Some nonlinear Brascamp–Lieb inequalities and applications to harmonic analysis

Bennett, Jonathan; Bez, Richard

DOI:
10.1016/j.jfa.2010.07.015

SOME NONLINEAR BRASCAMP–LIEB INEQUALITIES AND APPLICATIONS TO HARMONIC ANALYSIS

JONATHAN BENNETT AND NEAL BEZ

Abstract. We use the method of induction-on-scales to prove certain diffeomorphism invariant nonlinear Brascamp–Lieb inequalities. We provide applications to multilinear convolution inequalities and the restriction theory for the Fourier transform, extending to higher dimensions recent work of Bejenaru–Herr–Tataru and Bennett–Carbery–Wright.

1. Introduction

The purpose of this paper is to obtain nonlinear generalisations of certain Brascamp–Lieb inequalities and apply them to some well-known problems in euclidean harmonic analysis. Our particular approach to such inequalities is by induction-on-scales, and builds on the recent work of Bejenaru, Herr and Tataru [4].

The Brascamp–Lieb inequalities simultaneously generalise important classical inequalities such as the multilinear Hölder, sharp Young convolution and Loomis–Whitney inequalities. They may be formulated as follows. Suppose $m \geq 2$ and $d, d_1, \ldots, d_m$ are positive integers, and for each $1 \leq j \leq m$, $B_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$ is a linear surjection and $p_j \in [0,1]$. The Brascamp–Lieb inequality associated with these objects takes the form

$$\int_{\mathbb{R}^d} \prod_{j=1}^m (f_j \circ B_j)^{p_j} \leq C \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} f_j \right)^{p_j} \tag{1}$$

for all nonnegative $f_j \in L^1(\mathbb{R}^{d_j})$, $1 \leq j \leq m$. Here C denotes a constant depending on the datum $(\mathbf{B}, \mathbf{p}) := ((B_j), (p_j))$, which at this level of generality may of course be infinite. For nonnegative functions $f_j \in L^1(\mathbb{R}^{d_j})$ satisfying $0 < \int f_j < \infty$, we define the quantity

$$\mathrm{BL}(\mathbf{B}, \mathbf{p}; \mathbf{f}) = \frac{\int_{\mathbb{R}^d} \prod_{j=1}^m (f_j \circ B_j)^{p_j}}{\prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} f_j \right)^{p_j}},$$

where $\mathbf{f} := (f_j)$. We may then define the Brascamp–Lieb constant $0 < \mathrm{BL}(\mathbf{B}, \mathbf{p}) \leq \infty$ to be the supremum of $\mathrm{BL}(\mathbf{B}, \mathbf{p}; \mathbf{f})$ over all such inputs $\mathbf{f}$. The quantity $\mathrm{BL}(\mathbf{B}, \mathbf{p})$ is of course the smallest $0 < C \leq \infty$ for which (1) holds. It should be noted here that there is a natural equivalence relation on Brascamp–Lieb data, where $(\mathbf{B}, \mathbf{p}) \sim$

Date: 29th June 2010.

2000 *Mathematics Subject Classification.* 44A12; 42B10; 44A35.

Key words and phrases. Brascamp–Lieb inequalities, induction-on-scales, Fourier extension estimates.

Both authors were supported by EPSRC grant EP/E022340/1.

$(\mathbf{B}', \mathbf{p}')$ if $\mathbf{p} = \mathbf{p}'$ and there exist invertible linear transformations $C : \mathbb{R}^d \to \mathbb{R}^d$ and $C_j : \mathbb{R}^{d_j} \to \mathbb{R}^{d_j}$ such that $B_j' = C_j^{-1} B_j C$ for all j; we refer to C and C_j as the intertwining transformations. In this case, simple changes of variables show that

$$\mathrm{BL}(\mathbf{B}', \mathbf{p}') = \frac{\prod_{j=1}^m |\det C_j|^{p_j}}{|\det C|} \mathrm{BL}(\mathbf{B}, \mathbf{p}),$$

and thus $\mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty$ if and only if $\mathrm{BL}(\mathbf{B}', \mathbf{p}') < \infty$. This terminology is taken from [5].

The generality of this setup of course raises questions, many of which have been addressed in the literature. In [15] Lieb showed that the supremum above is exhausted by centred gaussian inputs, prompting further investigation into issues including the finiteness of $\mathrm{BL}(\mathbf{B}, \mathbf{p})$ and the extremisability/gaussian-extremisability of $\mathrm{BL}(\mathbf{B}, \mathbf{p}; \mathbf{f})$. A fuller description of the literature is not appropriate for the purposes of this paper. The reader is referred to the survey article [2] and the references there.

A large number of problems in harmonic analysis require nonlinear versions of inequalities belonging to this family; see [3], [4], [7], [14], [18], and [23] for instance. The generalisations we seek here are local in nature, and amount to allowing the maps B_j to be nonlinear submersions in a neighbourhood of a point $x_0 \in \mathbb{R}^d$, and then looking for a neighbourhood U of x_0 such that if ψ is a cutoff function supported in U, there exists a constant $C > 0$ for which

$$\int_{\mathbb{R}^d} \prod_{j=1}^m f_j(B_j(x))^{p_j} \psi(x) \, \mathrm{d}x \leq C \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} f_j \right)^{p_j} \tag{2}$$

for all nonnegative $f_j \in L^1(\mathbb{R}^{d_j})$, $1 \leq j \leq m$. The applications of such inequalities invariably require more quantitative statements involving the sizes of the neighbourhood U and constant C, and also the nature of any smoothness/non-degeneracy conditions imposed on the nonlinear maps (B_j).

Notice that if $d_j = d$ for each j, then the nonlinear B_j are of course local diffeomorphisms. In this situation necessarily $p_1 + \cdots + p_m = 1$ and (2) follows from the m-linear Hölder inequality. Similar considerations allow to reduce matters to the case where $d_j < d$ for all j.

It is perhaps reasonable to expect to obtain an inequality of the form (2) for smooth nonlinear maps (B_j) and exponents (p_j) for which $\mathrm{BL}((\mathrm{d}B_j(x_0)), (p_j)) < \infty$. Here $\mathrm{d}B_j(x_0)$ denotes the derivative map of B_j at x_0. However, the techniques that we employ in this paper appear to require additional structural hypotheses on the maps $\mathrm{d}B_j(x_0)$, and so instead we seek to identify a natural class

$$\mathcal{C} \subseteq \{(\mathbf{B}, \mathbf{p}) : \text{each } B_j \text{ is } \textit{linear} \text{ and } \mathrm{BL}(\mathbf{B}, \mathbf{p}) < \infty\}$$

such that (2) holds for nonlinear (B_j) with $((\mathrm{d}B_j(x_0)), (p_j)) \in \mathcal{C}$. As will become clear in Section 2, a natural choice for consideration is

$$\mathcal{C} = \left\{ (\mathbf{B}, \mathbf{p}) : \bigoplus_{j=1}^m \ker B_j = \mathbb{R}^d, \;\; p_1 = \cdots = p_m = \tfrac{1}{m-1} \right\}. \tag{3}$$

This class contains the classical Loomis–Whitney datum [16], whereby $m = d$, $d_j = d-1$, $p_j = 1/(d-1)$ and $B_j(x_1, \dots, x_d) = (x_1, \dots, \widehat{x_j}, \dots, x_d)$ for all $1 \le j \le d$. Here $\widehat{\ }$ denotes omission.

The purpose of this paper is two-fold. Firstly, we establish an inequality of the form (2) whenever $((\mathrm{d}B_j(x_0)), (p_j)) \in \mathcal{C}$, where $\mathcal{C}$ is defined in (3). Secondly, we use these inequalities to deduce certain sharp multilinear convolution estimates, which in turn yield progress on the multilinear restriction conjecture for the Fourier transform. These applications can be found in Section 7.

Before stating our nonlinear Brascamp–Lieb inequalities, it is important that we discuss further the class $\mathcal{C}$ given in (3). Notice that the transversality hypothesis

$$\bigoplus_{j=1}^{m} \ker B_j = \mathbb{R}^d \tag{4}$$

is preserved under the equivalence relation on Brascamp–Lieb data; that is, it is invariant under $B_j \mapsto C_j^{-1} B_j C$ for invertible linear transformations $C : \mathbb{R}^d \to \mathbb{R}^d$ and $C_j : \mathbb{R}^{d_j} \to \mathbb{R}^{d_j}$. By choosing appropriate intertwining transformations C and C_j, an elementary calculation shows that if $(\mathbf{B}, \mathbf{p}) \in \mathcal{C}$ then $(\mathbf{B}, \mathbf{p}) \sim (\mathbf{\Pi}, \mathbf{p})$, where $\mathbf{\Pi} = (\Pi_j)_{j=1}^m$ are certain coordinate projections. In order to define Π_j we let $\mathcal{K}_j \subseteq \{1, \dots, d\}$ be given by

$$\mathcal{K}_j = \{d_1' + \cdots + d_{j-1}' + 1, \dots, d_1' + \cdots + d_{j-1}' + d_j'\},$$

where $d_j' = d - d_j$ denotes the dimension of the kernel of B_j, so that $\mathcal{K}_1, \dots, \mathcal{K}_m$ form a partition of $\{1, \dots, d\}$. Then we let $\Pi_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$ be given by

$$\Pi_j(x) = (x_k)_{k \in \mathcal{K}_j^c}. \tag{5}$$

Proposition 1.1. [13] *If* $\mathbf{p} = (\frac{1}{m-1}, \dots, \frac{1}{m-1})$ *then* $\mathrm{BL}(\mathbf{\Pi}, \mathbf{p}) = 1$*, and thus*

$$\int_{\mathbb{R}^d} \prod_{j=1}^{m} f_j(\Pi_j x)^{\frac{1}{m-1}} \, \mathrm{d}x \le \prod_{j=1}^{m} \Big(\int_{\mathbb{R}^{d_j}} f_j \Big)^{\frac{1}{m-1}} \tag{6}$$

holds for all nonnegative $f_j \in L^1(\mathbb{R}^{d_j}), 1 \le j \le m$.

Proposition 1.1 follows from work of Finner [13] where a stronger result was established for $\mathbf{\Pi}$ consisting of more general coordinate projections and in the broader setting of product measure spaces. In particular, this includes the discrete inequality

$$\sum_{n \in \mathbb{N}^d} \prod_{j=1}^{m} f_j(\Pi_j n)^{\frac{1}{m-1}} \le \prod_{j=1}^{m} \Big(\sum_{\ell \in \mathbb{N}^{d_j}} f_j(\ell) \Big)^{\frac{1}{m-1}} \tag{7}$$

which holds for all nonnegative $f_j \in \ell^1(\mathbb{N}^{d_j}), 1 \le j \le m$. We mention this case specifically as it will be important later in the paper.

We remark that (6) is a generalisation of the classical Loomis–Whitney inequality [16] whereby $m = d$ and $\mathcal{K}_j = \{j\}$ for $1 \le j \le d$.

In order for $\mathrm{BL}(\mathbf{\Pi}, \mathbf{p})$ to be finite it is necessary that $\mathbf{p} = (\frac{1}{m-1}, \dots, \frac{1}{m-1})$, and this follows by a straightforward scaling argument.

The standard proof of Proposition 1.1 proceeds via the multilinear Hölder inequality and induction (see [13]). This proof and, to the best of our knowledge, other established proofs of Proposition 1.1 rely heavily on the linearity of the Π_j and break down completely in the nonlinear setting.

Since we would like to state our main theorem regarding nonlinear B_j in a diffeomorphism-invariant way, it is appropriate that we first formulate an affine-invariant version of Proposition 1.1. In order to state this it is natural to use language from exterior algebra; the relevant concepts and terminology can be found in standard texts such as [12]. In particular, $\Lambda^n(\mathbb{R}^d)$ will denote the nth exterior algebra of $\mathbb{R}^d$ and $\star : \Lambda^n(\mathbb{R}^d) \to \Lambda^{d-n}(\mathbb{R}^d)$ will denote the Hodge star operator. (It is worth pointing out here that if the reader is prepared to sacrifice the explicit diffeomorphism-invariance that we seek, then they may effectively dispense with these exterior algebraic considerations.) Given $(\mathbf{B}, \mathbf{p}) \in \mathcal{C}$ define $X_j(B_j) \in \Lambda^{d_j}(\mathbb{R}^d)$ to be the wedge product of the rows of the $d_j \times d$ matrix B_j. By (4) it follows that

$$\star \bigwedge_{j=1}^{m} \star X_j(B_j) \in \mathbb{R}\backslash\{0\}. \tag{8}$$

The quantity in (8) is a certain determinant and should be viewed as a means of quantifying the transverality hypothesis (4).

Proposition 1.2. *If $(\mathbf{B}, \mathbf{p}) \in \mathcal{C}$ then*

$$\mathrm{BL}(\mathbf{B}, \mathbf{p}) = \left| \star \bigwedge_{j=1}^{m} \star X_j(B_j) \right|^{-\frac{1}{m-1}},$$

and thus

$$\int_{\mathbb{R}^d} \prod_{j=1}^{m} f_j(B_j x)^{\frac{1}{m-1}} \, \mathrm{d}x \leq \left| \star \bigwedge_{j=1}^{m} \star X_j(B_j) \right|^{-\frac{1}{m-1}} \prod_{j=1}^{m} \left(\int_{\mathbb{R}^{d_j}} f_j \right)^{\frac{1}{m-1}} \tag{9}$$

for all nonnegative $f_j \in L^1(\mathbb{R}^{d_j})$, $1 \leq j \leq m$.

One may reduce Proposition 1.2 to Proposition 1.1 by appropriate linear changes of variables; see Appendix A for full details of this argument which will be of further use in Section 4 for the nonlinear case.

Since the inequality (9) is affine-invariant, one should expect it to have a diffeomorphism-invariant nonlinear version. This is our main result with regard to nonlinear generalisations of Brascamp–Lieb inequalities.

Theorem 1.3. *Let $\beta, \varepsilon, \kappa > 0$ be given. Suppose that $B_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$ is a $C^{1,\beta}$ submersion satisfying $\|B_j\|_{C^{1,\beta}} \leq \kappa$ in a neighbourhood of a point $x_0 \in \mathbb{R}^d$ for each $1 \leq j \leq m$. Suppose further that*

$$\bigoplus_{j=1}^{m} \ker \mathrm{d}B_j(x_0) = \mathbb{R}^d \tag{10}$$

and

$$\left| \star \bigwedge_{j=1}^{m} \star X_j(\mathrm{d}B_j(x_0)) \right| \geq \varepsilon.$$

Then there exists a neighbourhood U of x_0 depending on at most $\beta, \varepsilon, \kappa$ and d, such that for all cutoff functions ψ supported in U, there is a constant C depending only on d and ψ such that

$$\int_{\mathbb{R}^d} \prod_{j=1}^{m} f_j(B_j(x))^{\frac{1}{m-1}} \psi(x)\,\mathrm{d}x \leq C\varepsilon^{-\frac{1}{m-1}} \prod_{j=1}^{m} \left(\int_{\mathbb{R}^{d_j}} f_j \right)^{\frac{1}{m-1}} \tag{11}$$

for all nonnegative $f_j \in L^1(\mathbb{R}^{d_j})$, $1 \leq j \leq m$.

