A GEOMETRIC APPROACH TO APRIORI ESTIMATES FOR OPTIMAL TRANSPORT MAPS

SIMON BRENDLE, FLAVIEN LÉGER, ROBERT J. MCCANN, AND CALE RANKIN

ABSTRACT. A key inequality which underpins the regularity theory of optimal transport for costs satisfying the Ma–Trudinger–Wang condition is the Pogorelov second derivative bound. This translates to an apriori interior C^1 estimate for smooth optimal maps. Here we give a new derivation of this estimate which relies in part on Kim, McCann and Warren's observation that the graph of an optimal map becomes a volume maximizing spacelike submanifold when the product of the source and target domains is endowed with a suitable pseudo-Riemannian geometry that combines both the marginal densities and the cost.

Keywords: Pogorelov a priori estimate, Monge-Ampère type, optimal transportation map, volume maximizing submanifold, split special Lagrangian geometry, pseudo-Riemannian, second boundary value, degenerate elliptic, interior regularity, Ma–Trudinger–Wang

MSC2020: 35B45, 35J96, 49Q22, 53C38, 53C50, 58J60,

1. INTRODUCTION

Apriori second-derivative estimates for solutions of the Monge–Ampère equation date back to Pogorelov [28, 29], who combined them with a third-derivative estimate of Calabi [4] to deduce the regularity of solutions to Minkowski's problem. Such estimates have a rich history [30, 31]. They play a key role in the regularity theory of optimal transportation, which involve Monge–Ampère type equations whose exact form depends on the cost optimized. For the quadratic cost function, this theory was developed by Delanoë [8] in the plane and by Urbas [36] in higher dimensions, parallel results being obtained using different techniques by Caffarelli [1, 2, 3]. At the same time, a regularity theory for reflector antenna design was developed by X.-J. Wang [39], which can also be seen as an optimal transport problem with the restriction of a logarithmic cost to the sphere [40]. For more general cost functions, Ma, Trudinger and Wang [24, 33] identified a sufficient condition for the regularity of optimal maps, whose necessity was deduced by Loeper

1

[22]. Although challenging to interpret, Ma, Trudinger and Wang's condition (A3) (and its weak variant (A3w)) were subsequently understood by Kim and McCann as the positivity (respectively non-negativity) of certain sectional curvatures of a pseudo-Riemannian metric induced on the product of the source and target domains by the transportation cost [16]. With Warren [18], Kim and McCann discovered that the graph of an optimal map becomes a volume maximizing spacelike submanifold when the same metric is conformally rescaled so that its volume gives the product of the source and target densities.

The purpose of the present article is to exploit this geometric perspective to give a new derivation of Pogorelov type apriori estimates for optimal maps with respect to costs satisfying the Ma–Trudinger–Wang condition. Our approach is different but related to the strategy developed for $m = 2 = n$ by Warren [41], who was indirectly inspired [20] by the same sources [38, 35] as us. It yields new insights, which we hope may lead to further progress: in particular, it allows us to separate the general structure of the maximum principle estimate from the particulars of the application.

Although it will not be shown here, interior Pogorelov estimates persist even under the weaker variant (A3w) of the Ma–Trudinger–Wang condition, but the argument is more involved [21]; regularity of the optimal transport map follows from appropriate domain convexity conditions [33]. For the quadratic cost, 2-uniformity of the convexity required [3, 36, 33] was recently relaxed by Chen, Liu and Wang [6, 7], though some smoothness remains necessary [15]. Finally, it is interesting to note that for the quadratic cost, Kim and McCann's metric [16] restricted to the graph of optimal map coincides with the Hessian metric of the solution of the Monge–Ampère equation, introduced by Calabi [4].

The next section of this note gives a general estimate for maximal spacelike submanifolds in a fixed pseudo-Riemannian background. The last section explains how to use this to recover and apply Ma, Trudinger and Wang's local C^2 estimate for optimal maps.

2. Estimates for maximal spacelike submanifolds

Our ambient space will be a fixed manifold $\hat{M}$ of dimension $n + m$. Let $\hat{g}$ be a pseudo-metric on $\hat{M}$ of signature (n, m). Let $\hat{D}$ denote the Levi-Civita connection on $\hat{M}$, and let $\hat{R}$ denote its Riemann curvature tensor. We adopt the sign convention where

$$\hat{R}(X, Y, Z, W) = -\hat{g}(\hat{D}_X \hat{D}_Y Z - \hat{D}_Y \hat{D}_X Z - \hat{D}_{[X,Y]} Z, W)$$

for all vector fields X, Y, Z, W on $\hat{M}$. Let $\hat{E}_1, \ldots, \hat{E}_{n+m}$ denote a local frame of vector fields on $\hat{M}$ which are linearly independent at each point. At each point on $\hat{M}$, the matrix $\hat{g}(\hat{E}_\alpha, \hat{E}_\beta)$ is invertible; we denote its inverse by $\hat{\sigma}^{\alpha\beta}$. We will also fix a smooth symmetric $(0, 2)$-tensor field $\hat{S}$ on $\hat{M}$.

We are interested in studying n-dimensional spacelike submanifolds of $\hat{M}$. Given such a submanifold M, let g and S denote the respective restrictions of $\hat{g}$ and $\hat{S}$ to M. Since M is spacelike, g is positive definite. As in [27], let D denote the Levi-Civita connection on M, and let $D^\perp$ denote the connection on the normal bundle of M. We denote by $II : TM \times TM \to T^\perp M$ the second fundamental form of M. In other words, if X and Y are two tangential vector fields on M, then

$$II(X,Y) = \hat{D}_X Y - D_X Y.$$

The mean curvature vector of M is defined as the trace of II, which we assume vanishes identically.

We will use the notation

$$(\hat{D}_V II)(X,Y) = \hat{D}_V(II(X,Y)) - II(D_V X, Y) - II(X, D_V Y)$$

and

$$(D_V^\perp II)(X,Y) = D_V^\perp(II(X,Y)) - II(D_V X, Y) - II(X, D_V Y)$$

for all tangential vector fields X, Y, V on M (compare [27], p. 114). Both expressions are tensorial in X, Y, V. Moreover, $(D_V^\perp II)(X,Y) \in T^\perp M$ is the normal component of $(\hat{D}_V II)(X,Y) \in T\hat{M}$.

In the following, we assume that $\{e_1, \ldots, e_n\}$ is a local orthonormal frame on M and $\{e_1^\perp, \ldots, e_m^\perp\}$ is a local orthonormal frame for the normal bundle of M. Then $\hat{g}(e_k, e_l) = \delta_{kl}$, $\hat{g}(e_p^\perp, e_q^\perp) = -\delta_{pq}$, and $\hat{g}(e_k, e_p^\perp) = 0$.

Given $(\hat{M}, \hat{g}, \hat{S})$, we consider an n-dimensional spacelike submanifold $M \subset \hat{M}$ with zero mean curvature. We will show that the restriction S of $\hat{S}$ to M satisfies an elliptic partial differential equation with coefficients and inhomogeneities given by $\hat{g}$ and $\hat{S}$ and their first two derivatives. This identity contains quadratic terms in the second fundamental form. Since $II(X,Y) \in T^\perp M$ and the metric on the normal space $T^\perp M$ is negative definite, these quadratic terms have a favorable sign, allowing us to apply the maximum principle. This calculation is inspired in part by identities found for hypersurfaces by Calabi [5], and in the split geometry of the next section by Mealy [26], Harvey and Lawson [14].

