Optimal transportation between unequal dimensions*

Robert J McCann† and Brendan Pass‡

Abstract

We establish that solving an optimal transportation problem in which the source and target densities are defined on spaces with different dimensions, is equivalent to solving a new nonlocal analog of the Monge-Ampère equation, introduced here for the first time. Under suitable topological conditions, we also establish that solutions are smooth if and only if a local variant of the same equation admits a smooth and uniformly elliptic solution. We show that this local equation is elliptic, and $C^{2,\alpha}$ solutions can therefore be bootstrapped to obtain higher regularity results, assuming smoothness of the corresponding differential operator, which we prove under simplifying assumptions. For one-dimensional targets, our sufficient criteria for regularity of solutions to the resulting ODE are considerably less restrictive than those required by earlier works.

*The authors are grateful to Toronto's Fields' Institute for the Mathematical Sciences for its kind hospitality during part of this work, and to two anonymous referees, for their careful reading and crucial comments. RJM acknowledges partial support of this research by Natural Sciences and Engineering Research Council of Canada Grant 217006-15, by a Simons Foundation Fellowship, and by the US National Science Foundation under Grant No. DMS-14401140 while in residence at the Mathematical Sciences Research Institute in Berkeley CA during January and February of 2016. He thanks S.-Y. Alice Chang for a stimulating conversation. BP is pleased to acknowledge support from Natural Sciences and Engineering Research Council of Canada Grants 412779-2012 and 04658-2018, as well as a University of Alberta start-up grant. He is also grateful to the Pacific Institute for the Mathematical Sciences, in Vancouver, BC, Canada, for its generous hospitality during his visit in February and March of 2017.

†Department of Mathematics, University of Toronto, Toronto, Ontario, Canada mccann@math.toronto.edu

‡Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada pass@ualberta.ca.

1 Introduction

Since the 1980s [13] [26] [36] and the celebrated work of Brenier [2] [3], it has been well-understood [32] that for the quadratic cost $c(x,y) = \frac{1}{2}|x-y|^2$ on $\mathbf{R}^n$, solving the Monge-Kantorovich optimal transportation problem is equivalent to solving a degenerate elliptic Monge-Ampère equation: that is, given two probability densities f and g on $\mathbf{R}^n$, the unique optimal map between them, $F = D\tilde{u}$, is given by a convex solution $\tilde{u}$ to the boundary value problem

$$g \circ D\tilde{u} \det D^2\tilde{u} = f \quad \text{[a.e.]}, \tag{1}$$

$$D\tilde{u} \in \operatorname{spt} g \quad \text{[a.e.]}, \tag{2}$$

where $\operatorname{spt} g \subset \mathbf{R}^n$ is the smallest closed set of full mass for g. Similarly, its inverse is given by the gradient of the convex solution $\tilde{v}$ to the boundary value problem

$$f \circ D\tilde{v} \det D^2\tilde{v} = g \quad \text{[a.e.]}, \tag{3}$$

$$D\tilde{v} \in \operatorname{spt} f \quad \text{[a.e.]}. \tag{4}$$

Notice the quadratic cost implicitly requires x and y to live in the same space. Subsequent work of Ma, Trudinger and Wang [31] leads to an analogous result for other cost functions $c(x,y) = -s(x,y)$ satisfying suitable conditions, still requiring x and y to live in spaces of the same dimension n; see also earlier works such as [5] [21] [28] [33] [41]. The purpose of the present article is to explore what can be said when $x \in \mathbf{R}^m$ and $y \in \mathbf{R}^n$ live in spaces with different dimensions $m > n$, as in e.g. [22] [10] [30].

Although the symmetry between x and y is destroyed, the duality theorem from linear programming, [27] [37] [4], strongly suggests that the problem can still be reduced to finding a single scalar potential $u(x)$ or $v(y)$ reflecting the relative scarcity of supply f at x (or demand g at y). Although this potential solves a minimization problem, it is not clear what equation, if any, selects it. Nor whether one expects its solution to be smoother than Lipschitz and semiconvex [21] [19]. These are among the questions addressed hereafter. Our primary results are as follows: We exhibit an integro-differential equation which selects $v(y)$. In contradistinction to the case investigated by Ma, Trudinger and Wang, our equation, though still fully nonlinear, is in general nonlocal. However, we also show this equation has two local analogs, one of which is at least degenerate-elliptic. These may or may not admit solutions: however under mild topological conditions, it turns out they admit a C^2 smooth, strongly elliptic solution if and only if the dual linear program admits C^2

minimizers. These locality criteria build upon our results with Chiappori [10] from $n = 1$, and extend the notion of nestedness introduced there to targets of arbitrary dimension. We also relax and refine the notion of nestedness, leading to regularity results for a large new class of examples even when $n = 1$. Our basic set-up is as follows. Fix $m \geq n \geq 1$ and open sets $X \subset \mathbf{R}^m$ and $Y \subset \mathbf{R}^n$ equipped with Borel probability densities f and g. We say $F : X \longrightarrow Y$ pushes f forward to $g = F_{\#}f$ if F is Borel and

$$\int_Y \psi(y)g(y)dy = \int_X \psi(F(x))f(x)dx, \tag{5}$$

for all bounded Borel test functions $\psi \in L^\infty(Y)$. If, in addition, F happens to be Lipschitz and its (n-dimensional) Jacobian $JF(x) := \det^{1/2}[DF(x)DF^T(x)]$ vanishes at most on a set of f measure zero, then the co-area formula yields

$$g(y) = \int_{F^{-1}(y)} \frac{f(x)}{JF(x)} d\mathcal{H}^{m-n}(x) \tag{6}$$

for a.e. $y \in Y$, where $\mathcal{H}^k$ denotes k-dimensional Hausdorff measure.

Given a surplus function $s \in C^2(X \times \bar{Y})$, Monge's problem is to compute

$$\bar{s}(f,g) := \sup_{F_{\#}f=g} \int_X s(x, F(x))f(x)dx, \tag{7}$$

where the supremum is taken over maps F pushing f forward to g. The supremum is well-known to be uniquely attained provided $X \times Y$ is open and s is *twisted* [40], meaning $D_x s(x,\cdot)$ acts injectively on $\bar{Y}$ for each $x \in X$; here $\bar{Y}$ denotes the closure of Y. It can be characterized through the Kantorovich dual problem

$$\bar{s}(f,g) = \min_{u(x)+v(y)\geq s(x,y)} \int_X u(x)f(x)dx + \int_Y v(y)g(y)dy, \tag{8}$$

where the minimum is taken over pairs $(u,v) \in L^1(f) \oplus L^1(g)$ satisfying $u \oplus v \geq s$ throughout $X \times Y$. Dual minimizers of the form $(u,v) = (v^s, u^{\tilde{s}})$ are known to exist [40], where $\tilde{s}(y,x) := s(x,y)$ and

$$v^s(x) = \sup_{y\in\bar{Y}} s(x,y) - v(y) \qquad u^{\tilde{s}}(y) = \sup_{x\in X} s(x,y) - u(x). \tag{9}$$

Such pairs of payoff functions are called s-conjugate, and u and v are said to be s- and $\tilde{s}$-convex, respectively.

To motivate our first result, here and hereafter let

$$Y \subset \mathbf{R}^n \text{ be open and bounded, } X \subset \mathbf{R}^m \text{ be open, } m \geq n \tag{10}$$
$$\text{and } s \in C^2(X \times \bar{Y}) \text{ be twisted, } \textit{non-degenerate}, \tag{11}$$

meaning in addition to the injectivity of $y \in \bar{Y} \mapsto D_x s(x,y)$ mentioned above that $D^2_{xy}s(x,y)$ has maximal rank throughout $X \times \bar{Y}$. Here $C^2(X \times \bar{Y})$ denotes the usual Banach space. Suppose F maximizes the primal problem (7) and $(u,v) = (v^s, u^{\tilde{s}})$ is an s-conjugate pair of payoffs minimizing the dual problem (8). Then $u(x) + v(y) - s(x,y) \geq 0$ on $X \times \bar{Y}$, with equality on graph(F). Thus

$$
\begin{aligned}
F^{-1}(y) &\subset \partial_{\tilde{s}} v(y) && (12)\\
&:= \{x \in X \mid s(x,y) - v(y) = \sup_{y' \in \bar{Y}} s(x,y') - v(y')\}. && (13)
\end{aligned}
$$

Note that since $\bar{Y}$ is bounded, v must be bounded below — otherwise $u = v^s \equiv +\infty$ and (u,v) cannot solve (8); this implies that u is locally Lipschitz as in [33, Lemma 1], hence the level set $\partial_{\tilde{s}} v(y)$ of $x \mapsto s(x,y) - u(x)$ is closed, a fortiori measurable. Since $s \in C^2(X \times \bar{Y})$, u and v admit second-order Taylor expansions Lebesgue a.e. as in e.g. [21] [40], and the first- and second-order conditions for equality on graph(F) imply

$$
\begin{aligned}
Dv(F(x)) &= D_y s(x, F(x)) \qquad [f\text{-a.e.] and} && (14)\\
D^2 v(F(x)) &\geq D^2_{yy} s(x, F(x)) \qquad [f\text{-a.e.}]. && (15)
\end{aligned}
$$

Differentiating the first-order condition yields

$$
[D^2 v(F(x)) - D^2_{yy} s(x, F(x))] DF(x) = D^2_{xy} s(x, F(x)) \qquad [f\text{-a.e.}] \quad (16)
$$

as in e.g. [31]. Since $D^2_{xy}s$ has full-rank, when F happens to be Lipschitz we identify its Jacobian f-a.e. as

$$
JF(x) = \frac{\sqrt{\det[D^2_{xy} s(x, F(x))(D^2_{xy} s(x, F(x)))^T]}}{\det[D^2 v(F(x)) - D^2_{yy} s(x, F(x))]}. \quad (17)
$$

In this case we can rewrite (6) in the form

$$
g(y) = \int_{F^{-1}(y)} \frac{\det[D^2 v(y) - D^2_{yy} s(x,y)]}{\sqrt{\det D^2_{xy} s(x,y)(D^2_{xy} s(x,y))^T}} f(x) d\mathcal{H}^{m-n}(x). \quad (18)
$$

Note that although we neither assume nor establish Lipschitz continuity of F in the sequel, for $s \in C^2$ twisted the s-convexity of u makes F countably Lipschitz, as in e.g. Theorem 3.16 of [37].

Except for the appearance of the map F in the domain of integration, this would be a partial differential equation relating v to the data (s, f, g). However, using twistedness of the surplus we'll show that the containment (12) is essentially saturated, in the sense that there exists a certain set $X(v^s) \subset \operatorname{dom} D^2 v^s$ of full $\mathcal{H}^m$ measure such that for a.e. $y \in Y$,

$$g(y) = \int_{\partial_{\tilde{s}} v(y) \cap X(v^s)} \frac{\det(D^2 v(y) - D^2_{yy} s(x,y))}{\sqrt{\det D^2_{xy} s(x,y) (D^2_{xy} s(x,y))^T}} f(x) d\mathcal{H}^{m-n}(x) \quad [\mathcal{H}^n\text{-a.e.}]. \tag{19}$$

Indeed, $u = v^s$ is known to be semiconvex, meaning there exists $k \in \mathbf{R}$ such that $u(x) + k|x|^2$ is convex on every ball $B \subset X$. Thus $Du \in BV_{loc}(X)$, and §6.6.1–2 of [17] imply that for each positive integer i, there is a continuously differentiable map H_i on X with $H_i = Du$ and $DH_i = D^2 u$ outside a set Z_i of volume $\mathcal{H}^m[Z_i] < 1/i$. Moreover, we may take $Z_{i+1} \subset Z_i$ and set $X(v^s) := (\operatorname{dom} D^2 v^s) \setminus Z_\infty$ where $Z_\infty := \cap_{i>0} Z_i$ has zero volume. Here $\operatorname{dom} D^2 u$ denotes the subset of X where u admits a second-order Taylor expansion.

Now (19) is an analog of the Monge-Ampère equation (1), familiar from the case $s(x,y) = -\frac{1}{2}|x-y|^2$, or equivalently $s(x,y) = x \cdot y$. Notice the boundary condition (2) for that case is automatically subsumed in formulation (19). However, unlike the case $m = n$, it is badly nonlocal since the essential domain of integration $\partial_{\tilde{s}} v(y)$ defined in (12) may potentially depend on $v(y')$ for all $y' \in Y$.

For twisted non-degenerate s and an s-convex v, our first result states that v satisfies (19) if and only if v combines with its conjugate $u = v^s$ to minimize (8); see Corollary 3 of §2. Since the optimal map F can be recovered from the first-order condition

$$D_x s(x, F(x)) = Du(x), \tag{20}$$

analogous to (14), this shows Monge's problem has been reduced to the solution of the partial differential equation (19) for the $\tilde{s}$-convex scalar function v.

Although non-locality makes this equation a challenge to solve, it turns out there is a class of problems for which (19) can be replaced by a local partial differential equation, as follows. Introduce the $m-n$ dimensional submanifold

$$X_1(y,p,Q) := X_1(y,p) := \{x \in X \mid D_y s(x,y) = p\}$$

of X and its closed subset

$$X_2(y,p,Q) := \{x \in X_1(y,p) \mid D^2_{yy} s(x,y) \le Q\}. \tag{21}$$

Now (14)–(15) imply

$$\partial_{\tilde{s}} v(y) \subset X_2(y, Dv(y), D^2 v(y)) \subset X_1(y, Dv(y)) \tag{22}$$

for all $y \in Y \cap \operatorname{dom} D^2 v$; here $\operatorname{dom} D^2 v$ denotes the subset of $\bar{Y}$ where v admits a second-order Taylor expansion. It is often the case that one or both of these containments becomes an equality, at least up to $\mathcal{H}^{m-n}$ negligible sets. In this case locality is restored: we can then write (19) in the form

$$G(y, Dv(y), D^2 v(y)) = g(y) \qquad [\mathcal{H}^n\text{-a.e. } y \in Y], \tag{23}$$

where

$$\begin{aligned} G(y,p,Q) &:= G_i(y,p,Q) \\ &:= \int_{X_i(y,p,Q)} \frac{|\det(Q - D^2_{yy} s(x,y))|}{\sqrt{\det D^2_{xy} s(x,y)(D^2_{xy} s(x,y))^T}} f(x) d\mathcal{H}^{m-n}(x) \end{aligned} \tag{24}$$

and either $i = 1$ or $i = 2$.

Our second result states any classical s-convex solution $v \in C^2(\bar{Y})$ to either local problem (23) also solves the nonlocal one (19); Corollary 4. Assuming connectedness of $X_i(y, Dv(y), D^2 v(y))$, we show such a solution exists in Theorem 6 of §3, if and only if the dual minimization (8) admits a C^2 solution $(v^s, v = v^{s\tilde{s}})$. For an $n = 1$ dimensional target, other necessary and sufficient conditions for the more restrictive variant $i = 1$ to admit an $\tilde{s}$-convex solution have been given in joint work with Chiappori [10]. There the ordinary differential equation (23) is also analyzed to show v inherits smoothness from suitable conditions on the data (s, f, g) in this so-called nested case. Although it is not be needed in what follows, we recall this definition for reference: when $n = 1$, (s, μ, ν) is said to be *nested* if for every $y_0 < y_1 \in Y$ with $\int_{y_0}^{y_1} g(y)dy > 0$ we have $X_\leq(y_0, k(y_0)) \subseteq X_<(y_1, k(y_1))$, where

$$X_\leq(y,k) := \{x \in X : D_y s(x,y) \leq k\} = \bigcup_{p \leq k} X_1(y,p),$$

$X_<(y,k) := X_\leq(y,k) \setminus X_1(y,k)$, and $k(y)$ is defined as any solution of the proportional population splitting equation:

$$\int_{X_\leq(y,k(y))} f(x)dx = \int_{-\infty}^{y} g(z)dz.$$

In this case, we show in [10] that any anti-derivative $v(y) = \int_{y_0}^{y} k(z)dz$ of k is $\tilde{s}$-convex, (v^s, v) minimizes (8), and that $\partial_{\tilde{s}} v(y) = X_1(y, k(y))$, so that v solves the $i = 1$ version of (23). It is therefore consistent to take the existence of a solution to (23) with $i = 1$ as the definition of *nestedness* in higher dimensions ($n > 1$).

We go on to show that the operator G_2 is degenerate elliptic in §5, and that the ellipticity is strict at points where $G_2 > 0$. As a consequence, we are able to deduce higher regularity of solutions v of (23) with $i = 2$ from $C^{2,\alpha}$ regularity in Theorem 13, provided G_2 is sufficiently smooth. In Theorem 14 of §6 we establish this smoothness for the simpler operator G_1, allowing for the passage from $C^{2,\alpha}$ to higher regularity when $G_2 = G_1$. For one-dimensional targets, we establish the smoothness of G_2 in Theorem 23 of §7, whether or not it coincides with G_1. The hypothesized second order smoothness and uniform ellipticity of v remain intriguing open questions — with partial resolutions known only in the cases $n = m$ of Ma, Trudinger and Wang [31] [38] (which built on earlier work of Caffarelli [6] [7], Delanoe [16] and Urbas [39]), and for $n = 1$ in the nested case [10]; to these we now add the non-nested cases which satisfy the local equations (23)–(24) with $(n, i) = (1, 2)$, resolved in §8 below. When regularity fails for $m = n$ the size of the singular set has been estimated by DePhilippis and Figalli [15], building on work of Figalli [18] with Kim [20]; for related results see Kitagawa and Kim [25] and the survey [14].

Remark 1 (Boundedness of the domains) *Throughout the rest of the paper $Y \subset \mathbf{R}^n$ is assumed bounded, while $X \subset \mathbf{R}^m$ is bounded from Section 6 onwards. We define $\partial X = \emptyset$ when $X = \mathbf{R}^m$. The source and target measures to be transported are assumed to be absolutely continuous with respect to Lebesgue, and given by densities f on X and g on Y. The boundedness of domains is mainly assumed for simplicity; as is often the case in optimal transport theory, we believe that many of the results here can be established on unbounded domains with suitable decay assumptions on the measures. However, we expect that on unbounded domains the class of problems satisfying the local equations will be greatly reduced, and in some cases trivial. To illustrate this expectation, we consider the simplest and best understood case of unequal dimensional optimal transport, the theory corresponding to the $i = 1$ equation with one dimensional target ($n = 1$) from [10]. If $s(x, y) = x \cdot H(y)$ for some curve $H : Y \subseteq \mathbf{R} \to \mathbf{R}^m$, then the sets $X_1(y, p)$ are hyperplanes orthogonal to $H'(y)$. Therefore, if $X = \mathbf{R}^m$, any two level sets $X_1(y, p)$ and $X_1(y', p')$ intersect in X unless $H'(y)$ and $H'(y')$ are parallel. By Corollary 5.6 in [10], the model cannot be nested for arbitrary f and g unless the direction of $H'(y)$ is constant; in this case, $H(Y)$ is contained in a line l and $s(x, y) = x_l \cdot H(y)$ (where x_l is the projection of x onto l) takes the index form known to reduce the problem to a optimal transport problem between the densities $f_l = (x \mapsto x_l)_\# f$ and g with one dimensional support.*

On the other hand, when Y is unbounded the density g cannot have a lower bound. In many cases, this precludes nestedness (see, for example, Theorem 3 in [35]).

2 A nonlocal partial differential equation for optimal transport

Given $X \subset \mathbf{R}^m$ and $Y \subset \mathbf{R}^n$, a Borel probability density f on X and a Borel map $F : X \longrightarrow Y$, we define the pushed-forward measure $\nu := F_{\#}f$ by

$$\int_Y \psi(y) d\nu(y) = \int_X \psi(F(x)) f(x) dx \tag{25}$$

for all bounded Borel functions $\psi \in L^\infty(Y)$. This definition extends (5) to the case where ν need not be absolutely continuous with respect to Lebesgue; however when ν is absolutely continous with Lebesgue density g, we abuse notation by writing $g = F_{\#}f$.

Recall $s \in C^2(X \times \bar{Y})$ is *twisted* if for each $x \in X$ the map $y \in \bar{Y} \mapsto D_x s(x,y)$ is injective. If

$$D_x s(x,y) = p$$

we can then deduce y uniquely from x and p, in which case we write $y = s\text{-}\exp_x p := D_x s(x,\cdot)^{-1}(p)$. The non-degeneracy (11) of s (full-rank of $D^2_{xy}s$) guarantees s-exp is a continuously differentiable function of (x,p) where defined, by the implicit function theorem. Thus for a twisted cost function, the first-order condition (20) allows us to identify the map $F = s\text{-}\exp \circ Du$ at points of X where u happens to be differentiable. We denote the set of such points by $\operatorname{dom} Du$. Similarly we denote the set of points where $F : X \longrightarrow \bar{Y}$ is approximately differentiable by $\operatorname{dom} DF$, and the set where u admits a second order Taylor expansion by $\operatorname{dom} D^2u$. When s is non-degenerate and twisted, (20) implies $\operatorname{dom} DF = \operatorname{dom} D^2u$.