Inequality (11) may be interpreted as a multilinear "Radon-like" transform estimate. This is made explicit in the following corollary, upon which our applications in Section 7 depend.

Corollary 1.4. *Let $\beta, \varepsilon, \kappa > 0$ be given. If $F : (\mathbb{R}^{d-1})^{d-1} \to \mathbb{R}$ is such that $\|F\|_{C^{1,\beta}} \leq \kappa$ and*

$$|\det(\nabla_{u_1} F(0), \ldots, \nabla_{u_{d-1}} F(0))| \geq \varepsilon,$$

then there exists a neighbourhood V of the origin in $(\mathbb{R}^{d-1})^{d-1}$, depending only on $\beta, \varepsilon, \kappa$ and d, and a constant C depending only on d, such that

$$\int_V f_1(u_1) \cdots f_{d-1}(u_{d-1}) f_d(u_1 + \cdots + u_{d-1}) \delta(F(u))\,\mathrm{d}u \leq C\varepsilon^{-\frac{1}{d-1}} \prod_{j=1}^{d} \|f_j\|_{(d-1)'} \tag{12}$$

for all nonnegative $f_j \in L^{(d-1)'}(\mathbb{R}^{d-1})$, $1 \leq j \leq m$.

The case $d = 3$ of Corollary 1.4 was proved in [7] as a consequence of the nonlinear Loomis–Whitney inequality.

It is perhaps interesting to view Corollary 1.4 in the light of the theory of multilinear weighted convolution inequalities for L^2 functions developed in [19]. Inequality (12) is an example of such a convolution inequality in an L^p setting and with a singular (distributional) weight.

We conclude this section with a number of remarks on Theorem 1.3.

As in the reduction of Proposition 1.2 to Proposition 1.1, a linear change of variables argument shows that Theorem 1.3 may be reduced to the case where each linear mapping $\mathrm{d}B_j(x_0)$ is equal to the coordinate projection Π_j given by (5), in which case

$$\star \bigwedge_{j=1}^{m} \star X_j(\mathrm{d}B_j(x_0)) = 1.$$

Although this reduction is not essential, it does lead to some conceptual and notational simplification in the subsequent analysis. The details of this reduction may be found in Section 4.

The core component of the proof of Theorem 1.3 that we present is based on [4] and uses the idea of induction-on-scales. This approach provides additional information

about the sizes of the neighbourhood U and constant C appearing in its statement; see Section 4 for further details of this. In Section 2 we offer an explanation of why the induction-on-scales approach is natural in the context of Brascamp–Lieb inequalities and why the class $\mathcal{C}$ given in (3) is a natural class for consideration. In Section 3, we provide an outline of the proof of Theorem 1.3 which should guide the reader through the full proof which is contained in Sections 4 and 5.

In the case where $d_j = d-1$ for all j, Theorem 1.3 reduces to the nonlinear Loomis–Whitney inequality in [7] except that the stronger hypothesis $B_j \in C^3$ is assumed in [7]. The proof of the result in [7] is quite different from the proof we give here, and is based on the so-called method of refinements of M. Christ [11]. We make some further remarks on the role of the smoothness of the mappings B_j at the end of Section 5.

The condition (10) is somewhat less restrictive than it may appear. For example, consider smooth mappings $B_j : \mathbb{R}^5 \to \mathbb{R}^2$ satisfying

$$\ker \mathrm{d}B_j(x_0) = \langle \{ e_j, e_{(j+1) \bmod 5}, e_{(j+2) \bmod 5} \} \rangle$$

for each $1 \leq j \leq 5$, where e_j denotes the jth standard basis vector in $\mathbb{R}^5$. Evidently the condition (10) is not satisfied. However we may write

$$\prod_{j=1}^{5} (f_j \circ B_j)^{1/2} = \prod_{j=1}^{5} (\widetilde{f}_j \circ \widetilde{B}_j)^{1/4},$$

where $\widetilde{f}_j := f_j \otimes f_{(j+2) \bmod 5} : \mathbb{R}^4 \to [0, \infty)$ and $\widetilde{B}_j := (B_j, B_{(j+2) \bmod 5}) : \mathbb{R}^5 \to \mathbb{R}^4$. Since $\ker \mathrm{d}\widetilde{B}_j(x_0) = \langle \{ e_{(j+2) \bmod 5} \} \rangle$ for each $1 \leq j \leq 5$, the mappings $\widetilde{B}_j$ do satisfy the condition (10), and so by Theorem 1.3

$$\begin{aligned} \int_{\mathbb{R}^5} \prod_{j=1}^{5} (f_j \circ B_j)^{1/2} \psi &= \int_{\mathbb{R}^5} \prod_{j=1}^{5} (\widetilde{f}_j \circ \widetilde{B}_j)^{1/4} \psi \\ &\leq C \prod_{j=1}^{5} \left(\int_{\mathbb{R}^4} \widetilde{f}_j \right)^{1/4} \\ &= C \prod_{j=1}^{5} \left(\int_{\mathbb{R}^2} f_j \right)^{1/2}. \end{aligned}$$

Here the cutoff function ψ and constant C are as in the statement of Theorem 1.3. This inequality is optimal in the sense that $\mathrm{BL}(\mathrm{d}B_j(x_0)), (p_j)) < \infty$ if and only if $p_1 = \cdots = p_5 = 1/2$ – see [13]. Similar considerations form an important part of the proof of Corollary 1.4 in dimensions $d \geq 4$.

Very recently, Stovall [18] considered inequalities of the type (2) for the case $d_j = d-1$ for all j where one does not necessarily have the transversality hypothesis (10). Here, curvature of the fibres of the B_j plays a crucial role. In [18], Stovall determined completely all data $(\mathbf{B}, \mathbf{p})$, up to endpoints in $\mathbf{p}$, for which inequality (2) holds when each $B_j : \mathbb{R}^d \to \mathbb{R}^{d-1}$ is a smooth submersion. The work in [18] generalised work of Tao and Wright [23] for the bilinear case $m = 2$, and both approaches are based on Christ's method of refinements. It would be interesting to complete the picture further and understand the case where one does not necessarily

have transversality and each d_j is not necessarily equal to $d-1$. We do not pursue this matter here.

Given that Theorem 1.3 is a local result it is natural to ask whether one may obtain global versions based on the assumption that hypothesis (4) holds at every point $x_0 \in \mathbb{R}^d$, possibly with the insertion of a suitable weight factor. Simple examples show that naive versions, involving weights which are powers of the quantity $\star \bigwedge_{j=1}^m \star X_j(\mathrm{d}B_j(x))$ cannot hold; see [7] for an explicit example.

Organisation of the paper. To recap, in the next section we give some justification for our choice of proof of Theorem 1.3 and the class $\mathcal{C}$. In Section 3 we give an outline of the proof of Theorem 1.3 by considering the special case of the nonlinear Loomis–Whitney inequality in three dimensions. The full proof begins in Section 4 where we make the reduction to the coordinate projection case. The proof for this case rests on the induction-on-scales argument which appears in Section 5. In Section 6 we give a proof of Corollary 1.4, and in Section 7 we provide applications to two closely related problems in harmonic analysis.

Acknowledgements. The authors would like to express gratitude to the anonymous referee for their careful reading of the manuscript and extremely helpful recommendations, and also to Steve Roper at the University of Glasgow for creating the figures in Section 3.

2. Induction-on-scales and the class $\mathcal{C}$

The Brascamp–Lieb inequalities (1) possess a certain self-similar structure that strongly suggests an approach to the corresponding nonlinear statements by induction-on-scales. Induction-on-scales arguments have been used with great success in harmonic analysis in recent years. Very closely related to the forthcoming discussion is the induction-on-scales approach to the Fourier restriction and Kakeya conjectures originating in work of Bourgain [8], and developed further by Wolff [24] and Tao [20]; see also the survey article [21]. This self-similarity manifests itself most elegantly in an elementary convolution inequality due to Ball [1] (see also [5]), which we now describe.

Let $(\mathbf{B}, \mathbf{p})$ be a Brascamp–Lieb datum where each B_j is linear. Let $\mathbf{f}$ and $\mathbf{f}'$ be two inputs and we assume, for clarity of exposition, that these inputs are L^1-normalised. For each $x \in \mathbb{R}^d$ and $1 \leq j \leq m$ let $g_j^x : \mathbb{R}^{d_j} \to [0, \infty)$ be given by

$$g_j^x(y) = f_j(B_j x - y) f_j'(y).$$

By Fubini's theorem and elementary considerations we have that

$$\begin{aligned}
\mathrm{BL}(\mathbf{B},\mathbf{p};\mathbf{f})\mathrm{BL}(\mathbf{B},\mathbf{p};\mathbf{f}') &= \int_{\mathbb{R}^d} \prod_{j=1}^m (f_j \circ B_j)^{p_j} * \prod_{j=1}^m (f_j' \circ B_j)^{p_j} \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \prod_{j=1}^m (g_j^x \circ B_j)^{p_j} \right) \mathrm{d}x \\
&\leq \int_{\mathbb{R}^d} \left(\mathrm{BL}(\mathbf{B},\mathbf{p};(g_j^x)) \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} g_j^x(y)\,\mathrm{d}y \right)^{p_j} \right) \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \left(\mathrm{BL}(\mathbf{B},\mathbf{p};(g_j^x)) \prod_{j=1}^m (f_j * f_j'(B_j x))^{p_j} \right) \mathrm{d}x
\end{aligned}$$

and therefore

$$\mathrm{BL}(\mathbf{B},\mathbf{p};\mathbf{f})\mathrm{BL}(\mathbf{B},\mathbf{p};\mathbf{f}') \leq \sup_{x\in\mathbb{R}^d} \mathrm{BL}(\mathbf{B},\mathbf{p};(g_j^x))\ \mathrm{BL}(\mathbf{B},\mathbf{p};\mathbf{f}*\mathbf{f}'), \tag{13}$$

where $\mathbf{f}*\mathbf{f}' := (f_j * f_j')$. Notice that if $\mathbf{f}'$ is an extremiser to (1), i.e.

$$\mathrm{BL}(\mathbf{B},\mathbf{p};\mathbf{f}') = \mathrm{BL}(\mathbf{B},\mathbf{p}),$$

then since

$$\mathrm{BL}(\mathbf{B},\mathbf{p};\mathbf{f}*\mathbf{f}') \leq \mathrm{BL}(\mathbf{B},\mathbf{p}),$$

we may deduce that

$$\mathrm{BL}(\mathbf{B},\mathbf{p};\mathbf{f}) \leq \sup_{x\in\mathbb{R}^d} \mathrm{BL}(\mathbf{B},\mathbf{p};(g_j^x)). \tag{14}$$

In particular, in the presence of an appropriately "localising" extremiser $\mathbf{f}'$ (such as of compact support), (14) suggests the viability of a proof of nonlinear inequalities such as (2) by induction on the "scale of the support" of $\mathbf{f}$. The point is that g_j^x may be thought of as the function f_j localised by f_j' to a neighbourhood of the general point $B_j x$.

With the above discussion in mind it is natural to restrict attention to data $(\mathbf{B},\mathbf{p})$ for which (1) has extremisers of the form $\mathbf{f} = (\chi_{E_j})$, where for each j, E_j is a subset of $\mathbb{R}^{d_j}$ which tiles by translation. Furthermore, given our aspirations, it is natural to choose a class of data which is affine-invariant and stable under linear perturbations of $\mathbf{B}$. These requirements lead us to the transversality hypothesis in (4). Indeed, as there are linear changes of variables which show that Proposition 1.2 follows from Proposition 1.1 (see Appendix A), it is straightforward to observe that characteristic functions of certain paralellepipeds are extremisers for (9). Such sets of course tile by translation.

We remark that there are other hypotheses on the datum $\mathbf{B}$ which fulfill our requirements. For example, one may replace (4) by

$$\bigoplus_{j=1}^m \operatorname{coker} B_j = \mathbb{R}^d.$$

However, after appropriate changes of variables, the corresponding nonlinear inequality (2) merely reduces to a statement of Fubini's theorem, and in particular,

$p_j = 1$ for all j. There are further alternatives which are hybrids of these and are similarly degenerate.

Remark 2.1. Notice that if $\mathbf{f}'$ is an extremiser to (1) then we may also deduce from (13) that

$$\mathrm{BL}(\mathbf{B}, \mathbf{p}; \mathbf{f}) \leq \mathrm{BL}(\mathbf{B}, \mathbf{p}; \mathbf{f} * \mathbf{f}'). \tag{15}$$

This inequality suggests the viability of a proof of nonlinear inequalities such as (2) by induction on the "scale of constancy" of f. Certain weak versions of inequality (2), where the resulting constant C has a mild dependence on the smoothness of the input $\mathbf{f}$, have already been treated in this way in [6] (see Remarks 6.3 and 6.6).

In certain situations, (15) leads to the monotonicity of $\mathrm{BL}(\mathbf{B}, \mathbf{p}; \mathbf{f})$ under the action of convolution semigroups on the input $\mathbf{f}$. In the context of heat-flow, this observation originates in [10] and [5]; see the latter for further discussion of this perspective.

3. An outline of the proof of Theorem 1.3

The purpose of this section is to bring out the key ideas in the proof of Theorem 1.3. It is also an opportunity to introduce some notation which will be adopted (modulo small modifications) in the full proof in Section 5. As it is an outline we will sometimes compromise rigour for the sake of clarity. Our approach is based on [4].

Since the induction-on-scales argument we use to prove Theorem 1.3 is guided by the underlying geometry, in this outline we will consider the Loomis–Whitney case where $d = 3$, $m = 3$ and

$$\mathrm{d}B_j(x_0) = \Pi_j \tag{16}$$

for $j = 1, 2, 3$. In particular, we have $\ker \mathrm{d}B_j(x_0) = \langle e_j \rangle$ where e_j denotes the jth standard basis vector in $\mathbb{R}^3$.

We shall use $Q(x, \delta)$ to denote the axis-parallel cube centred at x with sidelength equal to δ.

Fix a small sidelength $\delta_0 > 0$ which, in terms of the induction-on-scales argument, represents the largest or "global" scale.

For $\delta, M > 0$ we let $C(\delta, M)$ denote the best constant in the inequality

$$\int_Q f_1(B_1(x))^{\frac{1}{2}} f_2(B_2(x))^{\frac{1}{2}} f_3(B_3(x))^{\frac{1}{2}} \, \mathrm{d}x \leq C \Big(\int_{\mathbb{R}^2} f_1 \Big)^{\frac{1}{2}} \Big(\int_{\mathbb{R}^2} f_2 \Big)^{\frac{1}{2}} \Big(\int_{\mathbb{R}^2} f_3 \Big)^{\frac{1}{2}}$$

over all axis-parallel subcubes Q of $Q(x_0, \delta_0)$ of sidelength δ and all inputs $f_1, f_2, f_3 \in L^1(\mathbb{R}^2)$ which are "constant" at the scale M^{-1}. The goal is to prove that $C(\delta_0, M)$ is bounded above by a constant independent of M, allowing the use of a density argument to pass to general $f_1, f_2, f_3 \in L^1(\mathbb{R}^2)$.

As our proof proceeds by induction it consists of two distinct parts.