Proposition 1 (An elliptic identity for zero mean curvature submanifolds)**.** *Let $\hat{R}$ denote the Riemann curvature tensor and $\hat{D}$ the Levi-Civita connection of a signature (n, m) pseudo-Riemannian manifold $(\hat{M}, \hat{g})$. Let $M \subset \hat{M}$ be an n-dimensional spacelike submanifold whose second fundamental form $II : TM \times TM \to T^\perp M$ has vanishing trace. Let $\hat{S}$ be a symmetric $(0, 2)$-tensor field on $\hat{M}$, and let S denote its restriction to M. Let $\hat{E}_1, \ldots, \hat{E}_{n+m}$ denote a local frame on $\hat{M}$, and $e_1, \ldots, e_n$ denote a local orthonormal frame*

on M. Finally, let X, Y be tangential vector fields on M. Then

$$(\Delta S)(X, Y)$$

$$= \sum_{l=1}^{n} (\hat{D}_{e_l, e_l}^2 \hat{S})(X, Y)$$

$$+ 2 \sum_{l=1}^{n} (\hat{D}_{e_l} \hat{S})(II(e_l, X), Y) + 2 \sum_{l=1}^{n} (\hat{D}_{e_l} \hat{S})(X, II(e_l, Y))$$

$$+ 2 \sum_{l=1}^{n} \hat{S}(II(e_l, X), II(e_l, Y))$$

$$- \sum_{k,l=1}^{n} \hat{g}(II(e_l, X), II(e_l, e_k)) \, S(e_k, Y) - \sum_{k,l=1}^{n} \hat{g}(II(e_l, Y), II(e_l, e_k)) \, S(X, e_k)$$

$$- \sum_{\alpha,\beta=1}^{n+m} \sum_{l=1}^{n} \hat{\sigma}^{\alpha\beta} \, \hat{R}(e_l, X, e_l, \hat{E}_\alpha) \, \hat{S}(\hat{E}_\beta, Y) - \sum_{\alpha,\beta=1}^{n+m} \sum_{l=1}^{n} \hat{\sigma}^{\alpha\beta} \, \hat{R}(e_l, Y, e_l, \hat{E}_\alpha) \, \hat{S}(X, \hat{E}_\beta)$$

$$+ \sum_{k=1}^{n} \sum_{l=1}^{n} \hat{R}(e_l, X, e_l, e_k) \, S(e_k, Y) + \sum_{k=1}^{n} \sum_{l=1}^{n} \hat{R}(e_l, Y, e_l, e_k) \, S(X, e_k)$$

at each point on M, where $\hat{\sigma}^{\alpha\beta}$ denotes the inverse of the matrix $\hat{g}(\hat{E}_\alpha, \hat{E}_\beta)$.

Proof. In the following, X, Y, V, W will denote tangential vector fields on M. By definition,

$$S(X, Y) = \hat{S}(X, Y).$$

Differentiating this identity in direction W gives

$$(D_W S)(X, Y) = (\hat{D}_W \hat{S})(X, Y) + \hat{S}(II(W, X), Y) + \hat{S}(X, II(W, Y)).$$

Differentiating this identity in direction V, we obtain

$$\begin{aligned}
(D_{V,W}^2 S)(X, Y) = {} & (\hat{D}_{V,W}^2 \hat{S})(X, Y) + (\hat{D}_{II(V,W)} \hat{S})(X, Y) \\
& + (\hat{D}_W \hat{S})(II(V, X), Y) + (\hat{D}_W \hat{S})(X, II(V, Y)) \\
& + (\hat{D}_V \hat{S})(II(W, X), Y) + (\hat{D}_V \hat{S})(X, II(W, Y)) \\
& + \hat{S}(II(W, X), II(V, Y)) + \hat{S}(II(V, X), II(W, Y)) \\
& + \hat{S}((\hat{D}_V II)(W, X), Y) + \hat{S}(X, (\hat{D}_V II)(W, Y))
\end{aligned}$$

at each point on M. We decompose the term $(\hat{D}_V II)(W, X)$ into its tangential and normal components:

$$(1) \qquad (\hat{D}_V II)(W, X) = (D_V^\perp II)(W, X) + \sum_{k=1}^{n} \hat{g}((\hat{D}_V II)(W, X), e_k) \, e_k.$$

Using the Codazzi equation, we obtain

$$(2) \qquad (D_V^\perp II)(W, X) = (D_X^\perp II)(V, W) + \sum_{p=1}^{m} \hat{R}(V, X, W, e_p^\perp) \, e_p^\perp$$

(see [27], p. 115). Note that the curvature term on the right hand side in (2) comes with a plus sign, due to the fact that the metric on the normal space is negative definite. Moreover, differentiating the identity $\hat{g}(II(W, X), e_k) = 0$ in direction V gives

$$(3) \qquad \hat{g}((\hat{D}_V II)(W, X), e_k) = -\hat{g}(II(W, X), II(V, e_k)).$$

Substituting (2) and (3) into (1), we obtain

$$(\hat{D}_V II)(W, X) = (D_X^\perp II)(V, W) - \sum_{k=1}^{n} \hat{g}(II(W, X), II(V, e_k)) \, e_k$$

$$(4) \qquad\qquad + \sum_{p=1}^{m} \hat{R}(V, X, W, e_p^\perp) \, e_p^\perp$$

at each point on M. Therefore,

$$
\begin{aligned}
&(D_{V,W}^2 S)(X, Y) \\
&= (\hat{D}_{V,W}^2 \hat{S})(X, Y) + (\hat{D}_{II(V,W)} \hat{S})(X, Y) \\
&+ (\hat{D}_W \hat{S})(II(V, X), Y) + (\hat{D}_W \hat{S})(X, II(V, Y)) \\
&+ (\hat{D}_V \hat{S})(II(W, X), Y) + (\hat{D}_V \hat{S})(X, II(W, Y)) \\
&+ \hat{S}(II(W, X), II(V, Y)) + \hat{S}(II(V, X), II(W, Y)) \\
&+ \hat{S}((D_X^\perp II)(V, W), Y) + \hat{S}(X, (D_Y^\perp II)(V, W)) \\
&- \sum_{k=1}^{n} \hat{g}(II(W, X), II(V, e_k)) S(e_k, Y) - \sum_{k=1}^{n} \hat{g}(II(W, Y), II(V, e_k)) S(X, e_k) \\
&+ \sum_{p=1}^{m} \hat{R}(V, X, W, e_p^\perp) \hat{S}(e_p^\perp, Y) + \sum_{p=1}^{m} \hat{R}(V, Y, W, e_p^\perp) \hat{S}(X, e_p^\perp)
\end{aligned}
$$

at each point on M. Next, we take the trace over V, W and use the fact that the mean curvature vanishes to eliminate (derivatives of) the trace of

