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 manifolds 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. 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. © by the authors, May 30, 2019

†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 [10] [23] [32] and the celebrated work of Brenier [2] [3], it has been well-understood [29] that for the quadratic cost \( c(x, y) = \frac{1}{2}|x - y|^2 \) on \( \mathbb{R}^n \), 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 \( \mathbb{R}^n \), the unique optimal map between them, \( F = Du \), is given by a convex solution \( u \) to the boundary value problem

\[
\begin{align*}
g \circ Du \det D^2u &= f \quad \text{[a.e.],} \\
Du &\in \text{spt } g \quad \text{[a.e.],}
\end{align*}
\]

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

\[
\begin{align*}
f \circ Dv \det D^2v &= g \quad \text{[a.e.],} \\
Dv &\in \text{spt } f \quad \text{[a.e.].}
\end{align*}
\]

Notice the quadratic cost implicitly requires \( x \) and \( y \) to live in the same space. Subsequent work of Ma, Trudinger and Wang [28] 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] [18] [25] [30] [37]. The purpose of the present article is to explore what can be said when \( x \in \mathbb{R}^m \) and \( y \in \mathbb{R}^n \) live in spaces with different dimensions \( m > n \), as in e.g. [19] [8] [27].

Although the symmetry between \( x \) and \( y \) is destroyed, the duality theorem from linear programming, [24] [33] [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 [15] [16]. 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 Chiapponi [8] 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 sets $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$ equipped with Borel probability densities $f$ and $g$. We say $F : X \rightarrow Y$ pushes $f$ forward to $g = F_\# f$ if $F$ is Borel and

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

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

$$g(y) = \int_{F^{-1}(y)} f(x) JF(x)^{-1/2} dH^{m-n}(x)$$

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

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

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

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 [36], 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

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

where the minimum is taken over pairs $(u, v) \in L^1(f) \oplus L^1(g)$ satisfying $u + v \geq s$ throughout $X \times Y$. Dual minimizers of the form $(u, v) = (v^s, u^\bar{s})$ are known to exist [36], where

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

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

To motivate our first result, let $X \subset \mathbb{R}^m$ be open and $Y \subset \mathbb{R}^n$ be open and bounded, and $s \in C^2(X \times \bar{Y})$ twisted and non-degenerate,
meaning in addition to the injectivity of \( y \in \bar{Y} \mapsto D_x s(x, y) \) mentioned above that \( D_{xy}^2 s(x, y) \) has maximal rank throughout \( X \times \bar{Y} \). Suppose \( F \) maximizes the primal problem (7) and \( (u, v) = (v^*, u^*) \) are \( s \)-convex payoffs minimizing the dual problem (8). Then \( u(x) + v(y) - s(x, y) \geq 0 \) on \( X \times \bar{Y} \), with equality on \( \text{graph}(F) \). Thus

\[
F^{-1}(y) \subset \partial_v v(y) := \{ x \in X \mid s(x, y) - v(y) = \sup_{y' \in \bar{Y}} s(x, y') - v(y') \}. 
\]

(10)

Since \( s \in C^2(X \times \bar{Y}) \), \( u \) and \( v \) admit second-order Taylor expansions Lebesgue a.e. as in e.g. [18] [36], and the first- and second-order conditions for equality on \( \text{graph}(F) \) imply

\[
Dv(F(x)) = D_y s(x, F(x)) \quad [f\text{-a.e.}] \quad \text{and} \quad D^2v(F(x)) \geq D^2_{yy} s(x, F(x)) \quad [f\text{-a.e.}]. 
\]

(12)

(13)

Differentiating the first-order condition yields

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

(14)

as in e.g. [28]. Since \( D^2_{xy} s \) has full-rank, when \( F \) happens to be Lipschitz we identify its Jacobian \( f\text{-a.e.} \) as

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

(15)

In this case we can rewrite (3) in the form

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

(16)

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 for a.e. \( y \), the containment (10) is saturated up to an \( \mathcal{H}^{m-n} \) negligible set. Thus we arrive at

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

(17)

This 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 (17). However, unlike the case \( m = n \), it is badly nonlocal since the domain of integration \( \partial s v(y) \) defined in (10) 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 (17) if and only if \( v \) combines with its conjugate \( u = v^s \) to minimize \( \mathcal{H} \); see Corollary 2 of [2]. Since the optimal map \( F \) can be recovered from the first-order condition

\[
D_x s(x, F(x)) = Du(x),
\]

analogous to (12), this shows Monge’s problem has been reduced to the solution of the partial differential equation (17) for the \( s \)-convex scalar function \( v \). 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. [33].

Although non-locality makes this equation a challenge to solve, it turns out there is a class of problems for which (17) can be replaced by a local partial differential equation, as follows. Introduce the \( m - n \) dimensional submanifold

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

of \( X \) and its closed subset

\[
X_2(y, p, Q) := \{ x \in X_1(y, p) \mid D_{yy}^2 s(x, y) \leq Q \}. \tag{19}
\]

Now (12)–(13) imply

\[
\partial s v(y) \subset X_2(y, Dv(y), D^2 v(y)) \subset X_1(y, Dv(y)) \tag{20}
\]

for all \( y \in \text{dom} \ D^2 v \), 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 (17) in the form

\[
G(y, Dv(y), D^2 v(y)) = g(y) \quad \text{[a.e.],} \tag{21}
\]

where

\[
G(y, p, Q) := G_i(y, p, Q) \tag{22}
:= \int_{X_i(y, p, Q)} \frac{\det [Q - D_{yy}^2 s(x, y)]}{\sqrt{\det D_{xy}^2 s(x, y)(D_{xy}^2 s(x, y))^T}} f(x) d\mathcal{H}^{m-n}(x)
\]

and either \( i = 1 \) or \( i = 2 \).
Our second result states any classical \( s \)-convex solution \( v \in C^2(Y) \) to either local problem \((21)\) also solves the nonlocal one \((17)\); Corollary 3. Assuming connectedness of \( X_1(y, Dv(y)) \), we show such a solution exists and satisfies the uniform ellipticity criterion \( D^2 v - D^2_{yy} s > 0 \) if and only if the dual minimization \((5)\) admits a \( C^2 \) solution; Theorem 4 of \( \S 3 \).

For an \( n = 1 \) dimensional target, 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 [8]. There the ordinary differential equation \((21)\) is also analyzed to show \( v \) inherits smoothness from suitable conditions on the data \((s, f, g)\) in this so-called nested case. The existence of a solution to \((21)\) with \( i = 1 \) extends the notion of nestedness from \( n = 1 \) to higher dimensions.

We go on to show that the operator \( G_2 \) is degenerate elliptic in \( \S 5 \) 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 \((21)\) with \( i = 2 \) from \( C^{2,\alpha} \) regularity in Theorem 10 provided \( G_2 \) is sufficiently smooth. In Theorem 11 of \( \S 6 \) 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 18 of \( \S 7 \) 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 [28] [34] (which built on earlier work of Caffarelli [6] [7], Delanoe [13] and Urbas [35]), and for \( n = 1 \) in the nested case [8]; to these we now add the non-nested cases which satisfy the local equations \((21)\)–\((22)\) with \((n, i) = (1, 2)\), resolved in \( \S 8 \) below.

When regularity fails for \( m = n \) the size of the singular set has been estimated by DePhilippis and Figalli [12], building on work of Figalli [15] with Kim [17]; for related results see Kitagawa and Kim [22] and the survey [11].

2 A nonlocal partial differential equation for optimal transport

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

\[
\int_Y \psi(y) dv(y) = \int_X \psi(F(x)) f(x) dx
\]

for all bounded Borel functions \( \psi \in L^\infty(Y) \). This definition extends \((5)\) to the case where \( \nu \) need not be absolutely continuous with respect to Lebesgue; however when \( \nu \) is absolutely continuous with Lebesgue density
with respect to Lebesgue: its density given for Lebesgue a.e. $f$

where

$s$ guarantees living on the disjoint sets $Y$

we can then deduce $y$ uniquely from $x$ and $p$, in which case we write

$y = s\text{-exp}_p := D_x s(x, \cdot)^{-1}(p)$. The non-degeneracy of $s$ (full-rank of $D^2_x s$) guarantees $s\text{-exp}$ 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 (18) 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 $\text{dom } Du$. Similarly we denote the set of points where $F : X \to \tilde{Y}$ is approximately differentiable by $\text{dom } DF$, and the set where $u$ admits a second order Taylor expansion by $\text{dom } D^2 u$. When $s$ is non-degenerate and twisted, (18) implies $\text{dom } DF = \text{dom } D^2 u$. A function $u : X \subset \mathbb{R}^m$ is said to be semiconvex if there exists $k \in \mathbb{R}$ such that $u(x) + k|x|^2$ is the restriction to $X$ of a convex function on $\mathbb{R}^m$.

**Theorem 1 (Properties of potential maps)** Fix $m \geq n$, open sets $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$ with $Y$ bounded, and $s \in C^2(X \times \tilde{Y})$ (so $\|s\|_{C^2(X \times \tilde{Y})} < \infty$) twisted and non-degenerate. Any pair $(u, v) = (u^s, u^g)$ of $s$-conjugate functions (9) are semiconvex, Lipschitz, and have second-order Taylor expansions Lebesgue a.e. The map $F : \text{dom } Du \to \tilde{Y}$ satisfying (18) is unique and differentiable Lebesgue a.e. Decompose $\tilde{Y}$ into $Y_+ := \text{dom } D^2 v \subset \tilde{Y}$ and $Y_- = \tilde{Y} \setminus Y_+$ and set $X_+ := F^{-1}(Y_+)$. The Jacobian $JF(x) := \det^{1/2}[DF(x)DF(x)^T]$ is positive on $X_+ \cap \text{dom } DF$ and given there by

$$JF(x) = \frac{\sqrt{\det[D^2_{xy}s(x, F(x))D^2_{yx}s(x, F(x))^T]}}{\det[D^2v(F(x)) - D^2_{yy}s(x, F(x))]}. \quad (24)$$

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 for Lebesgue a.e. $y \in \tilde{Y}$ by

$$g_+(y) = \int_{F^{-1}(y)} \frac{f_+(x)}{JF(x)} dH^{m-n}(x) \quad (25)$$

$$= \int_{\partial_1 v(y)} \frac{\det[D^2v(y) - D^2_{yy}s(x, y)]}{\sqrt{\det D^2_{xy}s(x, y)(D^2_{xy}s(x, y))^T}} f(x) dH^{m-n}(x). \quad (26)$$

7
Proof. It is well-known that \( u = v^\circ \) and \( v = u^\circ \) are Lipschitz and semiconvex \cite{31} Lemma 3.1: they inherit distributional bounds such as \(|Du| \leq \sup_Y |D_x s|\) and \(D^2 u \geq \inf_Y D^2_{xx} s\) from \( s \in C^2\). This implies they extend continuously to \( \bar{X} \) and \( \bar{V} \), where they are twice differentiable a.e. by Alexandrov’s theorem \cite{36} Theorem 14.25; indeed, for \( x_0 \in \text{dom} \, D^2 u \) we have
\[
0 = \lim_{x \to x_0} \sup_{p \in Du(x)} \frac{p - Du(x_0) - D^2 u(x_0)(x - x_0)}{|x - x_0|} \tag{27}
\]
which asserts differentiability (rather than just approximate differentiability) of \( Du \) at \( x_0 \).