(i) *The base case*: For each $M > 0$, $C(\delta, M)$ is bounded by an absolute constant for all δ sufficiently small.

(ii) *The inductive step*: There exists $\gamma > 0$ and $\alpha > 1$ such that

$$C(\delta, M) \leq (1 + O(\delta^\gamma))C(2\delta^\alpha, M) \tag{17}$$

uniformly in $\delta \leq \delta_0$ and $M > 0$.

Claims (i) and (ii) quickly lead to the desired conclusion since on iterating (17) we find that $C(\delta_0, M)$ is bounded by a convergent product of factors of the form $(1 + O(\delta^\gamma))$ with $\delta \leq \delta_0$.

To see why the base case is true, let Q be any axis-parallel cube contained in $Q(x_0, \delta_0)$ with centre x_Q and sidelength δ, and let $f_1, f_2, f_3 \in L^1(\mathbb{R}^2)$ be constant at scale M^{-1}. Observe that if δ is sufficiently small then each f_j does not "see" the difference between $B_j(x)$ and $\mathrm{d}B_j(x_Q)x$ for $x \in Q$ in the sense that $f_j \circ B_j \sim f_j \circ \mathrm{d}B_j(x_Q)$ (up to harmless translations) on Q. Now, by (16) and the smoothness of the B_j we know that

$$|X_j(\mathrm{d}B_j(x_Q)) - e_j| = |X_j(\mathrm{d}B_j(x_Q)) - X_j(\Pi_j)| \leq 1/10$$

if δ_0 is sufficiently small. Hence by Proposition 1.2 it follows that $C(\delta, M)$ is bounded above by an absolute constant for such δ.

Turning to the inductive step, fix any axis-parallel cube Q contained in $Q(x_0, \delta_0)$ with centre x_Q and sidelength δ, and let $f_1, f_2, f_3 \in L^1(\mathbb{R}^2)$ be constant at scale M^{-1}. First we decompose $Q = \bigcup P(n)$, where the $P(n)$ are axis-parallel subcubes with equal sidelength δ^α, and $\alpha > 1$. We choose the natural indexing of the $P(n)$ by $n \in \mathbb{N}^3$. Unfortunately this decomposition is too naive to prove the inductive step but nevertheless it is instructive to see where the proof breaks down.

Observe that

$$\begin{aligned}
&\int_Q f_1(B_1(x))^{\frac{1}{2}} f_2(B_2(x))^{\frac{1}{2}} f_3(B_3(x))^{\frac{1}{2}} \,\mathrm{d}x \\
&= \sum_{n \in \mathbb{N}^3} \int_{P(n)} f_1(B_1(x))^{\frac{1}{2}} f_2(B_2(x))^{\frac{1}{2}} f_3(B_3(x))^{\frac{1}{2}} \,\mathrm{d}x \\
&\leq C(\delta^\alpha, M) \sum_{n \in \mathbb{N}^3} \left(\int_{B_1(P(n))} f_1 \right)^{\frac{1}{2}} \left(\int_{B_2(P(n))} f_2 \right)^{\frac{1}{2}} \left(\int_{B_3(P(n))} f_3 \right)^{\frac{1}{2}}. \qquad (18)
\end{aligned}$$

If $n = (n_1, n_2, n_3)$ then $\int_{B_1(P(n))} f_1$ is "almost" a function of n_2 and n_3. Indeed, if B_1 is linear and equal to Π_1 then

$$B_1(P(n)) = B_1(T_1(n_2, n_3))$$

where $T_1(n_2, n_3)$ is a cuboid (or "tube") with long side in the direction of e_1 and containing $P(n)$. A similar remark holds for $\int_{B_2(P(n))} f_2$ and $\int_{B_3(P(n))} f_3$.

For $j = 1, 2, 3$ this leads us to define cuboids

$$T_j(\ell) = \bigcup_{\substack{n \in \mathbb{N}^3: \\ \Pi_j n = \ell}} P(n)$$

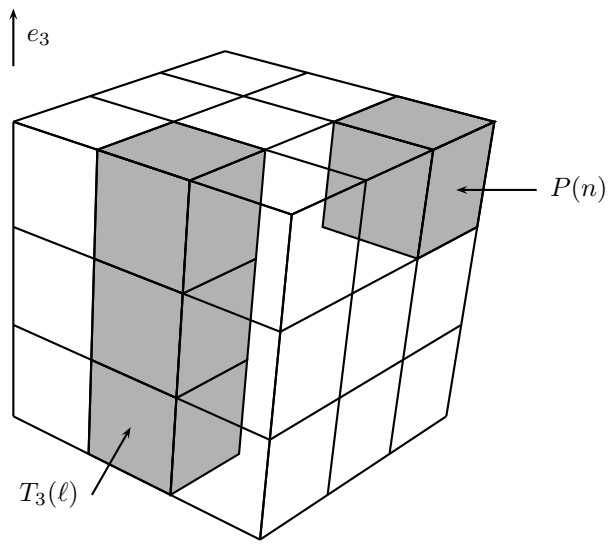

FIGURE 1. Subcubes $P(n)$ parametrised by $n \in \mathbb{N}^3$ and tubes $T_3(\ell)$ parametrised by $\ell \in \mathbb{N}^2$ with direction e_3.

for $\ell \in \mathbb{N}^2$. Note that $T_j(\ell)$ has direction e_j and its location is determined by $\ell \in \mathbb{N}^2$. In particular, for each $n \in \mathbb{N}^3$, $T_j(\Pi_j n)$ is a cuboid in the direction e_j which passes through $P(n)$. See Figure 1. Accordingly, we define

$$F_j(\ell) = \int_{B_j(T_j(\ell))} f_j$$

for $j = 1, 2, 3$ and $\ell \in \mathbb{N}^2$. Then by (18) and the discrete inequality (7),

$$\begin{aligned}\int_Q f_1(B_1(x))^{\frac{1}{2}} f_2(B_2(x))^{\frac{1}{2}} f_3(B_3(x))^{\frac{1}{2}}\,\mathrm{d}x \leq & C(\delta^\alpha, M) \sum_{n\in\mathbb{N}^3} F_1(\Pi_1 n)^{\frac{1}{2}} F_2(\Pi_2 n)^{\frac{1}{2}} F_3(\Pi_3 n)^{\frac{1}{2}} \\ \leq & C(\delta^\alpha, M) \|F_1\|_{\ell^1(\mathbb{N}^2)}^{\frac{1}{2}} \|F_2\|_{\ell^1(\mathbb{N}^2)}^{\frac{1}{2}} \|F_3\|_{\ell^1(\mathbb{N}^2)}^{\frac{1}{2}}.\end{aligned}$$

If we had disjointness in the sense that

$$B_j(T_j(\ell)) \cap B_j(T_j(\ell')) = \emptyset \qquad \text{whenever} \qquad \ell \neq \ell', \tag{19}$$

then

$$\|F_j\|_{\ell^1(\mathbb{N}^2)} \leq \int_{\mathbb{R}^2} f_j$$

would hold for each $j = 1, 2, 3$, and hence

$$\int_Q f_1(B_1(x))^{\frac{1}{2}} f_2(B_2(x))^{\frac{1}{2}} f_3(B_3(x))^{\frac{1}{2}}\,\mathrm{d}x \leq C(\delta^\alpha, M) \left(\int_{\mathbb{R}^2} f_1\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^2} f_2\right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^2} f_3\right)^{\frac{1}{2}} \tag{20}$$

would follow immediately. If each B_j is linear and equal to Π_j then (19) is of course true, although otherwise it is not. In order to achieve a version of (19) in general, it is necessary to modify our decomposition of Q.

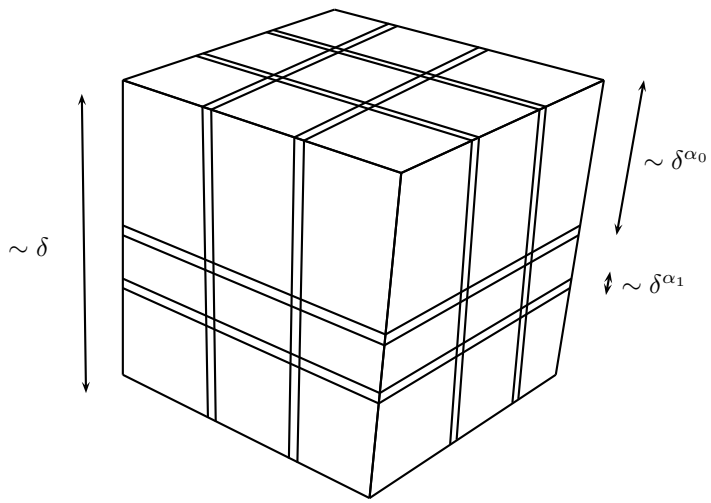

FIGURE 2. The modified decomposition of Q.

To better understand the location of each image $B_j(T_j(\Pi_j n))$ the $P(n)$ should in fact be parallelepipeds whose faces are given by pull-backs of certain lines in $\mathbb{R}^2$ under the linear maps $\mathrm{d}B_j(x_Q)$.

However, we still need to fully accommodate for the nonlinearity and in particular the difference between $B_j(T_j(\ell))$ and $\mathrm{d}B_j(x_Q)(T_j(\ell))$. Following the approach in [4] it is natural to insert relatively narrow "buffer zones" between the $P(n)$ to provide sufficient separation in order to guarantee the sought after disjointness property (19). Clearly this depends on the smoothness of the B_j and, since we assume $C^{1,\beta}$ regularity, we take the $P(n)$ to have sidelengths approximately δ^{α_0} and the buffer zones to have width approximately δ^{α_1} where

$$1 < \alpha_0 < \alpha_1 < 1 + \beta.$$

The decomposition of Q now has a "main component" from the $P(n)$ and a "error component" from the buffer zones. We would like to use the above argument which led to (20) on each component. However, in order for the error component to genuinely contribute an acceptable error term, we need to relax the regular decomposition (into equally sized $P(n)$) since a "large" amount of mass of the $f_j \circ B_j$ may lie on the buffer zones. Again following ideas from [4] we use a simple pigeonholing argument to position the buffer zones in an efficient location given the constraint that the $P(n)$ should have *essentially* the same sidelengths. See Figure 2. Putting the resulting estimates together yields the desired recursive inequality (17) with $\alpha = \alpha_0$ and some $\gamma > 0$.

See Section 5 for the complete details of this induction-on-scales argument in the full generality of Theorem 1.3.

4. Preparation and reduction to the orthogonal projection case

Recall the definition of $\Pi_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$ given by (5). In this section we shall prove that Theorem 1.3 is a consequence of the following nonlinear version of Proposition 1.1.

Proposition 4.1. *Suppose $\beta, \kappa > 0$ are given and α_0, α_1 satisfy $1 < \alpha_0 < \alpha_1 < 1 + \beta$. Let*

$$\delta_0 = \min\left\{ \left(\frac{c_d}{\kappa}\right)^{\frac{1}{1+\beta-\alpha_1}}, \left(\frac{1}{4}\right)^{\frac{1}{\min\{\alpha_0-1, \alpha_1-\alpha_0\}}} \right\}. \tag{21}$$

Suppose that $B_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$ is a $C^{1,\beta}$ submersion satisfying $\|B_j\|_{C^{1,\beta}} \leq \kappa$ in $Q(x_0, \delta_0)$ and $\mathrm{d}B_j(x_0) = \Pi_j$ for each $1 \leq j \leq m$. Then for $c_d \in (0, \kappa)$ sufficiently small,

$$\int_{Q(x_0,\delta_0)} \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \, \mathrm{d}x \leq 10^d \exp\left(\frac{10^d \delta_0^{\frac{\alpha_1-\alpha_0}{m-1}}}{1 - 2^{-\frac{\alpha_1-\alpha_0}{m-1}}} \right) \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} f_j \right)^{\frac{1}{m-1}}$$

for all nonnegative $f_j \in L^1(\mathbb{R}^{d_j})$, $1 \leq j \leq m$.

As mentioned already in the previous section, the proof of Proposition 4.1 will proceed by an induction-on-scales argument. For a cube at scale δ, we decompose into parallelepipeds of sidelength approximately δ^{α_0} and the buffer zones will have thickness approximately δ^{α_1}. We have stated Proposition 4.1 with this in mind and we have provided explicit information on how the size of the neighbourhood and the constant depend on the relevant parameters.

Deduction of Theorem 1.3 from Proposition 4.1. The argument which follows is similar to the argument given in Appendix A for the corresponding claim in the linear case. A little extra work is required to verify the uniformity claims in Theorem 1.3 concerning the neighbourhood and the constant.

Select any set of vectors $\{a_k : k \in \mathcal{K}_j\}$ forming an orthonormal basis for $\ker \mathrm{d}B_j(x_0)$. By definition of the Hodge star and orthogonality we get

$$\star X_j(\mathrm{d}B_j(x_0)) = \|X_j(\mathrm{d}B_j(x_0))\|_{\Lambda^{d_j}(\mathbb{R}^d)} \bigwedge_{k \in \mathcal{K}_j} a_k. \tag{22}$$

Let A be the $d \times d$ matrix whose ith column is equal to a_i for each $1 \leq i \leq d$. Finally, let C_j be the $d_j \times d_j$ matrix given by

$$C_j = \mathrm{d}B_j(x_0) A_j,$$

where A_j is the $d \times d_j$ matrix obtained by deleting from A the columns a_k for each $k \in \mathcal{K}_j$.

Then, by construction, the map $\widetilde{B}_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$ given by

$$\widetilde{B}_j(x) = C_j^{-1} B_j(Ax)$$

satisfies

$$\mathrm{d}\widetilde{B}_j(\widetilde{x}_0) = C_j^{-1} \mathrm{d}B_j(x_0) A = \Pi_j, \tag{23}$$

where $\widetilde{x}_0 = A^{-1}x_0$. Since we are assuming (4) and since B_j is a submersion at x_0 we know that the matrices A and C_j are invertible.

Let U be some neighbourhood of x_0 and ψ a cutoff function supported in U. Using A to change variables one obtains

$$\int_{\mathbb{R}^d} \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \psi(x)\,\mathrm{d}x = |\det(A)| \int_{\mathbb{R}^d} \prod_{j=1}^m \widetilde{f}_j(\widetilde{B}_j(x))^{\frac{1}{m-1}} \widetilde{\psi}(x)\,\mathrm{d}x, \tag{24}$$

where $\widetilde{\psi} = \psi \circ A$ is a cutoff function supported in $A^{-1}U$ and $\widetilde{f}_j = f_j \circ C_j$, $1 \leq j \leq m$. Of course, we know that $\mathrm{d}\widetilde{B}_j(\widetilde{x}_0) = \Pi_j$ by (23). Notice also that

$$\|\mathrm{d}\widetilde{B}_j(x) - \mathrm{d}\widetilde{B}_j(y)\| = \|C_j^{-1}(\mathrm{d}B_j(Ax) - \mathrm{d}B_j(Ay))A\| \leq C\kappa \|C_j^{-1}\| |x-y|^\beta,$$

where the constant C depends on at most d. To show that we may choose the neighbourhood U and the constant in the claimed uniform manner we need to show that suitable upper bounds hold for the norms of A^{-1} and each C_j^{-1}.