II. Thus,

$$(\Delta S)(X, Y)$$

$$= \sum_{l=1}^{n} (\hat{D}_{e_l, e_l}^2 \hat{S})(X, Y)$$

$$+ 2 \sum_{l=1}^{n} (\hat{D}_{e_l} \hat{S})(II(e_l, X), Y) + 2 \sum_{l=1}^{n} (\hat{D}_{e_l} \hat{S})(X, II(e_l, Y))$$

$$+ 2 \sum_{l=1}^{n} \hat{S}(II(e_l, X), II(e_l, Y))$$

$$- \sum_{k,l=1}^{n} \hat{g}(II(e_l, X), II(e_l, e_k)) \, S(e_k, Y) - \sum_{k,l=1}^{n} \hat{g}(II(e_l, Y), II(e_l, e_k)) \, S(X, e_k)$$

$$+ \sum_{p=1}^{m} \sum_{l=1}^{n} \hat{R}(e_l, X, e_l, e_p^{\perp}) \, \hat{S}(e_p^{\perp}, Y) + \sum_{p=1}^{m} \sum_{l=1}^{n} \hat{R}(e_l, Y, e_l, e_p^{\perp}) \, \hat{S}(X, e_p^{\perp})$$

at each point on M. Using the identity

$$\sum_{\alpha, \beta = 1}^{n+m} \hat{\sigma}^{\alpha\beta} \, \hat{E}_\alpha \otimes \hat{E}_\beta = \sum_{k=1}^{n} e_k \otimes e_k - \sum_{p=1}^{m} e_p^{\perp} \otimes e_p^{\perp},$$

the assertion follows. $\qquad\square$

In the remainder of this section, we assume that $\hat{S}$ is positive definite. In other words, $\hat{S}$ is a Riemannian metric on the ambient manifold $\hat{M}$.

Proposition 2. *Under the hypotheses of Proposition 1, assume that $\hat{S}$ is a positive definite symmetric $(0, 2)$-tensor field on $\hat{M}$, and let S denote its restriction to M. Suppose that p_0 is a point on M and $\{e_1, \ldots, e_n\} \subset T_{p_0} M$ is an orthonormal basis with respect to g that diagonalizes S. Then there exists a constant $C = C(\|\hat{g}, \hat{g}^{-1}, \hat{S}\|_{C^2(\{p_0\})})$ such that*

$$(\Delta S)(e_n, e_n) \geq 2 \sum_{l=1}^{n} (\hat{R}(e_l, e_n, e_l, e_n) - C S(e_l, e_l)) \, S(e_n, e_n)$$

at the point p_0.

Proof. Proposition 1 implies

$$(\Delta S)(e_n, e_n)$$
$$= \sum_{l=1}^{n} (\hat{D}^2_{e_l, e_l} \hat{S})(e_n, e_n) + 4 \sum_{l=1}^{n} (\hat{D}_{e_l} \hat{S})(II(e_l, e_n), e_n)$$
$$+ 2 \sum_{l=1}^{n} \hat{S}(II(e_l, e_n), II(e_l, e_n)) - 2 \sum_{l=1}^{n} \hat{g}(II(e_l, e_n), II(e_l, e_n)) S(e_n, e_n)$$
$$- 2 \sum_{\alpha, \beta = 1}^{n+m} \sum_{l=1}^{n} \hat{\sigma}^{\alpha\beta} \, \hat{R}(e_l, e_n, e_l, \hat{E}_\alpha) \, \hat{S}(\hat{E}_\beta, e_n) + 2 \sum_{l=1}^{n} \hat{R}(e_l, e_n, e_l, e_n) S(e_n, e_n)$$

at the point p_0, where $\hat{\sigma}^{\alpha\beta}$ denotes the inverse of the matrix $\hat{g}(\hat{E}_\alpha, \hat{E}_\beta)$. Since the preceding formula does not depend on the choice of the frame $\hat{E}_1, \ldots, \hat{E}_{n+m}$, we may assume that the frame $\hat{E}_1, \ldots, \hat{E}_{n+m}$ is chosen to be orthonormal with respect to the fixed Riemannian metric $\hat{S}$ on ambient space.

Clearly, $II(e_l, e_n)$ is a normal vector for each $l = 1, \ldots, n$. Since the metric on the normal space is negative definite, it follows that

$$\hat{g}(II(e_l, e_n), II(e_l, e_n)) \leq 0$$

for each $l = 1, \ldots, n$. Therefore,

$$-2 \sum_{l=1}^{n} \hat{g}(II(e_l, e_n), II(e_l, e_n)) S(e_n, e_n) \geq 0.$$

We estimate the following multilinear expressions using the fixed Riemannian metric $\hat{S}$ on ambient space (which reduces to S on the tangent space to M):

$$\sum_{l=1}^{n} (\hat{D}^2_{e_l, e_l} \hat{S})(e_n, e_n) \geq -C \sum_{l=1}^{n} S(e_l, e_l) \, S(e_n, e_n),$$

$$-2 \sum_{\alpha, \beta = 1}^{n+m} \sum_{l=1}^{n} \hat{\sigma}^{\alpha\beta} \, \hat{R}(e_l, e_n, e_l, \hat{E}_\alpha) \, \hat{S}(\hat{E}_\beta, e_n) \geq -C \sum_{l=1}^{n} S(e_l, e_l) \, S(e_n, e_n),$$

and

$$4 \sum_{l=1}^{n} (\hat{D}_{e_l} \hat{S})(II(e_l, e_n), e_n)$$
$$\geq -C \sum_{l=1}^{n} S(e_l, e_l)^{\frac{1}{2}} \, S(e_n, e_n)^{\frac{1}{2}} \, \hat{S}(II(e_l, e_n), II(e_l, e_n))^{\frac{1}{2}}$$

at the point p_0. Putting everything together, we obtain

$$
\begin{aligned}
(\Delta S)(e_n, e_n) \geq {} & -C \sum_{l=1}^{n} S(e_l, e_l)\, S(e_n, e_n) \\
& - C \sum_{l=1}^{n} S(e_l, e_l)^{\frac{1}{2}} S(e_n, e_n)^{\frac{1}{2}} \hat{S}(II(e_l, e_n), II(e_l, e_n))^{\frac{1}{2}} \\
& + 2 \sum_{l=1}^{n} \hat{S}(II(e_l, e_n), II(e_l, e_n)) \\
& + 2 \sum_{l=1}^{n} \hat{R}(e_l, e_n, e_l, e_n)\, S(e_n, e_n)
\end{aligned}
$$

at the point p_0. The assertion now follows from Young's inequality. $\square$

Corollary 3. *Under the hypotheses of Proposition 1, assume that $\hat{S}$ is a positive definite symmetric $(0,2)$-tensor field on $\hat{M}$, and let S denote its restriction to M. Let $\hat{\varphi}$ be a nonnegative smooth function on $\hat{M}$, and let φ denote the restriction of $\hat{\varphi}$ to M. Suppose that p_0 is a point on M where the largest eigenvalue of $\varphi^{2n-2}S$ with respect to the metric g attains its maximum. Suppose that $\{e_1, \ldots, e_n\} \subset T_{p_0}M$ is an orthonormal basis with respect to g that diagonalizes S, and $S(e_n, e_n)$ is the largest eigenvalue of S with respect to g at the point p_0. Then*

$$
\sum_{l=1}^{n} \hat{R}(e_l, e_n, e_l, e_n) \leq C\varphi^{-2}\, S(e_n, e_n)
$$

at the point p_0. Here $C = C(\|\hat{g}, \hat{g}^{-1}, \hat{S}, \hat{\varphi}\|_{C^2(\{p_0\})})$.

