space  $T_x \mathbf{S}^2 = \{y \in \mathbf{R}^3 : x \cdot y = 0\}$ . For  $y \in T_x \mathbf{S}^2$  with  $\|y\| = 1$ , we consider  $\exp_x(ty) = x \cos t + y \sin t$  so that  $\exp_x(0) = x$ ,  $\|\exp_x(ty)\| = 1$  and  $(d/dt)_{t=0} \exp_x(ty) = y$ ; hence  $\exp_x : T_x \mathbf{S}^2 \rightarrow \mathbf{S}^2$  gives the exponential map. We let  $J_{\exp_x}$  be the Jacobian determinant of this map.

Suppose that  $\mu(dx) = e^{-U(x)}dx$  is a probability measure and  $\nu$  is a probability measure that is absolutely continuous with respect to  $\mu$ , so  $d\nu = v d\mu$ . We say that a Borel function  $\Psi : \mathbf{S}^2 \rightarrow \mathbf{S}^2$  induces  $\nu$  from  $\mu$  if  $\int f(y)\nu(dy) = \int f(\Psi(x))\mu(dx)$  for all  $f \in C(\mathbf{S}^2; \mathbf{R})$ . McCann [12] showed that there exists  $\Psi$  that gives the optimal transport strategy for the  $W_2$  metric; further, there exists a Lipschitz function  $\psi : \mathbf{S}^2 \rightarrow \mathbf{R}$  such that  $\Psi(x) = \exp_x(\nabla\psi(x))$ ; so that

$$W_2(\nu, \mu)^2 = \int_{\mathbf{S}^2} d(\Psi(x), x)^2 \mu(dx) = \int_{\mathbf{S}^2} \|\nabla\psi(x)\|^2 \mu(dx). \quad (1.6)$$

Talagrand developed  $T_p$  inequalities in which  $W_p(\nu, \mu)^p$  is bounded in terms of  $\text{Ent}(\nu | \mu)$ , as in [14] p 569. In [5] and [6], the authors obtain some functional inequalities that are related to  $T_p$  inequalities. Here we offer an approach that is more direct, and uses only basic differential geometry to augment McCann's fundamental result. The key point is an explicit formula for the relative entropy in terms of the optimal transport maps.

LEMMA 1.2. *Suppose that  $\nu$  has finite relative entropy with respect to  $\mu$ , and let*

$$H = \text{Hess}_x \psi(x) \quad \text{and} \quad A = \text{Hess}_x d(x, y)^2/2 \quad \text{at} \quad y = \Psi(x); \quad (1.7)$$

let  $\Psi_t(x) = \exp_x(t\nabla\psi(x))$  for  $t \in [0, 1]$ . Then the relative entropy satisfies

$$\begin{aligned} \text{Ent}(\nu | \mu) \geq \int_{\mathbf{S}^2} & \left( \text{trace}(H - \log(A + H)) - \log J_{\exp_x}(\nabla\psi(x)) \right. \\ & \left. + \int_0^1 (1-t) \frac{d^2}{dt^2} U(\Psi_t(x)) dt \right) \mu(dx). \end{aligned} \quad (1.8)$$

where  $A$  is positive definite,  $H$  is symmetric and  $A + H$  is also positive definite, and

$$\text{trace}(H - \log(A + H)) \geq 0. \quad (1.9)$$

If  $\psi \in C^2$ , then equality holds in (1.8).

*Proof.* To express the relative entropy in terms of the transportation map, we adapt an argument from [1]. We have  $\text{Ent}(\nu | \mu) = \int_{\mathbf{S}^2} \log v(\Psi(x)) \mu(dx)$ , where the integrand is

$$\log v(\Psi(x)) = U(\Psi(x)) - U(x) - \log J_{\Psi}(x), \quad (1.10)$$

where the final term arises from the Jacobian of the change of variable  $y = \Psi(x)$ , where  $\Psi = \Psi_1$  and  $\Psi_t(x) = \exp_x(t\nabla\psi(x))$ . We compute this Jacobian by the chain rule for derivatives with respect to  $x$ . Specifically by [6] p 622, we have  $\text{Hess}(\psi(x) + d(x, y)^2/2) \geq 0$  and