For A^{-1}, we note that

$$\star \bigwedge_{j=1}^m \star X_j(\mathrm{d}B_j(x_0)) = \prod_{j=1}^m \|X_j(\mathrm{d}B_j(x_0))\|_{\Lambda^{d_j}(\mathbb{R}^d)} \star \bigwedge_{j=1}^m \bigwedge_{k \in \mathcal{K}_j} a_k$$

by (22) and therefore

$$\star \bigwedge_{j=1}^m \star X_j(\mathrm{d}B_j(x_0)) = \det(A) \prod_{j=1}^m \|X_j(\mathrm{d}B_j(x_0))\|_{\Lambda^{d_j}(\mathbb{R}^d)}. \tag{25}$$

Since $\|B_j\|_{C^{1,\beta}} \leq \kappa$ it follows that

$$\left| \star \bigwedge_{j=1}^m \star X_j(\mathrm{d}B_j(x_0)) \right| \leq C|\det(A)|$$

for some constant C depending on κ and d. Since each column of A is a unit vector, it follows that the norm of A^{-1} is bounded above by a constant depending on ε, κ and d.

For C_j^{-1}, from (22) we get

$$|\det(C_j)| = \|X_j(\mathrm{d}B_j(x_0))\|_{\Lambda^{d_j}(\mathbb{R}^d)} |\det(A)|. \tag{26}$$

By (25),

$$\varepsilon \leq C\|X_j(\mathrm{d}B_j(x_0))\|_{\Lambda^{d_j}(\mathbb{R}^{d_j})} |\det(A)|,$$

for some constant C depending on κ and d. It follows that the norm of C_j^{-1} is also bounded above by a constant depending on ε, κ and d.

Applying Proposition 4.1 it follows that there exists a neighbourhood U of x_0 depending on at most $\beta, \varepsilon, \kappa$ and d such that

$$\int_{\mathbb{R}^d} \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \psi(x)\,\mathrm{d}x \leq C|\det(A)| \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} \widetilde{f}_j \right)^{\frac{1}{m-1}},$$

where C depends on at most d and ψ. Thus

$$\int_{\mathbb{R}^d} \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \psi(x)\,\mathrm{d}x \leq C \frac{|\det(A)|}{\left(\prod_{j=1}^m |\det(C_j)|\right)^{\frac{1}{m-1}}} \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} f_j\right)^{\frac{1}{m-1}}$$

$$= C \left| \star \bigwedge_{j=1}^m \star X_j(\mathrm{d}B_j(x_0)) \right|^{-\frac{1}{m-1}} \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} f_j\right)^{\frac{1}{m-1}},$$

where the equality holds because of (25) and (26). Theorem 1.3 now follows.

For the various constants appearing in the above proof, one may easily obtain some explicit dependence in terms of the relevant parameters. Combined with Proposition 4.1, this gives additional information on the sizes of the neighbourhood U and constant C appearing in the statement of Theorem 1.3. We do not pursue this matter further here.

5. Proof of Proposition 4.1: Induction-on-scales

Before stating the main induction lemma we use to prove Proposition 4.1, we need to fix some further notation. For each $1 \leq j \leq m$ and $M > 0$, let $L^1_M(\mathbb{R}^{d_j})$ denote those nonnegative $f \in L^1(\mathbb{R}^{d_j})$ satisfying $f(y_1) \leq 2f(y_2)$ whenever y_1 and y_2 are in the support of f and $|y_1 - y_2| \leq M^{-1}$; that is, those f which are effectively constant at the scale M^{-1}. One may easily check that if μ is a finite measure on $\mathbb{R}^{d_j}$ then $P^{(d_j)}_{c/M} * \mu \in L^1_M(\mathbb{R}^{d_j})$, where $P^{(d_j)}_{c/M}$ denotes the Poisson kernel on $\mathbb{R}^{d_j}$ at height c/M. Here c is a suitably large constant depending only on d_j. By an elementary density argument, it will be enough to prove Proposition 4.1 for $f_j \in L^1_M(\mathbb{R}^{d_j})$, $1 \leq j \leq m$, with neighbourhood U and constant C independent of M. As we shall shortly see, we consider such a subclass of functions in order to provide a "base case" for the inductive argument.

For $\beta, \kappa > 0$, $1 < \alpha_0 < \alpha_1 < 1 + \beta$ and $x_0 \in \mathbb{R}^d$ we let $\mathcal{B}(\beta, \kappa, \alpha_0, \alpha_1, x_0)$ be the family of data $\mathbf{B}$ such that B_j belongs to $C^{1,\beta}(Q(x_0, \delta_0))$ with $\|B_j\|_{C^{1,\beta}} \leq \kappa$ and satisfies $\mathrm{d}B_j(x_0) = \Pi_j$, $1 \leq j \leq m$. Here, δ_0 is given by (21).

Now let $C(\delta, M)$ denote the best constant in the inequality

$$\int_Q \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}}\,\mathrm{d}x \leq C \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} f_j\right)^{\frac{1}{m-1}}$$

over all $\mathbf{B} \in \mathcal{B}(\beta, \kappa, \alpha_0, \alpha_1, x_0)$, all axis-parallel subcubes Q of $Q(x_0, \delta_0)$ with side-length equal to δ and all inputs $\mathbf{f}$ such that f_j belongs to $L^1_M(\mathbb{R}^{d_j})$, $1 \leq j \leq m$.

We note that the constant $C(\delta, M)$ also depends on the parameters β, κ, α_0 and α_1, although there is little to be gained in what follows from making this dependence explicit. The main induction-on-scales lemma is the following.

Lemma 5.1. *For all $0 < \delta \leq \delta_0$ we have*

$$C(\delta, M) \leq (1 + 10^d \delta^{\frac{\alpha_1 - \alpha_0}{m-1}}) C(2\delta^{\alpha_0}, M).$$

The proof of Lemma 5.1 is a little lengthy. Before giving the proof we show how Lemma 5.1 implies Proposition 4.1.

Deduction of Proposition 4.1 from Lemma 5.1. Firstly we claim that the "base case" inequality

$$C(\delta_0/2^N, M) \leq 10^d \tag{27}$$

holds for sufficiently large N. To see (27), suppose $\mathbf{B} \in \mathcal{B}(\beta, \kappa, \alpha_0, \alpha_1, x_0)$, Q is a subcube of $Q(x_0, \delta_0)$ with centre x_Q and sidelength $\delta_0/2^N$, and the input $\mathbf{f}$ is such that f_j belongs to $L^1_M(\mathbb{R}^{d_j})$, $1 \leq j \leq m$. For any $x \in Q$,

$$|B_j(x) - (B_j(x_Q) + \mathrm{d}B_j(x_Q)(x - x_Q))| \leq \kappa |x - x_Q|^{1+\beta} \leq 1/M$$

if N is sufficiently large (depending on β, κ, d and M). Since $f_j \in L^1_M(\mathbb{R}^{d_j})$ it follows that

$$\int_Q \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \, \mathrm{d}x \leq 2^m \int_{Q - \{x_Q\}} \prod_{j=1}^m f_j(\cdot + B_j(x_Q))(\mathrm{d}B_j(x_Q)x)^{\frac{1}{m-1}} \, \mathrm{d}x.$$

Now

$$\|\mathrm{d}B_j(x_Q) - \Pi_j\| = \|\mathrm{d}B_j(x_Q) - \mathrm{d}B_j(x_0)\| \leq \frac{1}{100^d},$$

which implies that

$$\star \bigwedge_{j=1}^m \star X_j(\mathrm{d}B_j(x_Q)) \geq \frac{1}{2},$$

and therefore

$$\int_Q \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \, \mathrm{d}x \leq 10^d \prod_{j=1}^m \Big(\int_{\mathbb{R}^{d_j}} f_j \Big)^{\frac{1}{m-1}}$$

by Proposition 1.2. Hence, (27) holds.

For $0 < \delta \leq \delta_0 \leq (1/4)^{1/\alpha_0 - 1}$ it follows from Lemma 5.1 that

$$C(\delta, M) \leq (1 + 10^d \delta^{\frac{\alpha_1 - \alpha_0}{m-1}}) C(\delta/2, M). \tag{28}$$

Applying (28) iteratively N times we see that

$$C(\delta_0, M) \leq C(\delta_0/2^N, M) \prod_{r=0}^{N-1} \big(1 + 10^d (\delta_0/2^r)^{\frac{\alpha_1 - \alpha_0}{m-1}}\big).$$

The product term is under control uniformly in N because

$$\begin{aligned} \log \prod_{r=0}^{N-1} \big(1 + 10^d (\delta_0/2^r)^{\frac{\alpha_1 - \alpha_0}{m-1}}\big) &= \sum_{r=0}^{N-1} \log \big(1 + 10^d (\delta_0/2^r)^{\frac{\alpha_1 - \alpha_0}{m-1}}\big) \\ &\leq 10^d \delta_0^{\frac{\alpha_1 - \alpha_0}{m-1}} \sum_{r=0}^{\infty} 2^{-\frac{\alpha_1 - \alpha_0}{m-1} r} \\ &\leq \frac{10^d \delta_0^{\frac{\alpha_1 - \alpha_0}{m-1}}}{1 - 2^{-\frac{\alpha_1 - \alpha_0}{m-1}}}. \end{aligned}$$

From the base case (27) it follows that

$$C(\delta_0, M) \leq 10^d \exp\left(\frac{10^d \delta_0^{\frac{\alpha_1-\alpha_0}{m-1}}}{1-2^{-\frac{\alpha_1-\alpha_0}{m-1}}}\right);$$

that is,

$$\int_{Q(x_0,\delta_0)} \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}}\,\mathrm{d}x \leq 10^d \exp\left(\frac{10^d \delta_0^{\frac{\alpha_1-\alpha_0}{m-1}}}{1-2^{-\frac{\alpha_1-\alpha_0}{m-1}}}\right) \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} f_j\right)^{\frac{1}{m-1}} \tag{29}$$

for all $f_j \in L^1_M(\mathbb{R}^{d_j})$, $1 \leq j \leq m$. Since the constant in (29) is independent of M, it follows that the inequality is valid for all $f_j \in L^1(\mathbb{R}^{d_j})$. This completes our proof of Proposition 4.1.

Proof of Lemma 5.1. Suppose $\mathbf{B} = (B_j) \in \mathcal{B}(\beta, \kappa, \alpha_0, \alpha_1, x_0)$, Q is an axis-parallel subcube of $Q(x_0, \delta_0)$ with sidelength equal to δ and centre x_Q, and suppose $\mathbf{f} = (f_j)$ is such that f_j belongs to $L^1_M(\mathbb{R}^{d_j})$, $1 \leq j \leq m$. Notice that the desired inequality

$$\int_Q \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}}\,\mathrm{d}x \leq (1+10^d\delta^{\frac{\alpha_1-\alpha_0}{m-1}})C(2\delta^{\alpha_0}, M) \prod_{j=1}^m \left(\int_{\mathbb{R}^{d_j}} f_j\right)^{\frac{1}{m-1}} \tag{30}$$

is invariant under the transformation $(\mathbf{B}, \mathbf{f}, Q) \mapsto (\widetilde{\mathbf{B}}, \widetilde{\mathbf{f}}, \widetilde{Q})$ where $\widetilde{B}_j = B_j(\cdot + x_Q) - B_j(x_Q)$, $\widetilde{Q} = Q - \{x_Q\}$ and $\widetilde{f}_j = f_j(\cdot + B_j(x_Q))$. Hence, without loss of generality, $Q = Q(0, \delta)$ and $B_j(0) = 0$ for $1 \leq j \leq m$. This reduction is merely for notational convenience; in particular, it ensures

$$|B_j(x) - \mathrm{d}B_j(0)x| \leq \kappa |x|^{1+\beta}.$$

By the smoothness hypothesis, we have that

$$\|\mathrm{d}B_j(0) - \Pi_j\| \leq \frac{1}{100^d} \tag{31}$$

for sufficiently small c_d. Since

$$\ker \Pi_j = \langle \{e_k : k \in \mathcal{K}_j\}\rangle,$$

it follows that for each $1 \leq k \leq d$ there exist $a_k \in \mathbb{R}^d$ such that

$$|a_k - e_k| \leq \frac{1}{10^d}, \tag{32}$$

and

$$\ker \mathrm{d}B_j(0) = \langle \{a_k : k \in \mathcal{K}_j\}\rangle$$

for each $1 \leq j \leq m$. Here, e_k denotes the kth standard basis vector in $\mathbb{R}^d$.

The proof of Lemma 5.1 naturally divides into four steps.

Step I: Foliations of $\mathbb{R}^d$

For each $1 \leq i \leq d$ consider the one-parameter family of hypersurfaces

$$\langle \{a_k : k \neq i\}\rangle + \left\{ s \star \bigwedge_{k \neq i} a_k \right\} \tag{33}$$

where $s \in \mathbb{R}$. We point out that $\star \bigwedge_{k \neq i} a_k$ is simply the cross product of the vectors $\{a_k : k \neq i\}$, yielding a vector normal to $\langle \{a_k : k \neq i\} \rangle$. The set of vectors $\{\star \bigwedge_{k \neq i} a_k : 1 \leq i \leq d\}$ in $\mathbb{R}^d$ is linearly independent since the same is true of $\{a_i : 1 \leq i \leq d\}$. Consequently, we may decompose $\mathbb{R}^d$ into parallelepipeds whose faces are contained in hyperplanes of the form (33), $1 \leq i \leq d$. We will use this to decompose the cube Q. As we shall see in the steps that follow, an important feature of these hypersurfaces is that they may be expressed as inverse images of hypersurfaces under the mappings $\mathrm{d}B_j(0)$. To this end, let $\sigma : \{1, \ldots, d\} \to \{1, \ldots, m\}$ be the map given by

$$\sigma(i) = (j+1) \bmod m$$

for $i \in \mathcal{K}_j$. As will become apparent under closer inspection, there is some freedom in our choice of this map; all that we require of σ is that $j \mapsto \sigma(\mathcal{K}_j)$ is a permutation of $\{1, 2, \ldots, m\}$ with no fixed points.

For each $1 \leq i \leq d$ and $J \subset \mathbb{R}$ we define the set

$$\Sigma(i, J) = \mathrm{d}B_{\sigma(i)}(0)\langle \{a_k : k \neq i\} \rangle + \Big\{ s \, \mathrm{d}B_{\sigma(i)}(0)\Big(\star \bigwedge_{k \neq i} a_k \Big) : s \in J \Big\}. \tag{34}$$

If $J = \{s\}$ is a singleton set then

$$\Sigma(i, \{s\}) = \mathrm{d}B_{\sigma(i)}(0)\langle \{a_k : k \neq i\} \rangle + \Big\{ s \, \mathrm{d}B_{\sigma(i)}(0)\Big(\star \bigwedge_{k \neq i} a_k \Big) \Big\}$$

is a hyperplane in $\mathbb{R}^{d_{\sigma(i)}}$ since $\ker \mathrm{d}B_{\sigma(i)}(0) \subseteq \langle \{a_k : k \neq i\} \rangle$. Similarly,

$$\mathrm{d}B_{\sigma(i)}(0)^{-1}\Sigma(i, \{s\}) = \langle \{a_k : k \neq i\} \rangle + \Big\{ s \star \bigwedge_{k \neq i} a_k \Big\} \tag{35}$$

which is of course the hyperplane (33).