Proof. Since the largest eigenvalue of $\varphi^{2n-2}S$ with respect to the metric g attains its maximum at p_0, we know that

$$
(5) \qquad (D_{e_l}(\varphi^{2n-2}S))(e_n, e_n) = 0
$$

for $l = 1, \ldots, n$ and

$$
(6) \qquad (\Delta(\varphi^{2n-2}S))(e_n, e_n) \leq 0
$$

at the point p_0. Combining (5) and (6), we obtain

$$
\begin{aligned}
0 \geq {} & \varphi^{-2n+2}\, (\Delta(\varphi^{2n-2}S))(e_n, e_n) \\
& - (4n-4)\, \varphi^{-2n+1} \sum_{l=1}^{n} d\varphi(e_l)\, (D_{e_l}(\varphi^{2n-2}S))(e_n, e_n) \\
(7) \qquad = {} & (\Delta S)(e_n, e_n) + (2n-2)\, \varphi^{-1}\, (\Delta\varphi)\, S(e_n, e_n) \\
& - (2n-2)(2n-1)\, \varphi^{-2} \sum_{l=1}^{n} (d\varphi(e_l))^2\, S(e_n, e_n)
\end{aligned}
$$

at the point p_0. On the other hand, Proposition 2 gives

$$(8) \qquad (\Delta S)(e_n, e_n) \geq 2 \sum_{l=1}^{n} (\hat{R}(e_l, e_n, e_l, e_n) - C S(e_l, e_l)) \, S(e_n, e_n)$$

at the point p_0. Since the mean curvature of M vanishes, we obtain

$$(9) \qquad \Delta \varphi = \sum_{l=1}^{n} (\hat{D}^2 \hat{\varphi})(e_l, e_l) \geq -C \sum_{l=1}^{n} S(e_l, e_l)$$

Moreover,

$$(10) \qquad \sum_{l=1}^{n} (d\varphi(e_l))^2 = \sum_{l=1}^{n} (d\hat{\varphi}(e_l))^2 \leq C \sum_{l=1}^{n} S(e_l, e_l)$$

at the point p_0. Substituting (8), (9), and (10) into (7), we conclude that

$$0 \geq 2 \sum_{l=1}^{n} (\hat{R}(e_l, e_n, e_l, e_n) - C\varphi^{-2} S(e_l, e_l)) \, S(e_n, e_n)$$

at the point p_0. Since $S(e_l, e_l) \leq S(e_n, e_n)$ for $l = 1, \ldots, n$, the assertion follows. $\qquad\square$

3. Application to optimal transport.

Let X and $\bar{X}$ be connected orientable manifolds having the same dimension n, equipped with nowhere vanishing smooth volume forms ρ and $\bar{\rho}$ satisfying $\int_X \rho = 1 = \int_{\bar{X}} \bar{\rho}$. Although we shall often take $\hat{M} = X \times \bar{X}$, it is useful to allow $\hat{M} \subset X \times \bar{X}$ to be any open domain equipped with a bounded smooth cost function $c \in C^\infty(\hat{M})$, and extend c to $(X \times \bar{X}) \setminus \hat{M}$ lower semicontinuously (and boundedly). The optimal transportation problem of Kantorovich is to minimize

$$(11) \qquad \int_{X \times \bar{X}} c \, d\gamma$$

among joint measures $\gamma \geq 0$ on $X \times \bar{X}$ having ρ and $\bar{\rho}$ as their left and right marginals. The basic insight gleaned from linear programming duality is the existence of a pair of functions $u \in L^1(X, \rho)$ and $\bar{u} \in L^1(\bar{X}, \bar{\rho})$ satisfying

$$(12) \qquad u(x) + \bar{u}(\bar{x}) + c(x, \bar{x}) \geq 0 \quad \text{for all} \quad (x, \bar{x}) \in X \times \bar{X}$$

such that γ minimizes (11) if and only if it vanishes outside the zero set of (12), e.g. Theorem 5.10 of [37].

Monge's version of the same problem is to find a Borel map $F : X \to \bar{X}$ which pulls $\bar{\rho}$ back to $\pm\rho$ (depending on the sign of $\det(-\frac{\partial^2 c}{\partial x^i \partial \bar{x}^j})$), such that $\gamma = (id \times F)_\# \rho$ attains the minimum (11). We assume that the matrix $\frac{\partial^2 c}{\partial x^i \partial \bar{x}^k} \, dx^i \otimes d\bar{x}^k$ is invertible at each point of $\hat{M}$, which is hypothesis (A2) of Ma, Trudinger and Wang [24].

We define a function $\chi : \hat{M} \to (0, \infty)$ so that

(13)

$$\chi(x, \bar{x})^n \det \left(\frac{\partial^2 c}{\partial x^i \partial \bar{x}^k}(x, \bar{x}) \right) dx_1 \wedge \ldots dx_n \wedge d\bar{x}_1 \wedge \ldots \wedge d\bar{x}_n = \pm \rho(x) \wedge \bar{\rho}(\bar{x}).$$

Note that χ is independent of the choice of coordinates. With this understood,

(14)
$$\hat{g} = -\chi \sum_{i,k=1}^{n} \frac{\partial^2 c}{\partial x^i \partial \bar{x}^k} \left(dx^i \otimes d\bar{x}^k + d\bar{x}^k \otimes dx^i \right)$$

becomes the Kim–McCann–Warren [18] pseudo-metric on $\hat{M} \subset X \times \bar{X}$. The induced volume form of $\hat{g}$ is given by $\pm\rho \wedge \bar{\rho}$. In its original (strong) form, the Ma–Trudinger–Wang condition (A3) is locally equivalent to the assertion that

(15)
$$\hat{R}(\xi \oplus \bar{0}, 0 \oplus \bar{\xi}, \xi \oplus \bar{0}, 0 \oplus \bar{\xi}) > 0$$

for all points $(x, \bar{x}) \in \hat{M}$ and all tangent vectors $\xi \in T_x X$ and $\bar{\xi} \in T_{\bar{x}} \bar{X}$ satisfying $\hat{g}(\xi \oplus \bar{0}, 0 \oplus \bar{\xi}) = 0$. Here, $0 \in T_x X$ and $\bar{0} \in T_{\bar{x}} \bar{X}$ denote zero vectors, by Remark 4.2 of [18].

We next define a Riemannian metric $\hat{S}$ on $\hat{M}$. It is convenient to choose the Riemannian metric $\hat{S}$ in a particular way. To explain this, let us fix a Riemannian metric h on X and let vol_h denote its associated volume form. We define a symmetric $(0, 2)$-tensor field $\bar{h}$ on $\hat{M}$ by

(16)
$$\bar{h} = \chi^2 \sum_{k,l,p,q=1}^{n} h^{pq} \frac{\partial^2 c}{\partial x^p \partial \bar{x}^k} \frac{\partial^2 c}{\partial x^q \partial \bar{x}^l} d\bar{x}^k \otimes d\bar{x}^l.$$

For each point $x \in X$, $\bar{h}$ defines a Riemannian metric on $\bar{X}$; denote its volume form by $\text{vol}_{\bar{h}}$. In view of (13), the volume forms of h and $\bar{h}$ are related by $\text{vol}_h \wedge \text{vol}_{\bar{h}} = \pm\rho \wedge \bar{\rho}$. With this understood, we define a Riemannian metric $\hat{S}$ on $\hat{M}$ by

(17)
$$\hat{S} = h + \bar{h}.$$

Note that $\hat{S}$ is not a product metric, as $\bar{h}$ depends on $x \in X$. The induced volume form of $\hat{S}$, like $\hat{g}$, is given by $\pm\rho \wedge \bar{\rho}$.