$$\log J_{\Psi}(x) = \log J_{\exp_x}(\nabla\psi(x)) + \log \det \text{Hess}(\psi(x) + d(x, y)^2/2) \quad (1.11)$$

where  $J_{\exp_x}$  is the Jacobian of  $\exp_x : T_x \mathbf{S}^2 \rightarrow \mathbf{S}^2$  and  $\text{Hess} = D_x^2$  is the Hessian, where the expression is evaluated at  $y = \exp_x(\nabla\psi(x))$ . For  $x \in \mathbf{S}^2$  and  $\tau \in \mathbf{R}^3$  such that  $x \cdot \tau = 0$ , we have  $\tau \in T_x \mathbf{S}^2$  and

$$\exp_x(\tau) = \cos(\|\tau\|)x + \frac{\sin(\|\tau\|)}{\|\tau\|}\tau; \quad (1.12)$$

see [5]. By a vector calculus computation, which we replicate from [5], one finds

$$J_{\exp_x}(\|\nabla\psi(x)\|) = \frac{\sin\|\nabla\psi(x)\|}{\|\nabla\psi(x)\|}. \quad (1.13)$$

With  $\psi : \mathbf{S}^2 \rightarrow \mathbf{R}$  we have  $\nabla\psi(x) \perp x$ , so  $0 = x \cdot \nabla\psi(x)$ , hence  $0 = \nabla\psi(x) + \text{Hess}(\psi(x))x$ . We write  $\theta = \|\nabla\psi(x)\|$  for the angle between  $x$  and  $\Psi(x)$  so

$$\Psi(x) = \exp_x(\nabla\psi(x)) = x \cos \theta + \frac{\sin \theta}{\theta} \nabla\psi(x);$$

let  $v = x \times \theta^{-1} \nabla\psi(x)$  where  $\times$  denotes the usual vector product; then  $\{x, \theta^{-1} \nabla\psi(x), v\}$  gives an orthonormal basis of  $\mathbf{R}^3$ . Hence

$$\frac{\partial \Psi}{\partial v} = v \cos \theta - \sin \theta \langle \nabla\theta, v \rangle x + \left( \cos \theta - \frac{\sin \theta}{\theta} \right) \langle \nabla\theta, v \rangle \frac{\nabla\psi(x)}{\theta} + \frac{\sin \theta}{\theta} \text{Hess}\psi(x)v,$$

and we obtain (1.13) from the final factor. Then by spherical trigonometry, we have

$$\cos d(\exp_x(\tau), y) = (\cos \|\tau\|) \cos d(x, y) + \frac{\sin \|\tau\|}{\|\tau\|} \langle \tau, y \rangle, \quad (1.14)$$

so we have  $\langle \nabla_x \cos d(x, y), \tau \rangle = \langle y, \tau \rangle$  and  $\langle \text{Hess}_x \cos d(x, y) \tau, \tau \rangle = -(\cos d(x, y)) \|\tau\|^2$ ; so

$$\langle A\tau, \tau \rangle = \frac{1}{2} \langle \text{Hess}_x \cos d(x, y)^2 \tau, \tau \rangle = \frac{d(x, y)}{\tan d(x, y)} \|\tau\|^2 + \left( 1 - \frac{d(x, y)}{\tan d(x, y)} \right) \frac{\langle y, \tau \rangle^2}{\sin^2 d(x, y)}; \quad (1.15)$$

hence  $A$  is positive definite and is a rank-one perturbation of a multiple of the identity matrix. Note that the formulas degenerate on the cut locus  $d(x, y) = \pi$ ; consider the international date line opposite the Greenwich meridian.

We have

$$\text{Ent}(\nu | \mu) = \int_{\mathbf{S}^2} (U(\Psi(x)) - U(x) - \log J_{\Psi}(x)) e^{-U(x)} dx \quad (1.16)$$

in which

$$U(\Psi(x)) - U(x) = \langle \nabla U(x), \nabla\psi(x) \rangle + \int_0^1 (1-t) \frac{d^2}{dt^2} U(\Psi_t(x)) dt, \quad (1.17)$$

and we can combine the first two terms in (1.16) by the divergence theorem so

$$\int_{\mathbf{S}^2} \langle \nabla U(x), \nabla\psi(x) \rangle e^{-U(x)} dx = \int_{\mathbf{S}^2} \nabla \cdot \nabla\psi(x) e^{-U(x)} dx. \quad (1.18)$$