As outlined in Section 3, a regular decomposition of $\mathbb{R}^d$ into parallelepipeds of equal size and adapted to a lattice (where for each i, the sequence of parameters $s^{(i)}$ that we choose is in arithmetic progression) will not suffice to prove Lemma 5.1. Moreover, our decomposition will need to incorporate certain "buffer zones" between the parallelepipeds to create separation. In Step II below we determine the location of the buffer zones and thus the desired decomposition of Q.

Step II: The decomposition of Q

For each $1 \leq i \leq d$ we claim that there exists a sequence $(s_n^{(i)})_{n \geq 1}$ such that

$$s_n^{(i)} + \tfrac{1}{2}\delta^{\alpha_0} \leq s_{n+1}^{(i)} \leq s_n^{(i)} + \delta^{\alpha_0} \tag{36}$$

and

$$\int_{\Sigma(i, [s_{n+1}^{(i)}, s_{n+1}^{(i)} + \delta^{\alpha_1}])} f_{\sigma(i)} \chi_Q \leq 4\delta^{\alpha_1 - \alpha_0} \int_{\Sigma(i, [s_n^{(i)} + \frac{1}{2}\delta^{\alpha_0}, s_n^{(i)} + \delta^{\alpha_0}])} f_{\sigma(i)} \chi_Q. \tag{37}$$

To prove this, we shall choose the sequence $(s_n^{(i)})_{n \geq 1}$ iteratively. We begin by choosing $s_1^{(i)}$ to be any real number such that $B_{\sigma(i)}(Q) \subseteq \Sigma(i, [s_1^{(i)}, \infty))$. Suppose that we have chosen $s_1^{(i)}, \ldots, s_n^{(i)}$ for some $n \geq 1$. Now let N be the largest integer

which is less than or equal to $\frac{1}{2}\delta^{\alpha_0-\alpha_1}$. Set $\zeta_0^{(i)} = s_n^{(i)} + \frac{1}{2}\delta^{\alpha_0}$ and then define $\zeta_r^{(i)} = \zeta_{r-1}^{(i)} + \delta^{\alpha_1}$ iteratively for $1 \leq r \leq N$ so that

$$[s_n^{(i)} + \tfrac{1}{2}\delta^{\alpha_0}, s_n^{(i)} + \delta^{\alpha_0}] \supseteq [s_n^{(i)} + \tfrac{1}{2}\delta^{\alpha_0}, s_n^{(i)} + \tfrac{1}{2}\delta^{\alpha_0} + N\delta^{\alpha_1}] = \bigcup_{r=1}^{N} [\zeta_{r-1}^{(i)}, \zeta_r^{(i)}].$$

Then,

$$\int_{\Sigma(i,[s_n^{(i)}+\frac{1}{2}\delta^{\alpha_0}, s_n^{(i)}+\delta^{\alpha_0}])} f_{\sigma(i)}\chi_Q \geq \sum_{r=1}^{N} \int_{\Sigma(i,[\zeta_{r-1}^{(i)},\zeta_r^{(i)}])} f_{\sigma(i)}\chi_Q,$$

and therefore by the choice of δ_0 in (21) and the pigeonhole principle, there exists $s_{n+1}^{(i)}$ such that (36) holds and

$$\int_{\Sigma(i,[s_n^{(i)}+\frac{1}{2}\delta^{\alpha_0}, s_n^{(i)}+\delta^{\alpha_0}])} f_{\sigma(i)}\chi_Q \geq \tfrac{1}{4}\delta^{\alpha_0-\alpha_1} \int_{\Sigma(i,[s_{n+1}^{(i)}, s_{n+1}^{(i)}+\delta^{\alpha_1}])} f_{\sigma(i)}\chi_Q;$$

that is, (37) also holds.

We shall use the notation $J(i,n,0)$ and $J(i,n,1)$ for the intervals given by

$$J(i,n,0) = (s_n^{(i)} + \tfrac{2}{3}\delta^{\alpha_1}, s_{n+1}^{(i)} + \tfrac{1}{3}\delta^{\alpha_1}] \tag{38}$$

and

$$J(i,n,1) = (s_n^{(i)} + \tfrac{1}{3}\delta^{\alpha_1}, s_n^{(i)} + \tfrac{2}{3}\delta^{\alpha_1}]. \tag{39}$$

Notice that the lengths of $J(i,n,0)$ and $J(i,n,1)$ are comparable to δ^{α_0} and δ^{α_1} respectively.

By construction, the sets $\Sigma(i, J(i,n,1))$ contain a relatively small amount of the mass of the function $f_{\sigma(i)}$ in the sense of (37). Furthermore, the inverse images of these sets,

$$\mathrm{d}B_{\sigma(i)}(0)^{-1}\Sigma(i, J(i,n,1)), \tag{40}$$

are $O(\delta^{\alpha_1})$ neighbourhoods of hyperplanes in $\mathbb{R}^d$, which as n varies are separated by $O(\delta^{\alpha_0})$. We refer to the sets (40) as buffer zones.

The decomposition of Q we use is given by

$$Q = \bigcup_{\chi\in\{0,1\}^d} \bigcup_{n\in\mathbb{N}^d} P(n,\chi) \tag{41}$$

where

$$P(n,\chi) = \bigcap_{i=1}^{d} \mathrm{d}B_{\sigma(i)}(0)^{-1}\Sigma(i, J(i,n_i,\chi_i)) \cap Q. \tag{42}$$

When $\chi = 0$, the $P(n,\chi)$ are large parallelepipeds (intersected with Q) with sidelength approximately δ^{α_0} which form the main part of our decomposition. For $\chi \neq 0$, the $P(n,\chi)$ are small parallelepipeds (intersected with Q) with at least one sidelength approximately δ^{α_1}, which decompose the buffer zones.

Step III: Disjointness

In this step we make precise the role of the buffer zones. For each $1 \leq j \leq m$, $\ell \in \mathbb{N}^{d_j}$ and $\chi \in \{0,1\}^d$ let

$$T_j(\ell,\chi) = \bigcup_{\substack{n\in\mathbb{N}^d:\\ \Pi_j n=\ell}} P(n,\chi).$$

It is the disjointness of the images of such sets under the mapping B_j that is crucial to the induction-on-scales argument which follows in Step IV.

Proposition 5.2. *Fix j with $1 \leq j \leq m$ and $\chi \in \{0,1\}^d$. If $\ell, \ell' \in \mathbb{N}^{d_j}$ are distinct then*

$$B_j(T_j(\ell,\chi)) \cap B_j(T_j(\ell',\chi)) = \emptyset. \tag{43}$$

To prove Proposition 5.2 we use the following.

Lemma 5.3. *For each $1 \leq j \leq m$ there exists a map $\Phi_j : \mathbb{R}^d \to \mathbb{R}^d$ such that*

(i) $\Phi_j(0) = 0$ *and* $\mathrm{d}\Phi_j(0)$ *is equal to the identity matrix* I_d,
(ii) $B_j = \mathrm{d}B_j(0) \circ \Phi_j$,
(iii) $\|\mathrm{d}\Phi_j(x) - \mathrm{d}\Phi_j(y)\| \leq 2\kappa|x-y|^\beta$ *for each* $x, y \in Q$,
(iv) $|x - \Phi_j(x)| \leq 2d\kappa\delta^{1+\beta}$ *for each* $x \in Q$.

Proof. Let $\widetilde{I}_{d_j}$ be the invertible $d_j \times d_j$ matrix obtained by deleting the kth column of $\mathrm{d}B_j(0)$ for each $k \in \mathcal{K}_j$. For $k \in \mathcal{K}_j$ define the kth component of $\Phi_j(x)$ to be x_k. Define the remaining d_j components of $\Phi_j(x)$ by stipulating that the element of $\mathbb{R}^{d_j}$ obtained by deleting the kth components of $\Phi_j(x)$ for $k \in \mathcal{K}_j$ is equal to

$$\widetilde{I}_{d_j}^{-1}\Big(B_j(x) - \sum_{k\in\mathcal{K}_j} x_k \mathrm{d}B_j(0)e_k\Big).$$

Then a direct computation verifies that Properties (i) and (ii) hold for Φ_j. Also,

$$\|\mathrm{d}\Phi_j(x) - \mathrm{d}\Phi_j(y)\| = \|\widetilde{I}_{d_j}^{-1}(\mathrm{d}B_j(x) - \mathrm{d}B_j(y))\| \leq 2\kappa|x-y|^\beta,$$

since $\|\widetilde{I}_{d_j} - I_{d_j}\| \leq 1/10$, and therefore (iii) holds. Finally, Property (iv) follows from Properties (i) and (iii), and the mean value theorem. □

Proof of Proposition 5.2. Suppose $\ell \neq \ell'$ and, for a contradiction, suppose that $z = B_j(x) = B_j(y)$ where $x \in T_j(\ell,\chi)$ and $y \in T_j(\ell',\chi)$. Then $x \in P(n,\chi)$ and $y \in P(n',\chi)$ for some $n, n' \in \mathbb{N}^d$ satisfying $\Pi_j n = \ell$ and $\Pi_j n' = \ell'$. Since $\Pi_j n \neq \Pi_j n'$ there exists $i \in \mathcal{K}_j^c$ such that $n_i \neq n_i'$.

By (42) and (34) it follows that there exist $s(x) \in J(i, n_i, \chi_i)$ and $s(y) \in J(i, n_i', \chi_i)$ such that

$$\Big\langle x, \star \bigwedge_{k\neq i} a_k \Big\rangle = s(x)\Big|\star \bigwedge_{k\neq i} a_k\Big|^2 \qquad \text{and} \qquad \Big\langle y, \star \bigwedge_{k\neq i} a_k \Big\rangle = s(y)\Big|\star \bigwedge_{k\neq i} a_k\Big|^2.$$

Therefore

$$\Big|\Big\langle x-y, \star \bigwedge_{k\neq i} a_k \Big\rangle\Big| = |s(x)-s(y)|\Big|\star \bigwedge_{k\neq i} a_k\Big|^2 \geq \tfrac{1}{3}\delta^{\alpha_1}\Big|\star \bigwedge_{k\neq i} a_k\Big|^2$$

where the inequality follows from (36), (38) and (39) since $n_i \neq n_i'$.

On the other hand, since x and y belong to the fibre $B_j^{-1}(z)$, it follows from Lemma 5.3(ii) that $\Phi_j(x)$ and $\Phi_j(y)$ belong to $\mathrm{d}B_j(0)^{-1}(z)$ and thus $\Phi_j(x) - \Phi_j(y) \in \ker \mathrm{d}B_j(0)$. Since $i \in \mathcal{K}_j^c$ and $\ker \mathrm{d}B_j(0) = \langle \{a_r : r \in \mathcal{K}_j\} \rangle$ the vector $\star \bigwedge_{k \neq i} a_k$ belongs to the orthogonal complement of $\ker \mathrm{d}B_j(0)$. Therefore,

$$\left\langle x - y, \star \bigwedge_{k \neq i} a_k \right\rangle = \left\langle x - \Phi_j(x), \star \bigwedge_{k \neq i} a_k \right\rangle - \left\langle y - \Phi_j(y), \star \bigwedge_{k \neq i} a_k \right\rangle,$$

and so by the Cauchy–Schwarz inequality and Lemma 5.3(iv) it follows that

$$\left| \left\langle x - y, \star \bigwedge_{k \neq i} a_k \right\rangle \right| \leq 4d\kappa\delta^{1+\beta} \left| \star \bigwedge_{k \neq i} a_k \right|.$$

Since $|\star \bigwedge_{k \neq i} a_k| \geq 1/2$ we conclude that $24d\kappa\delta^{1+\beta} \geq \delta^{\alpha_1}$. For a sufficiently small choice of c_d, this is our desired contradiction. □

Step IV: The conclusion via the discrete inequality

Using the decomposition in Step II,

$$\int_Q \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \, \mathrm{d}x = \sum_{\chi \in \{0,1\}^d} \sum_{n \in \mathbb{N}^d} \int_{P(n,\chi)} \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \, \mathrm{d}x.$$

By (32),

$$\left| \star \bigwedge_{k \neq i} a_k - e_i \right| \leq \frac{1}{10},$$

and thus each $P(n, \chi)$ is contained in an axis-parallel cube with sidelength equal to $2\delta^{\alpha_0}$.

The main term: $\chi = 0$. It follows that

$$\begin{aligned} \sum_{n \in \mathbb{N}^d} \int_{P(n,0)} \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \, \mathrm{d}x &\leq C(2\delta^{\alpha_0}, M) \sum_{n \in \mathbb{N}^d} \prod_{j=1}^m \left(\int_{B_j(P(n,0))} f_j \right)^{\frac{1}{m-1}} \\ &\leq C(2\delta^{\alpha_0}, M) \sum_{n \in \mathbb{N}^d} \prod_{j=1}^m F_j(\Pi_j n)^{\frac{1}{m-1}} \end{aligned}$$

where $F_j : \mathbb{N}^{d_j} \to [0, \infty)$ is given by

$$F_j(\ell) = \int_{B_j(T_j(\ell,0))} f_j.$$

Hence, by (7),

$$\sum_{n \in \mathbb{N}^d} \int_{P(n,0)} \prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}} \, \mathrm{d}x \leq C(2\delta^{\alpha_0}, M) \prod_{j=1}^m \|F_j\|_{\ell^1(\mathbb{N}^{d_j})}^{\frac{1}{m-1}}.$$

Consequently, by Proposition 5.2,

$$\sum_{n\in\mathbb{N}^d}\int_{P(n,0)}\prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}}\,\mathrm{d}x \leq C(2\delta^{\alpha_0},M)\prod_{j=1}^m\left(\int_{\mathbb{R}^{d_j}} f_j\right)^{\frac{1}{m-1}}. \tag{44}$$

The remaining terms: $\chi\neq 0$. To allow us to capitalise on the pigeonholing in Step II we need the following.

Lemma 5.4. *For each $1\leq i\leq d$ we have*

$$\mathrm{d}B_{\sigma(i)}(0)^{-1}\Sigma(i,J(i,n_i,1))\cap Q\subseteq B_{\sigma(i)}^{-1}\Sigma(i,[s_{n_i}^{(i)},s_{n_i}^{(i)}+\delta^{\alpha_1}])\cap Q.$$

Note here that $[s_{n_i}^{(i)},s_{n_i}^{(i)}+\delta^{\alpha_1}]$ is simply the "concentric triple" of $J(i,n_i,1)$.