Suppose that a diffeomorphism $F : X \to \bar{X}$ solves Monge's transport problem, so that the minimizer γ of (11) vanishes outside $\text{Graph}(M) := \{(x, F(x)) : x \in X\}$. Denote by g and S the restrictions of $\hat{g}$ and $\hat{S}$ to $M := \hat{M} \cap \text{Graph}(F)$. Kim, McCann and Warren [18] show that M is locally volume maximizing with respect to the metric $\hat{g}$, hence has zero mean curvature with respect to $\hat{g}$.

Theorem 4 (Apriori local estimates for optimal diffeomorphisms)**.** *Let $\hat{M} \subset X \times \bar{X}$ be an open domain. Suppose that $0 \leq \hat{\varphi} \in C^\infty(\hat{M})$ has compact*

support $\operatorname{spt}\hat{\varphi}\subset\hat{M}$. *Assume the Ma–Trudinger–Wang conditions* (A2)-(A3) *hold on* $\hat{M}$. *Let* κ *be a positive constant with the property that*

$$(18) \qquad \hat{R}(\xi\oplus\bar{0},0\oplus\bar{\xi},\xi\oplus\bar{0},0\oplus\bar{\xi}) \geq \kappa\,h(\xi,\xi)\,\bar{h}(\bar{\xi},\bar{\xi})$$

for all points $(x,\bar{x})\in\operatorname{spt}\hat{\varphi}$ *and all tangent vectors* $\xi\in T_xX$ *and* $\bar{\xi}\in T_{\bar{x}}\bar{X}$ *satisfying* $\hat{g}(\xi\oplus\bar{0},0\oplus\bar{\xi})=0$. *Let* $F:X\to\bar{X}$ *be a diffeomorphism that minimizes Monge's cost* c *between* ρ *and* $\bar{\rho}$. *Then there is a constant* C, *depending on* $\|\hat{g},\hat{g}^{-1},\hat{S},\hat{\varphi}\|_{C^2(\operatorname{spt}\hat{\varphi})}$ *and* $\|\log(\frac{\rho}{\operatorname{vol}_h})\|_{C^0(\operatorname{spt}\hat{\varphi})}$, *such that*

$$\kappa^{n-1}\varphi^{2n-2}S \leq Cg$$

on M. *Here,* φ *denotes the restriction of* $\hat{\varphi}$ *to* $M=\hat{M}\cap\operatorname{Graph}(F)$, *and* g *and* S *denote the restrictions of* $\hat{g}$ *and* $\hat{S}$ *from* (14) *and* (17) *to* M.

Proof. We consider a point $p_0=(x_0,F(x_0))\in M$ where the largest eigenvalue of $\varphi^{2n-2}S$ attains its maximum. Such a point exists by the smoothness of the objects in question and the compact support of $\hat{\varphi}$. Let $(x^1,\ldots,x^n)$ be a local coordinate system on X such that $h_{ij}=\delta_{ij}$ at x_0 and let $(\bar{x}^1,\ldots,\bar{x}^n)$ denote a local coordinate system on $\bar{X}$ such that $\bar{h}_{kl}=\delta_{kl}$ at $(x_0,F(x_0))$.

Recall there exist $u\in L^1(X,\rho)$ and $\bar{u}\in L^1(\bar{X},\bar{\rho})$ satisfying the inequality (12) with equality on M:

$$(19) \qquad u(x)+\bar{u}(F(x))+c(x,F(x))=0.$$

From smoothness of c at $(x,F(x))$ it follows that u is continuous and semiconvex in a neighbourhood of x, as in [25]. The first and second-order conditions for (12),(19) assert

$$(20) \qquad \frac{\partial u}{\partial x^i}+\frac{\partial c}{\partial x^i}\Big|_{\bar{x}=F(x)}=0 \qquad i=1,\ldots,n,$$

and symmetric nonnegative definiteness of $(B_{ij})_{1\leq i,j\leq n}$ given by

$$(21) \qquad B_{ij}:=\frac{\partial^2 u}{\partial x^i\partial x^j}+\frac{\partial^2 c}{\partial x^i\partial x^j}\Big|_{\bar{x}=F(x)}.$$

A priori these conditions hold almost everywhere, but smoothness of u then follows from the hypothesized smoothness of F and (20), whence both conditions actually hold everywhere. Differentiating (20) yields

$$(22) \qquad B_{ij}=-\sum_{k=1}^{n}\frac{\partial^2 c}{\partial x^i\partial\bar{x}^k}\,\frac{\partial F^k}{\partial x^j}.$$

Since $F^{\#}\bar{\rho}=\pm\rho$ by hypothesis, we find

$$\det(B)\,dx^1\wedge\cdots\wedge dx^n\wedge\bar{\rho}\Big|_{\bar{x}=F(x)}=\pm\det\left(\frac{\partial^2 c}{\partial x^i\partial\bar{x}^k}\right)\Big|_{\bar{x}=F(x)}\rho\wedge d\bar{x}^1\wedge\cdots\wedge d\bar{x}^n$$

As (A2) ensures the last term is non-vanishing on $\hat{M}$, nonnegative definiteness of the symmetric matrix B improves to positive definiteness on M.

In the coordinates we have chosen, $\mathrm{vol}_h = \pm dx^1 \wedge \cdots \wedge dx^n$ and $\mathrm{vol}_{\bar{h}} = \pm d\bar{x}^1 \wedge \cdots d\bar{x}^n$ at the point $(x_0, F(x_0))$. Therefore,

$$\det(B)\,\frac{\bar{\rho}}{\mathrm{vol}_{\bar{h}}} = \pm \det\left(\frac{\partial^2 c}{\partial x^i \partial \bar{x}^k}\right)\frac{\rho}{\mathrm{vol}_h}$$

at the point $(x_0, F(x_0))$. Moreover, (13) implies $\chi^n \det(\frac{\partial^2 c}{\partial x^i \partial \bar{x}^k}) = \pm\frac{\rho}{\mathrm{vol}_h}\frac{\bar{\rho}}{\mathrm{vol}_{\bar{h}}}$ at the point $(x_0, F(x_0))$. This gives

$$\chi^n \det(B) = \left(\frac{\rho}{\mathrm{vol}_h}\right)^2$$

at the point x_0. Hence, if we put $A := \chi B$, then

$$(23) \qquad \det(A) = \left(\frac{\rho}{\mathrm{vol}_h}\right)^2$$

at the point x_0. In the next step, we relate the eigenvalues of A with respect to h to the eigenvalues of S with respect to g. By a suitable choice of the coordinates $(x^1, \ldots, x^n)$, we can arrange that $A_{ij} = \lambda_i\,\delta_{ij}$ as well as $h_{ij} = \delta_{ij}$ at x_0. We define

$$(24) \qquad \xi_i = \frac{\partial}{\partial x_i} \in T_{x_0}X, \qquad \bar{\xi}_i = \lambda_i^{-1}\sum_{k=1}^{n}\frac{\partial F^k}{\partial x^i}\frac{\partial}{\partial \bar{x}^k} \in T_{F(x_0)}\bar{X}$$

for $i = 1, \ldots, n$. Then

$$(25) \qquad \hat{g}(\xi_i \oplus \bar{0}, 0 \oplus \bar{\xi}_j) = -\lambda_j^{-1}\chi\sum_{k=1}^{n}\frac{\partial^2 c}{\partial x^i \partial \bar{x}^k}\frac{\partial F^k}{\partial x^j} = \lambda_j^{-1}A_{ij} = \delta_{ij}$$