Recall \( u(x) + v(y) - s(x, y) \geq 0 \) on \( X \times \bar{Y} \). For each \( x \in \text{dom} \, Du \) at least one \( y \in \bar{Y} \) produces equality, since the maximum \cite{19} defining \( v^\circ(x) \) 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 \( x \in \text{dom} \, F \), and the map \( s\text{-exp} \) is well-defined and continuously differentiable in a neighbourhood of \( (x, Du(x)) \) by the twist and non-degeneracy of \( s \). Recall for any \( \epsilon > 0 \) the semiconvex function \( u \) agrees with a \( C^2 \) smooth function \( u_\epsilon \) outside a set of volume \( \epsilon \). As a result we see the extension \( \bar{F} \) is \( C^1 \) in a neighbourhood of \( \text{dom} \, Du \), except on a set of arbitrarily small volume, hence is countably Lipschitz (and approximately differentiable Lebesgue a.e.) The fact that it is actually differentiable a.e. follows from \( s\text{-exp} \in C^1 \) and (27).

Since \( u(x) + v(y) - s(x, y) \geq 0 \) vanishes at \( (x, F(x)) \in X \times \bar{Y} \) for each \( x \in X_+ \), we can differentiate (12) if \( x \in \text{dom} \, DF \) to obtain (14). 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 (13) establishes (24).

Decomposing a probability density \( f = f_+ + f_- \) on \( X \) into \( f_\pm = f_1 X_\pm \), the statements of mutually singularity follow from \( Y_+ \cap Y_- = \emptyset = X_+ \cap X_- \). Now decompose \( X \setminus X_\infty = \cup_{i=1}^{\infty} X_i \) into countably many disjoint Borel sets \( X_i \subset \mathbb{R}^m \) on which \( F \) is \( C^1 \) with \( JF(x) > 1/i \) on \( X_i \), plus an \( f_+ \) negligible set \( X_\infty \). Let \( f_i = f_1 X_i \) denote the restriction of \( f_+ \) to \( X_i \), and \( g_i := \bar{F}_i \circ \# f_i \) the density of the push-forward of \( f_i \). It costs no generality to assume \( f = 0 \) on \( X_\infty \cup \partial X \). Let \( F_i \) denote a Lipschitz extension of \( F \) from \( X_i \) to \( \mathbb{R}^m \). For each \( \phi \in L^1(\mathbb{R}^m) \), the co-area formula \cite{14} §3.4.3 implies
\[
\int_{X_i} \phi(x) JF_i(x) dx = \int_{\mathbb{R}^n} dy \int_{X_i \cap F_i^{-1}(y)} \phi d\mathcal{H}^{m-n}.
\]
Given \( \psi \in L^\infty(\mathbb{R}^n) \) with bounded support ensures \( \phi = f_i \psi \circ F_i / JF_i \in \).
Recalling $F_i = F$ on $X_i$, we infer 

$$g_i(y) = \int_{F^{-1}(y)} f_i(x) \frac{J_F(x)}{f_i} d\mathcal{H}^{m-n}(x).$$

a.e. since $\psi \in L^\infty$ had bounded support but was otherwise arbitrary.

Summing on $i$, the disjointness of $X_i$ and the fact that $f = 0$ on $X_\infty$ yields (25).

Now, $y \in Y_+$ implies $f = f_+$ on $F^{-1}(y) \subset X_+$. Since $Y_+ = \text{dom } D^2v$ has full measure in $Y$, we can replace $f$ by $f_+$ in (25). On the other hand, $X_+ \subset \partial s v(Y_+)$ with the difference satisfying $\partial s v(Y_+) \setminus X_+ \subset X \setminus \text{dom } Du$. This shows $f$ vanishes on $\partial s v(y) \setminus F^{-1}(y)$. Thus we can also replace $F^{-1}(y)$ by $\partial s v(y)$. Finally, (24) relates (25) to (26).

**Corollary 2 (Equivalence of optimal transport to nonlocal PDE)**

Under the hypotheses of Theorem 1, let $f$ and $g$ denote probability densities on $X$ and $Y$. If $v = v^{s\tilde{s}}$ satisfies the nonlocal equation (17) then $(v^s, v)$ minimize Kantorovich’s dual problem (8). Conversely, if $(u, v) = (v^s, u^\tilde{v})$ minimize (8) then $v$ satisfies (17).

**Proof.** First suppose $v = v^{s\tilde{s}}$ satisfies the nonlocal PDE (17). Setting $u = v^s$ implies for each $x \in \text{dom } Du$ the inequality

$$u(x) + v(y) - s(x, y) \geq 0$$

(28)

is saturated by some $y \in \tilde{Y}$. Identifying $F(x) = y$ we have the first-order condition (18), whence $F = s\exp \circ Du$ on $\text{dom } Du$. 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

$$\int_X u f + \int_Y v g = \int_X s(x, F(x)) f(x) dx,$$

which in turn shows $F$ maximizes (7) and $(u, v)$ minimizes (8) as desired. Comparing (17) with (26) we see $g_+ = g$ is a probability measure, hence
has the same total mass as \( f \). This implies \( g_- = 0 \) and \( F_\# f = g \) as desired.

Conversely, suppose \((u,v) = (v^s, u^s)\) minimizes (8). Since twistedness of \( s \) implies (7) is attained, there is some map \( F : X \to Y \) pushing \( f \) forward to \( g \) such that (28) becomes an equality \( f \text{-a.e. on } \text{Graph}(F) \). This ensures \( F = s \text{-exp} \circ Du \) holds \( f \text{-a.e.} \). Since \( Y_+ := \text{dom } D^2 v \subset \bar{Y} \) 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 (17) follows from (26) as desired. ■

**Corollary 3 (Optimal transport via local PDE)** Under the hypotheses of Theorem 1, let \( f \) and \( g \) denote probability densities on \( X \) and \( Y \). Fix \( i = 2 \) and let \( v = v^{s\bar{s}} \) have the property that \((s \text{-exp} \circ Dv^s)_# f \) vanishes on \( \bar{Y} \setminus \text{dom } D^2 v \) (as when \( e.g. \ v \in C^2(\bar{V}) \)). If the local equation (21) holds \( \mathcal{H}^n \text{-a.e.} \) then \((v^s, v)\) minimize Kantorovich’s dual problem (5).

**Proof.** Fix \( i = 2 \) and suppose \( v = v^{s\bar{s}} \) satisfies the local PDE (21). As in the preceding proof, setting \( u = v^s \) implies for each \( x \in \text{dom } Du \) the inequality

\[
    u(x) + v(y) - s(x,y) \geq 0
\]

is saturated by some \( y \in \bar{Y} \). Setting \( F(x) = y \) we have the first-order condition (13), whence \( F = s \text{-exp} \circ Du \) on \( \text{dom } Du \).

The present hypotheses assert the \( Y_+ = \text{dom } D^2 v \) forms a set of full measure for \( F_\# f \). Thus \( f_- = 0 \), while \( f = f_+ \) and \( g_+ \) are both probability densities in Theorem 1. Recalling \( \partial_s v(y) \subset X_2(y, Dv(y), D^2 v(y)) \) from (20), we deduce \( g \geq g_+ \) by comparing (21) with (26). Since both densities integrate to 1, this implies \( g = g_+ \) a.e. Thus (17) is actually satisfied and Corollary 2 asserts \((v^s, v)\) minimizes (8). ■

### 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 (21) to admit solutions, it is sufficient that the Kantorovich dual problem have a smooth minimizer \((u, v)\), with connected potential indifference sets \( X_i(y, Dv(y), D^2 v(y)) \) — in which case \( v \) also solves (21).

**Theorem 4 (When a smooth minimizer implies nestedness)** Fix \( m \geq n \), probability densities \( f \) and \( g \) on open sets \( X \subset \mathbb{R}^m \) and \( Y \subset \mathbb{R}^n \) with \( Y \) bounded, and \( s \in C^2(X \times \bar{Y}) \) twisted and non-degenerate. Let \( i \in \{1, 2\} \). If \((u, v) = (v^s, u^\bar{s}) \in C^2(X) \times C^2(Y)\) minimizes the Kantorovich dual (8) and \( X_i(y, Dv(y), D^2 v(y)) \) is connected for \( g \text{-a.e. } y \in Y \), then the local equation (21) holds \( g \text{-a.e.} \).
Proof. Corollary 2 implies $v$ solves the non-local equation (17). The local equation $G = g$ follows from equality in the inclusion

$$\partial_s v(y) \subset X_i(y, Dv(y), D^2v(y))$$

from (20) for $\mathcal{H}^n$-a.e. $y$.

We now derive this equality for all $y' \in Y$ with $\partial_s v(y')$ non-empty and $X'_i := X_i(y', Dv(y'), D^2v(y'))$ connected. This set contains the full mass of $g$.

Observe both $\partial_s v(y')$ and $X'_i$ are relatively closed subsets of $X$. Thus $\partial_s v(y')$ is also closed relative to $X'_i$. To show it is relatively open, let $x' \in \partial_s v(y')$. 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 [20], 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,

$$[\{y'\} \times \mathbb{R}^{m-n}] \cap U = F^{-1}(y') \cap U \subset \partial_s v(y') \cap U \subset X'_i \cap U \subset X'_i \cap U$$

follows from (30). But Proposition 2 of [8] shows $X'_i$ 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 $\partial_s v(y')$ relative to $X'_i$, concluding the proof that $\partial_s v(y')$ is relatively open. Thus $\partial_s v(y') = X'_i$ since the former is open, closed and non-empty and the latter is connected. Equality in (30) has been established for $g$-a.e. $y$, concluding the proof. $\blacksquare$

The following example shows that the level set connectivity assumption is required to deduce nestedness; it also illustrates why it may be necessary to consider the $i = 2$ case of the local equation.

Example 5 (Annulus to circle) Consider transporting uniform mass on the annulus, $X = \{x \in \mathbb{R}^2 : 1/2 \leq |x| \leq 1\}$ to uniform measure on the circle, $Y = \{y \in \mathbb{R}^2 : |y| = 1\}$ with the bilinear surplus, $s(x, y) = x \cdot y$. It is easy to see that $x \cdot y \leq |x|$, with equality only when $y = \frac{x}{|x|}$, implying that the optimal map takes the form $F(x) = \frac{x}{|x|}$ and the potentials $u(x) = |x|, v(y) = 0$. These are smooth on the regions $X$ and $Y$. Note that $X_2(y, Dv(y), D^2v(y)) = \{x \in X : \frac{x}{|x|} = y\}$ is connected and coincides with $\partial^s v(y)$ (as is guaranteed by the preceding theorem). On the other hand, $X_1(y, Dv(y), D^2v(y)) = \{x \in X : \frac{x}{|x|} = y\} \cup \{x \in X : \frac{x}{|x|} = -y\}$ is disconnected and the inclusion $\partial^s v(y) \subset X_1(y, Dv(y), D^2v(y))$ is strict.
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 [26]. This section is purely s-convex analytic; no measures are mentioned.

Lemma 6 (Continuity of maps (local)) Fix \( m \geq n \), open sets \( X \subset \mathbb{R}^m \) and \( Y \subset \mathbb{R}^n \) with \( Y \) bounded, and \( s \in C^2(X \times Y) \) twisted and non-degenerate. Let \( (u,v) = (v^x,u^x) \) and \( D^2v(y) > D^2_{yy}s(x,y) \) for some \( (x,y) \in X \times [\partial_s v^s(x) \cap \text{dom} D^2v] \). The isolated point \( y \) forms a \( C^1 \)-path-connected component of \( \partial_s u(x) \). Thus \( x \in \text{dom } Du \) if, in addition, \( \partial_s u(x) \) is \( C^1 \)-path-connected.

Proof. Fix \( (u,v) \) and \( (x,y) \) as in the lemma. Let \( y : t \in [0,1] \mapsto y(t) \in \partial_s u(x) \) be a continuously differentiable curve 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 \( \text{[1]} \) implies \( x \in \text{dom } Du \) provided we can establish convergence of \( Du(x_k) \) to a unique limit whenever \( x_k \in \text{dom } Du \) converges to \( x \). Therefore, let \( x_k \in \text{dom } Du \) converge to \( x \), and choose \( y_k \in \partial_s v(x_k) \). Any accumulation point \( y_\infty \) of the \( y_k \) satisfies \( y_\infty \in \partial_s v(x) = \{ y \} \). 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 \( x \in \text{dom } Du \).