Theorem 2 (Properties of potential maps) *Fix $X \subset \mathbf{R}^m, Y \subset \mathbf{R}^n$ and s as in (10)-(11). Any pair $(u,v) = (v^s, u^{\tilde{s}})$ of s-conjugate functions (9) are semiconvex, Lipschitz, and have second-order Taylor expansions Lebesgue a.e. The map $F : \operatorname{dom} Du \longrightarrow \bar{Y}$ satisfying (20) is unique and differentiable Lebesgue a.e. Decompose $\bar{Y}$ into $Y_+ := Y \cap \operatorname{dom} D^2 v$ and $Y_- = \bar{Y} \setminus Y_+$ and set $X_\pm := F^{-1}(Y_\pm)$. The Jacobian $JF(x) := \det^{1/2}[DF(x)DF(x)^T]$ is positive on $X_+ \cap \operatorname{dom} DF$ and given there by*

$$JF(x) = \frac{\sqrt{\det[D^2_{xy}s(x,F(x)) D^2_{xy}s(x,F(x))^T]}}{\det[D^2 v(F(x)) - D^2_{yy}s(x,F(x))]}. \tag{26}$$

Any Borel probability density on X can be decomposed as $f = f_+ + f_-$, where $f_\pm = f 1_{X_\pm}$ are mutually singular. Their images $F_\#(f_\pm)$ are measures living on the disjoint sets $Y_\pm$. Here $F_\#(f_+)$ is absolutely continuous with respect to Lebesgue: its density given by

$$g_+(y) = \int_{F^{-1}(y)\cap X(v^s)} \frac{f_+(x)}{JF(x)} d\mathcal{H}^{m-n}(x) \qquad [\mathcal{H}^n\text{-a.e. } y \in \bar{Y}] \tag{27}$$

$$= \int_{\partial_{\tilde{s}} v(y)\cap X(v^s)} \frac{f_+(x)\det[D^2 v(y) - D^2_{yy} s(x,y)]}{\sqrt{\det D^2_{xy} s(x,y)(D^2_{xy} s(x,y))^T}} d\mathcal{H}^{m-n}(x), \tag{28}$$

where $X(v^s) \subset \operatorname{dom} D^2 v^s$ is the set defined after (19). If $y \in Y^+$ then $f_+ = f$ on $\partial_{\tilde{s}} v(y) \cap X(v^s)$. Moreover, if $F_\#(f_-)$ is absolutely continuous with respect to Lebesgue and assigns zero mass to ∂Y, then $f = f_+$ in (27)–(28).

Proof. It is well-known that $u = v^s$ and $v = u^{\tilde{s}}$ are Lipschitz and semiconvex [34, Lemma 3.1]: they inherit distributional bounds such as $|Du| \le \sup_Y |D_x s|$ and $D^2 u \ge \inf_Y D^2_{xx} s$ from $s \in C^2$. This implies they extend continuously to $\bar{X}$ and $\bar{Y}$, where they are twice differentiable a.e. by Alexandrov's theorem [40, Theorem 14.25]; indeed, for $x_0 \in \operatorname{dom} D^2 u$ we have

$$0 = \lim_{x\to x_0} \sup_{p\in\partial u(x)} \frac{p - Du(x_0) - D^2 u(x_0)(x - x_0)}{|x - x_0|} \tag{29}$$

which asserts differentiability (rather than just approximate differentiability) of Du at x_0.

Recall $u(x) + v(y) - s(x, y) \ge 0$ on $X \times \bar{Y}$. For each $x \in \operatorname{dom} Du$ at least one $y \in \bar{Y}$ produces equality, since the maximum (9) defining $v^s(x)$ is attained. This y satisfies the first order condition $D_x s(x, y) = Du(x)$, which identifies it as $y = F(x)$ by the twist condition. We abbreviate $F = s\text{-}\exp \circ Du$. We note Du is differentiable a.e. in a neighbourhood of $x \in \operatorname{dom} F$, and the map s-exp is well-defined and continuously differentiable in a neighourhood of $(x, Du(x))$ by the twist and non-degeneracy of s. From the definition of $X(v^s)$ following (19), recall that for each positive integer i, there is a C^1 smooth map H_i and set Z_i of volume $\mathcal{H}^m[Z_i] < 1/i$, such that both $Du = H_i$ and $D^2 u = DH_i$ hold outside Z_i. Thus there is a map $F_i \in C^1(X, \bar{Y})$ with $F = F_i$ and $DF = DF_i$ outside Z_i, which admits a Lipschitz extension (also denoted F_i) to all of $\mathbf{R}^m$. As a result F is countably Lipschitz (and approximately differentiable Lebesgue a.e.); the fact that it is actually differentiable a.e. follows from s-exp $\in C^1$ and (29). As remarked after (19), the sets $Z_{i+1} \subset Z_i$ may be taken to be nested. We define $Z_\infty := \cap_{i>0} Z_i$ and $X(v^s) := \operatorname{dom} D^2 v^s \setminus Z_\infty$.

For each $\phi \in L^1(\mathbf{R}^m)$, the co-area formula [17, §3.4.3] asserts

$$\int_{\mathbf{R}^m} \phi(x) JF_i(x) dx = \int_{\mathbf{R}^n} dy \int_{F_i^{-1}(y)} \phi d\mathcal{H}^{m-n}. \tag{30}$$

If ϕ vanishes outside $X \setminus Z_i$ we may drop the subscripts i in the formula above. Thus if $\mathcal{H}^m[Z] = 0$ for any $Z \subset X(v^s)$, we conclude $\mathcal{H}^{m-n}[Z \cap F^{-1}(y)] = 0$ for almost all $y \in \mathbf{R}^n$, and similary if $Z \subset X(v^s)$ is f-negligible, then $Z \cap F^{-1}(y)$ is $fd\mathcal{H}^{m-n}$-negligible for almost all $y \in \mathbf{R}^n$.

Since $u(x) + v(y) - s(x, y) \geq 0$ vanishes at $(x, F(x)) \in X \times Y$ for each $x \in X_+$, we can differentiate (14) if $x \in \operatorname{dom} DF$ to obtain (16). Since the right hand side has rank n we conclude both factors on the left must have rank n as well. This shows $JF(x) > 0$ and noting (15) establishes (26).

Decomposing a probability density $f = f_+ + f_- + f_0$ on X into $f_\pm = f 1_{X_\pm}$ and $f_0 = 1_{X_0}$, where $X_0 := X \setminus \operatorname{Dom} Dv^s$ is Lebesgue negligible, the asserted mutual singularity follows from $Y_+ \cap Y_- = \emptyset = X_+ \cap X_-$. Moreover, the bounded increasing sequence $X_i^+ := \{x \in X_+ \setminus Z_i \mid |x| \leq i, JF(x) > 1/i\}$ of sets exhausts $X_+ \setminus Z_\infty$. Setting $X_0^+ := \emptyset$ and $X_i := X_i^+ \setminus X_{i-1}^+$ decomposes $X_+ \setminus Z_\infty = \cup_{i=1}^\infty X_i$ into countably many disjoint Borel sets $X_i \subset \mathbf{R}^m$ on which F is C^1 with $JF(x) > 1/i$ on X_i. Set $X_\infty := X_+ \cap Z_\infty$ and let $f_i = f_+ 1_{X_i}$ denote the restriction of f_+ to X_i, and $g_i := F_\# f_i$ the density of the push-forward of f_i. The absolute continuity of $F_\# f_i$ with respect to Lebesgue follows since $JF(x) > 1/i$ on X_i. Given $\psi \in L^\infty(\mathbf{R}^n)$ with bounded support ensures $\phi = f_i \psi \circ F_i / JF_i \in L^1(\mathbf{R}^m)$ hence (30) implies

$$\begin{aligned}
\int_{\mathbf{R}^n} g_i \psi &= \int_{\mathbf{R}^m} f_i \psi \circ F_i \\
&= \int_{\mathbf{R}^n} dy \psi(y) \int_{X_i \cap F_i^{-1}(y)} \frac{f_i}{JF_i} d\mathcal{H}^{m-n}.
\end{aligned}$$

Recalling $F_i = F$ and $DF_i = DF$ on $X_i \subset X(v^s)$, we infer

$$g_i(y) = \int_{F^{-1}(y) \cap X(v^s)} \frac{f_i(x)}{JF(x)} d\mathcal{H}^{m-n}(x). \tag{31}$$

a.e. since $\psi \in L^\infty$ had bounded support but was otherwise arbitrary. Summing (31) on i, the disjointness of X_i yields $1_{X_+} = 1_{X_\infty} + \sum 1_{X_i}$, hence $f_+ = f_\infty + \sum f_i$ and $g_+ = \sum g_i$. Since $f_\infty = 0$ on $X(v^s)$, (27) holds for Lebesgue a.e. $y \in \bar{Y}$. By the monotone convergence theorem, $g_+ \in L^1(\bar{Y}, d\mathcal{H}^n)$ and has mass at most one; thus for a.e. $y \in \bar{Y}$ its Lebesgue density $g_+(y)$ is finite and implies finiteness of the integral in (27).

To establish (28), it suffices to observe (26) holds wherever $f_+ > 0$ and to verify $F^{-1}(y) = \partial_{\tilde{s}} v(y) \cap \mathrm{Dom}\, Dv^s$. The containment $\subset$ follows from (12), so we need only consider the reverse inclusion. Given $x \in \partial_{\tilde{s}} v(y) \cap \mathrm{Dom}\, Dv^s$, the twist condition (11) implies $y = F(x)$, hence $x \in F^{-1}(y)$.

Now $y \in Y_+$ implies $f = f_+$ on $F^{-1}(y) \subset X_+$, hence on $\partial_{\tilde{s}} v(y) \cap X(v^s)$ by the previous sentence. If f charges $X \setminus X_+$, both f_- and $F_\# f_-$ are non-zero, in which case $F_\# f_-$ charges either ∂Y or the Lebesgue negligible set $Y \setminus \mathrm{dom}\, D^2 v$ which comprise Y_-. When the latter possibilities are ruled out by hypothesis, then $f_- = 0$ a.e., implying $f = f_+$ holds in (27)–(28). ■

Corollary 3 (Equivalence of optimal transport to nonlocal PDE) *Under the hypotheses of Theorem 2, let f and g denote probability densities on X and Y. If $v = v^{s\tilde{s}}$ satisfies the nonlocal equation* (19) *[a.e. on Y], then (v^s, v) minimize Kantorovich's dual problem* (8)*. Conversely, if $(u, v) = (v^s, u^{\tilde{s}})$ minimize* (8) *then v satisfies* (19) *[a.e. on $\bar{Y}$].*

Proof. First suppose $v = v^{s\tilde{s}}$ satisfies the nonlocal PDE (19) on Y. Setting $u = v^s$ implies for each $x \in \mathrm{dom}\, Du$ the inequality

$$u(x) + v(y) - s(x, y) \geq 0 \tag{32}$$

is saturated by some $y \in \bar{Y}$. Identifying $F(x) = y$ we have the first-order condition (20), whence $F = s\text{-}\exp \circ Du$ on $\mathrm{dom}\, Du$. We claim it is enough to show $F_\# f = g$: if so, integrating

$$u(x) + v(F(x)) = s(x, F(x))$$

against f yields

$$\int_X uf + \int_Y vg = \int_X s(x, F(x)) f(x) dx,$$

which in turn shows F maximizes (7) and (u, v) minimizes (8) as desired. To establish $F_\# f = g$, comparing (19) with (28) yields $g_+ \leq g$ on Y, with equality holding throughout $Y_+ := Y \cap \mathrm{dom}\, D^2 v$. Since Y_+ and Y differ by a Lebesgue negligible set, $g_+ = g$ is a probability measure. This implies $F_\# f_- = 0$, hence $g = g_+ = F_\# f$ as desired.

Conversely, suppose $(u, v) = (v^s, u^{\tilde{s}})$ minimizes (8). Since twistedness of s implies (7) is attained, there is some map $F : X \longrightarrow \bar{Y}$ pushing f forward to g such that (32) becomes an equality f-a.e. on $Graph(F)$. This ensures $F = s\text{-}\exp \circ Du$ holds f-a.e. Since $Y_+ := Y \cap \mathrm{dom}\, D^2 v$ is a set of full measure for g, we conclude $X_+ = F^{-1}(Y_+)$ has full measure for f, whence $f_+ := f 1_{X_+} = f$ and $g_+ := F_\#(f_+) = g$. Now (19) follows from (28) as desired. ■

Corollary 4 (Optimal transport via local PDE) *Under the hypotheses of Theorem 2, let f and g denote probability densities on X and Y. Fix $i \in \{1,2\}$. If $v = v^{s\tilde{s}}$ satisfies the local equation* (23)–(24) [a.e.] *then the following three statements become equivalent:*
(a) (v^s, v) minimizes Kantorovich's dual problem (8)*;*
(b) $(s\text{-}\exp \circ Dv^s)_{\#} f$ vanishes on $\bar{Y} \setminus \operatorname{dom} D^2 v$;
(c)

$$\int_{X_i(y,Dv(y),D^2v(y))\setminus(\partial_{\tilde{s}}v(y)\cap X(v^s))} f(x) d\mathcal{H}^{m-n}(x) = 0 \qquad [\mathcal{H}^n\text{-a.e. } y \in Y]. \tag{33}$$

Proof. Fix $i \in \{1,2\}$ and suppose $v = v^{s\tilde{s}}$ satisfies the local PDE (23). As in the preceding proof, setting $u = v^s$ implies for each $x \in \operatorname{dom} Du$ the inequality

$$u(x) + v(y) - s(x,y) \geq 0 \tag{34}$$

is saturated by some $y \in \bar{Y}$. Setting $F(x) = y$ we have the first-order condition (20), whence $F = s\text{-}\exp_x \circ Du$ on $\operatorname{dom} Du$.

(b) $\Rightarrow$ (a). Hypothesis (b) asserts that $Y_+ = Y \cap \operatorname{dom} D^2 v$ forms a set of full measure for $F_{\#} f$. Thus $f_- = 0$, while $f = f_+$ and g_+ are both probability densities in Theorem 2. Recalling $\partial_{\tilde{s}} v(y) \subset X_2(y, Dv(y), D^2 v(y))$ for $y \in Y_+$ from (22), we deduce $g \geq g_+$ by comparing (23) with (28). Since both densities integrate to 1, this implies $g = g_+$ a.e. Thus (19) is satisfied and Corollary 3 asserts (v^s, v) minimizes (8).

(a) $\Rightarrow$ (c). Conversely, when (a) holds, Corollary 3 implies (19), hence $\mathcal{H}^n$-a.e. $y \in Y$ satisfies

$$\begin{aligned} g(y) &= \int_{\partial_{\tilde{s}} v(y) \cap X(v^s)} \frac{f(x)}{JF(x)} d\mathcal{H}^{m-n}(x) && (35) \\ &\leq \int_{X_i(y,Dv(y),D^2v(y))} \frac{f(x)}{JF(x)} d\mathcal{H}^{m-n}(x) && (36) \\ &= g(y) && (37) \end{aligned}$$

where (22) and (23) have been used. Now (c) follows immediately.

(c) $\Rightarrow$ (b). Conversely, (c) and (23) imply (35)–(37), in which case Corollary 3 and its proof assert $g = F_{\#} f$. Since g vanishes on $\bar{Y} \setminus \operatorname{dom} D^2 v$ by hypothesis, the desired conclusion (b) is established. ∎

Remark 5 (Notes and queries) *Fix probability densities f, g as above. When a minimizing pair $(u, v) = (v^s, u^{\tilde{s}})$ for* (8) *happens to satisfy the local equation* (23)–(24) *— as in the nested case — then the containment* (22) *shows that we may replace $X(v^s)$ by $\mathbf{R}^m$ in the nonlocal equation* (19)*. The original draft of the present manuscript claimed this* *was true more generally, but our argument there suffered from a gap (which we would be glad to know how to close): although $X \setminus X(v^s)$ is $\mathcal{H}^m$-negligible, we cannot be sure that its intersection with $\partial_{\tilde{s}} v(y)$ is $\mathcal{H}^{m-n}$-negligible for $\mathcal{H}^n$-a.e. $y \in Y$ — unless the map $F(x) = s\text{-}\exp Dv^s$ happens to be Lipschitz instead of countably Lipschitz.*

When $m = n$ and both s and $\tilde{s}(y,x) = s(x,y)$ are twisted, it follows e.g. from Theorem 11.1 of [40] that any minimizing pair $(u,v) = (v^s, u^{\tilde{s}})$ satisfies the local equation with $i = 1$. If, in addition, s satisfies Ma-Trudinger-Wang condition (A3w) of [31] [38], the converse can be shown: if $v = v^{s\tilde{s}}$ solves the $i = 1$ local equation a.e. then (v^s, v) minimizes (8)*. (Here (A3w) is to deduce a.e. injectivity of F from $JF(x) > 0$, using the connectedness of $\partial_s u(x)$ shown by Loeper [29].) In this case (a) follows from the other hypotheses of Corollary 4. We don't know whether or not the conditions (a)-(c) are similarly redundant under weaker hypotheses, as when $m > n$. If not, then:*

Since (b) holds whenever $v \in C^2(\bar{Y})$, it might conceivably turn out to be a criterion for selecting (i.e. or defining) an appropriate notion of weak solution among nonsmooth functions $v = v^{s\tilde{s}}$ satisfying (23)–(24).

3 Local PDE from optimal transport

As a partial converse to the preceding corollary, we assert that for either the more restrictive ($i = 1$) or less restrictive ($i = 2$) local partial differential equation (23) to admit solutions, it is sufficient that the Kantorovich dual problem admit a smooth minimizer (u, v), with connected potential indifference sets $X_i(y, Dv(y), D^2v(y))$ — in which case v also solves (23).

Theorem 6 (When a smooth minimizer implies nestedness) *Fix s and probability densities f and g on $X \subset \mathbf{R}^m$ and $Y \subset \mathbf{R}^n$ as in* (10)–(11)*. Let $i \in \{1, 2\}$. If $(u,v) = (v^s, u^{\tilde{s}}) \in C^2(X) \times C^2(Y)$ minimizes the Kantorovich dual* (8) *then equation* (23) *holds* $\mathcal{H}^n$-a.e. *on any measurable $Y' \subset Y$ having $X_i(y, Dv(y), D^2v(y))$ connected for all $y \in Y'$.*

The assumed smoothness of u and v is essential. When the dual problem (8) has no smooth optimizers, Remark 5 shows the local equation (23) cannot have smooth c-convex solutions, neither for $i = 2$ nor for $i = 1$. For example, the explicit solution computed with Chiappori in Section 3.3.3 of [9] solves the non-local equation (19) but neither local version (23). Conversely, having solutions to either local equation will often imply smoothness of v, as in the nested case [10] when $n = 1$ and $i = 1$, and the last section of the present paper when $n = 1$ and $i = 2$. It is quite possible, however, for smooth solutions v to the $i = 2$ local equation to produce non-smooth $u = v^s$, as Example 8 below illustrates.

Proof. Corollary 3 implies v solves the non-local equation (19) a.e., with $X(v^s) = X$ since $u \in C^2(X)$ by hypothesis. The local equation $G = g$ follows wherever we have equality in the inclusion

$$\partial_{\tilde{s}} v(y) \subset X_i(y, Dv(y), D^2 v(y)). \tag{38}$$

We now derive this equality for all $y' \in Y$ with $\partial_{\tilde{s}} v(y')$ non-empty and $X_i' := X_i(y', Dv(y'), D^2 v(y'))$ connected.

Observe both $\partial_{\tilde{s}} v(y')$ and X_i' are relatively closed subsets of X. Thus $\partial_{\tilde{s}} v(y')$ is also closed relative to X_i'. To show it is relatively open, let $x' \in \partial_{\tilde{s}} v(y')$. Since $u, v \in C^2$ we see $F \in C^1(X)$ and DF has full rank at x'. By the Local Submersion Theorem [23], this means we can find a C^1 coordinate chart on a neighbourhood $U \subset X$ of x' in which F acts as the canonical submersion: $F(x_1, \ldots, x_n, x_{n+1}, \ldots, x_m) = (x_1, \ldots, x_n)$. In these coordinates,

$$[\{y'\} \times \mathbf{R}^{m-n}] \cap U = F^{-1}(y') \cap U \subset \partial_{\tilde{s}} v(y') \cap U \subset X_i' \cap U \subset X_1' \cap U$$

follows from (38). But Proposition 3.2 of [10] shows X_1' to be an $m - n$ dimensional submanifold of X, so equality must hold in this chain of inclusions (at least if U is a ball in the new coordinates). This shows x' lies in the interior of $\partial_{\tilde{s}} v(y')$ relative to X_i', concluding the proof that $\partial_{\tilde{s}} v(y')$ is relatively open. Thus $\partial_{\tilde{s}} v(y') = X_i'$ since the former is open, closed and non-empty and the latter is connected. Equality in (38) has been established whenever $X_i(y, Dv(y), D^2 v(y))$ is connected, concluding the proof. ■

The following example shows that the level set connectivity assumption in the preceding theorem is required to deduce that smooth solutions to the dual problem solve the local equation; it also illustrates why it may be necessary to consider the $i = 2$ case. In the example, the smooth s-conjugate dual potentials (u, v) solve the $i = 2$ but not $i = 1$ equation; note that each $X_2(y, Dv(y), D^2 v(y)) = \partial_{\tilde{s}} v(y)$ is connected whereas each $X_1(y, Dv(y))$ has two connected components — one is a segment on a ray through the origin and the other its negation.