for $i, j = 1, \ldots, n$. Moreover, $h(\xi_i, \xi_j) = \delta_{ij}$ and

$$\bar{h}(\bar{\xi}_i, \bar{\xi}_j) = \lambda_i^{-1}\lambda_j^{-1}\chi^2\sum_{k,l,p,q=1}^{n} h^{pq}\frac{\partial^2 c}{\partial x^p \partial \bar{x}^k}\frac{\partial^2 c}{\partial x^q \partial \bar{x}^l}\frac{\partial F^k}{\partial x^i}\frac{\partial F^l}{\partial x^j}$$

$$(26) \qquad\qquad = \lambda_i^{-1}\lambda_j^{-1}\sum_{p,q=1}^{n} h^{pq}A_{pi}A_{qj}$$

$$= \delta_{ij}$$

for $i, j = 1, \ldots, n$. We next define

$$e_i = (2\lambda_i)^{-\frac{1}{2}}(\xi_i \oplus \lambda_i\,\bar{\xi}_i) \in T_{(x_0, F(x_0))}M$$

and

$$\mu_i = \frac{1}{2}(\lambda_i + \lambda_i^{-1})$$

for $i = 1, \ldots, n$. Then

$$\hat{g}(e_i, e_j) = (2\lambda_i)^{-\frac{1}{2}}(2\lambda_j)^{-\frac{1}{2}}(\lambda_j\,\hat{g}(\xi_i \oplus \bar{0}, 0 \oplus \bar{\xi}_j) + \lambda_i\,\hat{g}(0 \oplus \bar{\xi}_i, \xi_j \oplus \bar{0})) = \delta_{ij}$$

and

$$S(e_i, e_j) = (2\lambda_i)^{-\frac{1}{2}}(2\lambda_j)^{-\frac{1}{2}}(h(\xi_i, \xi_j) + \lambda_i\lambda_j\,\bar{h}(\bar{\xi}_i, \bar{\xi}_j)) = \mu_i\,\delta_{ij}$$

for $i, j = 1, \ldots, n$. Thus, $\{e_1, \ldots, e_n\} \subset T_{(x_0, F(x_0))}M$ is an orthonormal basis with respect to g which diagonalizes S. Moreover, the eigenvalues of S with respect to g at the point p_0 are given by $\mu_1, \ldots, \mu_n$. Without loss of generality, we may assume that $\mu_i \leq \mu_n$ for $i = 1, \ldots, n-1$, so that μ_n is the largest eigenvalue of S with respect to g at the point p_0.

Note that $\hat{g}(\xi_i \oplus \bar{0}, 0 \oplus \bar{\xi}_j) = \hat{g}(\xi_n \oplus \bar{0}, 0 \oplus \bar{\xi}_j) = 0$ for $i = 1, \ldots, n-1$. Moreover, it follows from (24) and (26) that $h(\xi_i, \xi_i) = h(\xi_n, \xi_n) = \bar{h}(\bar{\xi}_i, \bar{\xi}_i) = \bar{h}(\bar{\xi}_n, \bar{\xi}_n) = 1$ for $i = 1, \ldots, n-1$. Therefore, the uniform Ma–Trudinger–Wang condition (18) yields

$$\hat{R}(\xi_i \oplus \bar{0}, 0 \oplus \bar{\xi}_n, \xi_i \oplus \bar{0}, 0 \oplus \bar{\xi}_n) \geq \kappa\, h(\xi_i, \xi_i)\, \bar{h}(\bar{\xi}_n, \bar{\xi}_n) = \kappa$$

and

$$\hat{R}(\xi_n \oplus \bar{0}, 0 \oplus \bar{\xi}_i, \xi_n \oplus \bar{0}, 0 \oplus \bar{\xi}_i) \geq \kappa\, h(\xi_n, \xi_n)\, \bar{h}(\bar{\xi}_i, \bar{\xi}_i) = \kappa$$

for $i = 1, \ldots, n-1$. Using the special structure of the curvature tensor $\hat{R}$ described in Remark 4.2 of [18] and Lemma 4.1 of [16], we obtain

$$\hat{R}(\xi_i \oplus \bar{0}, \xi_n \oplus \bar{0}, \xi_i \oplus \bar{0}, \xi_n \oplus \bar{0}) = 0$$

and

$$\hat{R}(0 \oplus \bar{\xi}_i, 0 \oplus \bar{\xi}_n, 0 \oplus \bar{\xi}_i, 0 \oplus \bar{\xi}_n) = 0.$$

This implies

$$\hat{R}(\xi_i \oplus \lambda_i \bar{\xi}_i, \xi_n \oplus \lambda_n \bar{\xi}_n, \xi_i \oplus \lambda_i \bar{\xi}_i, \xi_n \oplus \lambda_n \bar{\xi}_n)$$
$$\geq \hat{R}(\xi_i \oplus \bar{0}, \xi_n \oplus \bar{0}, \xi_i \oplus \bar{0}, \xi_n \oplus \bar{0}) + \lambda_n^2\, \hat{R}(\xi_i \oplus \bar{0}, 0 \oplus \bar{\xi}_n, \xi_i \oplus \bar{0}, 0 \oplus \bar{\xi}_n)$$
$$+ \lambda_i^2\, \hat{R}(0 \oplus \bar{\xi}_i, \xi_n \oplus \bar{0}, 0 \oplus \bar{\xi}_i, \xi_n \oplus \bar{0}) + \lambda_i^2 \lambda_n^2\, \hat{R}(0 \oplus \bar{\xi}_i, 0 \oplus \bar{\xi}_n, 0 \oplus \bar{\xi}_i, 0 \oplus \bar{\xi}_n)$$
$$- C\lambda_n - C\lambda_i - C\lambda_i \lambda_n - C\lambda_i \lambda_n^2 - C\lambda_i^2 \lambda_n$$
$$\geq \kappa\,(\lambda_n^2 + \lambda_i^2) - C\lambda_n - C\lambda_i - C\lambda_i \lambda_n - C\lambda_i \lambda_n^2 - C\lambda_i^2 \lambda_n$$

for $i = 1, \ldots, n-1$. This implies

$$\hat{R}(e_i, e_n, e_i, e_n) \geq \frac{\kappa}{4}\,(\lambda_i^{-1}\lambda_n + \lambda_i \lambda_n^{-1}) - C\,(\lambda_n + \lambda_n^{-1}) - C\,(\lambda_i + \lambda_i^{-1}) - C$$

for $i = 1, \ldots, n-1$. In the next step, we sum over $i = 1, \ldots, n-1$. Since $1 \leq \mu_i \leq \mu_n$ for $i = 1, \ldots, n-1$, it follows that

$$(27) \qquad \sum_{i=1}^{n-1} \hat{R}(e_i, e_n, e_i, e_n) \geq \frac{\kappa}{4} \sum_{i=1}^{n-1} (\lambda_i^{-1}\lambda_n + \lambda_i \lambda_n^{-1}) - C\mu_n.$$

Using (23), we obtain $\prod_{i=1}^{n} \lambda_i = \det A = \left(\frac{\rho}{\mathrm{vol}_h}\right)^2$ at $(x_0, F(x_0))$. Using the arithmetic-geometric mean inequality, we obtain

$$(28) \qquad \sum_{i=1}^{n-1} \lambda_i^{-1} \geq (n-1) \left(\prod_{i=1}^{n-1} \lambda_i^{-1}\right)^{\frac{1}{n-1}} \geq \frac{1}{C}\, \lambda_n^{\frac{1}{n-1}}$$

and

$$(29) \qquad \sum_{i=1}^{n-1} \lambda_i \geq (n-1) \left(\prod_{i=1}^{n-1} \lambda_i \right)^{\frac{1}{n-1}} \geq \frac{1}{C} \lambda_n^{-\frac{1}{n-1}}.$$