Corollary 7 (Continuity of maps (global)) Fix \( m \geq n \), open sets \( X \subset \mathbb{R}^m \) and \( Y \subset \mathbb{R}^n \) with \( Y \) bounded, and \( s \in C^2(X \times Y) \) twisted and non-degenerate. Let \( (u,v) = (v^x,u^x) \) with \( v \in C^2(Y) \). 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 \( y \in \partial_s u(x) \).

Proof. Lemma \( \text{[6]} \) implies \( X = \text{dom } Du \) under the hypotheses of Corollary \( \text{[4]} \). Since semiconvexity of \( u \) was shown in Theorem \( \text{[1]} \) this is sufficient to conclude \( u \in C^1(X) \).

Proposition 8 (Criteria for differentiability of maps) Fix \( m \geq n \), open sets \( X \subset \mathbb{R}^m \) and \( Y \subset \mathbb{R}^n \) with \( Y \) bounded, and \( s \in C^2(X \times Y) \) twisted and non-degenerate. Use \( (u,v) = (v^x,u^x) \) with \( u \in C^1(X) \) to define \( F : X \to \bar{Y} \) through \( \text{[18]} \). Then both \( F \) and \( D_x s_y(\cdot,F(\cdot)) \) are in
$$(BV_{loc} \cap C)(X, \mathbb{R}^n)$$. If, in addition, \(v \in C^{1,1}(Y)\) then \(F \in \text{Dom} \, D^2 v\) on a set of \(|DF|\) full measure, and as measures

$$\left(D^2 v(F(x)) - s_{yy}(x, F(x))\right) DF(x) = D_x s_y(x, F(x)). \quad (31)$$

In this case, \(F\) is Lipschitz in any open subset of \(X\) where

$$D^2 v(F(x)) - D^2_{yy} s(x, F(x)) \geq \epsilon I > 0 \quad (32)$$

is uniformly positive definite (and \(F\) inherits higher differentiability from \(v\) and \(s\) in this case).

**Proof.** Recalling

$$Du(x) = D_x s(x, F(x)), \quad (33)$$

the continuity \(F = s \circ Du\) follows from \(u \in C^1(X)\) and the twisted-ness and non-degeneracy of \(s\).

Since \(u\) from Theorem 1 is also semiconvex, its directional derivatives lie in \(BV(X)\) and its gradient in \(BV(X, \mathbb{R}^2)\). We shall use (33) to deduce \(F \in BV_{loc}(X)\), 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 (33) 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 \(F \in BV_{loc}(X, \mathbb{R}^n) \quad [1]\).

On the other hand, when \(Dv\) is assumed Lipschitz, Ambrosio and Dal Maso [1] assert \(F \in \text{Dom} \, D^2 v\) on a set of \(|DF|\) full measure, and differentiating \(Dv(F(x)) = D_y s(x, F(x))\) yields (31) 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 \(Dv\) follows immediately by rewriting (31)–(32) in the form

$$DF(x) = (D^2 v(F(x)) - s_{yy}(x, F(x)))^{-1} D_x s_y(x, F(x)) \in L^\infty(X).$$

---

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 \( C^{2,\alpha}_{\text{loc}} \), to parallel known results beginning with [6, 28] for \( n = m \); we identify conditions ensuring this when \( n = 1 \) in the last two sections (see Remark 28). Recall that a second-order differential operator \( G(y,p,Q) \) is said to be \textit{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 \textit{strict} at \( (y,p,Q) \) if there is a constant \( \lambda = \lambda(y,p,Q) > 0 \) called the \textit{ellipticity constant} such that \( Q' \geq Q \) implies

\[
G(y,p,Q') - G(y,p,Q) \geq \lambda \text{tr}[Q' - Q].
\]

(34)

**Lemma 9 (Strict ellipticity)** The operator \( G \) defined by (19) and (22) 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 \((y,p) \in Y \times \mathbb{R}^m\) 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

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

for all \( Q' \geq Q \). From here we deduce

\[
\det[I + (Q - D^2_{yy}s)^{-1}(Q' - Q)] \geq 1 + \Theta^{-1}\text{tr}[Q' - Q].
\]

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

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

as desired. 

**Theorem 10 (Bootstrapping regularity using Schauder theory)** Fix \( 0 < \alpha < 1 \) and integer \( k \geq 2 \). If \( g > \epsilon > 0 \) on some smooth domain \( Y' \) compactly contained in \( Y^0 \) 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 (21)–(22) with \( i = 2 \) implies \( v \in C^{k+1,\alpha}(Y') \).

**Proof.** Since \( v \in C^{2,\alpha}(Y') \), (21) 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

\[
a^{ij}(y)D^2_{ij}w + b^i(y)D_iw = d(y)
\]

(35)
for \( w = \partial v / \partial y^k \) whose coefficients

\[
a^{ij}(y) := \frac{\partial G}{\partial Q_{ij}} \bigg|_{(y,Dv(y),D^2v(y))} \\
\]

\[
b^{i}(y) := \frac{\partial G}{\partial q_i} \bigg|_{(y,Dv(y),D^2v(y))} \\
\]

and inhomogeneity

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

have (i) \( C^{k-2,\alpha^2} \) norm controlled by \( \|G - g\|_{C^{k-1,\alpha}} \|v\|_{C^{k,\alpha}} \) 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 (35) in the viscosity sense described e.g. in [9]. From Lemma 4 we see the matrix \( (a^{ij}) \) is bounded below by \( \epsilon I / \|v\|_{C^2(Y')} \); it is bounded above by \( \|G\|_{C^1(Y)} \). Thus the equation satisfied by \( w \) on \( Y' \) is uniformly elliptic. Since the coefficient of \( w \) vanishes in (35), the Dirichlet problem with continuous boundary data on any ball in \( Y' \) is known to admit a unique (viscosity) solution [9]; moreover, this solution (i) \( C^{k,\alpha^2} \) (by e.g. Gilbarg & Trudinger Theorems 6.13 \((k = 2)\) or 6.17 \((k > 2)\)). Thus we infer \( v \in C^{k+1,\alpha^2}(Y') \). Applying the same argument again starting from the improved estimates (ii) now established yields \( v \in C^{k+1,\alpha}(Y') \). 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 (21), the mean value theorem yields \( h^*(y) \in [0,h] \) lower semicontinuous such that

\[
0 = \Delta_h^k [G(y, Dv(y), D^2v(y)) - g(y)] \\
= a^{ij}_h(y) D^2_{ij} w_h + b^i_h(y) D_i w_h - d_h(y).
\]

Here \( w_h = \Delta_h^k v \) and the coefficients

\[
a^{ij}_h(y) := \frac{\partial G}{\partial Q_{ij}} \bigg|_{(I+h^*(y)\Delta_h^k y, Dv(y), D^2v(y))} \\
\]

\[
b^i_h(y) := \frac{\partial G}{\partial q_i} \bigg|_{(I+h^*(y)\Delta_h^k y, Dv(y), D^2v(y))} \\
\]

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

15
are measurable and converge uniformly to \((a^i, b^i, d)\) as \(h \to 0\). The solutions \(w_h = \Delta_h v \in C^{2, \alpha}\), being finite differences, converge to \(\partial v/\partial y_k\) in \(C^{1, \alpha}(Y')\). Lemma 6.1 and Remark 6.3 of [9] show this partial derivative \(w = \partial v/\partial y_k\) must then be the required viscosity solution of the limiting equation \(\partial_t v = \Delta v\). □

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 \(G_2\) under the simplifying hypothesis that \(G_2 = G_1\). Our main result is Theorem 11. Though the assumed smoothness of \(v\) is addressed for \((n, i) = (1, 2)\) in Section 5, we consider neither the smoothness nor the uniform convexity of \(v\) for higher dimensional targets, which as we have noted, remain interesting open questions.

Our joint work with Chiappori [8] establishes regularity of \(G_1\) when \(n = 1\); in this section, we focus on this smoothness for higher dimensional targets. 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 10 of the last section, and \(C^2\)-smoothness of \(u = v^s\), implies that \(G_1 = G_2\) by Theorem [10] so in many cases of interest it is enough to address smoothness of \(G_1\). 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 \(G_2\).

We let \(Y', P' \subseteq \mathbb{R}^n\) be bounded open sets and set \(X' = \bigcup_{(y,p) \in Y' \times P'} X_1(y, p)\). For technical reasons it is convenient to assume that \(y \mapsto s(x, y)\) is uniformly convex throughout this section; that is, \(D_{yy}s(x, y) \geq CI > 0\) throughout \(X' \times Y'\). Note that this assumption can always be achieved by adding a sufficiently convex function of \(y\) to \(s\).

**Theorem 11 (Smoothness of \(G_1\))** Let \(r \geq 1\). Then \(||G_1||_{C^{r,1}(Y' \times P')}\) is controlled by \(||f||_{C^{r,1}(X')}, ||D_y s||_{C^{r+1,1}(Y' \times X')}, ||\hat{n}_X||_{C^{r-1,1}}\) and

\[
\inf_{(x,y) \in X' \times Y'} \min_{v \in \mathbb{R}^n, |v| = 1} |D_{xy}s(x, y) \cdot v| \quad \text{(non-degeneracy),} \quad (36)
\]

\[
\inf_{(x,y,p) \in (\partial X \cap X') \times Y' \times P'} |(\hat{n}_X)_{T_x X_1(y, p)}| \quad \text{(transversality),} \quad (37)
\]

\[
\sup_{(y,p) \in Y' \times P'} \mathcal{H}^{m-n}(X_1(y, p)) \quad \text{(size of level sets), and of} \quad (38)
\]

\[
\sup_{(y,p) \in Y' \times P'} \mathcal{H}^{m-n-1}(X_1(y, p) \cap \partial X) \quad \text{(boundary intersections),} \quad (39)
\]
assuming finiteness and positivity of each quantity above. Here \((\hat{n}_X)_T X_1(y,p)\) denotes the projection of the outward unit normal \(\hat{n}_X\) to \(X\) onto the tangent space \(T_x X_1(y,p)\).

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

For \(i \in \{1, \ldots, n\}\), the set \(X^i(y,p) := \{x \mid s_{y_i} \leq p_i, s_{y_j} = p_j \forall j \neq i\}\) is a submanifold of \(X\) whose relative boundary is given by \(X_1(y,p)\). Then \(X^i(y,p) \subseteq X^i(y,p) := \{x \mid s_{y_j} = p_j \forall j \neq i\}\), while with an analogous definition \(X^i_1(y,p)\) 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 12 (Submanifold transversality)** The submanifolds \(\partial X^i = \bar{X}^i \cap \partial X\) and \(X_1\) intersect transversally in \(X^i\).