Example 7 (Annulus to circle) *Consider transporting uniform mass on the annulus, $X = \{x \in \mathbf{R}^2 : 1/2 \leq |x| \leq 1\}$ to uniform measure on the punctured circle, $C = \{(-1, 0) \neq \hat{y} \in \mathbf{R}^2 : |\hat{y}| = 1\}$ with the bilinear surplus, $x \cdot \hat{y}$. It is easy to see that $x \cdot \hat{y} \leq |x|$, with equality only when $\hat{y} = \frac{x}{|x|}$, implying that the optimal map takes the form $x \in X \mapsto \frac{x}{|x|} \in \bar{C}$ has a convex potential $u(x) = |x|$ which is smooth on the annulus X. Parameterizing C by $\hat{y}(\theta) = (\cos(\theta) \sin(\theta))$ for $\theta \in Y := (-\pi, \pi)$ places this problem within our framework. In these coordinates,* *setting* $s(x,\theta) = x \cdot \hat{y}(\theta)$ *we find* $D_\theta s(x,\theta) = x \cdot (-\sin(\theta), \cos(\theta))$ *and* $D^2_{\theta\theta} s(x,\theta) = -x \cdot \hat{y}(\theta)$ *and* $v = u^{\tilde{s}} = 0$. *This means that*

$$\begin{aligned} X_1(\theta, Dv(\theta), D^2v(\theta)) &= \{x \mid x \cdot (-\sin(\theta), \cos(\theta)) = 0\} \\ &= \{x \in X \mid \frac{x}{|x|} = \hat{y}(\theta)\} \cup \{x \in X \mid \frac{x}{|x|} = -\hat{y}(\theta)\} \end{aligned}$$

is disconnected. On the other hand,

$$\begin{aligned} X_2(\theta, Dv(\theta), D^2v(\theta)) &= X_1(\theta, Dv(\theta), D^2v(\theta)) \cap \{x \mid -x \cdot \hat{y}(\theta) \leq 0\} \\ &= \{x \in X \mid \frac{x}{|x|} = \hat{y}(\theta)\} \end{aligned}$$

is connected and coincides with $\partial_{\tilde{s}} v(\theta)$ *(as is guaranteed by the preceding theorem), whereas the inclusion* $\partial_{\tilde{s}} v(\theta) \subseteq X_1(\theta, Dv(\theta), D^2v(\theta))$ *is strict.*

We next alter the preceding example slightly by augmenting X so that the sets X_1 are connected. In this case, we still have a solution to the $i = 2$ but not $i = 1$ local equation (23). Now it is the smoothness of u (rather than connectedness of X_1) required by Theorem 6 that fails to hold.

Example 8 (Disk to circle) *Take* $C = \{(-1,0) \neq \hat{y} \in \mathbf{R}^2 : |\hat{y}| = 1\}$, ν *and* $s(x,\theta) = x \cdot \hat{y}(\theta)$ *as in the preceding example, but now choose* f *to be uniform measure on the disk* $\{x \in \mathbf{R}^2 \mid |x| \leq 1\}$. *The solutions* $(u,v) = (|x|, 0)$ *to* (8) *are as in the last example, but now*

$$X_1(\theta, Dv(\theta)) = \{x \in X \mid \frac{x}{|x|} = \pm\hat{y}(\theta)\} \cup \{0\}$$

is connected, as is

$$X_2(\theta, Dv(\theta), D^2v(\theta)) = \{x \in X \mid \frac{x}{|x|} = \hat{y}(\theta)\} \cup \{0\}.$$

As before, v *solves the* $i = 2$ *but not* $i = 1$ *version of equation* (23).

Concerning the $i = 1$ *case, whereas the connectedness hypothesis on* X_1 *is now satisfied, the regularity* $u \in C^2(X)$ *assumed there now fails, since the singularity of* $u(x) = |x|$ *at the origin* $x = 0$ *is now included in the domain* X.

4 Concerning the regularity of maps

This section collects some conditional results which illustrate how strong s-convexity of v plus a connectedness condition can imply the continuity and differentiability of optimal maps. In the case of equal dimensions, a related connectedness requirement appears in work of Loeper [29]. This section is purely s-convex analytic; no measures are mentioned.

Lemma 9 (Continuity of maps (local)) *Fix X, Y and s as in* (10)–(11). *Let $(u, v) = (v^s, u^{\tilde{s}})$ and $D^2v(y) > D^2_{yy}s(x, y)$ for some $(x, y) \in X \times [Y \cap \partial_s v^s(x) \cap \operatorname{dom} D^2 v]$. Then any C^1 curve in $\partial_s u(x)$ passing through y is constant; in particular, if $\partial_s u(x)$ is C^1-path-connected then $x \in \operatorname{dom} Du$.*

Proof. Fix (u, v) and (x, y) as in the lemma. The proof is by contradiction; if the lemma is false, then there exists a continuously differentiable curve $y : t \in [0, 1] \mapsto y(t) \in \partial_s u(x)$ departing from $y(0) = y$ with non-zero velocity $y'(0) \neq 0$. Since the non-negative function $u(x) + v(\cdot) - s(x, \cdot) \geq 0$ vanishes on this curve, differentiation shows $y'(0)$ to be in the nullspace of $D^2v(y) - D^2_{yy}s(x, y)$. This contradicts the positive-definiteness assertion and shows no such curve can exist.

Thus C^1-path connectedness implies $\partial_s u(x) = \{y\}$. The semiconvexity of u shown in Theorem 2 implies $x \in \operatorname{dom} Du$ provided we can establish convergence of $Du(x_k)$ to a unique limit whenever $x_k \in \operatorname{dom} Du$ converges to x. Therefore, let $x_k \in \operatorname{dom} Du$ converge to x, and choose $y_k \in \partial_s u(x_k)$. Any accumulation point y_∞ of the y_k satisfies $y_\infty \in \partial_s u(x) = \{y\}$. Now letting $k \to \infty$ in $Du(x_k) = D_x s(x_k, y_k)$ yields $Du(x_k) \to D_x s(x, y)$ to establish $x \in \operatorname{dom} Du$. ■

Corollary 10 (Continuity of maps (global)) *Fix X, Y and s as in* (10)–(11). *Let $(u, v) = (v^s, u^{\tilde{s}})$ with $v \in C^2(\bar{Y})$. Then $u \in C^1(X)$ if for each $x \in X$: $\partial_s u(x)$ is C^1-path connected and $D^2v(y) > D^2_{yy}s(x, y)$ for some $y \in Y \cap \partial_s u(x) \cap \operatorname{dom} D^2 v$.*

Proof. Lemma 9 implies $X = \operatorname{dom} Du$ under the hypotheses of Corollary 10. Since semiconvexity of u was shown in Theorem 2, this is sufficient to conclude $u \in C^1(X)$. ■

Proposition 11 (Criteria for differentiability of maps) *Fix X, Y and s as in* (10)–(11). *Use $(u, v) = (v^s, u^{\tilde{s}})$ with $u \in C^1(X)$ to define $F : X \longrightarrow \bar{Y}$ through* (20). *Then both F and $D^2_{xy}s(\cdot, F(\cdot))$ are in $(BV_{loc} \cap C)(X, \mathbf{R}^n)$. If, in addition, $v \in C^{1,1}(Y)$ then $F(x) \in \operatorname{Dom} D^2 v$ for all x in a set of $|DF|$ full measure, and as measures*

$$(D^2v(F(x)) - D^2_{yy}s(x, F(x)))DF(x) = D^2_{xy}s(x, F(x)). \tag{39}$$

In this case, F is Lipschitz in any open subset Z of X admitting $\epsilon > 0$ for which

$$D^2v(F(x)) - D^2_{yy}s(x, F(x)) \geq \epsilon I \tag{40}$$

holds for all $x \in Z$; (moreover, F inherits higher differentiability from v and s in this case).

Proof. Recalling

$$Du(x) = D_x s(x, F(x)), \tag{41}$$

the continuity $F = s\text{-exp} \circ Du$ follows from $u \in C^1(X)$ and the twistedness and non-degeneracy of s.

Since u from Theorem 2 is also semiconvex, its directional derivatives lie in $BV(X)$ and its gradient in $BV(X, \mathbf{R}^m)$. We shall use (41) to deduce $F \in BV_{loc}(X, \mathbf{R}^n)$, which means its directional weak derivatives are signed Radon measures on X. Fix $x' \in X$ and set $y' = F(x') \in Y$. Since $D^2_{xy}s$ has full rank, we can invert (41) to express

$$F(x) = [D_x s(x, \cdot)]^{-1}\, Du(x)$$

as the composition of a C^1_{loc} map and a componentwise BV map. This shows $F \in BV_{loc}(X, \mathbf{R}^n)$ [1].

On the other hand, when Dv is assumed Lipschitz, Ambrosio and Dal Maso [1] assert $F(x) \in \operatorname{Dom} D^2 v$ on a set of $|DF|$ full measure, and differentiating $Dv(F(x)) = D_y s(x, F(x))$ yields (39) in the sense of measures; DF has no jump part since F is continuous. The fact that F inherits the Lipschitz smoothness (and higher differentiability) from Dv on Z follows immediately by rewriting (39)–(40) in the form

$$DF(x) = (D^2 v(F(x)) - s_{yy}(x, F(x)))^{-1} D^2_{xy} s(x, F(x)) \in L^\infty(Z).$$

■

5 Ellipticity and potential regularity beyond $C^{2,\alpha}$

The previous sections show optimal transportation is often equivalent to solving a nonlinear partial differential equation — local or nonlocal. As an application of this reformulation we show how higher regularity of the solution v on the lower dimensional domain can be bootstrapped from its first $2+\alpha$ derivatives. This application, though well-known when $n = m$, is novel in unequal dimensions. It also highlights the need for a theory which explains when v can be expected to be $C^{2,\alpha}_{loc}$, to parallel known results beginning with [6] [31] for $n = m$; we identify conditions ensuring this when $n = 1$ in the last two sections (see Remark 36). Recall that a second-order differential operator $G(y, p, Q)$ is said to be *degenerate elliptic* if $G(y, p, Q') \geq G(y, p, Q)$ whenever $Q' \geq Q$, i.e. whenever $Q' - Q$ is non-negative definite and both Q and Q' are symmetric. We say the ellipticity is *strict* at (y, p, Q) if there is a constant $\lambda = \lambda(y, p, Q) > 0$ called the *ellipticity constant* such that $Q' \geq Q$ implies

$$G(y, p, Q') - G(y, p, Q) \geq \lambda \operatorname{tr}[Q' - Q]. \tag{42}$$

Note that for C^1 operators G, ellipticity is equivalent to everywhere positive semi-definiteness of the matrix $\left(\frac{\partial G}{\partial Q_{ij}}\right)$. Uniform positive definiteness implies strict ellipticity; any lower bound λ on $\left(\frac{\partial G}{\partial Q_{ij}}\right)$ is an ellipticity constant.

Lemma 12 (Strict ellipticity) *Fix X, Y and s as in* (10)–(11). *The operator G defined by* (21) *and* (24) *with $i = 2$ is degenerate elliptic. Moreover, if $G(y, p, Q) > 0$, and there exists $\Theta > 0$ such that $Q - D^2_{yy}s(x, y) \leq \Theta I$ for all $x \in X_2(y, p, Q)$, then the ellipticity constant of G at (y, p, Q) is given by $\lambda = G(y, p, Q)/\Theta$.*

Proof. Fixing $(y, p) \in \bar{Y} \times \mathbf{R}^m$ and $m \times m$ symmetric matrices $Q' \geq Q$, degenerate ellipticity of G follows from the facts that $f \geq 0$, $X_2(y, p, Q) \subset X_2(y, p, Q')$, and $Q' - D^2_{yy}s(x, y) \geq Q - D^2_{yy}s(x, y) \geq 0$ for all $x \in X_2(y, p, Q)$.

Now suppose also $Q - D^2_{yy}s(x, y) \leq \Theta I < \infty$ for all $x \in X_2(y, p, Q)$, so that the product $((Q - D^2_{yy}s(x, y))^{-1} - \Theta^{-1}I)(Q' - Q)$ of non-negative definite matrices has all non-negative eigenvalues, and therefore,

$$\mathrm{tr}[(Q - D^2_{yy}s(x, y))^{-1}(Q' - Q)] \geq \Theta^{-1}\mathrm{tr}[Q' - Q]$$

for all $Q' \geq Q$. From here, letting $\lambda_i \geq 0$ denote the eigenvalues of $(Q - D^2_{yy}s(x, y))^{-1}(Q' - Q)$ we deduce

$$\begin{aligned} \det[I + (Q - D^2_{yy}s)^{-1}(Q' - Q)] &= \Pi^n_{i=1}(1 + \lambda_i) \\ &\geq 1 + \sum_{i=1}^n \lambda_i \\ &= 1 + \mathrm{tr}[(Q - D^2_{yy}s(x, y))^{-1}(Q' - Q)] \\ &\geq 1 + \Theta^{-1}\mathrm{tr}[Q' - Q]. \end{aligned}$$

This can be integrated against $\det[Q - D^2_{yy}s] f d\mathcal{H}^{m-n} / \det[D^2_{xy}s D^2_{xy}s^T]$ over $X_2(y, p, Q)$ to find

$$\frac{G(y, p, Q')}{G(y, p, Q)} \geq 1 + \Theta^{-1}\mathrm{tr}[Q' - Q].$$

as desired. ■

Theorem 13 (Bootstrapping regularity using Schauder theory) *Fix $0 < \alpha < 1$, an integer $k \geq 2$, and X, Y and s as in* (10)–(11). *If $g > \epsilon > 0$ on some smooth domain Y' compactly contained in $Y \subset \mathbf{R}^n$ where $v \in C^{k,\alpha}(Y')$, and $G - g \in C^{k-1,\alpha}$ in a neighbourhood N of the 2-jet of v over Y', then* (23)–(24) *with $i = 2$ implies $v \in C^{k+1,\alpha}(Y')$.*

Proof. Since $v \in C^{2,\alpha}(Y')$, (23) holds in the classical sense. If $k \geq 3$, we can differentiate the equation in (say) the $\hat{e}_k$ direction to obtain a linear second-order elliptic equation

$$a^{ij}(y)D^2_{ij}w + b^i(y)D_i w = d(y) \tag{43}$$

for $w = \partial v/\partial y^k$ whose coefficients

$$\begin{aligned} a^{ij}(y) &:= \frac{\partial G}{\partial Q_{ij}}\bigg|_{(y,Dv(y),D^2v(y))} \\ b^i(y) &:= \frac{\partial G}{\partial p_i}\bigg|_{(y,Dv(y),D^2v(y))} \end{aligned}$$

and inhomogeneity

$$d(y) = \frac{\partial g}{\partial y_k}\bigg|_y - \frac{\partial G}{\partial y_k}\bigg|_{(y,Dv(y),D^2v(y))}$$

have (i) C^{k-2,α^2} norm controlled by $\|G-g\|_{C^{k-1,\alpha}}\|v\|^{\alpha}_{C^{k,\alpha}}$ and (ii) $C^{k-2,\alpha}$ norm controlled by $\|G-g\|_{C^{k-1,\alpha}}\|v\|_{C^{k,1}}$. In case $k=2$, we shall argue below that $w \in C^{1,\alpha}$ solves (43) in the *viscosity* sense described e.g. in [12]. From Lemma 12 we see the matrix (a^{ij}) is bounded below by $\epsilon I/\|v-s\|_{C^2(X\times Y')}$; it is bounded above by $\|G\|_{C^1(N)}$. Thus the equation satisfied by w on Y' is uniformly elliptic. Since the coefficient of w vanishes in (43), the Dirichlet problem with continuous boundary data on any ball in Y' is known to admit a unique (viscosity) solution [12]; moreover, this solution is (i) C^{k,α^2}_{loc} (by e.g. Gilbarg & Trudinger Theorems 6.13 ($k=2$) or 6.17 ($k>2$). Thus we infer $v \in C^{k+1,\alpha^2}_{loc}(Y')$. Applying the same argument again starting from the improved estimates (ii) now established yields $v \in C^{k+1,\alpha}_{loc}(Y')$. At this point we have gained the desired derivative of smoothness for v; starting from a neighbourhood slightly larger than Y' yields $v \in C^{k+1,\alpha}(Y')$.

In case $k=2$, applying the finite difference operator $\Delta^h_k v(y) := [v(y+h\hat{e}_k) - v(y)]/h$ to the equation (23), the mean value theorem yields $h^*(y) \in [0,h]$ lower semicontinuous such that

$$\begin{aligned} 0 &= \Delta^h_k[G(y,Dv(y),D^2v(y)) - g(y)] \\ &= a^{ij}_h(y)D^2_{ij}w_h + b^i_h(y)D_i w_h - d_h(y). \end{aligned}$$

Here $w_h = \Delta_k^h v$ and the coefficients

$$a_h^{ij}(y) := \frac{\partial G}{\partial Q_{ij}}\bigg|_{(I+h^*(y)\Delta_k^h)(y,Dv(y),D^2v(y))}$$

$$b_h^i(y) := \frac{\partial G}{\partial p_i}\bigg|_{(I+h^*(y)\Delta_k^h)(y,Dv(y),D^2v(y))}$$

$$d_h(y) = \frac{\partial g}{\partial y_k}\bigg|_{y+h^*(y)\hat{e}_k} - \frac{\partial G}{\partial y_k}\bigg|_{(I+h^*(y)\Delta_k^h)(y,Dv(y),D^2v(y))}.$$

are measurable and converge uniformly to (a^{ij}, b^i, d) as $h \to 0$. The solutions $w_h = \Delta_k^h v \in C^{2,\alpha}$, being finite differences, converge to $\partial v/\partial y_k$ in $C^{1,\alpha}(Y')$. Lemma 6.1 and Remark 6.3 of [12] show this partial derivative $w = \partial v/\partial y_k$ must then be the required viscosity solution of the limiting equation (43). ■

Notice G_2 is degenerate elliptic even when evaluated on functions which are not s-convex.

6 On smoothness of the nonlinear operators G_i

The preceding section illustrates how one can bootstrap from $v \in C^{2,\alpha}$ to higher regularity, assuming smoothness of the nonlinear elliptic operator G_2. We now turn our attention to verifying the assumed smoothness of G_2, at least (for simplicity) on the set where $G_2 = G_1$. Our main result is Theorem 14. For $(n, i) = (1, 2)$, the initial smoothness assumed of v is addressed in Section 8, but we establish neither the initial smoothness nor the uniform convexity of v for higher dimensional targets; as we have noted, these remain interesting open questions.

Our joint work with Chiappori [10] establishes regularity of G_1 (and v) when $n = 1 = i$; in this section, we focus on this smoothness for higher dimensional targets $n > 1$. We note that connectedness of almost every level set $X_2(y, Dv(y), D^2v(y))$, plus the C^2-smoothness of v hypothesized in Theorem 13 of the last section, and C^2-smoothness of $u = v^s$, implies that $G_1 = G_2$ by Theorem 6, so in many cases of interest it is enough to address smoothness of G_1. When $n = 1$, nested examples in our earlier work with Chiappori [10] [9] [8] satisfy the $i = 1$ version of (23), implying $G_1(y, Dv(y), D^2v(y)) = G_2(y, Dv(y), D^2v(y))$ for almost every y. Other v's for which $G_1 = G_2$ with $n > 1$ arise in Example 15. Note however that when $G_1(y, Dv(y), D^2v(y)) \neq G_2(y, Dv(y), D^2v(y))$, as can happen, for instance, when the $X_2(y, Dv(y), D^2v(y))$ are disconnected, the results in this section by themselves yield little information about G_2.

For technical reasons it is convenient to assume that $y \mapsto s(x, y)$ is uniformly convex, throughout this section. Note that this assumption can always be achieved by adding a sufficiently convex function of y to s. Henceforth we'll also require smoothness of s to extend to $\bar{X}$, to impose transversality conditions at its boundary. Given bounded open sets $Y', P' \subseteq \mathbf{R}^n$, throughout this section we therefore set $X' = \cup_{(y,p)\in Y'\times P'} X_1(y,p)$ and augment (10)–(11) by assuming:

$$\text{Assume } s \in C^2(\bar{X}\times\bar{Y}) \text{ in (10)–(11)}, \tag{44}$$

$$X \subset \mathbf{R}^m \text{ is bounded, } \partial X \in C^1, \tag{45}$$

$$\text{there exists } C > 0 \text{ such that } D^2_{yy} s(x,y) \geq CI \text{ on } X' \times Y', \tag{46}$$

$$\text{and (48)–(51) are all finite and positive.} \tag{47}$$

We define a smoothed version

$$\tilde{G}_1(y,p,Q) := \int_{X_i(y,p,Q)} \frac{\det(Q - D^2_{yy}s(x,y))}{\sqrt{\det D^2_{xy}s(x,y)(D^2_{xy}s(x,y))^T}} f(x)\, d\mathcal{H}^{m-n}(x),$$

of G_1, which is distinguished from the original only by the removal of the absolute value signs from the determinant in (24). On the set of (y,p,Q) where $G_1 = G_2$, the definition of X_2 makes these absolute value signs redundant, hence $\tilde{G}_1 = G_1$ on this set.

Theorem 14 (Smoothness of $\tilde{G}_1$) *Let $r \geq 1$ and assume (44)–(47). Then $||\tilde{G}_1||_{C^{r,1}(Y'\times P')}$ is controlled by $||f||_{C^{r,1}(\bar{X}')}$, $||D_y s||_{C^{r+1,1}(Y'\times\bar{X}')}$, $||\hat{n}_X||_{C^{r-1,1}(\partial X\cap\bar{X}')}$ and*

$$\inf_{(x,y)\in X'\times Y'} \min_{v\in\mathbf{R}^n, |v|=1} |D^2_{xy}s(x,y)\cdot v| \quad \text{(non-degeneracy)}, \tag{48}$$

$$\inf_{(x,y,p)\in(\partial X\cap\bar{X}')\times Y'\times P'} |(\hat{n}_X)_{T_x X_1(y,p)}| \quad \text{(transversality)}, \tag{49}$$

$$\sup_{(y,p)\in Y'\times P'} \mathcal{H}^{m-n}(X_1(y,p)) \quad \text{(size of level sets)}, \tag{50}$$

$$\text{and } \sup_{(y,p)\in Y'\times P'} \mathcal{H}^{m-n-1}(\overline{X_1(y,p)} \cap \partial X) \quad \text{(boundary intersections)} \tag{51}$$

Here $(\hat{n}_X)_{T_x X_1(y,p)}$ denotes the projection of the outward unit normal $\hat{n}_X$ to X onto the tangent space $T_x X_1(y,p)$.