Substituting (28) and (29) into (27) gives

$$(30) \qquad \sum_{i=1}^{n-1} \hat{R}(e_i, e_n, e_i, e_n) \geq \frac{\kappa}{C} \mu_n^{\frac{n}{n-1}} - C\mu_n.$$

On the other hand, Corollary 3 implies

$$(31) \qquad \sum_{i=1}^{n-1} \hat{R}(e_i, e_n, e_i, e_n) \leq C\varphi^{-2}\mu_n.$$

Combining (30) and (31), we conclude that $\kappa\mu_n^{\frac{n}{n-1}} \leq C\varphi^{-2}\mu_n + C\mu_n$. Since $\varphi \leq C$, it follows that $\kappa^{n-1}\varphi^{2n-2}\mu_n \leq C$. This shows that the largest eigenvalue of $\kappa^{n-1}\varphi^{2n-2}S$ is uniformly bounded from above at the point $(x_0, F(x_0))$, and hence on all of spt $\hat{\varphi}$. $\qquad\square$

Remark 5 (Quantitatively spacelike, slope, recovery of C^1 bounds). We may view the apriori lower bound $\kappa^{n-1}\varphi^{2n-2}S \leq Cg$ given by Theorem 4 as quantifying the spacelikeness of Graph(F). Since $\hat{g}$ vanishes on the horizontal fibres $X \times \{\bar{x}\}$ and vertical fibres $\{x\} \times \bar{X}$, a fortiori this quantifies the transversality with which F intersects these fibres, thus translating to a C^1 bound for F^{-1} and for F. More explicitly, the estimate in Theorem 4 is equivalent to the inequality

$$(32) \qquad \lambda_i + \lambda_i^{-1} \leq C\kappa^{1-n}\varphi^{2-2n}$$

for $i = 1, \dots, n$, where $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of $A = \chi B$ with respect to the metric h. The inequality (32) implies two separate bounds for λ_i and for λ_i^{-1}. The bound for λ_i quantifies the transversality with which the graph of F intersects the vertical fibers $\{x\} \times \bar{X}$; it corresponds to an interior C^1-estimate for F. The bound for λ_i^{-1} quantifies the transversality with which the graph of F intersects the horizontal fibers $X \times \{\bar{x}\}$; it corresponds to an interior C^1-estimate for F^{-1}. In the setting of Ma–Trudinger–Wang, we may take $X, \overline{X}$ as open subsets of $\mathbf{R}^n$ and h as the usual Euclidean metric; the definition (21) of B then implies the Ma–Trudinger–Wang estimates as they appear in [24].

Remark 6 (Optimal maps and dual potentials). The potentials $u \in L^1(X, \rho)$ and $\bar{u} \in L^1(\bar{X}, \bar{\rho})$ from (12) which certify optimality of γ in (11) can be interpreted as Lagrange multipliers for the marginal constraints ρ and $\bar{\rho}$. When $\hat{M} = X \times \bar{X}$, then u is well-known to become continuous and semiconvex if the second x derivatives of c are bounded locally on X uniformly with respect to $\bar{x} \in \bar{X}$; similarly $\bar{u}$ is semiconvex if the corresponding condition holds under the interchange $x \leftrightarrow \bar{x}$, e.g. [25]. A sufficient condition for existence of a Borel solution $F : X \to \bar{X}$ to Monge's problem is then that

$x \in X \mapsto c(x, \bar{x}) - c(x, \bar{x}')$ be free from critical points for all $\bar{x}' \neq \bar{x} \in \bar{X}$. Under hypothesis (A1) of [24], which asserts that both c and $\tilde{c}(\bar{x}, x) = c(x, \bar{x})$ satisfy this restriction, the solution F is injective outside a ρ-negligible set. If, in addition, F happens to be continuously differentiable, then it follows from non-degeneracy (A2) that F is a diffeomorphism, arguing as in the second paragraph of the proof of Theorem 4. Higher regularity can be bootstrapped by applying Evans [11] or Krylov [19] and then standard elliptic theory [13].

Remark 7 (Regularity of weak solutions). Diffeomorphisms $F : X \to \bar{X}$ solving Monge's problem may or may not exist [1, 22], depending on the choice of volume forms $\rho, \bar{\rho}$ and the cost c. Under an additional hypothesis — equivalent to geodesic convexity of the vertical fibers $\{x\} \times \bar{X}$ with respect to the Kim-McCann metric [16], or equivalently the Kim-McCann-Warren metric $\hat{g}$ — Ma, Trudinger and Wang construct a diffeomorphism by starting from a c-convex potential u which solves the dual problem and showing it uniquely solves the associated Monge–Ampère type equation in the sense of Alexandrov. To establish interior regularity of u, Ma, Trudinger, and Wang use the continuity method to construct a smooth solution of the Monge–Ampère type equation which is defined on a small ball and agrees with u on the boundary of the ball a priori, and a posteriori on its interior. This involves deforming the boundary data given by u to a new set of boundary data which is smooth and c-convex; this argument was extended to generated Jacobian equations in [34]. Independently of the regularization parameter, the C^2-estimate ensures the solution of the Monge–Ampère type equation is uniformly elliptic, hence c-convex [32, 16], after which higher regularity follows from Evans–Krylov and standard elliptic theory as in Remark 6.

Remark 8 (Boundary behaviour, global regularity, compact manifolds). Theorem 4 only addresses interior regularity. Studying the boundary regularity requires additional techniques, as in e.g. [2, 3, 36, 33, 7].

If X and $\bar{X}$ are compact manifolds without boundary, taking $\hat{M} = X \times \bar{X}$ is inconsistent with global non-degeneracy (A2) of $\hat{g}$. However, in some situations it is possible to identify an open subset $\hat{M} \subset X \times \bar{X}$, where the conditions (A2)-(A3) are satisfied. If we know that $\mathrm{Graph}(F)$ is contained in such a subset $\hat{M}$, then Theorem 4 yields a global estimate, extending [41] to dimensions $n > 2$.

As an example, suppose that we take X and $\bar{X}$ to be the round sphere $\mathbf{S}^n$ and define the cost function c by $c(x, \bar{x}) = d^2(x, \bar{x})/2$. In this case, we may take $\hat{M}$ to be the set where c is smooth (i.e. the complement of the cut locus). If the densities ρ and $\bar{\rho}$ are bounded from above and below, it is known [10] that $\mathrm{Graph}(F)$ is disjoint from the cut locus. Regularity then follows [23]. This line of reasoning can be generalized to perturbations [9], submersions [17] or products [12] of round spheres.

References

[1] Luis A. Caffarelli. The regularity of mappings with a convex potential. *J. Amer. Math. Soc.*, 5(1):99–104, 1992.

[2] Luis A. Caffarelli. Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.*, 45(9):1141–1151, 1992.

[3] Luis A. Caffarelli. Boundary regularity of maps with convex potentials. II. *Ann. of Math. (2)*, 144(3):453–496, 1996.

[4] Eugenio Calabi. Improper affine hyperspheres of convex type and a generalization of a theorem by K. Jörgens. *Michigan Math. J.*, 5:105–126, 1958.

[5] Eugenio Calabi. Examples of Bernstein problems for some nonlinear equations. In *Global Analysis (Proc. Sympos. Pure Math., Vols. XIV, XV, XVI, Berkeley, Calif., 1968)*, Proc. Sympos. Pure Math., XIV-XVI, pages 223–230. Amer. Math. Soc., Providence, RI, 1970.