**Proof.** The proof is straightforward linear algebra. Transversal intersection of \(X_1\) and \(\partial X\) in \(\mathbb{R}^m\) easily implies transversal intersection of \(X^i\) and \(\partial X\), and so \(T_x(\partial X) = T_x(\partial X) \cap T_x(X^i)\) at each point of intersection \(x \in \bar{X}^i \cap \partial X\). 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 \(p \in T_x X^i \subset \mathbb{R}^m = T_x(\partial X) + T_x X_1\) (by transversality), we write \(p = p_1 + p_0\), with \(p_1 \in T_x X^i \subseteq T_x X^i\) and \(p_0 \in T_x(\partial X)\). But then \(p_0 = p - p_1 \in T_x X_1\), and so \(p_0 \in [T_x(\partial X) \cap T_x(X^i)]\), implying the containment \(T_x X^i \subseteq [T_x(\partial X) \cap T_x(X^i)] + T_x X_1\). \(\blacksquare\)

Given \(f \in L^\infty\), Lemma 5.1 of [3] implies that

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

has a Lipschitz dependence on \(p\), with

\[ \frac{\partial \Phi^i}{\partial p_i}(y,p) = \int_{X^i(y,p)} \frac{f(x,y)}{|D_x X_1 y_i|} d\mathcal{H}^{m-n}(x) \quad [\text{a.e.}], \quad (40) \]

where \(D_x X_1 y_i\) is the differential of \(s_{y_i}\) along the submanifold \(X^i\), nonzero by the nondegeneracy assumption:
Lemma 13 (Restriction non-degeneracy) The differential $D_{X^i} s_{y_i}$ of $s_{y_i}$ along the manifold $X^i$ satisfies

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

Proof. Note that $D_{X^i} s_{y_i}$ is $D_x s_{y_i}$, minus its projection onto the span of the other $D_x s_{y_j}$, and so

$$|D_{X^i} s_{y_i}| = \min_{v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n} |D_x s_{y_i} - \sum_{j \neq i} v_j D_x s_{y_j}|$$

$$= \min_{v = (v_1, \ldots, v_n) \in \mathbb{R}^n, v_i = 1} |D_{xy}^2 s \cdot v|$$

$$\geq \min_{v \in \mathbb{R}^n, |v| = 1} |D_{xy}^2 s \cdot v|.$$

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

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

and the normal velocity of $X^i_1(y, p)$ 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

$$D_{X^i} s_{y_i}(x, y) := D_{X^i(y,p)} s_{y_i}(x, y) \bigg|_{p = p_0(x, y)}$$

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 $(X^i_{\leq}(y, p)) \cap \partial X$ in $(X^i(y, p)) \cap \partial X$ 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 transversality (note $|(\hat{n}_X)_{T_x X^i}| \geq |(\hat{n}_X)_{T_x X^i}|$). This is the outward unit normal to $X^i_{\leq}(y, p) \cap X$ in $\overline{X^i}(y, p)$.

We have that

$$\hat{n}^i_\partial = \frac{\hat{n}^i - (\hat{n}^i_X \cdot \hat{n}^i) \hat{n}^i_X}{\sqrt{1 - (\hat{n}^i_X \cdot \hat{n}^i)^2}}.$$
Lemma 14 (Derivative bounds along submanifolds) Given functions $a : X' \times Y' \to \mathbb{R}$, $b : \partial X \times Y \to \mathbb{R}$ and vector fields $v : X' \times Y' \to TX$ and $w : (X' \cap \partial X) \times Y \to T\partial X$ such that $v(x, y) \in T_xX^i(x, D_y\hat{s}(x, y))$ and $w(x, y) \in T_x(X^i(x, D_y\hat{s}(x, y)) \cap \partial X)$ everywhere, we have:

1. $\|D_{X^i(y, D_y\hat{s}(x, y))}a(x, y)\|_{C^{k,1}(X' \times Y')} = \|a\|_{C^{k,1,1}(X' \times Y')}$, and nondegeneracy.
2. $\|\nabla_{X^i(x, D_y\hat{s}(x, y))} \cdot v\|_{C^{k,1}(X' \times Y')}$ is controlled by $\|v\|_{C^{k,1,1}(X' \times Y')}$. $D_{y\hat{s}}$ is controlled by $\|y\|_{C^{k,1,1}(X' \times Y')}$. nondegeneracy, transversality and $\|\hat{n}_X\|_{C^{k,1}(X' \cap \partial X)}$.
3. $\|D_{X^i(y, D_y\hat{s}(x, y))} \cap \partial X b(x, y)\|_{C^{k,1}((X' \cap \partial X) \times Y')} = \|b\|_{C^{k,1,1}((X' \cap \partial X) \times Y')}$, nondegeneracy, transversality and $\|\hat{n}_X\|_{C^{k,1}(X' \cap \partial X)}$.

4. $\|D_{X^i(x, D_y\hat{s}(x, y))} \cap \partial X w\|_{C^{k,1}((X' \cap \partial X) \times Y')}$ is controlled by $\|w\|_{C^{k,1,1}((X' \cap \partial X) \times Y')}$.

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

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

where the $e_j(x, y)$ are an orthonormal basis for the span of $\{D_x\hat{s}_{y_j}(x, y)\}_{j \neq i}$. The $e_j$ can then be written explicitly as functions of the $D_x\hat{s}_{y_j}(x, y)$, using for instance the Gram-Schmidt procedure; the definition of $e_j$ involves projections onto the $e_j$ for $j < j$, which are controlled by nondegeneracy.
The second implication follows by noting that the divergence \( \nabla_{X^i} (x, D_y s(x, y)) \cdot v(x, y) \) 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 \( \{ D_y s_y(x, y) \}_{j \neq i} \cup \{ \hat{n}_X \} \); 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 [8].

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 15 (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 \( s \):

\[
\inf_{|u|=1} |D_{xp} s^* (x, p) \cdot u| \geq \frac{\inf_{|v|=1} |D_{xy} s(x, y) \cdot v|}{||D_{yy} s(x, y)||}
\]

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

\[
D_{xp}^2 s^* (x, D_y s(x, y)) = -D_{xy}^2 s(x, y) D_{pp}^2 s^* (x, D_y s(x, y))
\]

and so invertibility of \( D_{pp}^2 s^* \) and nondegeneracy of \( s \) imply nondegeneracy of \( s^* \), and we have, for \(|u| = 1\),

\[
D_{xp}^2 s^* (x, D_y s(x, y)) \cdot u = -D_{xy}^2 s(x, y) D_{pp}^2 s^* (x, D_y s(x, y)) \cdot u
\]

\[
= -D_{xy}^2 s(x, y) \frac{D_{pp}^2 s^* (x, D_y s(x, y)) \cdot u}{D_{pp}^2 s^* (x, D_y s(x, y)) \cdot u} |D_{pp}^2 s^* (x, D_y s(x, y)) \cdot u|.
\]

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

\[
|v| \geq \frac{1}{||D_{yy}^2 s(x, y)||}.
\]
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 \( X^{*i}_i(y, p) \), \( X^{**}_i(y, p) \) and \( \Phi^{**}_i \) analogously to above, and compute

\[
\frac{\partial \Phi^{**}_i}{\partial y_i} = \int_{X^{**}_i(y, p)} \frac{f(x, y)}{|D_{X^{**}_i y_i}|} d\mathcal{H}^{m-n-1}(x) + \int_{X^{**}_i(y, p)} \frac{\partial f}{\partial y_i}(x, y) d\mathcal{H}^{m-n}(x)
\]

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

Analogs of Lemmas 12, 13 and 14 when \( s(x, y) \) is replaced by \( s^*(x, p) \) then follow immediately. We note that

\[
D_{X^{**}_i} s^*_i(x, y) := D_{X^{**}_i(y, p)} s^*_i(x, p)
\]

is defined throughout \( X' \times Y' \). We define \( \hat{n}^{* i}, V^{* i}, \hat{n}_\partial^{* i}, \hat{n}^{** i}, V^{** i} \) 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 16 (Flux derivatives through moving surfaces)** Use \( a : X' \times Y' \times \mathbb{R}^2 \rightarrow \mathbb{R} \) Lipschitz to define \( \Phi(y, p) := \int_{X_1(y, p)} a(x, y, p) d\mathcal{H}^{m-n}(x) \) and \( \Psi(y, p) := \int_{X_1(y, p) \cap \partial X} a(x, y, p) d\mathcal{H}^{m-n-1}(x) \). Then \( \Phi \) and \( \Psi \) are Lipschitz with partial derivatives given almost everywhere by:

\[
\frac{\partial \Phi(y, p)}{\partial p_i} = \int_{X_1(y, p)} \left[ \nabla X^{**}_i(y, p) \cdot \left( a(x, y, p) \frac{D_{X^{**}_i} y_i}{|D_{X^{**}_i} y_i|} \right) V^i \cdot \hat{n}_{\partial}^i \right]_{p = D_y s(x, y)} d\mathcal{H}^{m-n-1}(x)
\]

\[
\frac{\partial \Psi(y, p)}{\partial p_i} = \int_{X_1(y, p) \cap \partial X} \left[ \nabla X^{**}_i(y, p) \cdot \left( a(x, y, p) \frac{D_{X^{**}_i} y_i}{|D_{X^{**}_i} y_i|} \right) V^i \cdot \hat{n}_{\partial}^i \right]_{p = D_y s(x, y)} d\mathcal{H}^{m-n}(x)
\]

(42)
\[ \frac{\partial \Phi(y, p)}{\partial y_i} = \int_{X_1(y,p)} \left[ \nabla_{X^* (y,p)} \cdot \left( a(x, y, p) \frac{D_{X^* s_{y_i}}}{D_{X^* s_{y_i}}} \right) V^* \cdot \hat{n}^* \right]_{p=D_y s(x,y)} \cdot d\mathcal{H}^{m-n}(x) \]
\[ - \int_{X_1(y,p)} \left[ \left( a(x, y, p) \frac{D_{X^* s_{y_i}}}{D_{X^* s_{y_i}}} \right) \cdot \hat{n}_X V_\partial \cdot \hat{n}_\partial \right]_{p=D_y s(x,y)} \cdot d\mathcal{H}^{m-n-1}(x) \]
\[ + \int_{X_1(y,p)} \frac{\partial a(x, y, p)}{\partial y_i} \cdot d\mathcal{H}^{m-n}(x), \] (45)

and
\[ \frac{\partial \Psi(y, p)}{\partial y_i} = \]
\[ \int_{X_1(y,p)} \left[ \nabla_{X^* (y,p)} \cdot \left( a(x, y, p) \frac{D_{X^* s_{y_i}}}{D_{X^* s_{y_i}}} \right) V^* \cdot \hat{n}^* \right]_{p=D_y s(x,y)} \cdot d\mathcal{H}^{m-n}(x) \]
\[ + \int_{X_1(y,p)} \frac{\partial a(x, y, p)}{\partial y_i} \cdot d\mathcal{H}^{m-n-1}(x). \] (46)

**Proof.** We begin by establishing the formulas assuming \( a \in C^{1,1}(X^\prime \times Y^\prime \times P^\prime) \).
Using the generalized divergence theorem [21 Proposition 27]

\[ \Phi(y, p) = \int_{X_1(y,p)} \left( a(x, y, p) \frac{D_{X^* s_{y_i}}}{D_{X^* s_{y_i}}} \right) \cdot \hat{n}^i \cdot d\mathcal{H}^{m-n}(x) \]
\[ = \int_{X_1(y,p) \setminus X_1(y,p)} \nabla_{X^* (y,p)} \cdot \left( a(x, y, p) \frac{D_{X^* s_{y_i}}}{D_{X^* s_{y_i}}} \right) \cdot d\mathcal{H}^{m-n+1}(x) \]
\[ + \int_{X_1(y,p) \setminus X_1(y,p)} \left( a(x, y, p) \frac{D_{X^* s_{y_i}}}{D_{X^* s_{y_i}}} \right) \cdot \hat{n}^i \cdot d\mathcal{H}^{m-n}(x) \]
\[ - \int_{X_1(y,p) \setminus X_1(y,p)} \nabla_{X^* (y,p)} \cdot \left( a(x, y, p) \frac{D_{X^* s_{y_i}}}{D_{X^* s_{y_i}}} \right) \cdot \hat{n}^i \cdot d\mathcal{H}^{m-n}(x). \]

Noting that the integrands in the first and third terms above are bounded, one can then combine the chain rule with (10) and (11) to differentiate with respect to \( p_i \), getting

\[ \frac{\partial \Phi(y, p)}{\partial p_i} = \int_{X_1(y,p) \setminus X_1(y,p)} \nabla_{X^* (y,p)} \cdot \left( a(x, y, p) \frac{D_{X^* s_{y_i}}}{D_{X^* s_{y_i}}} \right) \cdot V^i \cdot \hat{n}^* \cdot d\mathcal{H}^{m-n}(x) \]
\[ - \int_{X_1(y,p) \setminus X_1(y,p)} \nabla_{X^* (y,p)} \cdot \left( a(x, y, p) \frac{D_{X^* s_{y_i}}}{D_{X^* s_{y_i}}} \right) \cdot \hat{n}_X V_\partial \cdot \hat{n}_\partial \cdot d\mathcal{H}^{m-n-1}(x) \]
\[ + \int_{X_1(y,p)} \frac{\partial a(x, y, p)}{\partial p_i} \cdot d\mathcal{H}^{m-n}(x). \]
Finally, notice that one may substitute \( p = D_ys(x, y) \) in each integrand, as each region of integration is contained in \( X_1(y, p) \), to establish (13) for \( a \in C^{1,1} \).