Example 15 (Bilinear cost to an embedded target) *Let $s(x,y) = x\cdot H(y)$, where $X \subseteq \mathbf{R}^m$, $Y \subseteq \mathbf{R}^n$ and $H : \bar{Y} \to \mathbf{R}^m$ parametrizes a smooth n-dimensional submanifold. Then the convex function $u(x) = \max_{y\in\bar{Y}} x\cdot H(y)$ is s-convex with $v(y) = u^{\tilde{s}}(y) = 0$. In this case $X_1(y, Dv(y)) \subset X$ is given by the nullspace of $DH(y)$, and coincides with $X_2(y, Dv(y), D^2v(y))$ if $y \mapsto s(x,y)$ is concave for each $x \in X_1(y, Dv(y))$. More generally, if $\|v\|_{C^2(\bar{Y})} \leq \epsilon$, then $X_1(y, Dv(y)) = X_2(y, Dv(y), D^2v(y))$ provided*

$xD^2H(y) \leq -\epsilon I$ for each $x \in X_1(y, Dv(y))$. In either case $G_1 = G_2$ at $(y, Dv(y), D^2v(y))$ in (24)*.*

One can easily verify the other conditions in Theorem 14; noting that $D^2_{xy}s = DH(y)$, we see the nondegeneracy condition holds (since the parameterization H admits a smooth inverse on its image). Since $X_1(y, Dv(y))$ is the intersection of X with an $m-n$ dimensional affine subspace passing near the origin and orthogonal to $T_{H(y)}H(Y)$, it is not hard to check whether a given domain X satisfies the transversality, size of level sets and boundary intersections required by Theorem 14.

Remark 16 (Comparing these hypotheses to our earlier work) *We expect the preceding theorem (and similarly Theorem 23) to remain true when the hypothesis $\partial X \in C^1$ is replaced by $X' \cap \partial X \in C^1$, or when $r = 0$, as in [10], provided X assumed to have finite perimeter. However, apart from Corollary 29, Section 6 and 7 address only the smoothness of G_1 and not of v, so we won't need the lower bounds on the size of X_1 or the density of f that were required in Theorem 7.1 of [10] until Corollary 29 (and in Section 8). Of course, hypothesis* (47) *remains crucial. For example if there exist $(x', y') \in X' \times Y'$ and $0 \neq v' \in \mathbf{R}^n$ with $D^2_{xy}s(x', y')v' = 0$ then $X_1(y, p)$ may fatten (increase dimension) at $(y', p') = (y', D_y s(x', y'))$; similarly if $x' \in \partial X$ and $(\hat{n}_X)_{T_x X_1(y', p')} = 0$ then $\mathcal{H}^{m-n}[\bar{X}_1(y, p)]$ may jump discontinuously at (y', p'), due to its non-transerval intersection with ∂X. In either case, smoothness of G_1 would be expected to fail at (y', p').*

Before proving the theorem, we develop some notation and establish a few preliminary lemmas.

For $i \in \{1, \ldots, n\}$, the set $X^i_{\leq}(y, p) := \{x \mid s_{y_i} \leq p_i, s_{y_j} = p_j \forall j \neq i\}$ is a submanifold of X whose relative boundary is given by $X_1(y, p)$. Then $X^i_{\leq}(y, p) \subseteq X^i(y, p) := \{x \mid s_{y_j} = p_j \forall j \neq i\}$, while with an analogous definition $X^i_{=}(y, p)$ coincides with $X_1(y, p)$.

Nondegeneracy of s makes $X_1(y, p)$ a codimension one submanifold of the codimension $n-1$ submanifold $X^i(y, p)$ of X. By the implicit function theorem, these submanifolds are each one derivative less smooth than s.

Lemma 17 (Submanifold transversality) *The submanifold $\partial \bar{X}^i = \bar{X}^i \cap \partial X$ and submanifold-with-boundary $\bar{X}_1 \subset \bar{X}^i$ intersect transversally in $\bar{X}^i$.*

Proof. The proof is straightforward linear algebra. Since $X_1 \subset X^i$, the transversal intersection of $\bar{X}_1$ and ∂X in $\mathbf{R}^m$ guaranteed by positivity of (49) implies transversal intersection of $\bar{X}^i$ and ∂X, and so $T_x(\partial \bar{X}^i) =$

$T_x(\partial X) \cap T_x(\bar{X}^i)$, at each point of intersection $x \in \bar{X}_1 \cap \partial X$. We then need to show

$$[T_x(\partial X) \cap T_x(X^i)] + T_x X_1 = T_x X^i.$$

The containment $[T_x(\partial X) \cap T_x(X^i)] + T_x X_1 \subseteq T_x X^i$ is immediate, as each of the summands is contained in $T_x X^i$. On the other hand, if $p \in T_x X^i \subset \mathbf{R}^m = T_x(\partial X) + T_x X_1$ (by transversality), we write $p = p_1 + p_\partial$, with $p_1 \in T_x X_1 \subseteq T_x X^i$ and $p_\partial \in T_x(\partial X)$. But then, since $p_\partial = p - p_1$, both $p, p_1 \in T_x X^i$, and $T_x X^i$ is a vector space, we must have $p_\partial \in T_x X^i$, so $p_\partial \in [T_x(\partial X) \cap T_x(X^i)]$, implying the containment $T_x X^i \subseteq [T_x(\partial X) \cap T_x(X^i)] + T_x X_1$. ■

Given $f \in L^\infty$, we note that, keeping y and p_j for all $j \neq i$ fixed and working on the $m-n+1$ dimensional submanifold $X^i(y,p)$ of X allows us to use Lemma 5.1 of [10] to conclude that

$$\Phi^i(y,p) := \int_{X^i_{\leq}(y,p)} f(x,y) d\mathcal{H}^{m-n+1}(x)$$

has a Lipschitz dependence on p_i, with

$$\frac{\partial \Phi^i}{\partial p_i}(y,p) = \int_{X^i_{=}(y,p)} \frac{f(x,y)}{|D_{X^i} s_{y_i}|} d\mathcal{H}^{m-n}(x) \qquad \text{[a.e.]}, \tag{52}$$

where $D_{X^i} s_{y_i}$ is the differential of s_{y_i} along the submanifold X^i, nonzero by the nondegeneracy assumption:

Lemma 18 (Restriction non-degeneracy) *The differential $D_{X^i} s_{y_i}$ of s_{y_i} along the manifold X^i satisfies*

$$|D_{X^i} s_{y_i}| \geq \min_{v \in \mathbf{R}^n, |v|=1} |D^2_{xy} s \cdot v|.$$

Proof. Note that $D_{X^i} s_{y_i}$ is $D^2_{xy_i} s$, minus its projection onto the span of the other $D^2_{xy_j} s$, and so

$$\begin{aligned}
|D_{X^i} s_{y_i}| &= \min_{v_1, v_2, \dots v_{i-1}, v_{i+1} \dots v_n} |D^2_{xy_i} s - \sum_{j \neq i} v_j D^2_{xy_j} s| \\
&= \min_{v=(v_1,\dots,v_n) \in \mathbf{R}^n, v_i = 1} |D^2_{xy} s \cdot v| \\
&\geq \min_{v \in \mathbf{R}^n, |v|=1} |D^2_{xy} s \cdot v|.
\end{aligned}$$

■

Note that the outward unit normal to $X^i_{\leq}(y,p)$ in $X^i(y,p)$ is

$$\hat{n}^i := \frac{D_{X^i} s_{y_i}}{|D_{X^i} s_{y_i}|}$$

and the normal velocity of $X_1(y,p)$ in $X^i(y,p)$ as p_i is varied is

$$V^i = \frac{\hat{n}^i}{|D_{X^i} s_{y_i}|}.$$

Here $D_{X^i} s_{y_i} = D_{X^i(y,p)} s_{y_i}(x,y)$, and objects defined in terms of it, such as, $\hat{n}^i = \hat{n}^i(x,y,p)$ are defined only for $x \in X^i(y,p)$. We will denote

$$D_{X^i} s_{y_i}(x,y) := D_{X^i(y,p)} s_{y_i}(x,y)\Big|_{p=D_y s(x,y)}$$

which is defined globally on $X' \times Y'$. Expressions such as $\hat{n}^i(x,y)$ are defined analogously.

Similarly, the outward unit normal to $\left(\overline{X^i_{\leq}(y,p)}\right) \cap \partial X$ in $\left(\overline{X^i(y,p)}\right) \cap \partial X$ will be denoted $\hat{n}^i_\partial$. Denote by $\hat{n}^i_X = \frac{(\hat{n}_X)_{T_x X^i}}{|(\hat{n}_X)_{T_x X^i}|}$ the (renormalized) projection of $\hat{n}_X$ onto $T_x X^i$, which is well-defined by tranversality (note $|(\hat{n}_X)_{T_x X^i}| \geq |(\hat{n}_X)_{T_x X_1}|$). This is the outward unit normal to $\overline{X^i(y,p)} \cap X$ in $\overline{X^i(y,p)}$.

We have that

$$\hat{n}^i_\partial = \frac{\hat{n}^i - (\hat{n}^i_X \cdot \hat{n}^i)\hat{n}^i_X}{\sqrt{1 - (\hat{n}^i_X \cdot \hat{n}^i)^2}}.$$

Note that

$$V^i_\partial := \frac{|V^i|}{\sqrt{1 - (\hat{n}^i_X \cdot \hat{n}^i)^2}} \hat{n}^i_\partial$$

represents the normal velocity of $\left(\overline{X_1(y,p)}\right) \cap \partial X$ in $\left(\overline{X^i(y,p)}\right) \cap \partial X$. The denominator is bounded away from 0 by the transversality assumption.

Analogously to (52), working in the $m-n$ dimensional submanifold ∂X^i with y and each p_j for $j \neq i$ fixed, Lemma 5.1 in [10] implies for $g \in L^\infty$ that $\Psi^i(y,p) := \int_{X^i_{\leq}(y,p) \cap \partial X} g(x,y) d\mathcal{H}^{m-n}(x)$ has Lipschitz dependence on p_i, and

$$\frac{\partial \Psi^i}{\partial p_i}(y,p) = \int_{\overline{X^i_{=}(y,p)} \cap \partial X} g(x,y) |V^i_\partial| d\mathcal{H}^{m-n-1}(x) \qquad \text{[a.e.]}. \tag{53}$$

Lemma 19 (Derivative bounds along submanifolds) *Given functions $a : X' \times Y' \to \mathbf{R}$, $b : \partial X \times Y \to \mathbf{R}$ and $v : \bar{X}' \times Y' \to TX$ and $w : (\overline{X'} \cap \partial X) \times Y \to T\partial X$ such that $v(x,y) \in T_x X^i(x, D_y s(x,y))$ and $w(x,y) \in T_x(\bar{X}^i(x, D_y s(x,y)) \cap \partial X)$ everywhere, we have:*

1. $||D_{X^i(y,D_ys(x,y))}a(x,y)||_{C^{k,1}(X'\times Y')}$ *is controlled by* $||a||_{C^{k+1,1}(X'\times Y')}$, $||D_ys||_{C^{k,1}(X'\times Y')}$, *and nondegeneracy.*

2. $||\nabla_{X^i(x,D_ys(x,y))}\cdot v||_{C^{k,1}(X'\times Y')}$ *is controlled by* $||v||_{C^{k+1,1}(X'\times Y')}$.

3. $||D_{\overline{X^i(y,D_ys(x,y))}\cap\partial X}b(x,y)||_{C^{k,1}((\overline{X'}\cap\partial X)\times Y')}$ *is controlled by* $||b||_{C^{k+1,1}((\overline{X'}\cap\partial X)\times Y')}$, $||D_ys||_{C^{k,1}((\overline{X'}\cap\partial X)\times Y')}$, *nondegeneracy, tranversality and* $||\hat{n}_X||_{C^{k,1}(\overline{X'}\cap\partial X)}$

4. $||\nabla_{\overline{X^i(x,D_ys(x,y))}\cap\partial X}\cdot w||_{C^{k,1}((\overline{X'}\cap\partial X)\times Y')}$ *is controlled by* $||w||_{C^{k+1,1}((\overline{X'}\cap\partial X)\times Y')}$ *and* $||\hat{n}_X||_{C^{k+1,1}(\overline{X'}\cap\partial X)}$.

Proof. First we prove the first implication. Note that $D_{X^i(y,D_xs(x,y))}a(x,y)$ is equal to $D_xa(x,y)$, minus it's projection onto the span of the $D^2_{xy_j}s$ for $j\neq i$; that is

$$D_{X^i(y,D_ys(x,y))}a(x,y)=D_xa(x,y)-\sum_{j=1}^{n-1}[D_xa(x,y)\cdot e_j(x,y)]e_j(x,y)$$

where the $e_j(x,y)$ are an orthonormal basis for the span of $\{D^2_{xy_j}s(x,y)\}_{j\neq i}$. The e_j can then be written explicitly as functions of the $D^2_{xy_j}s(x,y)$, using for instance the Gram-Schmidt procedure; the definition of e_j involves projections onto the $e_{\bar{j}}$ for $\bar{j}<j$, which are controlled by nondegeneracy.

The second implication follows by noting that the divergence $\nabla_{X^i(x,D_ys(x,y))}\cdot v(x,y)$ coincides with $\nabla_X\cdot v(x,y)$.

The proof of the third implication is identical to that of the first, except that we subtract the projection onto the span of $\{D^2_{xy_j}s(x,y)\}_{j\neq i}\cup\{\hat{n}_X\}$; this is controlled by nondegeneracy and transversality, as well as the smoothness of these basis vectors.

Finally, the proof of the fourth assertion is almost the same as the second; the divergence coincides with $\nabla_{\partial X}\cdot w(x,y)$, which involves first derivatives of the metric, and hence of $\hat{n}_X$, as in the remarks preceding Lemma 7.2 in [10]. ■

Now, we define $s^*(x,p)$ to be the Legendre transformation of s with respect to the y variable:

$$s^*(x,p)=\sup_y(y\cdot p-s(x,y)).$$

Lemma 20 (Smoothness and non-degeneracy for Legendre duals) *The transformation s^* inherits the same smoothness as s, and is non-degenerate. Further, its non-degeneracy is quantitatively controlled by the non-degeneracy and C^2 norm of s:*

$$\inf_{|u|=1} |D^2_{xp}s^*(x,p)\cdot u| \geq \frac{\inf_{|v|=1} |D^2_{xy}s(x,y)\cdot v|}{||D^2_{yy}s(x,y)||}$$

for $p = D_y s(x,y)$.

Proof. Uniform convexity implies that s^* is continuously twice differentiable with respect to p. The implicit function theorem combined with the identity $D_p s^*(x, D_y s(x,y)) = y$ implies the smoothness of s^*. In particular, differentiating with respect to x yields

$$D^2_{xp}s^*(x, D_y s(x,y)) = -D^2_{xy}s(x,y) D^2_{pp}s^*(x, D_y s(x,y))$$

and so invertibility of $D^2_{pp}s^*$ and nondegeneracy of s imply nondegeneracy of s^*, and we have, for $|u| = 1$,

$$\begin{aligned} D^2_{xp}s^*(x, D_y s(x,y))\cdot u &= -D^2_{xy}s(x,y) D^2_{pp}s^*(x, D_y s(x,y))\cdot u \\ &= -D^2_{xy}s(x,y)\frac{D^2_{pp}s^*(x, D_y s(x,y))\cdot u}{|D^2_{pp}s^*(x, D_y s(x,y))\cdot u|}|D^2_{pp}s^*(x, D_y s(x,y))\cdot u|. \end{aligned}$$

Now note that setting $v = D^2_{pp}s^*(x, D_y s(x,y))\cdot u = [D^2_{yy}s(x,y)]^{-1}\cdot u$, so that $1 = |u| = |D^2_{yy}s(x,y)\cdot v| \leq ||D^2_{yy}s(x,y)||\cdot|v|$. Therefore

$$|v| \geq \frac{1}{||D^2_{yy}s(x,y)||}$$

and the result follows. ■

Now, we can identify the set $X_1(y,p) = \{x \mid D_p s^*(x,p) = y\}$. We then define $X^{*i}_{\leq}(y,p)$, $X^{*i}(y,p)$ and Φ^{*i} analogously to above, and compute

$$\frac{\partial \Phi^{*i}}{\partial y_i} = \int_{X^{*i}_{=}(y,p)} \frac{f(x,y)}{|D_{X^{*i}}s^*_{p_i}|} d\mathcal{H}^{m-n-1}(x) + \int_{X^{*i}_{\leq}(y,p)} \frac{\partial f}{\partial y_i}(x,y) d\mathcal{H}^{m-n}(x) \tag{54}$$

for a.e. (y,p) as long as f and f_{y_i} are Lipschitz.

Analogs of Lemmas 17, 18 and 19 when $s(x,y)$ is replaced by $s^*(x,p)$ then follow immediately. We note that

$$D_{X^{i*}}s^*_{p_i}(x,y) := D_{X^{i*}(y,p)}s^*_{p_i}(x,p)\Big|_{p=D_y s(x,y)}$$

is defined throughout $X' \times Y'$. We define $\hat{n}^{*i}$, V^{*i}, $\hat{n}^{*i}_{\partial}$, $\hat{n}^{*i}_X$, V^{*i} analogously to their un-starred counterparts and note that upon evaluating at $p = D_y s(x,y)$, each can be considered a function on $X' \times Y'$ or $\partial X' \times Y'$.

Lemma 21 (Flux derivatives through moving surfaces) *Use* $a : \overline{X' \times Y' \times P'} \to \mathbf{R}$ *Lipschitz to define* $\Phi(y,p) := \int_{X_1(y,p)} a(x,y,p) d\mathcal{H}^{m-n}(x)$ *and* $\Psi(y,p) := \int_{\overline{X_1(y,p)} \cap \partial X} a(x,y,p) d\mathcal{H}^{m-n-1}(x)$*. Then* Φ *and* Ψ *are Lipschitz with partial derivatives given almost everywhere by:*

$$\begin{aligned}
\frac{\partial \Phi(y,p)}{\partial p_i} = &\int_{X_1(y,p)} \Big[\nabla_{X^i(y,p)} \cdot \Big(a(x,y,p)\frac{D_{X^i} s_{y_i}}{|D_{X^i} s_{y_i}|}\Big) V^i \cdot \hat{n}^i\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n}(x) \\
&- \int_{\left(\overline{X_1(y,p)}\right) \cap \partial X} \Big[\Big(a(x,y,p)\frac{D_{X^i} s_{y_i}}{|D_{X^i} s_{y_i}|}\Big) \cdot \hat{n}^i_X V^i_\partial \cdot \hat{n}^i_\partial\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n-1}(x) \\
&+ \int_{X_1(y,p)} \Big[\frac{\partial a(x,y,p)}{\partial p^i}\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n}(x), \qquad (55)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Psi(y,p)}{\partial p_i} = &\\
&\int_{\overline{X_1(y,p)} \cap \partial X} \Big[\nabla_{\overline{X^i(y,p)} \cap \partial X} \cdot \Big(a(x,y,p)\frac{D_{\overline{X_i(y,p)} \cap \partial X} s_{y_i}}{|D_{\overline{X_i(y,p)} \cap \partial X} s_{y_i}|}\Big) V^i_\partial \cdot \hat{n}^i_\partial\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n}(x) \\
&+ \int_{\overline{X_1(y,p)} \cap \partial X} \Big[\frac{\partial a(x,y,p)}{\partial p^i}\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n-1}(x), \qquad (56)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Phi(y,p)}{\partial y_i} = &\int_{X_1(y,p)} \Big[\nabla_{X^{*i}(y,p)} \cdot \Big(a(x,y,p)\frac{D_{X^{*i}} s^*_{p_i}}{|D_{X^{*i}} s^*_{p_i}|}\Big) V^{*i} \cdot \hat{n}^{*i}\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n}(x) \\
&- \int_{\left(\overline{X_1(y,p)}\right) \cap \partial X} \Big[\Big(a(x,y,p)\frac{D_{X^{*i}} s^*_{p_i}}{|D_{X^{*i}} s_{p_i}|}\Big) \cdot \hat{n}^{*i}_X V^{*i}_\partial \cdot \hat{n}^{*i}_\partial\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n-1}(x) \\
&+ \int_{X_1(y,p)} \Big[\frac{\partial a(x,y,p)}{\partial y^i}\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n}(x), \qquad (57)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \Psi(y,p)}{\partial y_i} = &\\
&\int_{\overline{X_1(y,p)} \cap \partial X} \Big[\nabla_{\overline{X^{*i}(y,p)} \cap \partial X} \cdot \Big(a(x,y,p)\frac{D_{\overline{X^{*i}(y,p)} \cap \partial X} s^*_{p_i}}{|D_{\overline{X^{*i}(y,p)} \cap \partial X} s^*_{p_i}|}\Big) V^{*i}_\partial \cdot \hat{n}^{*i\partial}\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n}(x) \\
&+ \int_{\overline{X_1(y,p)} \cap \partial X} \Big[\frac{\partial a(x,y,p)}{\partial y^i}\Big]_{p=D_y s(x,y)} d\mathcal{H}^{m-n-1}(x). \qquad (58)
\end{aligned}$$

Proof. We begin by establishing the formulas assuming $a \in C^{1,1}(\overline{X' \times Y' \times P'})$.