Hence from (1.11) we have

$$\text{Ent}(\nu | \mu) = \int_{\mathbf{S}^2} (\nabla \cdot \nabla \psi(x) - \log J_{\Psi}(x)) \mu(dx) + \int_0^1 (1-t) \frac{d^2}{dt^2} U(\Psi_t(x)) dt \mu(dx), \quad (1.19)$$

in which the Alexandrov Hessian [6], [14] p 363 satisfies

$$\text{trace Hess}_x \psi(x) \leq \nabla \cdot \nabla \psi(x) = \Delta_D \psi(x), \quad (1.20)$$

where  $\Delta_D \psi$  is the distributional derivative of the Lipschitz function  $\psi$ ; so we recognise (1.8).

We have an orthonormal basis

$$\left\{ x, \frac{\nabla \psi(x)}{\|\nabla \psi(x)\|}, x \times \frac{\nabla \psi(x)}{\|\nabla \psi(x)\|} \right\} \quad (1.21)$$

for  $\mathbf{R}^3$  in which the final two vectors give an orthonormal basis for  $T_x \mathbf{S}^2$ . Then

$$\left\langle A \frac{\nabla \psi(x)}{\|\nabla \psi(x)\|}, \frac{\nabla \psi(x)}{\|\nabla \psi(x)\|} \right\rangle = 1 \quad (1.22)$$

and

$$\left\langle A \left( x \times \frac{\nabla \psi(x)}{\|\nabla \psi(x)\|} \right), x \times \frac{\nabla \psi(x)}{\|\nabla \psi(x)\|} \right\rangle = \frac{d(x, y)}{\tan d(x, y)}, \quad (1.23)$$

hence  $A$  and  $H$  have the form

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \frac{\|\nabla \psi(x)\|}{\tan \|\nabla \psi(x)\|} \end{bmatrix}, \quad H = \begin{bmatrix} h & \beta \\ \beta & k \end{bmatrix} \quad (1.24)$$

with respect to the stated basis of  $T_x \mathbf{S}^2$ .

The function  $f(x) = x - 1 - \log x$  for  $x > 0$  is convex and takes its minimum value at  $f(1) = 0$ . Let  $T$  be a self-adjoint matrix with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_n$  where  $\lambda_n > -1$ ; then the Carleman determinant of  $I + T$  is  $\det_2(I + T) = \prod_{j=1}^n (1 + \lambda_j) e^{-\lambda_j}$ . Since  $A + H$  is positive definite, as in [1] Corollary 4.3, we can apply the spectral theorem to compute the Carleman determinant and show that

$$-\log \det_2(A + H) = \text{trace}(A + H - I - \log(A + H)) \geq 0 \quad (1.25)$$

so

$$\begin{aligned} \text{trace}(H - \log(A + H)) &= \text{trace}(A + H - I - \log(A + H)) + \text{trace}(I - A) \\ &\geq 0 + 1 - \frac{\|\nabla \psi(x)\|}{\tan \|\nabla \psi(x)\|} \geq 0. \end{aligned} \quad (1.26)$$

PROPOSITION 1.3. *Suppose that the Hessian matrix of  $U$  satisfies*

$$\text{Hess } U(x) + I \geq \kappa_U I \quad (x \in \mathbf{S}^2) \quad (1.27)$$

for some  $\kappa_U > 0$ . Then  $\mu$  satisfies the transportation inequality

$$\text{Ent}(\nu \mid \mu) \geq \frac{\kappa_U}{2} W_2(\nu, \mu)^2. \quad (1.28)$$

This applies in particular when  $\mu$  is normalized surface area measure.