Now, note that the formula (13) 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 \( ||\Phi||_{C^{0,1}([X^1 \times X_0])} \) 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 (14) for \( \frac{\partial \Psi(y, p)}{\partial p_i} \), while essentially identical arguments apply to the \( y \) derivatives, yielding (15) and (16).

**Corollary 17 (Iterated derivative bounds)** The operators

\[
A_{p_i} : (a, b) \mapsto (a^i_p, b^i_p) \quad \text{and} \quad A_{y_i} : (a, b) \mapsto (a^i_y, b^i_y),
\]

given by

\[
a^i_p := \left[ \nabla X^1(y, p) \cdot \left( a(x, y) \frac{D_X^s y_i}{D_X^s y_i} \right) V^i \cdot \hat{n}_\partial \right]_{p = D_ys(x, y)},
\]
\[
b^i_p := \left[ \left( a(x, y) \frac{D_X^s y_i}{D_X^s y_i} \right) \cdot \hat{n}_X V^i \cdot \hat{n}_\partial \right.
\]
\[+ \left. \nabla X^1(y, p) \cdot \left( b(x, y) \frac{D_X^s y_i}{D_X^s y_i} \right) V^i \cdot \hat{n}_\partial \right]_{p = D_ys(x, y)},
\]
\[
a^i_y := \left[ \nabla X^1(y, p) \cdot \left( a(x, y) \frac{D_X^s s^p_i}{D_X^s s^p_i} \right) V^i \cdot \hat{n}_\partial + \frac{\partial a(x, y)}{\partial y^i} \right]_{p = D_ys(x, y)}, \quad \text{and}
\]
\[
b^i_y := \left[ \left( a(x, y) \frac{D_X^s s^p_i}{D_X^s s^p_i} \right) \cdot \hat{n}_X V^i \cdot \hat{n}_\partial \right.
\]
\[+ \left. \nabla X^1(y, p) \cdot \left( b(x, y) \frac{D_X^s s^p_i}{D_X^s s^p_i} \right) V^i \cdot \hat{n}_\partial \right]_{p = D_ys(x, y)},
\]

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

\[
||A_{p_i}|| \leq \frac{1}{||D_X^s y_i||} ||C^{k-1,1}(X' \times Y')|| \hat{n}_\partial ||C^{k-1,1}(X' \times Y')
\]
\[+ ||\hat{n}_\partial ||_{C^{k-1,1}(X' \times Y')} ||\hat{n}_X||_{C^{k-1,1}((X' \cap \partial X) \times Y')} ||V^i||_{C^{k-1,1}((X' \cap \partial X) \times Y')} ||\hat{n}_\partial||_{C^{k-1,1}((X' \cap \partial X) \times Y')}
\]
\[+ \frac{D_X^s y_i}{D_X^s y_i} ||C^{k-1,1}((X' \cap \partial X) \times Y')|| V^i ||C^{k-1,1}((X' \cap \partial X) \times Y')|| \hat{n}_\partial ||C^{k-1,1}((X' \cap \partial X) \times Y')
\]

23
and

\[ ||A_y|| \leq \left|\frac{1}{|D_{X^i}S_{y_i}|}||C^{k-1,1}(X'\times Y')||\hat{n}^{y_i}||C^{k-1,1}(X'\times Y') + 1 \right| \]

\[ + ||\hat{n}^{y_i}||C^{k-1,1}(X'\times Y')||\hat{n}^{y_i}||C^{k-1,1}(\overline{\nabla'\partial X')\times Y'})||V^{y_i}_\theta \cdot \hat{n}^{y_i}||C^{k-1,1}(\overline{\nabla'\partial X')\times Y'}) \]

\[ + \left|\frac{1}{|D_{X^i}S_{y_i}|}||C^{k-1,1}(X'\times Y')||\hat{n}^{y_i}||C^{k-1,1}(X'\times Y') + 1 \right| \]

controlled by \( ||D_y s||_{C^{k,1}} ||\hat{x}||_{C^{k,1}} \), non-degeneracy and transversality.

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

\[ ||A_y||_{C^{k,1}(X'\times Y')\oplus\{0\} \to B_{k-1}} \leq \left|\frac{1}{|D_{X^i}S_{y_i}|}||C^{k-1,1}(X'\times Y')||\hat{n}^{y_i}||C^{k-1,1}(X'\times Y') \right| \]

and

\[ ||A_y||_{C^{k,1}(X'\times Y')\oplus\{0\} \to B_{k-1}} \leq \left|\frac{1}{|D_{X^i}S_{y_i}|}||C^{k-1,1}(X'\times Y')||\hat{n}^{y_i}||C^{k-1,1}(X'\times Y') + 1 \right| \]

\[ + ||\hat{n}^{y_i}||C^{k-1,1}(X'\times Y')||\hat{n}^{y_i}||C^{k-1,1}(\overline{\nabla'\partial X')\times Y'})||V^{y_i}_\theta \cdot \hat{n}^{y_i}||C^{k-1,1}(\overline{\nabla'\partial X')\times Y'}) \]

are controlled by \( ||D_y s||_{C^{k,1}} ||\hat{x}||_{C^{k,1,1}}, \) 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 [13, 14, 15] and closure of the Hölder spaces \( C^{k,1} \) under composition.

We now prove the result announced at the beginning of this section: **Proof Theorem [11]**. First note that as \( Q \) enters the definition of \( G_1 \) only through the integrand, whose dependence on \( Q \) is smooth, computing derivatives with respect to \( Q \) is straightforward.

Corollary [17] 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 \( \gamma = (\gamma_1, \gamma_2, ... , \gamma_m) \) with \( |\alpha| + |\beta| + |\gamma| = \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \beta_i + \sum_{i=1}^{n} \gamma_i = k \leq r \), then Lemma [16] and Corollary [17] allow us to compute

\[ \frac{\partial^k G_1}{\partial p^\alpha \partial y^\beta \partial Q^\gamma} = \int_{X_1(y,p)} a^{\alpha,\beta} d\mathcal{H}^{m-n} + \sum_{\partial X_1(y,p) \cap \partial X_1} b^{\alpha,\beta} d\mathcal{H}^{m-n-1} \]

(47)

where \( (a^{\alpha,\beta}, b^{\alpha,\beta}) = A^\alpha A^\beta (\frac{\partial^{1\gamma} h}{\partial \gamma^\gamma}, 0) \in B_{r-k} \), with \( h(x, y, p) = \frac{\text{det}[Q - D^2_{xy} s(x, y)]}{\text{det}D^2_{xy} s(x, y)} f(x) \)

being the original integrand in the definition of \( G_1(y, p, Q) \), and \( A^\alpha = A_{p_1}^{\alpha_1} ... A_{p_n}^{\alpha_n}, A^\beta = A_{y_1}^{\beta_1} ... A_{y_m}^{\beta_m} \). Now, Corollary [17] implies that

\[ ||(a^{\alpha,\beta}, b^{\alpha,\beta})||_{C^{r-k,1}} \]

24
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 (47) that $\frac{\delta^k G_1}{\partial y^\alpha \partial y^\beta \partial Q^\gamma}$ 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 10?

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 adopt the approach in [5], 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 $C^{1,1}_{\text{loc}}$ solutions to the $i = 2$ equation are in fact $C^{2,1}_{\text{loc}}$; 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 regions $Y', P'$ and $Q'$ in $\mathbb{R}$, we set $X' = \bigcup_{(y,p,q) \in Y' \times P' \times Q'} X_2(y,p,q)$ Assume $D_x s_y$ and $D_x s_{yy}$ are linearly independent throughout $X' \times Y' \subseteq X \times Y$.

As $p$ is increased, the domain $W_{\leq}(y,p) := \{x \in X \mid s_y \leq p\}$ expands monotonically outward with normal velocity $w(x,y) := |D_x s_y|^{-1}$ 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 $z(x,y) := |D_x s_{yy}|^{-1}$ along its interface $Z = \{x \in X \mid s_{yy} = q\}$; its normal velocity with respect to changes in $y$ is $-s_{yyy} z$. Our linear independence
assumption guarantees these velocities are finite and \( W_\leq \) intersects \( W_\geq \) transversally. Notice \( X_2(y,p,q) = W_\geq (y,p) \cap Z_\leq (y,q) \). Assume also, in the same region of interest, that both \( W_\leq \cap Z_\leq \) and \( W_\geq \cap Z_\geq \) intersect \( \partial X \) transversally. We denote by \( \hat{n}_W = wD_x s_y \) and \( \hat{n}_Z = zD_x s_y \), 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 \( Z_\geq \), when \( \hat{n}_Z \cdot \hat{n}_W = \cos \theta \).

Our main theorem of this section is the following.

**Theorem 18 (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 \( \| f \|_{C^{r+1}(X')} \), \( \| s_y \|_{C^{r+2,1}(Y' \times X')} \), \( \| \hat{n}_X \|_{C^{r-1,1}} \) and

\[
\inf_{(x,y) \in X' \times Y'} \min \{ |D_x s_y(x,y)|, |D_x s_y(y,x)| \} \quad \text{(non-degeneracy)},
\]

\[
\inf_{(x,y) \in X' \times Y'} 1 - (\hat{n}_W \cdot \hat{n}_Z)^2 \quad \text{(transversality)},
\]

\[
\inf_{(x,y) \in (\partial X \cap X') \times Y'} |\lambda_1 \hat{n}_W + \lambda_2 \hat{n}_Z + \lambda_3 \hat{n}_X| \quad \text{(linear independence)},
\]

\[
\sup_{(y,p,q) \in Y' \times P' \times Q'} H^{m-1}(W_\geq (y,p) \cap Z_\leq (y,q)) \quad \text{(1st level set size)},
\]

\[
\sup_{(y,p,q) \in Y' \times P' \times Q'} H^{m-1}(W_\leq (y,p) \cap Z_\leq (y,q)) \quad \text{(2nd level set size)},
\]

\[
\sup_{(y,p,q) \in Y' \times P' \times Q'} H^{m-2}((W_\leq (y,p) \cap Z_\leq (y,q)) \cap \partial X) \quad \text{(1st boundary level size)},
\]

\[
\sup_{(y,p,q) \in Y' \times P' \times Q'} H^{m-2}((W_\geq (y,p) \cap Z_\geq (y,q)) \cap \partial X) \quad \text{(2nd boundary level size)},
\]

\[
\sup_{(y,p,q) \in Y' \times P' \times Q'} H^{m-1}((W_\geq (y,p) \cap Z_\leq (y,q)) \cap \partial X) \quad \text{(boundary sublevels size)},
\]

assuming finiteness and positivity of each quantity above.