Using the generalized divergence theorem [24, Proposition 5.8] we have, for fixed $\bar{p}^i < p^i$, denoting by $p^{(i)}$ the vector whose i-th entry is $\bar{p}^i$ and all other entries are equal to those of p,

$$\begin{aligned}
\Phi(y,p) = & \int_{X_1(y,p)} \Big(a(x,y,p) \frac{D_{X^i} s_{y_i}}{|D_{X^i} s_{y_i}|} \Big) \cdot \hat{n}^i d\mathcal{H}^{m-n}(x) \\
= & \int_{X^i_{\leq}(y,p) \setminus X^i_{\leq}(y,p^{(i)})} \nabla_{X^i(y,p)} \cdot \Big(a(x,y,p) \frac{D_{X^i} s_{y_i}}{|D_{X^i} s_{y_i}|} \Big) d\mathcal{H}^{m-n+1}(x) \\
& + \int_{X^i_{=}(y,p^{(i)})} \Big(a(x,y,p) \frac{D_{X^i} s_{y_i}}{|D_{X^i} s_{y_i}|} \Big) \cdot \hat{n}^i(x,y) d\mathcal{H}^{m-n}(x) \\
& - \int_{\Big(X^i_{\leq}(y,p) \setminus X^i_{\leq}(y,p^{(i)}) \Big) \cap \partial X} \Big(a(x,y,p) \frac{D_{X^i} s_{y_i}}{|D_{X^i} s_{y_i}|} \Big) \cdot \hat{n}^i_X(x,y) d\mathcal{H}^{m-n}(x).
\end{aligned}$$

Noting that the integrands in the first and third terms above are bounded, one can then combine the chain rule with (52) and (53) to differentiate with respect to p_i, getting

$$\begin{aligned}
\frac{\partial \Phi(y,p)}{\partial p_i} = & \int_{X^i_{=}(y,p)} \nabla_{X^i(y,p)} \cdot \Big(a(x,y,p) \frac{D_{X^i} s_{y_i}}{|D_{X^i} s_{y_i}|} \Big) V^i \cdot \hat{n}^i d\mathcal{H}^{m-n}(x) \\
& - \int_{\Big(\overline{X^i_{=}(y,p)} \Big) \cap \partial X} \Big(a(x,y,p) \frac{D_{X^i} s_{y_i}}{|D_{X^i} s_{y_i}|} \Big) \cdot \hat{n}^i_X V^i_\partial \cdot \hat{n}^i_\partial d\mathcal{H}^{m-n-1}(x) \\
& + \int_{X_1(y,p)} \frac{\partial a(x,y,p)}{\partial p^i} d\mathcal{H}^{m-n}(x).
\end{aligned}$$

Finally, notice that one may substitute $p = D_y s(x,y)$ in each integrand, as each region of integration is contained in $\overline{X_1(y,p)}$, to establish (55) for $a \in C^{1,1}$.

Now, note that the formula (55) for $\frac{\partial \Phi(y,p)}{\partial p_i}$ is controlled by $||a||_{C^{0,1}}$ (that is, it does not depend on $||a||_{C^{1,1}}$). For a merely Lipschitz, we can therefore choose a sequence $a_n \in C^{1,1}$ converging to a in the $C^{0,1}$ norm; passing to the limit implies that $||\Phi||_{C^{0,1}(\overline{Y' \times P'})}$ is controlled by $||a||_{C^{0,1}}$, and, using the dominated convergence theorem, one obtains the desired formula.

A similar argument applies to the boundary integral terms to produce the desired formula (56) for $\frac{\partial \Psi(y,p)}{\partial p_i}$, while essentially identical arguments apply to the y derivatives, yielding (57) and (58). ■

Corollary 22 (Iterated derivative bounds) *The operators*

$$A_{p_i} : (a,b) \mapsto (a^i_p, b^i_p) \quad \textit{and} \quad A_{y_i} : (a,b) \mapsto (a^i_y, b^i_y),$$

given by

$$
\begin{aligned}
a_p^i &:= \Big[\nabla_{X^i(y,p)} \cdot \Big(a(x,y)\frac{D_{X^i}s_{y_i}}{|D_{X^i}s_{y_i}|}\Big)V^i \cdot \hat{n}^i\Big]_{p=D_y s(x,y)},\\
b_p^i &:= \Big[\Big(a(x,y)\frac{D_{X^i}s_{y_i}}{|D_{X^i}s_{y_i}|}\Big) \cdot \hat{n}_X^i V_\partial^i \cdot \hat{n}_\partial^i\\
&\qquad +\nabla_{\overline{X^i(y,p)}\cap\partial X} \cdot \Big(b(x,y)\frac{D_{\overline{X^i(y,p)}\cap\partial X}s_{y_i}}{|D_{\overline{X^i(y,p)}\cap\partial X}s_{y_i}|}\Big)V_\partial^i \cdot \hat{n}_\partial^i\Big]_{p=D_y s(x,y)},\\
a_y^i &:= \Big[\nabla_{X^{*i}(y,p)} \cdot \Big(a(x,y)\frac{D_{X^{*i}}s^*_{p_i}}{|D_{X^{*i}}s^*_{p_i}|}\Big)V^{*i} \cdot \hat{n}^{*i} + \frac{\partial a(x,y)}{\partial y^i}\Big]_{p=D_y s(x,y)}, \textit{and}\\
b_y^i &:= \Big[\frac{\partial b(x,y)}{\partial y^i} + \Big(a(x,y)\frac{D_{X^{*i}}s^*_{p_i}}{|D_{X^{*i}}s^*_{p_i}|}\Big) \cdot \hat{n}_X^{*i} V_\partial^{*i} \cdot \hat{n}_\partial^{*i}\\
&\qquad +\nabla_{\overline{X^{*i}(y,p)}\cap\partial X} \cdot \Big(b(x,y)\frac{D_{\overline{X^{*i}(y,p)}\cap\partial X}s^*_{p_i}}{|D_{\overline{X^{*i}(y,p)}\cap\partial X}s^*_{p_i}|}\Big)V_\partial^{*i} \cdot \hat{n}^{*i\partial}\Big]_{p=D_y s(x,y)},
\end{aligned}
$$

define mappings $A_{p_i} : B_k \to B_{k-1}$ *and* $A_{y_i} : B_k \to B_{k-1}$ *between Banach spaces defined by*

$$B_k := C^{k,1}(X' \times Y') \oplus C^{k,1}([X' \cap \partial X] \times Y')$$

with norms

$$
\begin{aligned}
||A_{p_i}|| \le{}& ||\frac{1}{|D_{X^i}s_{y_i}|}||_{C^{k-1,1}(X'\times Y')}||\hat{n}^i||_{C^{k-1,1}(X'\times Y')}\\
&+ ||\hat{n}^i||_{C^{k-1,1}(X'\times Y')}||\hat{n}_X^i||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}||V_\partial^i \cdot \hat{n}_\partial^i||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}\\
&+ ||\frac{D_{\overline{X^i(y,p)}\cap\partial X}s_{y_i}}{|D_{\overline{X^i(y,p)}\cap\partial X}s_{y_i}|}||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}||V_\partial^i||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}||\hat{n}_\partial^i||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}
\end{aligned}
$$

and

$$
\begin{aligned}
||A_{y_i}|| \le{}& ||\frac{1}{|D_{X^{*i}}s^*_{p_i}|}||_{C^{k-1,1}(X'\times Y')}||\hat{n}^{*i}||_{C^{k-1,1}(X'\times Y')} + 1\\
&+||\hat{n}^{*i}||_{C^{k-1,1}(X'\times Y')}||\hat{n}_X^{*i}||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}||V_\partial^{*i} \cdot \hat{n}_\partial^{*i}||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}\\
&+||\frac{D_{\overline{X^{*i}(y,p)}\cap\partial X}s^*_{p_i}}{|D_{\overline{X^{*i}(y,p)}\cap\partial X}s^*_{p_i}|}||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}||V_\partial^{*i}||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}||\hat{n}_\partial^{*i}||_{C^{k-1,1}((\overline{X'}\cap\partial X)'\times Y')}
\end{aligned}
$$

controlled by $||D_y s||_{C^{k,1}}$, $||\hat{n}_X||_{C^{k,1}}$, *non-degeneracy and transversality.*

Furthermore, restricted to the subspace $C^{k,1}(X'\times Y')\oplus\{0\}$, *the norms*

$$
\begin{aligned}
||A_{p_i}||_{C^{k,1}(X'\times Y')\oplus\{0\}\to B_{k-1}} \le{}& ||\frac{1}{|D_{X^i}s_{y_i}|}||_{C^{k-1,1}(X'\times Y')}||\hat{n}^i||_{C^{k-1,1}(X'\times Y')}\\
&+ ||\hat{n}^i||_{C^{k-1,1}(X'\times Y')}||\hat{n}_X^i||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}||V_\partial^i \cdot \hat{n}_\partial^i||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}
\end{aligned}
$$

and

$$||A_{y_i}||_{C^{k,1}(X'\times Y')\oplus\{0\}\to B_{k-1}} \leq ||\frac{1}{|D_{X^{*i}}s^*_{p_i}|}||_{C^{k-1,1}(X'\times Y')}||\hat{n}^{*i}||_{C^{k-1,1}(X'\times Y')}+1$$
$$+||\hat{n}^{*i}||_{C^{k-1,1}(X'\times Y')}||\hat{n}^{*i}_X||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}||V^{*i}_\partial\cdot\hat{n}^{*i}_\partial||_{C^{k-1,1}((\overline{X'}\cap\partial X)\times Y')}$$

are controlled by $||D_y s||_{C^{k,1}}$*,* $||\hat{n}_X||_{C^{k-1,1}}$*, non-degeneracy and transversality.*

Proof. The estimates on the norms follow by simple calculations. The control on the various quantities in the estimates relies on Lemmas 18, 19, 20, and closure of the Hölder spaces $C^{k-1,1}$ under composition. ■

We now prove the result announced at the beginning of this section:
Proof Theorem 14. First note that as Q enters the definition of of $\tilde{G}_1$ only through the integrand, whose dependence on Q is smooth, computing derivatives with respect to Q is straightforward.

Corollary 22 allows us to iterate derivatives with respect to the other variables; given multi indices $\alpha=(\alpha_1,\alpha_2,...\alpha_n)$, $\beta=(\beta_1,\beta_2,...,\beta_n)$, and $\gamma=(\gamma_1,\gamma_2,....,\gamma_{n^2})$ with $|\alpha|+|\beta|+|\gamma|=\sum_{i=1}^n\alpha_i+\sum_{i=1}^n\beta_i+\sum_{i=1}^{n^2}\gamma_i=k\leq r$, then Lemma 21 and Corollary 22 allow us to compute

$$\frac{\partial^k\tilde{G}_1}{\partial p^\alpha\partial y^\beta\partial Q^\gamma}=\int_{X_1(y,p)}a^{\alpha,\beta}d\mathcal{H}^{m-n}+\int_{\partial\bar{X}_1(y,p)\cap\partial X_1}b^{\alpha,\beta}d\mathcal{H}^{m-n-1} \tag{59}$$

where $(a^{\alpha,\beta},b^{\alpha,\beta})=A^\alpha A^\beta(\frac{\partial^{|\gamma|}h}{\partial Q^\gamma},0)\in B_{r-k}$, with $h(x,y,Q)=\frac{\det[Q-D^2_{yy}s(x,y)]}{\sqrt{\det D^2_{xy}s(x,y)(D^2_{xy}s(x,y))^T}}f(x)$ being the original integrand in the definition of $\tilde{G}_1(y,p,Q)$, and $A^\alpha=A^{\alpha_1}_{p_1}....A^{\alpha_n}_{p_n}$, $A^\beta=A^{\beta_1}_{y_1}....A^{\beta_n}_{y_n}$. Now, Corollary 22 implies that $||(a^{\alpha,\beta},b^{\alpha,\beta})||_{C^{r-k,1}}$ is controlled by $||f||_{C^{r,1}}$, $||D_y s||_{C^{r+1,1}}$, $||\hat{n}_X||_{C^{r-1,1}}$, non-degeneracy and transversality.

It then follows from (59) that $\frac{\partial^k\tilde{G}_1}{\partial p^\alpha\partial y^\beta\partial Q^\gamma}$ is controlled by the quantities listed in the statement of the present theorem for $k\leq r$, as desired. ■

7 Smoothness of the local operator G_2 for one dimensional targets

Taken together, the two preceding sections allow one to bootstrap from $C^{2,\alpha}$ to higher regularity, when $X_2=X_1$. This raises the following natural questions:

1. When X_2 and X_1 differ (in which case the results in the previous subsection do not tell us much about solutions of the $i=2$ equation), under what conditions is the elliptic operator G_2 smooth?

2. When can we confirm solutions are $C^{2,\alpha}$, allowing one to apply Theorem 13?

The goal of this section and the next is to fill these gaps for one dimensional targets, $n = 1$. In this section, we identify conditions under which G_2 is smooth. As in the previous section, where regularity of G_1 for higher dimensional targets was considered, the general strategy is to adapt the approach in [10], using the divergence theorem to convert integrals over regions to those over boundaries, and differentiating the latter using the calculus of moving boundaries. These results, combined with general ODE theory, imply that $C^{1,1}_{loc}$ solutions to the $i = 2$ equation are in fact $C^{2,1}_{loc}$; higher order regularity estimates on G_2 in turn yield higher order regularity of these solutions.

The second question above is deferred to Section 8, where we find conditions under which any almost everywhere solution to the $i = 2$ equation with the one dimensional targets is locally $C^{1,1}$; the results of the present section then imply that these solutions are smoother, depending on the degree of regularity of G_2.

Given open regions Y', P' and Q' in $\mathbf{R}$, throughout this section we set $X' = \cup_{(y,p,q)\in Y'\times P'\times Q'} \overline{X_2(y,p,q)}$ and assume:

$$\text{Assume } m \geq n = 1 \text{ in (10)–(11)}, \ s_y := D_y s \in C^2(\bar{X} \times \bar{Y}), \tag{60}$$
$$X \subset \mathbf{R}^m \text{ is bounded}, \ \partial X \in C^1, \tag{61}$$
$$D^2_{xy}s \text{ and } D^3_{xyy}s \text{ are linearly independent throughout } \overline{X'} \times \overline{Y'}, \tag{62}$$
$$\text{and (65)–(68) below are all finite and positive.} \tag{63}$$

As p is increased, the domain $W_{\leq}(y,p) := \{x \in X \mid s_y \leq p\}$ expands monotonically outward with normal velocity $w(x,y) := |D^2_{xy}s|^{-1}$ along its interface $W_= = X_1$. Its normal velocity with respect to changes in y is $-s_{yy}w$. Similarly, as q is increased $Z_{\leq}(y,q) := \{x \in X \mid s_{yy} \leq q\}$ expands monotonically outward with normal velocity $z(x,y) := |D^3_{xyy}s|^{-1}$ along its interface $Z_=(y,q) := \{x \in X \mid s_{yy} = q\}$; its normal velocity with respect to changes in y is $-s_{yyy}z$. Our linear independence assumption guarantees these velocities are finite and $W_=$ intersects $Z_=$ transversally. Notice $X_2(y,p,q) = W_=(y,p) \cap Z_{\leq}(y,q)$. Also, in the same region of interest, (63) implies that both $\overline{W_= \cap Z_{\leq}}$ and $\overline{W_{\leq} \cap Z_=}$ intersect ∂X transversally. We denote by $\hat{n}_W = wD^2_{xy}s$ and $\hat{n}_Z = zD^3_{xyy}s$ the outer normals to $W_{\leq}$ and $Z_{\leq}$ respectively, and observe that the frontier of e.g. $W_{\leq}$ moves with velocity $w/\sin\theta$ in $Z_=$, when $\hat{n}_Z \cdot \hat{n}_W = \cos\theta$.

Our main result in this section is the following theorem.

Theorem 23 (Smoothness of the ODE given by G_2) *If $n = 1$ and $r \geq 0$ is an integer then $||G_2||_{C^{r,1}(Y'\times P'\times Q')}$ is controlled by $||f||_{C^{r,1}(X')}$,*

$||s_y||_{C^{r+2,1}(Y'\times X')}$, $||\hat{n}_X||_{(C^{r-1,1}\cap C^0\cap W^{1,1})(X'\cap\partial X)}$ *and*

$$\inf_{(x,y)\in X'\times Y'} \min\{|D^2_{xy}s(x,y)|, |D^3_{xyy}s(x,y)|\} \quad \text{(non-degeneracy)}, \tag{64}$$

$$\inf_{(x,y)\in(X'\cap\partial X)\times Y'} 1-(\hat{n}_W\cdot\hat{n}_X)^2 \quad (p\text{/boundary transversality}), \tag{65}$$

$$\inf_{y\in Y',\ x\in\partial X\cap\overline{W_{\le}(y,P')}\cap\overline{Z_{=}(y,Q')}} 1-(\hat{n}_Z\cdot\hat{n}_X)^2 \quad (q\text{/boundary transversality}), \tag{66}$$

$$\inf_{y\in Y',\ x\in W_{=}(y,P')\cap Z_{=}(y,Q')} 1-(\hat{n}_W\cdot\hat{n}_Z)^2 \quad (p/q \text{ transversality}), \tag{67}$$

$$\inf_{\substack{y\in Y',\ x\in\partial X\cap\overline{W_{=}(y,P')}\cap\overline{Z_{=}(y,Q')}\\ \lambda_1^2+\lambda_2^2+\lambda_3^2=1}} |\lambda_1\hat{n}_W+\lambda_2\hat{n}_Z+\lambda_3\hat{n}_X| \quad \text{(linear independence)}, \tag{68}$$

$$\sup_{(y,p,q)\in Y'\times P'\times Q'} \mathcal{H}^{m-1}(W_{=}(y,p)\cap Z_{\le}(y,q)) \quad \text{(1st level set size)}, \tag{69}$$

$$\sup_{(y,p,q)\in Y'\times P'\times Q'} \mathcal{H}^{m-1}(W_{\le}(y,p)\cap Z_{=}(y,q)) \quad \text{(2nd level set size)}, \tag{70}$$

$$\sup_{(y,p,q)\in Y'\times P'\times Q'} \mathcal{H}^{m}(W_{\le}(y,p)\cap Z_{\le}(y,q)) \quad \text{(iterated sublevel size)}, \tag{71}$$

$$\sup_{(y,p,q)\in Y'\times P'\times Q'} \mathcal{H}^{m-2}((\overline{W_{=}(y,p)\cap Z_{\le}(y,q)})\cap\partial X) \quad \text{(1st boundary level size)}, \tag{72}$$

$$\sup_{(y,p,q)\in Y'\times P'\times Q'} \mathcal{H}^{m-2}((\overline{W_{\le}(y,p)\cap Z_{=}(y,q)})\cap\partial X) \quad \text{(2nd boundary level size)}, \tag{73}$$

and

$$\sup_{(y,p,q)\in Y'\times P'\times Q'} \mathcal{H}^{m-1}((\overline{W_{\le}(y,p)\cap Z_{\le}(y,q)})\cap\partial X) \quad \text{(boundary sublevels size)}. \tag{74}$$

We note that, as in [10], since the divergence operator on ∂X involves first derivatives of the metric, which is as smooth as the outward unit normal ∂X, we require $\hat{n}_X \in W^{1,1}$ to define $\nabla_{\partial X}\cdot$. By convention $C^{-1,1} := L^\infty$.

Remark 24 (Convention) *We interpret the infimum in the linear independence assumption* (68) *to be 1 when $\partial X\cap\overline{W_{=}(y,P')}\cap\overline{Z_{=}(y,Q')}$ is empty. When $m=2$, this is the only case for which the assumption can hold (since three vectors in two dimensions cannot be linearly independent). When $m=2$, positivity of* (67) *implies $W_{=}(y,p)\cap Z_{=}(y,q)$ is discrete; with the interpretation above, assumption* (68) *amounts to the condition that this set be disjoint from ∂X for all $(p,q)\in P'\times Q'$.*

Remark 25 (Simplified hypotheses) *Note the positivity of* (64) *follows from* (62)*, while* (71) *and* (74) *are controlled by $\mathcal{H}^m[X]$ and $\mathcal{H}^{m-1}[\partial X]$. Together with* (48) *and $\sup_{y\in Y'}\|s_y\|_{C^{1,1}(X)}$, the same quantities control* (69) *and* (70) *by Lemma 7.2 of [10]. Finiteness and positivity of the remaining quantities* (72)–(73) *listed in the theorem follows from the transversality hypothesized in* (63).

Example 26 (Annulus to circle revisited) *Consider the annulus to circle example of Example 7, on a domain* $Y' = Y = \{\theta \mid \theta \in (-\pi, \pi)\}$ *again embedded in* $\mathbf{R}^2$ *by* $\hat{y}(\theta) = (\cos(\theta), \sin(\theta))$*, with* $P' = Q' = (-\epsilon, \epsilon)$ *small neighbourhoods around the solution* $v(y) = 0$ *explicitly computed there. It remains to verify the transversality hypotheses* (63).