*Proof.* Let  $K : [0, \pi) \rightarrow \mathbf{R}$  be the function

$$K(\alpha) = 1 - \frac{\alpha}{\tan \alpha} + \log \frac{\alpha}{\sin \alpha} = \frac{d}{d\alpha} \left( \alpha \log \frac{\alpha}{\sin \alpha} \right). \quad (1.29)$$

Then from (1.13) and (1.26) we have

$$\int_{\mathbf{S}^2} \left( \nabla \cdot \nabla \psi(x) - \log J_{\Psi}(x) \right) \mu(dx) \geq \int_{\mathbf{S}^2} \left( -\log \det_2(A + H) + K(\|\nabla \psi(x)\|) \right) \mu(dx).$$

Considering the final integral in (1.8), we have

$$\frac{\partial \Psi_t(x)}{\partial t} = -\|\nabla \psi(x)\| \sin(t\|\nabla \psi(x)\|)x + \cos(t\|\nabla \psi(x)\|)\nabla \psi(x) \quad (1.30)$$

which has constant speed  $\|\frac{\partial \Psi_t(x)}{\partial t}\| = \|\nabla \psi(x)\|$  and  $\langle \frac{\partial \Psi_t(x)}{\partial t}, \Psi_t(x) \rangle = 0$ ; also

$$\frac{\partial^2}{\partial t^2} U(\Psi_t(x)) = \left\langle \text{Hess}U \circ \Psi_t(x) \frac{\partial \Psi_t(x)}{\partial t}, \frac{\partial \Psi_t(x)}{\partial t} \right\rangle - \|\nabla \psi(x)\|^2 \langle (\nabla U) \circ \Psi_t(x), \Psi_t(x) \rangle, \quad (1.31)$$

where the final term is zero since  $\nabla U \circ \Psi_t(x)$  is in the tangent space at  $\Psi_t(x)$ , hence is perpendicular to  $\Psi_t(x)$ . We therefore have the crucial inequality

$$\begin{aligned} \text{Ent}(\nu \mid \mu) &\geq \int_{\mathbf{S}^2} \left( -\log \det_2(A + H) + K(\|\nabla \psi(x)\|) \right. \\ &\quad \left. + \int_0^1 (1-t) \left\langle \text{Hess}U \circ \Psi_t(x) \frac{\partial \Psi_t(x)}{\partial t}, \frac{\partial \Psi_t(x)}{\partial t} \right\rangle dt \right) \mu(dx) \end{aligned} \quad (1.32)$$

To simplify the function  $K$ , we recall from [9] 8.342 the Maclaurin series

$$\begin{aligned} \log \frac{\alpha}{\sin \alpha} &= \log \Gamma\left(1 + \frac{\alpha}{\pi}\right) + \log \Gamma\left(1 - \frac{\alpha}{\pi}\right) \\ &= \sum_{m=1}^{\infty} \frac{\zeta(2m)}{\pi^{2m} m} \alpha^{2m} \quad (|\alpha| < \pi), \end{aligned} \quad (1.33)$$

where we have introduced Euler's  $\Gamma$  function and Riemann's  $\zeta$  function, so

$$K(\alpha) = \sum_{m=1}^{\infty} \frac{(2m+1)\zeta(2m)}{\pi^{2m} m} \alpha^{2m} \geq \frac{3\zeta(2)}{\pi^2} \alpha^2 = \frac{\alpha^2}{2}. \quad (1.34)$$

Now we consider (1.32) with the hypothesis (1.27) in force. The Carleman determinant contributes a nonnegative term as in (1.25), while the final integral in (1.32) combines with the integral of  $K(\|\nabla\psi(x)\|)$  to give

$$\begin{aligned} \text{Ent}(\nu | \mu) &\geq \int_{\mathbf{S}^2} (K(\|\nabla\psi(x)\|) + (1/2)\|\nabla\psi(x)\|^2)\mu(dx) \\ &\geq \frac{\kappa_U}{2} \int_{\mathbf{S}^2} \|\nabla\psi(x)\|^2\mu(dx) \\ &= \frac{\kappa_U}{2} W_2(\nu, \mu)^2. \end{aligned} \tag{1.35}$$

When  $\mu$  is normalized surface area,  $U$  is a constant and the hypothesis (1.27) holds with  $\kappa_U = 1$ .