**Remark 19** When \( m = 2 \), the linear independence assumption \([50]\) cannot hold. It can be replaced by the assumption that \( \{ x \in \bar{X} : (s_y(x,y), s_{yy}(x,y)) \in P' \times Q' \} \) does not intersect \( \partial X \) for any \( y \in Y' \); in this case, quantity \([50]\) should be replaced by the pairwise linear independence quantities

\[
\inf_{(x,y) \in (X' \cap \partial X) \times Y'} 1 - (\hat{n}_W \cdot \hat{n}_X)^2,
\]

and

\[
\inf_{(x,y) \in (X' \cap \partial X) \times Y'} 1 - (\hat{n}_X \cdot \hat{n}_Z)^2.
\]
Lemma 20 (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

\[
A(y, p, q) := \int_{X_2(y, p, q)} a(x, y, p, q) d\mathcal{H}^{m-1}(x),
\]
\[
B(y, p, q) := \int_{W_\leq(y, p) \cap Z_\leq(y, q)} b(x, y, p, q) d\mathcal{H}^m(x),
\]
\[
C(y, p, q) := \int_{W_\leq(y, p) \cap Z_{\leq}(y, q)} c(x, y, p, q) d\mathcal{H}^{m-1}(x),
\]
\[
A^\partial(y, p, q) := \int_{(W_\leq(y, p) \cap Z_\leq(y, q)) \cap \partial X} a^\partial(x, y, p, q) d\mathcal{H}^{m-2}(x),
\]
\[
B^\partial(y, p, q) := \int_{(W_\leq(y, p) \cap Z_\leq(y, q)) \cap \partial X} b^\partial(x, y, p, q) d\mathcal{H}^{m-1}(x)
\]
and
\[
C^\partial(y, p, q) := \int_{(W_\leq(y, p) \cap Z_{\leq}(y, q)) \cap \partial X} c^\partial(x, y, p, q) d\mathcal{H}^{m-2}(x)
\]

are all Lipschitz, with derivatives given almost everywhere by the formulæ in Appendix A. Here \( w := |D_x s_y|^{-1} \) and \( z := |D_x s_{zy}|^{-1} \) as above.

Proof. The proof is similar to the proofs of Lemma 7.4 in [8] and Lemma 16 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 [8]:

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

The result for Lipschitz functions can be obtained as in Lemma 7.4 in [8] and Lemma 16 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,

$$A^\partial(y, p, q) = \int_{(W \cap Z \cap \partial X)} a^\partial(x, y, p, q) d\mathcal{H}^{m-2}(x)$$

$$= \int_{(W \cap Z \cap \partial X)} \nabla_{\partial X} \cdot (a^\partial \hat{n}_{\partial, W}) d\mathcal{H}^{m-1}(x) - \int_{(W \cap Z \cap \partial X)} a^\partial \hat{n}_{\partial, W} \cdot \hat{n}_{\partial, Z} d\mathcal{H}^{m-2}(x),$$

where $\hat{n}_{\partial, W}$ and $\hat{n}_{\partial, Z}$ are defined in Appendix A, we can again use Lemma 5.1 in [8] to differentiate with respect to $p$. Similar arguments apply to all derivatives of $A^\partial, B^\partial$ and $C^\partial$. ■

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

$$\int_{W \cap Z} a(x, y, p, q) d\mathcal{H}^{m-2}(x) = \int_{W \cap Z} \nabla_{W} \cdot (a \hat{n}_{W, Z}) d\mathcal{H}^{m-1}(x)$$

$$- \int_{(W \cap Z \cap \partial X)} a\hat{n}_{W, Z} \cdot \hat{n}_{W, X} d\mathcal{H}^{m-2}(x),$$

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 $Z \subseteq$ and $X$, respectively.

Similarly,

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

where $\hat{n}_{\hat{W} \cap \partial X, Z}$ is the outward unit normal to $\hat{Z} \cap (\hat{W} \cap \partial X)$ in the codimension 2 submanifold $(\hat{W} \cap \partial X)$; 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 20 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 20.

**Lemma 21 (More iterated derivative bounds)** Set

$$B_r := C^{r,1}(\vec{X}' \times \vec{Y}' \times \vec{P}' \times \vec{Q}') \times C^{r,1}((\vec{X}' \cap \partial X) \times \vec{Y}' \times \vec{P}' \times \vec{Q}')$$
and consider the operators $M^p, M^q, M^y : (B_r)^3 \to (B_{r-1})^3$ defined by

\[
M^p : (a, b, c, a^\partial, b^\partial, c^\partial) \mapsto
\left(\begin{aligned}
&\left(a_p + \nabla \cdot (a\hat{n}_W)w + bw - \nabla_w \cdot \left(\frac{aw\hat{n}_W \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}}\right)\hat{n}_{W,z} + \nabla_w \cdot (cw\hat{n}_{W,x})\;\right), \\
&\left(b_p, c_p, a^\partial_w \cdot \nabla \cdot (a^\partial \hat{n}_W)w + w \nabla \cdot \left(\frac{aw^\partial \hat{n}_W \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}}\right)\hat{n}_{W,z} + \nabla_w \cdot (c^\partial \hat{n}_W, \hat{W}) + \frac{b^\partial w}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}}\right).
\end{aligned}\right)
\]

\[
M^q : (a, b, c, a^\partial, b^\partial, c^\partial) \mapsto
\left(\begin{aligned}
&\left(a_q + \nabla_w \cdot \left(\frac{az}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_W)^2}}\right)\hat{n}_{W,z} - \nabla_w \cdot \left(\frac{cz\hat{n}_Z \cdot \hat{n}_W}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_Z)^2}}\right)\hat{n}_{W,z}, \\
&\left(b_q, b_z + c_q + \nabla \cdot (c\hat{n}_W)z, \\
&a^\partial_q - \frac{az}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_W)^2}}\hat{n}_{W,x} + \frac{cz\hat{n}_Z \cdot \hat{n}_W}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_Z)^2}}\hat{n}_{W,x} + \frac{b^\partial z}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_W)^2}}\hat{n}_{W,x} + \frac{cz\hat{n}_Z \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_X)^2}}\nabla_w \cdot (c^\partial \hat{n}_\partial, x), \\
&\nabla_w \cdot (a^\partial \hat{n}_W \cdot \hat{n}_\partial, x) + \frac{cz\hat{n}_Z \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_X)^2}}\nabla_w \cdot (c^\partial \hat{n}_\partial, x)\right).
\end{aligned}\right)
\]

and
\(M^y : (a, b, c, a^\partial, b^\partial, c^\partial) \mapsto\)
\[
\begin{align*}
\left(a_y - \nabla \cdot (a\hat{n}_W) w s_{yy} - b w s_{yy} - c \frac{\partial \hat{n}_Z}{\partial y} \cdot \hat{n}_W, \\
\nabla \cdot \left(\frac{\partial \hat{n}_W}{\partial y}\right) + b_y + \nabla \cdot (c \frac{\partial \hat{n}_Z}{\partial y}), \\
-a \frac{\partial \hat{n}_W}{\partial y} \cdot \hat{n}_Z - b z s_{yy} + c_y - \nabla \cdot (c \hat{n}_Z) z s_{yy}. \\
\right)
\end{align*}
\]
\[
\begin{align*}
\frac{aws_{yy} \hat{n}_W \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}} + a^\partial + \frac{w s_{yy}}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}} \nabla \partial_X \cdot (a^\partial \partial_{\hat{n}_W}) - \frac{b^\partial w s_{yy}}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}} - \\
\frac{e^\partial \partial_{\hat{n}_R.\partial Z} \cdot \hat{n}_{\partial W}}{\partial_y} \\
-a \frac{\partial \hat{n}_W}{\partial y} \cdot \hat{n}_X - c \frac{\partial \hat{n}_Z}{\partial y} \cdot \hat{n}_X + \nabla \partial_X \cdot (a^\partial \partial_{\hat{n}_W}) + b^\partial + \nabla \partial_X \cdot (c^\partial \partial_{\hat{n}_R.\partial Z}), \\
+ \frac{c z s_{yy} \hat{n}_Z \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_X)^2}} \nabla \partial_X \cdot (c^\partial \partial_{\hat{n}_R.\partial Z}) + \frac{z s_{yy}}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_X)^2}}. \\
\end{align*}
\]

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:

\[
||M^p|| \leq 1 + ||\hat{n}_W||_{C^{r,1}}||w||_{C^{r,1}} + ||w||_{C^{r-1,1}} + \frac{||w \hat{n}_W \cdot \hat{n}_Z||}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_Z)^2}} ||\hat{n}_W \cdot \hat{n}_Z||_{C^{r,1}} + ||w \hat{n}_W \cdot \hat{n}_Z||_{C^{r,1}} + 1 + 1
\]

\[
+ ||\frac{w \hat{n}_W \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}}||_{C^{r-1,1}} + 1 + \frac{||w \hat{n}_W \cdot \hat{n}_Z}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_Z)^2}} ||\hat{n}_W \cdot \hat{n}_X||_{C^{r,1}} + \frac{w \hat{n}_W \cdot \hat{n}_Z}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_Z)^2}} \hat{n}_W \cdot \hat{n}_X + ||w \hat{n}_W \cdot \hat{n}_X||_{C^{r-1,1}}
\]

\[
+ ||\hat{n}_W \cdot \hat{n}_Z \sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2} ||\hat{n}_W \cdot \hat{n}_Z ||_{C^{r,1}}
\]

\[
+ \frac{w \hat{n}_W \cdot \hat{n}_Z}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_Z)^2}} \hat{n}_W \cdot \hat{n}_X = ||w \hat{n}_W \cdot \hat{n}_X||_{C^{r,1}} + 1 + 1.
\]

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. \(\blacksquare\)

We are now ready to prove the Theorem \[\text{18}\] on the regularity of \(G_2\).

**Proof.** The proof is similar to the proof of Theorem \[\text{11}\] for indices \(\alpha, \beta, \gamma\) with \(\alpha + \beta + \gamma \leq r\), we apply the iterated operators \((M^y)^\alpha (M^p)^\beta (M^q)^\gamma\)
to \((h, 0, 0, 0, 0, 0)\), where \(h = h(x, y, q) = f(x) \frac{g\circ y_u(x, y)}{|D_x s_y(x, y)|} \in C^{r,1}\) is the integrand in \(G_2\). Setting
\[(M^\alpha \circ (M^\beta \circ M^\gamma))(h, 0, 0, 0, 0, 0) = (a, b, c, a^\beta, b^\beta, c^\beta) \in B^{r-\gamma}_{\alpha+\beta}\)
we have that
\[
\frac{\partial^{\alpha+\beta+\gamma} G_2}{\partial y^\alpha \partial p^\beta \partial q^\gamma} = \int_{W \cap Z \leq} adH^{m-1}(x) + \int_{W \cap Z =} bdH^{m}(x) + \int_{W \cap Z =} cdH^{m-1}(x)
+ \int_{(W \cap Z \geq) \cap \partial X} a^\beta dH^{m-2}(x) + \int_{(W \cap Z \geq) \cap \partial X} b^\beta dH^{m-1}(x) + \int_{(W \cap Z \geq) \cap \partial X} c^\beta dH^{m-2}(x)
\]
is Lipschitz by Lemma 20 and it’s norm is controlled by the sizes of the domains of integration and the \(L^\infty\) norms of the integrands, which are in turn controlled by the desired quantities as a consequence of iterating Lemma 21. As in Theorem 11 in the previous section, and the corresponding result for one dimensional targets in [5], 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 22 (Bootstrapping smoothness for local ODE)** Assume that the conditions in Theorem 15 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 almost everywhere solution \(v \in C^{1,1}(Y')\) of the \(i = 2\) equation 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 \(H^{m-1}(X_2(y, k(y), k'(y))) \geq C > 0\), where \(C\) depends on \(L_g, U_f, \min |D_x s_y|, \max |D_x s_y|\) and the diameter of \(X\). Noting that
\[
\frac{\partial G_2}{\partial q}(y, p, q) = \int_{X_2(y, p, q)} \frac{f(x)}{|D_x s_y(x, y)|} dH^{m-1}(x),
\]
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, \(q \mapsto G_2(y, p, q)\) 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
\[
k'(y) = q(y, k(y), g(y)). \tag{57}
\]
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 (57) 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 \( v \in C^{r+2,1}(Y') \).