It is straightforward to see that $W_=(\hat{y}(\theta), p) = X_1(\hat{y}(\theta), p) = \{x \mid x \cdot (-\sin(\theta), \cos(\theta)) = p\}$ *and* $Z_=(\hat{y}(\theta), q) = \{x \mid x \cdot (-\cos(\theta), \sin(\theta)) = q\}$ *are orthogonal line segments passing near the origin; for* $\epsilon < 2^{-1/2}$ *they do not intersect on the boundary* $\partial X = \{x \mid 1 = |x|\}$ *of* X*, rendering condition* (68) *vacuous. Since the normals to* $W_=$ *and* $Z_=$ *are orthogonal to each other, and do not parallel* $\hat{n}_X$ *at the points where* $W_=$ *(respectively* $Z_=$*) intersect* ∂X*, the other three transversality conditions hold as well.*

Lemma 27 (Derivatives on moving submanifolds-with-boundary) *Given real-valued Lipschitz functions* a, b, c *on* $X' \times Y' \times P' \times Q'$*, and* $a^\partial, b^\partial, c^\partial$ *on* $\partial X' \times Y' \times P' \times Q'$*, the functions*

$$\begin{aligned}
A(y,p,q) &:= \int_{X_2(y,p,q)} a(x,y,p,q)\, d\mathcal{H}^{m-1}(x), \\
B(y,p,q) &:= \int_{W_{\leq}(y,p) \cap Z_{\leq}(y,q)} b(x,y,p,q)\, d\mathcal{H}^{m}(x), \\
C(y,p,q) &:= \int_{W_{\leq}(y,p) \cap Z_{=}(y,q)} c(x,y,p,q)\, d\mathcal{H}^{m-1}(x), \\
A^\partial(y,p,q) &:= \int_{(\overline{W_{=}(y,p) \cap Z_{\leq}(y,q)}) \cap \partial X} a^\partial(x,y,p,q)\, d\mathcal{H}^{m-2}(x), \\
B^\partial(y,p,q) &:= \int_{(\overline{W_{\leq}(y,p) \cap Z_{\leq}(y,q)}) \cap \partial X} b^\partial(x,y,p,q)\, d\mathcal{H}^{m-1}(x) \text{ and} \\
C^\partial(y,p,q) &:= \int_{(\overline{W_{\leq}(y,p) \cap Z_{=}(y,q)}) \cap \partial X} c^\partial(x,y,p,q)\, d\mathcal{H}^{m-2}(x)
\end{aligned}$$

are all Lipschitz, with derivatives given almost everywhere by the formulae in Appendix A. Here $w := |D^2_{xy} s|^{-1}$ *and* $z := |D^3_{xyy} s|^{-1}$ *as above.*

Proof. The proof is similar to the proofs of Lemma 7.4 in [10] and Lemma 21 in the present paper; we only described the main differences here. For a sufficiently smooth integrand, the derivative of A with respect to p, for example, includes a term capturing differentiation of the integrand with respect to a, and a term capturing the dependence of the region of integration, which we compute using the generalized divergence theorem and Lemma 5.1 in [10]:

$$
\begin{aligned}
A_p - \int_{X_2} a_p d\mathcal{H}^{m-1} &= \frac{\partial}{\partial \tilde{p}}\bigg|_{\tilde{p}=p} \int_{W_=(y,\tilde{p})\cap Z_\leq(y,q)} a(x,y,p,q)\hat{n}_W \cdot \hat{n}_W d\mathcal{H}^{m-1}(x) \\
&= \frac{\partial}{\partial \tilde{p}}\left[\int_{W_\leq \cap Z_\leq} \nabla \cdot (a\hat{n}_W) d\mathcal{H}^m - \int_{W_\leq \cap Z_=} a\hat{n}_W \cdot \hat{n}_Z d\mathcal{H}^{m-1} - \int_{\overline{W_\leq \cap Z_\leq}\cap \partial X} a\hat{n}_W \cdot \hat{n}_X d\mathcal{H}^{m-1}\right]_{\tilde{p}=p} \\
&= \int_{W_= \cap Z_\leq} \nabla \cdot (a\hat{n}_W) w d\mathcal{H}^{m-1} - \int_{W_= \cap Z_=} \frac{aw\hat{n}_W \cdot \hat{n}_Z}{\sqrt{1-(\hat{n}_W \cdot \hat{n}_Z)^2}} d\mathcal{H}^{m-2} \\
&\quad - \int_{\overline{W_= \cap Z_\leq}\cap \partial X} \frac{aw\hat{n}_W \cdot \hat{n}_X}{\sqrt{1-(\hat{n}_W \cdot \hat{n}_X)^2}} d\mathcal{H}^{m-2}.
\end{aligned}
$$

The result for Lipschitz functions can be obtained as in Lemma 7.4 in [10] and Lemma 21 here, via approximation by $C^{1,1}$ integrands and the dominated convergence theorem. The arguments for other derivatives of A, B and C are similar. We treat boundary integrals analogously. Noting that, for instance,

$$
\begin{aligned}
A^\partial(y,p,q) &= \int_{(\overline{W_= \cap Z_\leq})\cap \partial X)} a^\partial(x,y,p,q) d\mathcal{H}^{m-2}(x) \\
&= \int_{(\overline{W_\leq \cap Z_\leq})\cap \partial X)} \nabla_{\partial X} \cdot (a^\partial \hat{n}_{\partial,W}) d\mathcal{H}^{m-1}(x) - \int_{(\overline{W_\leq \cap Z_=})\cap \partial X)} a^\partial \hat{n}_{\partial,W} \cdot \hat{n}_{\partial,Z} d\mathcal{H}^{m-2}(x),
\end{aligned}
$$

where $\hat{n}_{\partial,W}$ and $\hat{n}_{\partial,Z}$ are defined in Appendix A, we can again use Lemma 5.1 in [10] to differentiate with respect to p. Similar arguments apply to all derivatives of A^∂, B^∂ and C^∂. ■

We note that all integrals over the domain $W_= \cap Z_=$ can be rewritten using the divergence theorem as follows:

$$
\begin{aligned}
\int_{W_= \cap Z_=} a(x,y,p,q) d\mathcal{H}^{m-2}(x) = &\int_{W_= \cap Z_\leq} \nabla_{W_=} \cdot (\alpha \hat{n}_{W_=,Z}) d\mathcal{H}^{m-1}(x) \\
&- \int_{(\overline{W_= \cap Z_\leq})\cap \partial X} a\hat{n}_{W_=,Z} \cdot \hat{n}_{W_=,X} d\mathcal{H}^{m-2}(x)
\end{aligned}
$$

where $\hat{n}_{W_=,Z} := \frac{\hat{n}_Z - (\hat{n}_Z \cdot \hat{n}_W)\hat{n}_W}{\sqrt{1-(\hat{n}_Z \cdot \hat{n}_W)^2}}$ and $\hat{n}_{W_=,X} := \frac{\hat{n}_X - (\hat{n}_X \cdot \hat{n}_W)\hat{n}_W}{\sqrt{1-(\hat{n}_X \cdot \hat{n}_W)^2}}$ are the outward unit normals in the submanifold $W_= \subseteq X$ to $Z_\leq$ and X, respectively.

Similarly,

$$
\int_{(\overline{W_= \cap Z_=})\cap \partial X} a(x,y,p,q) d\mathcal{H}^{m-3}(x) = \int_{(\overline{W_= \cap Z_\leq})\cap \partial X} \nabla_{\bar{W}_= \cap \partial X} \cdot (a\hat{n}_{\bar{W}_= \cap \partial X, Z}) d\mathcal{H}^{m-2}(x),
$$

where $\hat{n}_{\bar{W}_= \cap \partial X, Z}$ is the outward unit normal to $\bar{Z}_\leq \cap (\bar{W}_= \cap \partial X)$ in the codimension 2 submanifold $(\bar{W}_= \cap \partial X)$; alternatively, it is equal to $\hat{n}_Z$, minus its projection onto the span of $\hat{n}_X$ and $\hat{n}_W$. This means that differentiating a function of any of the types in Lemma 27 with respect to any of p, q or y results in a sum of functions of these same types; we can therefore iterate these operations to compute higher order derivatives. The following Lemma keeps track of the effect on regularity of differentiating the sum of the terms in Lemma 27.

Lemma 28 (More iterated derivative bounds) *Set*

$$B_r := C^{r,1}(\bar{X}' \times \bar{Y}' \times \bar{P}' \times \bar{Q}') \times C^{r,1}((\bar{X}' \cap \partial X) \times \bar{Y}' \times \bar{P}' \times \bar{Q}')$$

and consider the operators $M^p, M^q, M^y : (B_r)^3 \to (B_{r-1})^3$ defined by

$$\begin{aligned}
M^p &: (a, b, c, a^\partial, b^\partial, c^\partial) \mapsto \\
&\Big(a_p + \nabla \cdot (a\hat{n}_W)w + bw - \nabla_{W_=} \cdot (\frac{aw\hat{n}_W \cdot \hat{n}_Z}{\sqrt{1-(\hat{n}_W \cdot \hat{n}_Z)^2}}\hat{n}_{W_=,Z}) + \nabla_{W_=} \cdot (cw\hat{n}_{W_=,Z}), b_p, c_p, \\
&-\frac{aw\hat{n}_W \cdot \hat{n}_X}{\sqrt{1-(\hat{n}_W \cdot \hat{n}_X)^2}} + a_p^\partial + \frac{w}{\sqrt{1-(\hat{n}_W \cdot \hat{n}_X)^2}}\nabla_{\partial X} \cdot (a^\partial \hat{n}_{\partial,W}) + \frac{b^\partial w}{\sqrt{1-(\hat{n}_W \cdot \hat{n}_X)^2}} \\
&-\frac{aw\hat{n}_W \cdot \hat{n}_Z}{\sqrt{1-(\hat{n}_W \cdot \hat{n}_Z)^2}}\hat{n}_{W_=,Z} \cdot \hat{n}_{W_=,X} + cw\hat{n}_{W_=,Z} \cdot \hat{n}_{W_=,X} \\
&-\nabla_{\bar{W}_= \cap \partial X} \cdot (a^\partial \hat{n}_{\partial,W} \cdot \hat{n}_{\partial,Z} \frac{w}{\sqrt{[1-(\hat{n}_W \cdot \hat{n}_X)^2][1-(\hat{n}_W \cdot \hat{n}_Z)^2]}}\hat{n}_{\bar{W}_= \cap \partial X, Z}) \\
&+\nabla_{\bar{W}_= \cap \partial X} \cdot (c\frac{w}{\sqrt{[1-(\hat{n}_W \cdot \hat{n}_X)^2][1-(\hat{n}_Z \cdot \hat{n}_W)^2]}}\hat{n}_{\bar{W}_= \cap \partial X, Z}), b_p^\partial, c_p^\partial\Big),
\end{aligned}$$

$$\begin{aligned}
M^q &: (a, b, c, a^\partial, b^\partial, c^\partial) \mapsto \\
&\Big(a_q + \nabla_{W_=} \cdot (\frac{az}{\sqrt{1-(\hat{n}_Z \cdot \hat{n}_W)^2}}\hat{n}_{W_=,Z}) - \nabla_{W_=} \cdot (\frac{cz\hat{n}_Z \cdot \hat{n}_W}{\sqrt{1-(\hat{n}_W \cdot \hat{n}_Z)^2}}\hat{n}_{W_=,Z}), \\
&b_q, bz + c_q + \nabla \cdot (c\hat{n}_Z)z, \\
&a_q^\partial - \frac{az}{\sqrt{1-(\hat{n}_Z \cdot \hat{n}_W)^2}}\hat{n}_{W_=,Z} \cdot \hat{n}_{W_=,X} + \frac{cz\hat{n}_Z \cdot \hat{n}_W}{\sqrt{1-(\hat{n}_W \cdot \hat{n}_Z)^2}}\hat{n}_{W_=,Z} \cdot \hat{n}_{W_=,X} \\
&-\nabla_{\overline{W_=} \cap \partial X} \cdot (c^\partial \hat{n}_{\partial,Z} \cdot \hat{n}_{\partial,W} \frac{z}{\sqrt{[1-(\hat{n}_Z \cdot \hat{n}_X)^2][1-(\hat{n}_Z \cdot \hat{n}_W)^2]}}\hat{n}_{\bar{W}_= \cap \partial X, Z}) \\
&\nabla_{\bar{W}_= \cap \partial X} \cdot (a^\partial \frac{z}{\sqrt{[1-(\hat{n}_Z \cdot \hat{n}_X)^2][1-(\hat{n}_W \cdot \hat{n}_Z)^2]}} d\mathcal{H}^{m-3}(x))\hat{n}_{\bar{W}_= \cap \partial X, Z}, b_q^\partial, \\
&-\frac{cz\hat{n}_Z \cdot \hat{n}_X}{\sqrt{1-(\hat{n}_Z \cdot \hat{n}_Z)^2}} + c_q^\partial + \frac{b^\partial z}{\sqrt{1-(\hat{n}_Z \cdot \hat{n}_X)^2}} + \frac{z}{\sqrt{1-(\hat{n}_Z \cdot \hat{n}_X)^2}}\nabla_{\partial X} \cdot (c^\partial \hat{n}_{\partial,Z})\Big)
\end{aligned}$$

and

$$
\begin{aligned}
M^y :\ & (a,b,c,a^\partial,b^\partial,c^\partial) \mapsto \\
& \Big(a_y - \nabla\cdot(a\hat{n}_W)ws_{yy} - bws_{yy} - c\frac{\partial \hat{n}_Z}{\partial y}\cdot\hat{n}_W, \\
& \nabla\cdot(a\frac{\partial \hat{n}_W}{\partial y}) + b_y + \nabla\cdot(c\frac{\partial \hat{n}_Z}{\partial y}), \\
& -a\frac{\partial \hat{n}_W}{\partial y}\cdot\hat{n}_Z - bzs_{yyy} + c_y - \nabla\cdot(c\hat{n}_Z)zs_{yyy}, \\
& \frac{aws_{yy}\hat{n}_W\cdot\hat{n}_X}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}} + a^\partial_y + \frac{ws_{yy}}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}}\nabla_{\partial X}\cdot(a^\partial\hat{n}_{\partial,W}) - \frac{b^\partial ws_{yy}}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}} - \\
& c^\partial\frac{\partial\hat{n}_{\partial,Z}}{\partial y}\cdot\hat{n}_{\partial,W}, \\
& -a\frac{\partial \hat{n}_W}{\partial y}\cdot\hat{n}_X - c\frac{\partial \hat{n}_Z}{\partial y}\cdot\hat{n}_X + \nabla_{\partial X}\cdot(a^\partial\frac{\partial\hat{n}_{\partial,W}}{\partial y}) + b^\partial_y + \nabla_{\partial X}\cdot(c^\partial\frac{\partial\hat{n}_{\partial,Z}}{\partial y}), \\
& +\frac{czs_{yyy}\hat{n}_Z\cdot\hat{n}_X}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}} - a^\partial\frac{\partial\hat{n}_{\partial,W}}{\partial y}\cdot\hat{n}_{\partial,Z} - \frac{b^\partial zs_{yyy}}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}} + c^\partial_y \\
& -\frac{zs_{yyy}}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}}\nabla_{\partial X}\cdot(c^\partial\hat{n}_{\partial,Z})\Big)
\end{aligned}
$$

Then the norms $||M^p||$, $||M^q||$ and $||M^y||$ are controlled by non-degeneracy, transversality, linear independence, $||s_y||_{C^{r+2,1}}$ and $||\hat{n}_X||_{C^{r,1}}$.

Proof. It is straightforward to compute:

$$
\begin{aligned}
||M^p|| \le\ & 1 + ||\hat{n}_W||_{C^{r,1}}||w||_{C^{r,1}} + ||w||_{C^{r-1}} + ||\frac{w\hat{n}_W\cdot\hat{n}_Z}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_Z)^2}}\hat{n}_{W_=,Z}||_{C^{r,1}} + ||w\hat{n}_{W_=,Z}||_{C^{r,1}} + 1 + 1 \\
& + ||\frac{w\hat{n}_W\cdot\hat{n}_X}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}}||_{C^{r-1,1}} + 1 + ||\frac{w}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}}||_{C^{r-1,1}}||\hat{n}_{\partial,W}||_{C^{r,1}} \\
& + ||\frac{w}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}}||_{C^{r-1,1}} + ||\frac{w\hat{n}_W\cdot\hat{n}_Z}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_Z)^2}}\hat{n}_{W_=,Z}\cdot\hat{n}_{W_=,X} + ||w\hat{n}_{W_=,Z}\cdot\hat{n}_{W_=,X}||_{C^{r-1,1}} \\
& + ||\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z}\frac{w}{\sqrt{[1-(\hat{n}_W\cdot\hat{n}_X)^2][1-(\hat{n}_W\cdot\hat{n}_Z)^2]}}\hat{n}_{\bar{W}_=\cap\partial X,Z}||_{C^{r,1}} \\
& + ||\frac{w}{\sqrt{[1-(\hat{n}_W\cdot\hat{n}_X)^2][1-(\hat{n}_Z\cdot\hat{n}_W)^2]}}\hat{n}_{\bar{W}_=\cap\partial X,Z})||_{C^{r,1}} + 1 + 1.
\end{aligned}
$$

Similar estimates hold for M^q and M^y, and it is straightforward to see that the upper bounds are controlled by the indicated quantities. ■

We are now ready to prove the Theorem 23, on the regularity of G_2.
Proof. The proof is similar to the proof of Theorem 14; for indices α,β,γ with $\alpha+\beta+\gamma\le r$, we apply the iterated operators $(M^y)^\alpha(M^p)^\beta(M^q)^\gamma$

to $(h,0,0,0,0,0)$, where $h = h(x,y,q) = f(x)\frac{q - s_{yy}(x,y)}{|D^2_{xy}s(x,y)|} \in C^{r,1}$ is the integrand in G_2. Setting

$$(M^y)^\alpha (M^p)^\beta (M^q)^\gamma (h,0,0,0,0,0) = (a,b,c,a^\partial,b^\partial,c^\partial) \in B^{r-(\alpha+\beta+\gamma)},$$

we have that

$$\begin{aligned}\frac{\partial^{\alpha+\beta+\gamma} G_2}{\partial y^\alpha \partial p^\beta \partial q^\gamma} &= \int_{W_= \cap Z_\le} a d\mathcal{H}^{m-1}(x) + \int_{W_\le \cap Z_\le} b d\mathcal{H}^m(x) + \int_{W_\le \cap Z_=} c d\mathcal{H}^{m-1}(x) \\ &+ \int_{(\overline{W_= \cap Z_\le}) \cap \partial X} a^\partial d\mathcal{H}^{m-2}(x) + \int_{(\overline{W_\le \cap Z_\le}) \cap \partial X} b^\partial d\mathcal{H}^{m-1}(x) + \int_{(\overline{W_\le \cap Z_=}) \cap \partial X} c^\partial d\mathcal{H}^{m-2}(x)\end{aligned}$$

is Lipschitz by Lemma 27, and its norm is controlled by the sizes of the domains of integration and the L^∞ norms of the integrands, which are in turn controlled by the desired quantities as a consequence of iterating Lemma 28. As in Theorem 14 in the previous section, and the corresponding result for one dimensional targets in [10], we observe that since the initial function $(h,0,0,0,0,0)$ we apply the operators to does not include any boundary terms, the norm of the first application depends on $||\hat{n}_X||_{C^{r-1,1}}$ rather than $||\hat{n}_X||_{C^{r,1}}$, saving a derivative of smoothness in $\hat{n}_X$ in the final result. ■

Corollary 29 (Boostrapping smoothness for local ODE) *Assume that the conditions in Theorem 23 hold for Y', $P' = v'(Y')$ and $Q' = v''(Y')$ for some $r \geq 0$, that $g \in C^{r,1}(Y')$ is bounded from below on Y', $g \geq L_g > 0$, and that $f \in C^{r,1}(X')$ is bounded from above and below $\infty > U_f \geq f \geq L_f > 0$ on $X' = \cup_{(y,p,q) \in Y' \times P' \times Q'} X_2(y,p,q)$.*

Then any $v \in C^{1,1}(Y')$ solving (23)–(24) [a.e. on Y'] *with $i = 2$ is in fact in $C^{r+2,1}(Y')$.*

Proof. Setting $k(y) := v'(y)$, we have $g(y) = G_2(y,k(y),k'(y))$ a.e. At any y where this holds, we must have $\mathcal{H}^{m-1}(X_2(y,k(y),k'(y))) \geq C > 0$, where C depends on L_g, U_f, $\min|D^2_{xy}s|$, $\max|D^3_{xyy}s|$ and the diameter of X. Noting that

$$\begin{aligned}\frac{\partial G_2}{\partial q}(y,p,q) &= \int_{X_2(y,p,q)} \frac{f(x)}{|D^2_{xy}s(x,y)|} d\mathcal{H}^{m-1}(x) \\ &+ \int_{Z_=(y,p)\cap W_=(y,q)} \frac{(q - s_{yy}(x,y))f(x)z(x,y)}{|D^2_{xy}s(x,y)|} d\mathcal{H}^{m-2}(x), \\ &= \int_{X_2(y,p,q)} \frac{f(x)}{|D^2_{xy}s(x,y)|} d\mathcal{H}^{m-1}(x)\end{aligned}$$

(since $q - s_{yy} = 0$ along $Z_=$) this yields a lower bound on $\frac{\partial G_2}{\partial q}(y,p,q) > B$ in the region of interest.

Therefore, by the Clarke inverse function theorem [11, Theorem 7.1.1], $q \mapsto G_2(y, p, q)$ is invertible; denoting its inverse $q(y, p, \cdot)$, q is as smooth as G_2 (that is, $q \in C^{r,1}$) and we have, almost everywhere

$$k'(y) = q(y, k(y), g(y)). \tag{75}$$

The Lipschitz function k is then equal to the antiderivative of its derivative; for a fixed y_0, we have

$$k(y) - k(y_0) = \int_{y_0}^{y} k'(s)ds = \int_{y_0}^{y} q(s, k(s), g(s))ds.$$

The fundamental theorem of calculus then implies that k is *everywhere* differentiable, and that (75) holds *for all* y. In particular, k' is Lipschitz as $y \mapsto q(y, k(y), g(y))$, hence $v \in C^{2,1}(Y')$. If $r > 0$, one can immediately bootstrap to get $k' \in C^{r,1}(Y')$, hence $v \in C^{r+2,1}(Y')$. ■

Remark 30 (Is G_2 smooth for higher dimensional targets?) *It is natural to ask whether the proofs of Theorem 14 and 23 can be adapted to yield smoothness of G_2 when $n > 1$. While this is conceivable, there is a significant hurdle: the structure of the set $X_2(y, p, Q)$ is not amenable to our techniques, since it is no longer a manifold-with-boundary; it is at best a manifold-with-corners, and in the absence of additional restrictions might be worse. To see this, recall $X_2(y, p, Q) = X_1(y, p) \cap A(y, Q)$, where $A(y, Q) := \{x : Q - D^2_{yy}s(x, y) \geq 0\}$. Non-negative definitess of an $n \times n$ matrix is determined by a system of nonlinear inequalities whose saturation sets are not generally manifolds — nevermind intersecting transversally — unless $n = 1$. These inequalities force the eigenvalues of $Q - D^2_{yy}s$ into the positive orthant. Although the positive orthant is a manifold-with-corners, we know of no implicit function type theorem that describes the set of $x \in X_1(y, p)$ satisfying such a system.*

8 $C^{1,1}$ regularity of solutions to the $i = 2$ ODE

In this section, we'll assume that $n = 1$ (one dimensional target), and that the minimizer $(u, v) = (v^s, u^{\tilde{s}})$ to (8) satisfies the local equation (23)–(24) a.e. for $i = 2$, which also implies (33). Recall $v = v^{s\tilde{s}}$ is in $C^{0,1}(Y)$ and semiconvex from (11). Our goal is to identify a subset $Y' \subset Y$ on which the initial smoothness hypothesis $v \in C^{1,1}(Y')$ needed for the bootstrap (Corollary 29 of the preceding section) is satisfied.