## 2. Transportation on compact Riemannian manifolds

Let  $M$  be a connected, compact and  $C^\infty$  smooth Riemannian manifold of dimension  $n$  without boundary, and let  $g$  be the Riemannian metric tensor, giving metric  $d$ . Let  $\mu(dx) = e^{-U(x)}dx$  be a probability measure on  $M$  where  $dx$  is Riemannian measure and  $U \in C^2(M; \mathbf{R})$ . Suppose that  $\nu$  is a probability measure on  $M$  that is of finite relative entropy with respect to  $\mu$ . Then by McCann's theory [12], there exists a Lipschitz function  $\psi : M \rightarrow \mathbf{R}$  such that  $\Psi(x) = \exp_x(\nabla\psi(x))$  induces  $\nu$  from  $\mu$ . Then we let  $\Psi_t(x) = \exp_x(t\nabla\psi(x))$ . We proceed to compute quantities which we need for our extension of Lemma 1.2.

Given distinct points  $x, y \in M$ , we suppose that  $x = \exp_y(\xi)$ , and for  $w \in T_yM$  introduce

$$\gamma(s, t) = \exp_y(t(\xi + sw)) \tag{2.1}$$

so that  $t \mapsto \gamma(s, t)$  is a geodesic, and in particular  $\gamma(0, t)$  is the geodesic from  $y = \gamma(0, 0)$  to  $x = \gamma(0, 1)$ . When  $y = \exp_x(\nabla\psi(x))$  for a Lipschitz function  $\psi : M \rightarrow \mathbf{R}$ , we can determine  $\xi$  as follows. Let  $\phi(z) = -\psi(z)$  and introduce its infimal convolution

$$\phi^c(y) = \inf_w \{(1/2)d(y, w)^2 - \phi(w)\} \tag{2.2}$$

which is attained at  $x$  since  $y = \exp_x(\nabla\psi(x)) = \exp_x(-\nabla\phi(x))$ . Now  $\phi^{cc}(x) = \phi(x)$ , so

$$\phi(x) = \inf_w \{(1/2)d(x, w)^2 - \phi^c(w)\} \tag{2.3}$$

where the infimum is attained at  $y$  since  $\phi(x) + \phi^c(y) = d(x, y)^2/2$ . By Lemma 2 of [12],  $\phi^c$  is Lipschitz and

$$x = \exp_y(-\nabla\phi^c(y)). \tag{2.4}$$

The speed of  $\gamma(0, t)$  is given by

$$\left\| \frac{\partial \gamma}{\partial t} \right\| = \|\nabla \phi^c(y)\| = d(y, \exp_y(-\nabla \phi^c(y))) = d(x, y) = d(x, \exp_x(-\nabla \phi(x))) = \|\nabla \psi(x)\|. \quad (2.5)$$

Let  $R$  be the curvature of the Levi-Civita derivation  $\nabla$  so

$$R(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z - \nabla_{[Y, X]} Z \quad (X, Y, Z \in T_x M).$$

Then by [13] page 36, for all  $Y \in T_x M$ , the curvature operator  $R_Y : X \mapsto R(X, Y)Y$  is self-adjoint with respect to the scalar product on  $T_x M$ . Also

$$Y(s, t) = \frac{\partial}{\partial s} \gamma(s, t) \quad (2.6)$$

satisfies the initial conditions

$$Y(s, 0) = 0, \quad \frac{\partial Y}{\partial t}(0, 0) = w, \quad (2.7)$$

and Jacobi's differential equation [4] (2.43)

$$\frac{\partial^2 Y}{\partial t^2} + R\left(\frac{\partial \gamma}{\partial t}, Y\right) \frac{\partial \gamma}{\partial t} = 0. \quad (2.8)$$

By calculating the first variation of the length formula [13] p 161, one shows that

$$\frac{1}{2} \left\langle \text{Hess}_x d(x, y)^2 Y(0, 1), Y(0, 1) \right\rangle = g\left(\frac{\partial Y}{\partial t}(0, 1), Y(0, 1)\right). \quad (2.9)$$

Assume that there are no conjugate points on  $\gamma(s, t)$ . Then by varying  $w$ , we can make  $Y(0, 1)$  cover a neighbourhood of 0 in  $T_x M$ . Let

$$A = \frac{1}{2} \text{Hess}_x d(x, y)^2 \Big|_{y=\exp_x(\nabla \psi(x))}, \quad (2.10)$$

and

$$H = \text{Hess} \psi(x). \quad (2.11)$$

Let  $J_{\exp_x}(v)$  be the Jacobian of the map  $T_x M \rightarrow M$  given by  $v \mapsto \exp_x(v)$ , as in (3.4) of [3].