Proof. Suppose $x\in Q$ satisfies $\mathrm{d}B_{\sigma(i)}(0)x\in\Sigma(i,J(i,n_i,1))$ so that

$$\mathrm{d}B_{\sigma(i)}(0)x=\mathrm{d}B_{\sigma(i)}(0)y+s\mathrm{d}B_{\sigma(i)}(0)\Big(\star\bigwedge_{k\neq i}a_k\Big) \tag{45}$$

for some $s\in[s_{n_i}^{(i)}+\frac{1}{3}\delta^{\alpha_1},s_{n_i}^{(i)}+\frac{2}{3}\delta^{\alpha_1}]$ and $y\in\langle\{a_k:k\neq i\}\rangle$, by (39) and (34). By Lemma 5.3(ii),

$$B_{\sigma(i)}(x)=\mathrm{d}B_{\sigma(i)}(0)x+\mathrm{d}B_{\sigma(i)}(0)(\Phi_{\sigma(i)}(x)-x). \tag{46}$$

Now $\Phi_{\sigma(i)}(x)-x=y'+s'\star\bigwedge_{k\neq i}a_k$ for some $s'\in\mathbb{R}$ and $y'\in\langle\{a_k:k\neq i\}\rangle$, and thus

$$\Big\langle\Phi_{\sigma(i)}(x)-x,\star\bigwedge_{k\neq i}a_k\Big\rangle=s'\Big|\star\bigwedge_{k\neq i}a_k\Big|^2.$$

Since $|\star\bigwedge_{k\neq i}a_k|\geq 1/2$, and by the Cauchy–Schwarz inequality and Lemma 5.3(iv), it follows that $|s'|\leq 4d\kappa\delta^{1+\beta}$. Now $s+s'\in[s_{n_i}^{(i)},s_{n_i}^{(i)}+\delta^{\alpha_1}]$ for a sufficiently small choice of c_d. Therefore, by (45) and (46), $B_{\sigma(i)}(x)\in\Sigma(i,[s_{n_i}^{(i)},s_{n_i}^{(i)}+\delta^{\alpha_1}])$ as required. □

Fix $\chi\neq 0$ and any i such that $\chi_i=1$. As above for the main term, it follows from (7) that

$$\sum_{n\in\mathbb{N}^d}\int_{P(n,\chi)}\prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}}\,\mathrm{d}x\leq C(2\delta^{\alpha_0},M)\prod_{j=1}^m\|F_j\|_{\ell^1(\mathbb{N}^{d_j})}^{\frac{1}{m-1}}$$

where now

$$F_j(\ell)=\int_{B_j(T_j(\ell,\chi))}f_j.$$

By Proposition 5.2 it follows that

$$\sum_{n\in\mathbb{N}^d}\int_{P(n,\chi)}\prod_{j=1}^m f_j(B_j(x))^{\frac{1}{m-1}}\,\mathrm{d}x\leq C(2\delta^{\alpha_0},M)\|F_{\sigma(i)}\|_{\ell^1(\mathbb{N}^{d_{\sigma(i)}})}^{\frac{1}{m-1}}\prod_{j\neq\sigma(i)}\left(\int_{\mathbb{R}^{d_j}}f_j\right)^{\frac{1}{m-1}}$$

and thus it suffices to show that

$$\|F_{\sigma(i)}\|_{\ell^1(\mathbb{N}^{d_{\sigma(i)}})}\leq 4\delta^{\alpha_1-\alpha_0}. \tag{47}$$

To see (47), first set $j = \sigma(i)$. Given the choice of notation in Step II, it is convenient to write

$$\|F_j\|_{\ell^1(\mathbb{N}^{d_j})} = \sum_{\ell \in \mathbb{N}^{d_j}} \int_{B_j(T_j(\ell,\chi))} f_j = \sum_{\substack{n_k: \\ k \in \mathcal{K}_j^c}} \int_{B_j(T_j(\Pi_j n,\chi))} f_j.$$

Now, since $i \in \mathcal{K}_j^c$ we may write

$$\|F_j\|_{\ell^1(\mathbb{N}^{d_j})} = \sum_{n_i} \sum_{\substack{n_k: \\ k \in \mathcal{K}_j^c \setminus \{i\}}} \int_{B_j(T_j(\Pi_j n,\chi))} f_j.$$

By Lemma 5.4 it follows that

$$\bigcup_{\substack{n_k: \\ k \in \mathcal{K}_j^c \setminus \{i\}}} B_j(T_j(\Pi_j n,\chi)) \subseteq \Sigma(i, [s_{n_i}^{(i)}, s_{n_i}^{(i)} + \delta^{\alpha_1}]) \cap Q.$$

Therefore, by Proposition 5.2 and (37),

$$\begin{aligned} \sum_{\substack{n_k: \\ k \in \mathcal{K}_j^c \setminus \{i\}}} \int_{B_j(T_j(\Pi_j n,\chi))} f_j &\leq \int_{\Sigma(i,[s_{n_i}^{(i)}, s_{n_i}^{(i)} + \delta^{\alpha_1}])} f_j \chi_Q \\ &\leq 4\delta^{\alpha_1 - \alpha_0} \int_{\Sigma(i,[s_{n_i-1}^{(i)} + \frac{1}{2}\delta^{\alpha_0}, s_{n_i-1}^{(i)} + \delta^{\alpha_0}])} f_j \chi_Q, \end{aligned}$$

from which (47) follows by summing in n_i and disjointness. This completes the proof of Lemma 5.1.

Remark 5.5. In Theorem 1.3, the smoothness assumption that each mapping B_j belongs to $C^{1,\beta}$ may be weakened. Suppose that each B_j is a C^1 submersion in a neighbourhood of x_0 such that the modulus of continuity of $\mathrm{d}B_j$, which we denote by $\omega_{\mathrm{d}B_j}$, satisfies

$$\omega_{\mathrm{d}B_j}(\delta) \leq \kappa \Omega(\delta),$$

where, for some $0 < \eta < 1$, Ω satisfies the summability condition

$$\sum_{r=0}^{\infty} \Omega(2^{-r})^{1-\eta} < \infty \tag{48}$$

and κ is a positive constant. Without significantly altering the above proof, one can show that Theorem 1.3 holds under such a smoothness hypothesis. Of course, Theorem 1.3 corresponds to $\Omega(\delta) = \delta^\beta$ with $\beta > 0$. It is of course easy to choose Ω satisfying $\delta^\beta = o(\Omega(\delta))$ as $\delta \to 0$ for all $\beta > 0$, and still satisfying (48); for example, $\Omega(\delta) = (\log 1/\delta)^{-2}$. Naturally, one pays for allowing a lower level of smoothness in the size of the neighbourhood on which the estimate in (11) holds.

6. Proof of Corollary 1.4

Without loss of generality we may suppose that there is a point a belonging to a sufficiently small neighbourhood of the origin in $(\mathbb{R}^{d-1})^{d-1}$ (depending on at most d, β, ε and κ) such that $F(a) = 0$; otherwise the neighbourhood V in the statement of the corollary could be chosen so that the left-hand side of (12) vanishes. By considering a translation taking a to the origin, we may suppose that $a = 0$. (Here we are using the uniformity claim relating to the neighbourhood V.)

Furthermore, we may assume that

$$\nabla_{u_j} F(0) = e_j, \tag{49}$$

the jth standard basis vector in $\mathbb{R}^{d-1}$, for each $1 \leq j \leq d-1$. We shall see that the full generality of Corollary 1.4 follows from this case by a change of variables.

Fix nonnegative $f_j \in L^{(d-1)'}(\mathbb{R}^{d-1})$, $1 \leq j \leq m$. We proceed in a similar way to the proof of Proposition 7 of [7]. Since $\partial_{(u_{d-1})_{d-1}} F(0) = 1$ it follows that there exists a neighbourhood W of the origin in $\mathbb{R}^{d(d-2)}$ and a mapping $\eta : W \to \mathbb{R}$ such that for each

$$x = (u_1, \ldots, u_{d-2}, (u_{d-1})_1, \ldots, (u_{d-1})_{d-2}) \in W$$

we have

$$F(x, \eta(x)) = 0. \tag{50}$$

The neighbourhood W depends only on β and κ, and the mapping η satisfies $\|\eta\|_{C^{1,\beta}} \leq \widetilde{\kappa}$ for some constant $\widetilde{\kappa}$ which depends only on d, β and κ. Our claims follow from the implicit function theorem in quantitative form. For completeness we have included an adequate version in Appendix B.

Let $B_j : W \to \mathbb{R}^{d-1}$ be given by

$$B_j(x) = (x_{(d-1)j-d+2}, \ldots, x_{(d-1)j})$$

for $1 \leq j \leq d-2$,

$$B_{d-1}(x) = (x_{(d-1)^2-d+2}, \ldots, x_{(d-1)^2-1}, \eta(x)),$$

and

$$B_d = B_1 + \cdots + B_{d-1}.$$

We claim that there exists a neighbourhood U of the origin, with $U \subset W$, depending only on d, β and κ, and a constant C depending on d, such that

$$\int_U \prod_{j=1}^{d} f_j(B_j(x))\, \mathrm{d}x \leq C \prod_{j=1}^{d} \|f_j\|_{(d-1)'}. \tag{51}$$

Since the subspaces $\ker \mathrm{d}B_1(0), \ldots, \ker \mathrm{d}B_d(0)$ are such that at least one pair has a nontrivial intersection, we cannot directly apply Theorem 1.3 to $\mathbf{B} = (B_j)$ in order to prove (51) (except in the special case $d = 3$ – see [7]). It is, however, possible to construct mappings $B_j^{\oplus} : \mathbb{R}^{d(d-2)} \to \mathbb{R}^{(d-1)(d-2)}$ for $1 \leq j \leq d$ in block form so that

$$\bigoplus_{j=1}^{d} \ker \mathrm{d}B_j^{\oplus}(0) = \mathbb{R}^{d(d-2)}. \tag{52}$$

We fix $1 \leq j \leq d$ and define $B_j^{\oplus} : \mathbb{R}^{d(d-2)} \to \mathbb{R}^{(d-1)(d-2)}$ as follows. Let $S^{(j)}$ be the $(d-2)$-tuple obtained by deleting j and $j+1 \pmod d$ from the d-tuple $(1, \ldots, d)$. [1] Then define $B_j^{\oplus} : \mathbb{R}^{d(d-2)} \to \mathbb{R}^{(d-1)(d-2)}$ by

$$B_j^{\oplus}(x) = (B_{S_1^{(j)}}(x), \ldots, B_{S_{d-2}^{(j)}}(x)).$$

[1] There is some freedom in the choice of the $S^{(j)}$; we only require that the components of each $S^{(j)}$ are distinct and that for each fixed $k \in \{1, \ldots, d\}$ there are exactly $d-2$ occurrences of k over all the components of $S^{(1)}, \ldots, S^{(d)}$.

To see that (52) holds, we compute the required kernels using the fact that

$$\ker \mathrm{d}B_j^{\oplus}(0) = \bigcap_{l=1}^{d-2} \ker \mathrm{d}B_{S_l^{(j)}}(0)$$

and using straightforward considerations. In order to write these down we write elements of $\mathbb{R}^{d(d-2)}$ as

$$(u_1, u_2, \dots, u_{d-3}, u_{d-2}; \widetilde{u}_{d-1})$$

where each $u_j \in \mathbb{R}^{d-1}$ and $\widetilde{u}_{d-1} \in \mathbb{R}^{d-2}$. Then, using (49) and (50), we have

$$\begin{aligned}
\ker \mathrm{d}B_1^{\oplus}(0) &= \{(u, -u, 0, 0, \dots, 0, 0, 0; 0) : u \in \langle e_1 - e_2\rangle^{\perp}\},\\
\ker \mathrm{d}B_2^{\oplus}(0) &= \{(0, u, -u, 0, \dots, 0, 0, 0; 0) : u \in \langle e_2 - e_3\rangle^{\perp}\},\\
&\vdots\\
\ker \mathrm{d}B_{d-3}^{\oplus}(0) &= \{(0, 0, 0, 0, \dots, 0, u, -u; 0) : u \in \langle e_{d-3} - e_{d-2}\rangle^{\perp}\},\\
\ker \mathrm{d}B_{d-2}^{\oplus}(0) &= \{(0, 0, 0, 0, \dots, 0, 0, u; (-u_1, \dots, -u_{d-2})) : u \in \langle e_{d-2} - e_{d-1}\rangle^{\perp}\},\\
\ker \mathrm{d}B_{d-1}^{\oplus}(0) &= \{(0, 0, 0, 0, \dots, 0, 0, 0; \widetilde{u}) : \widetilde{u} \in \mathbb{R}^{d-2}\},\\
\ker \mathrm{d}B_{d}^{\oplus}(0) &= \{(u, 0, 0, 0, \dots, 0, 0, 0; 0) : u \in \langle e_1\rangle^{\perp}\}.
\end{aligned}$$

An elementary calculation now shows that (52) holds.

Consequently, it follows from Theorem 1.3 that there exists a neighbourhood U of the origin, depending on d, β and κ, and a constant C depending on d, such that

$$\int_U \prod_{j=1}^{d} g_j(B_j^{\oplus}(x))\,\mathrm{d}x \le C \prod_{j=1}^{d} \|g_j\|_{d-1} \tag{53}$$

for all $g_j \in L^{d-1}(\mathbb{R}^{(d-1)(d-2)})$. Now, if $f_j^{\otimes} \in L^{d-1}(\mathbb{R}^{(d-1)(d-2)})$ is given by

$$f_j^{\otimes} = \bigotimes_{l=1}^{d-2} f_{S_l^{(j)}}^{1/(d-2)}$$

then by construction,

$$\int_U \prod_{j=1}^{d} f_j^{\otimes}(B_j^{\oplus}(x))\,\mathrm{d}x = \int_U \prod_{j=1}^{d} f_j(B_j(x))\,\mathrm{d}x$$

and

$$\prod_{j=1}^{d} \|f_j^{\otimes}\|_{d-1} = \prod_{j=1}^{d} \|f_j\|_{(d-1)'}.$$

Thus, (51) follows immediately from (53).

Finally, by the mean value theorem, it is easy to see that there is a neighbourhood V of the origin in $(\mathbb{R}^{d-1})^{d-1}$, depending only on d, β and κ, such that

$$\int_V f_1(u_1)\cdots f_{d-1}(u_{d-1}) f_d(u_1 + \cdots + u_{d-1})\delta(F(u))\,\mathrm{d}u \le 2\int_U \prod_{j=1}^{d} f_j(B_j(x))\,\mathrm{d}x.$$

Hence, whenever $\nabla_{u_j} F(0) = e_j$ and $\|F\|_{C^{1,\beta}} \leq \kappa$ there exists a neighbourhood V of the origin in $(\mathbb{R}^{d-1})^{d-1}$, depending only on d, β and κ, and a constant C depending only on d, such that

$$\int_V f_1(u_1) \cdots f_{d-1}(u_{d-1}) f_d(u_1 + \cdots + u_{d-1}) \delta(F(u)) \, \mathrm{d}u \leq C \prod_{j=1}^{d} \|f_j\|_{(d-1)'} \tag{54}$$

for all $f_j \in L^{(d-1)'}(\mathbb{R}^{d-1})$.