Assume the probability densities f and g satisfy

$$\log f \in C(X) \text{ and } \log g \in C(Y), \text{ so that} \tag{76}$$

$$0 < L_f \leq f(x) \leq U_f < \infty \text{ for all } x \in X \tag{77}$$

$$0 < L_g \leq g(y) \leq U_g < \infty \text{ for all } y \in Y. \tag{78}$$

As in Section 7, we'll assume X, Y and s satisfy (60)–(61). We shall work on open regions $Y' \subset Y$ and $P' \subset P = \frac{\partial s}{\partial y}(X, Y)$ satisfying

$$\inf_{(y,p)\in Y'\times P'} \mathcal{H}^{m-1}(X_1(y,p)) > 0. \tag{79}$$

Taking $Q' = \mathbf{R}$ and $X' = \cup_{(y,p)\in Y'\times P'} X_1(y,p)$, we'll assume (46) and positivity of (48) as in Section 6. In particular, s is uniformly non-degenerate and — without loss of generality — convex with respect to $y \in Y'$, so that $k(y) := v'(y)$ is non-decreasing, and the upper bound (50) complementing (79) is finite by Remark 25. We shall also require the function $q(y,p,\beta)$ (defined before (75) and recalled in the next paragraph) to be continuous, and that s satisfy an enhanced form of the twist condition detailed in Assumption 32 below.

The functional

$$G_2(y,p,q) := \int_{X_2(y,p,q)} \frac{q - s_{yy}(x,y)}{|D^2_{xy}s(x,y)|} f(x) d\mathcal{H}^{m-1}(x). \tag{80}$$

is strictly increasing in q on $\{(y,p,q) \mid G_2 > 0\}$ and diverges as $q \to \infty$. For $\beta > 0$ and fixed y and p, as in Section 7, denote by $q(y,p,\beta)$ the unique solution of

$$q \mapsto G_2(y,p,q) = \beta \tag{81}$$

We therefore have $q(y,k(y),g(y)) = k'(y)$ almost everywhere. Under the assumptions of Theorem 23 for some $r \geq 0$, $q(y,p,\beta)$ is Lipschitz continuous in all its arguments, by the Clarke implicit function theorem, [11, Section 7.1]. Although it holds only almost everywhere, this formulation provides some intuition for why we expect k to be Lipschitz, since boundedness of k and g imply boundedness of k' wherever the equation $q(y,k(y),g(y)) = k'(y)$ holds.

The estimate below essentially controls the volume of the region $F^{-1}([y_0,y_1])$ mapped to an interval $[y_0,y_1]$ by the map F of Theorem 2 using the variation in k, compensated by a term reflecting the variation in y. Together with mass balance, this proposition easily implies that k is Lipschitz if continuous (and so $v \in C^{1,1}_{loc}$ if $v \in C^1_{loc}$), as we will show in Theorem 35.

The continuity of k, assumed in Proposition 31, will be confirmed in Proposition 34.

Proposition 31 (The derivative of v is Lipschitz if continuous) *Let $Y' \subset Y$ and $P' \subset P = \frac{\partial s}{\partial y}(X, Y)$ satisfy the hypotheses imposed between* (76) *and* (80)*. Then there exist positive constants C_1 and C_2 such that*

$$\mathcal{H}^m\left(\bigcup_{y\in[y_0,y_1]} X_2(y,k(y),q(y,k(y),g(y)))\right) \geq C_1|k(y_0)-k(y_1)|-C_2|y_0-y_1| \tag{82}$$

for any $y_0, y_1 \in Y'$ and monotone increasing, continuous function $k : Y' \to P'$.

Before proving the Proposition, it is instructive to provide some intuition. For a *fixed* y, the coarea formula yields $\mathcal{H}^m(\cup_{p\in[p_0,p_1]}(X_1(y,p))) \sim |p_1 - p_0|$, where the constants of proportionality depend on two-sided bounds for $|D^2_{xy}s|$ and $\mathcal{H}^{m-1}(X_1(y,p))$. The equation

$$0 < g(y) = \int_{X_2(y,p,q(y,p,g(y)))} \frac{q - s_{yy}(x,y)}{|D^2_{xy}s|} f(x) d\mathcal{H}^m(x)$$

forces $X_2(y,p,q(y,p,g(y)))$ to fill up a proportion of $X_1(y,p)$ which can be bounded in terms of the same bounds as before, L_g, U_f, $\|s\|_{C^2}$, and $q(y,p,U_g)$. Thus there exists $C > 0$ such that

$$\mathcal{H}^m\left(\bigcup_{p\in[p_0,p_1]} X_2(y,p,q(y,p,g(y)))\right) \geq C|p_1 - p_0|.$$

In Proposition 31, y is not fixed but varies within an interval $[y_0, y_1]$, and $p = k(y)$ is now a function. Continuity and monotonicity force the image $k([y_0, y_1])$ to match the interval $[k_0 = k(y_0), k_1 = k(y_1)]$. If the domains $X_2(y,k(y),q(y,k(y),g(y))$ were independent of y, the result would then follow immediately, without the second term on the right hand side of (82). The second term compensates for the possibility that as y and $k(y)$ change, the level curves bend in a way that reduces the volume on the left hand side.

Proof. Set $k_i = k(y_i)$ for $i = 0, 1$, and choose C such that $|s_y(x,y_0) - s_y(x,y)| \leq C|y_0 - y|$ for all $y \in Y'$ and $x \in X_1(Y', P')$.

Suppose that $k_0 \leq s_y(x,y_0) \leq k_1 - C|y_1 - y_0|$. Then

$$s_y(x,y_1) \leq s_y(x,y_0) + C|y_1 - y_0| \leq k_1$$

By the intermediate value theorem, $k(y) = s_y(x,y)$ for some $y \in [y_0, y_1]$; that is, $x \in X_1(y,k(y))$.

Now, suppose in addition that $q(y_0, s_y(x,y_0), g(y_0)) - s_{yy}(x,y_0) \geq \alpha > 0$; by uniform continuity, there is a $\delta > 0$ (depending on α but not y) such that, we have

$$q(y,k(y),g(y)) - s_{yy}(x,y) \geq 0;$$

that is, $x \in X_2(y, k(y), q(y, k(y), g(y)))$, if $|y - y_0| < \delta$.

Now, note if $|y_0 - y_1| \geq \delta$, the right hand side of (82) is negative for appropriate choices of C_1, C_2 (note that $k(y_0) - k(y_1)$ is less than or equal to the diameter of P'). We can therefore assume $|y_0 - y_1| < \delta$ without loss of generality, and the above argument then yields $x \in X_2(y, k(y), q(y, k(y), g(y)))$ for some $y \in [y_0, y_1]$.

It therefore follows that

$$\begin{aligned}\cup_{p \in [k_0, k_1 - C|y_1 - y_0|]}\{x \in X_1(y_0, p) \mid q(y_0, p, g(y_0)) - s_{yy}(x, y_0) \geq \alpha\} \quad (83)\\ \subseteq \cup_{y \in [y_0, y_1]} X_2(y, k(y), q(y, k(y), g(y)))\end{aligned}$$

Now our definition of q yields $g(y_0) = G_2(y_0, p, q(y_0, p, g(y_0)))$, which implies that for a small enough α, we have

$$\mathcal{H}^{m-1}(\{x \in X_1(y_0, k(y)) : q(y_0, k(y), g(y_0)) - s_{yy}(x, y_0) \geq \alpha\}) \geq B\mathcal{H}^{m-1}(X_1(y_0, k(y)))$$

for some $B > 0$ depending on the lower bound L_g for g, $\min |D^2_{xy}s|$, $\max |D^3_{xyy}s|$ and the size of the level sets, $\sup_{(y,p) \in Y \times k(Y)} \mathcal{H}^{m-1}(\overline{X_1(y, p)})$. It then follows that

$$\begin{aligned}&vol[\cup_{p \in [k_0, k_1 - C|y_1 - y_0|]}\{x \in X_1(y_0, p) \mid q(y_0, p, g(y_0)) - s_{yy}(x, y_0) \geq \alpha\}]\\ &\geq B vol[\cup_{p \in [k_0, k_1 - C|y_1 - y_0|]} X_1(y_0, p)] \quad (84)\end{aligned}$$

Now, if $k_1 - 2C|y_1 - y_0| \leq k_0$, then $|k_1 - k_0| - 2C|y_1 - y_0| < 0$ and there is nothing to prove, since the right hand side of (82) is negative for appropriate choices of the constants.

On the other hand, if $k_1 - 2C|y_1 - y_0| \geq k_0$, then $k_1 - C|y_1 - y_0| - k_0 \geq \frac{|k_1 - k_0|}{2}$, and so

$$vol[\cup_{p \in [k_0, k_1 - C|y_1 - y_0|]} X_1(y_0, p)] \geq D\frac{|k_1 - k_0|}{2},$$

where D depends on the size $\min_{p \in [k_0, k_1]} \mathcal{H}^{m-1}(\overline{X_1(y_0, p)})$ of the level sets, and the speed limit $\min_{s_y(x, y_0) \in [k_0, k_1]} |D^2_{xy}s(y_0, x)|$. This combined with (83) and (84) establishes the result. ∎

Our strategy is to combine Proposition 31 with mass balance to deduce a Lipschitz condition on k. To apply this Proposition, we must first show that k is continuous. We do this under the following strengthening of the twist condition:

Assumption 32 (Enhanced twist) *We say $s \in C^2(X \times Y)$ satisfies the enhanced twist condition if $D_x s(x, y) - D_x s(x, \bar{y})$ and $D^2_{xy}s(x, \bar{y})$ are linearly independent for each $x \in X$ and $y, \bar{y} \in Y \subset \mathbf{R}$ with $y \neq \bar{y}$.*

Note that the usual twist condition asserts injectivity of the mapping $y \mapsto D_x s(x, y)$; injectivity of the projection of $D_x s(x, y)$ onto the potential level sets $X_1(y, k)$ of the optimal map is sufficient to imply our enhanced twist condition.

Lemma 33 (Map continuity on interior of isodestination set) *Under the enhanced twist condition, the optimal map is uniquely defined (and therefore continuous) at any point x in the relative interior of $\partial_{\tilde{s}} v(\bar{y})$ in $X_1(\bar{y}, p)$.*

Proof. If x lies in the interior of $\partial_{\tilde{s}} v(\bar{y})$ relative to $X_1(\bar{y}, p)$ then $u(\tilde{x}) = s(\tilde{x}, \bar{y}) - v(\bar{y})$ for all $\tilde{x} \in X_1(\bar{y}, p)$ sufficiently close to x. Therefore, u is smooth along $X_1(\bar{y}, p)$ and differentiating we have

$$D_{X_1(\bar{y},p)} u(x) = D_{X_1(\bar{y},p)} s(x, \bar{y}).$$

On the other hand, if there is another $y \neq \bar{y}$ such that $x \in \partial_{\tilde{s}} v(y)$, the envelope condition yields

$$D_{X_1(\bar{y},p)} u(x) = D_{X_1(\bar{y},p)} s(x, y)$$

and so

$$D_{X_1(\bar{y},p)} s(x, y) = D_{X_1(\bar{y},p)} s(x, \bar{y}),$$

violating the enhanced twist condition. ■

Proposition 34 (Continuous differentiability of v) *Let $(u, v) = (v^s, u^{\tilde{s}})$ achieve the minimum* (8) *and satisfy* (23)–(24) *with $i = 2$. Under the hypotheses imposed between* (76) *and* (80)*, $k = v'$ is continuous on $Y' \subset Y$.*

Proof. Recall our assumption (46), which costs no generality, and implies $s_{yy} \geq 0$ so that k is monotone increasing as before. We need only rule out jump discontinuities. Suppose k has a jump discontinuity at $\bar{y}$, with left and right limits k_0 and k_1, respectively.

For any y where k is differentiable with $g(y) = G_2(y, k(y), k'(y))$ and $X_2(y, k(y), k'(y)) = \partial_{\tilde{s}} v(y)$, the mean value theorem for integrals yields $x \in \partial_{\tilde{s}} v(y)$ at which

$$0 < L_g \leq g(y) = \frac{k'(y) - s_{yy}(x, y)}{|D^2_{xy} s|} f(x) \mathcal{H}^{m-1}[X_2(y, k(y), k'(y))].$$

Letting $X'' := \{x : d(x, X') < \delta\}$ be a neighbourhood of $X' := (s\text{-}\exp \circ D v^s)^{-1}(Y')$ for some $\delta > 0$ implies $k'(y) - s_{yy}(x, y) \geq \beta$ holds throughout a ball of radius r in X, where $\beta := \frac{c L_g}{2 C U_f} > 0$ with $c :=$

$\min |D^2_{xy}s(x,y)|$, $C := \sup_{(y,k)\in Y\times k(Y)} \mathcal{H}^{m-1}(\overline{X_1(y,k)})$, and r depends only on δ and $B := \sup_{(x,y)\in X"\times Y'} |D_x s_{yy}(x,y)|$.

Now take a sequence $\{y_i\}$ with $y_i < \bar{y}$ for which this is true, converging to $\bar{y}$. We have that $k(y_i) \to k_0$ and, after passing to a subsequence, the centers x_i of the corresponding balls converge to an $\bar{x} \in X_1(\bar{y}, k_0) \cap \partial_{\tilde{s}} v(y)$. By continuity, we have $q(\bar{y}, k_0, g(\bar{y})) - s_{yy}(\bar{y}, \bar{x}) \geq 2\beta > 0$, and therefore, $q(y, k(y), g(y)) - s_{yy}(y, x) \geq \beta > 0$ for all (x, y) close to $(\bar{x}, \bar{y})$ with $y < \bar{y}$.

Therefore,

$$k'(y) - s_{yy}(y,x) \geq \beta > 0 \tag{85}$$

for all x near $\bar{x}$, and almost all $y < \bar{y}$ near $\bar{y}$.

In addition, since $B_r(x_i) \cap X_1(y_i, k(y_i)) \subseteq \partial_{\tilde{s}} v(y_i)$, and each $x \in B_r(\bar{x}) \cap X_1(\bar{y}, k_0)$ can be approximated by points $z_i(x) \in B_r(x_i) \cap X_1(\bar{y}_i, k(y_i))$, we can pass to the limit in the equality $u(z_i(x)) + v(y_i) = s(z_i(x), y_i)$ to obtain $u(x)+v(\bar{y}) = s(x, \bar{y})$; that is, $x \in \partial_{\tilde{s}} v(\bar{y})$. Therefore, $B_r(\bar{x}) \cap X_1(\bar{y}, k_0) \subset \partial_{\tilde{s}} v(\bar{y})$.

We have now shown that $\bar{x}$ is in the relative interior of $\partial_{\tilde{s}} v(\bar{y})$ in $X_1(\bar{y}, k_0)$. Lemma 33 therefore implies that the optimal map F is continuous at $\bar{x}$. We next show that all points in the open set $X_>(\bar{y}, k_0) := \{x \in X \mid s_y(x, \bar{y}) > k_0\}$ sufficiently near $\bar{x}$ must get mapped to $\bar{y}$; this violates mass balance and establishes the result.

Choose x with $s_y(x, \bar{y}) > k_0$ such that $|\bar{x} - x| < \epsilon$, and set $y = F(x)$. The continuity of F at $\bar{x}$ ensures y is close to $\bar{y}$; for $\epsilon > 0$ sufficiently small we shall prove it must actually be equal to $\bar{y}$. First observe $s_y(x, y) < k_1$ for $\epsilon > 0$ sufficiently small, since $s_y(\bar{x}, \bar{y}) = k_0 < k_1$.

If $y > \bar{y}$, then $k(y) > k_1$. In this case, $s_y(x, y) = k(y) > k_1$, immediately yielding a contradiction.

On the other hand, if $y < \bar{y}$, then (85) implies that

$$\begin{aligned} k_0 - s_y(\bar{x}, \bar{y}) - [k(y) - s_y(\bar{x}, y)] &\geq \int_y^{\bar{y}} [k'(s) - s_{yy}(\bar{x}, s)] ds \\ &\geq \beta |\bar{y} - y| \end{aligned}$$

As $k_0 = s_y(\bar{x}, \bar{y})$, this means, for almost every $y < \bar{y}$, with y close to $\bar{y}$

$$s_y(\bar{x}, y) - s_y(x, y) = s_y(\bar{x}, y) - k(y) \geq \beta |\bar{y} - y|.$$

Therefore,

$$\begin{aligned} s_y(\bar{x}, \bar{y}) - s_y(x, \bar{y}) &= s_y(\bar{x}, y) - s_y(x, y) + \int_y^{\bar{y}} [s_{yy}(\bar{x}, t) - s_{yy}(x, t)] dt \\ &\geq \beta |\bar{y} - y| - B |\bar{y} - y| |\bar{x} - x| \\ &= |\bar{y} - y| (\beta - B |\bar{x} - x|) > 0. \end{aligned}$$

for $|x - \bar{x}|$ sufficiently small. This contradicts the assumption $x \in X_>(\bar{y}, k_0)$.

To summarize, we have shown that for $x \in X_>(\bar{y}, k_0)$ close to $\bar{x}$, we cannot have $F(x) > \bar{y}$ or $F(x) < \bar{y}$; we must therefore have $F(x) = \bar{y}$. As this set has positive mass, and $\nu(\{\bar{y}\}) = 0$, this violates mass balance, establishing that k cannot have a jump discontinuity. ■

Theorem 35 (Lipschitz differentiability of v) *Let $(u, v) = (v^s, u^{\tilde{s}})$ achieve the minimum (8) and solve (23)–(24) a.e. on Y with $i = 2$. If $Y' \subset Y$ and $P' = v'(Y')$ are regions satisfying the hypotheses imposed between (76) and (80), then $v \in C^{1,1}(Y')$.*

Proof. Setting $k = v'$ yields $k'(y) = v''(y) = q(y, k(y), g(y))$ almost everywhere. Choose $y_0 < y_1$ and denote $k(y_i) = k_i$ for $i = 1, 2$. Since $\partial_{\tilde{s}} v(y) = X_2(y, k(y), k'(y)) = X_2(y, k(y), q(y, k(y), g(y)))$ for a.e. y, mass balance and Propositions 31 and 34 combine to imply

$$
\begin{aligned}
U_g|y_1 - y_0| &\geq \int_{y_0}^{y_1} g(y)dy && (86)\\
&= \int_{\cup_{y\in[y_0,y_1]} X_2(y,k(y),q(y,k(y),g(y)))} f(x)d\mathcal{H}^m(x) && (87)\\
&\geq L_f vol[\cup_{y\in[y_0,y_1]} X_2(y, k(y), q(y, k(y), g(y)))] && (88)\\
&\geq L_f(C_1|k(y_0) - k(y_1)| - C_2|y_0 - y_1|). && (89)
\end{aligned}
$$

This is the desired conclusion. ■

Remark 36 (Conclusion and comparison to previous work) *Combined with Corollary 29 of Section 7, Theorem 35 identifies sets Y' on which any solution $v = v^{s\tilde{s}}$ to the local equation (23) for $i = 2$ must be $C^{2,1}$, or in fact smoother, depending on G_2. The main advance in these sections over the theory developed in [10] is that solutions are assumed only to satisfy the weaker $i = 2$ version of the local equation (23), rather that the stronger $i = 1$ version implied by the nestedness condition assumed in [10]. This is illustrated by the examples below. On the other hand, additional assumptions on s are needed here, including the enhanced twist condition, the third-order linear independence relation (62), and various conditions involving the geometry of the level sets of $x \mapsto D^2_{yy}s(x, y)$ and their intersection with level sets of $D_y s(x, y)$ implied by the fact that (66)–(68) are assumed positive.*

Example 37 (Revisiting the annulus to circle match again) *Recalling Example 7, note that*

$$
\begin{aligned}
D_x s(x, \theta) &= \hat{y}(\theta) := (\cos\theta, \sin\theta) \qquad \text{and}\\
D^2_{x\theta} s(x, \theta) &= \hat{y}(\theta)^\perp := (-\sin\theta, \cos\theta)
\end{aligned}
$$

are perpendicular unit vectors. For distinct $\theta, \bar\theta \in Y$ the linear independence required by the enhanced twist (Assumption 32) asserts $\hat y(\theta) - \hat y(\bar\theta) \neq a \hat y(\bar\theta)^\perp$ for any scalar $a \in \mathbf{R}$. Taking inner products with $\hat y(\bar\theta)$ yields $\hat y(\theta) \cdot \hat y(\bar\theta) \neq 1$ which is satisfied unless $\theta = \bar\theta \mod 2\pi$, or equivalently, unless $\theta = \bar\theta$ since both lie in $Y := (-\pi, \pi)$. Thus this surplus obeys the enhanced twist condition. The rest of the conditions are easy to check, so that solutions $v = v^{s\tilde s}$ to the $i = 2$ equation with $\|v\|_{C^{0,1}(Y)} < 1$ belong to $C^{1,1}(Y)$ by Theorem 35, and are in fact smoother under appropriate conditions by Remark 36. From Example 26, the required conditions include a smallness restriction $\|v\|_{C^{1,1}(Y)} < 2^{-1/2}$; when $\|v\|_{C^{0,1}(Y)} << 1$ is small, this follows directly from $v''(\theta) = q(\theta, v'(\theta), g(\theta))$ and the hypothesized continuity of q around $q(y, 0, 1/2\pi) = 0$.

This applies, for instance, to the problem where the marginals are uniform, solved explicitly by $v = 0$ in Example 7. More generally, when a uniformly small enough perturbation of the density $g(\theta) = 1/2\pi$ yields an a.e. solution $v = v^{s\tilde s}$ to the $i = 2$ equation, as in the next example, the above analysis can be applied to deduce smoothness of v. Note that neither example is nested, so the regularity theory developed in [10] does not apply.