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

Throughout this section, we will assume that \( n = 1 \) (one dimensional target), and the minimizer \((u,v) = (v^s, u^s)\) to (8) satisfies the local equation (21)–(22) for \( i = 2 \) and a.e. \( y \in Y \); this implies that \( \partial^s v(y) = X_2(y, v'(y), v''(y)) \) for almost all \( y \).

We will assume \( s, s_y \in C^2(X \times Y) \), and the probability densities \( f \) and \( g \) are continuous on \( X \) and \( Y \), satisfying bounds

\[
0 < L_f \leq f(x) \leq U_f < \infty \text{ for all } x \in X \tag{58}
\]

\[
0 < L_g \leq g(y) \leq U_g < \infty \text{ for all } y \in Y. \tag{59}
\]

As in Section 6, we assume without loss of generality that \( s \) is everywhere convex with respect to \( y \), so that \( k(y) := v'(y) \) is non-decreasing.

The functional

\[
G_2(y, p, q) := \int_{X_2(y,p,q)} \frac{q - s_{yy}(x,y)}{|D_x s_y(x,y)|} f(x) d\mathcal{H}^{m-1}(x).
\]

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

\[
q \mapsto G_2(y, p, q) = \beta
\]

We therefore have \( q(y, k(y), g(y)) = k'(y) \) almost everywhere. Under the assumptions of Theorem 18 for some \( r \geq 0 \), \( q(y, p, \beta) \) is Lipschitz continuous in all its arguments, by the Clarke implicit function theorem. 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 1.
Proposition 23 (The derivative of $v$ is Lipschitz if continuous)

Let $Y' \subset Y$ and $P' \subset P = \partial_y(X,Y)$ be regions such that $\inf_{y \in Y', p \in P'} \mathcal{H}^{m-1}(X_1(y,p)) > 0$. Then there exist positive constants $C_1$ and $C_2$ such that

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

(60)

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 $\mathcal{H}^m(\bigcup_{p \in [p_0,p_1]} (X_1(y,p))) \sim |p_1 - p_0|$, where the constants of proportionality depend on two-sided bounds for $|D_x s_y|$ and $\mathcal{H}^{m-1}(X_1(y,p))$. The equation

$$0 < g(y) = \int_{X_2(y,p,q(y,p,g(y)))} \frac{q - s_{y,y}(x,y)}{|D_x s_y|} f(x) d\mathcal{H}^m(x)$$

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, $L_y, U_y, ||s||_{C^2}$, and $q(y,p,U_y)$. 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 the Proposition, $y$ 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 (60).

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 $q(y_0, k(y), g(y_0)) - s_{gy}(x, y_0) \geq \alpha > 0$; by uniform continuity, there is a $\delta > 0$ (depending on $\alpha$ but not $y$) such that, we have

$$q(y, k(y), g(y)) - s_{gy}(x, y, y_0) \geq 0;$$

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 (60) 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

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

(61)

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 $\alpha$, 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 $g$, $\min |D_x s_y|$, $\max |D_x s_{yy}|$ and the size of the level sets, $\sup_{(y,p) \in Y \times k(Y)} \mathcal{H}^{m-1}(X_1(y, p))$. It then follows that

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

(62)

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 (60) 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

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

where $D$ depends on the size $\min_{p \in [k_0, k_1]} \mathcal{H}^{m-1}(X_1(y_0, p))$ of the level sets, and the speed limit $\min_{s_y(y_0, x) \in [k_0, k_1]} |D_x s_y(y_0, x)|$. This combined with (61) and (62) establishes the result.
Our strategy is to combine Proposition 23 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 24 (Enhanced twist)** We say \( s \in C^2(\overline{X} \times \overline{Y}) \) satisfies the enhanced twist condition if \( x \in X \) and \( y, \bar{y} \in Y \) imply

\[
D_x s(x, y) - D_x s(x, \bar{y})
\]

cannot be a multiple of \( D_x s_y(x, \bar{y}) \) unless \( y = \bar{y} \).

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 25 (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 \( \partial^* v(\bar{y}) \) in \( X_1(\bar{y}, p) \).

**Proof.** If \( x \) lies in the interior of \( \partial^* v(\bar{y}) \) 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 \( x \in \partial^* v(y) \), 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. \( \blacksquare \)

**Proposition 26 (Continuous differentiability of \( v \))** Let \( (v^i, v) \) achieve the minimum (58) and solve (21)–(22) a.e. on \( Y \) with \( i = 2 \). Let \( Y' \subseteq Y \) and \( X'' := \{ x : d(x, X') < \delta \} \) be a neighbourhood of \( X' := (s^{-1} \circ D v^i)^{-1}(Y') \) for some \( \delta > 0 \). If \( \|s_y\|_{C^2(X'' \times Y')} < \infty \), the enhanced twist condition and bounds (58)–(59) on the continuous probability densities \( f \) and \( g \) imply continuity of \( k = v' \) on \( Y'' \).
**Proof.** Without loss of generality, assume \( s_{yy} \geq 0 \) so that \( k \) is monotone increasing as before. We need only rule out jump discontinuities. Suppose \( k \) has a jump discontinuity at \( y \), 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 \( X_2(y, k(y), k'(y)) = \partial^*v(y) \), the mean value theorem for integrals yields \( x \in \partial^*v(y) \) at which
\[
0 < L_g \leq g(y) = \frac{k'(y) - s_{yy}(x, y)}{|Dxs_y|} f(x) \mathcal{H}^{m-1}[X_2(y, k(y), k'(y))].
\]

Thus \( k'(y) - s_{yy}(x, y) \geq \beta \) holds throughout a ball of radius \( r \) in \( X \), where \( \beta := \frac{L_g}{2c(\nu)} > 0 \) with \( c := \min |Dxs_y(x, y)| \), \( C := \sup_{(y,k) \in Y \times k(Y)} \mathcal{H}^{m-1}(X_1(y, k)), \) and \( r \) depends only on \( \delta \) and \( B := \sup_{(x,y) \in X^r \times Y} |Dxs_{yy}(x, y)| \).

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 \( \bar{x} \in X_1(\bar{y}, k_0) \cap \partial^*v(y) \). 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,
\[
k'(y) - s_{yy}(y, x) \geq \beta > 0 \tag{63}
\]
for all \( x \) near \( \bar{x} \), and almost all \( y < \bar{y} \) near \( \bar{y} \).

In addition, since \( B_r(x_i) \cap X_1(y_i, k(y_i)) \subseteq \partial^*v(y_i) \), and each \( x \in B_r(\bar{x}) \cap X_1(\bar{y}, k_0) \) can be approximated by points \( z_i(x) \in B_r(x_i) \cap X_1(\bar{y}, k(y_i)) \), 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, \( x \in \partial^*v(\bar{y}) \). Therefore, \( B_r(\bar{x}) \cap X_1(\bar{y}, k_0) \subset \partial^*v(\bar{y}) \).

We have now shown that \( \bar{x} \) in the relative interior of \( \partial^*v(\bar{y}) \) in \( X_1(\bar{y}, k_0) \). Lemma 25 therefore implies that the optimal map \( F \) is continuous at \( \bar{x} \). We next show that all points in the open set \( X_\epsilon(\bar{y}, k_0) := \{x \in X \mid s_y(x, \bar{y}) > k_0\} \) 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 (63) implies that
\[
k_0 - s_y(\bar{x}, \bar{y}) - [k(y) - s_y(\bar{x}, y)] \geq \int_y^{\bar{y}} [k'(s) - s_{yy}(\bar{x}, s)]ds
\geq \beta|\bar{y} - y|
\]
As \( k_0 = s_y(\bar{x}, \bar{y}) \), this means, for almost every \( y < \bar{y} \), with \( y \) close to \( \bar{y} \)
\[s_y(\bar{x}, y) - s_y(\bar{x}, x, y) = s_y(\bar{x}, y) - k(y) \geq \beta|\bar{y} - y|.
\]
Therefore,
\[
s_y(\bar{x}, \bar{y}) - s_y(x, \bar{y}) = s_y(\bar{x}, y) - s_y(x, y) + \int_y^{\bar{y}} [s_{yy}(\bar{x}, y) - s_{yy}(x, t)]dt
\geq \beta|\bar{y} - y| - B|\bar{y} - y||\bar{x} - x|
= |\bar{y} - y|(\beta - B|\bar{x} - x|) > 0.
\]
for \(|x - \bar{x}| \) sufficiently small. This contradicts that assumption \( x \in X(y, k_0) \).

To summarize, we have shown that for \( x \in X(y, k_0) \) 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 27 (Lipschitz differentiability of \( v \))** Fix open sets \( X \subset \mathbb{R}^m \) and \( Y \subset \mathbb{R} \) equipped with continuous probability densities which are bounded away from zero and infinity. Let \((v^*, v)\) achieve the minimum \[\mathcal{L}\text{ and solve } (21) - (22) \text{ a.e. on } Y \text{ with } i = 2. \text{ Let } Y' \subset Y \text{ and } P' = v'(Y') \text{ be regions such that } \inf_{y \in Y', p \in P'} \mathcal{H}^{m-1}(X_1(y, p)) > 0 \text{ and } X'' \text{ as in Proposition 28. }
\]
Under the enhanced twist condition, if \( \|s_y\|_{C^2(\bar{X}_\nu \times \bar{Y}')} < \infty \) 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 \( \partial^2 v(y) = X_2(y, k(y), k'(y)) = X_2(y, k(y), q(y, k(y), g(y))) \) for a.e. \( y \), mass balance and Propositions 23 and 28 combine to imply
\[
U_g|_{y_1 - y_0} \geq \int_{y_0}^{y_1} g(y)dy
= \int_{\cup_{y \in [y_0, y_1]} X_2(y, k, q(y, k(y), g(y)))} f(x)d\mathcal{H}^m(x)
\geq L_f \text{vol}[\cup_{y \in [y_0, y_1]} X_2(y, k, q(y, k(y), g(y)))]
\geq L_f [C_1|k(y_0) - k(y_1)| - C_2|y_0 - y_1|] \quad (67)
\]
This is the desired conclusion. 

37
Remark 28 Combined with Corollary 22, this yields conditions under which any solution to the $i = 2$ local equation is $C^{2,1}$, or in fact smoother, depending on $G_2$