LEMMA 2.1. *Suppose that  $\Psi_t(x) = \exp_x(t\nabla \psi(x))$ , where  $\Psi_1$  induces the probability measure  $\nu$  from  $\mu$  and gives the optimal transport map for the  $W_2$  metric. Then the relative entropy satisfies*

$$\begin{aligned} \text{Ent}(\nu | \mu) &\geq \int_M \left( \text{trace}(H - \log(A + H)) - \log J_{\exp_x}(\nabla \psi(x)) \right. \\ &\quad \left. + \int_0^1 (1-t) \left\langle \text{Hess} U \circ \Psi_t(x) \frac{\partial \Psi_t(x)}{\partial t}, \frac{\partial \Psi_t(x)}{\partial t} \right\rangle dt \right) \mu(dx). \end{aligned} \quad (2.12)$$



where  $H$  is symmetric and  $A + H$  is also positive definite. If  $\psi \in C^2(M; \mathbf{R})$ , then equality holds in (2.12).

*Proof.* This is similar to Lemma 1.2. As in (125), we have

$$\begin{aligned} \text{trace}(H - \log(A + H)) &= -\log \det(A + H) + \text{trace}(I - A) \\ &\geq \text{trace}(I - A), \end{aligned} \quad (2.13)$$

and by standard calculations [13] p32 we have

$$\frac{\partial^2}{\partial t^2} U(\Psi_t(x)) = \left\langle \text{Hess}U \circ \Psi_t(x) \frac{\partial \Psi_t(x)}{\partial t}, \frac{\partial \Psi_t(x)}{\partial t} \right\rangle \quad (2.14)$$

since  $\Psi_t(x)$  is a geodesic.

The curvature operator is the symmetric operator  $R_Z : Y \mapsto R(Z, Y)Z$ . If  $M$  has nonnegative Ricci curvature so that  $R_Z \geq 0$  as a matrix for all  $Z$ , then we have

$$-\log J_{\exp_x}(\nabla \psi(x)) \geq 0. \quad (2.15)$$

by (3.4) of [Ca].

The following result recovers the Lichnérowicz integral, as in (4.16) of [1] and (1.1) of [7]. This integral also appears implicitly in the Hessian calculations in Appendix D of [11]. Let  $\|H\|_{HS}$  be the Hilbert–Schmidt norm of  $H$ .

**PROPOSITION 2.2.** *Suppose that  $\psi \in C^2(M; \mathbf{R})$  and  $\Psi_\tau(x) = \exp_x(\tau \nabla \psi(x))$  induces a probability measure  $\nu_\tau$  from  $\mu$  such that  $\Psi_\tau$  is the optimal transport map for the  $W_2$  metric. Then*

$$\begin{aligned} \text{Ent}(\nu_\tau | \mu) &= \frac{\tau^2}{2} \int_M \left( \|\text{Hess} \psi(x)\|_{HS}^2 + \text{trace} R_{\nabla \psi(x)} \right. \\ &\quad \left. + \langle \text{Hess} U(x) \nabla \psi(x), \nabla \psi(x) \rangle \right) \mu(dx) + O(\tau^3) \quad (\tau \rightarrow 0+). \end{aligned} \quad (2.16)$$

*Proof.* For small  $\tau > 0$ , we rescale  $\psi$  to  $\tau\psi$  and consider  $y = \exp_x(\tau \nabla \psi(x))$ ; then we return to  $x$  along a geodesic  $\gamma_\tau(t) = \exp_y(-t \nabla(-\tau\psi)^c(y))$  for  $0 \leq t \leq 1$  with constant speed  $\tau \|\nabla \psi(x)\|$ . Observe that  $\tau\psi(x) = (-\tau\psi)^c(y) - \tau^2 \|\nabla \psi(x)\|^2/2$ , and  $\nabla_x d(x, y)^2/2 = -\exp_x^{-1}(y) = -\tau \nabla \psi(x)$  and  $\nabla_y d(x, y)^2/2 = -\exp_y^{-1}(x) = \nabla(-\tau\psi)^c(y)$  by Gauss's Lemma. Recalling that the curvature operator is self-adjoint by page 36 of [13], we choose the basis of  $T_y M$  so that the first basis vector points along the direction of the geodesic  $\gamma_\tau(0)$ . Hence Jacobi's equation (2.8) can be expressed as a second order differential equation in block matrix form, with a symmetric matrix  $S_{-\nabla(-\tau\psi)^c(y)}$  given by components of the curvature tensor such that