Example 38 (Mapping the disk to a weighted interval) *Let $s(x, \theta) = x \cdot (\cos(\theta), \sin(\theta))$ on $X = B(1, 0) \subseteq \mathbf{R}^2$ and $Y = (-\pi, \pi)$. Let f be uniform and choose $\lambda \in (-1, 1)$. Then $v(\theta) = \lambda \cos(\theta)$ is $\tilde s$-convex and therefore solves the $i = 2$ (but not $i = 1$) equation when $g(\theta)$ is defined by*

$$g(\theta) = G_2(\theta, v'(\theta), v''(\theta)).$$

Here $X_1(\theta, v'(\theta))$ coincides with the line segment through $(\lambda, 0)$ making angle θ with the horizontal axis, while $X_2(\theta, v'(\theta), v''(\theta)) = \partial_{\tilde s} v(\theta)$ is the intersection of this line segment with $\{\theta x_2 \geq 0\}$. As in the preceding example, the conditions in Theorems 23 and 35 are easily verified, implying smoothness of v (in agreement with the explicit solution).

Appendices

Appendix A Formulas for partial derivatives

The partial derivatives of the functions defined in Lemma 27 are given almost everywhere by the following formulas, with $w(x,y) = |D^2_{xy}s|^{-1}$:

$$\begin{aligned}
A_p- \quad & \int_{X_2} a_p d\mathcal{H}^{m-1} = \frac{\partial}{\partial \tilde{p}}\bigg|_{\tilde{p}=p} \int_{W_=(y,\tilde{p})\cap Z_\le(y,q)} a(x,y,p,q)\hat{n}_W\cdot\hat{n}_W d\mathcal{H}^{m-1}(x) \\
&= \frac{\partial}{\partial\tilde{p}}\left[\int_{W_\le\cap Z_\le}\nabla\cdot(a\hat{n}_W)d\mathcal{H}^m - \int_{W_\le\cap Z_=} a\hat{n}_W\cdot\hat{n}_Z d\mathcal{H}^{m-1} - \int_{\overline{W_\le\cap Z_\le}\cap\partial X} a\hat{n}_W\cdot\hat{n}_X d\mathcal{H}^{m-1}\right]_{\tilde{p}=p} \\
&= \int_{W_=\cap Z_\le}\nabla\cdot(a\hat{n}_W)w d\mathcal{H}^{m-1} - \int_{W_=\cap Z_=}\frac{aw\hat{n}_W\cdot\hat{n}_Z}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_Z)^2}}d\mathcal{H}^{m-2} \\
&\quad - \int_{\overline{W_=\cap Z_\le}\cap\partial X}\frac{aw\hat{n}_W\cdot\hat{n}_X}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}}d\mathcal{H}^{m-2},
\end{aligned}$$

$$\begin{aligned}
A_q- \quad & \int_{X_2} a_q d\mathcal{H}^{m-1} = \frac{\partial}{\partial\tilde{q}}\bigg|_{\tilde{q}=q}\int_{W_=(y,p)\cap Z_\le(y,\tilde{q})} a(x,y,p,q)d\mathcal{H}^{m-1}(x) \\
&= \int_{W_=\cap Z_=}\frac{az}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_W)^2}}d\mathcal{H}^{m-2},
\end{aligned}$$

$$\begin{aligned}
A_y- \quad & \int_{X_2} a_y d\mathcal{H}^{m-1} + \int_{W_=\cap Z_=}\frac{azs_{yyy}}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_W)^2}}d\mathcal{H}^{m-2} = \frac{\partial}{\partial\tilde{y}}\bigg|_{\tilde{y}=y}\int_{W_=(\tilde{y},p)\cap Z_\le(y,q)} a(x,y,p,q)d\mathcal{H}^{m-1} \\
&= \frac{\partial}{\partial\tilde{y}}\left[\int_{\tilde{W}_\le\cap Z_\le}\nabla\cdot(a\hat{n}_{\tilde{W}})d\mathcal{H}^m - \int_{\tilde{W}_\le\cap Z_=} a\hat{n}_{\tilde{W}}\cdot\hat{n}_Z d\mathcal{H}^{m-1} - \int_{\overline{\tilde{W}_\le\cap Z_\le}\cap\partial X} a\hat{n}_{\tilde{W}}\cdot\hat{n}_X d\mathcal{H}^{m-1}\right]_{\tilde{y}=y} \\
&= \int_{W_\le\cap Z_\le}\nabla\cdot\Big(a\frac{\partial\hat{n}_W}{\partial y}\Big)d\mathcal{H}^m - \int_{W_\le\cap Z_=} a\frac{\partial\hat{n}_W}{\partial y}\cdot\hat{n}_Z d\mathcal{H}^{m-1} - \int_{\overline{W_\le\cap Z_\le}\cap\partial X} a\frac{\partial\hat{n}_W}{\partial y}\cdot\hat{n}_X d\mathcal{H}^{m-1} \\
&\quad - \int_{W_=\cap Z_\le}\nabla\cdot(a\hat{n}_W)ws_{yy}d\mathcal{H}^{m-1} + \int_{W_=\cap Z_=}\frac{aws_{yy}\hat{n}_W\cdot\hat{n}_Z}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_Z)^2}}d\mathcal{H}^{m-2} \\
&\quad + \int_{\overline{W_=\cap Z_\le}\cap\partial X}\frac{aws_{yy}\hat{n}_W\cdot\hat{n}_X}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}}d\mathcal{H}^{m-2}
\end{aligned}$$

where $\frac{\partial}{\partial y}\hat{n}_W = (\hat{n}_Z - \hat{n}_W(\hat{n}_W\cdot\hat{n}_Z))w/z$,

$$B_p = \int_{W_\le(y,p)\cap Z_\le(y,q)} b_p d\mathcal{H}^m(x) + \int_{W_=(y,p)\cap Z_\le(y,q)} bw d\mathcal{H}^{m-1}(x),$$

$$B_q = \int_{W_\le(y,p)\cap Z_\le(y,q)} b_q d\mathcal{H}^m(x) + \int_{W_\le(y,p)\cap Z_=(y,q)} bz d\mathcal{H}^{m-1}(x),$$

$$B_y = \int_{W_{\leq}(y,p)\cap Z_{\leq}(y,q)} b_y d\mathcal{H}^m(x) - \int_{W_{=}(y,p)\cap Z_{\leq}(y,q)} bws_{yy} d\mathcal{H}^{m-1}(x) - \int_{W_{\leq}(y,p)\cap Z_{=}(y,q)} bzs_{yyy} d\mathcal{H}^m(x),$$

$$C_p = \int_{W_{\leq}(y,p)\cap Z_{=}(y,q)} c_p d\mathcal{H}^{m-1}(x) + \int_{W_{=}(y,p)\cap Z_{=}(y,q)} cw d\mathcal{H}^{m-2}(x),$$

$$\begin{aligned} C_q = &\int_{W_{\leq}(y,p)\cap Z_{=}(y,q)} c_q d\mathcal{H}^{m-1}(x) + \int_{W_{\leq}(y,p)\cap Z_{=}(y,q)} \nabla\cdot(c\hat{n}_Z) z d\mathcal{H}^{m-1}(x) \\ &- \int_{W_{=}(y,p)\cap Z_{=}(y,q)} \frac{cz\hat{n}_Z\cdot\hat{n}_W}{\sqrt{1-(\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z})^2}} d\mathcal{H}^{m-2}(x) - \int_{\overline{W_{\leq}\cap Z_{=}}\cap\partial X} \frac{cz\hat{n}_Z\cdot\hat{n}_X}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}} d\mathcal{H}^{m-2}, \end{aligned}$$

$$\begin{aligned} C_y = &\int_{W_{\leq}\cap Z_{=}} c_y d\mathcal{H}^{m-1} - \int_{W_{=}\cap Z_{=}} \frac{cws_{yy}}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_W)^2}} d\mathcal{H}^{m-2} \\ &+ \int_{W_{\leq}\cap Z_{\leq}} \nabla\cdot(c\frac{\partial\hat{n}_Z}{\partial y}) d\mathcal{H}^m - \int_{W_{=}\cap Z_{\leq}} c\frac{\partial\hat{n}_Z}{\partial y}\cdot\hat{n}_W d\mathcal{H}^{m-1} - \int_{\overline{W_{\leq}\cap Z_{\leq}}\cap\partial X} c\frac{\partial\hat{n}_Z}{\partial y}\cdot\hat{n}_X d\mathcal{H}^{m-1} \\ &- \int_{W_{\leq}\cap Z_{=}} \nabla\cdot(c\hat{n}_Z) zs_{yyy} d\mathcal{H}^{m-1} + \int_{W_{=}\cap Z_{=}} \frac{czs_{yyy}\hat{n}_W\cdot\hat{n}_Z}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_Z)^2}} d\mathcal{H}^{m-2} \\ &+ \int_{\overline{W_{\leq}\cap Z_{=}}\cap\partial X} \frac{czs_{yyy}\hat{n}_Z\cdot\hat{n}_X}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}} d\mathcal{H}^{m-2}, \end{aligned}$$

$$\begin{aligned} A_p^\partial = &\int_{(\overline{W_{=}\cap Z_{\leq}})\cap\partial X)} a_p^\partial d\mathcal{H}^{m-2}(x) + \int_{(\overline{W_{=}\cap Z_{\leq}})\cap\partial X)} \frac{w}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}} \nabla_{\partial X}\cdot(a^\partial\hat{n}_{\partial,W}) d\mathcal{H}^{m-2}(x) \\ &- \int_{(\overline{W_{=}\cap Z_{=}})\cap\partial X)} a^\partial\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z} \frac{w}{\sqrt{[1-(\hat{n}_W\cdot\hat{n}_X)^2][1-(\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z})^2]}} d\mathcal{H}^{m-3}(x), \end{aligned}$$

$$A_q^\partial = \int_{(\overline{W_{=}\cap Z_{\leq}})\cap\partial X)} a_q^\partial d\mathcal{H}^{m-2}(x) + \int_{(\overline{W_{=}\cap Z_{=}})\cap\partial X)} a^\partial \frac{z}{\sqrt{[1-(\hat{n}_Z\cdot\hat{n}_X)^2][1-(\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z})^2]}} d\mathcal{H}^{m-3}(x),$$

$$\begin{aligned} A_y^\partial = &\int_{(\overline{W_{=}\cap Z_{\leq}})\cap\partial X)} a_y^\partial d\mathcal{H}^{m-2}(x) - \int_{(\overline{W_{=}\cap Z_{=}})\cap\partial X)} a^\partial \frac{zs_{yyy}}{\sqrt{[1-(\hat{n}_Z\cdot\hat{n}_X)^2][1-(\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z})^2]}} d\mathcal{H}^{m-3}(x) \\ &+ \int_{(\overline{W_{\leq}\cap Z_{\leq}})\cap\partial X)} \nabla_{\partial X}\cdot(a^\partial\frac{\partial\hat{n}_{\partial,W}}{\partial y}) d\mathcal{H}^{m-1}(x) - \int_{(\overline{W_{\leq}\cap Z_{=}}\cap\partial X)} a^\partial\frac{\partial\hat{n}_{\partial,W}}{\partial y}\cdot\hat{n}_{\partial,Z} d\mathcal{H}^{m-2}(x) \\ &+ \int_{(\overline{W_{=}\cap Z_{\leq}})\cap\partial X)} \frac{ws_{yy}}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}} \nabla_{\partial X}\cdot(a^\partial\hat{n}_{\partial,W}) d\mathcal{H}^{m-2}(x) \\ &- \int_{(\overline{W_{=}\cap Z_{=}})\cap\partial X)} \frac{ws_{yy}}{\sqrt{[1-(\hat{n}_W\cdot\hat{n}_X)^2][1-(\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z})^2]}} a^\partial\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z} d\mathcal{H}^{m-3}(x), \end{aligned}$$

where $\hat{n}_{\partial,W} := \frac{\hat{n}_W - (\hat{n}_W \cdot \hat{n}_X)\hat{n}_X}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}}$ and $\hat{n}_{\partial,Z} := \frac{\hat{n}_Z - (\hat{n}_Z \cdot \hat{n}_X)\hat{n}_X}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}}$ are the outward unit normals to $\bar{W}_{\leq} \cap \partial X$ and $\bar{Z}_{\leq} \cap \partial X$ in ∂X, respectively and

$$B_p^\partial = \int_{(\overline{W_\leq \cap Z_\leq}) \cap \partial X} b_p^\partial d\mathcal{H}^{m-1}(x) + \int_{(\overline{W_= \cap Z_\leq}) \cap \partial X} \frac{b^\partial w}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}} d\mathcal{H}^{m-2}(x),$$

$$B_q^\partial = \int_{(\overline{W_\leq \cap Z_\leq}) \cap \partial X} b_q^\partial d\mathcal{H}^{m-1}(x) + \int_{(\overline{W_\leq \cap Z_=}) \cap \partial X} \frac{b^\partial z}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}} d\mathcal{H}^{m-2}(x),$$

$$\begin{aligned}
B_y^\partial = &\int_{(\overline{W_\leq \cap Z_\leq}) \cap \partial X} b_y^\partial d\mathcal{H}^{m-1}(x) - \int_{(\overline{W_= \cap Z_\leq}) \cap \partial X} \frac{b^\partial w s_{yy}}{\sqrt{1-(\hat{n}_W\cdot\hat{n}_X)^2}} d\mathcal{H}^{m-2}(x) \\
&- \int_{(\overline{W_\leq \cap Z_=}) \cap \partial X} \frac{b^\partial z s_{yyy}}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}} d\mathcal{H}^{m-2}(x),
\end{aligned}$$

$$C_p^\partial = \int_{(\overline{W_\leq \cap Z_=}) \cap \partial X} c_p^\partial d\mathcal{H}^{m-2}(x) + \int_{(\overline{W_= \cap Z_=}) \cap \partial X} c^\partial \frac{w}{\sqrt{[1-(\hat{n}_W\cdot\hat{n}_X)^2][1-(\hat{n}_{\partial,Z}\cdot\hat{n}_{\partial,W})^2]}} d\mathcal{H}^{m-3}(x),$$

$$\begin{aligned}
C_q^\partial = &\int_{(\overline{W_\leq \cap Z_=}) \cap \partial X} c_q^\partial d\mathcal{H}^{m-2}(x) + \int_{(\overline{W_\leq \cap Z_=}) \cap \partial X} \frac{z}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}} \nabla_{\partial X} \cdot (c^\partial \hat{n}_{\partial,Z}) d\mathcal{H}^{m-2}(x) \\
&- \int_{(\overline{W_= \cap Z_=}) \cap \partial X} c^\partial \hat{n}_{\partial,Z} \cdot \hat{n}_{\partial,W} \frac{z}{\sqrt{[1-(\hat{n}_Z\cdot\hat{n}_X)^2][1-(\hat{n}_{\partial,Z}\cdot\hat{n}_{\partial,W})^2]}} d\mathcal{H}^{m-3}(x),
\end{aligned}$$

and

$$\begin{aligned}
C_y^\partial = &\int_{(\overline{W_\leq \cap Z_=}) \cap \partial X} c_y^\partial d\mathcal{H}^{m-2}(x) - \int_{(\overline{W_= \cap Z_=}) \cap \partial X} c^\partial \frac{w s_{yy}}{\sqrt{[1-(\hat{n}_W\cdot\hat{n}_X)^2][1-(\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z})^2]}} d\mathcal{H}^{m-3}(x) \\
&+ \int_{(\overline{W_\leq \cap Z_\leq}) \cap \partial X} \nabla_{\partial X} \cdot (c^\partial \frac{\partial \hat{n}_{\partial,Z}}{\partial y}) d\mathcal{H}^{m-1}(x) - \int_{(\overline{W_= \cap Z_\leq} \cap \partial X)} c^\partial \frac{\partial \hat{n}_{\partial,Z}}{\partial y} \cdot \hat{n}_{\partial,W} d\mathcal{H}^{m-2}(x) \\
&- \int_{(\overline{W_\leq \cap Z_=}) \cap \partial X} \frac{z s_{yyy}}{\sqrt{1-(\hat{n}_Z\cdot\hat{n}_X)^2}} \nabla_{\partial X} \cdot (c^\partial \hat{n}_{\partial,Z}) d\mathcal{H}^{m-1}(x) \\
&+ \int_{(\overline{W_= \cap Z_=}) \cap \partial X} \frac{z s_{yyy}}{\sqrt{[1-(\hat{n}_Z\cdot\hat{n}_X)^2][1-(\hat{n}_{\partial,W}\cdot\hat{n}_{\partial,Z})^2]}} c^\partial \hat{n}_{\partial,W} \cdot \hat{n}_{\partial,Z} d\mathcal{H}^{m-3}(x).
\end{aligned}$$

References

[1] Luigi Ambrosio and Gianni Dal Maso. A general chain rule for distributional derivatives. *Proc. Amer. Math. Soc.*, 108:691–702, 1990.

[2] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

[3] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44:375–417, 1991.

[4] H. Brezis. Remarks on the Monge-Kantorovich problem in the discrete setting. *C. R. Math. Acad. Sci. Paris*, 356:207–213, 2018.

[5] L. Caffarelli. Allocation maps with general cost functions. In P. Marcellini et al, editor, *Partial Differential Equations and Applications*, number 177 in Lecture Notes in Pure and Appl. Math., pages 29–35. Dekker, New York, 1996.

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

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

[8] Pierre-André Chiappori, Robert McCann, and Brendan Pass. Transition to nestedness in multi- to one-dimensional optimal transport. *European J. Appl. Math.*, 30(6):1220–1228, 2019.

[9] P.-A. Chiappori, R.J. McCann, and B. Pass. Multidimensional matching. *Preprint*.

[10] P.-A. Chiappori, R.J. McCann, and B. Pass. Multi- to one-dimensional optimal transport. *Comm. Pure Appl. Math.*, 70:2405–2444, 2017.

[11] F.H. Clarke. *Optimization and nonsmooth analysis.* Wiley, New York, 1983.

[12] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.

[13] M.J.P Cullen and R.J. Purser. An extended Lagrangian model of semi-geostrophic frontogenesis. *J. Atmos. Sci.*, 41:1477–1497, 1984.

[14] G. De Philippis and A. Figalli. The Monge-Ampère equation and its link to optimal transportation. *Bull. Amer. Math. Soc. (N.S.)*, 51(4):527–580, 2014.

[15] G. De Philippis and A. Figalli. Partial regularity for optimal transport maps. *Publ. Math. Inst. Hautes Études Sci.*, 121:81–112, 2015.

[16] P. Delanoë. Classical solvability in dimension two of the second boundary-value problem associated with the Monge-Ampère oper-

ator. *Ann. Inst. H. Poincarè Anal. Non Linèaire*, 8:443–457, 1991.

[17] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. Stud. Adv. Math. CRC Press, Boca Raton, 1992.

[18] A. Figalli. Regularity properties of optimal maps between nonconvex domains in the plane. *Comm. Partial Differential Equations*, 35:465–479, 2010.

[19] A. Figalli and N. Gigli. Local semiconvexity of Kantorovich potentials on non-compact manifolds. *ESAIM Control Optim. Calc. Var.*, 17:648–653, 2011.

[20] A. Figalli and Y.-H. Kim. Partial regularity of Brenier solutions of the Monge-Ampère equation. *Discrete Contin. Dyn. Syst.*, 28:559–565, 2010.

[21] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.*, 177:113–161, 1996.

[22] W. Gangbo and R.J. McCann. Shape recognition via Wasserstein distance. *Quart. Appl. Math.*, 58:705–737, 2000.

[23] Victor Guillemin and Alan Pollack. *Differential Topology*. Prentice-Hall, Englewood Cliffs, NJ, 1974.

[24] S. Hofmann, M. Mitrea and M. Taylor. Singular integrals and elliptic boundary problems on regular Semmes-Kenig-Toro domains. *Int. Math. Res. Not. IMRN*, 14:2567–2865, 2010.

[25] Y.-H. Kim and J. Kitagawa. Prohibiting isolated singularities in optimal transport. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 16:277–290, 2016.

[26] M. Knott and C.S. Smith. On the optimal mapping of distributions. *J. Optim. Theory Appl.*, 43:39–49, 1984.

[27] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Res. Logist. Quart.*, 2:83–97, 1955.

[28] V.L. Levin. Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem. *Set-valued Anal.*, 7:7–32, 1999.

[29] G. Loeper. On the regularity of solutions of optimal transportation problems. *Acta Math.*, 202:241–283, 2009.

[30] J. Lott. On tangent cones in Wasserstein space. *Proc. Amer. Math. Soc.*, 145:3127–3136, 2017.

[31] X.-N. Ma, N. Trudinger and X.-J. Wang. Regularity of potential functions of the optimal transportation problem. *Arch. Rational Mech. Anal.*, 177:151–183, 2005.

[32] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128:153–179, 1997.

[33] R.J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11:589–608, 2001.

[34] R.J. McCann and N. Guillen. Five lectures on optimal transportation: geometry, regularity, and applications. In G. Dafni et al, editor, *Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the Séminaire de Mathématiques Supérieure (SMS) Montréal 2011*, pages 145–180. American Mathematical Society, Providence, 2013.

[35] L. Nenna and B. Pass. Variational problems involving unequal dimensional optimal transport. To appear in *J. Math. Pures Appl.*

[36] L. Rüschendorf and S.T. Rachev. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, 32:48–54, 1990.

[37] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Birkhäuser/Springer, Cham, 2015.

[38] N.S. Trudinger and X.-J. Wang. On the second boundary value problem for Monge-Ampère type equations and optimal transportation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 8:1–32, 2009.

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

[40] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, New York, 2009.

[41] X.-J. Wang. On the design of a reflector antenna II. *Calc. Var. Partial Differential Equations*, 20:329–341, 2004.