Now suppose that $F : (\mathbb{R}^{d-1})^{d-1} \to \mathbb{R}$ is such that $\|F\|_{C^{1,\beta}} \leq \kappa$ and

$$|\det(\nabla_{u_1} F(0), \ldots, \nabla_{u_{d-1}} F(0))| > \varepsilon. \tag{55}$$

Let $A^{\oplus}$ be the block diagonal $(d-1)^2 \times (d-1)^2$ matrix with $d-1$ copies of the matrix

$$A = (\nabla_{u_1} F(0), \ldots, \nabla_{u_{d-1}} F(0))^T$$

along the diagonal. Then, by the change of variables $u \mapsto A^{\oplus} u$ it follows that

$$\begin{aligned} &\int_V f_1(u_1) \cdots f_{d-1}(u_{d-1}) f_d(u_1 + \cdots + u_{d-1}) \delta(F(u)) \, \mathrm{d}u \\ &= |\det(A)|^{-(d-1)} \int_{A^{\oplus}(V)} \widetilde{f}_1(u_1) \cdots \widetilde{f}_{d-1}(u_{d-1}) \widetilde{f}_d(u_1 + \cdots + u_{d-1}) \delta(\widetilde{F}(u)) \, \mathrm{d}u \end{aligned}$$

where $\widetilde{f}_j = f_j \circ A^{-1}$ and $\widetilde{F} = F \circ (A^{\oplus})^{-1}$. The neighbourhood V of the origin shall be chosen momentarily.

By (55) it follows that the norm of A^{-1} is bounded above by a constant depending on only d, ε and κ. It follows that the same conclusion holds for the $C^{1,\beta}$ norm of $\widetilde{F}$. Since, by construction, $\nabla_{u_j} \widetilde{F}(0) = e_j$, and by (54), it follows that there exists a neighbourhood V, depending on only d, β, ε and κ, and a constant C depending only on d, such that

$$\int_{A^{\oplus}(V)} \widetilde{f}_1(u_1) \cdots \widetilde{f}_{d-1}(u_{d-1}) \widetilde{f}_d(u_1 + \cdots + u_{d-1}) \delta(\widetilde{F}(u)) \, \mathrm{d}u \leq C \prod_{j=1}^{d} \|\widetilde{f}_j\|_{(d-1)'}.$$

Therefore, by (55),

$$\begin{aligned} &\int_V f_1(u_1) \cdots f_{d-1}(u_{d-1}) f_d(u_1 + \cdots + u_{d-1}) \delta(F(u)) \, \mathrm{d}u \\ &\qquad \leq C |\det(A)|^{-1/(d-1)} \prod_{j=1}^{d} \|f_j\|_{(d-1)'} \leq C \varepsilon^{-1/(d-1)} \prod_{j=1}^{d} \|f_j\|_{(d-1)'}. \end{aligned}$$

This concludes the proof.

7. Applications to harmonic analysis

7.1. **Multilinear singular convolution inequalities.** Given three transversal and sufficiently regular hypersurfaces in $\mathbb{R}^3$, the convolution of two L^2 functions supported on the first and second hypersurface, respectively, restricts to a well-defined L^2 function on the third. Under a $C^{1,\beta}$ regularity hypothesis and further scaleable assumptions, this was proved by Bejenaru, Herr and Tataru in [4]. We note

that the inequality underlying this restriction phenomenon also follows from the nonlinear Loomis–Whitney inequality in [7]; the precise versions of the underlying inequalities differ in [4] and [7] because a stronger regularity assumption is made in [7] and a uniform transversality assumption is made in [4]. Here we show that natural higher dimensional analogues of this phenomenon may be deduced from Corollary 1.4.

For $d \geq 2$ and $1 \leq j \leq d$, let U_j be a compact subset of $\mathbb{R}^{d-1}$ and $\Sigma_j : U_j \to \mathbb{R}^d$ parametrise a $C^{1,\beta}$ codimension-one submanifold S_j of $\mathbb{R}^d$. Let the measure $\mathrm{d}\sigma_j$ on $\mathbb{R}^d$ supported on S_j be given by

$$\int_{\mathbb{R}^d} \psi(x)\,\mathrm{d}\sigma_j(x) = \int_{U_j} \psi(\Sigma_j(x'))\,\mathrm{d}x',$$

where ψ denotes an arbitrary Borel measurable function on $\mathbb{R}^d$.

Theorem 7.1. *Suppose that the submanifolds $S_1, \dots, S_d$ are transversal in a neighbourhood of the origin, $1 \leq q \leq \infty$ and $p' \leq (d-1)q'$. Then there exists a constant C such that*

$$\|f_1 \mathrm{d}\sigma_1 * \cdots * f_d \mathrm{d}\sigma_d\|_{L^q(\mathbb{R}^d)} \leq C \prod_{j=1}^{d} \|f_j\|_{L^p(\mathrm{d}\sigma_j)} \tag{56}$$

for all $f_j \in L^p(\mathrm{d}\sigma_j)$ with support in a sufficiently small neighbourhood of the origin.

Remark 7.2. (i) By Hölder's inequality it suffices to prove Theorem 7.1 when $p' = (d-1)q'$. One can also verify that the exponents in Theorem 7.1 are optimal, as may be seen by taking f_j to be the characteristic function of a small cap on S_j. As such examples illustrate, at this level of multilinearity, the transversality hypothesis prevents any additional curvature hypotheses on the submanifolds S_j from giving rise to further improvement. See [6] for further discussion of such matters.

(ii) Certain bilinear versions of Theorem 7.1 are well-known and discussed in detail in [21]. In particular, it follows from [22] that for transversal S_1 and S_2 (as above), which are smooth with nonvanishing gaussian curvature, there is a constant C for which

$$\|f_1 \mathrm{d}\sigma_1 * f_2 \mathrm{d}\sigma_2\|_{L^2(\mathbb{R}^d)} \leq C \|f_1\|_{L^{\frac{4d}{3d-2}}(\mathrm{d}\sigma_1)} \|f_2\|_{L^{\frac{4d}{3d-2}}(\mathrm{d}\sigma_2)}.$$

The exponent $\frac{4d}{3d-2}$ here is optimal given the L^2 norm on the left-hand side. The case $d = 3$ of this inequality was obtained previously in [17]. See for instance [9] for earlier manifestations of such inequalities.

(iii) In particular, when $q = \infty$ inequality (56) implies that

$$f_1 \mathrm{d}\sigma_1 * \cdots * f_d \mathrm{d}\sigma_d(0) \leq C \prod_{j=1}^{d} \|f_j\|_{L^{(d-1)'}(\mathrm{d}\sigma_j)}.$$

By duality, this is equivalent to the statement that, provided $f_1, \dots, f_{d-1}$ have support restricted to a sufficiently small fixed neighbourhood of the origin, then the multilinear operator

$$(f_1, \dots, f_{d-1}) \mapsto f_1 \mathrm{d}\sigma_1 * \cdots * f_{d-1} \mathrm{d}\sigma_{d-1}\Big|_{S_d}$$

is bounded from $L^{(d-1)'}(\mathrm{d}\sigma_1) \times \cdots \times L^{(d-1)'}(\mathrm{d}\sigma_{d-1})$ to $L^{d-1}(\mathrm{d}\sigma_d)$. For $d = 3$ this is a local variant of the result in [4].

(iv) The proof of Theorem 7.1 (below) leads to a stronger uniform statement, whereby the sizes of the constant C and neighbourhood of the origin may be taken to depend only on natural transversality and smoothness parameters. We omit the details of this.

Proof of Theorem 7.1. By multilinear interpolation and the trivial estimate

$$\|f_1 \mathrm{d}\sigma_1 * \cdots * f_d \mathrm{d}\sigma_d\|_{L^1(\mathbb{R}^d)} \leq \prod_{j=1}^{d} \|f_j\|_{L^1(\mathrm{d}\sigma_j)},$$

it suffices to prove Theorem 7.1 for $q = \infty$.

By considering a rotation in $\mathbb{R}^d$, we may assume without loss of generality that the submanifolds S_j are hypersurfaces; i.e. given by $\Sigma_j(x') = (x', \phi_j(x'))$ for $C^{1,\beta}$ functions $\phi_j : U_j \to \mathbb{R}$. Now, for f_j supported on S_j for each $1 \leq j \leq d$, and any $y \in \mathbb{R}^d$ we may write

$$\begin{aligned}
&f_1 \mathrm{d}\sigma_1 * \cdots * f_d \mathrm{d}\sigma_d(y) \\
&= \int_{(\mathbb{R}^d)^d} \prod_{j=1}^{d} f_j(x_j)\delta(x_{jd} - \phi_j(x'_j))\delta(x_1 + \cdots + x_d - y)\, \mathrm{d}x_1 \cdots \mathrm{d}x_d \\
&= \int_{U_1 \times \cdots \times U_d} \prod_{j=1}^{d} f_j(x'_j, \phi_j(x'_j))\delta(x'_1 + \cdots + x'_d - y')\delta(\phi_1(x'_1) + \cdots + \phi_d(x'_d) - y_d)\, \mathrm{d}x'_1 \cdots \mathrm{d}x'_d \\
&= \int_{U_1 \times \cdots \times U_d} \prod_{j=1}^{d} g_j(x'_j)\delta(x'_1 + \cdots + x'_d - y')\delta(\phi_1(x'_1) + \cdots + \phi_d(x'_d) - y_d)\, \mathrm{d}x'_1 \cdots \mathrm{d}x'_d \\
&= \int_{U_1 \times \cdots \times U_{d-1}} \prod_{j=1}^{d-1} g_j(x'_j)\widetilde{g}_d(x'_1 + \cdots + x'_{d-1})\delta(F(x'_1, \ldots, x'_{d-1}))\, \mathrm{d}x'_1 \cdots \mathrm{d}x'_{d-1}
\end{aligned}$$

where

$$g_j(x'_j) := f_j(x'_j, \phi_j(x'_j)), \quad \widetilde{g}_d(u) := g_d(y' - u)$$

and

$$F(x'_1, \ldots, x'_{d-1}) = \phi_1(x'_1) + \cdots + \phi_{d-1}(x'_{d-1}) + \phi_d(y' - (x'_1 + \cdots + x'_{d-1})) - y_d.$$

Observe that $F \in C^{1,\beta}$ uniformly in y belonging to a sufficiently small neighbourhood of the origin, and that by the transversality hypothesis (combined with the smoothness hypothesis),

$$\det(\nabla_{x'_1} F(0), \cdots, \nabla_{x'_{d-1}} F(0)) = \det \begin{pmatrix} 1 & \cdots & 1 & 1 \\ \nabla\phi_1(0) & \cdots & \nabla\phi_{d-1}(0) & \nabla\phi_d(y') \end{pmatrix} \neq 0$$

similarly uniformly. Theorem 7.1 now follows by Corollary 1.4. □

Estimates of the type (56) are intimately related to the multilinear restriction theory for the Fourier transform, to which we now turn.

7.2. **A multilinear Fourier extension inequality.** Very much as before, let U be a compact neighbourhood of the origin in $\mathbb{R}^{d-1}$ and $\Sigma : U \to \mathbb{R}^d$ parametrise a $C^{1,\beta}$ codimension-one submanifold S of $\mathbb{R}^d$. To the mapping Σ we associate the operator $\mathcal{E}$, given by

$$\mathcal{E}g(\xi) = \int_U g(x) e^{i\langle \xi, \Sigma(x)\rangle}\, \mathrm{d}x;$$

here $g \in L^1(U)$ and $\xi \in \mathbb{R}^d$. We note that the formal adjoint $\mathcal{E}^*$ is given by the restriction $\mathcal{E}^* f = \widehat{f} \circ \Sigma$, where $\widehat{\ }$ denotes the Fourier transform on $\mathbb{R}^d$. The operator $\mathcal{E}$ is thus referred to as an *adjoint Fourier restriction operator* or *Fourier extension operator*.

Suppose that we have d such extension operators $\mathcal{E}_1, \dots, \mathcal{E}_d$, associated with mappings $\Sigma_1 : U_1 \to \mathbb{R}^d, \dots, \Sigma_d : U_d \to \mathbb{R}^d$ and submanifolds $S_1, \dots, S_d$.

Conjecture 7.3 (Multilinear Restriction [7], [6])**.** *Suppose that the submanifolds $S_1, \dots, S_d$ are transversal in a neighbourhood of the origin, $q \geq \frac{2d}{d-1}$ and $p' \leq \frac{d-1}{d} q$. Then there exists a constant C for which*

$$\Big\| \prod_{j=1}^d \mathcal{E}_j g_j \Big\|_{L^{q/d}(\mathbb{R}^d)} \leq C \prod_{j=1}^d \|g_j\|_{L^p(U_j)} \tag{57}$$

for all $g_1, \dots, g_d$ supported in a sufficiently small neighbourhood of the origin.

Remark 7.4. Conjecture 7.3 implies Theorem 7.1. To see this we first observe that for any function f_j on S_j, $\widehat{f_j \mathrm{d}\sigma_j} = \mathcal{E}_j g_j$ where $g_j = f_j \circ \Sigma_j$. Now, if $2 \leq q \leq \infty$ and $p' = (d-1)q'$, then by the Hausdorff–Young inequality followed by Conjecture 7.3,

$$\begin{aligned} \|f_1 \mathrm{d}\sigma_1 * \cdots * f_d \mathrm{d}\sigma_d\|_{L^q(\mathbb{R}^d)} &\leq \Big\| \prod_{j=1}^d \mathcal{E}_j g_j \Big\|_{L^{q'}(\mathbb{R}^d)} \\ &\leq C \prod_{j=1}^d \|g_j\|_{L^p(U_j)} = C \prod_{j=1}^d \|f_j\|_{L^p(\mathrm{d}\sigma_j)}. \end{aligned}$$

This link was observed for $d = 3$ in [4].

In [6] a local form of Conjecture 7.3 was proved with an ε-loss; namely for each $\varepsilon > 0$ the above conjecture was obtained with (57) replaced by

$$\Big\| \prod_{j=1}^d \mathcal{E}_j g_j \Big\|_{L^{q/d}(B(0,R))} \leq C_\varepsilon R^\varepsilon \prod_{j=1}^d \|g_j\|_{L^p(U_j)}, \tag{58}$$

for all $R > 0$. In [7] the global estimate (57) was obtained for $d = 3$ and $q = 6$. Here we extend this global result to all dimensions.

Theorem 7.5. *If $S_1, \dots, S_d$ are transversal in a neighbourhood of the origin then there exists a constant C such that*

$$\Big\| \prod_{j=1}^d \mathcal{E}_j g_j \Big\|_{L^2(\mathbb{R}^d)} \leq C \prod_{j=1}^d \|g_j\|_{L^{\frac{2d-2}{2d-3}}(U_j)} \tag{59}$$

for all $g_1, \dots, g_d$ supported in a sufficiently small neighbourhood of the origin.

Proof. By Plancherel's Theorem, (59) is equivalent to the estimate

$$\|f_1 \mathrm{d}\sigma_1 * \cdots * f_d \mathrm{d}\sigma_d\|_{L^2(\mathbb{R}^d)} \leq C \prod_{j=1}^{d} \|f_j\|_{L^{\frac{2d-2}{2d-3}}(\mathrm{d}\sigma_j)},$$

where as before we are identifying f_j with g_j by $g_j = f_j \circ \Sigma_j$. Theorem 7.5 now follows immediately from Theorem 7.1. □

Remark 7.6. The Lebesgue exponent $\frac{2d-2}{2d-3}$ on the right-hand side of (59) is best-possible given the L^2 norm on the left. Again, at this level of multilinearity, the transversality hypothesis prevents any additional curvature hypotheses from giving rise to further improvement. See [6] for further discussion.

Appendix A. Proposition 1.1 implies Proposition 1.2

Assume that, for each $1 \leq j \leq m$, $B_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$ is a linear surjection and (4) holds. Let $\Pi_j : \mathbb{R}^d \to \mathbb{R}^{d_j}$ be given by (5) where d_j' is the dimension of $\ker B_j$.