$$R\left(\frac{d\gamma_\tau}{dt}, Y\right) \frac{d\gamma_\tau}{dt} = \begin{bmatrix} 0 & 0 \\ 0 & S_{-\nabla(-\tau\psi)^c(y)} \end{bmatrix} Y \quad (0 < t < 1). \quad (2.17)$$

as in (2.4) of [6]. Then the Jacobi equation reduces to a first-order block matrix equation with blocks of shape  $(1 + (n - 1)) \times (1 + (n - 1))$  in a  $(2n) \times (2n)$  matrix

$$\frac{d}{dt} \begin{bmatrix} Y \\ V \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & I_{n-1} \\ 0 & 0 & 0 & 0 \\ 0 & -S_{-\nabla(-\tau\psi)^c(y)} & 0 & 0 \end{bmatrix} \begin{bmatrix} Y \\ V \end{bmatrix}; \quad \begin{bmatrix} Y(0) \\ V(0) \end{bmatrix} = \begin{bmatrix} 0 \\ w \end{bmatrix}. \quad (2.18)$$

To find the limit as  $\tau \rightarrow 0$ , we can assume that  $S_{-\nabla(-\tau\psi)^c(y)}$  is constant on the geodesic, and may be expressed as  $\tau^2 S$  where  $\tau^2 S = S_{\tau\nabla\psi(x)}$  has shape  $(n - 1) \times (n - 1)$ . The functions  $\cos \alpha$  and  $\sin \alpha / \alpha$  are entire and even, so  $\cos \sqrt{s}$  and  $\sin \sqrt{s} / \sqrt{s}$  are entire functions, hence they operate on complex matrices. Note that the matrix

$$T = \begin{bmatrix} 0 & 0 \\ 0 & S_{-\nabla(-\tau\psi)^c(y)} \end{bmatrix}$$

in the bottom left corner is symmetric, has rank less than or equal to  $n - 1$ , and does not depend upon  $t$ . Hence we consider the matrix

$$\begin{bmatrix} Y \\ V \end{bmatrix} = \begin{bmatrix} \cos(t\sqrt{T}) & \frac{\sin(t\sqrt{T})}{\sqrt{T}} \\ -\sqrt{T} \sin(t\sqrt{T}) & \cos(t\sqrt{T}) \end{bmatrix} \begin{bmatrix} Y_0 \\ V_0 \end{bmatrix}$$

which has derivative

$$\frac{d}{dt} \begin{bmatrix} Y \\ V \end{bmatrix} = \begin{bmatrix} 0 & I \\ -T & 0 \end{bmatrix} \begin{bmatrix} \cos(t\sqrt{T}) & \frac{\sin(t\sqrt{T})}{\sqrt{T}} \\ -\sqrt{T} \sin(t\sqrt{T}) & \cos(t\sqrt{T}) \end{bmatrix} \begin{bmatrix} Y_0 \\ V_0 \end{bmatrix}$$

so we can use this formula to solve (2.18). So the approximate differential equation has solution

$$\begin{bmatrix} Y(1) \\ V(1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & \cos \tau\sqrt{S} & 0 & \frac{\sin \tau\sqrt{S}}{\tau\sqrt{S}} \\ 0 & 0 & 1 & 0 \\ 0 & -\tau\sqrt{S} \sin \tau\sqrt{S} & 0 & \cos \tau\sqrt{S} \end{bmatrix} \begin{bmatrix} 0 \\ w \end{bmatrix}. \quad (2.19)$$