## Appendices

### Appendix A  Formulas for partial derivatives

The partial derivatives of the functions defined in Lemma 20 are given almost everywhere by the following formulas, with $w(x, y) = |D_x s_y|^{-1}$:

\[
A_p = \int_{X_2} a_p d\mathcal{H}^{m-1} = \frac{\partial}{\partial p} \bigg|_{p=p} \int_{W_=(y,\tilde{p}) \cap Z_{\le}(y,\tilde{q})} a(x, y, p, q) \hat{n}_W \cdot \hat{n}_W d\mathcal{H}^{m-1}(x)
\]

\[
= \frac{\partial}{\partial \tilde{p}} \left[ \int_{W_\le \cap Z_{\le}} \nabla \cdot (a\hat{n}_W) d\mathcal{H}^m - \int_{W_\le \cap Z_{=} \cap \partial X} a\hat{n}_W \cdot \hat{n}_Z d\mathcal{H}^{m-1} \right]_{\tilde{p}=p}
\]

\[
= \int_{W_\le \cap Z_{\le}} \nabla \cdot (a\hat{n}_W) w d\mathcal{H}^{m-1} - \int_{W_\le \cap Z_{=} \cap \partial X} \frac{aw\hat{n}_W \cdot \hat{n}_Z}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}} d\mathcal{H}^{m-2}
\]

\[
- \int_{W_\le \cap Z_{\le} \cap \partial X} \frac{aw\hat{n}_W \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}} d\mathcal{H}^{m-2},
\]

\[
A_q = \int_{X_2} a_q d\mathcal{H}^{m-1} = \frac{\partial}{\partial q} \bigg|_{q=q} \int_{W_=(y,\tilde{q}) \cap Z_{\le}(y,\tilde{q})} a(x, y, p, q) d\mathcal{H}^{m-1}(x)
\]

\[
= \int_{W_\le \cap Z_{=} \cap \partial X} \frac{aw\hat{n}_W \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}} d\mathcal{H}^{m-2},
\]

\[
A_y = \int_{X_2} a_y d\mathcal{H}^{m-1} + \int_{W_\le \cap Z_{=} \cap \partial X} \frac{aw s_{yy}}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_W)^2}} d\mathcal{H}^{m-2} = \frac{\partial}{\partial y} \bigg|_{y=y} \int_{W_=(y,\tilde{y}) \cap Z_{\le}(y,\tilde{q})} a(x, y, p, q) d\mathcal{H}^{m-1}(x)
\]

\[
= \frac{\partial}{\partial \tilde{y}} \left[ \int_{W_\le \cap Z_{\le}} \nabla \cdot (a\hat{n}_W) d\mathcal{H}^m - \int_{W_\le \cap Z_{=} \cap \partial X} a\hat{n}_W \cdot \hat{n}_Z d\mathcal{H}^{m-1} \right]_{\tilde{y}=y}
\]

\[
= \int_{W_\le \cap Z_{=} \cap \partial X} \frac{aw\hat{n}_W \cdot \hat{n}_Z}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_Z)^2}} d\mathcal{H}^{m-2}
\]

\[
- \int_{W_\le \cap Z_{=} \cap \partial X} \frac{aw s_{yy} \hat{n}_W \cdot \hat{n}_Z}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_Z)^2}} d\mathcal{H}^{m-2}
\]

\[
+ \int_{W_\le \cap Z_{=} \cap \partial X} \frac{aw s_{yy} \hat{n}_W \cdot \hat{n}_X}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}} d\mathcal{H}^{m-2},
\]

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

38
\[ B_p = \int_{W \leq (y,q) \cap Z \leq (y,q)} b_p dH^m(x) + \int_{W = (y,p) \cap Z \leq (y,q)} bw dH^{m-1}(x), \]

\[ B_q = \int_{W \leq (y,p) \cap Z \leq (y,q)} b_q dH^m(x) + \int_{W = (y,p) \cap Z = (y,q)} bz dH^{m-1}(x), \]

\[ B_y = \int_{W \leq (y,p) \cap Z \leq (y,q)} b_y dH^m(x) - \int_{W = (y,p) \cap Z \leq (y,q)} bw s_{yy} dH^{m-1}(x) - \int_{W = (y,p) \cap Z \leq (y,q)} b z s_{yy} dH^m(x), \]

\[ C_p = \int_{W \leq (y,p) \cap Z = (y,q)} c_p dH^{m-1}(x) + \int_{W = (y,p) \cap Z = (y,q)} cw dH^{m-2}(x), \]

\[ C_q = \int_{W \leq (y,p) \cap Z = (y,q)} c_q dH^{m-1}(x) + \int_{W = (y,p) \cap Z = (y,q)} \nabla \cdot (c \hat{\eta} Z) z dH^{m-1}(x) \]
\[ - \int_{W = (y,p) \cap Z = (y,q)} \frac{cz \hat{n} Z \cdot \hat{n} W}{\sqrt{1 - (\hat{n} W \cdot \hat{n} Z)^2}} dH^{m-2}(x) - \int_{W = (y,p) \cap Z \leq (y,q)} \frac{cz \hat{n} Z \cdot \hat{n} X}{\sqrt{1 - (\hat{n} Z \cdot \hat{n} X)^2}} dH^{m-2}, \]

\[ C_y = \int_{W \leq Z \leq (y,p)} c_y dH^{m-1} - \int_{W \cap Z \leq (y,p)} \frac{cz s_{yy} \hat{n} W \cdot \hat{n} Z}{\sqrt{1 - (\hat{n} W \cdot \hat{n} Z)^2}} dH^{m-2} \]
\[ + \int_{W \leq Z \leq (y,p)} \nabla \cdot (c \frac{\partial \hat{n} Z}{\partial y}) dH^m - \int_{W \cap Z \leq (y,p)} c \frac{\partial \hat{n} Z}{\partial y} \cdot \hat{n} W dH^{m-1} - \int_{W \cap Z \leq (y,p)} c \frac{\partial \hat{n} Z}{\partial y} \cdot \hat{n} X dH^{m-1} \]
\[ - \int_{W \cap Z \leq (y,p)} \frac{cz s_{yy} \hat{n} W \cdot \hat{n} Z}{\sqrt{1 - (\hat{n} W \cdot \hat{n} Z)^2}} dH^{m-2} + \int_{W \cap Z \leq (y,p)} \frac{cz s_{yy} \hat{n} W \cdot \hat{n} Z}{\sqrt{1 - (\hat{n} W \cdot \hat{n} Z)^2}} dH^{m-2}, \]

\[ A_p^0 = \int_{(W \cap Z \leq (y,p)) \cap \partial X} a_p^0 dH^{m-2}(x) + \int_{(W \cap Z \leq (y,p)) \cap \partial X} \frac{w}{\sqrt{1 - (\hat{n} W \cdot \hat{n} X)^2}} \nabla \cdot (a_p^0 \hat{n}_{\partial W} dH^{m-2}(x) \]
\[ - \int_{(W \cap Z \leq (y,p)) \cap \partial X} a_p^0 \hat{n}_{\partial W} \cdot \hat{n}_{\partial X} \frac{w}{\sqrt{1 - (\hat{n} W \cdot \hat{n} X)^2} \sqrt{1 - (\hat{n} W \cdot \hat{n} Z)^2}} dH^{m-3}(x), \]

\[ A_q^0 = \int_{(W \cap Z \leq (y,p)) \cap \partial X} a_q^0 dH^{m-2}(x) + \int_{(W \cap Z \leq (y,p)) \cap \partial X} \frac{z}{\sqrt{1 - (\hat{n} Z \cdot \hat{n} X)^2} \sqrt{1 - (\hat{n} W \cdot \hat{n} Z)^2}} dH^{m-3}(x), \]
\[
A_y^\partial = \int_{(W_\omega \cap Z_\omega) \cap \partial X} a_\partial^y d\mathcal{H}^{m-2}(x) - \int_{(W_\omega \cap Z_\omega) \cap \partial X} a_\partial^y \frac{z_{swy}}{\sqrt{1 - (\hat{n}_z \cdot \hat{n}_x)^2}} d\mathcal{H}^{m-3}(x) + \int_{(W_\omega \cap Z_\omega) \cap \partial X} \nabla_{\partial X} \cdot (a_\partial^y \hat{n}_\partial \omega \cdot \partial x) d\mathcal{H}^{m-2}(x) - \int_{(W_\omega \cap Z_\omega) \cap \partial X} a_\partial^y \frac{\partial \hat{n}_\partial \omega \cdot \hat{n}_\partial \omega \cdot \partial x}{\sqrt{1 - (\hat{n}_z \cdot \hat{n}_x)^2}} d\mathcal{H}^{m-2}(x)
\]

\[
B_p^\partial = \int_{(W_\omega \cap Z_\omega) \cap \partial X} b_p^\partial d\mathcal{H}^{m-1}(x) + \int_{(W_\omega \cap Z_\omega) \cap \partial X} b_p^\partial \frac{b_w}{\sqrt{1 - (\hat{n}_w \cdot \hat{n}_x)^2}} d\mathcal{H}^{m-2}(x),
\]

\[
B_q^\partial = \int_{(W_\omega \cap Z_\omega) \cap \partial X} b_q^\partial d\mathcal{H}^{m-1}(x) + \int_{(W_\omega \cap Z_\omega) \cap \partial X} b_q^\partial \frac{b_z}{\sqrt{1 - (\hat{n}_z \cdot \hat{n}_x)^2}} d\mathcal{H}^{m-2}(x),
\]

\[
B_y^\partial = \int_{(W_\omega \cap Z_\omega) \cap \partial X} b_y^\partial d\mathcal{H}^{m-1}(x) - \int_{(W_\omega \cap Z_\omega) \cap \partial X} b_y^\partial \frac{b_{swy}}{\sqrt{1 - (\hat{n}_w \cdot \hat{n}_x)^2}} d\mathcal{H}^{m-2}(x) - \int_{(W_\omega \cap Z_\omega) \cap \partial X} b_y^\partial \frac{z_{swy}}{\sqrt{1 - (\hat{n}_z \cdot \hat{n}_x)^2}} d\mathcal{H}^{m-2}(x),
\]

\[
C_p^\partial = \int_{(W_\omega \cap Z_\omega) \cap \partial X} c_p^\partial d\mathcal{H}^{m-2}(x) + \int_{(W_\omega \cap Z_\omega) \cap \partial X} c_p^\partial \frac{w}{\sqrt{1 - (\hat{n}_w \cdot \hat{n}_x)^2}} d\mathcal{H}^{m-3}(x),
\]

\[
C_q^\partial = \int_{(W_\omega \cap Z_\omega) \cap \partial X} c_q^\partial d\mathcal{H}^{m-2}(x) + \int_{(W_\omega \cap Z_\omega) \cap \partial X} c_q^\partial \frac{z}{\sqrt{1 - (\hat{n}_z \cdot \hat{n}_x)^2}} d\mathcal{H}^{m-2}(x) - \int_{(W_\omega \cap Z_\omega) \cap \partial X} c_q^\partial \frac{z_{\hat{n}_\partial \omega \cdot \hat{n}_\partial \omega \cdot \partial x}}{\sqrt{1 - (\hat{n}_z \cdot \hat{n}_x)^2}} d\mathcal{H}^{m-3}(x),
\]

where \( \hat{n}_\partial \omega := \frac{\hat{n}_w - (\hat{n}_w \cdot \hat{n}_x) \hat{n}_x}{\sqrt{1 - (\hat{n}_w \cdot \hat{n}_x)^2}} \) and \( \hat{n}_\partial \omega \cdot \hat{n}_\partial \omega \cdot \partial x \) are the outward unit normals to \( W_\omega \cap \partial X \) and \( Z_\omega \cap \partial X \) in \( \partial X \), respectively and
\[ C_y^\partial = \int_{(W \leq Z \cap \partial X)} c^\partial d\mathcal{H}^{m-2}(x) - \int_{(W = Z \cap \partial X)} c^\partial \frac{w S_{yy}}{\sqrt{1 - (\hat{n}_W \cdot \hat{n}_X)^2}} \frac{w S_{yy}}{\sqrt{1 - (\hat{n}_{\partial, W} \cdot \hat{n}_{\partial, Z})^2}} d\mathcal{H}^{m-3}(x) + \int_{(W = Z \cap \partial X)} \nabla_{\partial X} \cdot (c^\partial \hat{n}_{\partial, Z}) d\mathcal{H}^{m-1}(x) - \int_{(W = Z \cap \partial X)} c^\partial \frac{\hat{n}_{\partial, Z}}{\partial y} \cdot \hat{n}_{\partial, W} d\mathcal{H}^{m-2}(x) - \int_{(W \leq Z \cap \partial X)} \nabla_{\partial X} \cdot (c^\partial \hat{n}_{\partial, Z}) d\mathcal{H}^{m-1}(x) + \int_{(W \leq Z \cap \partial X)} \frac{z S_{yy}}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_X)^2}} \nabla_{\partial X} \cdot (c^\partial \hat{n}_{\partial, Z}) d\mathcal{H}^{m-3}(x) + \int_{(W \leq Z \cap \partial X)} \frac{z S_{yy}}{\sqrt{1 - (\hat{n}_Z \cdot \hat{n}_X)^2}} \frac{z S_{yy}}{\sqrt{1 - (\hat{n}_{\partial, W} \cdot \hat{n}_{\partial, Z})^2}} c^\partial \hat{n}_{\partial, W} \cdot \hat{n}_{\partial, Z} d\mathcal{H}^{m-3}(x). \]

**References**