Select *any* set of vectors $\{a_k : k \in \mathcal{K}_j\}$ forming an orthonormal basis for $\ker B_j$; that is, the orthogonal complement of the subspace spanned by the rows of B_j. By definition of the Hodge star and orthogonality considerations it follows that

$$\star X_j(B_j) = \|X_j(B_j)\|_{\Lambda^{d_j}(\mathbb{R}^d)} \bigwedge_{k \in \mathcal{K}_j} a_k. \tag{60}$$

Here, $\|\cdot\|_{\Lambda^{d_j}(\mathbb{R}^d)} : \Lambda^{d_j}(\mathbb{R}^d) \to [0,\infty)$ is the norm induced by the standard inner product $\langle \cdot, \cdot \rangle_{\Lambda^{d_j}(\mathbb{R}^d)} : \Lambda^{d_j}(\mathbb{R}^d) \times \Lambda^{d_j}(\mathbb{R}^d) \to \mathbb{R}$ given by

$$\langle u_1 \wedge \cdots \wedge u_{d_j}, v_1 \wedge \cdots \wedge v_{d_j} \rangle_{\Lambda^{d_j}(\mathbb{R}^d)} = \det(\langle u_k, v_\ell \rangle)_{1 \leq k, \ell \leq d_j}.$$

Let A be the $d \times d$ matrix whose ith column is equal to a_i for each $1 \leq i \leq d$ and let C_j be the $d_j \times d_j$ matrix given by

$$C_j = B_j A_j,$$

where A_j is the $d \times d_j$ matrix obtained by deleting from A the columns a_k for each $k \in \mathcal{K}_j$. Then, by construction,

$$\Pi_j = C_j^{-1} B_j A.$$

The matrices A and C_j are invertible by the hypothesis (4). Using A to change variables one obtains

$$\int_{\mathbb{R}^d} \prod_{j=1}^{m} f_j(B_j x)^{\frac{1}{m-1}} \, \mathrm{d}x = |\det(A)| \int_{\mathbb{R}^d} \prod_{j=1}^{m} \widetilde{f}_j(\Pi_j x)^{\frac{1}{m-1}} \, \mathrm{d}x,$$

where $\widetilde{f}_j = f_j \circ C_j$, $1 \leq j \leq m$. By Proposition 1.1 it follows that

$$\begin{aligned} \int_{\mathbb{R}^d} \prod_{j=1}^{m} f_j(B_j x)^{\frac{1}{m-1}} \, \mathrm{d}x &\leq |\det(A)| \prod_{j=1}^{m} \left(\int_{\mathbb{R}^{d_j}} \widetilde{f}_j \, \mathrm{d}x \right)^{\frac{1}{m-1}} \\ &= \frac{|\det(A)|}{\left(\prod_{j=1}^{m} |\det(C_j)| \right)^{\frac{1}{m-1}}} \prod_{j=1}^{m} \left(\int_{\mathbb{R}^{d_j}} f_j \right)^{\frac{1}{m-1}} \end{aligned}$$

and it remains to check that

$$\frac{|\det(A)|}{\left(\prod_{j=1}^m |\det(C_j)|\right)^{\frac{1}{m-1}}} = \left| \star \bigwedge_{j=1}^m \star X_j(B_j) \right|^{-\frac{1}{m-1}}. \tag{61}$$

To this end, note that

$$\star \bigwedge_{j=1}^m \star X_j(B_j) = \prod_{j=1}^m \|X_j(B_j)\|_{\Lambda^{d_j}(\mathbb{R}^d)} \star \bigwedge_{j=1}^m \bigwedge_{k\in\mathcal{K}_j} a_k$$

by (60) and therefore

$$\star \bigwedge_{j=1}^m \star X_j(B_j) = \det(A) \prod_{j=1}^m \|X_j(B_j)\|_{\Lambda^{d_j}(\mathbb{R}^d)} \tag{62}$$

since $\mathcal{K}_1, \dots, \mathcal{K}_m$ partitions $\{1, \dots, d\}$.

Again use (60) to write

$$\begin{aligned} |\det(C_j)| &= \left| \left\langle X_j(B_j), \bigwedge_{l\notin\mathcal{K}_j} a_l \right\rangle_{\Lambda^{d_j}(\mathbb{R}^d)} \right| \\ &= \|X_j(B_j)\|_{\Lambda^{d_j}(\mathbb{R}^d)} \left| \left\langle \star \left(\bigwedge_{k\in\mathcal{K}_j} a_k \right), \bigwedge_{l\notin\mathcal{K}_j} a_l \right\rangle_{\Lambda^{d_j}(\mathbb{R}^d)} \right| \end{aligned}$$

and therefore, by definition of the Hodge star,

$$|\det(C_j)| = \|X_j(B_j)\|_{\Lambda^{d_j}(\mathbb{R}^d)} |\det(A)|. \tag{63}$$

Now (61) follows from (62) and (63). This completes the reduction of Proposition 1.2 to Proposition 1.1.

Appendix B. A quantitative version of the implicit function theorem

We provide a quantitative version of the implicit function theorem for $C^{1,\beta}$ functions which we used in the proof of Proposition 1.4.

Below we use the notation $B(0,R)$ to denote the open euclidean ball centred at the origin with radius $R>0$ in either $\mathbb{R}^n$ or $\mathbb{R}$; the dimension of the ball will be clear from the context. Similarly, we denote by $\overline{B}(0,R)$ the closed euclidean ball centred at the origin with radius $R>0$.

Theorem B.1. *Suppose $n\in\mathbb{N}$ and $\beta,\kappa>0$ are given. Let $R_1,R_2>0$ be given by*

$$R_1 = \frac{1}{(100\kappa)^{1/\beta}} \min\left\{1, \frac{1}{10\kappa}\right\} \quad \text{and} \quad R_2 = \frac{1}{(100\kappa)^{1/\beta}}. \tag{64}$$

If $F:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$ is such that $\|F\|_{C^{1,\beta}}\le\kappa$, $F(0,0)=0$ and $\partial_{n+1}F(0,0)=1$ then there exists a function $\eta: B(0,R_1)\to\overline{B}(0,R_2)$ such that

$$F(x,\eta(x)) = 0 \quad \text{for each } x \text{ belonging to } B(0,R_1),$$

and a constant $\widetilde{\kappa}$, depending on at most n,β, and κ, such that $\|\eta\|_{C^{1,\beta}}\le\widetilde{\kappa}$.

Proof. The proof proceeds via a standard fixed point argument applied to the map $\Psi_x : \overline{B}(0,R_2) \to \mathbb{R}$ given by

$$\Psi_x(\eta) = \eta - F(x,\eta)$$

for fixed $x \in B(0,R_1)$. We shall prove that Ψ_x is a contraction which maps $\overline{B}(0,R_2)$ to itself.

Let $\Phi : (\mathbb{R}^n \times \mathbb{R})^2 \to \mathbb{R}$ be the map given by

$$\Phi((x_1,\eta_1),(x_2,\eta_2)) = \frac{F(x_2,\eta_2) - F(x_1,\eta_1) - dF(x_1,\eta_1)(x_2-x_1,\eta_2-\eta_1)}{|(x_2-x_1,\eta_2-\eta_1)|}$$

whenever $(x_1,\eta_1),(x_2,\eta_2) \in \mathbb{R}^n \times \mathbb{R}$ are distinct, and zero otherwise. By the mean value theorem and the fact that $\|F\|_{C^{1,\beta}} \leq \kappa$ it follows that Φ is everywhere continuous and

$$|\Phi((x_1,\eta_1),(x_2,\eta_2))| \leq 1/4 \quad \text{for all} \quad (x_j,\eta_j) \in \overline{B}(0,R_2) \times \overline{B}(0,R_2). \tag{65}$$

For each $\eta_1,\eta_2 \in \overline{B}(0,R_2)$ we have

$$\Psi_x(\eta_1) - \Psi_x(\eta_2) = (1 - \partial_{n+1}F(x,\eta_1))(\eta_1-\eta_2) + \Phi((x,\eta_1),(x,\eta_2))|\eta_1-\eta_2|.$$

Since $\partial_{n+1}F(0,0) = 1$ and $\|F\|_{C^{1,\beta}} \leq \kappa$ it follows that

$$|1 - \partial_{n+1}F(x,\eta)| \leq 1/4 \quad \text{whenever} \quad (x,\eta) \in \overline{B}(0,R_2) \times \overline{B}(0,R_2). \tag{66}$$

Hence, by (65) and (66) it follows that

$$|\Psi_x(\eta_1) - \Psi_x(\eta_2)| \leq \tfrac{1}{2}|\eta_1-\eta_2| \tag{67}$$

and Ψ_x is a contraction.

Now let $\eta \in \overline{B}(0,R_2)$. Using the hypothesis $\|F\|_{C^{1,\beta}} \leq \kappa$, along with (67) and (64), it follows that

$$|\Psi_x(\eta)| \leq |\Psi_x(\eta) - \Psi_x(0)| + |\Psi_x(0)| \leq R_2.$$

Hence $\Psi_x(\overline{B}(0,R_2)) \subseteq \overline{B}(0,R_2)$. By the Banach fixed point theorem, there exists a mapping $\eta : B(0,R_1) \to \overline{B}(0,R_2)$ such that $\Psi_x(\eta(x)) = \eta(x)$, or equivalently $F(x,\eta(x)) = 0$, for each $x \in B(0,R_1)$.

It remains to show that η belongs to $C^{1,\beta}$ and $\|\eta\|_{C^{1,\beta}} \leq \widetilde{\kappa}$ for some constant $\widetilde{\kappa}$ depending on at most n, β and κ. To see that η is differentiable, fix $x, h \in B(0,R_1)$ such that $x+h \in B(0,R_1)$. Since $F(x+h,\eta(x+h)) = F(x,\eta(x))$ it follows that

$$dF(x,\eta(x))(h,\eta(x+h)-\eta(x)) + \Phi((x,\eta(x)),(x+h,\eta(x+h)))|(h,\eta(x+h)-\eta(x))| = 0$$

and therefore

$$\begin{aligned}&\partial_{n+1}F(x,\eta(x))(\eta(x+h)-\eta(x))\\ &= -\langle \nabla_x F(x,\eta(x)),h\rangle - \Phi((x,\eta(x)),(x+h,\eta(x+h)))|(h,\eta(x+h)-\eta(x))|.\end{aligned}$$

Note that by (65) and (66) it follows that

$$|\eta(x+h) - \eta(x)| \leq C|h|$$

for some finite constant C independent of h. Moreover, Φ is continuous and vanishes along the diagonal. It follows that η is differentiable at x and

$$\nabla\eta(x) = -\frac{\nabla_x F(x,\eta(x))}{\partial_{n+1}F(x,\eta(x))}.$$

Using $\|F\|_{C^{1,\beta}} \leq \kappa$ and (66) one quickly obtains the inequality $\|\eta\|_{C^{1,\beta}} \leq \widetilde{\kappa}$ for some constant $\widetilde{\kappa}$ depending only on n, β and κ. □

References

[1] K. Ball, *Volumes of sections of cubes and related problems*, Geometric Aspects of Functional Analysis (J. Lindenstrauss, V. D. Milman, eds.) Springer Lecture Notes in Math. **1376** (1989), 251–260.

[2] F. Barthe, *The Brunn–Minkowski theorem and related geometric and functional inequalities*, International Congress of Mathematicians. Vol. II, 1529–1546, Eur. Math. Soc., Zürich, 2006.

[3] I. Bejenaru, S. Herr, J. Holmer, D. Tataru, *On the 2d Zakharov system with L^2 Schrödinger data*, Nonlinearity **22** (2009), 1063–1089.

[4] I. Bejenaru, S. Herr, D. Tataru, *A convolution estimate for two-dimensional hypersurfaces*, Rev. Mat. Iberoamericana (to appear), arXiv:0809.5091.

[5] J. Bennett, A. Carbery, M. Christ, T. Tao, *The Brascamp–Lieb inequalities: finiteness, structure and extremals*, Geom. Funct. Anal. **17** (2007), 1343–1415.

[6] J. Bennett, A. Carbery, T. Tao, *On the multilinear restriction and Kakeya conjectures*, Acta Math. **196** (2006), 261–302.

[7] J. Bennett, A. Carbery, J. Wright, *A nonlinear generalisation of the Loomis–Whitney inequality and applications*, Math. Res. Lett. **12** (2005), 443–457.

[8] J. Bourgain, *Besicovitch-type maximal operators and applications to Fourier analysis*, Geom. Funct. Anal. **22** (1991), 147–214.

[9] J. Bourgain, *On the restriction and multiplier problem in $\mathbb{R}^3$*, Lecture notes in Mathematics, **1469**, Springer Verlag, 1991.

[10] E. A. Carlen, E. H. Lieb, M. Loss, *A sharp analog of Young's inequality on S^N and related entropy inequalities*, Jour. Geom. Anal. **14** (2004), 487–520.

[11] M. Christ, *Convolution, curvature, and combinatorics: a case study*, Internat. Math. Res. Notices **19** (1998), 1033–1048.

[12] R. W. R. Darling, Differential forms and connections, Cambridge University Press, 1999.

[13] H. Finner, *A generalization of Hölder's inequality and some probability inequalities*, Ann. Probab. **20** (1992), 1893–1901.

[14] P. T. Gressman, *L^p-improving properties of averages on polynomial curves and related integral estimates*, Math. Res. Lett. **16** (2009), 971-989.

[15] E. H. Lieb, *Gaussian kernels have only Gaussian maximizers*, Invent. Math. **102** (1990), 179–208.

[16] L. H. Loomis and H. Whitney, *An inequality related to the isoperimetric inequality*, Bull. Amer. Math. Soc **55** (1949), 961–962.

[17] A. Moyua, A. Vargas and L. Vega, *Restriction theorems and maximal operators related to oscillatory integrals in $\mathbb{R}^3$*, Duke Math. J. **96** (1999), 547–574.

[18] B. Stovall, *L^p improving multilinear Radon-like transforms*, preprint.

[19] T. Tao, *Multilinear weighted convolution of L^2-functions, and applications to nonlinear dispersive equations*, Amer. J. Math. **123** (2001), 839–908.

[20] T. Tao, *A sharp bilinear restriction estimate for paraboloids*, Geom. Funct. Anal. **13** (2003), 1359–1384.

[21] T. Tao, *Recent progress on the Restriction conjecture*, Park City Proceedings, arXiv:math/0311181.

[22] T. Tao, A. Vargas and L. Vega, *A bilinear approach to the restriction and Kakeya conjectures*, J. Amer. Math. Soc. **11** (1998), 967–1000.

[23] T. Tao and J. Wright, *L^p improving bounds for averages along curves*, J. Amer. Math. Soc. **16** (2003), 605–638.

[24] T. H. Wolff, *A sharp bilinear cone restriction estimate*, Ann. of Math. **153** (2001), 661–698.

Jonathan Bennett and Neal Bez, School of Mathematics, The Watson Building, University of Birmingham, Edgbaston, Birmingham, B15 2TT, England.

E-mail address: J.Bennett@bham.ac.uk

E-mail address: N.Bez@bham.ac.uk