Hence by (2.9) we have

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \frac{\tau\sqrt{S}}{\tan \tau\sqrt{S}} \end{bmatrix} = (1 + O(\tau^2))I_n \quad (2.20)$$

which gives rise to the approximation

$$\text{trace}(I_n - A) = \text{trace}\left(I_{n-1} - \frac{\tau\sqrt{S}}{\tan \tau\sqrt{S}}\right) = \frac{\tau^2}{3} \text{trace}(S) + O(\tau^4) \quad (\tau \rightarrow 0+), \quad (2.21)$$

and likewise we obtain

$$-\log J_{\text{exp}_x}(\tau\nabla\psi(x)) = -\log \det \frac{\sin \tau\sqrt{S}}{\tau\sqrt{S}} = \frac{\tau^2}{6} \text{trace}(S) + O(\tau^4). \quad (2.22)$$

From (2.19), we have

$$\begin{aligned}
-\log \det_2(A + \tau H) &= \frac{1}{2} \text{trace}((A - I_n + \tau H)^2) + O(\tau^3) \\
&= \frac{\tau^2}{2} \text{trace}(H^2) + O(\tau^3) \\
&= \frac{\tau^2}{2} \|\text{Hess } \psi(x)\|_{HS}^2 + O(\tau^3),
\end{aligned} \tag{2.23}$$

so the result follows by Lemma 2.1.

We conclude with a transportation inequality which generalizes Proposition 1.3 to the unit spheres  $\mathbf{S}^n$ . See [2] for a discussion of measures on product spaces.

**THEOREM 2.3.** *Let  $M = \mathbf{S}^n$  for some  $n \geq 2$ , and suppose that*

$$(n-1)I + \text{Hess } U(x) \geq \kappa_U I \quad (x \in \mathbf{S}^n) \tag{2.24}$$

for some  $\kappa_U > 0$ . Then

$$\text{Ent}(\nu | \mu) \geq \frac{\kappa_U}{2} W_2(\nu, \mu)^2. \tag{2.25}$$

*Proof.* In this case, the curvature operator is constant, so we have  $S_{\nabla\psi(x)}Y = \|\nabla\psi(x)\|^2 Y$ , so

$$\text{trace } R_{\nabla\psi(x)} = (n-1)\|\nabla\psi(x)\|^2. \tag{2.26}$$

Thus the result follows with a similar proof to Proposition 1.3 using data from the proof of Proposition 2.2.

**Acknowledgement** I thank Graham Jameson for helpful remarks concerning inequalities which led to (1.34). I am also grateful to the referee, whose helpful comments improved the exposition.

## References

- [1] G. Blower, The Gaussian isoperimetric inequality and transportation, *Positivity* **7** (2003), 203-224.
- [2] G. Blower and F. Bolley, Concentration of measure on product spaces with applications to Markov processes, *Studia Math.* **175** (2006), 47-72.
- [3] X. Cabre, Nondivergent elliptic equations on manifolds with nonnegative curvature, *Comm. Pure Appl. Math.* **50** (1997), 623-665.
- [4] I. Chavel, *Riemannian Geometry: a modern introduction*, (Cambridge University Press, 1993).
- [5] D. Cordero-Erausquin, Prékopa–Leindler inequalities sur la sphère, *C.R. Acad. Sci. Paris* **329** (1999), 789-792.

- [6] D. Cordero-Erausquin, R.J. McCann and M. Schmuckensläger, Prékopa–Leindler type inequalities on Riemannian manifolds, Jacobi fields and optimal transport, *Annales. de la Fac. Sci. Toulouse Math.* **15** (2006), 613-635.
- [7] J.-D. Deuschel and D.W. Stroock, Hypercontractivity and spectral gap of symmetric diffusions with applications to stochastic Ising models, *J. Funct. Anal.* **92** (1990), 30-48.
- [8] R. M. Dudley, *Real Analysis and Probability*, second edition, (Cambridge University Press, 2004).
- [9] I. S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products*, (Academic Press, 1965).
- [10] Y. Kimura and H. Okamoto, Vortex motion on a sphere, *J. Phys. Soc. Japan* **56** (1987), 4203-4206.
- [11] J. Lott and C. Villani, Ricci curvature for metric measure spaces via optimal transport, *Annals of Math.* (2) **169** (2009), 903-991.
- [12] R. J. McCann, Polar factorization of maps on Riemannian manifolds, *Geom. Funct. Anal.* **11** (2001), 589-608.
- [13] P. Pedersen, *Riemannian Geometry*, second edition, (Springer, 2006).
- [14] C. Villani, *Optimal Transport: Old and New*, (Springer, 2009).