[6] Shibing Chen, Jiakun Liu, and Xu-Jia Wang. Global regularity of optimal mappings in non-convex domains. *Sci. China Math.*, 62(11):2057–2072, 2019.

[7] Shibing Chen, Jiakun Liu, and Xu-Jia Wang. Global regularity for the Monge-Ampère equation with natural boundary condition. *Ann. of Math. (2)*, 194(3):745–793, 2021.

[8] P. Delanoë. Classical solvability in dimension two of the second boundary-value problem associated with the Monge-Ampère operator. *Ann. Inst. H. Poincaré C Anal. Non Linéaire*, 8(5):443–457, 1991.

[9] Philippe Delanoë and Yuxin Ge. Regularity of optimal transport on compact, locally nearly spherical, manifolds. *J. Reine Angew. Math.*, 646:65–115, 2010.

[10] Philippe Delanoë and Grégoire Loeper. Gradient estimates for potentials of invertible gradient-mappings on the sphere. *Calc. Var. Partial Differential Equations*, 26(3):297–311, 2006.

[11] Lawrence C. Evans. Classical solutions of fully nonlinear, convex, second-order elliptic equations. *Comm. Pure Appl. Math.*, 35(3):333–363, 1982.

[12] Alessio Figalli, Young-Heon Kim, and Robert J. McCann. Regularity of optimal transport maps on multiple products of spheres. *J. Eur. Math. Soc. (JEMS)*, 15(4):1131–1166, 2013.

[13] David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[14] F. Reese Harvey and H. Blaine Lawson, Jr. Split special Lagrangian geometry. In *Metric and differential geometry*, volume 297 of *Progr. Math.*, pages 43–89. Birkhäuser/Springer, Basel, 2012.

[15] Yash Jhaveri. On the (in)stability of the identity map in optimal transportation. *Calc. Var. Partial Differential Equations*, 58(3):Paper No. 96, 25, 2019.

[16] Young-Heon Kim and Robert J. McCann. Continuity, curvature, and the general covariance of optimal transportation. *J. Eur. Math. Soc. (JEMS)*, 12(4):1009–1040, 2010.

[17] Young-Heon Kim and Robert J. McCann. Towards the smoothness of optimal maps on Riemannian submersions and Riemannian products (of round spheres in particular). *J. Reine Angew. Math.*, 664:1–27, 2012.

[18] Young-Heon Kim, Robert J. McCann, and Micah Warren. Pseudo-Riemannian geometry calibrates optimal transportation. *Math. Res. Lett.*, 17(6):1183–1197, 2010.

[19] N. V. Krylov. Boundedly inhomogeneous elliptic and parabolic equations. *Izv. Akad. Nauk SSSR Ser. Mat.*, 46(3):487–523, 670, 1982. English translation in *Math. USSR Izv.* 20 (1983) 459–492.

[20] Guanghan Li and Isabel M. C. Salavessa. Mean curvature flow of spacelike graphs. *Math. Z.*, 269(3-4):697–719, 2011.

[21] J. Liu, N.S. Trudinger, X.-J. Wang. Interior $C^{2,\alpha}$ regularity for potential functions in optimal transportation. *Comm. Partial Differential Equations*, 35:165–184, 2010.

[22] Grégoire Loeper. On the regularity of solutions of optimal transportation problems. *Acta Math.*, 202(2):241–283, 2009.

[23] Grégoire Loeper. Regularity of optimal maps on the sphere: the quadratic cost and the reflector antenna. *Arch. Ration. Mech. Anal.*, 199(1):269–289, 2011.

[24] Xi-Nan Ma, Neil S. Trudinger, and Xu-Jia Wang. Regularity of potential functions of the optimal transportation problem. *Arch. Ration. Mech. Anal.*, 177(2):151–183, 2005.

[25] Robert J. McCann and Nestor Guillen. Five lectures on optimal transportation: geometry, regularity and applications. In *Analysis and geometry of metric measure spaces*, volume 56 of *CRM Proc. Lecture Notes*, pages 145–180. Amer. Math. Soc., Providence, RI, 2013.

[26] Jack Mealy. Volume maximization in semi-Riemannian manifolds. *Indiana Univ. Math. J.*, 40(3):793–814, 1991.

[27] B. O'Neill, *Semi-Riemannian Geometry*. Academic Press, New York, 1983.

[28] A. V. Pogorelov. *Monge-Ampère equations of elliptic type*. P. Noordhoff Ltd., Groningen, russian edition, 1964.

[29] A. V. Pogorelov. The regularity of the generalized solutions of the equation $\det(\partial^2 u/\partial x^i \partial x^j) = \varphi(x^1, x^2, \ldots, x^n) > 0$. *Dokl. Akad. Nauk SSSR*, 200:534–537, 1971.

[30] Neil S. Trudinger and John I. E. Urbas. On second derivative estimates for equations of Monge-Ampère type. *Bull. Austral. Math. Soc.*, 30(3):321–334, 1984.

[31] Neil S. Trudinger and Xu-Jia Wang. The Monge-Ampère equation and its geometric applications. In *Handbook of geometric analysis. No. 1*, volume 7 of *Adv. Lect. Math. (ALM)*, pages 467–524. Int. Press, Somerville, MA, 2008.

[32] Neil S. Trudinger and Xu-Jia Wang. On strict convexity and continuous differentiability of potential functions in optimal transportation. *Arch. Ration. Mech. Anal.*, 192(3):403–418, 2009.

[33] Neil S. Trudinger and Xu-Jia Wang. On the second boundary value problem for Monge-Ampère type equations and optimal transportation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 8(1):143–174, 2009.

[34] Neil S. Trudinger. On the local theory of prescribed Jacobian equations. *Discrete Contin. Dyn. Syst. A*, 34(4):1663–1681, 2014.

[35] Mao-Pei Tsui and Mu-Tao Wang. Mean curvature flows and isotopy of maps between spheres. *Comm. Pure Appl. Math.*, 57(8):1110–1126, 2004.

[36] John Urbas. On the second boundary value problem for equations of Monge-Ampère type. *J. Reine Angew. Math.*, 487:115–124, 1997.

[37] Cédric Villani. *Optimal transport, Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

[38] Mu-Tao Wang. Long-time existence and convergence of graphic mean curvature flow in arbitrary codimension. *Invent. Math.*, 148(3):525–543, 2002.

[39] Xu-Jia Wang. On the design of a reflector antenna. *Inverse Problems*, 12(3):351–375, 1996.

[40] Xu-Jia Wang. On the design of a reflector antenna. II. *Calc. Var. Partial Differential Equations*, 20(3):329–341, 2004.

[41] Micah Warren. Minimal Lagrangian submanifolds of weighted Kim-McCann metrics. *Preprint at https://pages.uoregon.edu/micahw/papers/KM.pdf.*

18 SIMON BRENDLE, FLAVIEN LÉGER, ROBERT J. MCCANN, AND CALE RANKIN

Columbia University, 2290 Broadway, New York NY 10027, USA
Email address: simon.brendle@columbia.edu

INRIA Paris, France
Email address: flavien.leger@inria.fr

Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 2E4
Email address: mccann@math.toronto.edu

Fields Institute for the Mathematical Sciences and Department of Mathematics, University of Toronto, Toronto, Ontario, Canada M5S 2E4
Email address: cale.rankin@utoronto.ca